IMPACT OF COVID-19 ON CAR SALES IN INDIA

A PROJECT REPORT

SUBMITTED TO THE UNIVERSITY OF CALICUT IN A PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF THE DEGREE OF

MASTER OF SCIENCE IN STATISTICS

Submitted by

AJAY RAJ C

Reg.No:FKAUMST003

DEPARTMENT OF STATISTICS

FAROOK COLLEGE(Autonomous) CALICUT

UNIVERSITY OF CALICUT

UNIVERSITY OF CALICUT

DEPARTMENT OF STATISTICS

FAROOK COLLEGE(Autonomous)

CALICUT

CERTIFICATE

This is to certify that **Ajay Raj C** has done project work entitled **IMPACT OF COVID-19 ON CAR SALES IN INDIA** for the partial fulfillment of the requirements for the award of degree of Master of Science in Statistics of Calicut University under my supervision and guidance

Dr.Haritha.N.Haridas
Assistant Professor
Department of Statistics

Name of Examiners:

1

2

FAROOK COLLEGE

Date:

DECLARATION

I hereby declare that the project entitled **IMPACT OF COVID-19 ON CAR SALES IN INDIA** is a record of original work done by me under the supervision and guidance of Dr.Haritha.N.Haridas, Assistant Professor, Department of statistics, Farook College, Calicut in partial fulfilment of the degree of Master of Science in Statistics of Calicut University.

Ajay Raj C
M.Sc. Statistics

Farook College, Calicut

ACKNOWLEDGEMENT

On the very outset of this report, I would like to extend my sincere and heartfelt obligation towards all the personages who have helped me in this endeavour.Without their active guidance,help,cooperation and encouragement I would not have made headway in the project.

I record my profound thanks to Dr. K M Naseer, Principal, Farook College and the management for giving me an opportunity to persue M.Sc. degree in this prestigious institution and to undertake this project.

I extend my sincere thanks to Dr. R.M. Juvairiyya, HOD of Statistics department who provided all facilities and necessary encouragement during the project.

I am extremely thank full and pay my gratitude to Dr.Haritha.N.Haridas for her valuable guidance and support on completion of this project.

I also acknowledge with a deep sense of reverence,my gratitude towards my parents who always support me morally as well as economically.At last but not least gratitude goes to all my friends who directly or indirectly helped me to complete this project report.

Any omission in this brief acknowledgement does not mean lack of gratitude.

Ajay Raj C

M.Sc Statistics

Farook College,Calicut

IMPACT OF COVID 19 ON

CAR SALES IN INDIA

CONTENTS

CHAPTER 1

CHAPTER 2

Objective and Tools used in Study

Tools and Software

CHAPTER 3

Terminology

CHAPTER 4

ANALYSIS

Time Series Analysis

CHAPTER 5

ARIMA Modeling Analysis

Multiple Linear Regression

CHAPTER 6

Conclusion

CHAPTER 1

Introduction

Car manufacturing is one of the major industries in India, it contributes about 7 percent of the country's GDP with an annual turnover of 7.5 lakh crore and export of Rs.3.5 lakh crores. It supports the development of other industries by the procurement of raw materials like steel,metal,rubber,glass and so on.The automotive and mobility industries have certainly been among the hardest hit during the COVID-19 pandemic. Auto sales had been dropping for 12 to 15 months when the outbreak stalled production and overall economic activity. The industry now faces concerns on short-term liquidity as well as long-term growth in revenue and profitability, even as automakers restart production and dealerships record sales again with a gradual relaxation of the lockdown. As it emerges from this crisis, the industry will need to realign itself to some of the new realities of the post COVID-19 world that are outlined here.

Customers aspiring to buy entry-level cars or those looking for an upgrade to such models are postponing decisions as the COVID-19 pandemic has hit their income sentiment significantly, a report said on Monday. At the same time, sales of premium segment cars are expected to go up on account of resilient income of affluent buyers, while the share of higher-priced two-Wheeler's is likely to remain around 40 per cent.

In India the leading car manufacturer is Maruthi Suzuki India Limited, a subsidiary of Suzuki Motor Corporation,Japan. The eight Maruthi Suzuki models contribute over 8 percentage of the top 10 selling models and above 60 percent of total car sales in India is contributed by Maruthi. Ayukawa, the MD and CEO of Maruthi-Suzuki India Limited said in an online event(SIAM's 61st Annual convention) that even before COVID-19 the Indian Automotive Industry id facing a deep structural breakdown. The COVID-19 Pandemic has cased negative growth for the industry. Some of the immediate short-term concerns for the industry are pandemic-related uncertainties and health our people,global shortage of semiconductors,rising commodity prices,upcoming fuel-efficiency and Bharat-Stage VI(BS-VI) phase 2 regulations,short of shipping containers and import restrictions.

During COVID-19 the main problem people faced was transportation since public transportation were banned due to possibility of a COVID cluster. So people must use their private vehicles for mode or transportation and certain taxi were also available. The Showrooms were closed mostly due to COVID-19 restriction ad buffer zone or employee was tested COVID-19 and the showroom is closed after sterilizing which may take 5-10 days. Also due the problems mentioned above the available cars in showrooms were very short so the waiting period was very long

Time series is an important field in Statistics.It is used for non-stationary data that are constantly fluctuating over time or are affected by time. What sets time series data from other data is that the: analysis can show how variable changes over time. In other words,time is crucial variable because it shows how the data adjusts over the course of the data point as well as the final results. Time series data can be used for forecasting-predicting future based on the historical data. Time series forecasting is the process of analyzing time series data using Statistics and modeling to make predictions and inform strategic decision-making When Organizations analyze data over consistent intervals,they can use time series forecasting to predict the likelihood of future events. Time series forecasting is a part of predictive analytic.One of the most important algorithms in time series is Holt-Winters and and ARIMA modeling

A Study was done by lead data scientist Yuxin Zhao presently working as Associate consultant at Booz Allen Hamilton consulting firm with expertise in analytics, digital, engineering, and cyber, which help businesses, government, and military organizations transform. The analysis was done by R software to forecast the monthly sales of Volkswagen Lavida from the date April-2011 to October-2013. The forecasting used there was ARIMA forecasting. Tests like ADF and Ljung box test is used for testing stationarity

Regression analysis is a statistical method used for estimating the relationship between a depended variable to one or more independent variable. This includes several methods such as linear,multiple linear,logistic regression models.It is mainly used in finance,investments... . The benefit of regression analysis is that it allows you to analyse the data and make better decisions. The most common use of regression analysis in business is for forecasting future opportunities and threats. Demand analysis, for example, forecasts the amount of things a customer is likely to buy. When it comes to business, though, demand is not the only dependent variable. Regressive analysis can anticipate significantly more than just direct income.

We are using R software for the analysis in this project because R is one of the latest cutting-edge tools. R has a power-full package for data calculation, statistical analysis and data mining. Today, millions of analysts, researchers, and brands such as Facebook, Google, Bing, Accenture, and Wipro are using R to solve complex issues.

CHAPTER 2

OBJECTIVES AND TOOLS

OBJECTIVE

The main objective of this study is to analyze the car sales data in India during COVID-19. The data we collected is from May-2020 to Feb-2022. We are doing forecasting the next 5 months data using R software. Also we analyse the customer buying preference over car after COVID-19, here we are considering the first 10 cars in the ranking order of number of unit sold in May-2022. We use Multiple Linear Regression model to find out the significant factors that affect the sales.

Tools and Software

To achieve our objective of our study we need various statistical tools and software like Excel,RStudio etc. A brief description of these are given below.

MS Excel

Microsoft Excel is the industry leading spreadsheet software program, a powerful data visualization and analysis tool. It has been build for Windows,Mac and android Operating Systems.Excel is widely accepted software for saving and analysing data since the version 5 in 1993. It features calculations,graphing tools,pivot table etc.In our study we are collecting and storing our data in excel which is imported to our other software's for further analysis.

R Software

R software is a programming language for statistical computing and developed by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team. R is an open software which available free for all users and have a wide selection of packages for statistical analysis. The first official version of R software was released in June 1995, later the stable bet version was released in 29 February 2000. The current latest version 4.2.0. In this study we are using R Studio instead of R because RStudio is and IDE(Integrated Development Environment) that allows user to use R more steadily and user friendly.

Line graph

A Line graph or line chart is a graph with each data series is plotted as separate line. It is use-full to represent a time series to analyse the change of trend over time in time series.Excel has a built in line graph available. In our study we analyse these graphs through an online graph generator chartgo rather using from excel.The graphs generated from chartgo has more options available. It is similar to scatter plot except that the measurement points are ordered over interval of time, thus the line if often drawn chronologically.

Bar Chart

A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values they represent. It is a type if graph that uses the heights of the bars to convey the information. It is easy to draw conclusions from bar graph and gather information. The height of each bar represents the number of data values in that category(frequencies). Bar graph is generally plotted vertically or horizontally. As mentioned above we use chartgo for plotting bar graph.

Pie-Chart

A Pie Chart is is a type of graph that represent the data in circular graph. Where slices of the graph shows the size of data. It is one of the most commonly used graphs to represent data using the attributes of circles, spheres, and angular data to represent real-world information.The whole chart the total percent of the data. Since the chart is circular the total of the data is 360^o.When you interpret one pie chart, look for differences in the size of the slices. The size of a slice shows the proportion of observations that are in that group. When you compare multiple pie charts, look for differences in the size of slices for the same categories in all the pie charts.

CHAPTER 3

TERMINOLOGY

Time Series

A time series is a data set that tracks a sample over time. In particular, a time series allows one to see what factors influence certain variables from period to period. Time series analysis can be useful to see how a given asset, security, or economic variable changes over time. Time series can be tracked for short term such as the price of a security on the hour over the course of a business day, or the long term, such as the price of a security at close on the last day of every month over the course of five years ref. Time series can be classified into two stock and flow.A stock series is a measure of certain attributes at a point in time and can be thought of as stock-takes. For example monthly labour surveys is a stock measure because it takes stock of whether a person was employed in the reference week. Flow series are series which are a measure of activity over a given period. For example, surveys of Retail Trade activity. Manufacturing is also a flow measure because a certain amount is produced each day, and then these amounts are summed to give a total value for production for a given reporting period.

Stationary process

A Stationary process in time series is that the data does not depend at which the series is observed. In general a stationary time series will have no predictable pattern in long term. It does not mean that the series does not change over time, just that the way it changes does not itself change over time. For a stationary series we have constant mean and variance. Stationarity is importance because analytical tools and statistical tests and models rely on it.

