WELINGKAR INSTITUTE OF MANAGEMENT DEVELOPMENT & RESEARCH

TRIMESTER VI PROJECT REPORT

ON

"PREDICTION OF CAR MODEL PREFERENCES BY CUSTOMERS – A STUDY ON HYUNDAI MOTORS"

BY

ANSHIKA JAIN

PGDM - RESEARCH & BUSINESS ANALYTICS (2019 – 21)

ROLL NO : 26

PROJECT FACULTY GUIDE

PROF. SONAL DAULATLKAR

WELINGKAR INSTITUTE OF MANAGEMENT DEVELOPMENT & RESEARCH

TRIMESTER VI PROJECT REPORT

ON

"PREDICTION OF CAR MODEL PREFERENCES BY CUSTOMERS – A STUDY ON HYUNDAI MOTORS"

BY

ANSHIKA JAIN

PGDM - RESEARCH & BUSINESS ANALYTICS (2019 – 21)

ROLL NO : 26

PROJECT FACULTY GUIDE

PROF. SONAL DAULATLKAR

DECLARATION

I hereby declare that the project report entitled "Prediction of car models preferences – A study on Hyundai Motors" submitted to the S. P. Mandali's Welingkar Institute of Management Development & Research, Mumbai is a record of work done by me under the guidance of Prof. Dr. Sonal Daulatkar, Assistant Professor, Research & Business Analytics. This report has been submitted under the partial fulfilment of the requirements for the award of the post- graduation degree of Post Graduate Diploma in Management (Research & Business Analytics). The results embodied in this report have not been submitted to any other university or institute for the award of any degree or diploma.

_____

Name: Anshika Jain
Date: 20th March 2021

PROJECT COMPLETION CERTIFICATE

TO WHOSOEVER IT MAY CONCERN

This is to certify that project titled "**Prediction of car models preferences – A study on Hyundai Motors"** has successfully been completed by **Ms. Anshika Jain** in partial fulfilment of her two years full time course 'Post Graduation Diploma in Management – Research & Business Analytics' recognized by AICTE through the Print. L. N. Welingkar Institute of Management Development & Research, Matunga, Mumbai.

This project has been completed carried out under my guidance.

_____

(Signature of Faculty Guide)

Name: Dr. Sonal Daulatkar

Designation: Assistant Professor, Research and Business Analytics

SYNOPSIS

1. Name of the Student : Anshika Jain

2. Program & Year : PGDM – Research & Business Analytics – 2019-2021

3. Area of Project Research : Business Analytics

4. Name of the Faculty Guide : Dr. Sonal Daulatkar

5. Title of the Project : Predict the car model preferences – A study on Hyundai Company.

6. Project Details

(A) Objectives of study :

a) To classify and predict most preferred car model on the basis of multiple features like age,

income, location, family size etc.

b) To analyze the customer behavior and preferences.

c) The understand existing trend in the market in automobile industry

(B) Review of Literature :

Customers now do research for the differentiating parameters that helps them to choose among the alternative products available in the market. The purpose of the research paper is to know the identification of parameters that impacts customers purchase decisions. To know preferences for models and why customer abandon brand loyalty and switch to other brands that provide high perceived value. This research will help the manufacturers to design or upgrade models that suits customer's requirements and preferences and promote accordingly. (Beena et al., 2013).

(C) Research Methodology :

- Step I: Research Design

  The data collection will consists of primary research through a questionnaire. The dataset will consist of one target variable i.e. prediction of car model preferences of Hyundai. The features of the dataset will consist of safety, age, income, gender, family size, location, mileage. Predictions can be done using unsupervised algorithms like k means etc. Tools used for prediction will be python and tableau.

- Step II: Data Collection

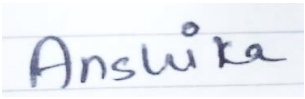
Collection of Primary data using questionnaire, personal visits etc.

(D) Findings/Results and Analysis :

The findings will help us find the most preferred car models of Hyundai. This will help in understanding the customer behaviour in the existing market.

(E) Conclusions and Recommendations:

This will help to know the individual's expectations, preferences for models with the reasons and how lifestyle changes affect their choices. It will also help to comprehend that why customer abandon brand loyalty and switch to other brands that provide high perceived value. This research will help the manufacturers to design or upgrade models that suit customer's requirements and preferences and promote accordingly.

Faculty Guide	**Student**
Name : Dr. Sonal Daulatkar	Name : Anshika Jain
Contact No: 8871156790	
Signature:	Signature:
	
Date: 08/02/2021	Date: 09/02/2021

ACKNOWLEDGEMENT

The successful result of this project was possible due to tremendous research, extensive work, and dedication towards the project. This wouldn't have been possible without the constant support received from peers & surroundings directly or indirectly. I would like to express a sincere thanking note and gratitude to all of them for their support.

I firstly thank my parents for all the support and motivation they have shown towards me. Any project is incomplete without the support of one's family.

I owe my deep gratitude to my mentor Prof. Dr. Sonal Daulatkar for mentoring me and guiding me. She has been my mentor for the project motivating me with very useful insights and suggestions despite the busy schedule. I would also like to thank all the Research and Business Analytics faculty members for their guidance.

Also thanking the Dean of our department, Prof. C. Y. Nimkar for all the insights and guidance.

Lastly, a thanking note to all my family members, friends, individuals around & mother nature for the support

Thanking all of them

Ms. Anshika Jain

Roll No. 26

NO PLAGIARISM DECLARATION BY THE STUDENT

I, the undersigned, hereby declare that the project titled "**Prediction of car models preferences – A study on Hyundai Motors".**

(a) Has been prepared by me towards the partial fulfilment for the award of Post-Graduation Diploma in Management – Research and Business Analytics under the guidance of Prof. Sonal Daulatkar, S. P. Mandali's Prin. L. N. Welingkar Institute of Management Development and Research, Mumbai.

(b) This work is original and has not been submitted for any degree/diploma in this or any other Institute/Organization.

(c) The information furnished in this dissertation is genuine and original to the best of my knowledge and belief. (d) I have not indulged in plagiarism. The project report has been checked for plagiarism and output report has been attached.

Student Name : Anshika Jain Signature :

Place: Mumbai Date:

PLAGARISM REPORT

Curiginal

Document Information

Analyzed document	Anshika Jain_PGDM RBA_2019-21.docx (D99548986)
Submitted	3/25/2021 7:35:00 AM
Submitted by	Anshika jain
Submitter email	anshika.jain1022@gmail.com
Similarity	9%
Analysis address	pgdmrba.wimdr@analysis.urkund.com

Sources included in the report

W	URL: https://www.ethz.ch/content/dam/ethz/special-interest/baug/ivt/ivt-dam/vpl/reports ... Fetched: 3/25/2021 7:36:00 AM	8
SA	**Priya. N.doc** Document Priya. N.doc (D33777711)	1
W	URL: https://www.ijser.in/archives/v5i9/IJSER171885.pdf Fetched: 1/5/2020 11:20:21 AM	1
SA	**Ranjan Sahu Thesis for Plag Check.docx** Document Ranjan Sahu Thesis for Plag Check.docx (D83383978)	4
W	URL: https://shodhganga.inflibnet.ac.in/jspui/bitstream/10603/231037/3/03_chapter_2.pdf Fetched: 5/29/2020 7:30:42 AM	2
SA	**Ranjan Sahu Thesis for Plag Check.docx** Document Ranjan Sahu Thesis for Plag Check.docx (D84737628)	1

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

In Indian economy, automobile sector comes as one of the important and growing industries. Currently in Indian market there are Indian as well as international brands of automobiles - Maruti, Honda, Skoda, Audi, Nissan, BMW, Hyundai etc. This leads to cutthroat competition among the companies therefore satisfying and delighting the customers become the differentiating factor in the competitive automobile industry. Companies are now becoming customer focused and are aiming to retain their existing customers and increase their market share in India. Due to constant changes in their attitudes, needs, lifestyles, expectations of the customers, companies need to adapt themselves and implement those strategies that will help them to remain competitive in the market.

a. **Background of the Problem:** Even Hyundai Motors, one of the leading automobile giants in India, is very much concerned about understanding the customer behaviour, their needs, and expectations. The present study throws light on how company should identify a mix of values, preferences and behaviour that impacts the consumer's purchase decisions and so the car model. After knowing the preferences and expectations they can decide on mix of products (vehicle models) and technology aspects (fuel type/engines/features/transmission) to be used. Hence the purpose of this study is to come up with a model that helps to predict the car model preferences of prospective or existing customers with high accuracy.

b. Objectives of the Study

i. To analyse the customer behaviour and preferences.

ii. Finding & fitting a predictive model that relates the predictors to the response, where we will try to predict the preferred car model based on multiple features.

