**1. INTRODUCTION**

Metro Bike Data is a dataset of a company named “Metro Bike” situated in Los Angeles ,the data comprises of various tabulation ranging from station names, geolocation, duration of rides, pass types, etc.

Bike sharing is a service which provides available bikes as a shared use for  
individuals on a short term basis, either free or at a reasonable price. A metro  
bike sharing system allows users to rent a bike from one station and return it  
at any other station within the system. Metro bike share system have been  
deployed in various cities around the world since the second half of 20th  
century and become more popular in recent years. These systems provide  
access to bicycles for short-distance trips as an alternative to private vehicles  
or motorized public transport such as bus or subway in an urban area. In  
addition, they help reduce the traffic congestion, air pollution and noise.  
Moreover, they have been considered to solve the “last mile” problem. Finally,  
they help bridge the gap between existing transportation modes such as  
subways and bus systems and connect users to public transit networks. Beside  
the benefits mentioned above, system face many problems, one of which is  
the availability imbalance. Since movements of customers are highly dynamic,  
the bike usage is non-stationary, changing markedly with time and location.  
Therefore, some stations may be short of available bikes for rent while some  
are full and do not have enough docks for returned bikes. A general approach  
to solve this problem is that the system should monitor and redistribute bikes  
between stations frequently using trucks or bike-trailers.  
Of all the 21st century smart-city data we have, virtually none of it directly  
answers one of the most important questions about a city: Where is people  
going?  
If you think about it, we regularly count where people are at a certain point.  
There’s data for what bus stops and train stations people use. There’s data for  
vehicle traffic throughout the city, taken at specific points. Sometimes we do  
pedestrian and bicycle counts, which are laborious. In short, we have a lot of  
Point As, but virtually no Point Bs.

The data helps us with coming up a newer outlook for maxmizing the profits of the company while minimizing the losses for the company. Our team with the help of Tools such as Excel and Tableau embarked to visualize the data and propose a Business Analytics Report for the company based of previous trends inculcated to present a future abiding the commpany’s benifit.

Programming Languages such as Python and R helped us extensively at a developer’s prospect to inculcate stastical modelling and Predictive ananlysis for the common inherent goal of company’s better future.

## EXCEL

Data was very first analyzed and visualized in MS-Excel for the track and display of company past performance in the sector.

## TABLEAU

Tableau as used later for the Business Analysis and understanding consumer behavior for the same. The analysis helps the Sales Rep in building relations with customers along with a marketing and sales approach for the same.

### R

R language was used in R Studio for inculcation and calculation of statistical model of the data that is the core benign of the developer’s work from top to bottom

### PYTHON

Python Language was used in Jupyter Notebook on Google Colab for inherence of predictive analysis of the Pass Type so to assist company and Sales Reps for the Customer approach due advantage and branding where necessary.

**2.** **SCOPE OF ANALYSIS**

Scope of Analysis tells us about the limitations of the capstone we present and that can not be altered just be worked upon or avoided in a smart efficient manner as the ball is not in our court during the following aspects:

**UNEVEN DESTRIBUTION**

The ununiform no. of rides from and to a specific station because of population, geolocation , popularity, rush, etc. pose a disturbance in locomotion for using these services.

Transport vehicles such as trucks have to be used for mobility of bikes as per as demand from a specified station to the other.

**NOT PRONE TO VANDALISM**

The data does not provide the vandalism occurring to the bikes at the expense of the company though the data can be take from other source for the locality criminal rate yet it would be a failure for analysis of the real situation.

**CAN’T CHANGE CONSUMER’S BEHAVIOUR**

Certain consumer behaviors such being uninterested in riding during winters, etc. can’t be changed and the company could only work upon cutting losses for that period and nothing of that issue can be solved.

**COST EFFECTIVENESS FROM COMPETITORS**

Electric bikes pose a grave threat to the normal bikes market because of their cost effective product and easy handling and low input of man labor and the great leisure. The profit model needs to be comprehending enough to outstand the threat and remain dynamic with the change in market.

1. **EXISTING SYSTEM**
   1. **INTRODUCTION**

Existing system of our proposed capstone gives us insight upon what has already been done on the data and give us a purpose of what new could be added to the same.

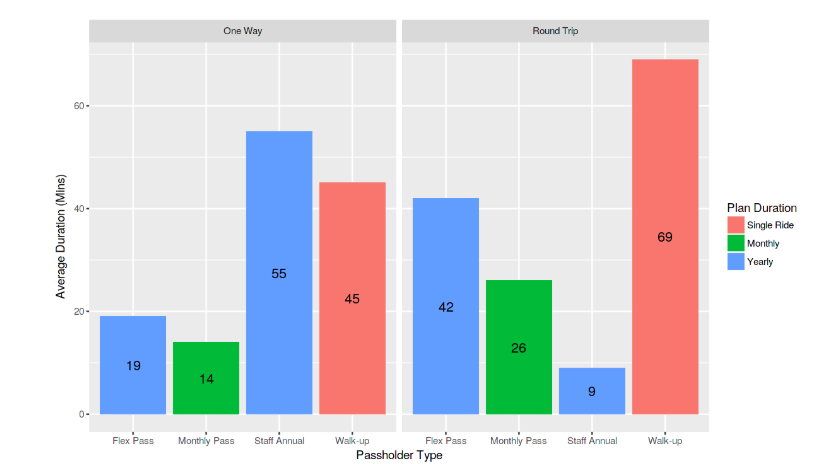
Hence marking the next step for evolution of information gathering and distribution.

3.2 **EXISTING ANALYSIS**

We picked the data from Kaggle and there are many kernels where people had already work upon the data as follows:

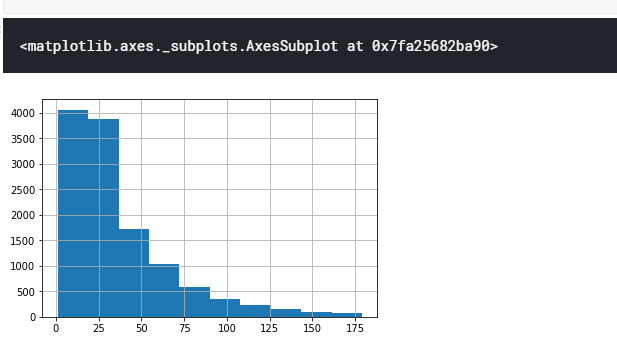
1. **Grouping and Histogram in R**

* This is one of the statistical modelling done in R which is written in one of the kernel using library knitr and repr.
* They have done in it using entities like Passholder type,
* Plan duration and Trip wise that which trip the customer will choose or prefer.
* They mostly used plan duration is monthly followed by Single ride and Yearly.
* They have used str and summary functions to understand the data.



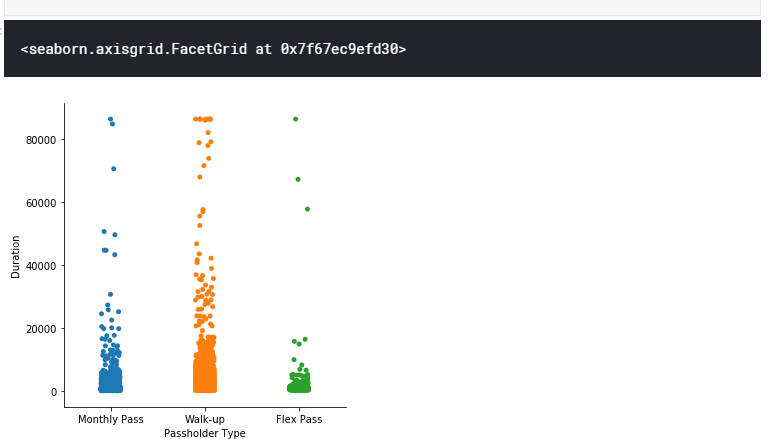
1. **Visualization in Python using Matplotlib**

* This is one of the visualization graph which is done in Python using library.
* The Visualization is done using panda, numpy and matplotlib libraries.
* They have done the Visualization using Duartion and Trip Id.



1. **Visualization in Python using Seaborn**

* In this data visualization is done in Python using Seaborn library.
* This visualization is done by taking the parameter of data like Passholder Type.
* They visualize the data upon the type of pass whether it is flex, monthly or walk up pass.
* They have use duration along with it to do the visualization.
* Some of the kernel has done the predictions in python.



1. **Prediction using Python**

* In this Kernel prediction has done using scikit learn library.
* First they had train and test the data.
* Then they used the split function to divide the data.
* They removed the NA’s value from the data.
* And then they used it for prediction in python over the type of trips.(One way/ Round)



**3.3 What’s new in the system is to be developed**

1. **Statistical Modelling using R**

=> Dropping Null Values

=> Linear Regression for contiguous data

=> Scatter Plots for contiguous data

=> Data Reshaping using melting and casting

=> Box Plots for contiguous data

=> Histograms for discrete data

=> Exploratory Data Analysis

1. **Data Visualization in Tableau**

**=>** Duration Contrast

=> Behaviour modelling

=> PassType Analysis

=> Station Congestion

=> Station Mapping

1. **Predictive Analysis in Python**

**=>** HeatMap exploration missing content

=> Distance Calculation using GeoLocation

=> SVM Prediction of Consumer Pass Type

=> Accuracy Sharending for optimum results

**4.PROBLEM ANALYSIS**

**4.1 PROBLEM DEFINITION**

* Product Title : Data Visualization and Analysis on Los Angeles Metro Bike Trip Data
* Product Type : Data Visualization and Analysis
* Live Project : No
* Company Name : Metro Bike
* Suitable For : Company
* Contents : Visualization, Data Analysis, Business Analytics
* Technology for Visualization : Tableau, Microsoft Excel
* Technology for Analysis : R Programming Language, Python
* Other Components : Predictive Analysis for Business Modelling
* Domain & Hosting : No

1. Microsoft Excel provides a simple interface of dashboard showing various visualizations and charts performed on the given data like :

* Monthly category Trip wise
* Sum total monthly wise rides
* Station Condition
* Trip category Passholder type wise

1. After performing data visualizations, several results were observed and concluded as :

* Most Busiest Month – August
* Least busiest month – February
* Busiest Station – 7th & Flower
* Least busy station – 12th & Grand
* Passes sold – 91,110
* Total Rides – 132,305
* Total Bikes – 763

1. Tableau platform provides the data visualizations in form of charts and live graphs going in to the depth of the data like:

* Passholder Type vs Starting Station
* Starting station vs Sum duration
* Trip route category vs Passholder type

1. After performing on Tableau, several results were observed and were concluded as:

* People prefer to use Monthly pass using One way trip
* People prefer to use Walk up pass for one way trip
* Usage of Metro Bike remains at top during summers (Mar-Aug)
* It gradually declines during autumn (Sep- Oct)
* People rarely tend to use Metro bike during winters (Nov-Jan)

V. R Studio used for Stastical Modelling:

* Stastical Modelling For developers’ better understanding.
* Shredding data to core binary data and visualization.
* Linear Regression model for data dependency.
* Scatter plots for interoperability among columns of data.
* Now developers can understand whole dataset without even opening the dataset.