Autocorrelation

Autocorrelation refers to the degree of correlation of the same variables between two successive time intervals. It measures how the lagged version of the value of a variable is related to the original version of it in a time series. Autocorrelation is also known as serial correlation.An autocorrelation of +1 represents a perfect positive correlation, while an autocorrelation of negative 1 represents a perfect negative correlation. The formula of Theoretical Autocorrelation is:

$$\rho_k = \frac{E(X_t - \mu)(X_{t-k} - \mu)}{\sigma_x^2}$$

The formula for sample Autocorrelation is:

$$r_k = \frac{\sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2}$$

Partial Autocorrelation

A partial autocorrelation is a summary of the relationship between an observation in a time series with observations at prior time steps with the relationships of intervening observations removed. This function plays an important role in data analysis aimed at identifying the extent of the lag in an auto-regressive model. The use of this function was introduced as part of the Box–Jenkins approach to time series modelling, whereby plotting the partial autocorrelation functions one could determine the appropriate lags p in an AR (p) model or in an extended ARIMA (p,d,q) model.Partial Autocorrelation for lag 2 is

$$\Phi_{22} = PACF(X_t, X_{t-2})$$

$$=\text{Cor}(X_t, X_{t-2}/Xt-1)$$

$$= \frac{Cov(X_t, X_{t-k}/Xt-1)}{\sqrt{Var(X_t/X_{t-1})}\sqrt{Var(X_{t-2}/X_{t-1})}}$$

Autoregressive Intergrated Moving Average(ARIMA)

An Autoregressive Intergrated Moving Average(ARIMA) model is a generalization of a simple Autoregressive Moving Average(ARMA) model. Both of these models are used to forecast or predict future points in the time-series data. ARIMA is a form of regression analysis that indicates the strength of a dependent variable relative to other changing variables.
ARIMA model is generally denoted as ARIMA(p,q,d) where p,d,q are defined as follow

- p: the lag order or the number of time lag of autoregressive model AR(p).
- d: degree of differencing or the number of times the data have had subtracted with past value.
- q: the order of moving average model MA(q).

Forecasting

Forecasting is a technique that use historical data as inputs to make informed estimates that are predictive in determining the direction of future trends. It is used by utilities for analyzing their data and predict whether their sales,share or profit trend is up or down. The forecasting is mainly used by retail,operations, marketing,Manufacturing ,Logistics etc.The main methods for time series forecasting is Exponential smoothing,Holt-Winter's Method and ARIMA modeling. In our study we are applying ARIMA modeling to forecast our data using R software.

Multiple Linear Regression Analysis

Multiple Linear Regression or simply known as multiple regression is a statistical technique that used several explanatory variables to predict the outcome of response variable.The multiple linear regression model is based on the following:

- There is a linear relationship between the dependent variables and the independent variables
- The independent variables are not too highly correlated with each other
- y_i observations are selected independently and randomly from the population
- Residuals should be normally distributed with a mean of 0 and variance σ

R-Squared (R^2 or the coefficient of determination) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other words, r-squared shows how well the data fit the regression mode.

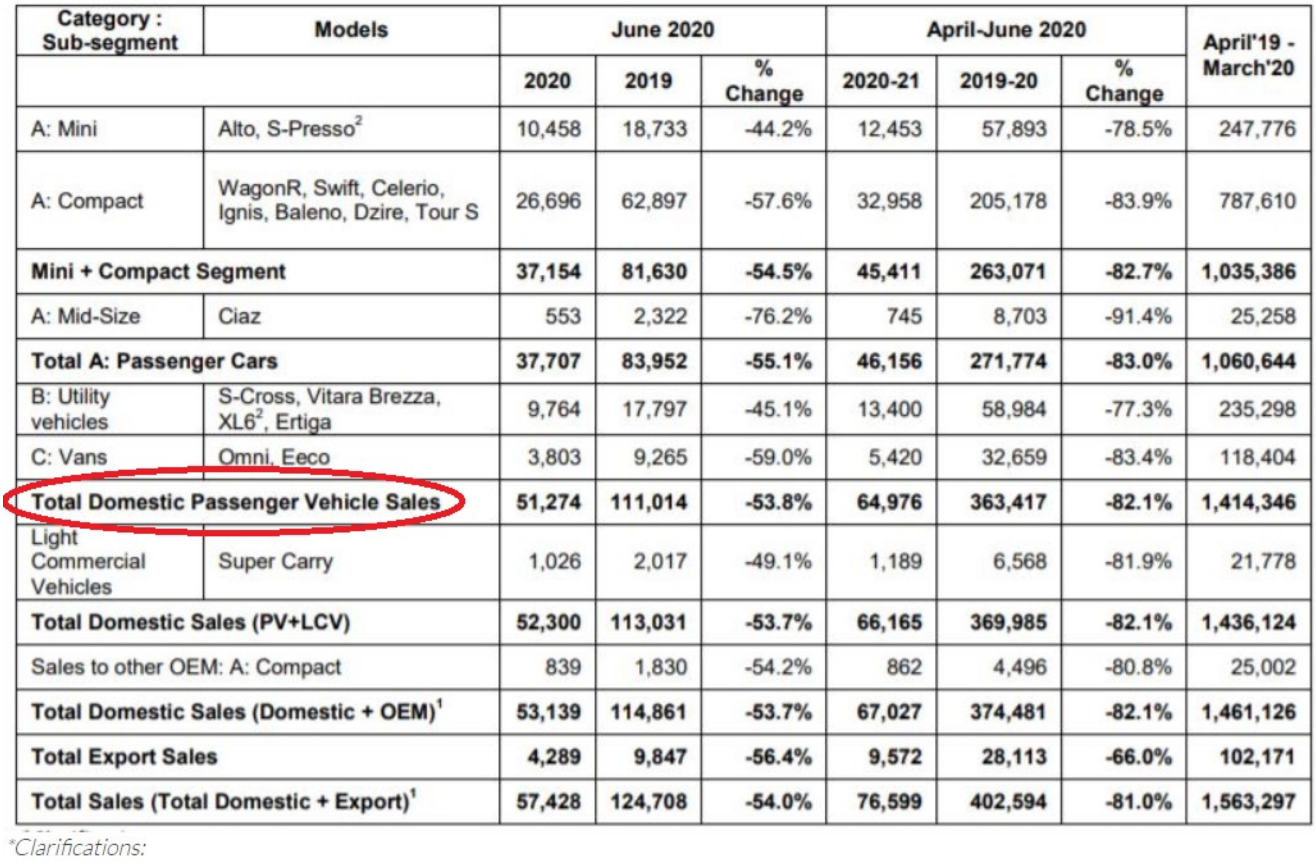
CHAPTER 3

ANALYSIS

Data Situations

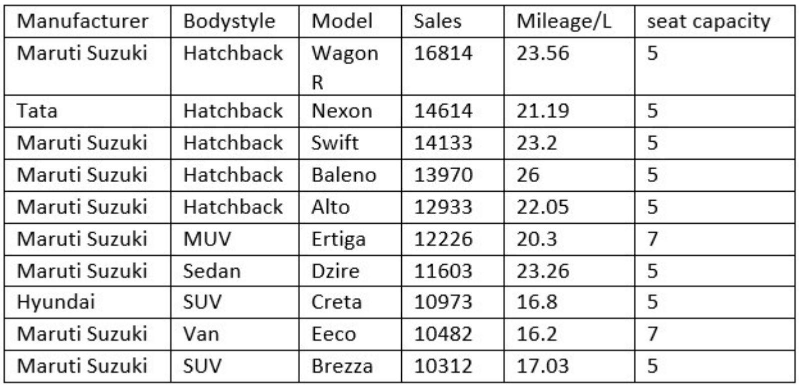
First ,We have to analysis and predict the Domestic Passenger segment which is the point of interest because it includes the India's most selling cars like Alto,Swift,Wagon R etc. The Data was collected from the Official Maruthi Suzuki Website and recorded the monthly sales of these domestic passengers from the time period 5,2020-2,2022. Now to do our regression analysis we collect the number of first 10 car models sold in the month may-2022 from the website. The data includes mileage,body-style,ground clearance,tyre size etc . The data format in Website of which the data is collected is shown below(fig:1 and fig:2) and Data is been stored in excel in the following (fig:3 and fig:4). In fig:4 we plotted the line graph and we can see the highest sales during this period was in October 2020 and the lowest sales is in May 2021. In fig:4 we plotted the line chart and we can see the data given is linear so we can under go multiple linear regression for our study.

From the fig:4,The line graph most of our attributes are linear rather we check correlation for selecting the factors To do multiple Linear Regression. From Fig:5 the Pie-Chart shows that the most cars sold in India is of Hatch-Back body-style. Also there is only one sedan in our 10, that to made this sales because it's a Maruthi Vehicle. In fig:7 we show the sales of May-2022 and we see that the Maruthi has 8 out of 10 first cars sold in this month. Fig:8 shows the number of sale figures in May-2022 and Wagon-R is the most selling car which sold 16814 units.

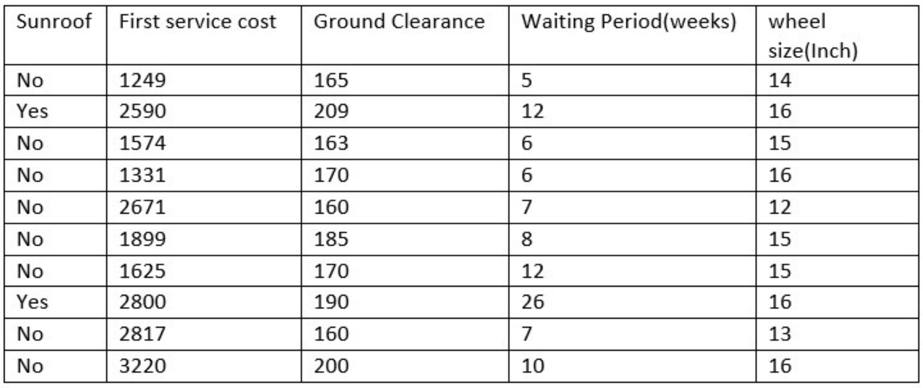


Category : Sub-segment	Models	June 2020			April-June 2020			April'19 - March'20
		2020	2019	% Change	2020-21	2019-20	% Change	
A: Mini	Alto, S-Presso²	10,458	18,733	-44.2%	12,453	57,893	-78.5%	247,776
A: Compact	WagonR, Swift, Celerio, Ignis, Baleno, Dzire, Tour S	26,696	62,897	-57.6%	32,958	205,178	-83.9%	787,610
Mini + Compact Segment		**37,154**	**81,630**	**-54.5%**	**45,411**	**263,071**	**-82.7%**	**1,035,386**
A: Mid-Size	Ciaz	553	2,322	-76.2%	745	8,703	-91.4%	25,258
Total A: Passenger Cars		**37,707**	**83,952**	**-55.1%**	**46,156**	**271,774**	**-83.0%**	**1,060,644**
B: Utility vehicles	S-Cross, Vitara Brezza, XL6², Ertiga	9,764	17,797	-45.1%	13,400	58,984	-77.3%	235,298
C: Vans	Omni, Eeco	3,803	9,265	-59.0%	5,420	32,659	-83.4%	118,404
Total Domestic Passenger Vehicle Sales		**51,274**	**111,014**	**-53.8%**	**64,976**	**363,417**	**-82.1%**	**1,414,346**
Light Commercial Vehicles	Super Carry	1,026	2,017	-49.1%	1,189	6,568	-81.9%	21,778
Total Domestic Sales (PV+LCV)		**52,300**	**113,031**	**-53.7%**	**66,165**	**369,985**	**-82.1%**	**1,436,124**
Sales to other OEM: A: Compact		839	1,830	-54.2%	862	4,496	-80.8%	25,002
Total Domestic Sales (Domestic + OEM)¹		**53,139**	**114,861**	**-53.7%**	**67,027**	**374,481**	**-82.1%**	**1,461,126**
Total Export Sales		**4,289**	**9,847**	**-56.4%**	**9,572**	**28,113**	**-66.0%**	**102,171**
Total Sales (Total Domestic + Export)¹		**57,428**	**124,708**	**-54.0%**	**76,599**	**402,594**	**-81.0%**	**1,563,297**