REVIEW OF LITERATURE

According to various researchers, the intention to purchase cars depends on three explanatory variables : psychological aspects, socio-demographics, and site-specific dummy variables **Belgiawan et al. (2017).**They measured the dependent variables using response rate on a 7-point Likert scale (very unlikely – very likely) and the type of cars is as follows: Small, Mid-size, Large, Luxury, Sports, Hybrid and Electric. There are 20 attitudinal variables toward cars obtained from the survey and to understand 18 better "emotional attachment" of students to cars, the attitudinal questions are analyzed and 19 grouped with PCA. Correlation model is used and found that psychological aspects are significant for different models of cars.

Raghu (2013) in his study proposed that marketing can play an important role and can be made effective by knowing the buying behavior of consumers and creating personality of brands and how companies are now using Total relationship management(TRM) philosophy in marketing their cars. And how competition is changing the business environment completely and the companies that are adapt strategize and react to these changes quickly are the most successful ones. Due to technological developments, importance of physical differences has decreased. Differentiation should be made on the meaning of products bear instead of on their physical features. A study by **Malhotra and Jain** suggested in **2017** that successful brand differentiation can be made possible by building personality. All this is possible because of brand personality and so customer sees brands as their friend with emotional benefits.

John et al. (2013*)* conducted research on small car preferences and found that demographic factors like age, gender, education, status, and income influence consumers indirectly for small car buying preference. There are factors one is Value consciousness and second is price quality inference which influences small car buying. Diesel cars were preferred by male consumers whereas Petrol cars were preferred by female consumers. They highlighted the fact that "India's passenger automobile car production is the sixth largest in the world, in keeping with recent reports, according to the Society of Indian Automobile Manufacturers, annual vehicle sales are projected to increase to five million by 2015 and more than nine million by 2020. By 2050, the car volumes will be at the top of the world in our country with approximately 611 million vehicles on the roads of the nation" (p. 02).

Joseph and Kamble (2011) evaluated the consumer behavior pattern shown by passenger car customers in the state - Karnataka. Data is collected from 525 passenger car owners consisting of professionals, businessmen, employees of public and private sector and agriculturist in Karnataka district. Through their study they found that availability of auto finance or consumer credit is one of the most important factors that influences purchase of passenger cars in India.

Chopra (2018) conducted research to compare the consumer preference towards passenger cars of Hyundai motors and Maruti Suzuki in Delhi Region. He used primary data for the

study was collected using a self-designed questionnaire on the sample size of 150 respondents using convenience sampling technique. It was followed by Independent T test analysis to compare the preferences of consumers towards the passenger vehicles of Maruti Suzuki and Hyundai Motors. A comparison on these brands was performed on five attributes namely, price, interior, color, after sales service and re sale value. From the study it was found that there is no difference in the consumer preference across these two brands on attributes such as interiors, color and after sales service. However, the preferences of the customers varied on factors like price & resale value .The main recommendation of this research is to adopt innovative techniques by the brands in order to distinguish themselves and build brand equity for them.

Fuel efficiency and safety comes at the top of the checklist when buying a car**. Armstrong (2018)** examined other factors that are taken into consideration when purchasing a car - quality , warranty and customer service, suitability for everyday use , driving comfort , design , low price, and spaciousness . **Mandeep Kaur and Sandhu** concluded in their report of passenger cars living in the major cities of the State of Punjab and the Union Territory of Chandigarh **(2006)** that safety and comfort were the most important features of the passenger car followed by luxuriousness.

Norwegian University (2013) conducted an online retrospective survey on 200 households revealed Socio-demographic and psychological variables explain purchases of new cars. The strong determinant of carbon dioxide emission levels is car type. Normative influences on car purchase are indirect. Intention to buy environmentally friendly cars directly impact on carbon dioxide emissions and women buy more fuel-efficient cars than men do.

RESEARCH DESIGN

a) **Type of Research design:** There are two types of Research Design – Quantitative and Qualitative.

   i. This Project follows mixed approach of Research Design, where both quantitative as well as qualitative data is obtained and used in tandem to predict the car model preference among the prospective customers.

b) **Data Collection:** There are primarily two types of Data collection – Primary and Secondary.

   i. This Project uses primary data collection method using structured questionnaire and/or personal visits to obtain data of owners of passenger cars to comprehend their views and experiences regarding various facets of the cars, which would help to study their perception and preferences in choosing the car.

   ii. **Sampling Technique** – Non-Probability Convenience Sampling Method. Such kind of sampling method often allows a respondent to self-select into the sample and allows the researchers to choose who, where, and when to collect the data. One of the drawbacks of this method is that the results may not represent the entire population. However, this sampling method is common in business & management research as it ensures a high response rate whereas probability sampling involves a lot of difficulty and costs (**Bryman & Bell 2003**)

   iii. **Sample Size** – 200.

   iv. Respondents' attitudes and perceptions is measured on a five-point Likert scale. We have evaluated customer's attitude using Likert-type scale as it has been used by researchers for over three decades with persuasion. Likert scale reports very reliable and satisfactory data (**Likert, R., 1932).**

   v. Before sending questionnaires ,pre-test is done. The purpose of the pre-test is to identify if the questionnaire can provide all the information as expected by the researchers (**Aaker et al. 2004, p327**).Another concern is to check if there any ambiguity exists and does questionnaire takes long time.

c) **Data Description:** The data set includes information about:

   i. **Demographic information of consumers** – age, gender, marital status, occupation , income, and family size.

   ii. **Car specific details**

   - **Usage of car** – shopping, picking up family members, short drive between the cities, social activities, commute to work, load luggage.
   - **Rating on importance while buying the car** - Brand, Price, Safety, Comfort, Mileage ,Engine, Colour , Equipment , Interior design, Exterior Design , Resale value.

iii. Amount spent , Transmission mode, Fuel Type.

iv. **Pre-purchase and post purchase factors** – source of purchase ,influencers while purchasing ,electric car awareness variable ,dealer contact , payment mode and satisfaction with car and after sales service.

d) **Statistical Techniques Used:** Multi-Class Classification Techniques such as Decision Trees, Random Forest, Naive Bayes and KNN etc. is used to develop statistical models that predicts the car model preferences. And for dimension reduction Random forest algorithm is used.

i. **Decision Tree Classifier –** Under this algorithm data is splitted continuously according to certain parameters. It has two entities decision nodes and leaves. Final outcomes are called leaves whereas where data is splitted are known as decision nodes.

- Simple to understand and effective.
- Based on one feature it splits the data at every node.
- Uses Gini Index as measure for split.

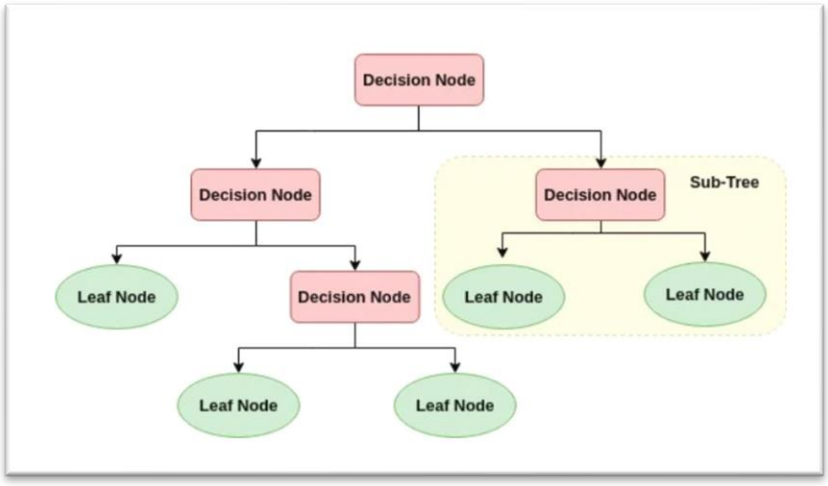


Figure 1

ii. **Random Forest Classifier –** As the name suggests this algorithm is made up of multiple decision trees on data samples and by means of voting selects the best solution.

- It helps to avoid overfitting problem by averaging out all results.

