Google Trends as a Complementary Tool for New Car Sales Forecasting: A Cross-Country Comparison along the Customer Journey.

A case for India and South Korea

Master of Business Studies

In

Marketing

at Massey University Albany, New Zealand

Mina Sayedi (Student ID 15369140)

Supervisor: Vishnu Menon

December 2020

Abstract

This dissertation aims to evaluate factors responsible for different consumption patterns and customer's preferences for goods and services in India and South Korea. It further analyzes several model frameworks that reflect the different variants of conceptual ideas of different authors. It gives special importance to customer-oriented study with the aim of improving customer-oriented analysis. It also discusses the assessment of economic conditions of India and South Korea as purchasing power of customers is directly proportional to the state of the economy. It further discusses challenges involved in pricing of new car models. The focus is given to quantitative research to evaluate the correlation between the customer preference for a particular model of car and data of new car sales. This purpose is achieved using several linear regression models and cross-sectional studies. It has been observed that seasonality has little impact on car prices in India as per the data retrieved from Google Trend Index but in South Korea seasonality has significant impact on car prices. This dissertation also discusses the significance of Google Trend Index in both the countries. It further provides that use of studies used in this dissertation can serve as a valuable tool for predicting the demand for various car models in both the countries which would not be accurately provided by use of traditional approaches.

Table of Contents

List of Abbreviations

NCS	New Car Sales
SPSS	Statistical Package for Social Scientists
UN	United Nations
SIAM	Society of Indian Automobile Association
IMF	International Monetary Fund

List of Figures

List of Tables

1.0 Introduction

This chapter provides a brief background of the study, problem statement, research objective, research questions, hypothesis, and definitions of relevant concepts in the thesis.

1.1 Background of the study

The automotive industry is one of the essential industries in India and South Korea since these two countries are emerging economies (Vosen & Schmidt, 2011). Forecasting of sales within the automotive industry is essential because new cars in the present generation are built to either deliver or forecast. However, forecast usually leads to the effect of bullwhip since there are demand uncertainty and inaccuracies in forecasting (Karlgren et al., 2012). Although some cars are built for delivery, the issue of accurate forecasting is important for managers and decision-makers to efficiently and effectively allocate and plan their resources.

Google Trend is a tool employed by various companies in both developed and developing countries to collect updated and large sums of data on how consumers respond to new car sales (Carley et al., 2013). After realizing how important this development is, many automotive companies have paid great attention to Google Search and other social media platforms to develop ideas and implement positive and significant decisions (Karlgren et al., 2012; Liu, 2012). Typically, consumers spend much of their time searching for new and potential vehicle models. Kandaswami and Tiwar (2014) conducted a study and discovered that the average time spent by people on Google to search for cars of their choices was 10 hours a day. In Asian countries, including China, India, South Korea, and Japan, the number approached 70% (Kandaswami & Tiwar, 2014). Currently, the registered search engines like Google.com have a more extensive search platform for decision information on new car sales. Past authors have conducted research on the forecasting power of Google Trends on car sales. However, their results differ from time to time (Barreira et al., 2013; Geva et al., 2017). Geva et al. (2017) discovered that the forecast models for best car models sales is a combination of Google Trends' data, forum mention and forum sentiment. In their research, they found that Google Trends and forum sentiment have the same predictive power and volume. According to information provided by Trading Economics global macro models, the use of Google Trends as a tool for forecasting new car sales across India and South Korea, the customer journey of AIDA model is equally essential (Gensler et al., 2017). AIDA describes the four procedures that the customer considers during the process of buying. First A means attention. Attention is always attained through publishing activities and sales promotion, that may be done, for instance, in Google adverts. Letter I means interest that is portrayed in the customer's search behaviours it may be registered through Google Trends (Vosen & Schmidt, 2011). D means desire. The desire evolves in people's sentiment that may be registered through positive and negative Google expressions. The last letter 'A' means action, which is the customer's real purchasing power.

1.2 Problem statement

The process of decision making in companies is affected by the efficiency of forecasting methods. This is because it reduces the dependency on probability and acts as a scientific way to cope with external events (Waheed et al., 2014). The manufacturers of different cars are forced to be prepared for future and plan for the increased demand for new car models, especially within countries with a high developing rate. The demand efficiency across countries leads to several problems because approaches are needed to handle a wide range of data that affects the functioning of the automotive industry (Waheed et al., 2014). The automotive industry is composed of fast-changing needs of customers reflected by the dynamics of demand patterns that are observed as a threat to predict future requirements. Dharmani et al. (2015) discussed on the value of current statistical forecasting tools in enhancing the planning and decision making and having knowledge of the whole market. The value of suitable forecasts is encouraged by the Institute of Business Forecasting and Planning. This is because reduction in forecast error by only a point reduces the average saving rate. However many car manufacturers still base their forecast predictions on traditional and local tools that may not be able to manage their rising complexity.

The decision to purchase a car is determined by extensive information search by the customer, although there are other factors that are relevant in the new car buying process, including the national culture. However, various sources, such as professional dealerships and personal contacts, are considered before the decision of buying is reached. Much attention is given to the presence of a time lag between the attention given to a car model on the internet and the purchase decision made by the decision that requires to be emphasized in the sales predictions. Researchers like Artola et al. (2015) explain that the common use of internet in recent years has changed the performance of traditional activities including the way how financial transactions are conducted together with the online buying of products. In addition, Ernst and Young (2015) explored that customers put much time in online searching for new car trends before buying a car in comparison to any other product (Goel et al., 2010). The decision of purchasing is always made in the store. Thence, the assumption that the interests of people are revealed in their behaviours online, and in the main words they send to search engines. Prediction accuracy has been improved through various fields of application by adding web data within a model. However, internet data, especially Google Trends, is rarely applied in forecasting sales that might arise out of issues of reliability and validity concerned with big data. Nevertheless, accessibility and volume of internet data act as a suitable solution to solve the increasing slow development in approaches of forecasting and help decision-makers to solve complex problems within the changing environment in the automotive sector.

1.3 Research objective and research questions

Comparing the search engine-based prediction of new car sales in India and South Korea has extraordinary potential. India produces about 3.8 million passenger cars annually, and South Korea about 3.7 million passenger cars per year (statista, 2020). This implies that these two countries produce approximately equivalent amounts of passenger vehicles annually, and both countries are considered to have different cultures according to the habits of consumers. In this paper, quantitative research is employed to assess the relationship between the interests of people for a specific car model and the data of new car sales. This is done using various linear regression models and analyses that are cross-sectional.

The main goal of this thesis is to determine whether Google Trends tools are efficient forecasting for New Car Sales in India and South Korea. The Google Trends tool enables customized search query data that connect to a given time frame, country and relevant keywords. Although identification of the difference in internet searches across different countries is important in forecasting sales, less attention has been given to this issue. The paper additionally investigates the prophetic power of best pause in programme knowledge as some authors previously verified the Twitter data and different social media phenomenon. In this research, time lag refers to the time between product information research on the internet and the final decision made on purchases. The incorporation of time lag into the model helps in quantifying inaccurate prediction of the model in India and South Korea. The current high value of search engine data by determining new relationships and patterns significantly increases sales performance and forecasting on new car models. Besides, the practical value of free internet data (Google Trends tools) as a complementary tool to determine economic variables across countries is equally reviewed. This research study also intends to create search engine raw data awareness to other researchers and decision-maker and provide ways on how to handle issues concerned with internet data that are reliable and valid. Taking theoretical framework into consideration also improves the prediction value that enables in determining the difference in a time lag of the buying process that is related to the price of cars, the segment of vehicles and the respective model. This study emphasizes Google Trends as a complementary tool for forecasting new car sales and car purchasing decisions made by people in India and South Korea. This is intended to develop insights that allow positive response to demand volatility across countries. During this study, the specific research question and sub-questions below will be answered.

1.4 Specific Research questions

- What impact does the state of economy have on Google Trends' ability to predict new car model sales?
- What is the relationship between the state of the economy and car prices in India and South Korea?
- Does the seasonality presence impact the ability of Google Trends to predict volume as well as prices of cars in both countries?
- Does the average time lag length differ between high-priced and low-priced cars in both India and South Korea?

- How does the accurate prediction of new car model sales differ in South Korea and India?

1.5 Significance of the study

This paper encourages the use of theoretical frameworks together with statistical models to predict and gain results which are of benefit. Analysis of the customer journey and the state of the country together with the Google Trends analysis relates the research streams. The literature of Google Trends, research stream of marketing, and the cross-country research promote possibilities of new research. This thesis improves the validity and accuracy of prediction models, including internet data and thus, it contributes to literature prediction. In addition, the raw and unprocessed search engine data is unsatisfactory, and so this report calls for the attention of researchers to make adjustments within the dataset to minimize the effect of random observations (Jerath et al., 2014). The study assesses the cross-country differences in the application of new tool Google Trends' properties that extends the literature of Google Trends within India and South Korea for the very first time. This contribution paves the way for further research that may discover more about the differences found in cross-country comparison along the customer journey (Kinski, 2016). The study also provides information on Google Trends literature by investigating the search query predictions value on the level of products that result in proper applicability and forecasting. By executing more search queries, this research paper gives scholars who intend to conduct sales forecasting with advanced variables to reduce forecasting errors in their future studies. Furthermore, the determinations of a time lag between internet data generation and sales data occurrence encourage the use of engine data research and researchers on social media also become well versed with the phenomenon. The pros and cons of internet data within the field of academics and research are related to the previous survey that enables researchers to assess the analysis appropriately. The study further outlines some guidelines on how to enhance the reliability, validity and value of Google Trends data to improve the accuracy of economic variables prediction in the study.

The study gives decision-makers within the automotive industry an evaluated tool to use when making cross-country forecast and thus reduces the usage of outdated and traditional approaches while forecasting. Also, the Google Trends tool can be applied as an innovative and complementary tool to increase the demand for new car models by adding an updated and efficient search engine data within the forecasting model. The patterns identified in the internet data are specified for some car models and countries and acts as a catalyst variable to justify changes in the planning capacity of a company (Dharmani et al., 2015). The concept of prediction depends on lead time that is required by the customer to search for information before making the final purchase decision. The use of search engine data to understand consumers' interests brings about a competitive advantage with competitors until they recognize the relevance of internet value (Gensler et al., 2017). However, a comparison of search engine-based prediction for India and South Korea particularly helps decision-makers to assess and analyze which country and car model is less valuable and less important. The obtained results support the prediction of customer demand each day that lead to increased company savings and serve as a method of managing demand volatility (Gensler et al., 2017). Better forecasts for specific new car models support car producers to negotiate with their suppliers and customers. Suppliers tend to increase car prices depending on the customer's accuracy in capacity planning (Dharmani et al., 2015). Car model classification into small size, medium-size and large luxury vehicles helps in determining the differences in the time of the car buying process. From the marketing perspective of car producers, the relationship with customer journey has played a role of increasing the understanding of their customers in India and South Korea. The consideration of customers' interests is relevant during marketing activities to reach further customers, especially during most influential periods on decisions of buying. The emphasis of the economy's state enables practitioners to be aware of operating in the dynamic purchasing decisions of cars as it adds a positive trend in marketing campaigns. The analysis of Google Trends is suitable for every car model portfolio and various countries through carrying out a few changes in the data request of Google Trends. In other words, this paper helps decision-makers plan and implement decisions past the automotive industry to determine the benefits and the limitations of huge amounts of data that is created on the internet.

Chapter 2: Literature Review

2.0 Introduction

The chapter reviews the literature about the study variables. The study variables, in this case, are the production, sales, supply chain of new model cars and the overall automobile industry in selected countries of South Korea and India. Different studies from various researchers and publications or statements from Google Trends aid the development of reliable literature in this case. The literature review is a vital step in the development of research work. The effectiveness of the entire research project is deeply associated with literature studies. It guides the researcher towards its research objectives. Selection of the articles, journals and books are the key tools to develop a constructive research paper. According to Alvarado-Herrera et al. (2017), a review process can be categorized into five steps or three steps. However, most of the research studies prefer to apply five stage models for detailed coursework. The five-step methodology is interactive and comprehensive in nature. Wolfswinkel (1967) is the developer of the five-stage model. He believes that literature review is subject to the grounded theory method. The theory is comprised of five different steps First, find out relevant research papers; second, select pertinent literature; third, analyse the selected literature, and fourth, present findings. Finally, the fifth option is to enlighten the insight view of the study, including the conclusion.

It has been observed that scarcity of the proper research paper is a considerable concern for any researcher. The literature work is restricted to relevant academic journals, articles or textbooks available in the last decade. The current paper is focused on discussing the first three aspects of the five-stage model. Meanwhile, the study includes the proposed review technique in support of the key literature study. For instance, the study does not include the data related to social media, including Facebook or Instagram, while considering the field review developed by the Google Trends Applications. This technique is followed to narrow down the scope of the research work as well as the volume of the literature papers. Therefore, the researcher needs to invest comparatively less time in the literature study along with that appropriate dataset is required to precede the research project: the database EBSCO data, Google Books, Scopus and Google Scholar. In the field of the research work, Google Scholar plays an essential role as it vividly mentions the author’s name along with the publication. The detailed introduction of the research paper helps the researcher to find out the appropriate research material to leverage the research area. A set of keywords and their combinations are used to select the most useful research literature. Most predominant keywords are Google Trend Prediction, Google Trend Analysis, Customer Journey, Kotler’s Five Stage Model, Cultural Influences on Consumption, Purchasing decision method, Cultural differences, India, South Korea, Hofstede’s Cultural Dimension, Forecasting, Web Data Predictions and Predictive Analysis. With the help of these key search words, the researcher is expected to find out appropriate literature studies. This research process is an essential task to cite significant research articles for the current research paper. The third step is dedicated to select suitable articles as per the need of the study. The in-texting process is comprised of backward and forward (Bartels & Onwezen, 2014). A research paper is generally focused on the abstract, introduction and a summary portion of every paper. This provides a sufficient idea about the paper. In this way, the researcher can get an overall idea of the paper and hence, the analysis is expected to be relatively strong. The role of comparative analysis is significant to evaluate the significance of the research project. It explains how one research study is different from another study. An appropriate differentiation between the studies raises the importance of the research project. The current literature survey is developed on a number of scientific articles, books, internet resources and insight resources available on the online platform.

2.1 Systematic Literature Review

Currently, the automobile industry around the globe is entailed with various risks and opportunities in the growing mature markets of new model cars. More so, profitable growth has become harder to acquire because of several challenges arising from the supply side to the consumers. Economies, especially of India and South Korea, have many competitors, much capacity, as well as overlap and more redundancy (Smith et al., 2013). However, the automobile industry in identified countries has been challenged with the problem of rising global price wars.

2.1.1 Production of new Car models

The big manufacturers of cars involve the production facility in various markets to be produced mainly for exports to other markets. Most important to note is that a big automotive industry in the country does not depend on a single factory which exports the new model cars to other economies (Lassen et al., 2014). Besides, new car models are not identical in several markets of countries. Many cars have a similar brand, technical, or design features, but differ according to the country. Cars are different in various economies due to the nature of customer demands and pricing strategies. For instance, Indians have higher incomes than South Koreans, which means that the needs of the customers as well as affordable cars (Smith et al., 2013). Also, in South Korea, customers consider more space in cars, which in turn helps the new car models to be successful (Lassen et al., 2014). On the other note, small cars are the most popular in India. This is because India has limited parking spaces due to its large population (Stein & Ramaseshan, 2016). Besides, it may not be possible for a high-volume market to initiate big cars to each and every part of the world. Therefore, different car manufacturers and suppliers should research the preferences of customers in this changing market. Effective research helps the car manufacturers to produce car models which are preferred by consumers. However, completely second-hand cars are mostly renovated and produced in various countries, though in small volumes. Such a situation is a barrier to exporting cars in several countries since they are demanded on smaller volumes (Waheed et al., 2014). Basing on large markets, the model of sales is high, which in turn requires several companies to own their own plants to produce and supply the cars.

Due to the varying customer demand and sharp competition, the process of developing the product has been instrumental compared to product architecture (Wang & Park, 2015). However, the cycles for the products continue to shorten as many companies gain simultaneous engineering strategies. Also, some advancement in computer-aided engineering tools is acquired to do what physical prototypes do, as well as processes of testing. Different automakers in both South Korea and India are attributed to many responsibilities to design the products, as well as allowing the production base to change to strategic locations for reduced costs (Wang & Park, 2015). Besides, due to inadequate standardization, several tiers during procurement of cars are not reduced. Some domestic matters are supposed to be taken into consideration especially when the model is replicated with the new physical development of the product. Therefore, such modifications are considered as legal, regulatory procedures and liabilities (Waheed et al. 2014). Moreover, there are several technical advancements on modules, especially the functional units which are self-contained with some standardized interfaces to act as blocks for building in different car models or any other product. However, the supply chain process in the automobile industry can be reshaped using modularization. This is because such designs of the components are shifted gradually to top supplier companies (Wang & Park, 2015). The Shifting of the design components helps the suppliers to increase efficiency, as well as reducing costs. However, the survey by an international motor vehicle program found out that there is an elusive cost saving in new car model production (Waheed et al. 2014). The architecture of automobile retarded the modularization probability of success leads to a visible module advantage. However, different factors affect the supply chain to operate towards modularization process, which includes the efforts of automakers to change investment risks to the producers. This, therefore, increases the usage of information and communication technology amongst vehicles, where the consumers are likely to get interests towards in built-to-order cars. Many researchers found out that customers gauge production from the ‘model’ and ‘make’ viewpoint only in reality, but the production of automobiles is considered as dependent on supplier layers to alter outputs for real assembly.