*Clarifications:

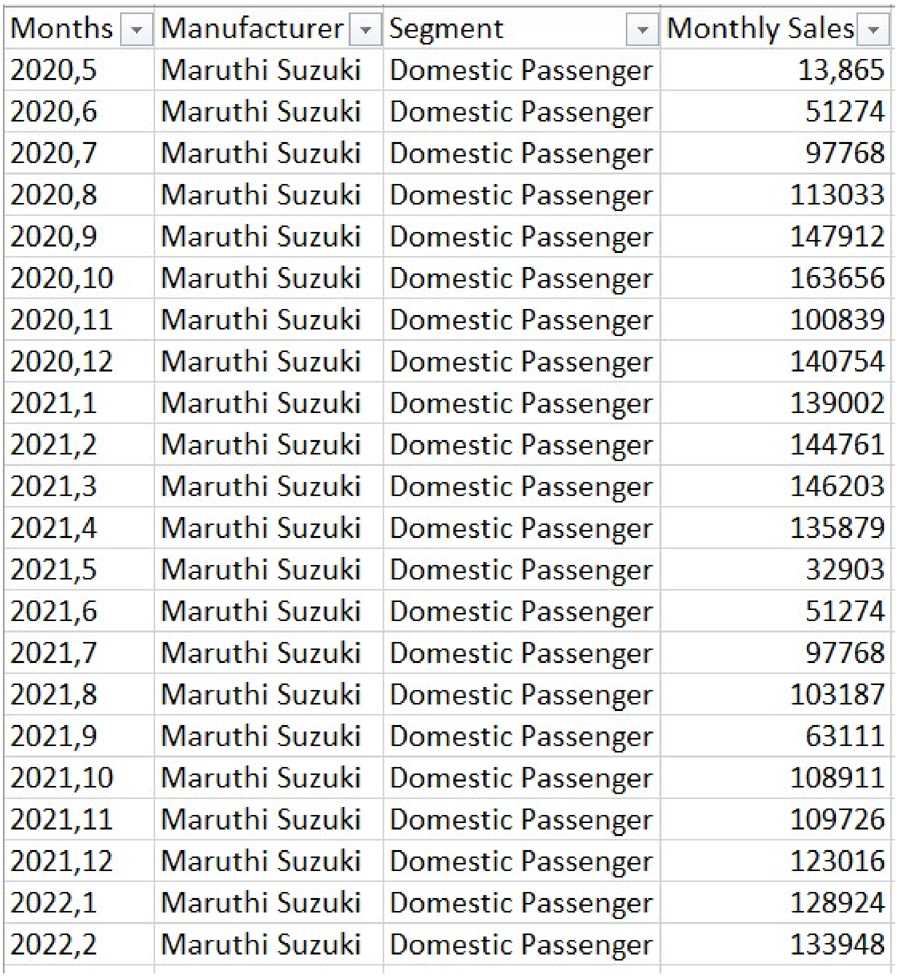


Manufacturer	Bodystyle	Model	Sales	Mileage/L	seat capacity
Maruti Suzuki	Hatchback	Wagon R	16814	23.56	5
Tata	Hatchback	Nexon	14614	21.19	5
Maruti Suzuki	Hatchback	Swift	14133	23.2	5
Maruti Suzuki	Hatchback	Baleno	13970	26	5
Maruti Suzuki	Hatchback	Alto	12933	22.05	5
Maruti Suzuki	MUV	Ertiga	12226	20.3	7
Maruti Suzuki	Sedan	Dzire	11603	23.26	5
Hyundai	SUV	Creta	10973	16.8	5
Maruti Suzuki	Van	Eeco	10482	16.2	7
Maruti Suzuki	SUV	Brezza	10312	17.03	5



Sunroof	First service cost	Ground Clearance	Waiting Period(weeks)	wheel size(Inch)
No	1249	165	5	14
Yes	2590	209	12	16
No	1574	163	6	15
No	1331	170	6	16
No	2671	160	7	12
No	1899	185	8	15
No	1625	170	12	15
Yes	2800	190	26	16
No	2817	160	7	13
No	3220	200	10	16

Table 1: Maruthi Car Sales data and May First 10 Car Sales



Months	Manufacturer	Segment	Monthly Sales
2020,5	Maruthi Suzuki	Domestic Passenger	13,865
2020,6	Maruthi Suzuki	Domestic Passenger	51274
2020,7	Maruthi Suzuki	Domestic Passenger	97768
2020,8	Maruthi Suzuki	Domestic Passenger	113033
2020,9	Maruthi Suzuki	Domestic Passenger	147912
2020,10	Maruthi Suzuki	Domestic Passenger	163656
2020,11	Maruthi Suzuki	Domestic Passenger	100839
2020,12	Maruthi Suzuki	Domestic Passenger	140754
2021,1	Maruthi Suzuki	Domestic Passenger	139002
2021,2	Maruthi Suzuki	Domestic Passenger	144761
2021,3	Maruthi Suzuki	Domestic Passenger	146203
2021,4	Maruthi Suzuki	Domestic Passenger	135879
2021,5	Maruthi Suzuki	Domestic Passenger	32903
2021,6	Maruthi Suzuki	Domestic Passenger	51274
2021,7	Maruthi Suzuki	Domestic Passenger	97768
2021,8	Maruthi Suzuki	Domestic Passenger	103187
2021,9	Maruthi Suzuki	Domestic Passenger	63111
2021,10	Maruthi Suzuki	Domestic Passenger	108911
2021,11	Maruthi Suzuki	Domestic Passenger	109726
2021,12	Maruthi Suzuki	Domestic Passenger	123016
2022,1	Maruthi Suzuki	Domestic Passenger	128924
2022,2	Maruthi Suzuki	Domestic Passenger	133948