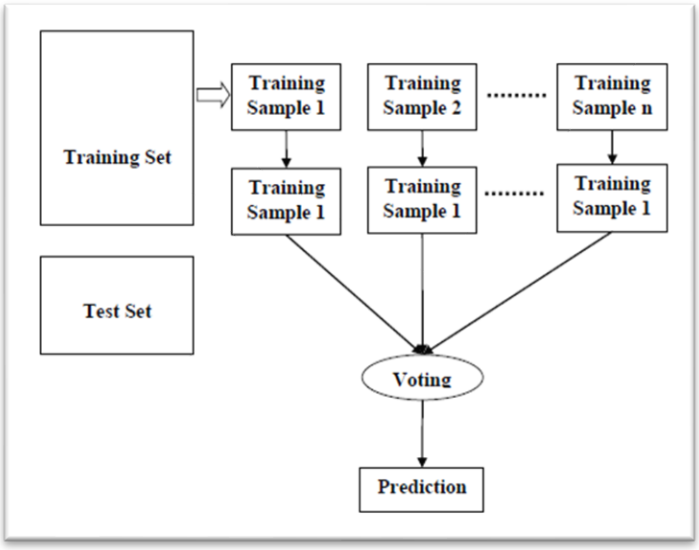


Figure 2

iii. **Support Vector Machine –** This algorithm can be used for both regression as well as classification problems. The main aim of this method is to create a decision boundary called as hyperplane which segregates the n-dimensional space into classes(where n is number of features) so that new data point is placed into the right category.

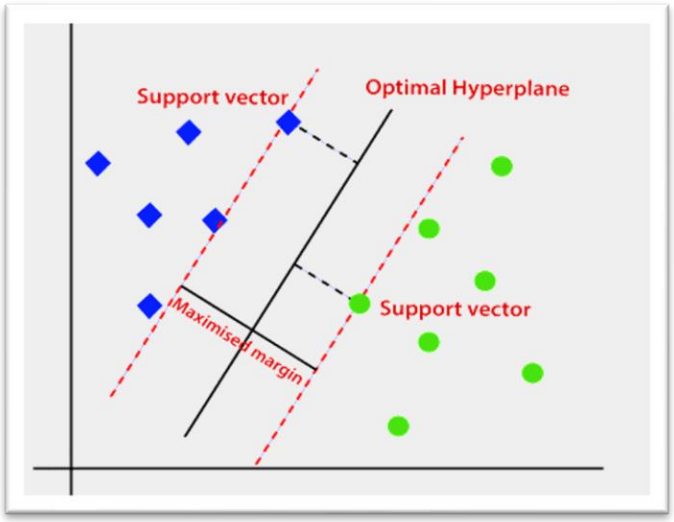


Figure 3

iv. **K-Nearest Neighbours –** It is a non-parametric algorithm and uses feature similarity function in predicting the values of new data points. For this, it first assigns the value of k i.e., the nearest data points. Then calculates the distance between the test data with all rows of training data and sorts in ascending order, and then finally assign the class to test data point based on most frequent class of these rows.

Figure 4

v. **Boosting Algorithms –** It is one of the techniques that uses ensemble learning method. These algorithm converts weak learners to strong learners.it combines various models and then generates the output.

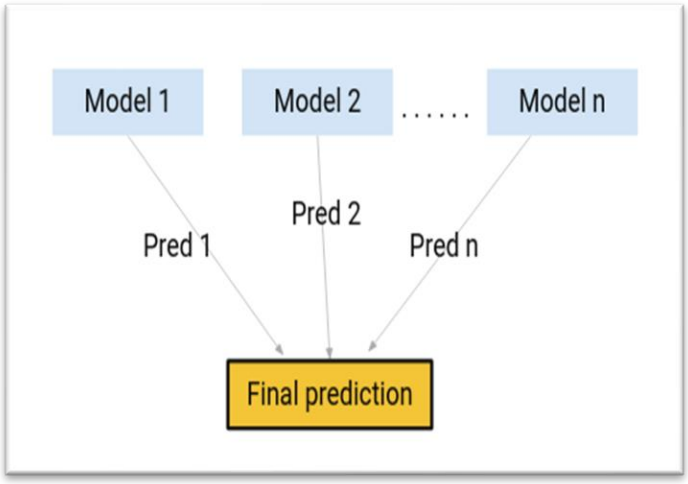


Figure 5

e) **<u>Fieldwork details:</u>** Face-to-face interviews with the owners are also conducted to collect data through questionnaire. Any personal feedback and suggestion from their side is also noted for future consideration.

DATA ANALYSIS

a. Exploratory Data Analysis

The main aim to conduct EDA is to clean the dataset, understand the variables and identify the relationships between them.

- Methods/Techniques: Univariate Analysis, Bivariate Analysis using Python language.
- Software Tools: MS Excel, Anaconda Navigator - Jupyter.

b. Predictive Analysis

The Churn prediction modeling techniques attempts to precisely comprehend customer behaviors and attributes which give signs about the risk and timing of customer churn.

- Methods/Techniques: Decision Tree, Random forest Classifier, Support Vector Machine(SVM), K-Nearest Neighbour(KNN), Gradient Descent algorithm and 2 other boosting algorithms under ensemble method.
- Software Tools: MS Excel, Anaconda Navigator - Jupyter.

a. Exploratory Data Analysis

Univariate analysis

i. Demographic Information of the Respondents

Characteristics	Measuring Groups	Frequency	Percentage
Gender	Male	152	76%
	Female	48	24%
Age	20-35	76	38%
	36-50	96	48%
	51-75	28	14%
	75 & above	0	0%
Marital Status	Single	62	31%
	Married	138	69%
Education level	Senior high /technical/professional school	44	22%
	College/university	69	35%
	Post graduate or above	87	44%
Occupation	Employed	71	36%
	Self - Employed	71	36%
	Student	83	42%
	Retired	7	4%
Annual Income	Upto 3 lacs.	52	26%
	3,00,000-8,00000	36	18%
	800,000-12,00,000	68	34%
	12,00,000-18,00,000	12	6%
	18,00,000-23,00,000	26	13%
	Above 23 lacs.	6	3%
Family Size	2 Members	24	12%
	2-4 members	131	66%
	4-10members	45	23%

Table 1

Observations: Among the total passenger cars customers majority of the customers are:

a. Male i.e.76%.
b. Belonging to age group 36-50 i.e. 48%.
c. Married i.e.69%
d. Post graduated and other high degree education i.e. 44%.
e. Belonging to Student class group i.e. 42%.
f. Having income between 8-12 lacs i.e. 34%.
g. Belonging to small and nuclear family size (2-4 members) i.e. 66%.

ii. Importance of Each Car Attribute

Attribute	Mean Values
Brand	4.355
Competitive Price and Discount	4.46
Safety	4.63
Comfort	4.625
Mileage	4.53
Engine	4.21
Color	3.835
Equipments	3.905
Exterior Design	4.005
Interior Design	4.025
Resale Value	4.04
Better After Sales services	4.475

Table 2

Observations

- From this result we can see that "Safety" is the most important car attribute that customers see while purchasing the car.

- Secondly, buyers take "Comfort" as the second most important car attribute when buying a car. This is probably because in our survey 70% are married and approximately 90% have family size of more than 3 people.

- Thirdly ,buyers take " Mileage" as the third most important car attribute. This is because Indian people have tendency to save money and be environment friendly.