VI. Python used for Prediction of PassType and Duration Calculation.

* Duration between Stations using Geo Location calculated.
* Heatmaps for visualization of missing data and after removal.
* Using SVM for Prediction of PassType for Sales Representative assistance.
* Scatter plots for interoperability among stations and PassType.
* Using Google colab for faster processing and reducing costs.

**4.2 FEASIBILITY ANALYSIS**

The feasibility analysis is as follows :

* Cost Estimation : 0
* Technically Feasible : Yes, all platform tools and languages are open source and easy to use.
* Profitable : Yes
* Time Feasibility : Yes, almost 1 year
* Availability : Yes, on R Studio, Tableau, Jupyter Notebook(Python)
* Accessible : Yes
* Requirements : R Studio, Tableau, Microsoft Excel, Jupyter Notebook(Python)

**4.3 PROJECT PLAN**

**Table 1: Gantt Chart for Project Management**

|  |  |  |  |
| --- | --- | --- | --- |
| TASK | EARLIEST START DATE | LENGTH | DEPENDENT UPON |
| 1. Feasibility Study | February 2019 | 1 month | None |
| 1. Excel (Data Visualization) | March 2019 | 2 months | A |
| 1. Tableau | May 2019 | 1 month | B |
| 1. R Studio | June 2019 | 2 months | C |
| 1. Python | August 2019 | 2.5 months | B,C |
| 1. Report | Oct-Nov 2019 | 1 month | ALL |

**Table 2: Duration of work represented with different colors**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **TASK** | **FEB** | **MAR** | **APR** | **MAY** | **JUNE** | **JUL** | **AUG** | **SEP** | **OCT** |
| **A** |  |  |  |  |  |  |  |  |  |
| **B** |  |  |  |  |  |  |  |  |  |
| **C** |  |  |  |  |  |  |  |  |  |
| **D** |  |  |  |  |  |  |  |  |  |
| **E** |  |  |  |  |  |  |  |  |  |

**5. SOFTWARE REQUIREMENT ANALYSIS**

**5.1 INTRODUCTION:**

A software requirement analysis or SRA is a detailed formal document which is prepared briefly in accordance of customer requirements to meet the software objectives.

In order to prepare SRA , developer meets customer to meet the software requirements. The software requirement analysis report consists of all the formal documents which are required for the entire development process of software.

To develop a software free from ambiguity, the developers should be clear with all the modules of the software requirement analysis report.

**5.2 GENERAL DESCRIPTION:**

Microsoft Excel helps to sort the raw data and remove all kinds of ambiguity and null values in the data. whereas an excel dashboard represents different kinds of graphs and charts which gives the clear results and more refined observations are concluded from the results which helps in business analytics to any organization.

Tableau is the most powerful, and fastest growing tool dedicated for the work of data visualization for the Business analysis. At first, it simplifies the raw data in much simple format. Data analysis with Tableau becomes much more faster as well as complex.charts and graphs are visualized dynamically with the help of in-built strut fitting algorithms.

R is the no. one stastical modelling tool currently in the field of Data Science which is used for the exploratory analysis of the data with a full back up of mathematical libraries which make the analysis pretty easy which would be serving as the backbone of our capstone project.

Google Colab is an open source initiative by Google which give free access to 12GB Ram and GPUs for free to individuals for running their model at fast pace.

Jupyter Notebook is a concise IDE that stores and lets you code and compile your model step by step in blocks backed by all Python libraries pre installed at your expense.

**5.3 SPECIFIC REQUIREMENTS:**

This subsection contains the requirements for the tools and platforms used. These requirements are organized by features of product function.

**TABLEAU**

* Microsoft Windows 7 or newer.
* 2 GB memory.
* 1.5 GB minimum free disk space.
* Microsoft server 2008 R2 or newer.

**R-STUDIO**

* 2 GB RAM.
* 10 GB free disk space.
* Microsoft Windows 7 or newer.

**JUPYTER NOTEBOOK(on Google Colab)**

* Good Internet Connection.
* Working PC
* Google Account

1. **IMPLEMENTATION**
2. **Implementation in R**

* We have done Statistical modelling using R.
* At first we have done Data Preprocessing.
* We have plotted the graphs according to round trip and one way trip using ggplot Library.
* First of all we find and removed the NA’s value from the data.
* Getting the structure and summary of the data using str and summary function.
* Calculate the dimension, number of rows of the data and Overview of data in R.
* Filtered the values route and pass wise .
* Reshape the data using melting and casting Functions.
* Linear Regression on data is done by using R.
* Plotly library is used to plot the graphs over Duration, Trip Route Category wise, Trip Id.
* Plotted the graphs according to ID, Pass and Route wise.
* Creating the Histogram over use time.
* We have done the exploratory data analysis.
* Created graphs for statistical modelling using Scatterplot and Boxplot.

1. **Implementation in Tableau**

* We used the tableau software to make the Business Analytics Model for the Company.
* At first we import the excel file using connect option in Tableau.
* We have two main data fields in Tableau Dimensions and Measures.
* We create the Duration Contrast in tableau by putting starting station in column and duration in minute in row attribute.
* This very visualization helps us to understand the duration in total that takes place during momentum. It will tell us about the market knowledge, profit and growth of the company.
* We have done Pass Visualization by putting Trip Route Category into column and Passholder type along with duration in row wise.
* It will tell us about categorization of our consumers based on their choice of ride or Pass for e.g Major customer choose walk-up ride over pass.
* We put the starting station into column and ending station into row wise attribute to create Station Rush.
* It will help us to know the stations relationship according to customers locomotion.
* This will help the company to grow in the market by creating new plans, customers and market expansion.
* Using plan duration and duration in minute with starting date we create the consumer behavior.
* It will tell us the actions and thinking of the consumer towards company.
* We have created the Station Mapping using the Trip route category along with Pass holder type over starting station,
* Which will help us to understand Sales Representative of the whole market model that how to increase profits.

1. **Implementation in Python**

* We have used Python for Predictive analysis of data using Jupyter Notebook and Google Colab.
* We are using numpy, panda and scikit learn libraries for implementation of predictive analytics.
* At first we imported all these libraries for e.g import panda as pd, import numpy as np.
* Heat Maps for missing data Visualization.
* Performed Data Cleaning.
* Calculated Distance Using Geolocation.
* Scatter Plot for Distance vs. PassType
* Trained Model.
* Tested Model.
* Working on random generation with split ration of 1:3
* Used SVM(Support Vector Machine) for Prediction of PassType.
* Achieved accuracy above 99.96%.

**7. PROJECT LEGACY**

**7.1 CURRENT STATUS OF PROJECT**

The Project is fully complete, we have met all set of requirements, goals and tasks that we originally aimed for.

The Project is working fine in all aspects with the set of goals that we once embarked upon at the beginning.

The prediction model for Machine Learning is working with great accuracy above than **99.96%.**

Statstical Modelling shows a positive outlook for what it was made for in R. Any developer can know now each and every bit of information without even looking at the dataset.

Business Analytics Report in Tableau shows an upward trend in area of improvement for the company and behavior modelling to sets par as confirmed by other datasets and analysis of the people of Los Angeles.

**7.2 TECHNICAL AND MANEGERIAL LESSONS LEARNT**

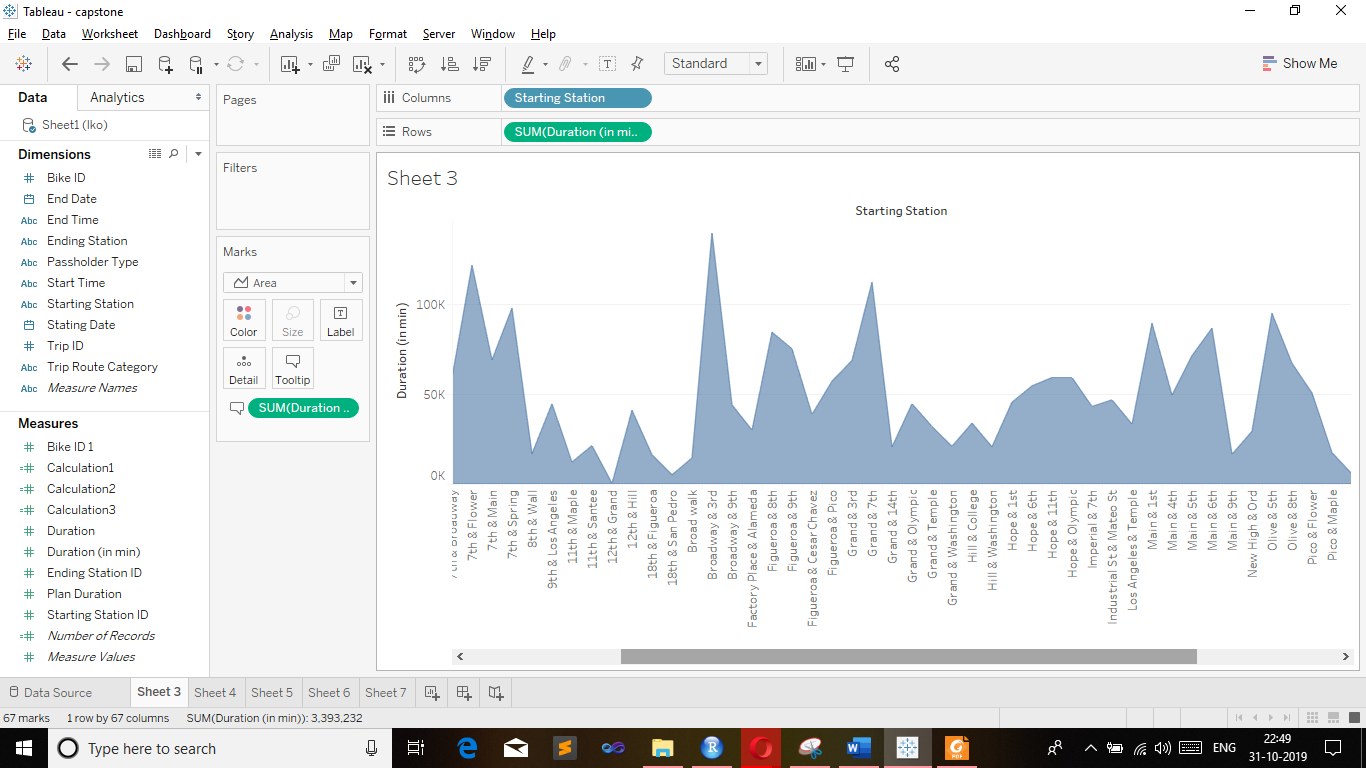
The very journey of our Capstone from beginning to end was a hayride full with challenges, lessons, achievements, benchmarks, stress, happiness, hardship, etc. The following are the lessons that we learnt during this journey:

* Google Colab is a wonder for model building that costs you nothing.
* Importance of Data Visualization in shaping up Company.
* Corporate Standards for coding.
* Team work is much fruitful than self saga.
* Python is having key role in the future.
* Importance of statistical modelling for a developer.
* When in problem, it’s ok to seek help.
* Experience has an upper hand over newly built enthusiasm.
* Importance of commenting your code in corporate sector.
* Willpower can make you emerge out as winners always.
* Code can break, you must not!
* Assignment of tasks as per skills make work easier.
* Importance of time management.
* In field of technology one needs to constantly upgrade himself with current market trends.

