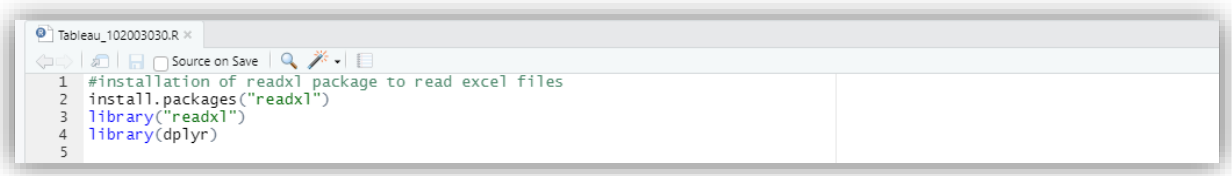
AIR TRAVEL TABLEAU PROJECT

NAME: MRITUNJAY DUBEY

SUBGROUP: 3CO2

ROLL NO: 102003030

Installation and Using Packages



Datasets:

Table 1

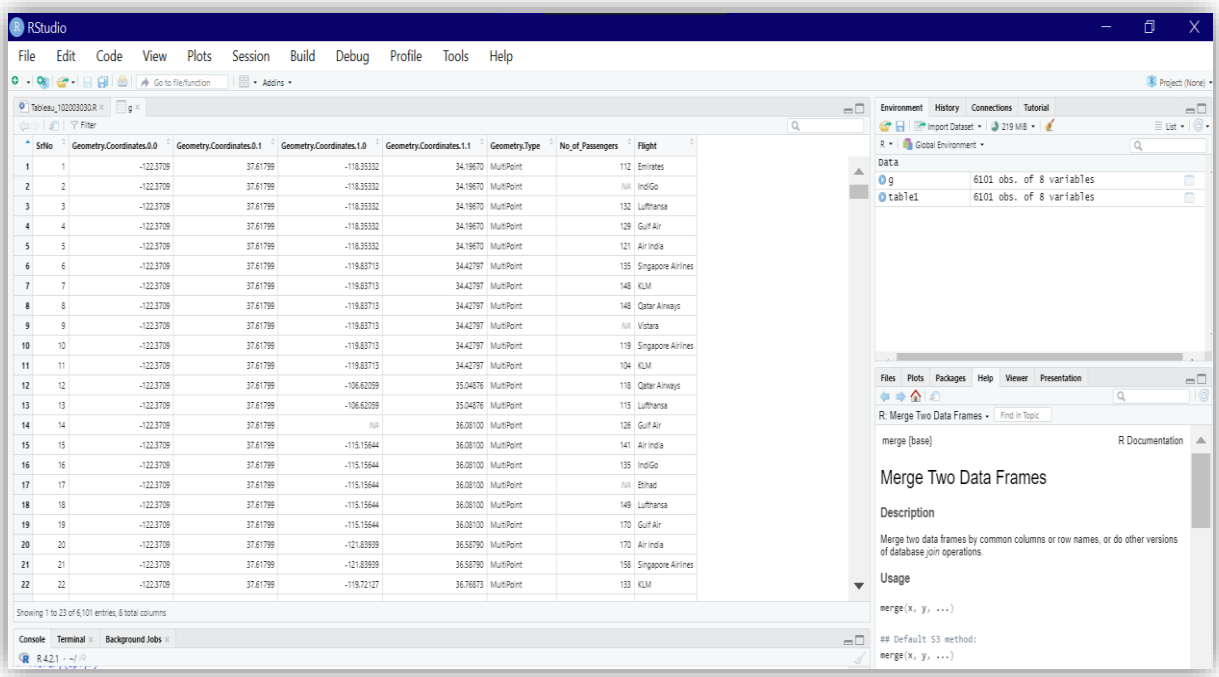
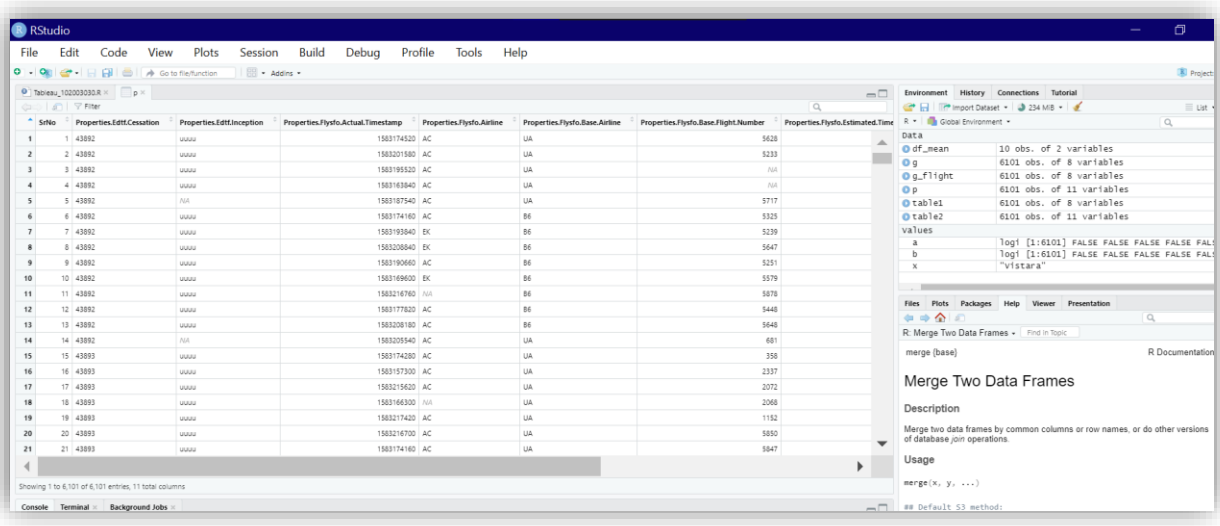


