Table of Contents

[Introduction 1](#_Toc97460363)

[Data Analysis 2](#_Toc97460364)

[Descriptive Statistics 3](#_Toc97460365)

[Exploratory Data Analysis 15](#_Toc97460366)

[Conclusion 18](#_Toc97460367)

[Bibliography 19](#_Toc97460368)

[Appendix 20](#_Toc97460369)

# Introduction

Data analytics is the process of analyzing raw data and generating actionable insights. It comprises of the processes, tools and techniques of data analysis and management, including the collection, organization, and storage of data. Organizations use data analytics to gain competitive advantage by improving their performance and operational efficiency. Data analytics is performed on a variety of big data sets, like transactions, click streams, server logs, electronic health records, insurance claims, etc. Different analytical techniques and algorithms can be applied on these data sets to accomplish different objectives. These different types of analytical techniques are colloquially called: -

1. **Descriptive analytics** - Summarizing the data to understand past events and performance
2. **Diagnostic analytics** - Investigating the root cause of certain events
3. **Predictive analytics** - Predicting the future for planning
4. **Prescriptive analytics** - Recommending the optimal outcomes

Irrespective of the type of analytics being performed, the basis of every method or algorithm in data analytics is descriptive/inferential statistics and machine learning. In this analysis report, we will leverage descriptive statistics to generate insights from the data.

**Problems Statement:**

CarDekho.com is India’s leading car search venture that helps users buy cars that are right for them. It’s website and app carry rich automotive content such as expert reviews, detailed specs and prices, comparisons as well as videos and pictures of all car brands and models available in India.

# Data Analysis

**Importing the libraries required for the analysis**

# Declaring the names of packages to be imported  
packageList <- c("tidyverse", "vtable", "RColorBrewer", "corrplot", "car", "psych", "stargazer", "scales", "DT")  
  
for (package in packageList) {  
 if (!package %in% rownames(installed.packages()))   
 { install.packages(package) }  
   
 # Import the package  
 library(package, character.only = TRUE)  
}

**Importing the data set for analysis**

## Descriptive Statistics

**Descriptive Statistics of ‘Total Votes’ in 2008, 2012, 2016**

totalVotesStats <- ElectionData %>%   
 select(v2008, v2012, v2016)  
  
# Kable Classic Method  
totalVotesStats <- totalVotesStats %>%   
 describe(quant = c(.25, .75), IQR = TRUE) %>%   
 mutate(year = c(2008, 2012, 2016)) %>%   
 relocate(year)  
round(totalVotesStats, 2) %>%   
 kbl(caption = "Table 1: Descriptive Statistics for Total Votes") %>%   
 kable\_classic(html\_font = "Cambria")

Table 1: Descriptive Statistics for Total Votes

**Descriptive Statistics of ‘Total Democratic Votes’ in 2008, 2012, 2016**

totalDemocraticVotesStats <- ElectionData %>%   
 select(vd2008, vd2012, vd2016) %>%   
 describe(quant = c(.25, .75), IQR = TRUE) %>%   
 mutate(year = c(2008, 2012, 2016)) %>%   
 relocate(year)  
# Kable Classic Method  
round(totalDemocraticVotesStats, 2) %>%   
 kbl(caption = "Table 2: Descriptive Statistics for Total Democratic Votes") %>%   
 kable\_classic(html\_font = "Cambria")

Table 2: Descriptive Statistics for Total Democratic Votes

**Descriptive Statistics of ‘Total Republican Votes’ in 2008, 2012, 2016**

totalRepublicanVotes <- ElectionData %>%   
 select(vg2008, vg2012, vg2016) %>%   
 describe(quant = c(.25, .75), IQR = TRUE) %>%   
 mutate(year = c(2008, 2012, 2016)) %>%   
 relocate(year)  
# Kable Classic Method  
round(totalRepublicanVotes, 2) %>%   
 kbl(caption = "Table 3: Descriptive Statistics for Total Republican Votes") %>%   
 kable\_classic(html\_font = "Cambria")

Table 3: Descriptive Statistics for Total Republican Votes

**Descriptive Statistics of ‘Total Unemployment Rate’ in 2011, 2012, 2013, 2014, 2015**

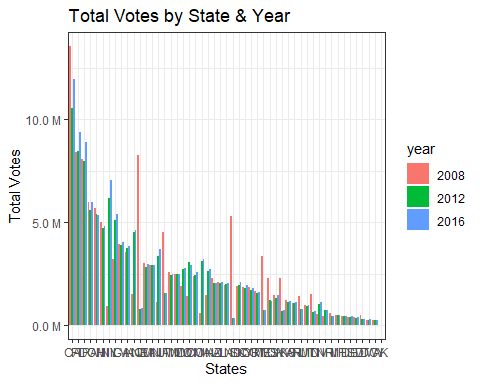
unemploymentRate <- ElectionData %>%   
 select(unemp11, unemp12, unemp13, unemp14, unemp15) %>%   
 describe(quant = c(.25, .75), IQR = TRUE) %>%   
 mutate(year = c(2011, 2012, 2013, 2014, 2015)) %>%   
 relocate(year)  
# Kable Classic Method  
round(unemploymentRate, 2) %>%   
 kbl(caption = "Table 4: Descriptive Statistics for Total Unemployment Rate") %>%   
 kable\_classic(html\_font = "Cambria")