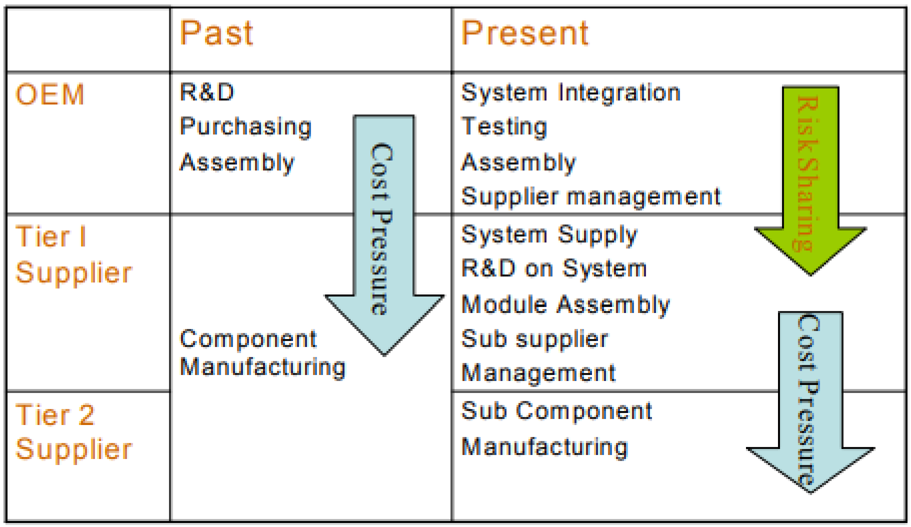
Besides, several automobile companies focus on the activities of assembling only whereas some are entailed by the vertical chains. The automobile industry is entailed with the great planning horizon, as well as high costs which are involved with the new car models (Waheed et al. 2014). The level of economies of scale in the automobile industry is much involved with technological flexibility to continue producing several models in a similar platform. Most important to note is that the core technological issues are fundamental in increasing competency and energy efficiency in both external and internal combustion engines. Such energy efficiency and competency in internal combustion engines are essential in combating vehicle weights, incorporation of high-tech safety features, among others (Wijnhoven & Plant, 2017). In addition, the standardization efforts are meant to reduce fuelling costs as well as maintaining the costs of production at lower measurements.

2.1.2 Supply chain of new model cars

Most of the automobile companies currently have adopted global strategies in their operations. It is clear that the growth of the transplant in 1990s accumulated the establishment of many competitors in the entire world (Wijnhoven & Bloemen, 2014). By putting much emphasis on interchangeable and common platforms, an automobile is likely to alter the easy and faster deployment of new strategies to achieve low costs from the products. This is essential to tailor vehicles to penetrate the tastes and preferences of the populace in the selected countries (Wijnhoven & Bloemen, 2014). However, proper product differentiation should be assured in order to attain a high-efficiency level, as well as managing the equity of the brand (Stein & Ramaseshan, 2016). It is also clear that great automakers are currently operating worldwide. Due to the established investments, the industries for automobiles have tried to replicate the structures of the supply chain, seeking for the suppliers to be allocated in new places especially near the plants (Shafi & Madhavaiah, 2013). Most important to note is that the auto industry supply chain has been varying over the years. Furthermore, the players in original equipment manufacturers around the globe have increased their focus on assembling and designing the operations and servicing after-sale markets. Therefore, after-sale markets are preferred because they deal with small numbers of big suppliers (Shafi & Madhavaiah, 2013).

Furthermore, the process of the supply chain has morphed in different sub-system integrators, commodity layers, and component makers. However, there is increased risk-sharing segregation which was termed as cost pressure by the car manufacturer (Wikström et al. 2016). It is clear that tier1 suppliers are engaged with the increasing risks from big automobile companies to shift the cost pressure to another supplier to focus on the sub-components production only. Figure 1 clearly demonstrates the process of shifting the cost pressure to other companies. Generally, vehicle suppliers are categorized into some specific groups (Wikström et al. 2016). These groups include; global standardized manufacturers systems, systems integrator, components specialist, and suppliers of raw materials (Shafi & Madhavaiah, 2013). Many vehicle companies including Toyota, Renault and Volkswagen think that some mono-suppliers' strategies, such as that of Ford, are bad but portraying large suppliers are considered better strategies. Ford makes sure that the suppliers are pushed to own the equipment and specific strategies to push risks involved with fluctuations of car models. Besides, suppliers must choose their schedules of amortization when issuing prices since the payback period for the investments is included in price (Shafi & Madhavaiah, 2013). Globally, companies like Renault and Volkswagen wish their suppliers to fully invest around their plants, as well as transferring information to indigenous players. Most important to note is that automobile companies advocate for quality standards, as well as affordable prices to develop the contracts between them and the suppliers. Generally, the value and length of the contract should be related to the targets of reducing prices that are supposed to be admitted by the suppliers (Shafi & Madhavaiah, 2013). On the other note, suppliers and assemblers of cars are also entitled to propose some methods which may yield similar results like the automobile companies.

Figure 1: Supply chain in the automobile industry



Source: Society for Indian Automobile Association (SIAM), (2004)

2.1.3 Pricing

Pricing of new car models is difficult since it is affected by fixed costs, technological aspects, and economies of scale. Also, consumer demand and competition among different companies are fundamental in the process of pricing (Hu et al. 2016). Recently, most of the companies in the automobile industry focus on pricing strategies to determine whether the company is to survive for a long period in the business. Among the pricing strategies, the price reduction is considered as an alternative approach to gain popularity and survive. Automobile companies require a lot of decisions at each phase of production and selling. The production and selling phase starts from supply chain and production, as well as management factors to aid negotiations with the dealers.

Most important to note is that price is considered as an important factor that aids the variability of sales for the goods and services in a significant way (Wu & Brynjolfsson, 2010). Several companies need effective policies to be practiced in an intelligent manner to yield better decisions (Somervuori & Ravaja, 2013). Besides, when prices are reduced, it does not necessarily increase profits but creates the popularity and reputation of the company. Therefore, the combination of quality preservation and product marketing decisions are essential when reducing prices to penetrate customers' tastes and preferences (Somervuori & Ravaja, 2013). Reduction of car prices attracts disloyal clients who do not value the company, but rather through price reduction in the brands. The companies' lifecycle helps them to gain many returns from the newly introduced car brands, which in turn reduces the costs of obtaining cars.

However, different companies use strategies of pricing entitled to various segmentation of products bearing in mind the changing expectations of the customers. Furthermore, automobile companies use innovative methods basically to increase profits without exploiting esteemed customers (Risius et al. 2015). Most important to note is that prices are adjusted in relation to the purchase volume, qualities, loyalty, profitability, and potential for development factor. When fixed costs are high, companies need to improvise with many and practical strategies or models to make proper decisions about total output (Wu & Brynjolfsson, 2010). It is also clear that many brands are aided by the level of economies of scale to put down production costs. In case when the carmakers offer prices to their brands depending on the average cost, several deviations are expected to alter many outputs in the short run to influence the pricing strategies without considering factor cost changes. It is also believed that fixed costs are subsets of total costs, which means that they can change into variable costs of production (Risius et al. 2015).

Among the car manufacturers, there is evidence that they face collusive pricings in their operations. However, the law about price disclosure in South Korea favoured several players to collaborate, which in turn led to collusion among themselves trying to fight for quality-adjusted prices. Moreover, big car companies were inhibited by the law in the struggle of providing price discounts. It is clear that the arguments provided by ability-motivation approaches or theories indicate that highly concentrated and stable environment markets with cooperative car producers create sustainability in the business. Research has shown that a market with current customers who are loyal, there is always motivation which helps them to aggressively create a base for new customers (Hu et al., 2017). Several firms believe that positive loyalty benefits from the customers enable them to survive in the business for a long period of time. Most important to note is that when the client's loyalty in the market increases, several gains are likely to increase as the market shares are altered by aggressive competition to maintain the profit margin. In general, car firms are entitled to have the motivation to cooperate with prices (Scott & Varian, 2013). The competition from foreign companies and eradication of local concentration has, therefore, changed the uniformity of prices. In summary, variation in pries helps the indigenous manufacturers in the specific countries to always coordinate with the counterpart firms during the process of setting prices for different car brands.

2.2 Buying Decision Models

2.2.1 Five-Stage Model of the Buying Process

Consumer behaviour is the key central idea of this model. Bartels and Onweze (2014) classify the buying decision into five different stages, including information search, evaluation of alternatives, purchasing strategy and purchasing behaviour. This model is known as Kotler's

Five Stage model. Every stage of the model is assigned to fulfil the different objectives of the research work. It has been observed that every stage of the model is interrelated to each other.

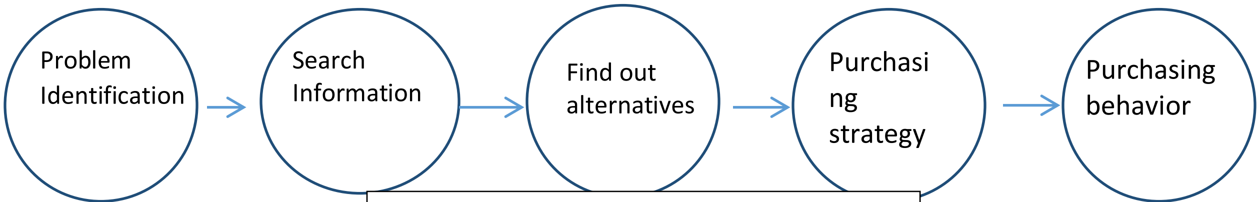


Figure 2: Kotler's Five Stage Model

Source: (as created by the author)

According to Kotler (2012), research problem needs to be recognized at the beginning of the research project. The model is marked as a traditional decision-making model. The model leads the customers to reach their final goals to buy their products and services. The consumers realize the need for purchasing goods and services step by step. Most of the consumers follow this strategy to optimize their consumption goals. Meanwhile, the intensity of the choice and preferences guides the customers to identify the objectives behind product choice. In this regard, another research theory by Engel, Kollet and Blackwell (1968) can be considered as the fundamental step of the five-stage model in order to analyse consumer behaviour. The model is comprised of two comprehensive stages. The first stage is dedicated to analysing the impacts of the marketing tools on consumer choice. These marketing tools include television, online social platforms and newspaper. After collection of the data, consumers enter into the second stage in the decision-making process. This second step analyses the external impacts of the influencing factors (Dhandra & Park, 2016). The information provided by the manufacturers for the customers plays the influencing role underlying the consideration for purchase. Branding of the product is the important parameter of the second stage. Branding reflects the standard of living of the consumers.

Meanwhile, ample information provides data on the alternative goods of the desired products. It has been observed that substitute goods play a significant role in determining the price of the desired product (Dhandra & Park, 2016). If many substitute products are available in the market, the consumer will get sufficient opportunity to transfer their choice of the products from one good to another good. For example, tea and coffee are substitute goods. In this case, the customer will prefer to buy coffee is the price of tea increases. A little change in the price of the goods brings a significant impact on the purchasing amount of the desired products. It is noticed that closed substitute goods are highly-priced elastic.

In this regard, some companies offer an experimental opportunity to realize the practical utility of the product. This practical experience genuinely helps the consumers to acknowledge the multiple characteristics of the product and services. This inclusion of different sources of information leverages the buying power of the consumers and reduces the risk factors related to the products. Contextually, the availability of the large size information builds a strong brand image for the company (Dwivedi, 2015). This ensures the company’s strong reputation in the product market and hence, the customers feel confident about purchasing of the desired product. The initial set of information reveals the key important characteristics of the product. The next step is counted as the third stage of the Kotler’s model. The latest modification of this model depicts that the multiple stages of information gathering depend on how a consumer collects the data. It is a non-singular approach. In this case, the diversity in information is highly expected.

Therefore, the customers need to wise while making their purchasing decision for the product. Both the negative and positive consequences need to be evaluated while finalizing the buying decision. The buyer should purchase the products when the benefits surpass the negative consequences related to the consumption of the products (Gupta & Sahu, 2015). The analysis needs to cover both the upside and downside aspects of the buying decision. Meanwhile, consumers often suffer while gathering information in an appropriate way. The post-purchase actions are greatly associated with pre-purchase actions. Keller notes that the customer's level of involvement is highly appreciable to get a successful outcome from the purchasing decision of the customers. The most effective utilization of consumer expenditure is only possible when the good is consumed in a proper way.

Moreover, mention that buying decision gets influenced by the purchasing pattern of other consumers, including friends and family members. Their choices and preferences are influential factors. The consumer desires to buy products which are not urgently required to meet the daily. The linear integration of the information gathering and analysis process is highly necessary to develop a link between the research objectives and desired outcomes. This requires additional adjustments tom recognize the future assessment for the next purchasing activity. This dynamic transmission helps the customers to deduce an effective purchasing strategy for the products and services.

2.3 Cultural Dimensions in India and South Korea

2.3.1 Cultural Difference and Consumption Behavior

The consumption pattern is significantly influenced by cultural values. Culture is the identifying parameter for every country. The people develop their choice and preference for the goods and services influenced by the culture. The people in India possess different taste and preferences for goods and services with respect to South Korea. Data reveals that the consumers in India expense more for spices than the South Korean customers. On the contrary, the consumption level of fast food is relatively higher in South Korea than India. Hur et al. (2015) urge that this difference in food consumption reflects the cultural influence on food habit. Meanwhile, it is observed that South Korean customers are more product conscious in respect of the Indian consumers. In South Korea, consumers spend considerable time to know the features and usefulness of the products. On the basis of limited availability of the information, the study evaluates the cultural impact on consumer behaviour through the Hofstede Model. According to the model developer, the cultural dimension is the influencing factor to determine the intensity of the difference in consumer behaviour. As South Korea is a technically advanced country, people have a growing tendency to buy electronic gadgets. This implies that product availability and consumer choice are intensely correlated across the two countries. As far as the model is concerned, the usage of technology can be mentioned to highlight the difference in the consumption pattern. South Korea is far way developed in the technology sector. On this account, the consumers of the country are accustomed to the advanced technology, which in turn increases the consumption of the electronic goods.

In the context of India, the traditional culture is the dominant feature. For this reason, spending on fast food is relatively less than the other sample country from Asia. The adoption and innovation of technology are strongly dependent on the cultural pattern of the country. Several studies have verified that culture has a significant influencing factor in determining the lifestyle of the people. Data asserts that average consumers in South Korea are wealthier than India. The great expense on the luxurious goods reveals that South Korean consumers are more sophisticated than India.

2.3.2 Hofstede's Cultural Dimensions

According to Hofstede, the model has five important dimensions from a cultural perspective. These five different aspects include individualism or collectivism (1), femininity/masculinity (2), uncertainty avoidance (3), power distance (4), long- or short-term coordination & (5) indulgence/restraint. The cumulative index number is assigned from 0 to 100. The study has been performed by the IBM employees. One hundred sixteen questionnaires are considered on the basis of seventy-two countries from 1967 to 1973. The model has an effective approach to the management and marketing research program. The model compares the mean value of the consumption pattern across the sample countries.

Along with that, the model attains to correlate the different cultural dimensions to justify the objective of the model. Despite the Hofstede model, several models have been developed to analyze the relative disparity in the cultural behaviour in respect of the individuals and organizations (Valaei et al. 2016). The next section briefly illustrates the significance of the five different models.

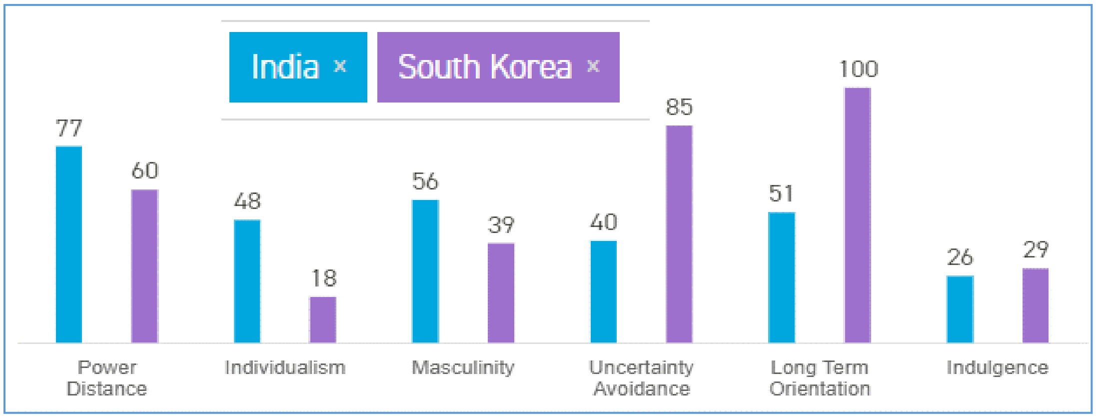


Figure 3: Hofstede's Cultural Dimension: India and South Korea

Source: (Hofstede-insights.com, 2020)

Considering the Figure 4, Hofstede's Cultural Dimensions in India and South Korea implies that there is a significant difference for uncertainty avoidance, long-term orientation, masculinity and individualism. The indices for power distance and indulgence are almost the same for both countries.

Power distance: This is regarded as a first dimension which explains that unequal power distribution makes some people powerful and the rest of the people gets controlled by those powerful persons. It is observed that powerful organization exhibits a large social gap between employees and employers in India as compared to South Korea. The organizational hierarchy triggers a great challenge to the productivity for India. The dominating attitude of the higher authority on the lower authority reflects the country's score of 77. The subordinates working under a boss felt a great distance in the workplace. This distance is an expected feature of a standard organization. This helps to maintain a cultural difference with respect to the designation. This is expected to provide a successful work environment for the employees. According to Yadav & Rahman (2017), most of the younger employees own relatively less power as compared to experienced employees. Purchasing tendency of high pricing car of the company owners clearly reflects the power distance among the society.