Table 2


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Environment History Connections Tutorial
Data
df_mean 10 obs. of 2 variables
g 6101 obs. of 8 variables
g_flight 6101 obs. of 8 variables
p 6101 obs. of 11 variables
table1 6101 obs. of 8 variables
table2 6101 obs. of 11 variables
values
a logi [1:6101] FALSE FALSE FALSE FALSE FAL
b logi [1:6101] FALSE FALSE FALSE FALSE FAL
x "vistara"
Files Plots Packages Help Viewer Presentation
R: Merge Two Data Frames
merge {base}
R Documentation
Merge Two Data Frames
Description
Merge two data frames by common columns or row names, or do other versions of database join operations.
Usage
merge(x, y, ...)
Default S3 method:

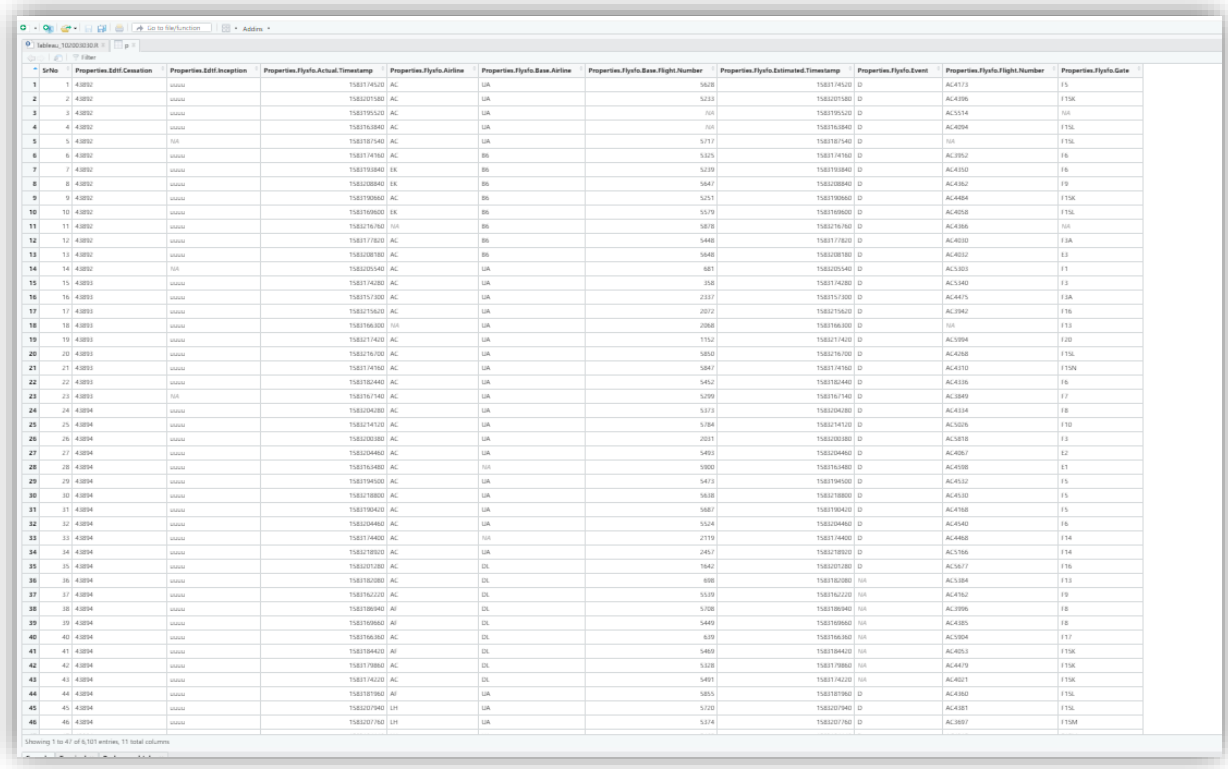
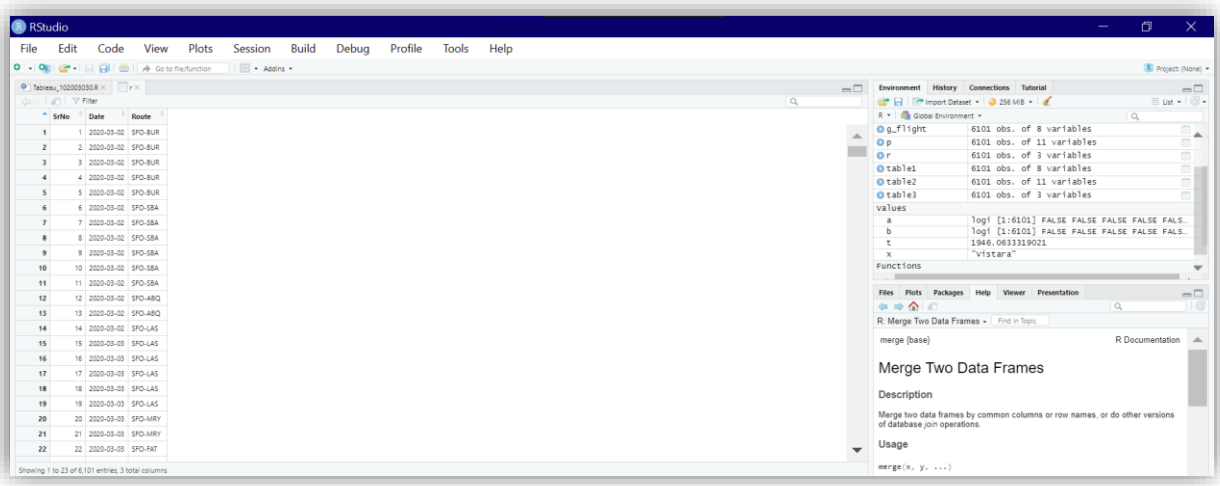