iii. Car users according to car body type

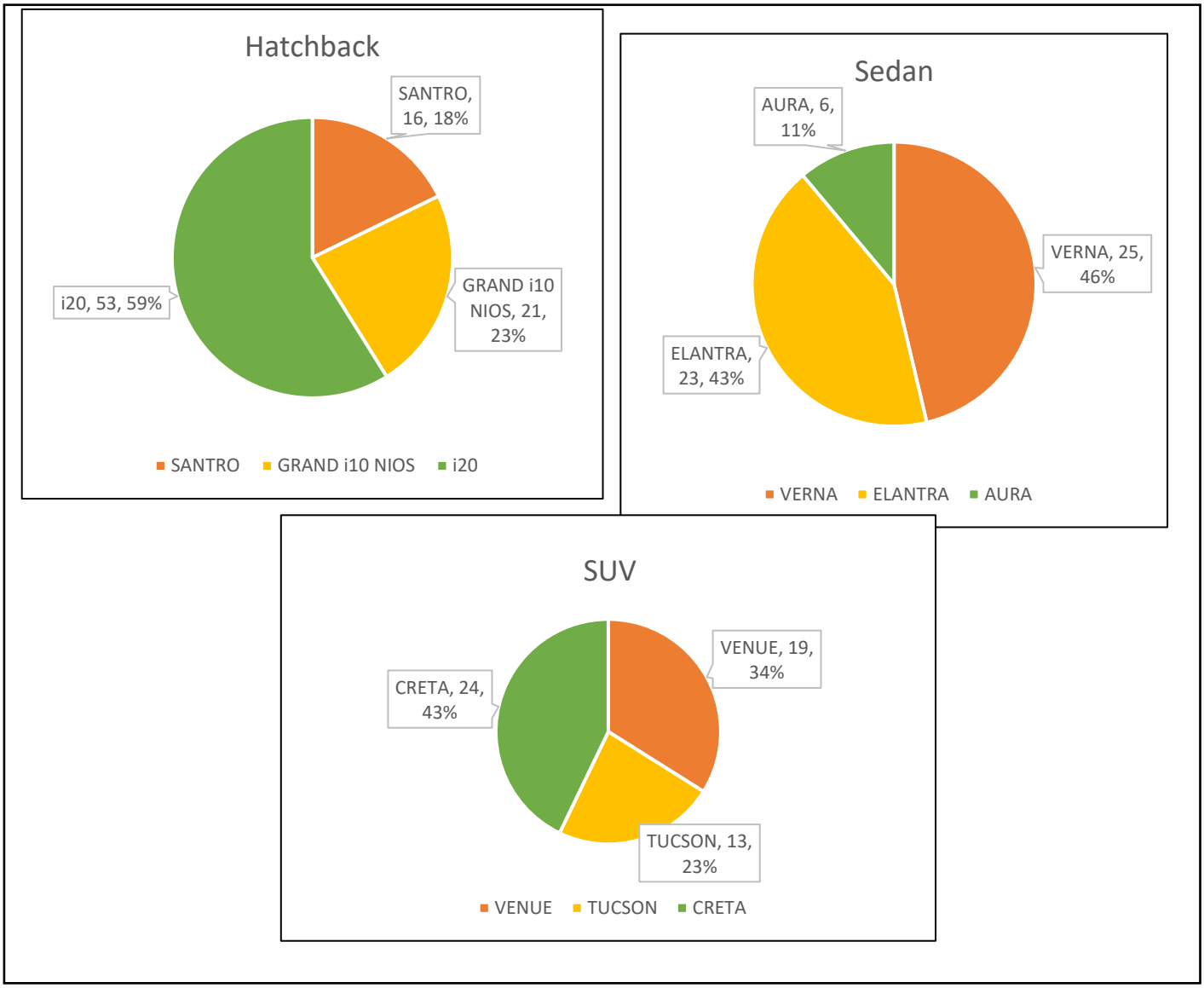


Figure 6

Observations: Highest share is clearly of the most preferred car – i20 among all kinds of cars.

- **Hatchback** – after i20 ,Grand i10 NIOS is preferred over Santro. i20 is feature-rich, power-packed hatchback leading to many variants for all types of buyers and hence reason for its high sale.
- **Sedan** – Verna is the most preferred car followed by Elantra, then Aura. Many people attached comfort and safety to verna model and so its final purchase.
- **SUV** – Creta is the most preferred car followed by Venue and then Tucson. There are many reasons that worked in favour of its sales are:
  a. Pricing – 9-18 lac variants and so able to lure buyers from low as well as high class segments.
  b. Drivetrain options - three types of engine options with four gearbox options, add many varieties to this model.
  c. Brand Value

d. Designs
e. Features

iv. Sources of Information used before purchasing a car

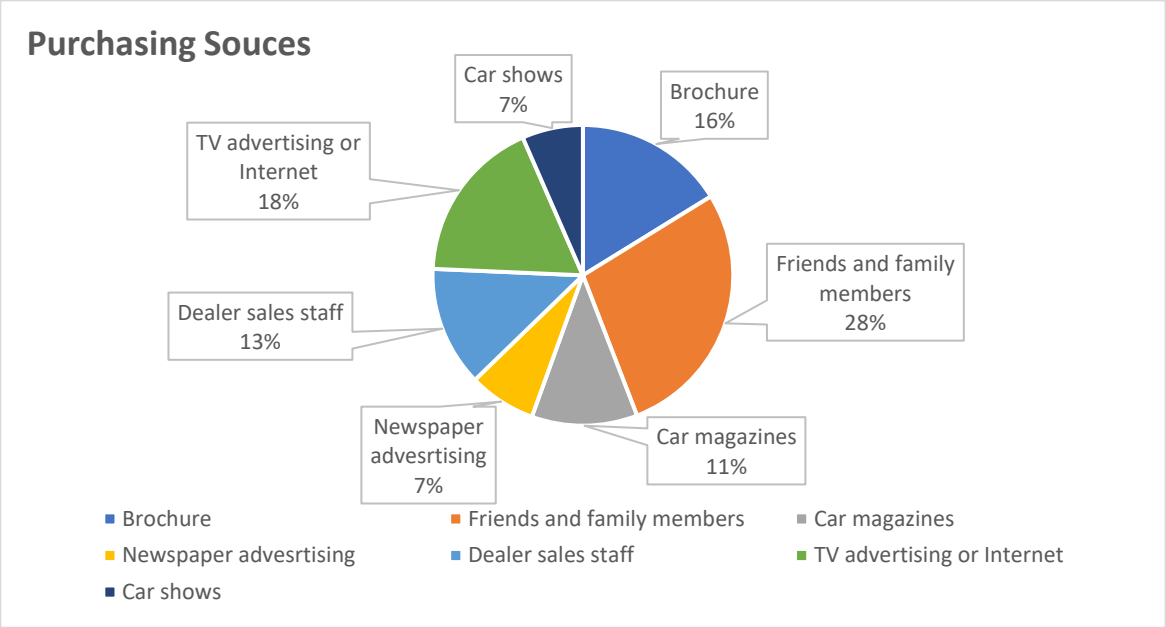


Figure 7

Observations: Friends and family members are the major source of information and many people do take their consideration before making buying decision. And this is followed by TV advertisement and internet due to presence of advanced social media marketing available nowadays.

v. Decision Influencers while buying a car

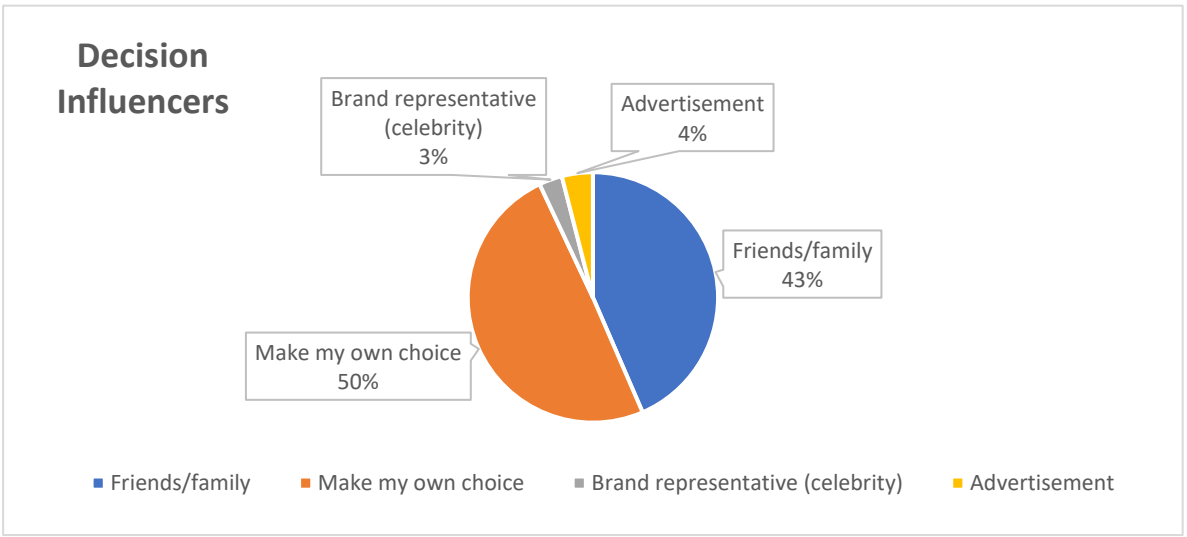


Figure 8

Observations: Half of the people makes their own choices while final buying decision followed by friends and family members and then advertisement. This suggest that if customer is able to fetch quality information from reliable sources with less time he would purchase the car in short period of time.