Table 2: Maruthi sales during 5-2020 to 2-2022

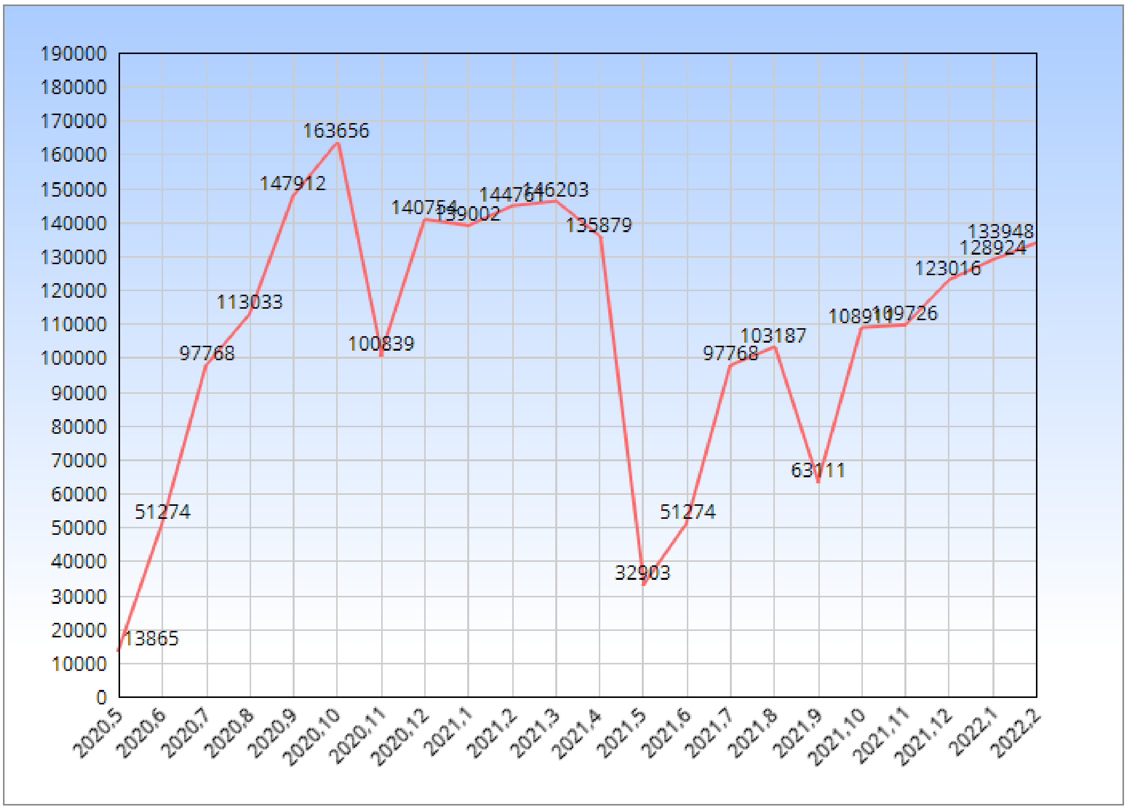


Figure 3: Sales vs time plot

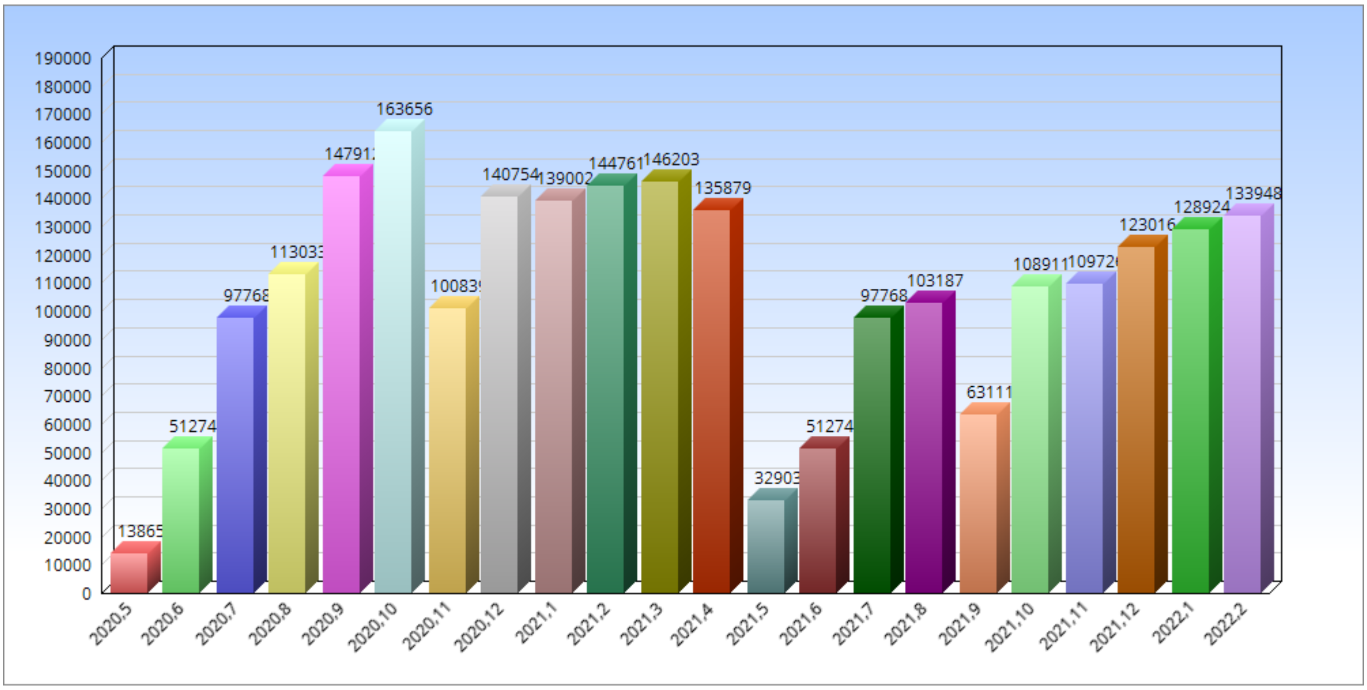


Figure 4: Bar Graph of Car Sales vs Time

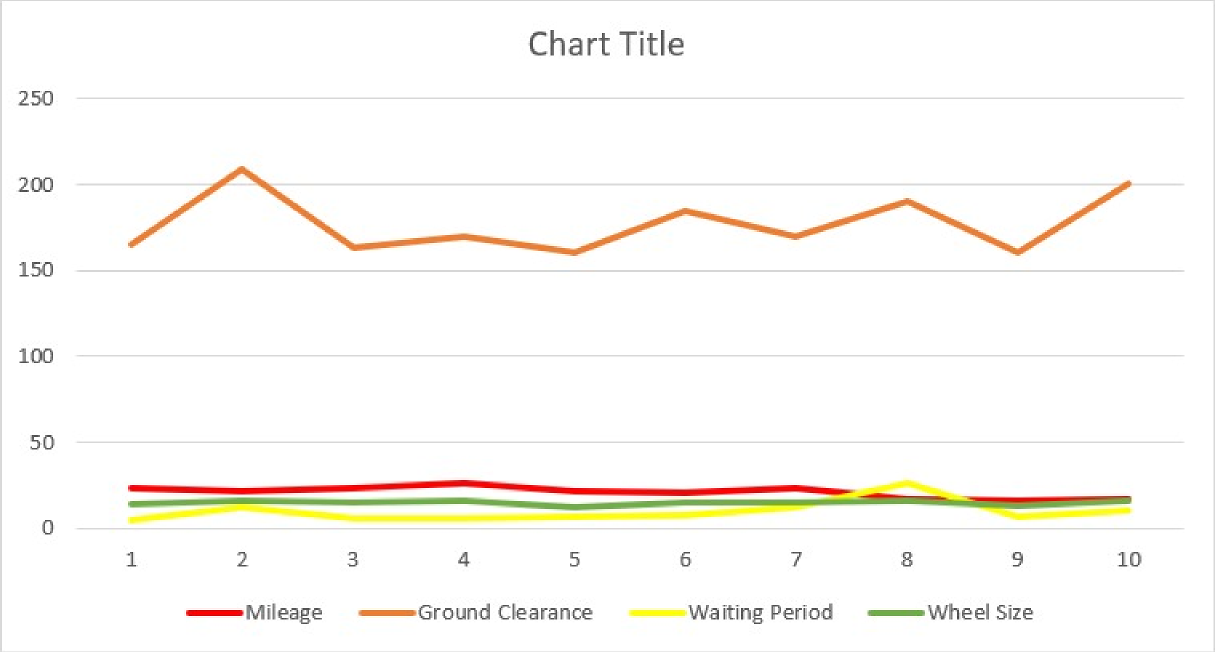


Figure 5: Line graph of the data

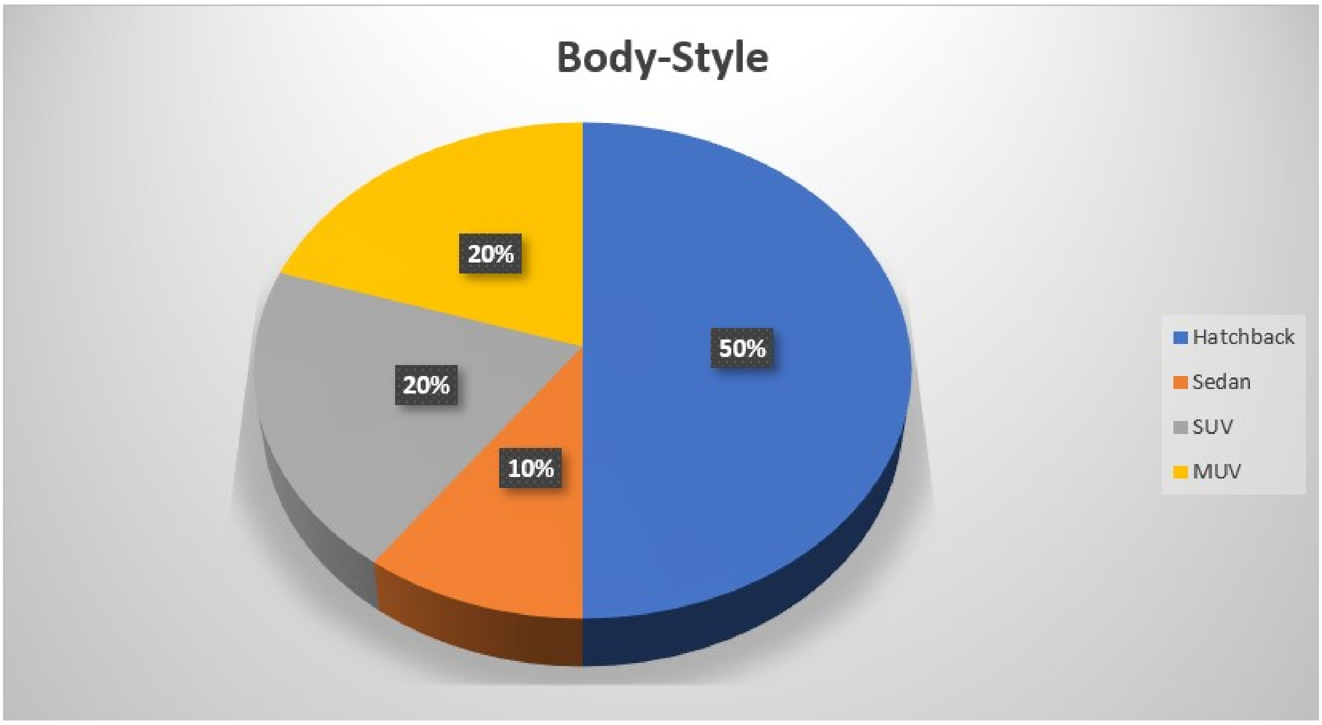


Figure 6: Pie-Chart of Body Style

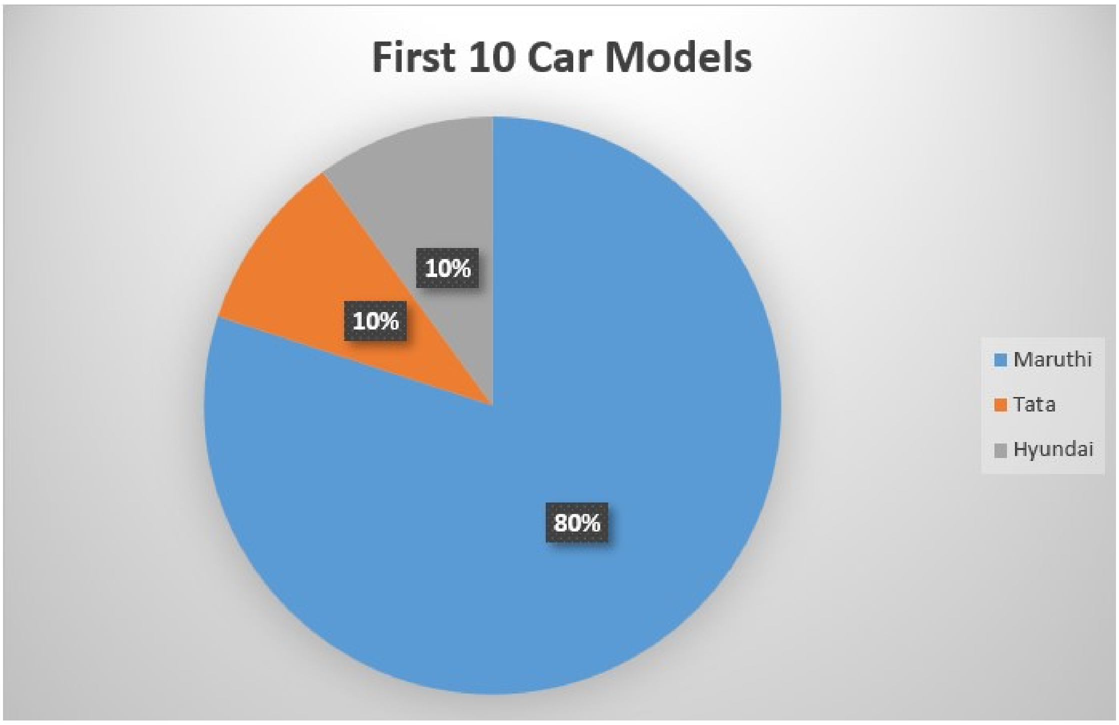


Figure 7: Pie-Chart of Manufacturer

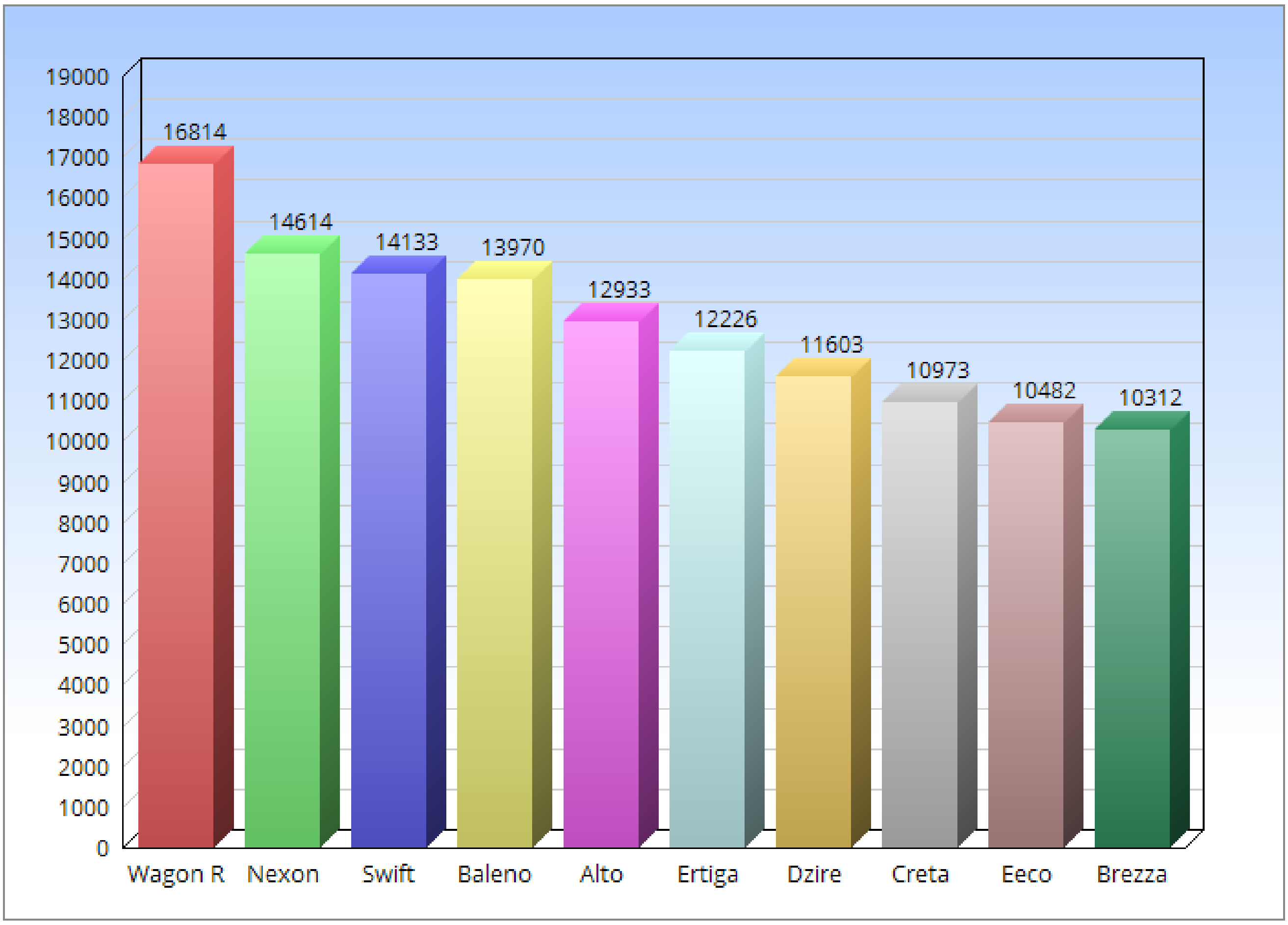


Figure 8: Car Sales in May-2022

Data Processing

First, we want to import the data into our R software from excel using the r code given below(fig:9). Later we plot the time series data in our r software and is given below(fig:10).

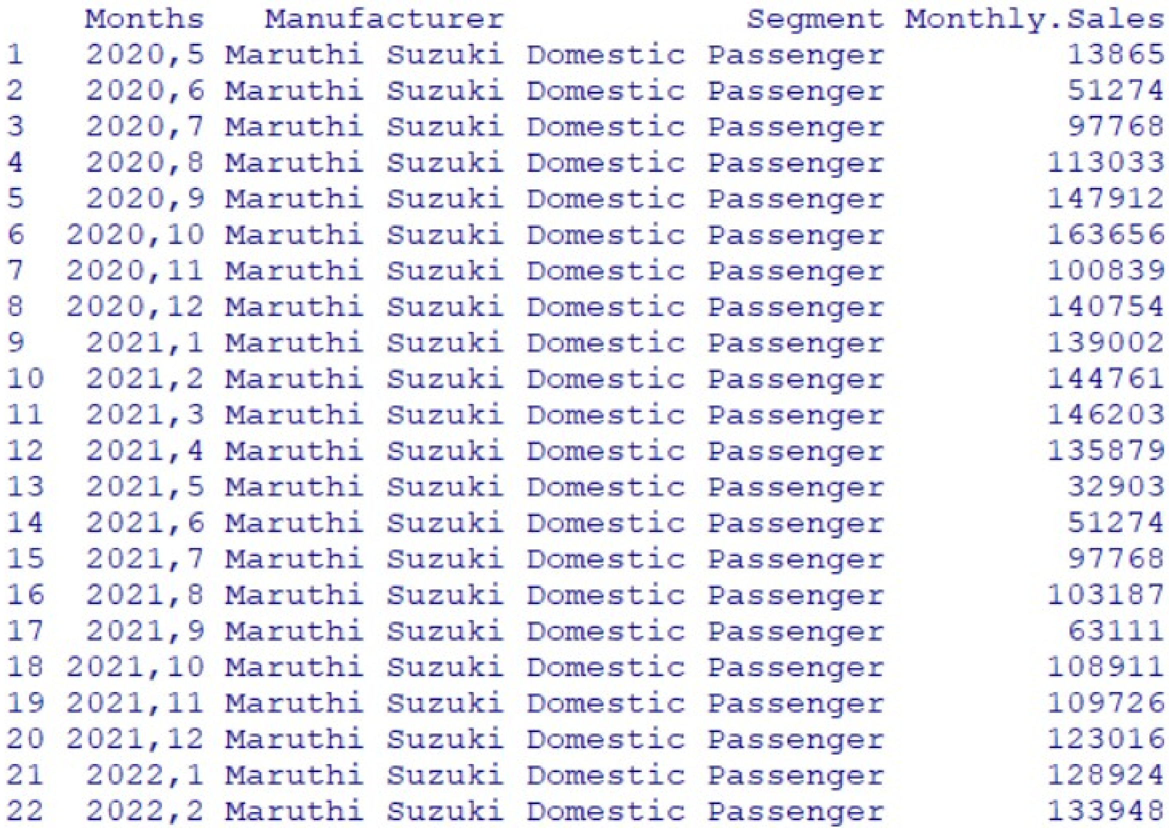
> data=read.csv(file.choose(),header=T)

> data

> ts.plot(Data$Monthly.Sales,xlab=”months”)

>acf(data$Monthly.Sales,lag.max=30)

>pacf(data$Monthly.Sales,lag.max=30)



```
     Months   Manufacturer            Segment Monthly.Sales
1    2020,5 Maruthi Suzuki Domestic Passenger         13865
2    2020,6 Maruthi Suzuki Domestic Passenger         51274
3    2020,7 Maruthi Suzuki Domestic Passenger         97768