Individualism: Individualism stands for the moral stance, ideology and social outlook. It can be defined as the right of the individual to follow its self-realization. It is an evolving concept for the self-realization power of individuals (Howes et al., 2016). This leads to the betterment of the individuals and further for society. In this aspect, India scores 48, and it is considerably higher than South Korea's score of 18. This further verifies that the disparity is driven by the cultural choice of the inhabitants (Sreen et al. 2018). According to the model, the characteristics of the consumers are influenced by cultural practice. Along with that, rational thinking is another parameter which plays an effective role to justify the consumption pattern provided with the dimensional aspects of the cultural activity.

Collectivism: It is the alternative concept of the individualism. In terms of the collectivism, the integration among the people is the essential criterion. The people are cohesive in a group. The decision is taken through a collective method. Zhao et al. (2016) urge that individual perspective is not an important factor to assure the development of the society. The dependency on the other people strengthens the bonding among the people. The loyalty exchange is the prime characteristic of the collectivism theory. It develops an unquestioning loyalty between the customers and sellers. Therefore, it prefers high dependency than self-supporting consumption pattern. The self-interested group is not the fundamental concept of societal development (Fumo & Biswas, 2015). Considering the fact, both South Korea is strongly emphasizing on the collectivism mechanism as it scores lower in the individualism index. For example, the consumer choice in South Korea is majorly controlled by the central government and therefore, the consumer choice gets compromised.

Masculinity: This term referred to as the dominant values of society. The acquisition of wealth without considering the utility of others shows the impact of the masculinity. The dominating attitude is the reflection of the masculinity (Billing et al., 2014). For example, men are passionate about buying two-wheelers instead of beauty products. On the contrary, the female prefers to purchase cosmetics even if costs extravagant prices. Here, India exhibits greater gender disparity on the consumer choice and preferences for the goods and services as its score is 56 in terms of the masculinity index. In contrast, the influence of masculinity index is lower for South Korea. According to Dwivedi (2015), the country having high femininity score exhibits a better quality of life. People care for each other when the country is dominated by the feminism concept. The considerable difference between India and South Korea in terms of the masculinity results in the wide gap between the power concentrations. According to economists, masculinity indices score high due to the overwhelming wealth inequality. In the context of India, inequality in wage and wealth distribution is a significant concern for the economy (Le at al., 2016). On this account, it can be mentioned that the majority of the Indian families own a family car.

Uncertainty prevention: Second highest difference gets reflected by the uncertainty avoidance. In this case, South Korea has again scored a comparatively high level with 85, whereas India has scored 40. This indicates that the customers in South Korea are risk-averse than Indian customers. The people in South Korea are not willing to take risks while purchasing products from the commodity market. Uncertainty avoidance countries exhibit less intention to purchase online products provided with a 5% level of significance. According to Dhandra and Park (2016), online purchasing activities are a comparatively, less attractive option for risk-averse customers. This dimension highlights the willingness of the people to avoid the uncertainty related to the product. This uncertainty reflects the ambiguous nature of the consumers towards their purchasing decision for the products. In the context of India, most of the consumers exhibit ambiguous nature while making the purchasing decision as its score point is 40 in comparison with 85 of South Korea (Hoffmann et al., 2016). Here, the low per capita income of the Indian people can be one of the reasons behind this uncertain nature.

Long-term orientation: The score of long-term orientation has scored 100 for South Korea with respect to 51 in India. This difference implies that individualism is about two times than India. This score of indices implies that the people of South Korea are concerned about future consumption than Indian consumers. Economic resources are used in a more sustainable way when the country owns high long-term orientation score. This emphasizes the future perspective rather than on the current moment. This idea assists the sustainable development of the society and economy. The factors of the production are consumed in such a way that future generation will not be deprived of consuming all those products. For example, growing to emphasize the consumption of non - renewable products are aimed at maintaining the ecological balance in the environment. Henceforth, the consumption pattern needs to be changed to achieve a long-term goal. On the other, short-term orientation is subject to the personal stability. This aspect is developed from the perspective. Singh & Venugopal (2015) point out that most of the Indian consumers are reported to have less saving habits compared to South Korea. Therefore, it can be assured that India has less available resources to continue the investment process for the future.

2.3.3 Cultural Differences in India and South Korea

There are several numbers of approaches which can be incorporated to analyze the role of the cultural values on the consumption style. The seven-dimensional approaches of Trompenaars examine the cross-cultural disparity across the countries. This measurement is acknowledged as Tightness – Looseness method, which is used to determine the strengths of social rules (Treviño et al., 2019). It further examines the tolerance power of consumers. This latest concept not only is based on the previous five-dimensional model also is concerned about politics and history. If tightness is high in a country, the social norms are relatively rigid. This model has been developed comparing the cultural differences across thirty-three countries (Montoya & Horton, 2014). Contextually, the South Korean government is popular for its strict rules and regulation. From this point of view, the consumption pattern of South Korea is highly influenced by government intervention. On the contrary, India is a democratic country and consumers can freely express their choice and preferences for the goods and services (Gupta & Sahu, 2015). This cross-cultural dimension can be related to daily life-style. In case of strong life situation, people face limited options to express their choices due to income constraint or limited access to the products. On the contrary, weak-day life offers limited behavioural options to the customers in terms of comprehensive access to the products and services. This implies that individuals belong to the tight countries lead strong everyday life-style due to the strong intervention of the government.

However, several numbers of studies have found several loopholes of the Hofstede model. According to McSweeney, the small sample size is selected for the analysis of this dimensional model. Moreover, the numbers of questionnaires are not sufficient for model development. The concerned IBM survey was taken for two times in 1968 and 1973 for more than one thousand respondents (Jiang, 2017). The survey was conducted in Pakistan over a hundred employees of IBM. However, this survey raises questions on the validity and reliability of the model. The author also explains that every employee selected for the sample survey is the representative of Pakistan, and all of them are working under the same organization (Nyadzayo & Khajehzadeh, 2016). Therefore, focusing on a certain company is expected to influence the respondents. Apart from that, all the respondents are employs. The respondents can be unemployed, poor and elite class people (Beugelsdijk et al., 2015). They all are representatives of the entire nation. Considering the fact, the people from different section will contribute their significant values on the change in consumption pattern. Most importantly, people belong to a single country share the same values and preferences. In order to find out the degree of cultural influence, the relationship between national culture and national states needs to be examined. This important factor has not been emphasized during the development of the Hofstede model. Nonetheless, the deduced correlation model on the consumption pattern and culture was studied a long time ago. This study needs to be revised. The cultural difference depends on the change of time period. Therefore, the time frame is a vital aspect to observe the cultural changes with respect to the time period. The cultural difference needs to be treated as a quantified object to analyze the consequences of the changes in the consumption style (Dessart et al., 2016). In the case of Indian and South Korea, people have been experiencing robust changes in the consumption pattern. The growing consumption of the fast-food clearly denotes to the shift in the food intake pattern. Fast food is basically dominating the culture of the Western countries. Our sample Asian countries are traditionally fond of cooked foods. This shift in the taste of foods highlights the importance of the time frame.

2.4 Google Trends Predictions and Forecasting

2.4.1 Predictions and Forecasting

Trend analysis is a popular statistical model to predict future values. This provides a broad idea which products should be bought or sold from the future perspective. Prediction provides possible logics in order to finalize the purchasing decision. As far as the study is concerned, the prediction is always not able to construct accurate decision regarding the purchasing habit. However, the appropriate application of the theoretical framework results in the effective outcome of the predictions. In general, the outcome of the prediction seems like an approximate outcome. It provides an expected outcome of the experimental work. The predicted value may exceed the original outcome due to the impact of the selected variables. The application of the statistical tool is crucial to deduce the desired outcome with respect to objectives. Meanwhile, the data mining system is another important aspect of future observations. In terms of Çifci et al. (2016), a predictive model is the best applicable model for the empirical data. A new observation can be added to the original data while computing the predictive model. They have vividly illustrated the difference between explanations and predictions. The explanations highlight the importance of the task, and the latter one analyses the outcome of the model. The empirical analysis evaluates the significance of the predictive model. The predictive model is used to test the hypothesis in order to support the theoretical work. It illustrates the constructive framework as per the requirement of the study. The vast explanation of prediction is required to justify the impact of predictive modelling. It focuses on the future basis of the sample observation. However, there is a fine difference between the prediction and hypothesis. The hypothesis is taken before initiating the research work, whereas the prediction refers to the outcome after the calculation is done. The understanding of prediction depends on the theoretical approach towards the model framework. It is important to understand the role of the predictive analysis as it reduces the uncertainty related to research work.

Fritzsch et al. (2019) surveyed about whether Google trends are used to forecast and improve the sales and value of a certain product. The combination of standard time series and data about search query are fundamental in forecasting the sales of the product. Fritzsch et al. (2019) included the company's search volume, and brand-related keywords, which are provided by the Google trends. The keywords in Google trends help in forecasting the sales of the product since they act as new predictors. The study used time-series data from January 2015 to 2016 December, which helped them to get evidence about the essence of Google trends towards the performance of conventional models.

Mavragan et al. (2018) assessed the tools, methods, as well as statistical models used in research about Google trends. In the information overload era, the evolution of big data analytics is essential to solve and manage the data available to be interpreted. The study discussed the usage of web-based data about public health on the behaviour for individuals. Therefore, Google Trends is a common tool used for gathering information about fluctuating issues such as sales, public health issues, among others. Web-based behaviour previously was used in analyzing and monitoring the actual behaviour of the subject under investigation. Therefore, the stakeholders by using web-based behaviour also aid in predicting, assessing, as well as preventing the obstacles of production. Mavragan et al. (2018) used a systematic review to report critical analysis of the methods and statistical models about Google trends, which were carried out to intervene with health-related issues from 2006 to 2016. The methods used in their study followed the guidelines of Meta-analyses by selecting the individual term "Google trends" from the database of PubMed.

Forecasting is the cumulative research work on the available data set to predict the future performance of the selected variables. The estimated probabilistic value offers to illustrate the possibility of getting the desired outcome. The prediction analysis has an important role in measuring the future growth of the business, population and other important aspects of the economy. For instance, a prediction on the cycle sales can be like this. The sales growth can be hiked by 50%, whereas there is a 10% for which car sales growth can be declined for a certain period. Several numbers of factors play as a catalyst role in this predictive model. On this account, it is considered as one of the complex data structures under the forecast model. Referring to the time period, the forecast model can be classified into two categories, such as short, medium and long term forecast. The short term forecast is done for one to three months, the medium-term prediction is made for five to ten years, and finally, the long-term forecast is developed for five to ten years. According to the experts, the forecasting approach is relatively complex than prediction. The accuracy of the sample variables is essential to attain an accurate outcome. A proper justification of the forecast is derived from the work of the forecast model. The insight view of forecasting model assures that theoretical has the great potentiality to equip with the advanced predictive analysis (Zhang & Oczkowski, 2016). As far as the literature survey is concerned, both forecast and predictions have a different approach for new observations. Moreover, it distinguishes between out-of-sample incidents and forecasts for future prediction with the help of the supportive theory. Meanwhile, it can find out the impacts of new additional variables into the model. Predictions provide an outcome for the short-term instead of applying the probabilistic approach (Xu & Perron, 2014). The forecast technique underlying the current study incorporates the Google Trends Data. This data acts as an independent variable for the study. In general, both terms are used in an interchangeable way without considering the theoretical difference.

Mavragan et al. (2018) used 109 published papers by focusing on their correlations, seasonality, and visualizations features which are used to determine the category of issues under investigation. The study examined the performance of Google trends by using time series analysis, the definition of the key terms. Mavragan et al. (2018) observed that 23% of the studies used data from Google Trends to assess seasonality, whereas 40% of the studies used Google trends to get modelling and correlations. Online queries monitoring and assessments provide critical insight into the researchers' behaviours on how to make decisions about a specific matter under investigation.

2.4.2 Google Trends data as a tool of now-casting and forecasting

Internet data application is categorized into now-casting and forecasting measurements, which go beyond what other approaches, can perform. Now-casting, in this case, refers to the nature of prediction basing on current occurrences (Pieper, 2016). Now-casting helps the researchers to predict models which are based on short time period usually three weeks (Carrière-Swallow & Labbé, 2013). Now-casting helps the decision-makers to timely use time series always weekly and on a quarterly basis (Labbe & Swallow, 2013). The published data sometimes provide rough estimation, which, therefore, yield unreliable decision making (Mitchell, 2011). On the other note, data composition can be distinguished by using periods since some data do not appear at a specific period of time.

2.4.3 Shifts in the Forecasting Context

Over the last two decades, the introduction of the internet made commendable changes in the forecasting context. Globalization and comprehensive forecasting model have also led the forecasting system to the advanced level with the help of a software program. However, deficiency in technical knowledge is often acknowledged as drawbacks of the forecasting context. The inefficient training system, coupled with a knowledge gap misinterprets the outcome of the forecasting model (Zhang, Wei & Zhou, 2017). Altogether, it reduces the accuracy of the forecasting result. According to an article, forecasting analysis is unable to improve the sales prediction on a car on the basis of data over a decade. This inappropriate outcome calls for considerable improvement in the forecasting system (Mestre et al., 2015). On this account, new statistical tools are invented to obtain more accurate outcome related to the sample objective. In nowadays, the overwhelming presence of the internet raises the need for predictive analysis of online products. This growing concept is known as “online product buzz”. The model expresses the interest of consumers for online products and services. This sort of forecasting analysis gets structured on the basis of the review, blogs and online resources related to products available on the search engines. It has been observed that the accuracy of the predictive model has been improved by a significant level considering the online resources. All the online data helps in developing an advanced sales forecasting through online interviews or reviews written by the customers on the online platform (Huang & Kitani, 2014). It allows the customers to give reviews after the usage of the goods. Henceforth, online reviews are expected to give a significant contribution to enhancing the predictive study.

2.4.4 Judgment Forecasting as a Source of Inaccuracies

Several research papers have examined that sluggish improvement in the forecasting analysis results in the development of the complexities in the forecasting models. As a consequence of that, the desired outcome is not achieved as per the expected level. Lack of accuracy in the statistical tool misinterprets the statistical properties, which are the fundamentals of the model. Judgment is a fundamental step to judge human characteristics. The forecast will be appropriate if the experts do not get the available data as per the requirement of the model. Moreover, judgment is often reported to surpass the pragmatic outcome of the statistical analysis. Currently, more than 80% of commercial organizations are dependent on the forecast tool (Hendry & Mizon, 2014). Thereafter, the derived outcome gets adjusted in accordance with the objectives of the model. The degree of adjustment is subject to the knowledge of the market experts and the requirements of the company. Biasness and misinterpretations are the most common consequences of the inaccurate forecast model.

Optimism bias is regarded as the psychological tendency to influence the judicial process towards a definite direction. The insight view of business reflects the large involvement of the personal perspectives, such as personal benefits, clients' objectives and surrounding situations of the society and environment. The effect of optimism biases can be controlled through the selection of the accurate empirical database. Further, an effective comparison with a similar type of projects can help in fetching the desired outcome (de Carvalho et al., 2015). The expert finds out that political interference and organizational pressure misinterprets the survey outcomes, which are acknowledged as a part of strategic misrepresentation of the dataset. A survey conducted on Scandinavia explains that traffic forecasting outcomes get misinterpreted due to the presence of the incentive (Howes et al., 2016). There should be sufficient logic behind the initiation of the incentive policy in order to expand the capacity of the organization. The negotiation of the funding often misrepresents the outcome of the forecasting methods. Contextually, the reward policy can be introduced to encourage the analysts to calculate an accurate forecast. The managers are generally bribed to leather the original outcomes. The less number of inaccurate sources the high chance of transparent results. Considering the facts, the analysts always prefer to use transparent statistical tool to investigate the role of the likelihood functions and outliers (Lana et al., 2018). A lot of research works have been formed to highlight the role of the forecast methodologies to justify the adjustments of the forecast outcomes. It has been found that 99% of the forecast model uses statistical tools to verify the predictive outcome. The common approach used in the forecast model is the combination of the statistical inference and judgmental adjustments. Along with that, the respondent outcomes are incorporate as to cover all the influencing factors into the discussion. On average, it has been noted that a 10% error in the forecasting method corresponds to the human error. The use of Google trend application needs human interaction. The data extraction and interpretation are made by a human. On this account, a considerable portion of mistakes is expected in terms of human errors. These potential errors must be ignored while interoperating the forecasting analysis. Therefore, it is mandatory action needs to be taken to tackle the computation error related to the human research work (Brugler et al., 2015). Meanwhile, data extraction can be modified and controlled to avert the misinterpretation of the forecast module.

Prediction of economic variables and private consumption

A paper observes that Google develops the plan for a holiday trip on the basis of the travel destinations and the model of transport available for the customers. In the same way, the housing sales and motor vehicle shop are analyzed on the basis of the searching trends (Vaijayanthi & Shreenivasan, 2017). The paper explains that there is a strong correlation between the search frequency for several countries from India to South Korea and the data on the actual visitors come for the tours. This available data set helps the experts to set the trend for the tourists mitigating the uncertainty along with affordable expenditure (Fatma et al., 2016). The sales data reveals that car selling is highly influenced by online research work. Buyers do intensive research work before purchasing a car. The paper has constructed a regression model to find out a relationship between the cars and light truck sales of the basis of twenty-seven predetermined car makers (Peng et al., 2014). The available sales data act as a dependent variable in this tidy. The paper has the Google prediction to analyze the sale trend of the automobile market and retail sector. Trend data accumulate sales data on a monthly basis. The Google trend analysis makes the researchers acknowledged with the market trend. In this way, the management can structure the business plan for the future. It has been observed that the regression model under Google trend can improve the trend performance of the selected variables up to 20% with respect to those models which are not using the Google data (Dale et al., 2014). Google trend analysis measures the prediction using the mean absolute errors. Two different types of comparisons get developed, excluding the Google data as well as including the Google data.