Bivariate Analysis

vi. Customer Age and Car model

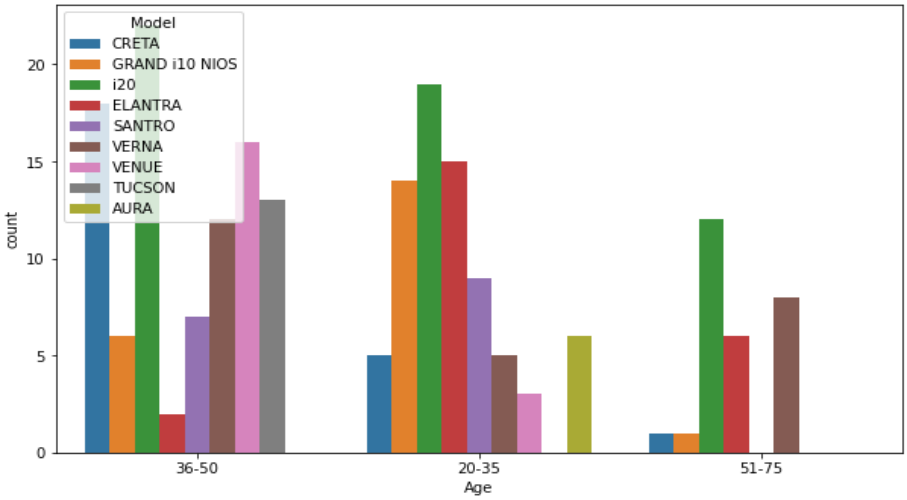


Figure 9

Observations

- i20 is the most purchased car among all age groups.
- In the age group of 20-35 Elantra and Grand i10 NIOS is highly preferred.
- In the age group of 36-50 Creta then Venue is more preferred than other models.
- In the age Group of 51-75 sedan is chosen after i20. Verna and Elantra is more preferred than other models.

vii. Family Size and Car model

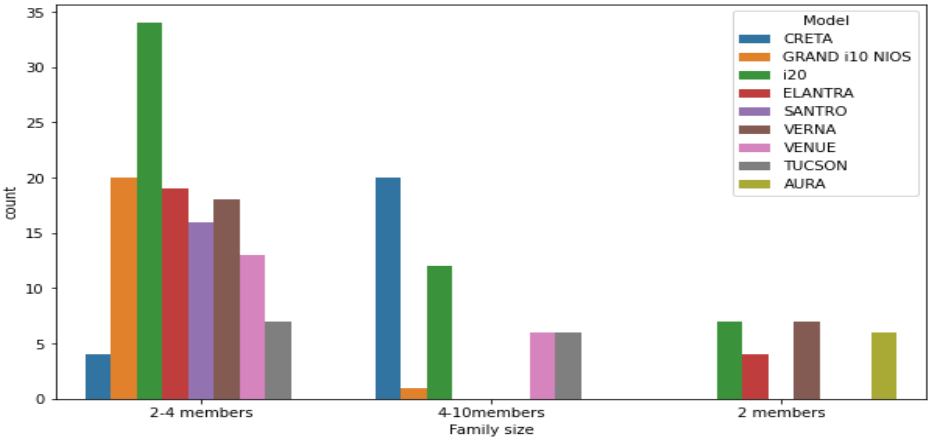


Figure 10

Observations

- In all types of family again i20 is the most preferred car because of many variants available in the market serving all types of buyers.
- Family of 2 members chose either hatchback or Sedan – i20,Verna,Aura and Elantra.
- Family of 2-4 members has highest no. of users of i20 followed by Grand i10 NIOS, Elantra, Verna etc.

viii. Occupation and Car model

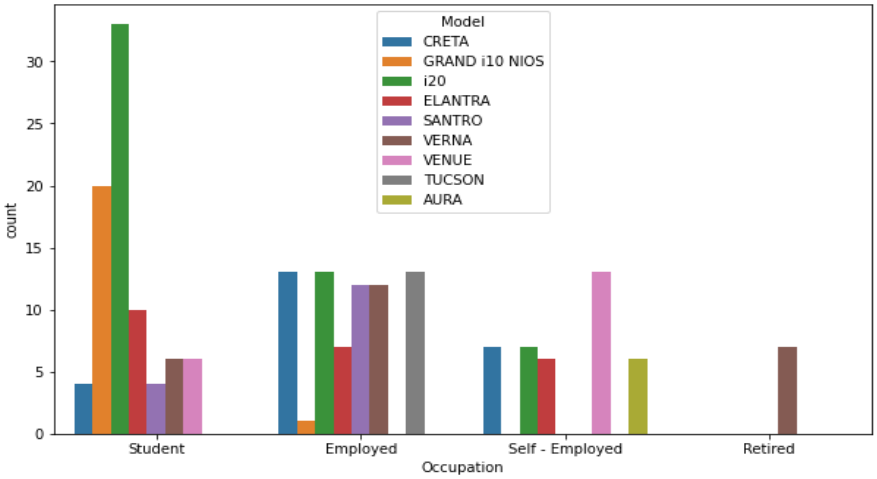


Figure 11

Observations:

- Among students, hatchback body type is the most preferred where i20 comes at the top followed by Grand i10 NIOS and Elantra.
- Most of the Employed people uses Creta,i20 and Tucson in their cars.
- Most of the self-employed people uses SUV-Venue model over others.

ix. Annual Income and Car model

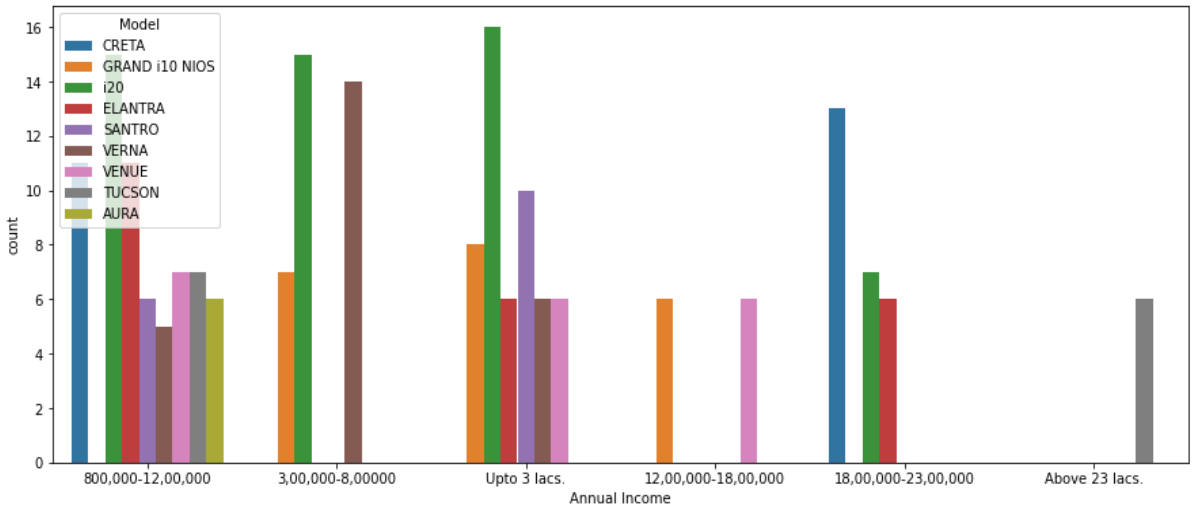


Figure 12

Observations: Customers with income between 3-12 lacs prefers all types of models in hatchback, Sedan and SUVs. But people with income more than 23 lacs mostly purchases Tucson model.

x. Fuel Type and Car model

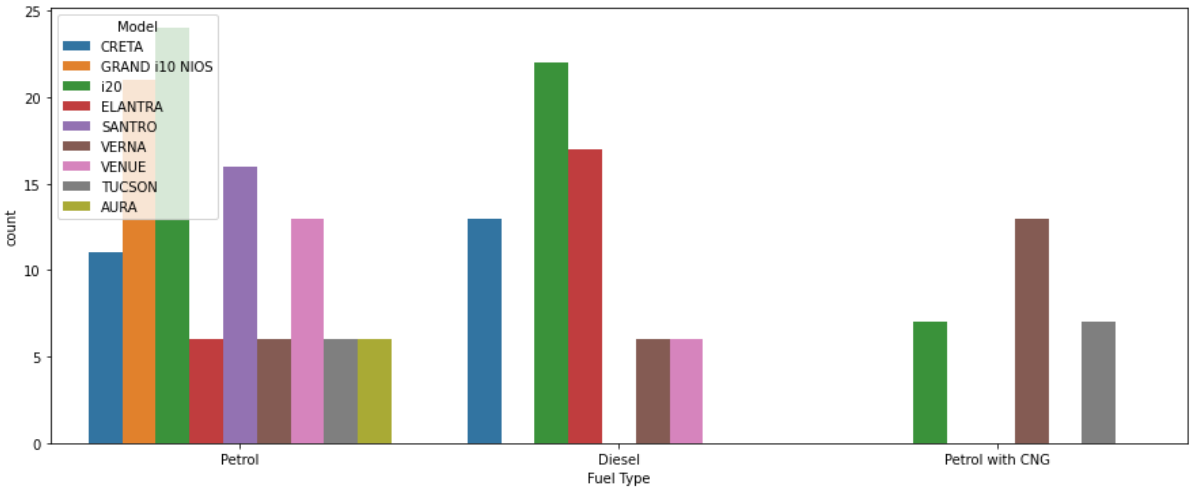


Figure 13

Observations: Petrol type is preferred in all types of car models .And only 5 models are preferred in Diesel type and those are : Creta,i20,Elantra,Verna and Venue while in Petrol with CNG Models only 3 models are preferred by customers i.e., Verna, Tucson and i20.

xi. How long information is collected and Amount spent.