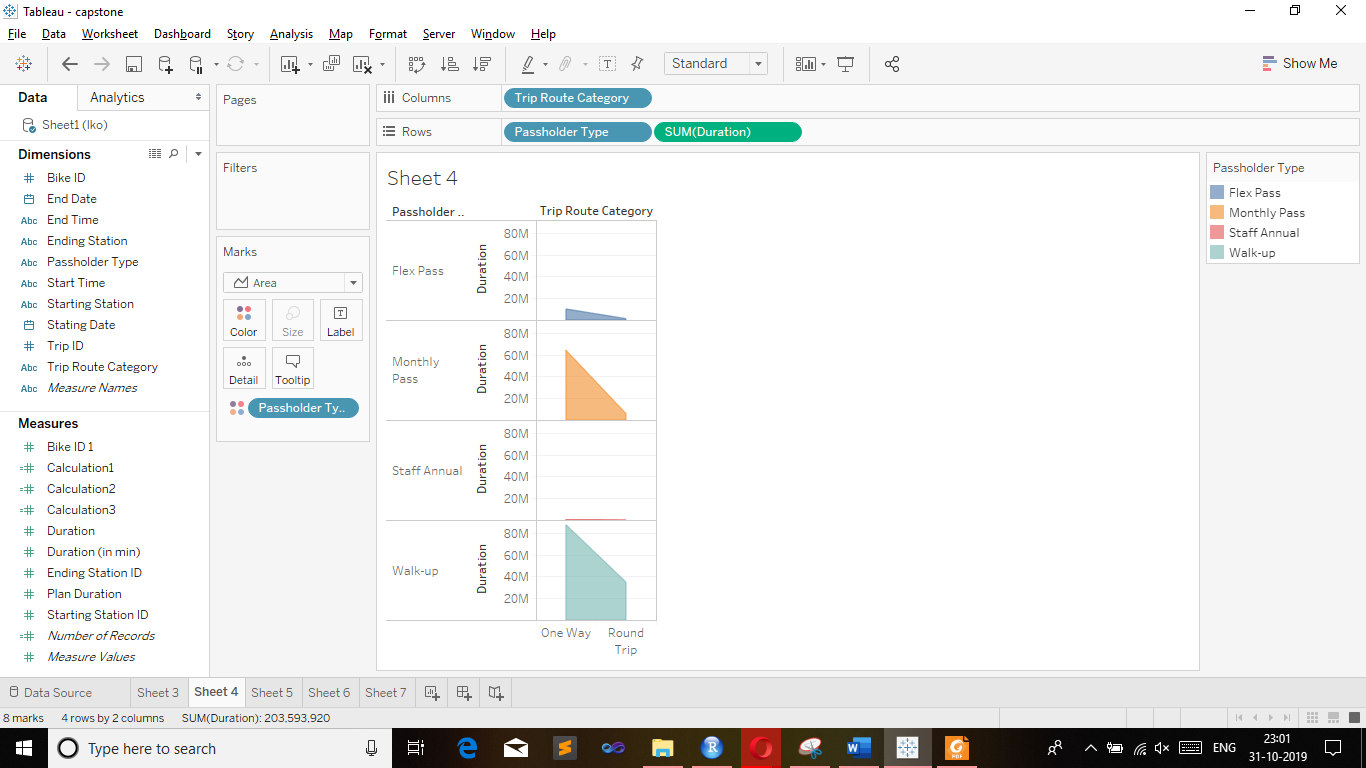
**8. SOURCE CODE**

**8.1 TABLEAU**

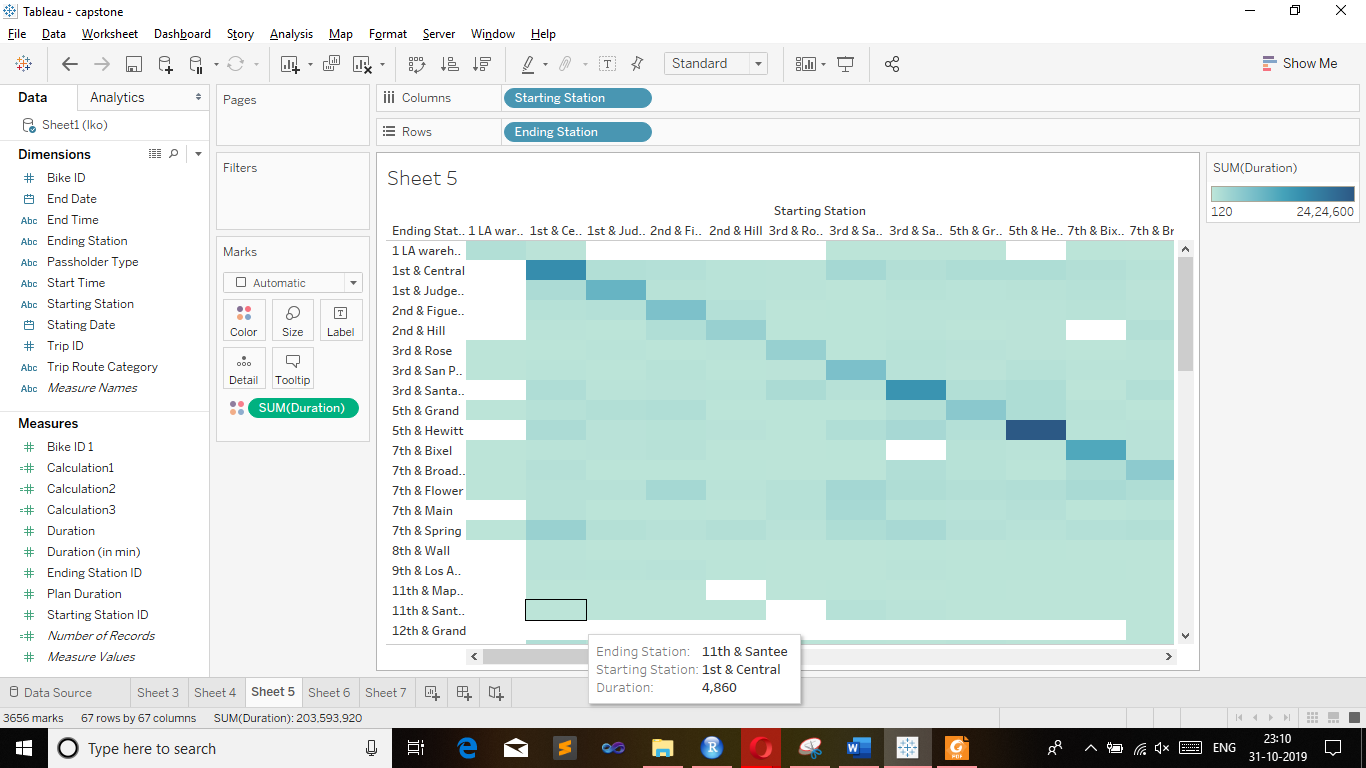
**8.1.1 Duration Contrast**

****

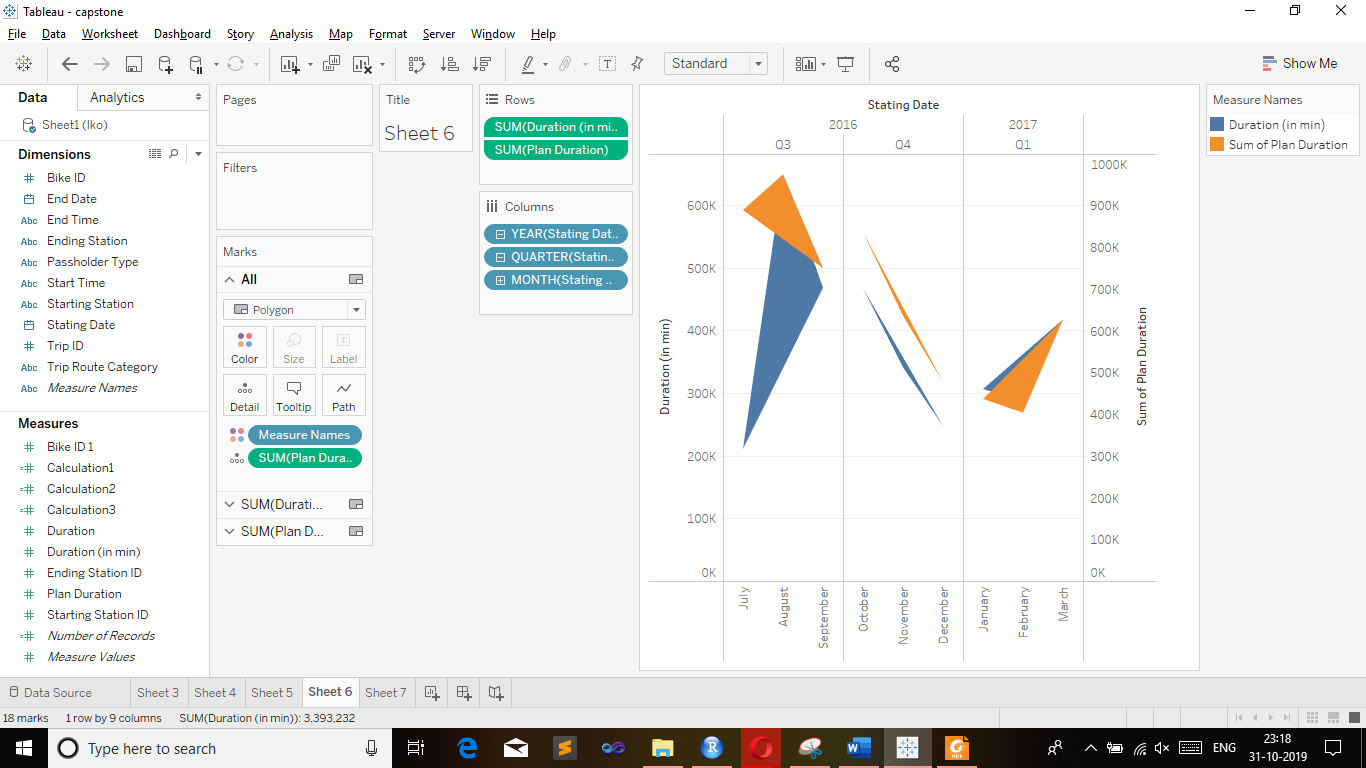
**8.1.2 Pass Visualization**



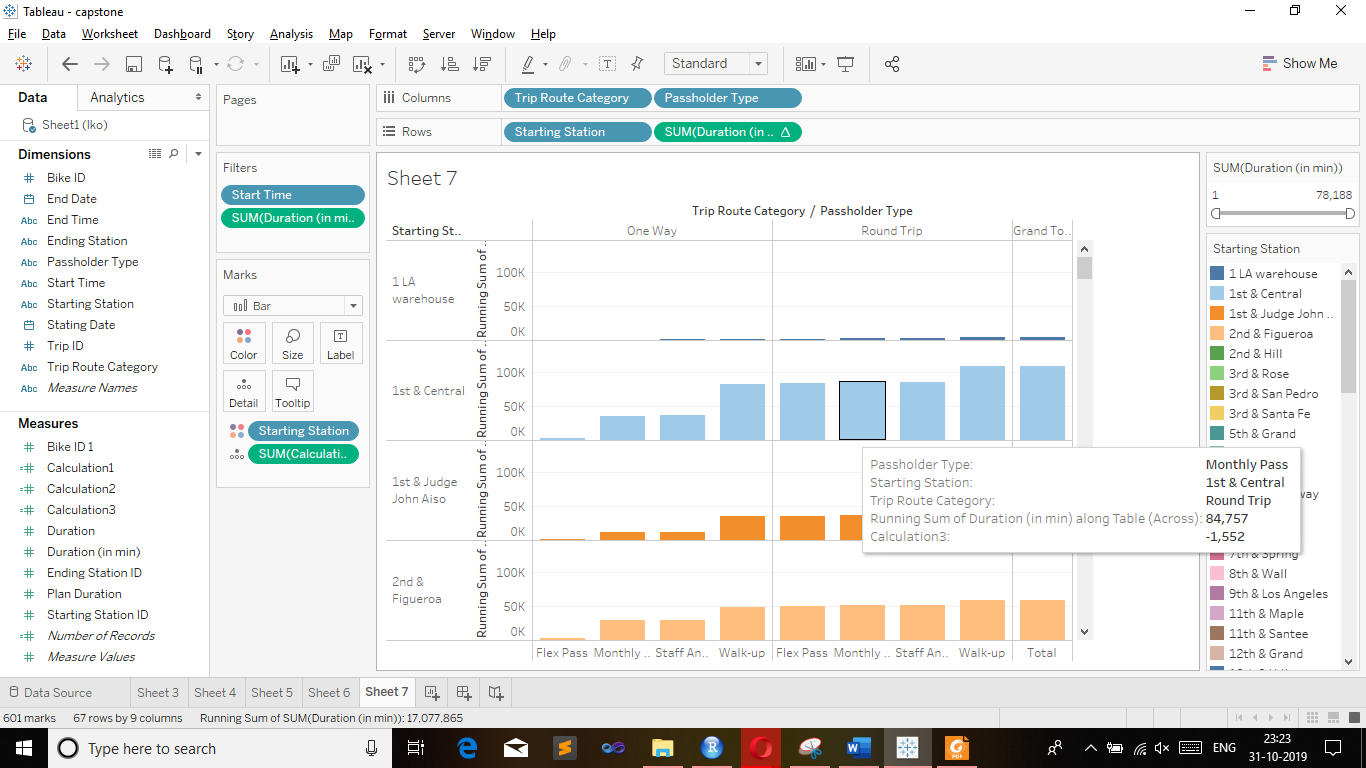
**8.1.3 Station Rush**



**8.1.4 Consumer Behavior**



**8.1.5 Station Mapping**



**8.2 R STUDIO CODE**

# metapackage with lots of helpful functions

library(tidyverse)

library(ggplot2)

library(data.table)

library(dplyr)

library(tidyr)

library(plotly)

#Reading the Data

data=read.csv(file.choose(), header = TRUE)

View(data)

#Data Preprocessing

nrow(data)

dim(data)

colnames(data)

head(data)

tail(data)

str(data)

glimpse(data)

#Getting summary of Data

summary(data)

#Missing Values in Dataset

sapply(data, function(x) sum(is.na(x)))

sapply(data, FUN = class)

#New dataset without missing values

newdataset = data[complete.cases(data),]

View(newdataset)

#Exploratory Data Analysis

#Browsing the Dataset

library(DT)

datatable(data =newdataset ,

rownames = FALSE,

filter = "top",

options = list(autoWidth = TRUE))

#plotting

table(data$Passholder.Type)

qplot(data$Passholder.Type)

table(data$Trip.Route.Category)

qplot(data$Trip.Route.Category)

head(data)

head(data$Start.Time)

library(lubridate)

data$usetime<-ymd\_hms(data$End.Time)-ymd\_hms(data$Start.Time)

data$usetime<-as.numeric(data$usetime)

print(data$usetime)

summary(data$usetime)

#boxplot

boxplot(data$usetime)$stats

#Checking NA values

data$usetime<-ifelse(data$usetime<=1|data$usetime>=36,NA, data$usetime)

print(data$usetime)

table(is.na(data$usetime))

bike<- data %>% filter(!is.na(usetime))

print(bike)

ggplot(data=data, aes(x=Passholder.Type, y=usetime))+geom\_boxplot()+ggtitle("usetime by pass holder")

#Histogram

hist(data$usetime, breaks = 30,

main = "Using time", col = "Grey", xlab="time", ylab = "number")

#Scatterplot between duration and starting date

attach(data)

plot(Stating.Date,Duration,main="Scatterplot",xlab="Duration",ylab="Starting Date",pch=10)

#group\_by

ggplot(data=data, aes(x=Passholder.Type, y=usetime))+geom\_boxplot()+ggtitle("summary using time by Passholder Type")

#Roundtrip

round\_journey<- bike %>%

filter(Trip.Route.Category=="Round Trip")

round\_journey %>%

group\_by(Passholder.Type) %>%

summarise(mean=mean(usetime))

head(round\_journey)

#passholder type round\_journey

ggplot(data=round\_journey, aes(x=Passholder.Type, y=usetime))+geom\_boxplot()+ggtitle("round\_journey")

#Onewaytrip

oneway\_journey<- bike %>%

filter(Trip.Route.Category=="One Way")

oneway\_journey %>%

group\_by(Passholder.Type) %>%

summarise(mean=mean(usetime))

#passholder type Oneway\_journey

ggplot(data=oneway\_journey, aes(x=Passholder.Type, y=usetime))+geom\_boxplot()+ggtitle("oneway\_journey")

#oneway\_journey starting station & Ending station

tail(oneway\_journey)