Another study reveals the simplicity and accuracy of the web search data are useful to predict the house sales data. This will help to determine the market price of the South Korea housing sector (Youn& Cho, 2016). The report states that around 35% of the absolute mean error can be expected without using the Google data, while 5% of the absolute mean error can be expected using the Google data (Andreano et al. 2017). Further, Google trend analysis is reported to generate useful data for the emerging market condition. A set of research papers have pointed out that the merger of Google data with other available data resources develops multivariate data analysis (Lassoued & Hobbs, 2015). This predictive model helps in developing a long-term forecast up to 24 months which is ahead for the first time. In this context, a number of studies have criticized that the time horizon of the previous model is less than a few months. This analysis is developed on the basis of the data provided by the Federal Motor Transport

Authority in Germany. The data is comprised of both foreign and local car producers (Jindal et al., 2019). The data includes seasonally adjusted GDP data, unemployment rate and petrol price. The report concludes that macroeconomic variables are able to outperform on the basis of long-term. In the short-term, they have been failed to perform beyond the expectation. Nonetheless, parsimonious hardly considers the complexity of the long-term forecast with the help of Google data (Puška et al., 2018).

The relationship between unemployment and Google keywords is studied by other scholars in India. The study is limited to four predetermined sets of search keywords, including unemployment office, Monster. According to the scholars, there exists a strong association between the unemployment level and Google queries (Shen et al., 2018). Henceforth, many studies attain to predict the unemployment situation on the basis of key search words used on the internet. The similar exercise is done for the unemployment level among the French youth. It is concluded that the intensity of the job search gets doubled at the beginning of the first ten weeks.

Google Trends Predictions in Comparison to Traditional Survey

The analysis of Google Trends prediction as compared to that of a traditional survey in research is not a new phenomenon. Several scholars and researchers have touched on the aspect in previous times, and it is still a topic of wide concern in the world of academia and research. With the traditional methods of data collection, there is always the need for a massive input of resources. These resources come in terms of time, logistics, materials and human resources hence making the whole process very demanding, expensive and labour-intensive (Mellon, 2013). These logistical implications are among the reasons that posed the urge for researchers to look for an alternative means of data collection.

The internet has, in the recent past, proved to be a meaningful resource in collecting such data with the implication of a very minimal resource allocation to the process. Google trends are such a platform which proved to be a meaningful resource in the aspect of sourcing internet derived information and gathering data needed by scholars and researchers. It allows users to interact with internet-based data which is potentially useful and easily accessible upon demand.

The process of purchasing a specific product or brand is influenced by several factors, especially price, packaging, design, consumer knowledge and marketing strategy such as celebrity endorsements (Sun et al. 2015). Most developed countries like South Korea have got mature markets for retail products. Indians and Koreans are always keen to understand the brand, which is prevailing in the market (Mavragan et al. 2018). Do people pose different questions about the prevailing brands such as how will the product fulfil their wants? Research has shown that consumers' decision making on a specific product is based on the current information for both external and internal environment. The internal information about the product is gathered from advertisements, whereas external information emanates from market places or peers.

Decision making of purchasing a product is influenced by different brand selection procedures for specific products. Most important to note is that brand selection is based on the nature of the cohesiveness of the product. Therefore, car companies have to focus on the packaging of their new cars to affect the buying decisions of individuals (Zhang et al. 2017). Some of the areas on which the automobile companies need to improve are; consumer perception, customer information, marketing strategies (e.g., celebrity endorsement), and design.

2.5 State of the economy

2.5.1 Relationship between GDP and new car sales in India and South Korea

There exists a positive and direct correlation between the gross domestic products (GDP) of a country and its automotive industry. In this paper, GDP includes consumption, investment, and government expenditure within a specified time frame. According to Saberi (2018), the automotive industry is a knowledge-intensive and capital-intensive industry that is very important in the social-economic development of a country. In the current era, the industry is booming across the world, and so many countries have adopted the business of car production. However, the alignment of forces within the market of the automotive industry is dynamic (Geva et al. 2017). All economies growing at a faster rate, including India and South Korea, there is a need for efficient transportation means and faster mobility (Wikström et al. 2016). This forces the production and sale of more vehicles that, in turn, attract investment. In this situation, the automotive industry gains from the cycle. So (2020) explains that the manufacturers of several car brands like General Motors, Honda, Toyota and Ford experienced increased sales growth in Asia when countries in the West were hit by an economic crisis. He further explains that during 2019, the Gross Domestic Product (GDP) of South Korea amounted to 1642.38 billion US dollars. He derived this information from the World Bank data and projections from trading economies. South Korea's GDP, therefore, contributes a portion of 1.37 per cent of the World's total GDP. Currently, the automobile industry of South Korea is ranked the fifth largest producer for passenger cars across the world. Relating the this industry with the country's GDP, the automobile industry contributes 13 percent of the manufacturing plant, 12 percent of value-added, and as an industrial area, it generates over 12 percent employment opportunities for South Koreans (Barreira et al. 2013). The automotive sector is the central industry that influences all sectors under industrialization from materials like glass, steel, nonferrous metals to advertising, transportation, construction and financial services. For the past decades, the automotive industry of South Korea has gone through a revolution and grown, making it one of the economic growth indicators of the country.

For the case of India, the figures released in the dismissal April automobile sales by the Society of Indian Automobile Manufacturers (SIAM) indicate that whole vehicle production ecosystem in India could be a core growth indicator for the nation. SIAM explains that the production of vehicles within India reduced tremendously from the FY 2019 to FY 2020. They claim that the problem might have been caused by the presence of a global pandemic (Corona virus). This reduction in car production affected all the car sales segments (NetApplications.com, 2016). In the past, sales of cars in India have a weak relationship with the GDP growth due to factors like consumer sentiment.

Figure 5: India's economy and car sales

Source: Goyal, 2019

The figure above indicates the relationship between GDP growth and car sales growth and two-wheel sales growth. It shows that recently in 2019, the trend of car sales with GDP growth is directly proportional to the GDP growth compared to the two-wheel sales growth.

Figure 6: Relationship between Indian economy and commercial car sales

Source: Goyal, 2019

Figure 2 above portrays the impact of GDP growth on commercial vehicle sales in India from FY 2006 to FY 2019. Goyal (2019) discusses that the sales of commercial vehicles have a strong relationship with GDP growth. Goyal (2019) explains that although the annual car sales growth has been showing a positive trend, many car producing companies have been seen fighting for space on roads and this greatly reduces the growth of car sale.

2.6 Theoretical Framework

2.6.1 The customer Journey using the AIDA Model

AIDA model is developed to illustrate the influencing impact of the demographic transition of the consumers. The full form of AIDA is Attention (A), Interest (I), Desire (D) and Action (A). Continuing with the very initial stage, the product needs to be attracted by the customer, as stated by Lee & Hoffman (2015). Then the customer should focus on the interest of the products (Wijnhoven, & Plant, 2017). This makes the products attractive for the consumers. The awareness for the products gets evolved among the consumers. In this way, the consumers attain to focus on the consequences following the consumption of the products. Considerable awareness for the product results in the improvement of the purchasing amount for that particular product. Customer awareness develops when customers devote considerable time to collect product information (Pashootanizadeh & Khalilian, 2018). The simple approach of the model makes it more popular as compared to the five-stage model. The model explicitly illustrates in which way a company can improvise its product features in accordance with the growing consumer demand. The hierarchy-effects are the principle strategy to make both supplier and consumer awareness of the current market condition (Gnann et al., 2015). Apart from considering the direct involvement of the consumers, the theory also considers the purchasing activity of the consumers. The model emphasizes on the practical experience of the consumers. The complete framework of the model gets accomplished only after analyzing the reaction of the consumers.

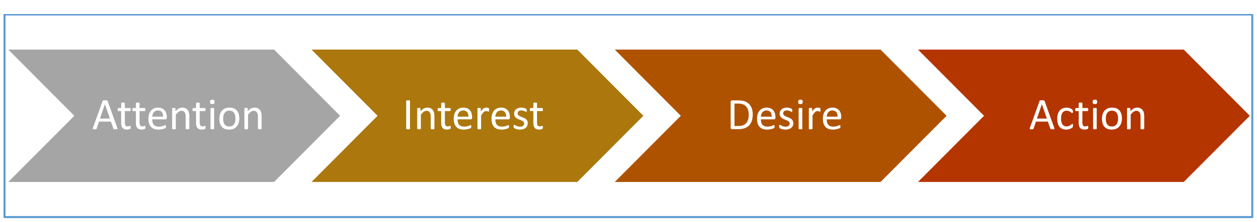


Figure 7: AIDA model

Source: (as created by the author)

Figure 2 exhibits a linear structure of four different stages of the AIDA model. The concept of the customer journey is conventional, combining with a number of alternative theories and marketing activities. It narrows down the broad concept of the buying decision towards a definite objective (Hassan et al., 2015). The model follows a customer-driven approach, and on this account, the theory is often called a customer journey model, as mentioned by Puzach & Hass (2014).

The AIDA model is the purchasing model that reveals the entire decision process of a customer in the form of information search, decisions and processing. The AIDA model contains four processes that include attention, interest, desire, action respectively that customers follow from the point of searching for a product to the point of deciding to buy it. Google Trends usually represent the stage of interest that real customers have for a product or service. Research from different scholars shows that search activities exhibit the intention of buying and also predict the behaviour of a consumer and also the sales of the products (Wijnhoven & Plant, 2017). However, some researches indicate that Google Trends are not suitable for showing the desire for a product since they do not portray negative or positive sentiments (Liu, 2012). Basing on the AIDA model, interest strongly indicates the intention of a customer to buy more than attention (Wijnhoven & Plant, 2017). On the other hand, desire has a stronger intention to buy compared to interest. Thus, there is always a stronger correlation between desire and sales compared to that of interest and attention. Furthermore, minimum time lags are expected to prevail between desire peak moments and the sales peak moments than for attention and interest

(Wijnhoven & Plant, 2017). Nevertheless, Google search has a strong influence on decisions made after the search results. This is very common when conditions are so indecisive at the stage of searching (Epstein & Robertson, 2015). Since high-priced purchases need more research, Google Trends tool can be a suitable forecasting tool for the volume of new sales. (Yang et al. 2015).

Nonetheless, the impulse buying behaviour of the consumer is highlighted by Hawkins Stern in 1962. The impulsive behaviour of the consumer makes him concerned about the product features when he comes to know about the product for the first time. Stern categorizes impulsive buying nature into four angles. First one is an impulsive purchase, the second one is reminding impulsive buying nature, the third one is suggestive impulsive buying nature, and the final one is planned impulsive buying nature. Impulsive buying nature provides ample opportunities for marketers to understand consumer behaviour.

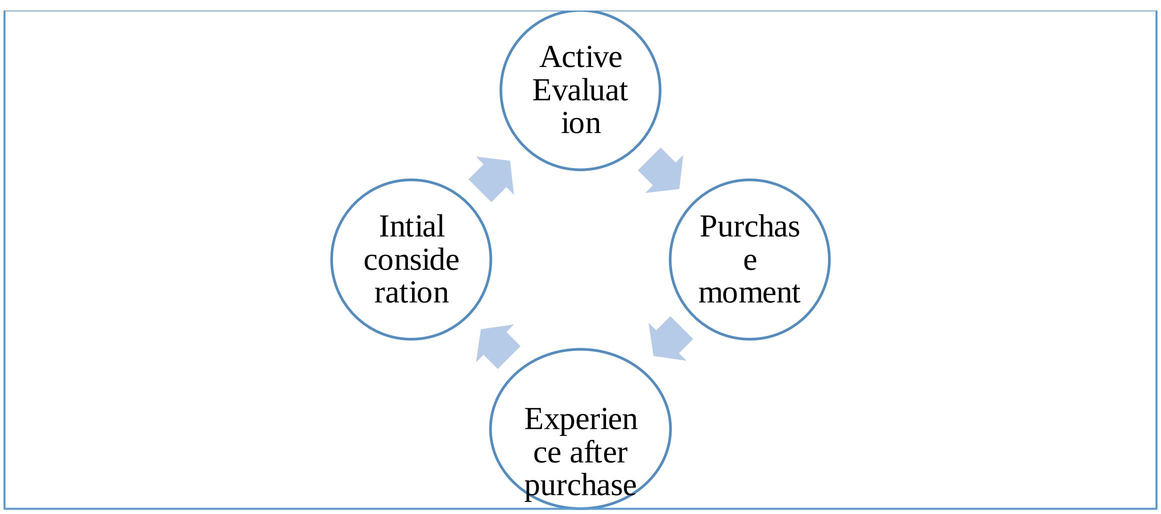


Figure 8: Customer Journey

Source: (as created by the author)

Brand consideration is the main stage of this Customer Journey model. A single product can be purchased by a different number of manufacturing companies. Initially, there are a potential number of customers for around a hundred numbers of car companies. The customers need to phase out car brands as per the requirements. This will narrow down the choices of car companies, and this will further develop a constructive comparative analysis among the sorted number of car brands. Henceforth, the evaluation phase is much dependent on the customer and choice of preferences. According to Rosenbaum et al. (2017), the evaluation process is driven by online research work. This active search process is developed through a number of selected brands. The next stage defines the moment of the purchase. This focuses on the attitude of the customers while they initiate their purchasing process. This process encourages customers to accelerate the purchasing activity. This is an integral part of the post-purchase activity. In this way, a strong brand loyalty gets formed among the customers. This reflects on the proactive recommendation for the purchasing decision. The proactive suggestion is to account for the important aspect of brand loyalty for the customers. It highlights the importance of brand loyalty in case of easing the development process of the customers. The cumulative actions of this research process help the customers to examine the presence of alternative brands (Diţoiu & Căruntu, 2014). In the view of modification, a considerable amount of information is required to justify the changes in the model. Lemon and Verhoef (2016) identify that the importance of the customer-driven purchasing activity is the key significant part of generating public reviews about the available commodities. Meanwhile, it supports the product evaluation strategy. It outlines the availability of substitute goods for the desired product. From the customer’s point of view, it is important to focus on the nature of the substitute goods and their purpose of services. The circular motion highlights the interdependency of the products.

Nymand-Andersen and Pantelidis (2018) conducted research on Google econometrics on new car sales forecasting and the requirements of big data quality. In their words, the authors said that "Big Data" is turning into a useful issue in the daily lives as it's a digital source of information that helps to cover a large population (Nymand-Andersen & Pantelidis, 2018). The information sources may help in the formulation of new datasets leading to proper insights into the use of digital behaviour and the present actions of economic agents. This, in turn, may help the promotion of early indicators of both financial and economic trend in activities. The authors based their paper on investigating the importance of Google search data in nowcasting of car sales within the euro area (Nymand-Andersen & Pantelidis, 2018). This acted as a major macroeconomic indicator and took the quality requirements for employing the new data sources as a tool for policymaking and sound decisions. The results revealed that although Google data may be capable of being predictive for now-casting car sales within the euro area, there was a need to make improvements in the sources of data requires so as to go further in experimental statistics (Nymand-Andersen & Pantelidis, 2018). The authors concluded that quality requirements could be achieved and the resulting improvements in knowledge and theory of interpreting big data help to re-shape people's thoughts in both socio-economic issues and behaviour.

Wijnhoven and Plant (2017) explored the importance of sentiment analysis and Google Trends data while forecasting car sales. They explain that previous authors had conducted on how the two techniques may be used in forecasting sales but did not base on goods like cars and so they covered that gap by forecasting sales for cars. In their study, the researchers employed over 500,000 social media posts for the eleven car models in the market of Dutch. The analysis was conducted with models of linear regression. In addition, the study relates the outcomes from the analysis with Google Trends predictive power (Wijnhoven & Plant, 2017). The findings show that sentiments of social media have less predictive power on car sales as well as Google Trends data and social mention volume significantly have more predictive power while suited in the prediction model. They conclude by providing a prediction model having time lags developed with a decision tree regression that could be employed by the automotive industry besides the traditional methods of forecasting (Wijnhoven & Plant, 2017).

Chapter 3: Methodology

3.0 Introduction

This chapter provides an overview of some methodologies that will be applied during analysis (Couper, 2013). It gives information on the concept of Google Trends and methods to be used and tests whether Google trends tool acts as a complementary tool for forecasting new car sales compared to traditional surveys. Also, this third chapter discusses issues of reliability and validity of Google Trends data (Hyndman & Athanasopoulos, 2014). It provides a research design for the study and discusses how the data measurement was carried out. Lastly, regression analysis is the most common method employed to explain the correlation between economic variables and the Internet data.

3.1 Google Trends Tool

The Internet is among the suitable information sources as trillions of queries are input into search engines every day. Google has several platform tools that help researchers to have access to information about Google data for different fields. Google web queries presented what is known as "Treasure House for web data mining" since the interests of people's concerns are mirrored

(Perrin & Duggan, 2015). A search query is defined as a complete term employed by the user of Google search. Google provides a query segmentation into groups like "Computer and electronics" and available data. According to Stephens-Davidowitz and Varian (2014), Google Trends creates numerous keywords through publishing values ranging from zero to one hundred. Therefore, the value of 100 shows the highest query volume for the predetermined group. In 2006, Google introduced a search analysis website for Google Trends. The commonly known tool gives information concerning potential individual searches, as indicated in the search volume index (Mislove et al. 2011). Therefore, Google never reports absolute numbers of data but rather provides relative popularity for the search. To get the index, data points of a query are divided by the total search volumes of a particular area, and the period is also put into consideration.

Google Search is essential in retrieving information, as well as keying in queries into specific search engines. Google is endowed with several tools that aid the researchers to use and access data from Google and use it for many purposes. It is a Google user who instructs the search engine with keywords to produce reliable results, which can be used to make decisions after analysis. The information or data from Google is segmented into various categories, especially such as automobile shopping and data is accessed depending on the time you wish to study on. Google trends always normalize the search volume of queries instead of displaying absolute searches to aid comparable data across industries or regions. Most important to note is that Google sums up all the data which is searched about a specific term or region within a period.