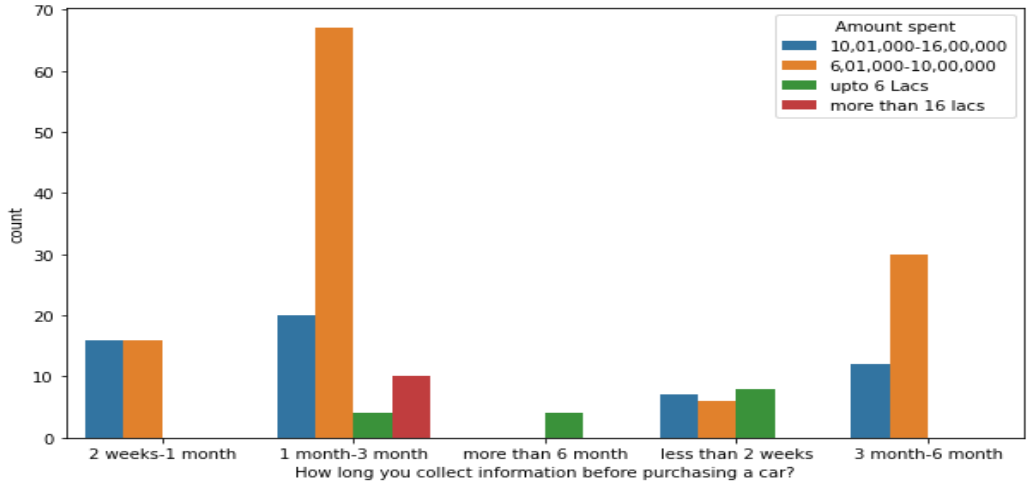


Figure 14

Observations

- Price sensitive people invest more time in collecting information before making a

purchasing decision i.e., if they are buying a car up to 6 lacs rupees they take more than 6 months of time.

- Then comes the range of 6-10 lacs cars where customers take 1-3 months in collecting the information and so on.

xii. Payment Mode and Amount Spent

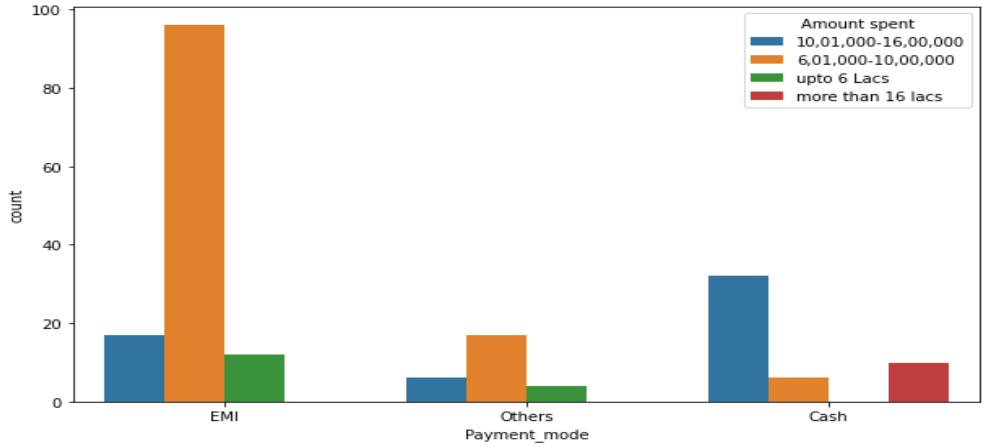


Figure 15

Observations: Most of the users in India buy cars on EMI because of easy finance options available at low price and hence we can see that most of the cars ranging from 6-16 lacs are bought on EMI option and very less on Cash.

xiii. Gender and Transmission mode used.

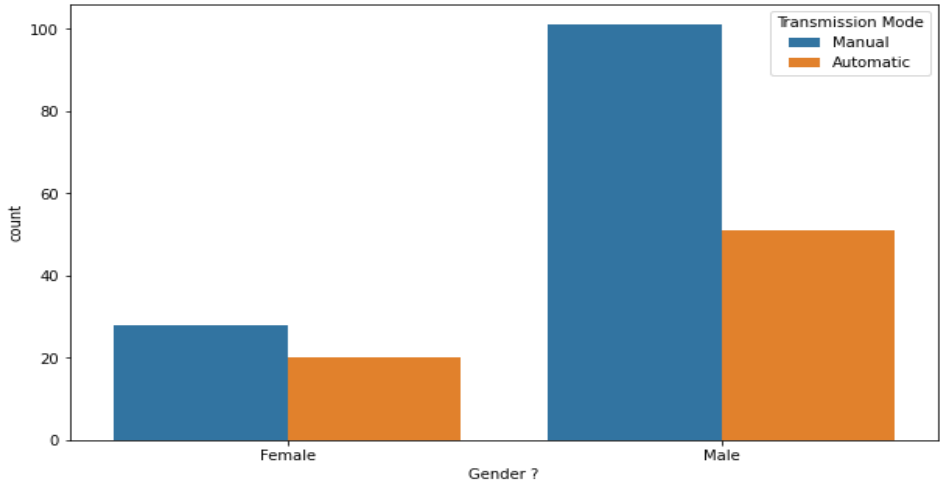


Figure 16

Observations: Generally, the car enthusiasts prefer manual transmission mode and so among males 65% uses Manual transmission mode while among female 60% uses manual mode.

xiv. Electric car awareness and its future use

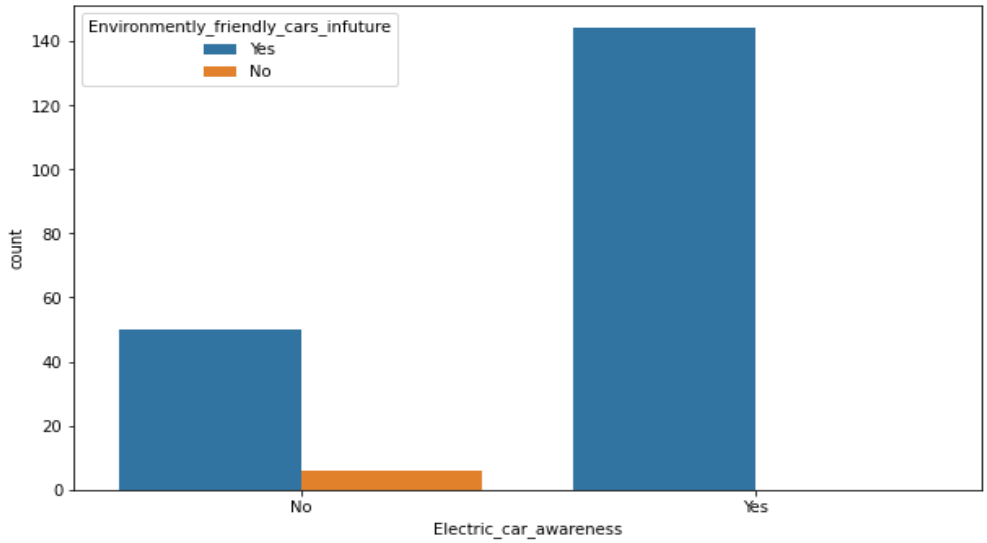


Figure 17

Observations: Almost 80% people are aware of electric cars and 20% are unaware but among all of them 94% will prefer in future to buy it in future.

Multivariate Analysis

xv. Correlation of car features' importance as perceived by customers.