table(oneway\_journey$Starting.Station.ID)

table(oneway\_journey$Ending.Station.ID)

oneway\_starting\_station\_freq<-ggplot(data=oneway\_journey,

aes(x=oneway\_journey$Starting.Station.ID))+geom\_bar(fill='#FF9900')+

coord\_cartesian(xlim = c(3000, 3100))

ggplotly(oneway\_starting\_station\_freq)

oneway\_ending\_station\_freq<-ggplot(data=oneway\_journey,

aes(x=oneway\_journey$Ending.Station.ID))+geom\_bar()+coord\_cartesian(xlim = c(3000, 3100))

ggplotly(oneway\_ending\_station\_freq)

#oneway starting.station id | passholder type

oneway\_starting\_station\_id<- bike %>%

filter(Trip.Route.Category=="One Way") %>%

select(Starting.Station.ID,Passholder.Type)

#starting station ID

oneway\_starting\_station\_count<-oneway\_starting\_station\_id %>%

group\_by(Starting.Station.ID) %>%

tally()

#oneway ending station id | passholder type

oneway\_ending\_station\_id<- bike %>%

filter(Trip.Route.Category=="One Way") %>%

select(Ending.Station.ID, Passholder.Type)

#ending station id

oneway\_ending\_station\_count<-oneway\_ending\_station\_id %>%

group\_by(Ending.Station.ID) %>%

tally()

#

starting\_count\_new<-oneway\_starting\_station\_count

starting\_count\_new<-rename(starting\_count\_new, station = Starting.Station.ID)

starting\_count\_new<-rename(starting\_count\_new, startcount = n)

#

ending\_count\_new<-oneway\_ending\_station\_count

#

ending\_count\_new<-rename(ending\_count\_new, station = Ending.Station.ID)

ending\_count\_new<-rename(ending\_count\_new, endingcount = n)

#

starting\_count\_new<-starting\_count\_new[-c(65,66),]

ending\_count\_new<-ending\_count\_new[-c(65,66),]

#Left Join

total<-left\_join(starting\_count\_new, ending\_count\_new, by="station")

library(reshape2)

total\_1<-melt(total, id=(c("station")))

f1<-ggplot(data=total\_1, aes(x=station))

f2<-f1+geom\_bar(aes(y=startcount), colour="Red", size=1)

f3<-f2+geom\_bar(aes(y=endingcount), colour="blue", size=1)

head(total\_1)

h1<-ggplot(data=total\_1, aes(x=station, y=value, fill=variable))+geom\_bar(stat="identity", position = "dodge")+xlim(x=c(3000,3100))

ggplotly(h1)

h1

#Linear Regression

x<-c(data$Ending.Station)

print(x)

y<-c(data$Duration)

print(y)

relation<-lm(y~x)

summary(relation)

result<-predict(relation)

print(result)

#Plotting graphs using plotly

library(plotly)

plot\_ly(data=data,x=~Trip.ID,y=~Duration,type ="scatter",mode="markers")

plot\_ly(data=data,x=~Trip.ID,y=~Duration,type ="scatter",mode="markers",

color=I("black"))

plot\_ly(data=data,x=Trip.ID,y=~Duration,type ="scatter",mode="markers",

marker=list

(color="green",size=2))

plot\_ly(data=data,x=~Trip.ID,y=~Duration,type ="scatter",mode="markers",

marker = list(size = 7,

color = 'rgba(255, 182, 193, .9)',

line = list(color = 'rgba(152, 0, 0, .8)',

width = 0.5)))

route=plot\_ly(data=data, x = ~Trip.Route.Category, color = ~Stating.Date, type = "box")

print(route)