Ever since it was incepted in 2006, Google trends have served as a source of in-depth comparisons of information and data concerning different searches on varied topics across the board. Billions of search queries are entered into the Google search engine on a daily basis. In this case, a search query refers to the set of words that a user or searcher keys into the Google search bar to generate results or more information about the search as expected by the user.

When the searcher keys in a query, Google searches for the best possible response from the millions of data sets therein so as to respond to the searcher with the most relevant response to the keyed in a query, apart from the keywords used, this tool considers several other aspects also including the geographical location of the searcher so as to give out the best match. A search originating from India or South Korea will also want to concentrate its findings within the locality. This is to increase the relevancy of the search to the searcher based on where they are (Mellon, 2014).

It does not publish the outright number of searches for a particular keyword, but rather, it makes it familiar for the search volume of queries to create search terms comparable across regions (Google, 2015). Google Trends makes use of a fraction of searches for the specific term, that is, the keyword. It then, according to a given geographical location and a defined timeframe analyses the Google search outcome. A relative search volume is then assigned to the keyword, standardizing it from 0 to 100, where 100 represent the highest share of the term over a time series.

The figure below makes a comparison for the popularity of the search term “Toyota” between the searchers in India and those in South Korea, for the period starting May 2019 to May 2020.

The search popularity for the word Toyota in India is relatively higher than the same in South Korea all through the entire period of comparison.

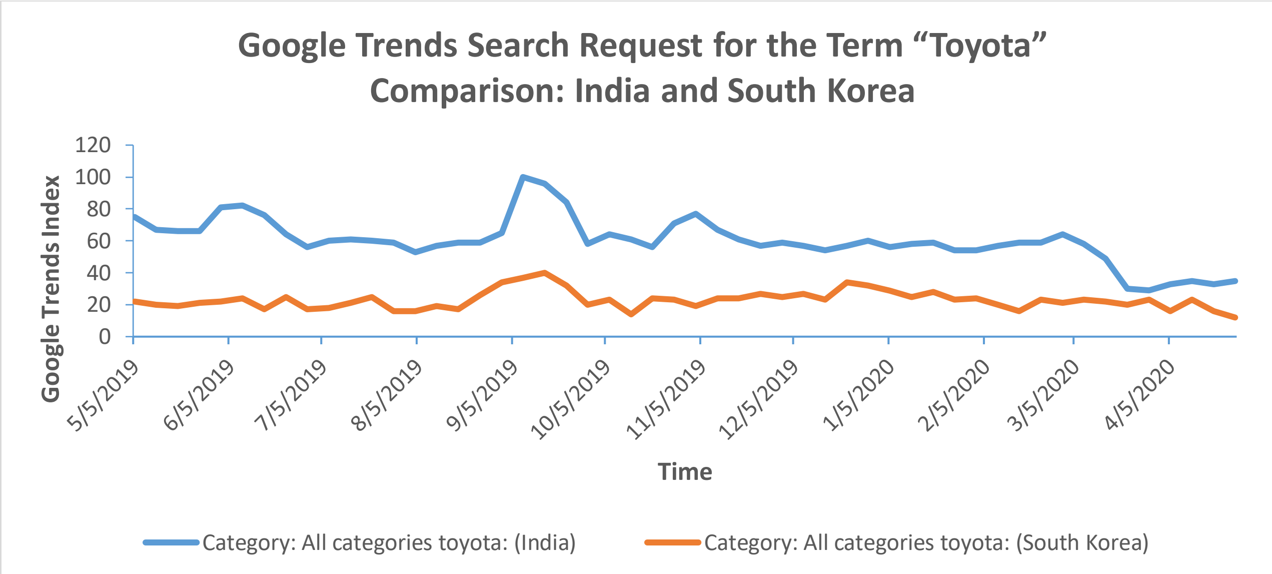


Fig 9: ***Google Trends Search Request for the Term "Toyota" comparison: India and South Korea.***

3.2 Research Design

Monthly data was collected from July 2019 to July 2020 for different car brands and car models. These were combined with the critical selected keywords from Google Ads. The focus was put on the manufacturers of car brands like Honda, Toyota, Nissan, Kia, Volkswagen, BMW, Audi, and Mercedes Benz as our research car brands since they are large sellers in both India and South Korea. To achieve the Google Trends data for all the eight brands, each car brand was given its best-selling car models that determine the rate of car sales in India and South Korea within a given period (Appendix 1). This was intended to attain a suitable sample size for search queries of new car models. To avoid including the searches that do not match with the automotive industry, the indices of search queries are grouped within the "Autos and vehicles". However, all the data collected was limited to India and South Korea between the periods of July 2019 to July 2020. The analysis in this paper was based on two databases having monthly car sales in India and South Korea, respectively. The data used was obtained from the analysis of the automotive industry website CarSaleBase where information on car sales used for the analysis was sought. To ensure that there was consistency in both datasets, we checked that all the required information including car prices, and Google Trends data for all car models available (Wu, and Brynjolfsson, 2015). The economic variables selected in this paper were selected systematically based on previous literature. The variables show changes in the price paid by customers to purchase a car, the impact of demand on car sales, and the state of both India and South Korea economies. To estimate the linear regression models, we employed data from July 2019 to July 2020.

3.3 Data measurement

Linear regression analysis was employed in this research since it is an appropriate tool to determine the relationship among various variables. It is also relevant during hypotheses testing and time lag implementation. The regression technique is most applicable for a set of analyses having internet data and a coefficient of determination R^2 to measure the quality of the prediction model (Choi & Varian, 2012). The linear regression analysis estimation provides the coefficients and can be applied in investigating the quality of implementing a time lag within the search engine data (Fantazziniand & Toktamysova, 2015). Therefore, the regression analysis method ensures that there exist comparability and reliability for the quality of the model in related studies.

In 2015, Fantazzini and Toktamysoya employed an advanced multivariate model having various variables such as GDP, seasonality, and state of the economy minus taking into account a single time lag. However, the advanced regression models base much on the quality of input variables, and there is an assumption that the value of making improvements on raw search engine data is well shown in less complicated models. When regressing, the SPSS software helped in testing the hypotheses tested. In the model, the Google Trends data is the independent variable well as the new car sales are the dependent variable (Gensler et al. 2017). After the preparation and arrangement of Google Trends data and recent car sales within the Excel file, data was successfully transferred into the Statistical Package for Social Scientists (SPSS).

Within the first regression hypothesis, a single linear regression for the whole new car sales dataset and the Google Trends dataset was used minus considering the country and the car model. While using the SPSS package, the steps followed were analysis, regression, and linear to predict the dataset relationship. The 95% confidence interval was taken while testing the relationship between Google Trends data and new car model sales while the other 5% ($p = 0.5$) acts as the significant level of the estimation and determine whether the relationship is significant at 95% confidence interval. The analysis also calculates the correlation coefficient R and R^2. This is intended to determine the power of prediction and direction of the relationship between the variables investigated. The value of R^2 ranges from 0% to 100% and determines the accuracy of the model. The positive correlation is revealed when R, R^2, $P< 0.05$ between the two coefficients of Google Trends data and new car sales.

3.4. The Reliability and Validity of Google Trends Data

The use of internet data needs specific attention to both validity and reliability to ensure that web data shows exactly what the researchers want. Babbie (2015) explains that to predict the relevance of any measurement, it should both valid and reliable. Nevertheless, Google Trends literature seems to have less attention to these two concepts. Reliability is related to the number of probability errors within the measurement and whether repetitive usage of a particular technique leads to similar results (Voortman, 2015). Therefore, reliability refers to measurement consistency (Babbie, 2015).

The study conducted by Choi and Varian (2009) acts as an appropriate example to repeatedly increase measurement as their measurements illustrates the necessity if information while using their methodology. This study also encourages the trends of evaluation or relative changes other than focusing on a particular date in time to illustrate the real picture of reliability due to the increased volatility of users of the internet. Google Trends data regulates the data that influences the tool of reliability (Google, 2015). Baker and Fradkin, (2013) made a study where they reduced the reliability through averaging data collected from different samples. They employed the data from Google Trends data for a similar keyword but in four different weeks intending to average the samples plus reduce internet data noise. Babbie (2015) suggests that the use of already established measures approved due to appropriateness that can improve the level of research reliability (Moore, 2015). Therefore, different researchers have compared the results of their analysis from Google Trends with surveys that are conventional to determine the new tool value (Geva et al. 2017). On the other side, Zhu in 2012 discovered that there were significant negative differences in the number of internet users compared to non-internet users as an indication of the search query's reliability.

Validity is defined as the degree to which a measure that is empirical properly and accurately reflects the true meaning of the concept discussed Babbie (2015). It is hard to prove the validity of the measurement, although numerous criteria may be employed to assess the accurateness. To test the relevancy of internet search data, Mellon (2014) differentiated between criterion-related validity, face validity, and content validity. Stephens-Davidowitz and Varian (2014) formulated guidelines on how to use the Google Trends tool to prevent ambiguity in terms of searching by appropriately selecting categories to get data that is valid. Zhu et al (2012) discuss the importance of the process of keyword selection to ensure research validity. The choosing of categories like "Automobile and vehicles" and limiting it to a single country raises the validity of content by removing topics that are separate from cars or a country of interest (Yang et al. 2015). The use of internet data is influenced by issues of reliability and validity. However, various techniques are present that reduce the effect of such threats and appropriately increase Google Trends data value (Wijnhoven & Bloemen, 2014). In this case, the reliability test was done by using Cronbach's alpha coefficient. The items used in the Cronbach's alpha coefficient test comprised of the time series data on the variables in the study.

Items	**Alpha Coefficient**
30	0.79

By using the Cronbach's alpha coefficient to test reliability and validity of the data used for analysis, the criterion is that when the coefficient is greater than 0.70, the instruments or data used are reliable and valid. In this case, by considering 30 items (car models), the alphas values is 0.79. This implies that the data used in the study was valid and reliable.

Reliability can be defined as the general consistency in a given measure. Generally, a measure is expected to give consistent results if subjected to constant conditions. As thus, one that will produce identical results with all the conditions constant is said to have high reliability. The internet has always been a key source of information on varied topics. Internet-derived information has always been recognized as a valuable tool for sourcing key information, data among other online resources. Google Trends, a Google Inc. portal, generates comparative data on different topics according to the search words keyed in.

When making use of information from the internet, one has to have enough confidence to acquire information that is both relevant and valid. There is a lot of corrupt and inauthentic data doing rounds on the internet, and this could pose to be a great impediment in drawing correct information, especially so if the searcher is not keen enough. Trends tend to normalize data that affect the tool's reliability (Zhu et al. 2012).

Some researchers had taken on to compare the trend of their Google trend analysis against results obtained from previous research that had employed the use of traditional survey methods to see the difference. However, no great differences were noticed to prove that the internet usage of Google trends is equally reliable in the research process. Validity refers to the extent as to which an idea, conclusion or measurement is legitimate as per its claim and is a likely inaccurate agreement to the real world. The validity of a measurement tool, in this case, Google trends, refers to the degree to which the tool can actually measure accurately whatever it claims to measure (Zhu et al. 2012). It is a general universal agreement that the concept of scientific validity uses statistical measures to address the nature of reality.

Despite its difficulty in proving, different criteria can be affected to check how appropriate it is as explained below;

i. **Face validity-**

Face validity tries to determine the extent to which a given test is independently perceived to cover a given purpose as it is supposed to measure. Face validity gives an indication of the relevance of a test as it appears to test whatever it is supposed to measure (Holden, 2010). For example, an IQ test is supposed to measure intelligence. The test would only be valid if it accurately tested the intelligence. Mellon said that face validity is defined as the choosing of a satisfactory group of keywords that tend to have face validity. Babbie (2015) added that a measure like that is valid on its face. Back in 2009, some researchers used the job database “monster” as a keyword that enforces face validity to determine the rate of unemployment as it had an assumption that job seekers apply these terms persistently than the employed people.

ii. Criterion-related validity

Criterion-related validity can also be referred to as predictive validity. It refers to the extent to which a measure is directly related to an outcome (Bäckström et al., 2014). An example here is the measure of a student’s score in a certain subject in high school and its relation to the same student’s score in the same unit in college. Criterion-related validity also called predictive validity, is defined as the level at which a measure connects to some external criterion (Babbie, 2015). The values of the written driver's test are connected with the records of driving for a person who serves as the external criterion to assess the validity of results written.

iii. Content validity

It is also known as logic validity. It refers to the extent as to which a measure will represent all the sides or will highlight or the dimensions of a given concept. Identification and exclusion of phrases or words that are not fully related to the topic under research will help in increasing the validity of the Google search data. In this case, keyword selection is very common (Heale & Twycross, 2015). The concept of content validity relies on the logical relationship between variables as a valid criterion. It is predicted that people with interest in purchasing new cars in a limited period are likely to use the internet to search for information on a specific car model which is concerned with the increase in the volume of the search query (Reyes et al. 2012). On the other side, content validity is defined as the amount of measure that covers different meanings present in a concept (Vosen & Schmidt, 2011). The concept of content validity determines the accuracy of the internet searches, and if the keywords evaluated measure exactly what they are supposed to (Mellon, 2014). Google search data for content validity may be increased by determining and removing terms that do not accept under investigation.

Also, the use of ambiguous words or those with more than one meaning could quickly reduce the chances of gaining a valid search result. For example, the search for “Toyota” and having filtered the results to a specific geographical location like in our case, we filtered it to specifically India and the second to South Korea. This will improve the validity of the search results as it will confine its findings to the specified search in the specified location only.

Generally, reliability and validity are a significant issue that faces the use of the internet and online data. However with correct usage and putting in place the key considerations as discussed in the topic, the issues are positively addressed, and more accurate data can be retrieved

3.5. Advantages and Disadvantages of Google Trends in Comparison to a Survey

Advantages of Google Trends in Survey

Google Trends is a free tool; it is available, easily accessible and free to use. All a user needs to have is the internet, a computer and relevant keywords. Google trends do not bear any heavy financial and human effort implications of physically going to the fields to collect data and later analyze it to make a finding. Google Trends gives already analyzed information, thus cutting down the time and resources that would have been used if the survey would have been done by use of the traditional methods.

The data acquired can be easily exported for use in research. Google Trends presents its data in already plotted charts that that can be easily exported and replotted on your work, coming along with all the relevant data that is encompassed within the charts provided. This is also beneficial as sometimes a researcher could make a single mistake while recording data, and in the end, come up with charts that are misleading or irrelevant.

Provision of updated data; there is no survey that can be updated with data coming in with every click of a second. The data provided by Google Trends is updated by the time and can be presented in whichever language the user wants and is available for, from and across all geographical borders. The value of the data presented by Google Trends is very high, and this is owed to its continuous data collection across the globe (Choi & Varian, 2012).

Relevance; Google is arguably the most and vastly used search engine in the world. Google Trends, being a product of Google, uses it (Google) as the basic source of its data. This then means that Google Trends data represent almost a complete fraction of all people with access to free internet.

Privacy Concerns; unlike other sources of internet data, the privacy concerns of individuals have been highly addressed hence reducing the risk in that the search queries cannot be traced down to individuals, a specific location or the IP address.

Google Trends generally provides updated internet data with adequate coverage and also a huge sample size making the results more precise, and this gives a researcher the advantage of working with a detailed analysis.

Demerits of Google Trends in Survey

Relativity; Google Trends can only provide relative numbers, and it’s almost impossible to attain the absolute numbers for the data presented. This prompts searchers to also perform a compared search so as to get results which can now be compared to the previous finding (Arora & Stuckler, 2019)

The question of Internet access; Google trends will mostly draw its data from that on the internet and compute the accurate findings from it. However, not all people have access to the internet. Worst so could be the case where the majority of the population in the works of the researcher does not have the access. This could obviously generate neither misleading report due to inaccurate data online stemming from the fact that the data attained is nor representing a random sample as compared to the actual state of affairs in reality. Data on Google Trends will only be from people who have access to the internet, that is, limited to areas that have a high internet penetration rate.

The interest of the searcher; there is generally no clarity in the context of the searches made. As far as it is assumed that interest is the main reason for the search to be done, there could be very many reasons, not even including interest, that will make an individual perform a search on Google. Goggle does not give any information about the person performing the search online, and thus it is impossible to profile them.

The ambiguity of Search Terms; there is a great need to do a thorough analysis of the context of the content. Several words have more than one meaning, and this could make it hard to establish if the finding is really for the specific finding that you are interested in or it includes the searches for the same word but with a different meaning from what you are interested in. although there is the provision to search according to categories, the big question remains to be how these categories were defined and formed and also the metrics used in matching searches together as relevant.

Inaccuracies due to a low search quota; search items whose search quota is extremely low might not be displayed. However, by the fact that there are very many searches for terms, even a very small number as such cannot be assumed, especially when dealing in absolute numbers. Nevertheless, such data cannot be attained as Google Trends does not give it.

Pre-determined search terms; data from Google Trends is highly dependent on the search terms used with the results being affected even by the smallest changes in the keywords. The ability of a searcher to use relevant keywords so as to attain the desired search result will really get to affect the quality of the internet data.

The table below represents Google Trends advantages and disadvantages as compared to those advantages and disadvantages of the traditional methods of survey, expressed in terms of their reliability and validity.

Google Trends Compared to Traditional Methods of Survey					
Google Trends	**Traditional Methods**	**Validity**		**Reliability**	
Relative data (has a single variable)	Absolute data (has multiple Variables)	-	✓	-	✓
Data from a large sample	Data is from a small sample	✓	✓	✓	-
It's not so vulnerable to desirability bias	It is high vulnerability to desirability bias	✓	-	✓	-

Easily replicable	Hard to replicate	✓	✓	✓	-
Uses predetermined keywords	Uses predetermined questions	-	-	-	-
Dependency: Internet User	Dependency: Respondents		-	-	-
Portrays short-term trends	Portrays both short and long-term trends	✓	✓	✓	✓
Highly volatile data	Constant Data	-	-	-	✓

Source: *Wang & Park, 2015*

Table 1: Google Trends advantages and disadvantages as compared to those advantages and disadvantages of the traditional methods of survey, expressed in terms of their reliability and validity

The next table below seeks to analyse the practicality and the simplicity of Google Trends as compared to the traditional methods of survey.