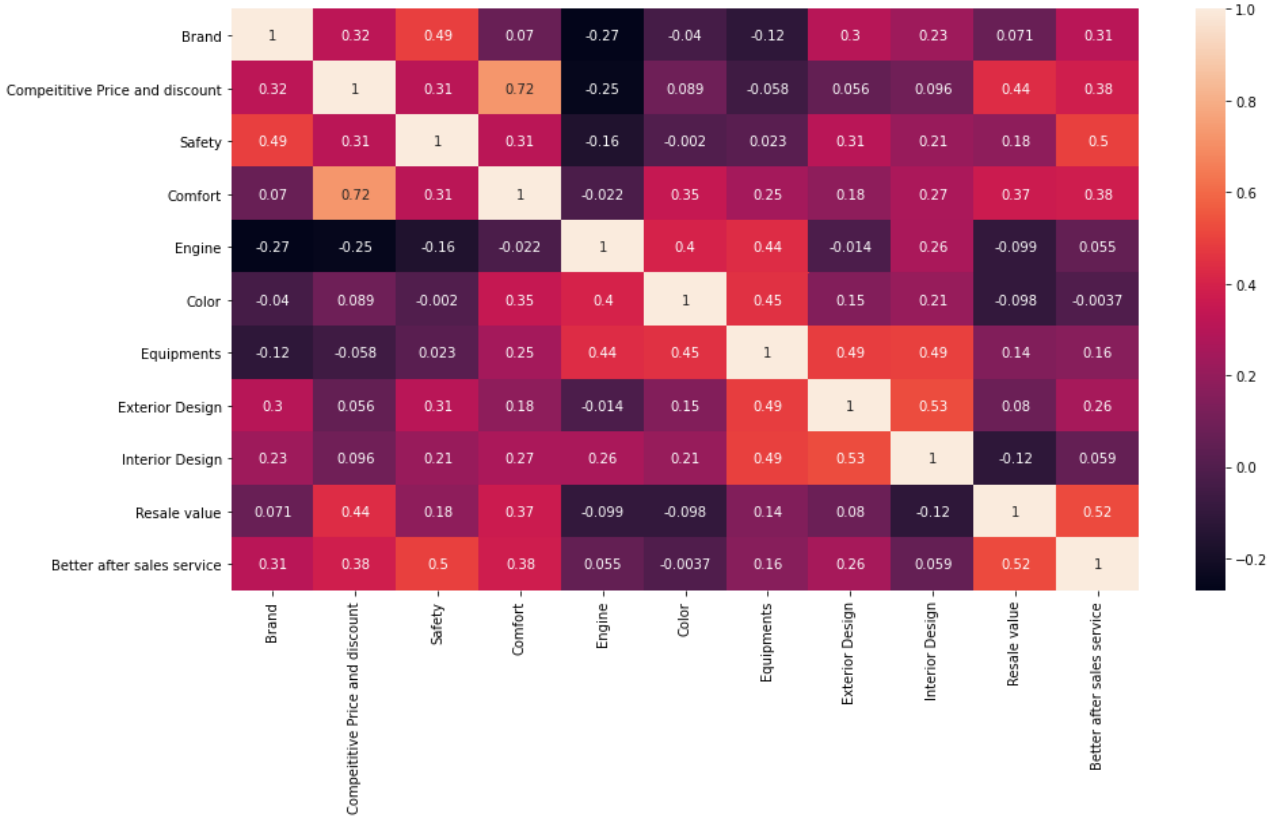


Figure 18

Observations

a. "Brand" and "Engine" shows negative correlation and so importance of brand increases there will be decrease in importance of engine as they rely on the Hyundai brand.
b. There is negative correlation between "Competitive Price and discounts" and "Engine" as if they are giving more importance to competitive price and discounts, they will not give much importance to engine used in the car.
c. We can observe that there is positive correlation between "Competitive price & discount" and the "Comfort" demanded by customers as we can see the i20 is the most preferred car among all Hyundai models.

b. Predictive Analysis

i. Flow of the Analysis

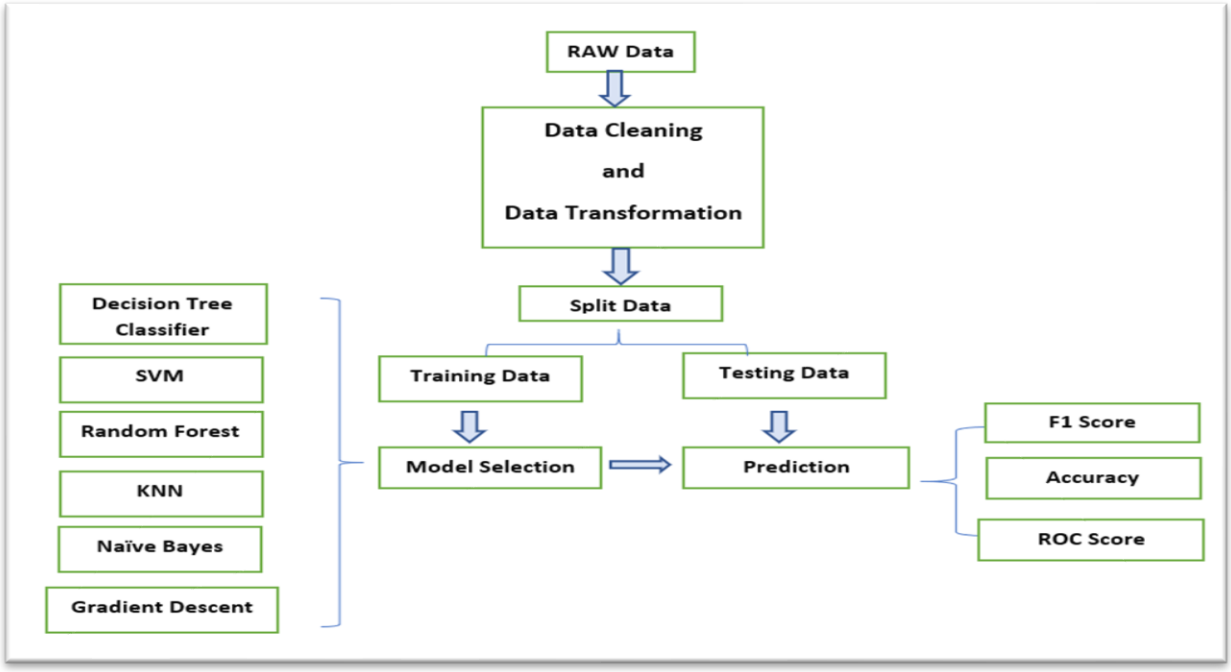


Figure 19

ii. Feature Selection - To know which features among 34 are important, Random forest algorithm is used and following result is generated –

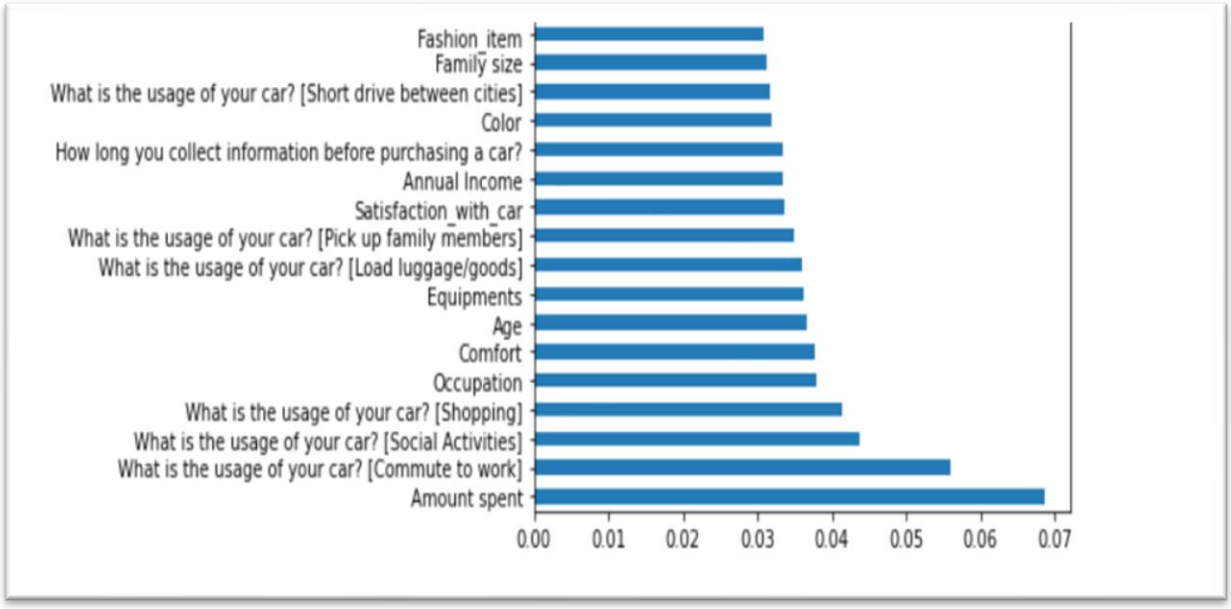


Figure 20

Observations

a. We can observe from the Random forest algorithm that the Amount spent,Commute to work , Social activities are the most important predictor variables to predict the car model.

b. In total they are 17 predictor variables majorly impacting our target variable that is car model.

iii. Predictive Machine Learning Models for Classification - To predict the car models that will be preferred by prospective customers various Supervised Machine Learning algorithms are deployed to know which performs the best using various metrics.