Table 4: Descriptive Statistics for Total Unemployment Rate

## Exploratory Data Analysis

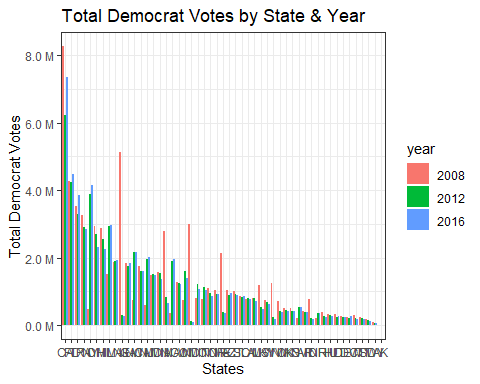
**Plot of outcome variable ‘Total Votes’ by state and year**

totalVotesL <- ElectionData %>%   
 select(state, v2008, v2012, v2016) %>% group\_by(state) %>% summarise('2008' = sum(v2008, na.rm = TRUE),  
 '2012' = sum(v2012, na.rm = TRUE),  
 '2016' = sum(v2016, na.rm = TRUE)) %>%  
 gather(year, tVotes, c('2008', '2012', '2016'))  
  
ggplot(data = totalVotesL, mapping = aes(x = reorder(factor(state), tVotes, function(x) -1\*sum(x)), y = tVotes, fill = year)) +  
 geom\_bar(position = "dodge", stat = "identity") +  
 labs(title = "Total Votes by State & Year") +   
 scale\_x\_discrete(name ="States") +   
 scale\_y\_continuous(name = "Total Votes", labels = label\_number(suffix = " M", scale = 1e-6)) +  
 theme\_bw()



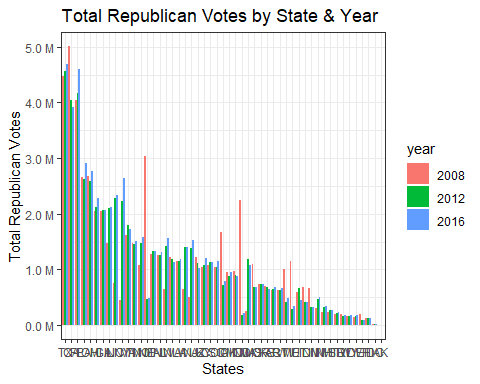
**Plot of outcome variable ‘Total Democratic Votes’ by state and year**

totalDVotesL <- ElectionData %>%   
 select(state, vd2008, vd2012, vd2016) %>% group\_by(state) %>% summarise('2008' = sum(vd2008, na.rm = TRUE),  
 '2012' = sum(vd2012, na.rm = TRUE),  
 '2016' = sum(vd2016, na.rm = TRUE)) %>%  
 gather(year, tdVotes, c('2008', '2012', '2016'))  
  
ggplot(data = totalDVotesL, mapping = aes(x = reorder(factor(state), tdVotes, function(x) -1\*sum(x)), y = tdVotes, fill = year)) +  
 geom\_bar(position = "dodge", stat = "identity") +  
 labs(title = "Total Democrat Votes by State & Year") +   
 scale\_x\_discrete(name ="States") +   
 scale\_y\_continuous(name = "Total Democrat Votes", labels = label\_number(suffix = " M", scale = 1e-6)) +  
 theme\_bw()



**Plot of outcome variable ‘Total Republican Votes’ by state and year**

totalRVotesL <- ElectionData %>%   
 select(state, vg2008, vg2012, vg2016) %>% group\_by(state) %>% summarise('2008' = sum(vg2008, na.rm = TRUE),  
 '2012' = sum(vg2012, na.rm = TRUE),  
 '2016' = sum(vg2016, na.rm = TRUE)) %>%  
 gather(year, tgVotes, c('2008', '2012', '2016'))  
  
ggplot(data = totalRVotesL, mapping = aes(x = reorder(factor(state), tgVotes, function(x) -1\*sum(x)), y = tgVotes, fill = year)) +  
 geom\_bar(position = "dodge", stat = "identity") +  
 labs(title = "Total Republican Votes by State & Year") +   
 scale\_x\_discrete(name ="States") +   
 scale\_y\_continuous(name = "Total Republican Votes", labels = label\_number(suffix = " M", scale = 1e-6)) +  
 theme\_bw()



# Conclusion

The data set of car sales by CarDekho.com has provided various insights about the types of cars sold in the car industry and the patterns between them. The data set contains 8128 data points along with 13 features related to car details, engine details, and sale details.

# Bibliography

Vehicle dataset. (2020, October 24). Kaggle. <https://www.kaggle.com/nehalbirla/vehicle-dataset-from-cardekho>

Xie, Y. C. D. (2021, October 7). R Markdown Cookbook. R Markdown. Retrieved October 30, 2021, from <https://bookdown.org/yihui/rmarkdown-cookbook/>

Bluman, A. (2017). Elementary Statistics: A Step By Step Approach (10th ed.). McGraw-Hill Education.

Kabacoff, R., I. (2022). R in Action, Third Edition. Manning.

CarDekho. (2021). About Us | CarDekho.com. <https://www.cardekho.com/info/about_us>

F. (2021, April 2). tidyverse in r – Complete Tutorial. R-Bloggers. Retrieved October 30, 2021, from <https://www.r-bloggers.com/2021/04/tidyverse-in-r-complete-tutorial/>

D. (2021a, March 22). Data Analytics for Car Dealers. Automated Metrics. <https://www.automatedmetrics.io/data-analytics-for-car-dealers/>

# Appendix

The RMD file of the analysis is included with the analysis report.