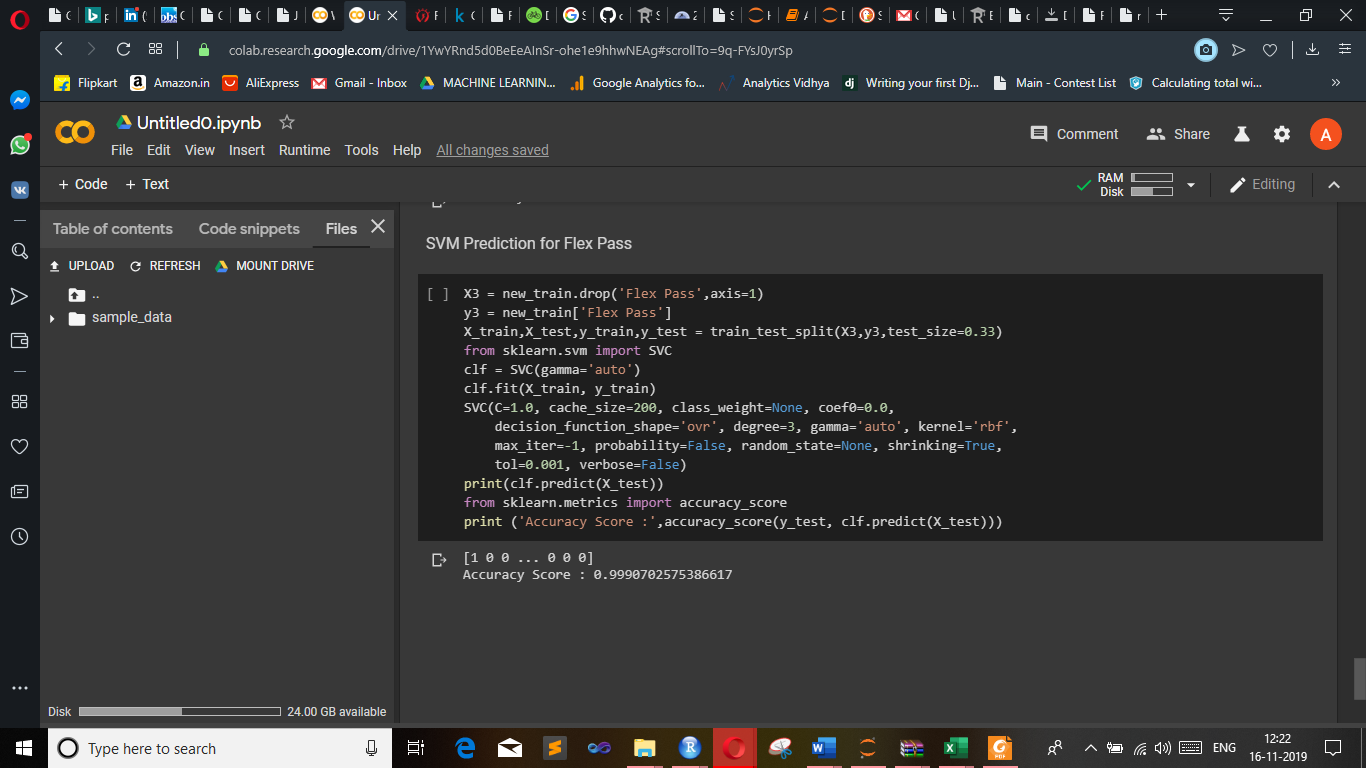
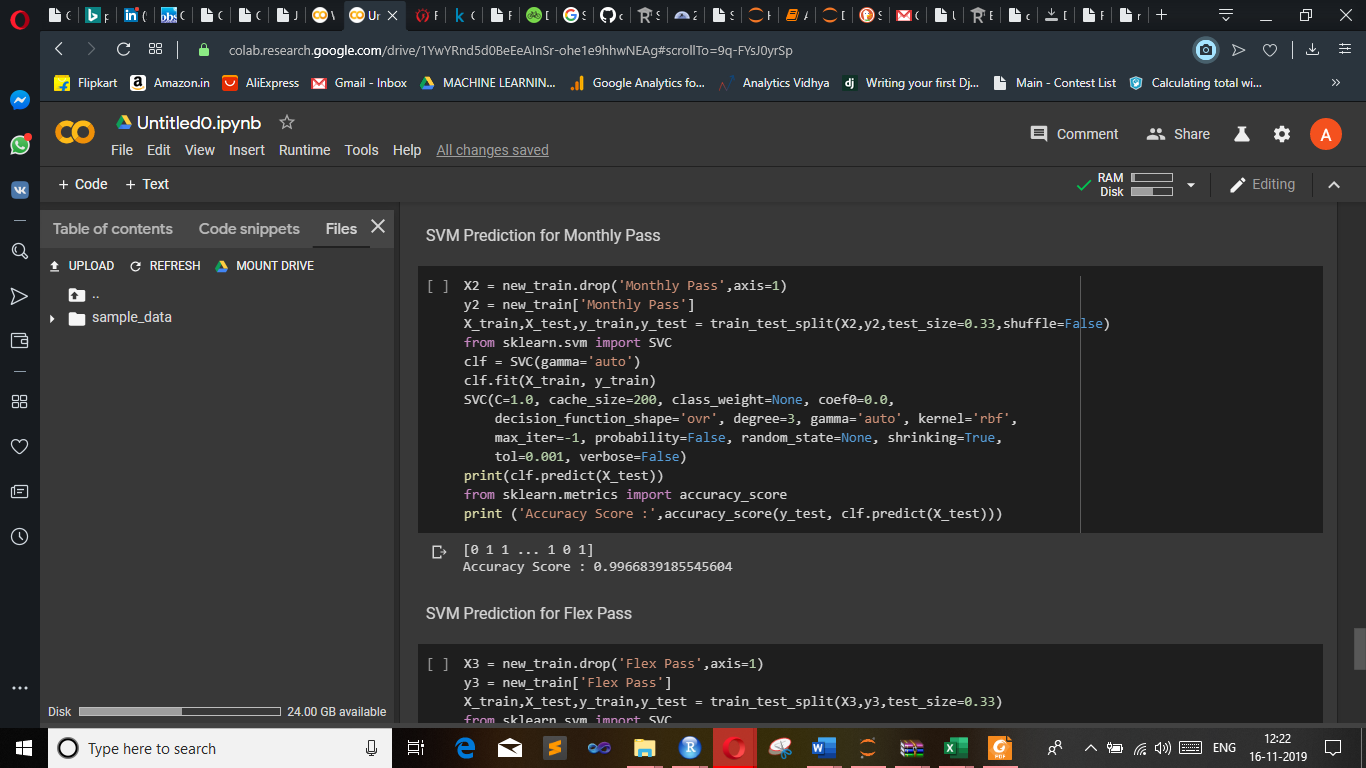
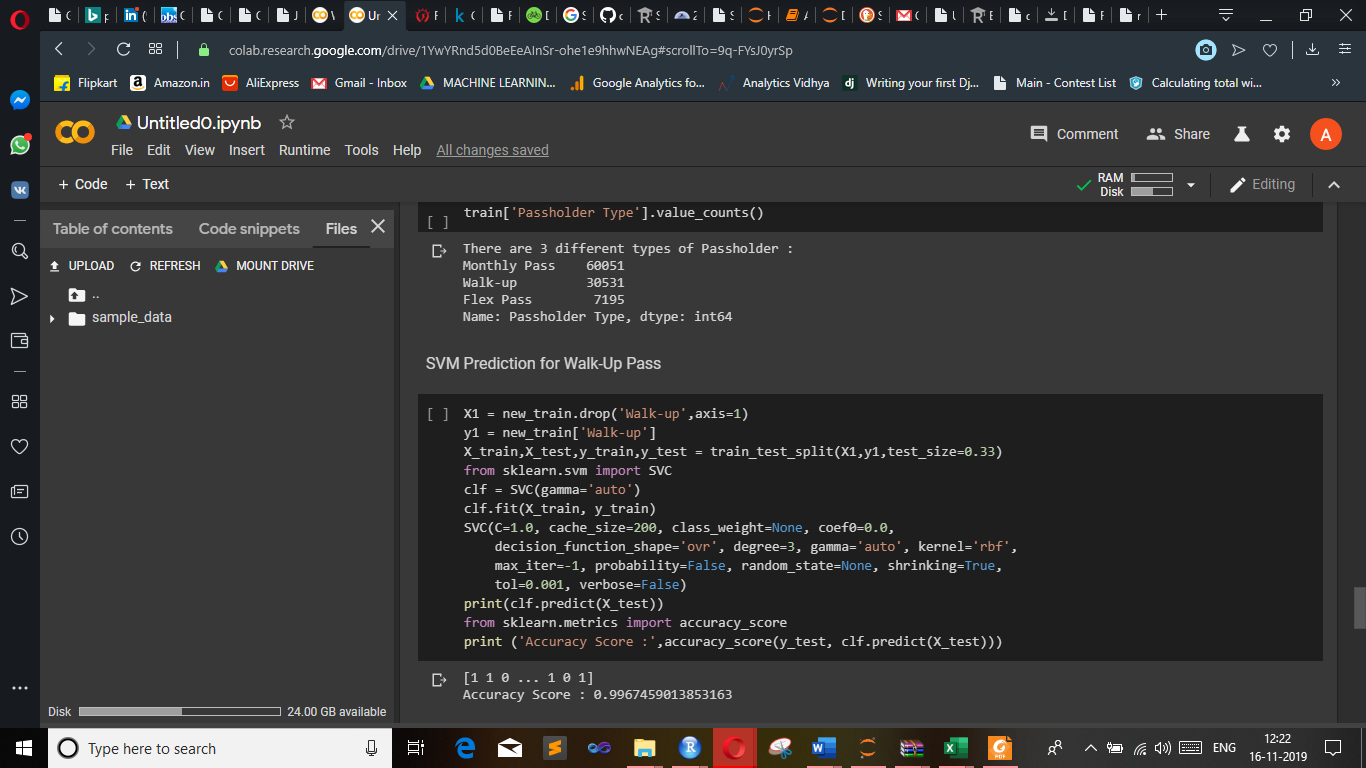
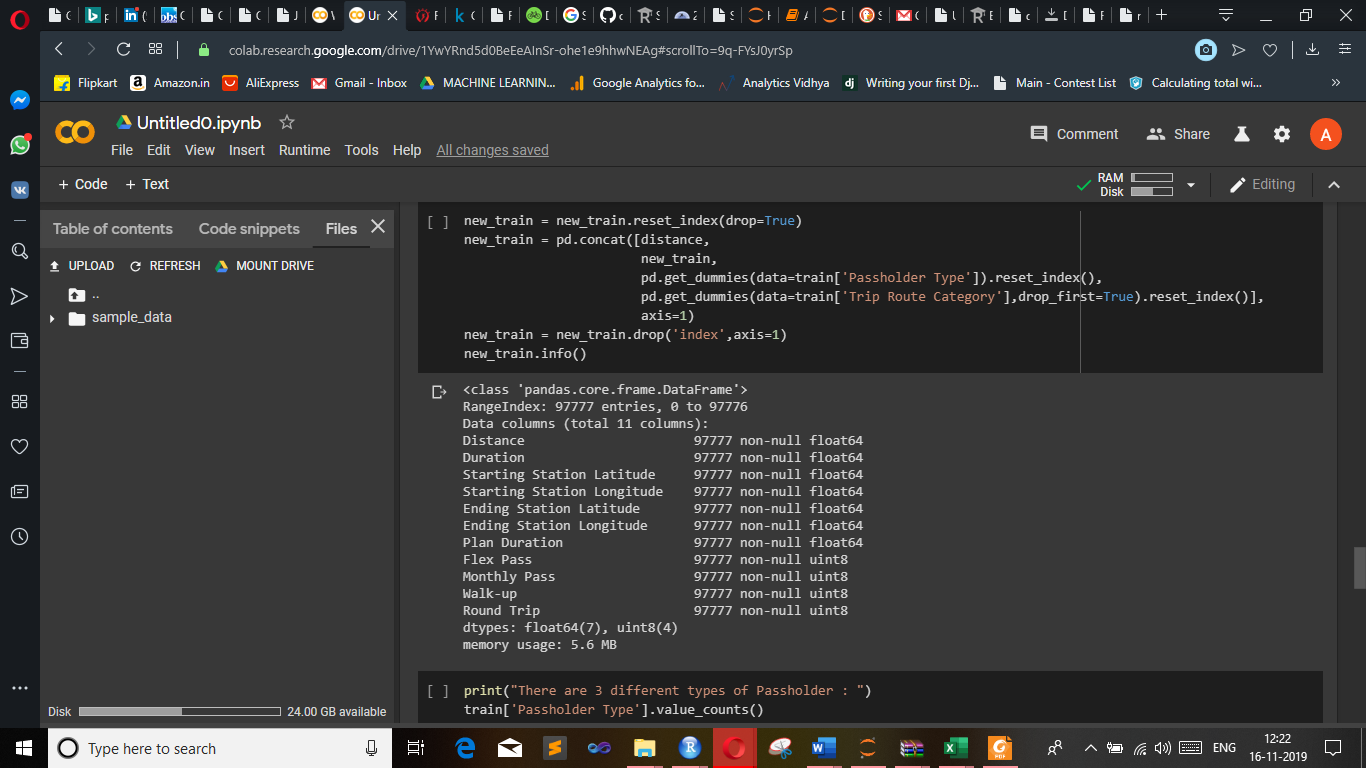
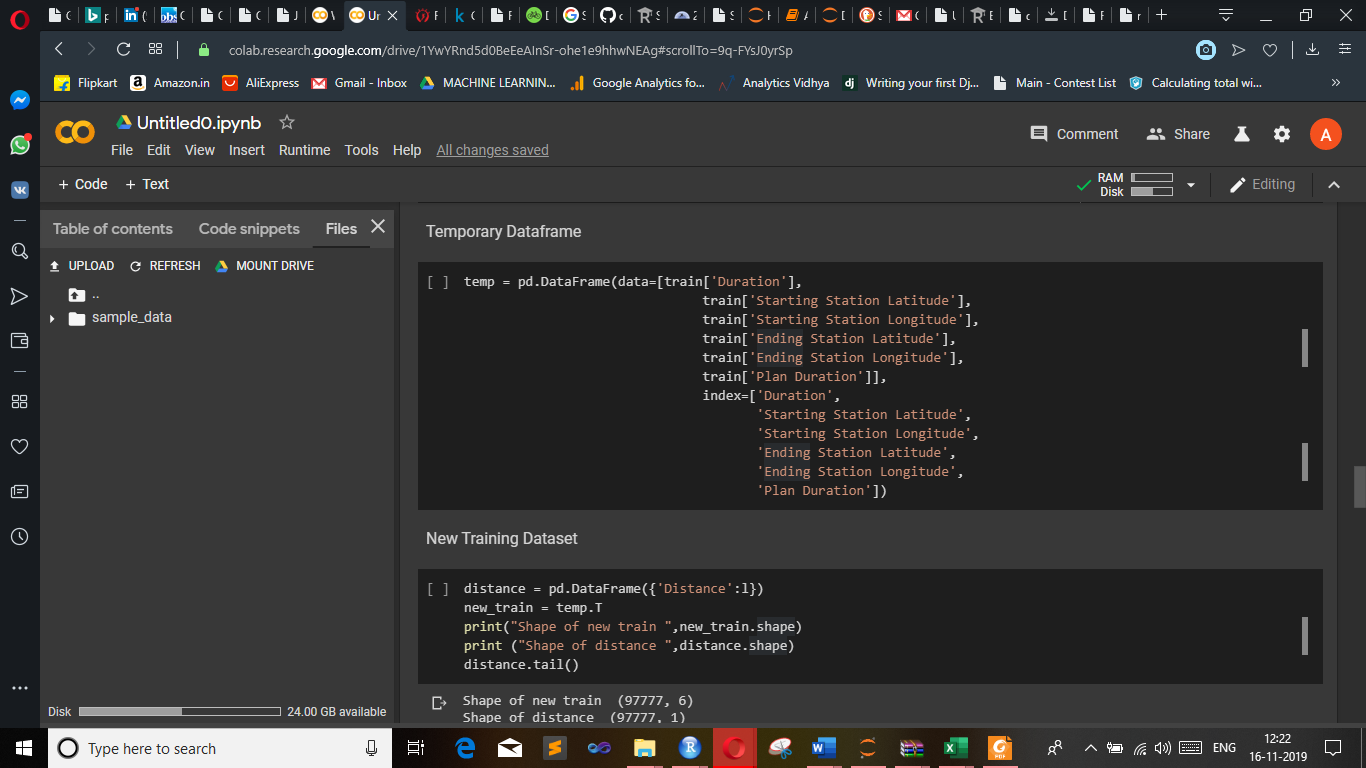