4    2020,8 Maruthi Suzuki Domestic Passenger        113033
5    2020,9 Maruthi Suzuki Domestic Passenger        147912
6   2020,10 Maruthi Suzuki Domestic Passenger        163656
7   2020,11 Maruthi Suzuki Domestic Passenger        100839
8   2020,12 Maruthi Suzuki Domestic Passenger        140754
9    2021,1 Maruthi Suzuki Domestic Passenger        139002
10   2021,2 Maruthi Suzuki Domestic Passenger        144761
11   2021,3 Maruthi Suzuki Domestic Passenger        146203
12   2021,4 Maruthi Suzuki Domestic Passenger        135879
13   2021,5 Maruthi Suzuki Domestic Passenger         32903
14   2021,6 Maruthi Suzuki Domestic Passenger         51274
15   2021,7 Maruthi Suzuki Domestic Passenger         97768
16   2021,8 Maruthi Suzuki Domestic Passenger        103187
17   2021,9 Maruthi Suzuki Domestic Passenger         63111
18  2021,10 Maruthi Suzuki Domestic Passenger        108911
19  2021,11 Maruthi Suzuki Domestic Passenger        109726
20  2021,12 Maruthi Suzuki Domestic Passenger        123016
21   2022,1 Maruthi Suzuki Domestic Passenger        128924
22   2022,2 Maruthi Suzuki Domestic Passenger        133948
```