The Practicality and Simplicity of the Methods			
Google Trends	**Traditional Methods**		
Access to the data is free and over a short period of time.	Access to the data requires a lot of time, is labour-intensive and has a huge financial implication	✓	-

The setting can easily be changed	The setting is more rigid as it is hard to change	✓	-
There is always an ongoing collection of data passively	The collection of data is done once as is actively done.	✓	-

Source: Wang & Park, 2015

Table 2: ***The practicality and the simplicity of Google Trends as compared to the traditional methods of survey***

3.6 Regression Analysis in the Context of Internet Data

Regression analysis refers to a general statistical procedure for making estimates in the relationship between a dependent variable (outcome variable) and one or more than one independent variables.

This analysis gets to sort out mathematically by determining which of the variables actually has an impact (Sperandei, 2014).

It is used to provide answers to a set of guide questions, which are;

i. Which one of the factors matters most?
ii. Which one of the factors can we ignore?
iii. What is the interactivity of these factors with each other?
iv. What is our certainty about all of these factors?

These factors, in regression analysis, are the variables. There are two variables; the dependent variable which is the main factor that one tries to understand and there is the independent variable which are the other factors that might be having an impact on the dependent variable. (Roediger et al., 2001).

The simplest linear regression model consists of one independent variable as well as one dependent variable (Milton, 1986), as shown below:

$Y = a + bX + \epsilon$

Where Y =Dependent variable

X = Independent (explanatory) variable

a =Intercept

b =Slope

ϵ = Residual (error)

Here, the dependent variable Y is a linear function of the independent variable X with the slope b that indicates the rising-rate of the regression line once the dependent variable increases.

The best-fitting regression line then is the line that most accurately estimates the relationship that exists between the dependent and the independent variable (Guerard, 2013). A regression line is only aimed at giving estimates and thus does not predict the exact value for that given period.

After the relationship between the dependent and the independent variable has been determined, the independent variable can then is forecasted with the help of the known variable. In various application fields, daily, weekly, monthly, and annual collection and observation of data is important. Time series is referred to as "a sequence of values arranged by a time perimeter". The application fields are several rights from monthly values for economic purposes to sociology (Nguyen et al. 2015). Google Trends data is also known as the time series determined by location and time of the search query volumes entered. The objective of a time series is to provide the researchers with knowledge and reasons for past patterns of data series and values of the future forecast. Thus, analysis of time series as a method for quantitative forecasting can be used in case of the presence of historical data in a numerical format with an assumption that the past patterns will again happen in the future.

According to Frechtling (2011), regression analysis is a suitable method for determining the time-series correlation. Other scholars also emphasize the importance of regression analysis as a statistical tool to predict parameters, extrapolate trends, and determine quantitative data. The regression analysis model contains the dependent variable usually expressed as (X) and the independent variable expressed as (Y). Regression analysis aims to determine the relationship between new car sales and Google Trends tool in India and South Korea.

The most suitable regression line is that which estimates the relationship between the independent variable and the dependent variable very accurately (Guerard, 2013). The analysis of linear regression relies on the assumption that the relationship that exists between the dependent variable and the independent variable is relatively linear. Usually, the method of ordinary least squares is taken as the simplest estimation to determine how perfectly the line suits within the dataset investigated. The line having the least number of deviation squares from the real points of data is regarded as a suitable fitting line. The difference that exists between the estimated value and the real data point is revealed in the random or real error (e). Since the regression analysis aims at providing estimations, it never forecasts the real values for a specific period. However, after identification of the relationship between variables, Forecasting of the dependent variable like the number of sales becomes simpler with a known variable (Elshendy et al. 2017). Nevertheless, a particular criterion that confirms the value and the degree to which the model suits data guides the regression models.

There also exists a correlation coefficient *R* that measures the correlation between -1 (strong negative correlation) and +1 (strong positive correlation). The determination coefficient R^2explains the degree at which the variance independent variable (Y) is explained by the independent variable (X). The level p significance shows whether there is a statistically significant relationship within a given level of confidence. A significant level that is less than 0.05 (5%) is related to lower chances or probability (< 5%) that the relationship investigated is false (Wang and Park, 2015). Many scholars discovered a strong fit for their models where their independent variable was internet data to forecast the economic values (Shmueli, 2010).

Lassen et al. (2014) conducted research and estimated the linear regression model to determine the number of sales for the iPhone with several iPhone tweets. The results from their analysis were related to models that were well established by the investment bankers were the R^2 coefficient was 0.96. Furthermore, in 2012, Choi and Varian assessed the relationship between the sales of motor vehicle parts and the Google Trends Index. The results of the determination coefficient were 0.80. This indicated that 80% of the changes within the sales of motor vehicle parts (dependent variable) might be examined by changes in the Google Trends index (independent variable). Asur and Huberman (2010) made a study where they estimated linear regression and Twitter data, which acted as the independent variable to predict revenues of movies before releasing the real number with significant results. According to the good performance from prediction where $R^2 = 0.8$ and the correlation coefficient R of 90%, the scholars explained that attention given to a particular movie depends on its future status. Some researchers explain that regression analysis quality depends on the selected sample size and that the minimum data points to be used are at least 30.

A regression analysis has more than one independent variable is known as multiple regression (Sang & Bekhet, 2015). Some scholars reveal that complex models are normally error-prone and they support the application of simple models in conditions of the increased level of uncertainty. Armstrong and Green (2015) assessed the level of appropriateness of simple versus complex forecasting models in terms of being accurate. They discovered that the complex of forecasting models also affect the accuracy by raising the error of forecasting on an average of 27% (Vosen & Schmidt, 2011). However, regression analysis also has limitations such as the assumption that past patterns persist in the future (Epstein & Robertson. 2015). Outcomes that are not accurate may happen in case the patterns identified do not occur in the future and also in situations where unpredicted events never existed in the past data. Nevertheless, Choi and Varian (2012) showed that Google Trends data might be employed to reduce the effect of these "turning points" in economic time series. The regression analysis performance also depends on the sample size though the presence of past Google Trends data is limited (Choi & Varian, 2012). This threat may bring about missing relevant patterns that happen every decade on in long term cycles. Although it is impossible to estimate trends that are out of the dataset, considering internet data is important in determining current signals. Some researchers reveal that linear regression never includes patterns that are seasonal because of the linearity and dataset assumption. The use of the transformational methodology like time lags implementation within time series data may decrease the above limitation caused by the analysis of linear regression.

Chapter 4: Results and Presentation

The chapter depicts the results and elaborates more on the outputs of regression models, as well as cross-correlations about the stated hypotheses. Besides, the chapter assesses the accuracy prediction of Google trends in India and South Korea. Most important to note is that the chapter portrays the level at which time lag is affected by the car prices.

4.1 Data Analysis and presentation: State of the Economy and Car prices across Countries

The results in this section show the relationship between the state of economies and car prices. Most important to note is that the state of the economy was gauged using Gross Domestic Product for both countries in the study for the last 12 months. The Gross Domestic Product using web searches from both countries aided the process of regression with car price searches. The Google trends bout GDP were considered as dependent variables, whereas new car prices as independent variables. The study investigated 36 new car models from 8 car brands of Toyota, Nissan, Mercedes-Benz, BMW, Audi, Volkswagen, Kia, and Honda. The state of economies was gauged using the last 12 months for both countries. Therefore, the transformed values for all variables aided the regression analysis in this case. The variables, in this case, are supply, demand, and prices of new car models. Table 4.1.1 below shows the model summary between car prices and State of economic indicators within the selected period of the last 12 months. Basing on the multiple R (R=0.071), the results show that there is a weak positive relationship between the state of the economies and car prices. This implies that gross domestic product has less impact, which in turn impacts on car prices according to the results of the current study. Also, the results of R-square (0.005) show that 0.5% of the changes in the state of the economy are explained by the car price changes using the Google trend indices.

Table 4.1.1: Model Summary (All variables)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.071[a]	.005	-.015	18.492

Furthermore, table 4.1.2 depicts the results of regression coefficients about the study variables. The null hypothesis is stated as; *there is no relationship between the state of economy and new car model prices across the countries in the study*. The criterion is that when the Sig value is less than the level of significance, the variables are considered to be significant, thus rejecting the null hypothesis (Ghirvu, 2013). In this case, the sig value (0.618) is greater than the level of significance (0.05), implying that the null hypothesis is accepted. When the null is accepted, it implies that there is no relationship between the state of the economy and new car prices among the variables in the selected period.

Table 4.1.2. Regression Coefficients Results

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	9.479	3.249		2.918	.005
	Transformed variables	.074	.148	.071	.503	.618

Second Hypothesis: *The prediction accuracy of Google Trends is higher in India about new car models than in South Korea.* The transformed variables, in this case, are; car supply, demand, and new car model prices.

Table: 4.1.3: The accuracy rate of selected car models for both India and South Korea

Car Model	**Linear Regression Analysis without ‘Time lag’**	
	India	***South Korea***
	R-square	***R-Square***
Liva Dual Tone	0.023	0.013
Limited Edition	0.369	0.210
Etios Liva	0.492	0.546
Dual Tone Liva	0.507	0.554

Platinum Etios	0.075	0.113
Etios Cross	0.421	0.473
Glanza	0.701	0.541
Yaris	0.034	0.014
Innova Touring Sport	0.397	0.201
C-Class Cabriolet	0.038	0.092
Amaze	0.507	0.500
Jazz	0.353	0.310
BR-V	0.019	0.014
Micra Active	0.682	0.651
Kicks	0.019	0.008
Micra	0.030	0.043
Seltos	0.032	0.085
New Polo	0.508	0.490
New Vento	0.329	0.301
Brio	0.482	0.407
Polo GT	0.564	0.535
Audi A6	0.437	0.396
Audi A8	0.621	0.580
BMW X1	0.352	0.330
BMW 3 Series	0.675	0.593
BMW X6	0.679	0.632
CLS Coupe	0.002	0.017
Terrano	0.050	0.421
G-Class	0.329	0.275
Mercedes-AMG C 43 Coupe	0.504	0.501

The table above shows the comparison between India's and South Korea's predictability accuracy rate of selected new car models. The comparison is aided by the R-square values, which show the difference in prediction accuracy. From the selected new car models from both countries, it is evident that India has got higher prediction accuracy compared to South Korea due to a high predictability level by R-square values. Therefore, the hypothesis, in this case, is rejected.

However, table 4.1.4 shows the analysis of Google trends for the selected car models for both India and South Korea. The selection criteria covered 83.3% of the total car models in India and

South Korea. Under the p-value columns, the significant car models are highlighted in pink concerning the countries in the study (Gelfand et al. 2011). From the output in Table 4.1.4 below, the coefficient of the '***Micra Active'*** car model in India depicts the highest model (R^2 = 0.651) in terms of quality compared to other car models. Also, the Micra Active car model obtained a high positive correlation of 0.682 with the absence of time lag. In comparison with the first results, the model shows that there is a great improvement in variation by about 0.5% (R^2= 0.005). Besides, C-Class Cabriolet shows a strong and highest model with a degree of variation of 69%, as well as the correlation coefficient (R) of 0.703 in South Korea.

Car Model	**Linear Regression Analysis without 'Time lag'**				**Linear Regression Analysis without 'Time lag'**		
India				**South Korea**			
	R	***R^2***	***P-value***		***R***	***R^2***	***P-value***
Liva Dual Tone	0.023	0.013	0.162	Liva Dual Tone	0.041	0.015	0.231
Limited Edition	0.369	0.210	0.512	Limited Edition	0.460	0.494	0.013
Etios Liva	0.492	0.546	0.004	Etios Liva	0.512	0.520	0.019
Dual Tone Liva	0.507	0.554	0.031	Dual Tone Liva	0.291	0.357	0.014
Platinum Etios	0.075	0.113	0.931	Platinum Etios	0.044	0.014	0.882
Etios Cross	0.421	0.473	0.019	Etios Cross	0.713	0.507	0.000
Glanza	0.701	0.541	0.024	Glanza	0.604	0.540	0.002
Yaris	0.034	0.014	0.064	Yaris	0.021	0.115	0.161
Innova Touring Sport	0.397	0.201	0.001	Innova Touring Sport	0.305	0.309	0.035
C-Class Cabriolet	0.038	0.092	0.342	C-Class Cabriolet	0.703	0.690	0.002
Amaze	0.507	0.500	0.000	Amaze	0.521	0.462	0.084
Jazz	0.353	0.310	0.020	Jazz	0.412	0.303	0.139
BR-V	0.019	0.014	0.635	BR-V	0.403	0.371	0.033
Micra Active	0.682	0.651	0.002	Micra Active	0.547	0.460	0.120
Kicks	0.019	0.008	0.344	Kicks	0.201	0.182	0.021
Micra	0.030	0.043	0.043	Micra	0.108	0.059	0.093

Seltos	0.032	0.085	0.472	Seltos	0.574	0.521	0.007
New Polo	0.508	0.490	0.018	New Polo	0.508	0.490	0.018
New Vento	0.329	0.301	0.480	New Vento	0.751	0.612	0.040
Brio	0.482	0.407	0.036	Brio	0.408	0.414	0.019
Polo GT	0.564	0.535	0.004	Polo GT	0.390	0.270	0.032
Audi A6	0.437	0.396	0.015	Audi A6	0.194	0.083	0.105
Audi A8	0.621	0.580	0.029	Audi A8	0.511	0.496	0.037
BMW X1	0.352	0.330	0.643	BMW X1	0.400	0.472	0.050
BMW 3 Series	0.675	0.593	0.041	BMW 3 Series	0.710	0.644	0.005
BMW X6	0.679	0.632	0.003	BMW X6	0.680	0.740	0.001
CLS Coupe	0.002	0.017	0.772	CLS Coupe	0.454	0.012	0.302
Terrano	0.050	0.421	0.006	Terrano	0.114	0.023	0.037
G-Class	0.329	0.275	0.283	G-Class	0.611	0.591	0.091
Mercedes-AMG C 43 Coupe	0.504	0.501	0.005	Mercedes-AMG C 43 Coupe	0.591	0.463	0.015
Number of significant results			**18**	***Number of significant results***			**20**
Average significant results	0.341	0.328		Average significant results	0.423	0.366	

Table 4.1.4: Regression analysis with the absence of time lag for selected car models in both countries

Furthermore, South Korea shows that it has got 20 car models out of 30 (67%) that are statistically significant basing on 5% confidence interval. Most important to note is that South Korea has an average correlation coefficient of 0.42, as well as the mean prediction accuracy of 0.366. On the other note, 18 out of 30 (60%) car models depict statistically significant outcomes in India with a weak average relationship (R=0.341), as well as the level of variation of 32.8% (R^2 = 0.328). Therefore, 32.2% of the new car model changes in India are explained by the variation in Google trend Indices. In general, South Korea has got greater prediction accuracy with R-square 36.6% compared to India, which means that the second hypothesis is rejected. Also, it is believed that many people in South Korea are interested in searching for new car models more than India. For the last 12 months, South Koreans are more interested in new car model searches compared to Indians due to the level of variability and correlation levels.

4.2 Seasonality impacts on Car prices

Hypothesis three: *there are no significant impacts of variation on car prices in selected countries.* The data about seasonality was retrieved from the Google Trend index for the last 12 months in both India and South Korea.

Table 4.2.1: Variation in car sales and car prices in India

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.078[a]	.006	-.014	13.706

a. Predictors: (Constant), Seasonality

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	66.203	5.458		12.130	.000
	Seasonality	-.324	.587	-.078	-.551	.584

a. Dependent Variable: Car prices (India)

Table 4.2.1 above shows the distribution of the relationship between seasonality and car prices in India. The correlation coefficient, R (0.078) shows that there is a weak positive relationship between seasonality and variation in car prices. Also, the coefficient output shows negative relationship seasonality and car prices in India for the period of the last 12 months. By considering the Sig or P-value, the criterion is that when the p-value is less than the level of significance (5%), the null hypothesis is rejected. In this context, the p-value (0.854) is greater than 0.05, which implies that the null hypothesis is accepted. Therefore, there is no significant relationship between seasonality and prices for the case of India.

Table 4.2.2: Seasonality and car prices for the case of South Korea

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.078[a]	.006	-.014	22.12769

a. Predictors: (Constant), Seasonality for South Korea

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	T	Sig.
1	(Constant)	14.606	3.708		3.939	.000
	Seasonality for South Korea	.112	.204	.078	.550	.585

a. *Dependent Variable: Car prices for South Korea*

From the output in the table above, 7.8% explains the variation of seasonality using Google trends on car pr price changes. In comparison with table 4.2.1, the output, in this case, shows that the p-value (0.585) is greater than the 5% level of significance. Therefore, the null hypothesis is also accepted in the context of South Korea. However, seasonality and price variation in South Korea have got a positive relation implying that a change in season leads to the increase in car prices by 0.112.

Chapter 5: Conclusion, Discussion, and Future Recommendations

5.1 Overview

This chapter focuses on the basic results of the current study by discussing and deducing on study hypotheses, and research questions. The discussion emphasizes on comparison with the results found by other researchers about the same research topic. Also, the related Google trends' comparison of car sales gives insight as an underlying score for the study. Most important to note is that the chapter provides future recommendations with limitations encountered during the thesis.