Model Building -

a. Decision Tree Classifier
b. Random Forest
c. Naïve Bayes
d. Support Vector Machine
e. K-Nearest Neighbour
f. Gradient Boosting
g. Light GBM
h. CatBoost

iv. Model Comparison -

a. **Training Score** – It shows how well our model is generalized and fitted on training data.
b. **Testing Score** – when our model is ready this score is calculated to check how well it is working on real data set ,higher the score better is our model.
c. **Accuracy** – It is the ratio of correctly identified observations to total observations.
d. **F1-Score** – It is the weighted average of precision and recall and works best on uneven class distribution as in our study of car models.
e. **ROC-AUC Score** – it is one of the best metrics to evaluate the multi-class classification models as it explains how much our model is able to distinguish among the classes.
f. **Hyperparameter Tuning and Optimal Hyperparameters** – Grid Search CV is used for hyperparameter tuning, the optimal hyperparameters obtained from this method are then again applied to all algorithms to implement the best model architecture.

Algorithm	Training		Testing		Hyper Parameters Tuned	Optimal hyperparameters	AUC
	Accuracy	F1 Score	Accuracy	F1 Score			
Decision Tree	72.50%	72.50%	75%	75%	{'criterion': ['gini','entropy'], 'max_leaf_nodes': list(range(2, 50)), 'min_samples_split': list(range(2, 5)), 'max_depth': [2,3,4]}	{'criterion': 'entropy', 'max_depth': 4, 'max_leaf_nodes': 14, 'min_samples_split': 2}	88.10%
Random Forest	84.3	84.37%	77.50%	77.50%	{'n_estimators': [10,20,50,100,200],'min_samples_leaf': [5,10,15,20],'min_samples_split': [4, 6,8,10],'max_depth': [5,6,7]}	n_estimators=200,min_samples_leaf=5,'min_samples_split': 4,'max_depth': 5	88.10%
Naïve Bayes	50.63%	50.63%	40%	40%	-	-	73.00%
SVM	85.63%	85.63%	82.50%	82.50%	{'C': [0.1, 1, 10, 100, 1000],'gamma': [1, 0.1, 0.01, 0.001, 0.0001], 'kernel': ['rbf']}	{'C': 1, 'gamma': 1, 'kernel': 'rbf'}	92.39%
KNN	77.50%	77.50%	52.50%	52.50%	{'K':[2,3,4,5,6,7,8,9,10,11,12]	K=7	79.50%
Gradient Boosting	85.63%	85.63%	82.50%	82.50%	{"learning_rate": [0.01, 0.05, 0.1, 0.2],"max_depth":[2, 3, 5, 8],"max_features":["log2","sqrt"],"n_estimators":[10,20,50,100]}	{'learning_rate': 0.05, 'max_depth': 25, 'n_estimators': 200, 'num_leaves': 300}	91.79%
Light GBM Boost	85.63%	85.65%	85%	85%	"max_depth": [25,50, 75],"learning_rate" : [0.01,0.05,0.1], "num_leaves": [300,900,1200], "n_estimators": [200]	{'learning_rate': 0.05, 'max_depth': 25, 'n_estimators': 200, 'num_leaves': 300}	92.39%
Cat Boost	84%	84%	78%	78%	{'learning_rate': [0.03, 0.1],'depth': [4, 6, 10],'l2_leaf_reg': [1, 3, 5, 7, 9], 'iterations':(10,50)}	{'depth': 6, 'iterations': 50, 'l2_leaf_reg': 5, 'learning_rate': 0.1}	88.10%

Table 3

Observations : SVM Model scores highest in ROC-AUC score with 92.39% and the training and testing accuracy is far good with 85.63% and 82.50% accuracy and F1 score respectively.

The hyperparameters that are tuned for each algorithm with its respective optimal hyperparameters is shown and then that optimal hyperparameters are again placed into their respective models to obtain their accuracy and f1 score metrics to later compare all models.

Hence SVM Model and then LIGHT GBM Boosting Algorithm is best and reliable model that can be deployed to predict the car model preferences by the prospective customers in future.

CONCLUSIONS

For Hyundai motors, to increase its market share it is critical to understand the needs and preferences of consumers.

a. Existing customers are moderately satisfied with the product and after sales services and these customers being the major source of information , do also have potential to influence decision of new buyers. As satisfied customers are assets to any organization, hence Hyundai Motors should **provide better after sales services and other support to its customers** which will lead to more loyal and new customers.
b. As internet being the 2nd most preferred source of information and half of the people make their own choice at final step, therefore it becomes important for the company to create readily and reliable information available through websites, Hyundai Care application , social media, and other sources. It also means that car manufacturers and dealers should think about **innovative digital marketing strategies** that should be implemented in different decision-making phases of the prospective customers. For example, getting information about the customer through queries generated on any platform, it can predict the car models that customer may buy and hence suggest only those cars' details through personalized marketing strategies which will lead to quick purchase as compared to general marketing strategies.
c. Company should put special focus on "**safety**" and "**riding comfort**" in Indian market, for example promoting in advertising to attract customers. These aspects along with mileage should also be considered while building new models as customers do give more weightage to these features while buying the cars.
d. Company can release the **electric models** in future with confidence and expect its existing as well as prospective customers , that they will buy their product as from the study it is found that irrespective of awareness of electric car people will buy the product.
e. More than half of consumers uses EMI facility to purchase the car ,so company should **tie up with more finance providing firms** and promote the product not only on basis of features but also easy payment options available which will lead to quick purchases.
f. Through EDA and Predictive Analysis, it was found that there are total seventeen features that influences the customers buying behaviour so keeping those predictor variables in mind company can retain and acquire its new customers. The prediction models should have high accuracy and ROC-AUC score to be reliable and implementable , therefore **Support Vector Machine and Light GBM Boosting are found to be most reliable Machine learning model** of classification.

RECOMMENDATIONS

a. It is important to know the psychographic profile of customers too along with the demographic profile of the buyers which will help the company to better understand their attitude , behaviour, mindset and emotional connect with the car as well with the company.

b. Advanced prediction models can also be deployed like Extra tree Classifier , Neural Networks etc.

REFERENCES

- Raghu,G. (2013) - *Car Market and Consumer Behavior - A Study of Consumer Perception.*
- John,B., and Pragadeeswaran,S. (2013) - A study of small car consumer preference in pune city.
- Malhotra and Jain (2017) - *Buying Preferences of consumers for small passenger cars: Maruti Suzuki Ltd. & Hyundai Motor India.*
- Belgiawan,P. , Schmöcker,J. & Zeid,M. (2017) - *Analysis of car type preferences among students.*
- Hoen, A., and Koetse, M.J.(2014) *A choice experiment on alternative fuel vehicle preferences of 25 private car owners in the Netherlands.* Transportation Research Part A, Vol. 61,pp. 26 199–215.
- Armstrong, M.(2018) - Most important factors when buying a car[online]Available at: https://www.statista.com/chart/13075/most-important-factors-when-buying-a-car/
- Kumar,S. and Dixit,P. (2019) *Identify the Factors Influencing Customer's Preferences for Segment of the Cars* [online] Available at : http://ignited.in/I/a/211172
- Ericsson Blog (2020) *consumer trends in the automotive industry: is this the end of the consumer car?* [online] Available at : https://www.ericsson.com/en/blog/2020/12/consumer-trends-automotive-industry-2020
- Consumer Trends for the Auto Industry (2020) [online] Available at : https://www.brandwatch.com/blog/consumer-trends-auto-industry-2020
- Belgiawan et al.,.(2016) Modelling social 5 norms: Case study of students' car purchase intentions. Working paper.
- Yang, D., and Timmermans, H.(2013) *Analysis of influence of fuel price on individual activity-travel time expenditure*. Transport Policy, Vol. 30, pp. 40–55.
- Higgins, Christopher, D. & Mohamed (2017) *How vehicle body type affects consumer preferences for electric vehicles* Transportation Research Part A: Policy and Practice, Elsevier, vol. 100(C), pages 182-201.
- Nayum, A. and Prugsamatz.,S.(2013*). Influences of car type class and carbon 19 dioxide emission levels on purchases of new cars:* A retrospective analysis of car purchases in 20 Norway. Transportation Research Part A, Vol. 48, pp. 96–108.
- Likert, S. (1932) – Article: *A Technique for the Measurement of Attitudes,* in Achieves of Psychology.
- Bryman, A. and Bell, E (2003) , Business Research Methods ,Oxford University Press.