Figure 9: Maruthi Monthly Sales in R

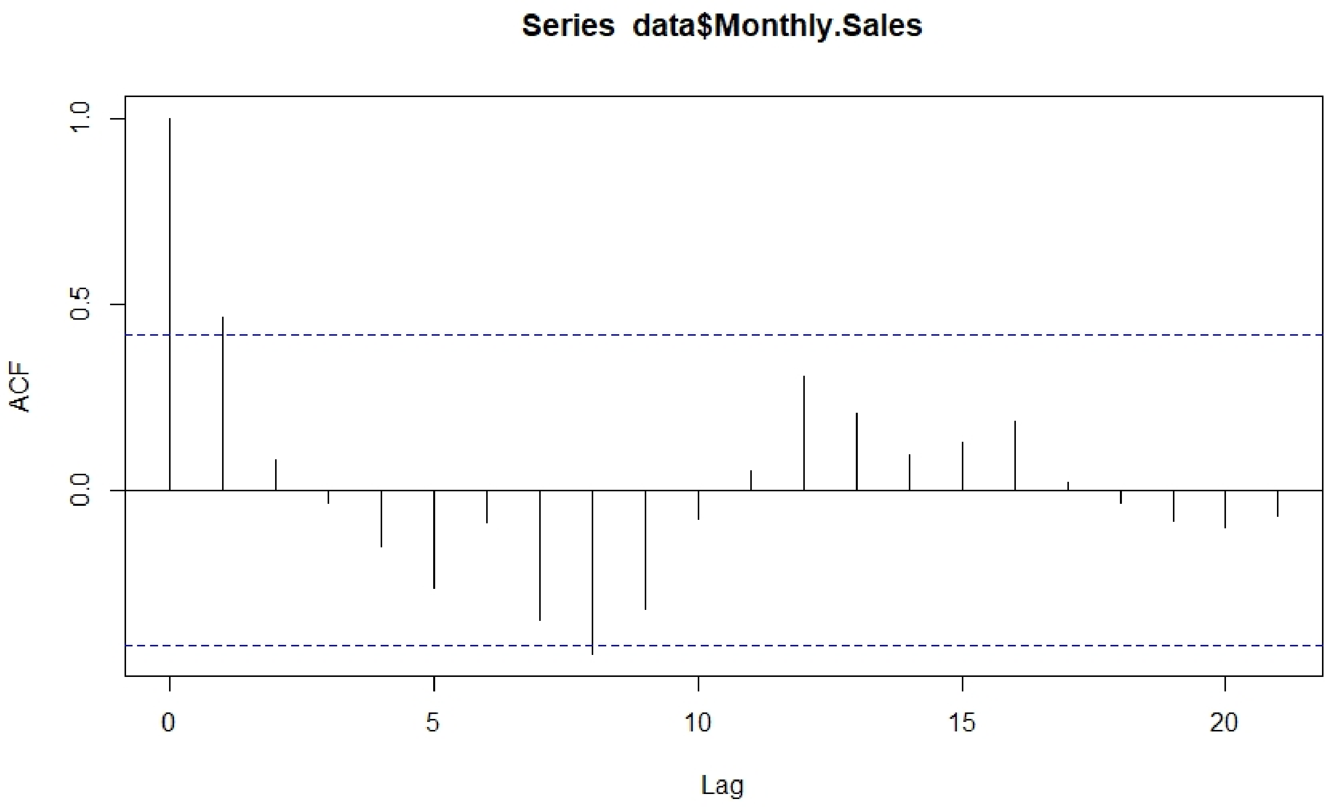


Figure 10: Autocorrelation graph of the sequence

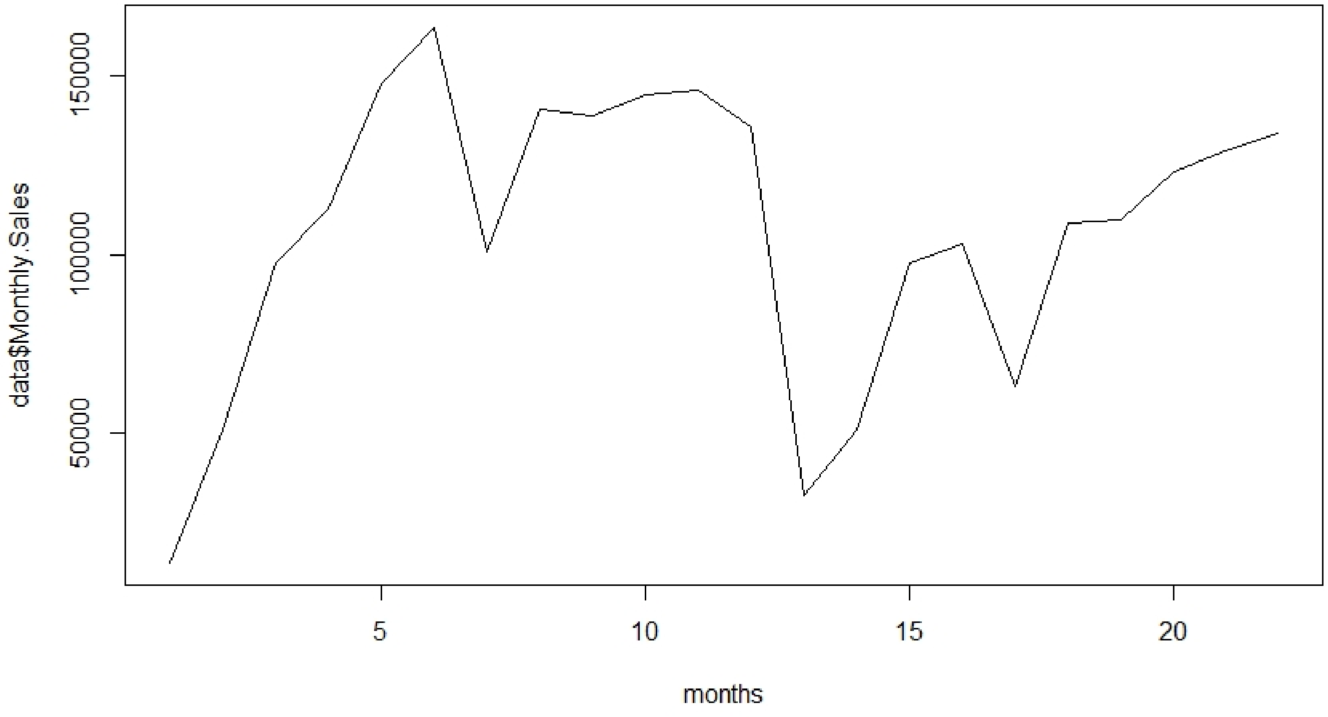


Figure 11: Time series plot

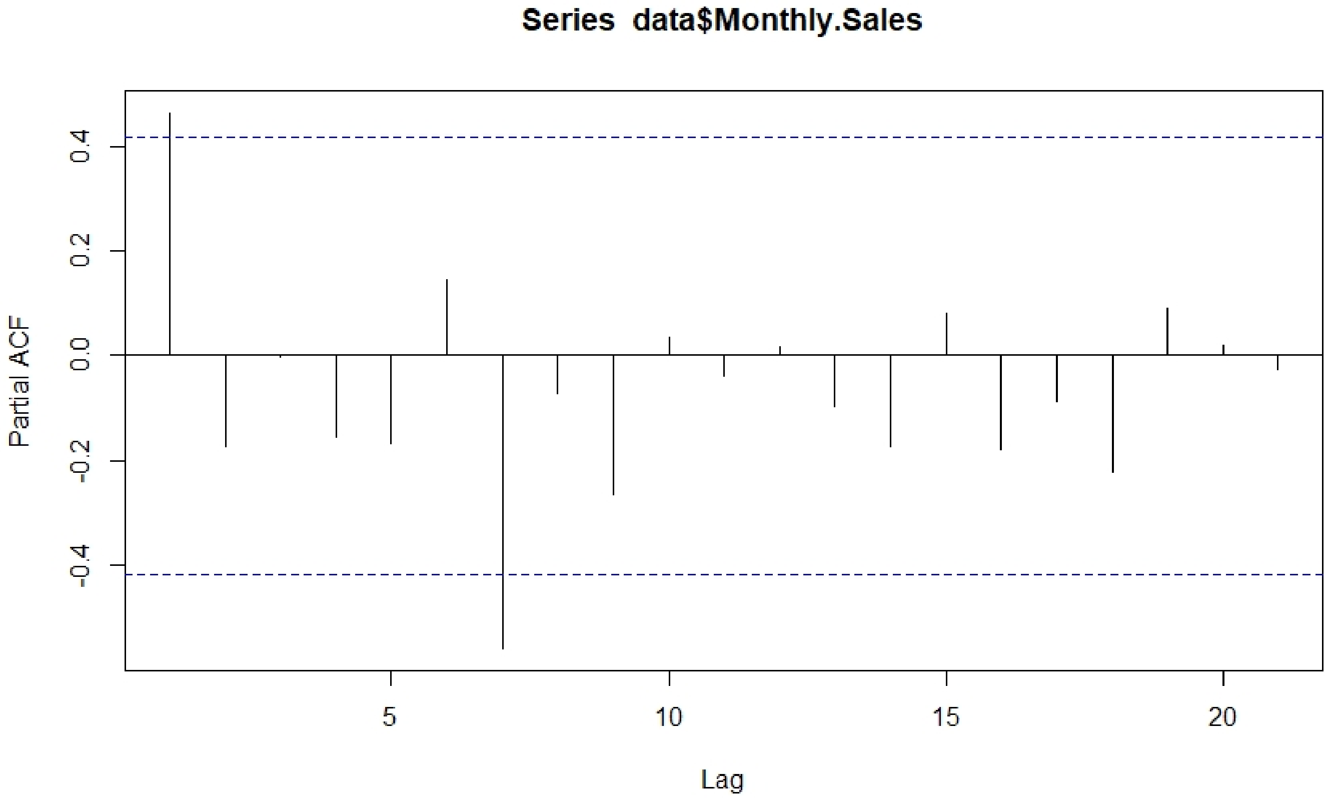


Figure 12: Partial ACF graph of the sequence

In the above figure we can see the monthly sales data of the passenger segment in Maruthi Suzuki of 22 Months from the beginning of COVID-19(2020,5). Clearly we can notice the hike in sales after the lockdown restrictions have been lifted out. When the second wave of COVID-19 occur during May-2021 there is a significant drop in sales. But when we plot the ACF and PACF of the respective data, we observe that the ACF and PACF graph tends to zero so we assume that the given data is stationary and make more evidence for our statement.

Time Series Analysis

Autocorrelation Test

For stationary series of data, the ACF autocorrelation graph(fig:10) immediately tend to zero. The two dashed lines in the autocorrelation graph represent the confidence bounds, which are the upper and lower bounds of the autocorrelation coefficients. The R code for plotting ACF is given below

> acf(Data$Monthly.Sales,lag.max=30)#fig:10

Unit Root test

Unit root is a characteristic of a time series that makes it non-stationary. That is a unit root is said to exist in a time series of the value α=1 in the below equation

$$Y_t = \alpha Y_{t-1} + \beta X_e + \epsilon$$

Where Y_t s the value of the time series at time 't' and X_e is a separate explanatory variable, which is also a time series. The presence of unit root means the time series is non-stationary. Besides, the number of unit roots contained in the series corresponds to the number of differencing operations required to make the series stationary.

KPSS TEST

Kwiatkowski-Phillips-Schmidt-Shin (**KPSS**) Test is a statistical test for testing whether a data is stationary or not.We define the null and alternate hypothesis as

H_0:Series is stationary or has no unit root.

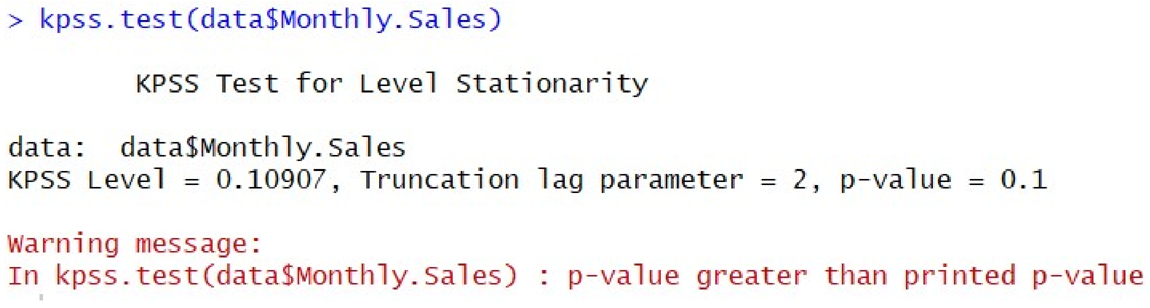
H_1:Series is non stationary or has a unit root.

We fail to reject null hypothesis if the p values is greater than significant level(0.05). The key difference of KPSS test from tests like ADF is that the null hypothesis is the given series is stationary whereas ADF test is vice-versa.

KPSS Test in R Software

The R code for KPSS test is

>kpss.test(data$Monthly.Sales)



```
> kpss.test(data$Monthly.Sales)

        KPSS Test for Level Stationarity

data:  data$Monthly.Sales
KPSS Level = 0.10907, Truncation lag parameter = 2, p-value = 0.1

Warning message:
In kpss.test(data$Monthly.Sales) : p-value greater than printed p-value
```

Figure 13: KPSS Test in R

The p-value is greater than 0.05 so we fail to reject the null hypothesis. So let us conclude that the given data is stationarity.