5.2 Key Findings and Data Analyses

The analysis in this study and the related literature about the relationship between new car model sales and Google trends across India and South Korea were based on different research hypotheses. The first hypothesis stated as; ***there is no relationship between the state of economy and new car model prices across the countries in the study***. The literature review showed that there is a high correlation between the state of the economy and the new car model sales. The state of the economy can be categorized by inflation, GDP, prices, among others. In this context, the results showed that there was a weak positive relationship between the states of the economy and new car model sales across countries identified in the study (Fondeur & Karamé, 2013). Besides, Google trends showed that people or values of interest on the internet were focused on forecasting the sales of cars about the state of the economy. Most important to note is that the study results were contrary to the results depicted by the related studies in the literature. The current study's literature found out that there is no relationship between the state of the economy and new car sales since the stated hypothesis was accepted under a 5% level of significance.

Basing on the first question or hypothesis, multiple R (R=0.071) results showed that there was a weak positive relationship between the state of the economies and car prices. This implies that gross domestic product had less impact, which in turn impacts on car prices according to the results of the current study. Also, the results of R-square (0.005) showed that 0.5% of the changes in the state of the economy are explained by the car price changes using the Google trend indices. Basing on the second hypothesis, the results showed that that South Korea had an average correlation coefficient of 0.42, as well as the mean prediction accuracy of 0.366. Besides 18 out of 30 (60%), car models depict statistically significant outcomes in India with a weak average relationship (R=0.341), as well as the level of variation of 32.8% (R^2 = 0.328).

Also, by answering the **third and fourth hypothesis,** which states that there are no significant impacts of seasonality on car prices in selected countries, the results were obtained. The correlation coefficient, R (0.078) shows that there is a weak positive relationship between seasonality and variation in car prices. Also, the coefficient output shows negative relationship seasonality and car prices in India for the period of the last 12 months (Hudson & Thal, 2013). By considering the P-value, the criterion was that when the p-value is less than the level of significance (5%), the null hypothesis is rejected. In this context, the p-value (0.854) was greater than 0.05, which implied that the null hypothesis was accepted. Therefore, the study showed that there was no significant relationship between seasonality and prices for the case of India (Hudson et al. 2011).

5.3 Discussion

Discussion about Forecasting and Prediction with Google search

Research has shown that much emphasis is put on the predictive analyses compared to current evaluations of different phenomena (Fantazzini, 2014). To increase the relationship between the variables such as the state of the economy and new car model sales, the prevailing theoretical models need to be improved (iCrossing. 2015). Google trends index emphasizes the time lag existence since it alters the accuracy of predictability of a certain phenomenon. Most important to note is that variation in time may alter people's intentions or interests of searching a given item using internet data, which in turn influences a given variable under evaluation (Fantazzini & Toktamysova, 2015). This is because seasonality, especially in time lags, in this case, strongly influences the new car model sales in South Korea and India (Fantazzini, 2014). Therefore, this study emphasizes on the essentiality of time lag for forecasting and predicting internet data since it alters explanatory power or Google searches.

Research indicates that there is a significant positive relationship between the state of the economy and car prices in the economy. Therefore, the results for this study are contrary to other researches which indicate that there is a high positive relationship between the study variables (Google. 2015). Most important to note is that when the state of the economic indicators such as Gross Domestic product increases, the new car prices also increase.

Gelis and Onay (2018) surveyed about the macroeconomic effect on the sales of automobiles. The forces of the economy affect the industry of automotive. The results of Gelis and Onay found out that real GDP has a significant but positive impact on the sales of new model cars among the selected countries. Some researchers have got inconsistent results which are contrary to the current study. However, Gelis and Onay also found out that gasoline prices negatively affect the sales of cars. Therefore, such a situation is reflected by the GDP per capita since people's income fluctuates. When the income of the people in the country fluctuates, people tend to focus on other essential products than purchasing luxury commodities such as new car models.

Google trends Index Literature

Most researchers have used Google trend Index to forecast various car model sales with special key terms in the search engines (Knaub, 2015). Most important to note is that car brands are a good example of searching about car sales via the internet, which in turn shows the value of Google trends tool in the process of forecasting and prediction (Fantazzini & Toktamysova, 2015). However, specific keywords available on the internet are under-valued by decision-makers, especially in the automotive industry. Also, the current study used the real car models or car brands as key terms to search about their variations in prices and seasonality across countries in the study (Fondeur & Karamé, 2013). Most important to note is that price is considered an important factor that aids the variability of sales for goods and services in a significant way. Several companies need effective policies to be used intelligently to yield better decisions (Scott & Varian, 2013). Besides, when prices are reduced, it does not necessarily increase profits but creates the popularity and reputation of the company.

5.4 Significance of the study

The results of the study could help decision-makers within the automotive industry by providing an evaluation tool to use to make accurate predictions across countries and thus reduce the usage of outdated and traditional approaches to forecasting. In addition, the Google Trends tool can be applied as an innovative and complementary tool to increase the demand for new car model by adding an updated and efficient search engine data within the forecasting model. The patterns identified in the internet data are specified for some cars models and countries and acts as a catalyst variable to justify changes in the planning capacity of a company. The concept of prediction depends on lead time that is required by the customer to search for information before making the final purchase decision. Most important to note is that the automobile industry should understand that customers change with the variation in car models considering that other economic factors are constant.

5.5 Limitation and Future Research

Limitations

The study was limited by the comparison of only two countries of India and South Korea using Google trends index, leaving out some of the useful automobile industries across the world. Also, the Google trends index comparison was based only on specific and popular car models, which left out other up-coming brands in the countries (Kotler & Keller, 2012). This implies that the study missed out on some important data on other car brands. In this context, the study was limited to 30 car models from only eight car brands in both countries, which limited the presence of sales data about other brands.

Future Recommendations

The selection of Google trend's filter should be used to investigate the selected variables to increase the reliability and validity of results rather than using the entire Google searches.

Most important to note is that Google trends analysis is fundamental in investigating short-term variations due to the nature of internet users. Therefore, Google trends index should be used only for short-term variations like for 2 weeks, not a list of many years to increase the accuracy of results about the specified phenomenon of interest.

To get more precise, reliable, and practical results about keywords, stable search items should be used with a specific time interval to avoid inconsistency of results.

Furthermore, the search engine data can yield good results when over 70-time lags are used. This helps in reducing the dependency of probability results and risks of discovering random observations to predict study variables.

References

Alvarado-Herrera, A., Bigne, E., Aldas-Manzano, J., & Curras-Perez, R. (2017). A scale for measuring consumer perceptions of corporate social responsibility following the sustainable development paradigm. *Journal of Business Ethics*, *140*(2), 243-262.

Andreano, M., Benedetti, R., Postiglione, P., & Savio, G. (2017). On the use of Google Trend data as covariates in now-casting: Sampling and modeling issues. *SIS*.

Armstrong, J. S., Green, K. C., & Graefe, A. (2010). Forecasting principles. In M. Lovric (Ed), *International Encyclopedia of Statistical Science* (pp. 527-534). Springer Berlin Heidelberg

Arora, V. S., McKee, M., & Stuckler, D. (2019). Google Trends: Opportunities and limitations in health and health policy research. *Health Policy*, *123*(3), 338-341.

Artola, C., Pinto, F., & de Pedraza García, P. (2015). Can internet searches forecast tourism inflows? *International Journal of Manpower, 36*(1), 103-116.

Asur, S., & Huberman, B. (2010). Predicting the future with social media. *Web Intelligence and Intelligent Agent Technology*, *1*, 492-499

Babbie, E. (2015). *The practice of social research* (14th ed.). Cengage Learning.

Bäckström, M., Björklund, F., & Larsson, M. R. (2014). Criterion validity is maintained when items are evaluatively neutralized: Evidence from a full-scale five-factor model inventory. *European Journal of Personality*, *28*(6), 620-633.

Baker, S. R., & Fradkin, A. (2013). The impact of unemployment insurance on job search: evidence from google search data. *Social Science Research Network*. https://doi.org/10.2139/ssrn.2251548

Barreira, N., Godinho, P., & Melo, P. (2013). Now-casting unemployment rate and new car sales in southwestern Europe with Google Trends. *NETNOMICS: Economic Research and Electronic Networking*, *14*(3), 129–165.

Bartels, J., & Onwezen, M. C. (2014). Consumers' willingness to buy products with environmental and ethical claims: the roles of social representations and social identity. *International Journal of Consumer Studies*, *38*(1), 82-89.

Benyamin, D., & Chu, A. (2015). *U.S. Patent No. 9,002,892*. Washington, DC: U.S. Patent and Trademark Office.

Beugelsdijk, S., Maseland, R., & Van Hoorn, A. (2015). Are Scores on H ofstede's Dimensions of National Culture Stable over Time? A Cohort Analysis. *Global Strategy Journal*, *5*(3), 223-240.

Billing, T. K., Bhagat, R. S., Babakus, E., Krishnan, B., Ford Jr, D. L., Srivastava, B. N., ... & Setiadi, B. (2014). Work–family conflict and organisationally valued outcomes: The moderating role of decision latitude in five national contexts. *Applied Psychology*, *63*(1), 62-95.

Brugler, T. S., Craig, R. A., Conzola, V. C., Eischeid, T. M., & Molander, M. E. (2015). *U.S. Patent No. 8,959,003*. Washington, DC: U.S. Patent and Trademark Office.

Carley S, Krause RM, Lane BW, & Graham JD. (2013). Intent to purchase a plug-in electric vehicle: A survey of early impressions in large US cites. *Transportation Research Part D: Transport & Environment*, *18*(1), 39–45.

Chen, C., Wen, S., Zhang, J., Xiang, Y., Oliver, J., Alelaiwi, A., & Hassan, M. M. (2016). Investigating the deceptive information in Twitter spam. *Future Generation Computer Systems*, *79*, 319-326. https://doi.org/10.1016/j.future.2016.05.036

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic record*, *88*, 2-9.

Çifci, S., Ekinci, Y., Whyatt, G., Japutra, A., Molinillo, S., & Siala, H. (2016). A cross validation of Consumer-Based Brand Equity models: Driving customer equity in retail brands. *Journal of Business Research*, *69*(9), 3740-3747.

Couper, M. P. (2013). Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys. *Survey Research Methods*, *7*(3), 145-156.

Dale, M., Wicks, J., Mylne, K., Pappenberger, F., Laeger, S., & Taylor, S. (2014). Probabilistic flood forecasting and decision-making: an innovative risk-based approach. *Natural hazards*, *70*(1), 159-172.

de Carvalho, B. L., de Fátima Salgueiro, M., & Rita, P. (2015). Consumer Sustainability Consciousness: A five dimensional construct. *Ecological indicators*, *58*, 402-410.

Dessart, L., Veloutsou, C., & Morgan-Thomas, A. (2016). Capturing consumer engagement: duality, dimensionality and measurement. *Journal of Marketing Management*, *32*(5-6), 399-426.

Dhandra, T. K., & Park, H. J. (2016). A Comparative Study of Consumer Ethicality in Korea and India. *Journal of International Trade & Commerce*, *12*(5), 33-48.

Dharmani, S., Anand, D., & Demirici, M. (2015). *Shifting gear: Capacity management in the automotive industry*. http://www.ey.com/Publication/vwLUAssets/EY-shifting-gear-capacity-management-in-the-automotive

Diţoiu, M. C., & Căruntu, A. L. (2014). Sensory experiences regarding five-dimensional brand destination. *Procedia-Social and Behavioral Sciences*, *109*(2), 301-306.

Dolara, A., Leva, S., & Manzolini, G. (2015). Comparison of different physical models for PV power output prediction. *Solar energy*, *119*, 83-99.

Dwivedi, A. (2015). A higher-order model of consumer brand engagement and its impact on loyalty intentions. *Journal of Retailing and Consumer Services*, *24*, 100-109.

Elshendy, M., Colladon, A. F., Battistoni, E., & Gloor, P. A. (2017). Using four different online media sources to forecast the crude oil price. *Journal of Information Science*, *44*(3), 408-421. https://doi/10.1177/0165551517698298

Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(33), E4512-21. https://doi.org/10.1073/pnas.1419828112

Erkan, I., & Evans, C. (2016). The influence of eWOM in social media on consumers' purchase intentions: An extended approach to information adoption. *Computers in Human Behavior, 61*, 47–55.

Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(33), E4512-21. https://doi/10.1073/pnas.1419828112

Fantazzini (2014). Now-casting and forecasting the monthly food stamps data in the US using online search data. *PloS One, 9*(11). http://dx.doi.org/10.1371/journal.pone.0111894

Fantazzini, D., & Toktamysova, Z. (2015). Forecasting German car sales using Google data and multivariate models. *International Journal of Production Economics*, *170*, 97-135. https://doi/10.1016/j.ijpe.2015.09.010

Fatma, M., Rahman, Z., & Khan, I. (2016). Measuring consumer perception of CSR in tourism industry: Scale development and validation. *Journal of Hospitality and Tourism Management*, *27*, 39-48.

Fondeur, Y., & Karamé, F. (2013). Can Google data help predict French youth unemployment? *Economic Modelling*, *30*, 117-125.

Frechtling, D. (2011). *Forecasting Tourism Demand: methodologies and strategies*. Routledge. https://www.routledge.com/Forecasting-Tourism-Demand/Frechtling/p/book/9780750651707

Fritzsch, B., Wegner, K., Sibbertsen, P., & Ullmann. G. (2019). Can google trends improve sales forecasts on a product level?. *Journal of Applied Economics Letter*. https://www.tandfonline.com/doi/abs/10.1080/13504851.2019.1686110

Fritzsch, B., Wenger, K., Sibbertsen, P. &Ullmann, G. (2019). Can Google trends improve sales forecasts on a product level?

https://www.tandfonline.com/doi/abs/10.1080/13504851.2019.1686110

Fumo, N., & Biswas, M. R. (2015). Regression analysis for prediction of residential energy consumption. *Renewable and sustainable energy reviews*, *47*, 332-343.

Gao, W., Darvishan, A., Toghani, M., Mohammadi, M., Abedinia, O., & Ghadimi, N. (2019). Different states of multi-block based forecast engine for price and load prediction. *International Journal of Electrical Power & Energy Systems, 104*, 423-435.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., & Aycan, Z. (2011). Differences between Tight and Loose Cultures: A 33-Nation Study. *Science*, *332*(6033), 1100-1104.

Gensler, S., Neslin, S. A., & Verhoef, P. C. (2017). The Showrooming Phenomenon: It's More than Just About Price. *Journal Of Interactive Marketing, 3,,* 29–43.

Geva, T., Oestreicher-Singer, G., Efron, N., & Shimshoni, Y. (2017). Using Forum and Search Data for Sales Prediction of High-Involvement Products. *Management Information Systems Quarterly*, *41*(1), 65–82. https://doi.org/10.25300/misq/2017/41.1.04

Ghirvu, A. I. (2013). The AIDA Model for AdverGames. *The USV Annals of Economics and Public Administration*, *13*(1(17)), 90-98.

Gnann, T., Plötz, P., Funke, S., & Wietschel, M. (2015). What is the market potential of plug-in electric vehicles as commercial passenger cars? A case study from Germany. *Transportation Research Part D: Transport & Environment*, *37*, 171–87.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with web search. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(41), 17486–17490. http://www.jstor.org/stable/20780485

Google. (2015). Search Google Trends Help: How Trends data is adjusted. Retrieved May 19, 2016. https://support.google.com/trends/answer/4365533?hl=en&ref_topic=4365599

Google. (2016). How trends data is adjusted - Trends Help. https://support.google.com/trends/answer/4365533?hl=en&ref_topic=4365599

Google. (2016). Privacy Policy - Privacy & Terms. https://www.google.de/intl/en/policies/privacy/

Google. (2016). Search tips for Trends. https://support.google.com/trends/answer/4359582?hl=en&ref_topic=4365530

Google. (2016). Where Trends data comes from. https://support.google.com/trends/answer/4355213?hl=en&ref_topic=4365599

Goyal, M. (2019). *With India's economy growing at about 7%, why the auto industry is hurting so badly?* Economic Times. https://economictimes.indiatimes.com/industry/auto/autonews/when-indias-economy-is-growing-at-about-7-then-how-could-auto-industry-be-hurting-so-badly/articleshow/69075048.cms

Guerard, J. (2013). *Introduction to Financial Forecasting in Investment Analysis.* Springer Science & Business Media.

Guerard, J. B. (2013). Regression Analysis and Multicollinearity: Two Case Studies. In *Introduction to Financial Forecasting in Investment Analysis* (pp. 73-96).

Gupta, A., & Sahu, G. P. (2015). Exploring relationship marketing dimensions and their effect on customer loyalty-a study of Indian mobile telecom market. *International Journal of Business Innovation and Research*, *9*(4), 375-395.

Hassan, S., Nadzim, A., Zaleha, S., & Shiratuddin, N. (2015). Strategic use of social media for small business based on the AIDA model. *Procedia-Social and Behavioral Sciences*, *172*, 262-269.

Hassan, S., Nadzim, S. Z. A., & Shiratuddin, N. (2014). Strategic Use of Social Media for Small Business Based on the AIDA Model. The Global Conference on Business & Social Science-2014, Kuala Lumpur.

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence based nursing*, *18*(3), 66-67.

Hendry, D. F., & Mizon, G. E. (2014). Unpredictability in economic analysis, econometric modeling and forecasting. *Journal of Econometrics, 182*(1), 186-195.

Hoffmann, C., Wulf, T., & Stubner, S. (2016). Understanding the performance consequences of family involvement in the top management team: The role of long-term orientation. *International Small Business Journal, 34*(3), 345-368.

Hofstede-insights.com. (2020). *Country Comparison*. Hofstede Insights.

https://www.hofstede-insights.com/country-comparison/india,south-korea/

Holden, R. R. (2010). Face validity. *The corsini encyclopedia of psychology*, 1-2.

Howes, J. Z., Solderitsch, J., LaSalle, R. M., Rozmiarek, D. W., & Ellett, E. A. (2016). *U.S. Patent No. 9,355,172*. Washington, DC: U.S. Patent and Trademark Office.

Hu, X., Martinez, C. M., & Yang, Y. (2017). Charging, power management, and battery degradation mitigation in plug-in hybrid electric vehicles: A unified cost-optimal approach. *Mechanical Systems & Signal Processing*, *87*, 4–16.