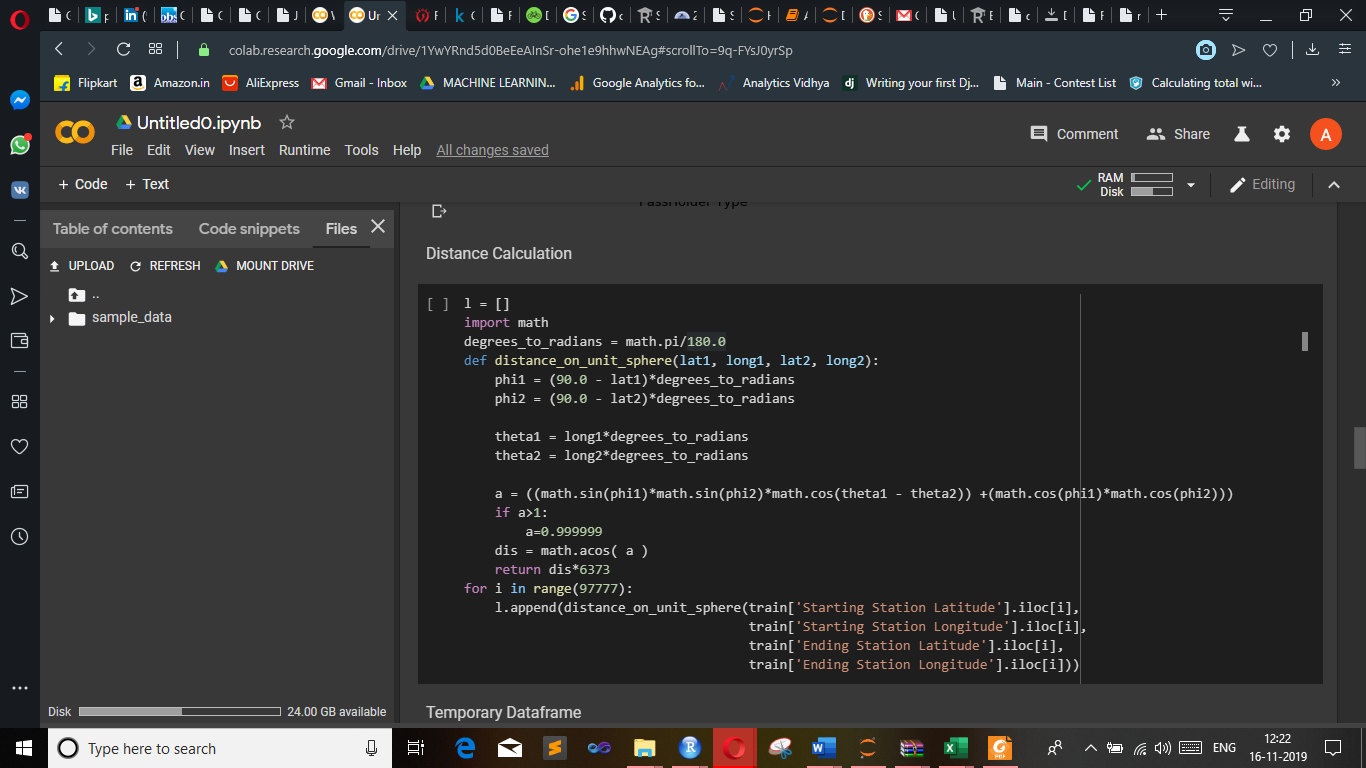
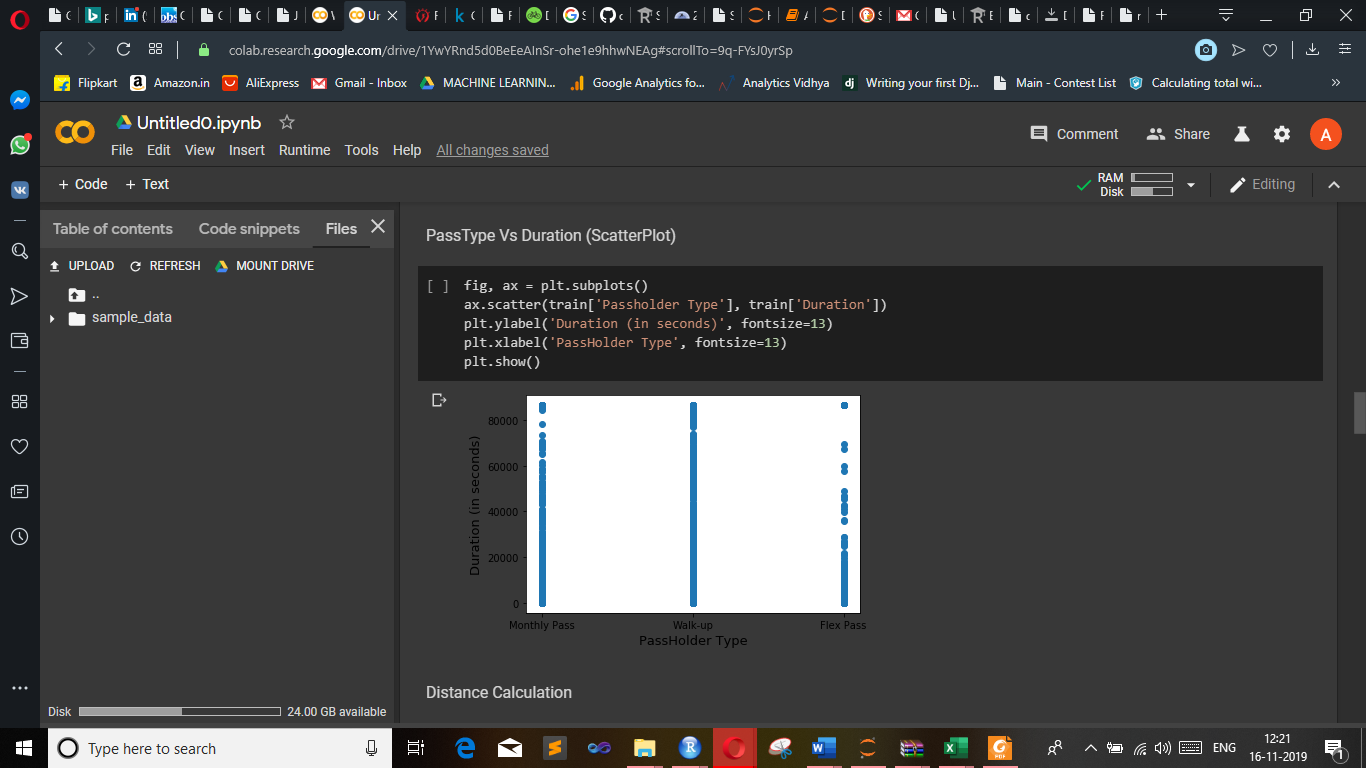
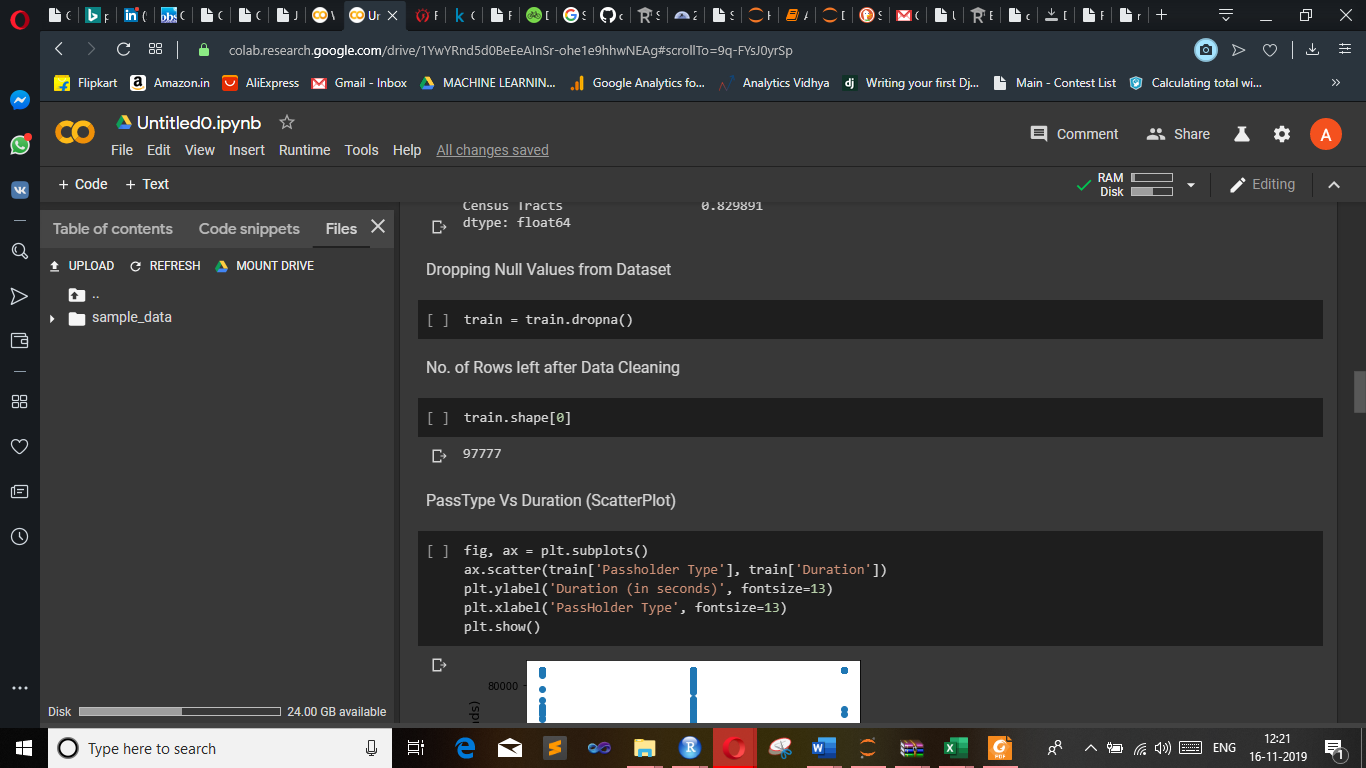
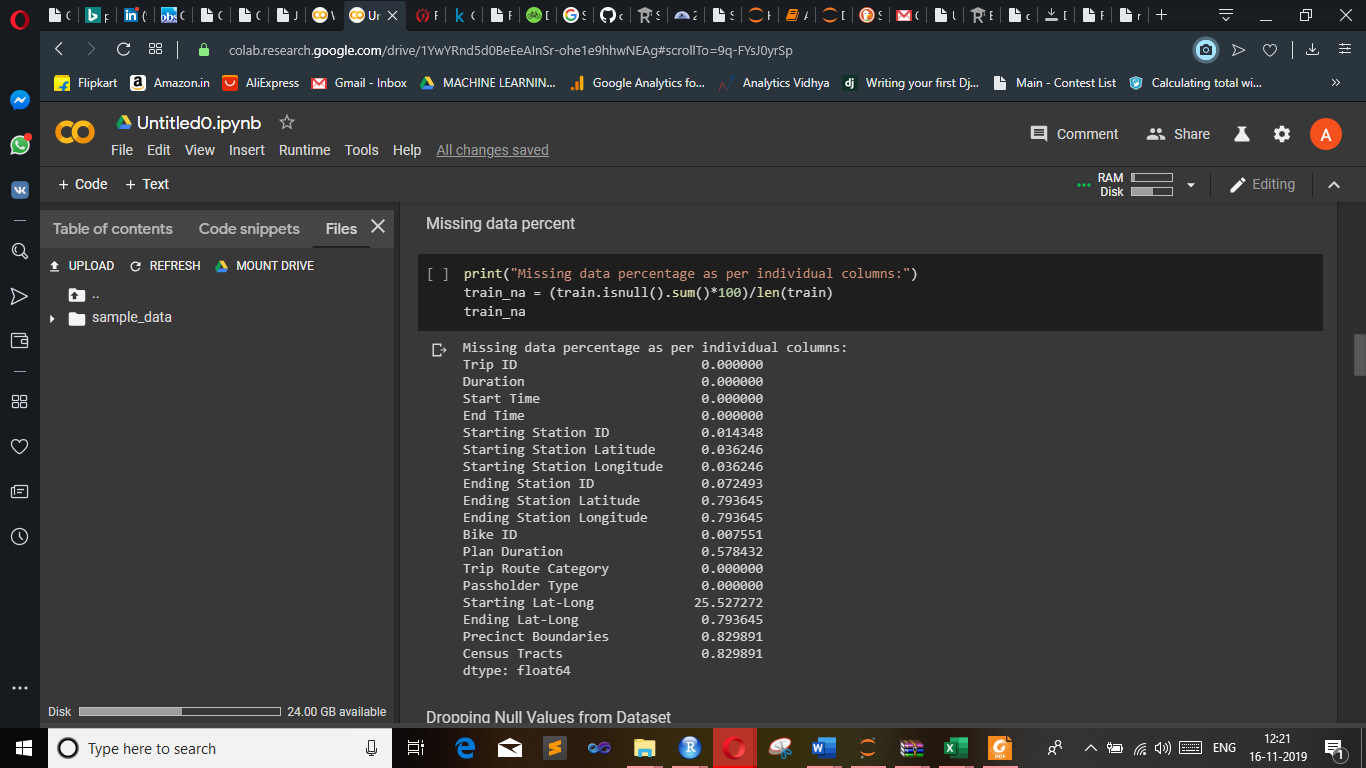