CHAPTER 5

ARIMA Modeling Analysis

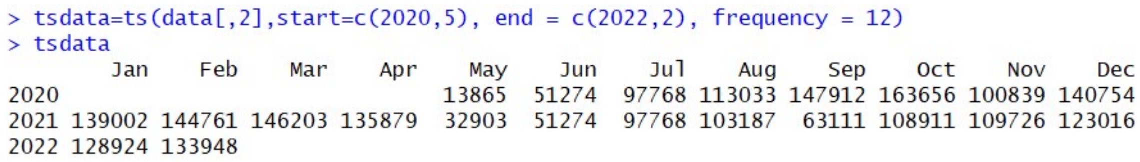
Identifying Order

Since we have obtained ACF and PACF of the data. In ACF we have two significant values and we have only one significant value in PCF.Since we haven't done differencing in our data, we assume three models.First we assume three models and check their Akaike and Bayesian Information Criterion(AIC & BIC).AIC estimates the quality of each model, when we represent a data while modeling some information about the data may be lost.AIC estimates the relative amount lost during this modeling.while fitting models to increase the likelihood we add more parameters which may fail to fit the model and predict adequately.The model with least AIC and BIC is the best model.Later we check the model using auto.arima() function and conclude our arima model and go to forecasting.

ARIMA Modeling in R

As mentioned above first we execute auto arima function in R.Before executing auto arima function we must convert our data into time series data in R. There are several methods for converting our data into a time series data in R. R code for converting the data into time series is

```
>tsdata=ts(data[,2],start=c(2020,5), end = c(2022,2), frequency = 12)
```



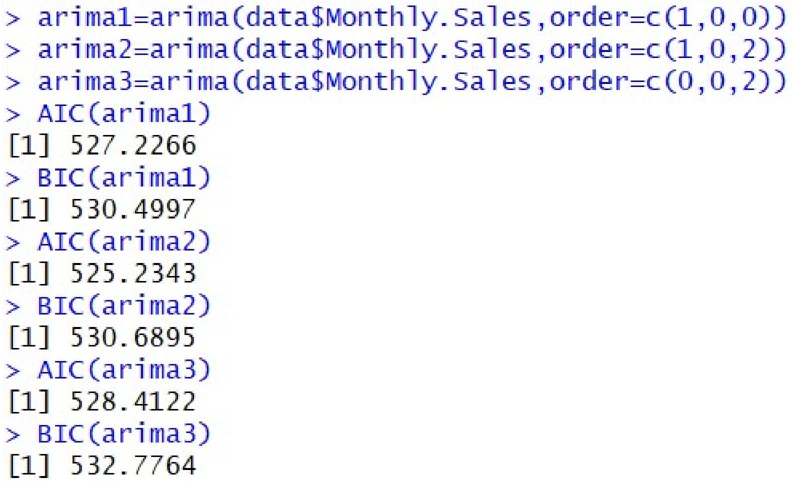
```
> tsdata=ts(data[,2],start=c(2020,5), end = c(2022,2), frequency = 12)
> tsdata
        Jan    Feb    Mar    Apr    May    Jun    Jul    Aug    Sep    Oct    Nov    Dec
2020                             13865  51274  97768 113033 147912 163656 100839 140754
2021 139002 144761 146203 135879  32903  51274  97768 103187  63111 108911 109726 123016
2022 128924 133948
```

Figure 14: data converted to time series

Now we call the three ARIMA model

- ARIMA(1,0,0)
- ARIMA(1,0,1)
- ARIMA(0,0,2)

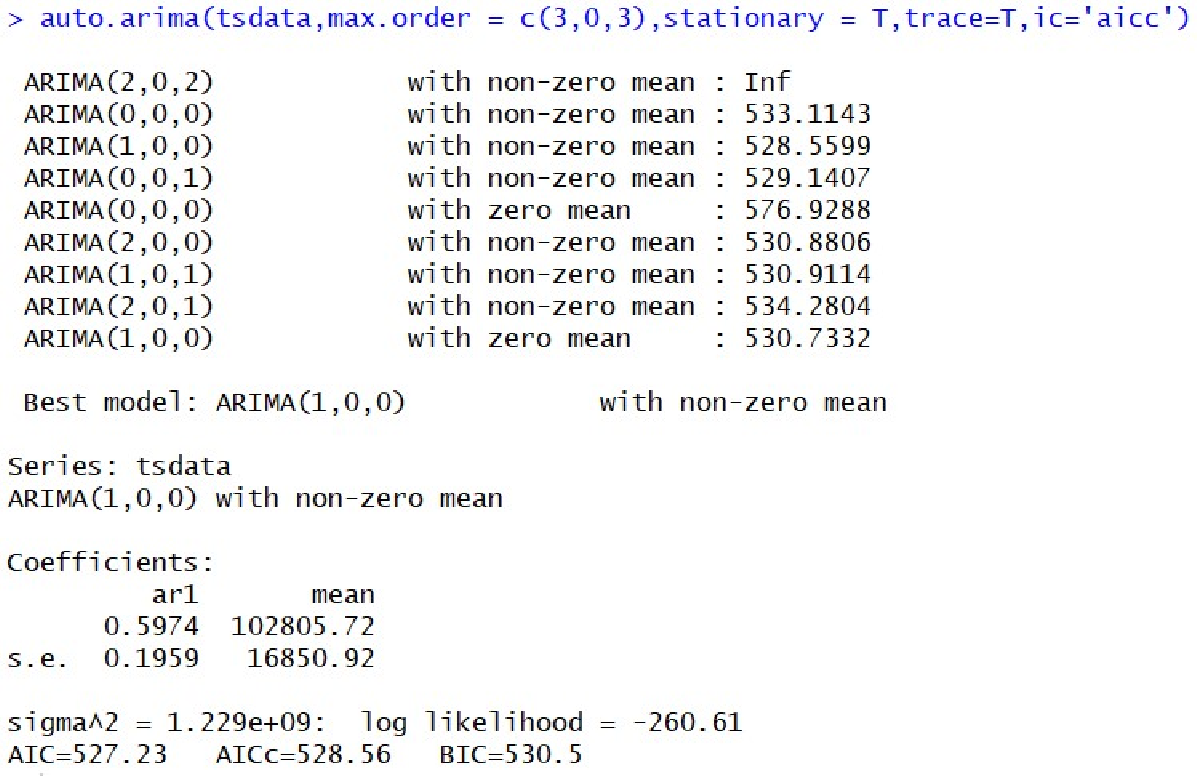
R code for calling ARIMA model and to find the AIC and BIC values
>arima1=arima(data$Monthly.Sales,order=c(1,0,1))
>AIC(arima1)
>BIC(arima1)



```
> arima1=arima(data$Monthly.Sales,order=c(1,0,0))
> arima2=arima(data$Monthly.Sales,order=c(1,0,2))
> arima3=arima(data$Monthly.Sales,order=c(0,0,2))
> AIC(arima1)
[1] 527.2266
> BIC(arima1)
[1] 530.4997
> AIC(arima2)
[1] 525.2343
> BIC(arima2)
[1] 530.6895
> AIC(arima3)
[1] 528.4122
> BIC(arima3)
[1] 532.7764
```

Figure 15: AIC and BIC of ARIMA

We find the BIC values of ARIMA(1,0,0) is less than the other two models whereas the AIC value of ARIMA(1,0,2) is less than the above model, therefore for more confirmation we do the auto arima function.



```
> auto.arima(tsdata,max.order = c(3,0,3),stationary = T,trace=T,ic='aicc')

 ARIMA(2,0,2)           with non-zero mean : Inf
 ARIMA(0,0,0)           with non-zero mean : 533.1143
 ARIMA(1,0,0)           with non-zero mean : 528.5599
 ARIMA(0,0,1)           with non-zero mean : 529.1407
 ARIMA(0,0,0)           with zero mean     : 576.9288
 ARIMA(2,0,0)           with non-zero mean : 530.8806
 ARIMA(1,0,1)           with non-zero mean : 530.9114
 ARIMA(2,0,1)           with non-zero mean : 534.2804
 ARIMA(1,0,0)           with zero mean     : 530.7332

 Best model: ARIMA(1,0,0)           with non-zero mean

Series: tsdata
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1       mean
      0.5974  102805.72
s.e.  0.1959   16850.92

sigma^2 = 1.229e+09:  log likelihood = -260.61
AIC=527.23   AICc=528.56   BIC=530.5
```

Figure 16: Auto ARIMA in R

Here the auto arima function checked the models up to the order (3,0,3) and found out that the best order for your ARIMA is (1,0,0) that is Auto Regressive model of order 1. Now we want to forecast and predict using this model.

FORECASTING in R

Now we have identified our ARIMA model with evidence. Our next procedure is to forecast the sales data in R.In our forecasting procedure, we predict the next 5 months data using R.The R code for forcasting is

```
>forecast=forecast:::forecast.Arima(arima,h=5,level=c(80,90))
>forecast
>plot(forecast)
```

We have forecast the 5 month sales data, since the data we choose is from 5-2020 to 2-2022 we can use Mean absolute percentage error (MAPE) to measure forecast accuracy.MAPE is the sum of the individual absolute errors divided by the demand, that is the average of the percentage error.We have the 3 months data (March,April,May) after 2-2020 so we compare these original data and our predicted data. A good MAPE value score is as follows:

- <10 percent:Very Good
- 10-20 percent:Good
- 20-50 percent:Ok
- >50 percent:Bad

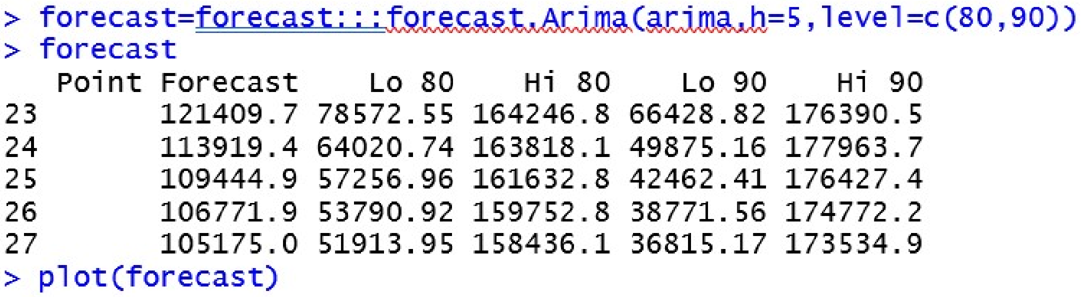
$$M = \frac{1}{n}\sum_{t=1}^{n}|\frac{A_t-F_t}{A_t}|$$

M=mean absolute percentage error

n=number of times the summation iteration happens

A_1=actual value

F_t=predicted value



```
> forecast=forecast:::forecast.Arima(arima,h=5,level=c(80,90))
> forecast
   Point Forecast     Lo 80    Hi 80    Lo 90    Hi 90
23       121409.7 78572.55 164246.8 66428.82 176390.5
24       113919.4 64020.74 163818.1 49875.16 177963.7
25       109444.9 57256.96 161632.8 42462.41 176427.4
26       106771.9 53790.92 159752.8 38771.56 174772.2
27       105175.0 51913.95 158436.1 36815.17 173534.9
> plot(forecast)
```