Hu, X., Zou, Y., & Yang, Y. (2016). Greener plug-in hybrid electric vehicles incorporating renewable energy and rapid system optimization. *Energy*, *111*:971–80

Huang, D. A., & Kitani, K. M. (2014, September). Action-reaction: Forecasting the dynamics of human interaction. In *European Conference on Computer Vision* (pp. 489-504). Springer, Cham.

Hudson, S., & Thal, K. (2013). The Impact of Social Media on the Consumer Decision Process: Implications for Tourism Marketing. *Journal of Travel & Tourism Marketing*, *30*(1-2), 156-160.

Hudson, S., Wang, Y., & Gil, S. M. (2011). The Influence of a Film on Destination Image and the Desire to Travel: a Cross-Cultural Comparison. *Internationa*l *Journal of Tourism Research*, *13*(2), 177-190.

Hur, W. M., Kang, S., & Kim, M. (2015). The moderating role of Hofstede's cultural dimensions in the customer-brand relationship in China and India. *Cross Cultural Management*.

Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: Principles & Practice*. Otexts.com. https://otexts.com/fpp2/

iCrossing. (2015). *Search engine infographic 2015: The countries that stand between Google and total world domination.* http://www.icrossing.com/uk/ideas/search-engine-infographic-2015-countries-stand-between-google-and-total-world-domination

Jeon, J. (2015). The strengths and limitations of the statistical modeling of complex social phenomenon: Focusing on SEM, path analysis, or multiple regression models. *International Journal of Economics and Management Engineering*, *9*(5), 1594-1602.

Jerath, K., Ma, L., & Park, Y. H. (2014). Consumer Click Behavior at a Search Engine: The Role of Keyword Popularity. *Journal of Marketing Research*, *51*(4), 480-486.

Jiang, X. (2017). Discussion on the Methodology of Hofstede's National Cultural Theory. In *Asia International Symposium on Language, Literature and Translation* (p. 443).

Jindal, H., Faujdar, D. S., & Kumar Aggarwal, A. (2019). Rising Public Interest of Hookah and E-cigarettes after Tobacco Cessation Law in India: A Google Trend Analysis. International *Journal of Health Systems and Implementation Research*, *3*(1), 56-64.

Kandaswami, K., & Tiwar, A. (2014). Deloitte. Driving through the consumer's mind: Steps in the buying process. *Deloitte Touch India.* http://www2.deloitte.com/content/dam/Deloitte/in/Documents/manufacturing/in-mfg-dtcmsteps-in-the-buying-process-noexp.pdf

Karlgren, J., Sahlgren, M., Olsson, F., Espinoza, F., & Hamfors, O. (2012, April). Usefulness of sentiment analysis. In *European Conference on Information Retrieval* (pp. 426–435).

Kinski, A. (2016). *Google trends as complementary tool for new car sales forecasting: a cross-country comparison along the customer journey* (Master's thesis, University of Twente).

Knaub. (2015). *When Prediction is Not Time Series Forecasting: Note on Forecasting v Prediction in Samples for Continuous Data*. https://doi.org/10.13140/RG.2.1.3412.4328

Kotler, P., & Keller, K. L. (2012). Marketing Management.14th edition.

Kumar, V., & Syan, A. S. (2017). Consumer Decision Making of Solar Water Heaters: An Empirical Analysis.

Labbe, F & Swallow, Y. (2013). Now-casting With Google Trends in an Emerging Market. *Journal of Forecasting*. https://www.researchgate.net/publication/46449441_Nowcasting_With_Google_Trends_in_an_Emerging_Market

Lana, I., Del Ser, J., Velez, M., & Vlahogianni, E. I. (2018). Road traffic forecasting: Recent advances and new challenges. *IEEE Intelligent Transportation Systems Magazine*, *10*(2), 93-109.

Lassen, N. B., Madsen, R., & Vatrapu, R. (2014, September). Predicting iPhone Sales from iPhone Tweets. In *2014 IEEE 18th International Enterprise Distributed Object Computing Conference* (pp. 81-90). IEEE.

Lassoued, R., & Hobbs, J. E. (2015). Consumer confidence in credence attributes: The role of brand trust. *Food Policy*, *52*, 99-107.

Le, K., Narayanan, D., Tran, V., Ho, M., & Steinberg, H. (2016). CUSTOMER SERVICE & HOFSTEDE'S CULTURAL DIMENSIONS IN INDIA, SOUTH KOREA, VIETNAM, & THE USA. In *Allied Academies International Conference. Academy for Studies in International Business*. *Proceedings* (Vol. 16, No. 1, p. 11). Jordan Whitney Enterprises, Inc.

Lee, S. H., & Hoffman, K. D. (2015). Learning the ShamWow: Creating infomercials to teach the AIDA model. *Marketing Education Review*, *25*(1), 9-14.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of marketing*, *80*(6), 69-96.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167. https://doi.org/10.2200/s00416ed1v01y201204hlt016

Mavragani, A., Ochoa, G., & Tsagarakis, K. P. (2018). Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. https://www.jmir.org/2018/11/e270/

medica, *24*(1), 12-18.

Mellon, J. (2013). Where and When Can We Use Google Trends to Measure Issue Salience?. *PS: Political Science & Politics*, *46*(2), 280-290.

Mellon, J. (2014). Internet Search Data and Issue Salience: The Properties of Google Trends as a Measure of Issue Salience. *Journal of Elections, Public Opinion & Parties*, *24*(1), 45-72.

Mestre, F., Pita, R., Paupério, J., Martins, F. M., Alves, P. C., Mira, A., & Beja, P. (2015). Combining distribution modelling and non-invasive genetics to improve range shift forecasting. *Ecological Modelling*, *297*, 171-179.

Milton, S. (1986). A Sample Size Formula for Multiple Regression Studies. *Public Opinion Quarterly*, *50*(1), 112-118.

Mislove, A., Lehmann, S., Ahn, Y., Onnela, J., and Rosenquist, J. 2011. "*Understanding the Demographics of Twitter Users,*" Fifth International AAAI Conference on Weblogs and Social Media (N. Nicolov and J. Shanaha, eds.), Barcelona: AAAI digital library, pp. 17–21

Mitchell, J. (2011*). Google Searches vs. Sentiment Analysis*: Which Is The Real Zeitgeist? http://readwrite.com/2011/12/22/google_searches_vs_sentiment_analysis_whi ch_is_the/

Montoya, R. M., & Horton, R. S. (2014). A Two-Dimensional Model for the Study of Interpersonal Attraction. *Personality and Social Psychology Review*, *18*(1), 59-86..

Moon, K. H., & Kim, J. M. (2019). Analysis of preference convergence by analyzing search words for oralcare products: Using the Google trend. *Journal of the Korea Convergence Society*, *10*(6), 59-64.

Moore, J. (2015). Pragmatism, mathematical models, and the scientific ideal of prediction and control.*Behavioural Processes*, *114*, 2-13. https://doi.org/10.1016/j.beproc.2015.01.007

Mousavi, M., & Moghe, R. (2014). *U.S. Patent No. 8,682,612*. Washington, DC: U.S. Patent and Trademark Office.

NetApplications.com. (2016). *Search Engine Market Share*. https://www.netmarketshare.com/search-engine-marketshare.aspx?qprid=4&qpcustomd=0

Nguyen, T. H., Shirai, K., & Velcin, J. 2015. "Sentiment analysis on social media for stock movement prediction," *Expert Systems with Applications 42*(24), 9603–9611. https://doi.org/10.1016/j.eswa.2015.07.052

Nyadzayo, M. W., & Khajehzadeh, S. (2016). The antecedents of customer loyalty: A moderated mediation model of customer relationship management quality and brand image. *Journal of Retailing and Consumer Services, 30*, 262-270.

Nymand-Andersen, P., & Pantelidis, E. (2018). *Google econometrics: Nowcasting euro area car sales and big data quality requirements*. Statistics Paper Series, No. 30, pp. 2-33

Pashootanizadeh, M., & Khalilian, S. (2018). Application of the AIDA model. *Information and Learning Science*.

Peng, B., Song, H., & Crouch, G. I. (2014). A meta-analysis of international tourism demand forecasting and implications for practice. *Tourism Management*, *45*, 181-193.

Perrin, A., & Duggan, M. (2015). *Americans' Internet Access*: 2000-2015. http://www.pewinternet.org/2015/06/26/americansinternet-access-2000-2015/

phenomenon: Focusing on SEM, path analysis, or multiple regression models. *Int J Soc Behav Educ Econ Bus Ind Eng*, *9*(5), 1594-1602.

Pieper, A. T. (2016). *Detecting Review Spam on Amazon with Reviewalarm*. (Bachelor's thesis, University of Twente).

Puška, A., Stojanović, I., Šadić, S., & Bečić, H. (2018). The influence of demographic characteristics of consumers on decisions to purchase technical products. *The European Journal of Applied Economics, 15*(2), 1-16.

Puzach, C., & Hass, A. (2014). Addressing the Needs of Juvenile Offenders: A Study of the Greene County Missouri Teen Court Model. *Int'l J. Soc. Sci. Stud*., *2*, 113.

Reyes, A., Rosso, P., & Buscaldi, D. (2012). From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*, *74*, 1–12.

Risius, M., Akolk, F., & Beck, R. (2015). Differential Emotions and the Stock Market – The Case of Company-specific Trading. *ECIS 2015 Completed Research Papers*. Paper 147.

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic bulletin & review*, *8*(3), 385-407.

Rosenbaum, M. S., Otalora, M. L., & Ramírez, G. C. (2017). How to create a realistic customer journey map. *Business Horizons*, *60*(1), 143-150.

Saberi, B. (2018).The role of the automobile industry in the economy of developed countries. *International Robotics & Automation Journal*, *4*(3), 179-180

Sang, Y. N., & Bekhet, H. A. (2015). Modelling electric vehicle usage intentions: an empirical study in Malaysia. *Journal of Cleaner Production*, *92*, 75–83.

Scott, S. L., & Varian, H. R. (2013). *Predicting the present with bayesian structural time series*. http://people.ischool.berkeley.edu/~hal/Papers/2013/pred-present-with-bsts.pdf

Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, *311*, 18– 38. https://doi.org/10.1016/j.ins.2015.03.040

Shafi, S. I., & Madhavaiah, C. (2013). The Influence of Brand Equity on Consumer Buying Behaviour of Organic Foods in India. *Journal Of Marketing & Communication*, *9*(2), 44-51.

Shen, H., Wu, L., Yi, S., & Xue, L. (2018). The effect of online interaction and trust on consumers' value co-creation behavior in the online travel community. *Journal of Travel & Tourism Marketing*, 1-11.

Shmueli, G. (2010). To Explain or to Predict?. *Statistical Science*, *25*(3), 289-310. https://doi.org/10.1214/10-sts330

Singh, H., Weickert, B., Sprague, M., Kross, M., Borochoff, A., Menon, P., & Harris, M. (2015). *U.S. Patent No. 9,177,344*. Washington, DC: U.S. Patent and Trademark Office.

Singh, R., & Venugopal, P. (2015). The impact of salesperson customer orientation on sales performance via mediating mechanism. *Journal of Business & Industrial Marketing*.

Smith, R., Deitz, G., Royne, M. B., Hansen, J. D., Grünhagen, M., & Witte, C. (2013). Cross-cultural examination of online shopping behaviour: A comparison of Norway, Germany, and the United States. *Journal of Business Research*, *66*(3), 328-335.

So, W. (2020). *Automotive industry in South Korea - Statistics & Facts*. Retrieved from: https://www.statista.com/topics/5249/automotive-industry-in-south-korea/

Somervuori, O., & Ravaja, N. (2013). Purchase Behavior and Psychophysiological Responses to Different Price Levels. *Psychology and Marketing*, *30*(6), 479-489. https://doi.org/10.1002/mar.20621

Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia medica: Biochemiamedica*, *24*(1), 12-18.

Sreen, N., Purbey, S., & Sadarangani, P. (2018). Impact of culture, behavior and gender on green purchase intention. *Journal of Retailing and Consumer Services*, *41*, 177-189.

Statista. (2020). https://www.statista.com/statistics/650919/south-korea-vehicle-production-by-type/

Stein, A., & Ramaseshan, B. (2016). Towards the identification of customer experience touch point elements. *Journal of Retailing and Consumer Services*, *30*, 8-19. https://doi.org/10.1016/j.jretconser.2015.12.001

Stephens-Davidowitz, S., & Varian, H. (2014). *A Hands-on Guide to Google Data*. http://people.ischool.berkeley.edu/~hal/Papers/2015/primer.pdf

Sun, C., Hu, X., Moura, S. J., & Sun, F. (2015). Velocity Predictors for Predictive Energy Management in Hybrid Electric Vehicles. *IEEE Transactions on Control Systems Technology*, *23*(3):1197–204

Suthikarnnarunai, N. (2010). “Automotive supply chain and logistics management,” Imecs 2008: *International Multiconference of Engineers and Computer Scientists*, Vols I and Ii, pp. 1800–1806.

Treviño, L. J., Egri, C. P., Ralston, D. A., Naoumova, I., Furrer, O., Li, Y., ... & de la Garza Carranza, M. T. (2019). A multi-country, multi-sector replication challenge to the validity of the cultural tightness-looseness measure. *Asia Pacific Journal of Management*, 1-30.

Tuarob, S., Tucker, C. S., & Asme. (2014). Fad or here to stay: predicting product market adoption and longevity using large scale, social media data. *Proceedings of the Asme International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 2013, Vol 2b (doi: V02bt02a012).

Vaijayanthi, P., & Shreenivasan, K. (2017). Measuring brand equity of cosmeceuticals â€ a case using consumer based brand equity model. *Asian Journal of pharmaceutical and clinical research*, *10*(6), 210-215.

Valaei, N., Rezaei, S., Ismail, W. K. W., & Oh, Y. M. (2016). The effect of culture on attitude towards online advertising and online brands: applying Hofstede's cultural factors to internet marketing. *International Journal of Internet Marketing and Advertising*, *10*(4), 270-301.

Vaughan, L., & Chen, Y. (2015). Data mining from web search queries: A comparison of google trends and baidu index. *Journal of the Association for Information Science and Technology*, *66*(1), 13-22.

Voortman, M. C. (2015). *Validity and reliability of web search based prediction for car sales*. (Master’s thesis, University of Twente).

Vosen, S., & Schmidt, T. (2011). Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends. *Journal of Forecasting*, *30*, 565-578. https://doi.org/10.1002/for.1213

Waheed, A., Mahasan, S. S., & Sandhu, M. A. (2014). Factor That Affects Consumer Buying Behavior: An Analysis of Some Selected Factors. *Middle-East Journal of Scientific Research*, *19*(5), 636-641.

Wang, G. T., & Park, K. (2015).*Student Research and Report Writing: From Topic Selection to the Complete Paper*. John Wiley & Sons.

Wijnhoven, F & Plant, O. (2017). Sentiment Analysis and Google Trends Data for Predicting Car Sales. *Thirty Eighth International Conference on Information Systems*, South Korea, pages 1-15. https://pdfs.semanticscholar.org/abb6/37b12887a8ee3f612af0706d8e7028f7681a.pdf

Wijnhoven, F., & Bloemen, O. (2014). External validity of sentiment mining reports: Can current methods identify demographic biases, event biases, and manipulation of reviews? *Decision Support Systems*, *59*, 262-273. https://doi.org/10.1016/j.dss.2013.12.005

Wikström, M., Eriksson, L., & Hansson, L. (2016). Introducing plug-in electric vehicles in public authorities. *Research in Transportation Business & Management*, *18*, 29–37

Wu, L., & Brynjolfsson, E. (2010). The Future of Prediction: How Google Searches Foreshadow Housing Prices and Quantities. *Paper presented at the ICIS 2010 Proceedings*, Phoenix, AZ.

Wu, L., & Brynjolfsson, E. (2015). The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales. *Economic Analysis of the Digital Economy*. Goldfarb, S. Greenstein, and C. Tucker (eds.), Chicago, Illinois: University of Chicago Press, pp. 89–118

Xu, J., & Perron, P. (2014). Forecasting return volatility: Level shifts with varying jump probability and mean reversion. *International Journal of Forecasting*, *30*(3), 449-463.

Yadav, M., & Rahman, Z. (2017). Measuring consumer perception of social media marketing activities in e-commerce industry: Scale development & validation. *Telematics and Informatics*, *34*(7), 1294-1307.

Carrière-Swallow, Y., & Labbé, F. (2013). Now-casting with Google Trends in an Emerging Market. *Journal of Forecasting*, *32*(4), 289–98.

Yang, X., Pan, B., Evans, J. A., and Lv, B. F. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management*, *46*, 386–397.

Youn, S., & Cho, H. C. (2016). Nowcast of TV market using Google Trend data. *Journal of Electrical Engineering & Technology*, *11*(1), 227-233.

Zhang, Y., & Oczkowski, E. (2016). Exploring the potential effects of expatriate adjustment direction. *Cross Cultural & Strategic Management*.

Zhang, Y., Wei, X., & Zhou, W. (2017). An asymmetric cross-cultural perspective on the mediating role of conflict management styles in expatriation. *International Journal of Conflict Management.*

Zhang. Y, Zhong. M, Geng. Z & Jian. Y. (2017). Forecasting electric vehicles sales with univariate and multivariate time series models: The case of China. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0176729

Zhao, H. Y., Kwon, J. W., & Yang, O. S. (2016). Updating Hofstede's Cultural Model and Tracking Changes in Cultural Indices. *Journal of International Trade & Commerce, 12*(5), 85-106.

Zhu, J. J., Wang, X., Qin, J., & Wu, L. (2012). Assessing Public Opinion Trends Based on User Search Queries: Validity, Reliability, and Practicality. *The Annual Conf. of the World Association for Public Opinion Research*, 1-7.