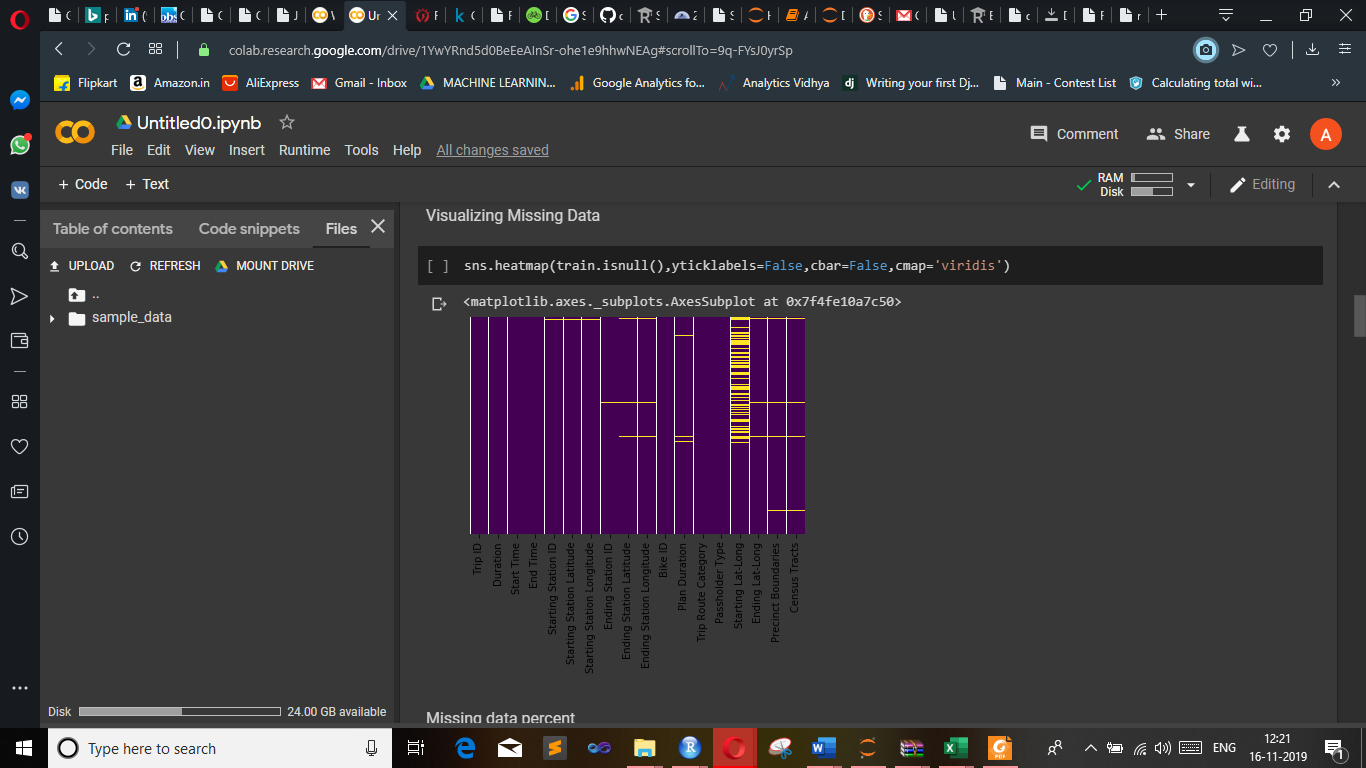
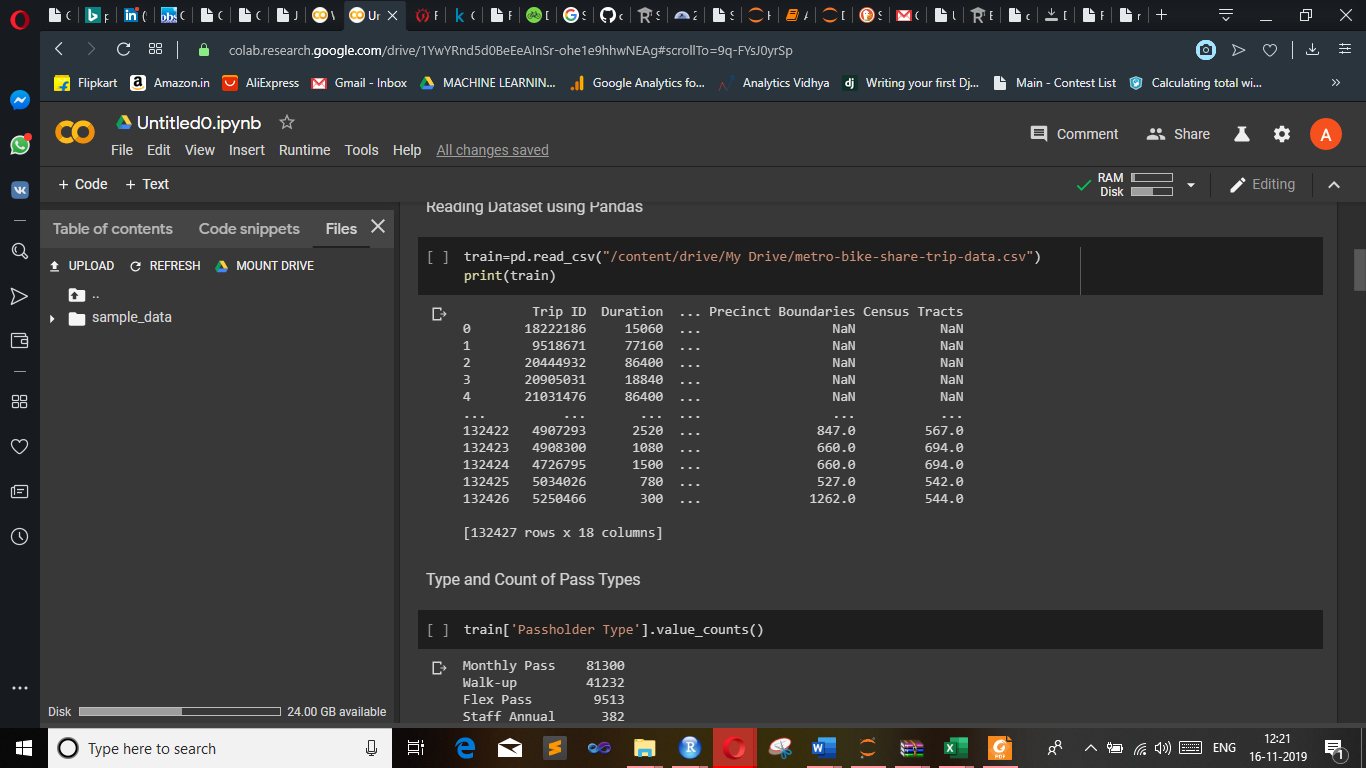
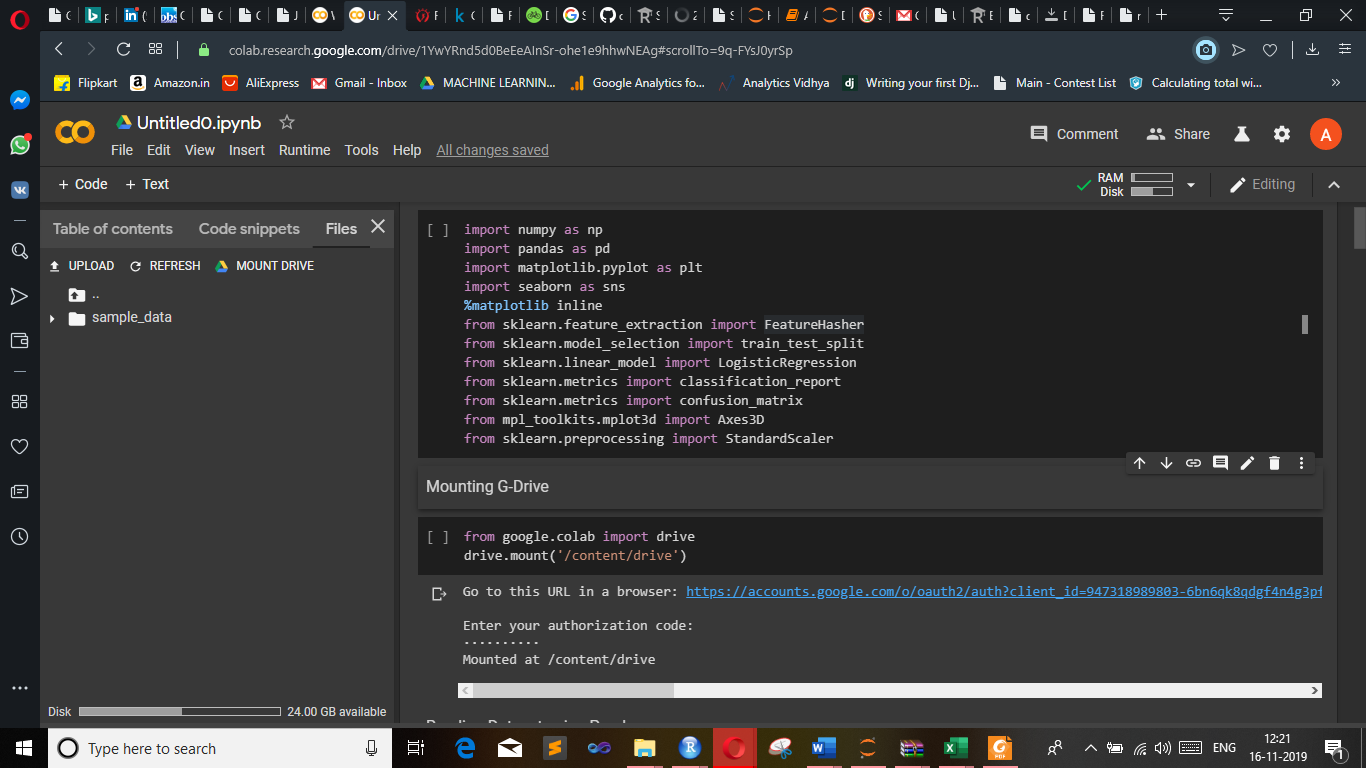
ID=plot\_ly(data=data, x = ~Trip.ID, color = ~Stating.Date, type = "box")