Figure 17: forecast of Sales

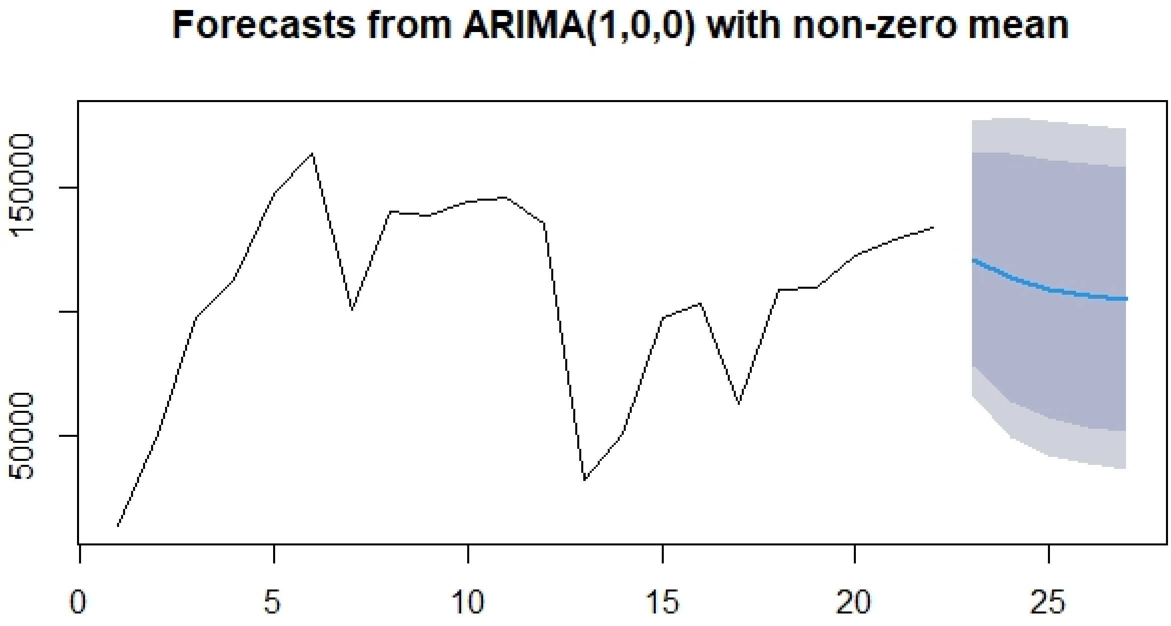
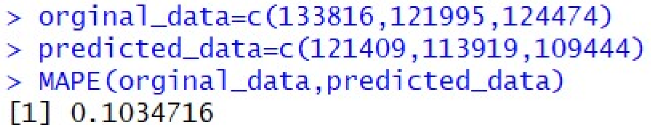


Figure 18: Plot of forecast



```
> orginal_data=c(133816,121995,124474)
> predicted_data=c(121409,113919,109444)
> MAPE(orginal_data,predicted_data)
[1] 0.1034716
```

Figure 19: MAPE test in R

Inference

From Fig:17 the predictions can be clearly seen,we draw the primitive and predictive graphics in r using plot command(Fig:18).Also in Fig:19, we get the value of MAPE Measure as 10.34 percent which indicate our forecasting is good.

Multiple Linear Regression

The model of multiple linear regression is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...$$

Where Y is the response variable,$X_1, X_2, ...$ are the regressor variable, β_0 is the Y-Intercept and $\beta_1, \beta_2, ...$ are the regression coefficient
We state the null and alternate hypothesis as

$$H_0 : \beta_1 = \beta_2 = ... = 0$$

$$H_1\text{:At least one } \beta_j \neq 0$$

Before modeling we want to know that whether our data is linear or not,for that we plot the data.Data which we are taking is the monthly sales(May-2022) of first 10 cars in INDIA,Our Objective is find the customer buying preference over car models From the graph Fig:4.

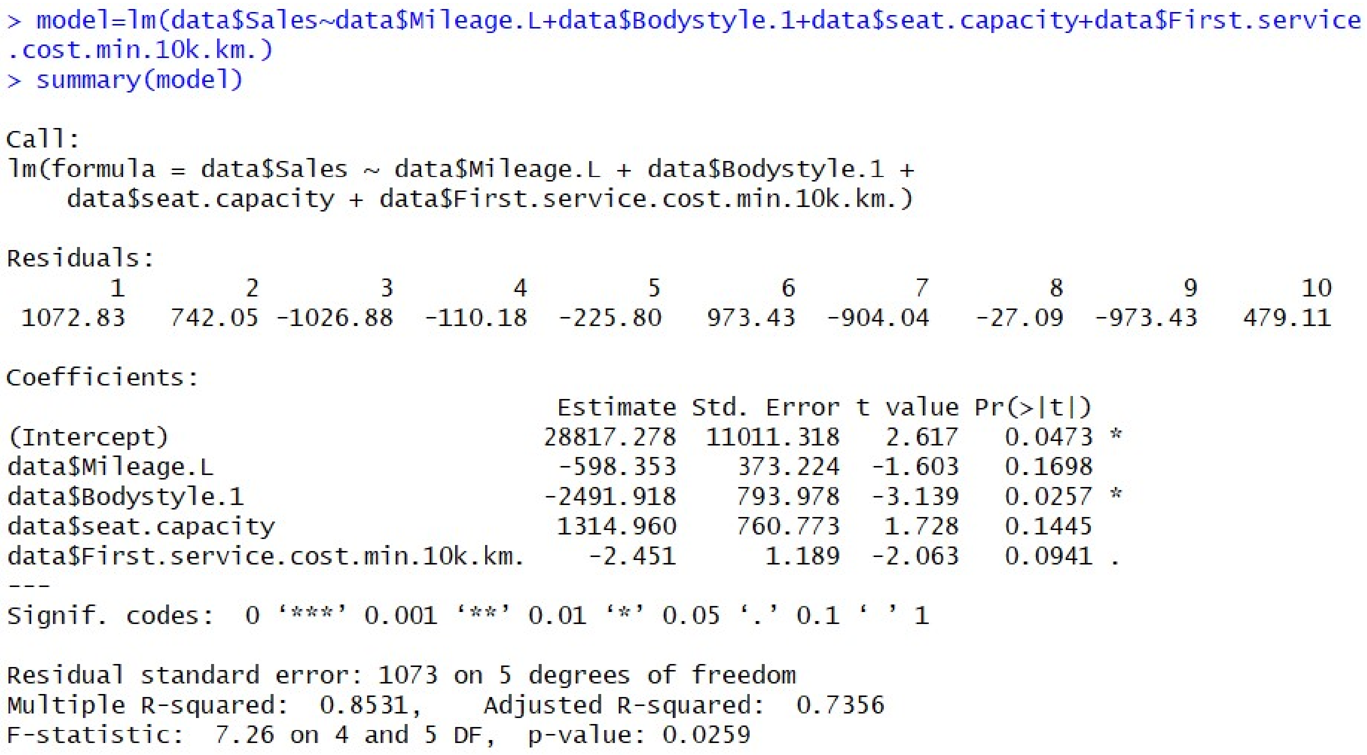
Multiple Linear Regression in R

After importing our data into R software, we check the correlations of the factors and choose the factors in which sales have more correlations.Later we fit the model using lm() function in R and call.If the p-value is >0.05 the we fail to reject null hypothesis or else we accept H_1.

The R code for fitting the model is

>model=lm(data$Sales data$Mileage.L+data$Bodystyle.1+data$seat.capacity+ data$First.service.cost.min.10k.km.)

>summary(model)



```
> model=lm(data$Sales~data$Mileage.L+data$Bodystyle.1+data$seat.capacity+data$First.service
.cost.min.10k.km.)
> summary(model)

Call:
lm(formula = data$Sales ~ data$Mileage.L + data$Bodystyle.1 + 
    data$seat.capacity + data$First.service.cost.min.10k.km.)

Residuals:
       1        2        3        4        5        6        7        8        9       10 
 1072.83   742.05 -1026.88  -110.18  -225.80   973.43  -904.04   -27.09  -973.43   479.11 

Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)  
(Intercept)                         28817.278  11011.318   2.617   0.0473 *
data$Mileage.L                       -598.353    373.224  -1.603   0.1698  
data$Bodystyle.1                    -2491.918    793.978  -3.139   0.0257 *
data$seat.capacity                   1314.960    760.773   1.728   0.1445  
data$First.service.cost.min.10k.km.    -2.451      1.189  -2.063   0.0941 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1073 on 5 degrees of freedom
Multiple R-squared:  0.8531,	Adjusted R-squared:  0.7356 
F-statistic:  7.26 on 4 and 5 DF,  p-value: 0.0259
```

Figure 20: Model fitting in R

Inference

From above test(Fig:20) p-value is less than 0.05 so we reject null hypothesis, that is we accept that some of the regression coefficients are significant.In our Model the Intercept and Body Style is significant Ṫhe regression coefficients are

β_0=28817.278

β_1=-2491.918

where β_0 is the y-intercept and β_1 is Body style.

Therefore our model is given by

$$Y = 28817.278 + -2491.918X_1$$

Chapter 6

Conclusion

The analysis of monthly sales data or Maruthi-Suzuki cars is done above by taking the data during the COVID-19 hits in our country. Interest of study is in sales of domestic passenger of Maruthi cars which dominated in first 10 cars sold in each and every month. R software is used here for time series analysis , that is to draw the sequence diagram to test whether it is a smooth test. After confirming that our data is stationary the use ARIMA method for analysis and modeling and forecast. MAPE measure is used to test whether our forecast is good or not.Then to know the customer buying preference over cars,Multiple Linear Regression modeling is used in R and tested significance of regression coefficients. After analysis,the best model is ARIMA(1,0,0) that is

the AR-1 model. Since our data is stationary we need not do differencing. This ARIMA(1,0,0) model is used for predict the short-term trend of domestic passenger segment cars of Maruthi-Suzuki. When MAPE is measured we got the forecasting is good.In Multiple Linear Regression modeling we confirmed that the factor body style is significant.The adjusted R-square 85 percent. The factors are selected according their correlations and modeling the model regression model.

Forecasting shows that there is a possibility of a decrease in sales. To overcome this the company must do new marketing strategies like giving exciting offers.Company must alsDuring COVID-19 restriction the company must give an option for online purchases and provide home delivery even in remote places.From monthly car sales data, it is seen that customer prefer Hatchback over other models. So to increase the number of sales the company must do more research and development on giving more features in hatchbacks.The Manufacturer must give more authorized customization options for their cars in both exterior and interior. Availing Engine modifications through an authorized service center attracts youth to buy their respective cars

References

- Shailesh Tiwary-Munesh C.Trivedi-Krishn Kumar Misra Smart Innovations in Communication and Computational Sciences (197–207)Link
- Ruoyu, L., Libo, L.: Analysis and prediction of tourist numbers based on ARIMA model. Comput. Telecommun
- Yuxin Zhao.,Analysis and Forecast based of Car Sales Based on R Language Link
- R-Bloggers Link
- Statology Link