print(ID)

Pass=plot\_ly(data=data, x = ~Passholder.Type, color = ~Stating.Date, type = "box")

print(Pass)

**8.3 JUPYTER NOTEBOOK (PYTHON3)**



**9. BIBLIOGRAPHY**

* [**https://www.edureka.co/blog/scikit-learn-machine-learning/**](https://www.edureka.co/blog/scikit-learn-machine-learning/)
* [**https://www.youtube.com/user/sentdex**](https://www.youtube.com/user/sentdex)
* [**https://www.datacamp.com/tracks/data-scientist-with-r**](https://www.datacamp.com/tracks/data-scientist-with-r)
* [**https://scikit-learn.org/stable/documentation.html**](https://scikit-learn.org/stable/documentation.html)
* [**https://www.rdocumentation.org/**](https://www.rdocumentation.org/)
* [**https://www.tableau.com/learn/training**](https://www.tableau.com/learn/training)
* [**https://www.tutorialspoint.com/tableau/index.htm**](https://www.tutorialspoint.com/tableau/index.htm)
* [**https://medium.com/deep-learning-turkey/google-colab-free-gpu-tutorial-e113627b9f5d**](https://medium.com/deep-learning-turkey/google-colab-free-gpu-tutorial-e113627b9f5d)
* [**https://github.com/changwookjun/StudyBook/blob/master/DeepLearningBooks/oreilly-hands-on-machine-learning-with-scikit-learn-and-tensorflow-1491962291.pdf**](https://github.com/changwookjun/StudyBook/blob/master/DeepLearningBooks/oreilly-hands-on-machine-learning-with-scikit-learn-and-tensorflow-1491962291.pdf)