Table 3



Data Pre-Processing:

```
#For Table 1
table1<- read_excel("C:\\Users\\MRITUNJAY\\Desktop\\Tableau Project\\Table1.xlsx")
g<-data.frame(table1)  #storing data into g as data frame
View(g)

#Replacing NA values with Mean
g$Geometry.Coordinates.1.0[is.na(g$Geometry.Coordinates.1.0)]<-mean(g$Geometry.Coordinates.1.0,na.rm

g_flight<- group_by(g,Flight)

df_mean<-summarise(g_flight,passengers=mean(No_of_Passengers, na.rm = TRUE))

for(x in df_mean$Flight)
{
  a<-is.na(g$No_of_Passengers)
  b<-g$Flight==x
  g$No_of_Passengers[a&b]=floor(df_mean$passengers[df_mean$Flight==x])
}
View(g)

#For Properties Table
table2<- read_excel("C:\\Users\\MRITUNJAY\\Desktop\\Tableau Project\\Table2.xlsx")
p<-data.frame(table2)  #storing data into p as data frame
View(p) #Before preprocessing
#Preprocessing of properties table-->

#Replacing NA values with Mean
t<-mean(p$Properties.Flysfo.Base.Flight.Number,na.rm=TRUE)
p$Properties.Flysfo.Base.Flight.Number[is.na(p$Properties.Flysfo.Base.Flight.Number)]<-trunc(t)
```

```r
#Declaring function to calculate Mode
Mode<-function(x){
  ux<-na.omit(unique(x))
  tab<-tabulate(match(x,ux)); ux[tab==max(tab)]
}

#Replacing NA values with Mode
p$Properties.Edtf.Inception[is.na(p$Properties.Edtf.Inception)]<-Mode(p$Properties.Edtf.Inception)

p$Properties.Flysfo.Airline[is.na(p$Properties.Flysfo.Airline)]<-Mode(p$Properties.Flysfo.Airline)

p$Properties.Flysfo.Base.Airline[is.na(p$Properties.Flysfo.Base.Airline)]<-Mode(p$Properties.Flysfo.

p$Properties.Flysfo.Event[is.na(p$Properties.Flysfo.Event)]<-Mode(p$Properties.Flysfo.Event)

p$Properties.Flysfo.Flight.Number[is.na(p$Properties.Flysfo.Flight.Number)]<-Mode(p$Properties.Flysf

p$Properties.Flysfo.Gate[is.na(p$Properties.Flysfo.Gate)]<-Mode(p$Properties.Flysfo.Gate)

View(p) #After preprocessing


#For Route Table
table3<- read_excel("C:\\Users\\MRITUNJAY\\Desktop\\Tableau Project\\Table3.xlsx")
r<-data.frame(table3)  #storing data into r as data frame
View(r)

#Spliting the Route column into two as "From" and "To" using stringr
library(stringr)
s<-str_split_fixed(r$Route, "-", 2)
```

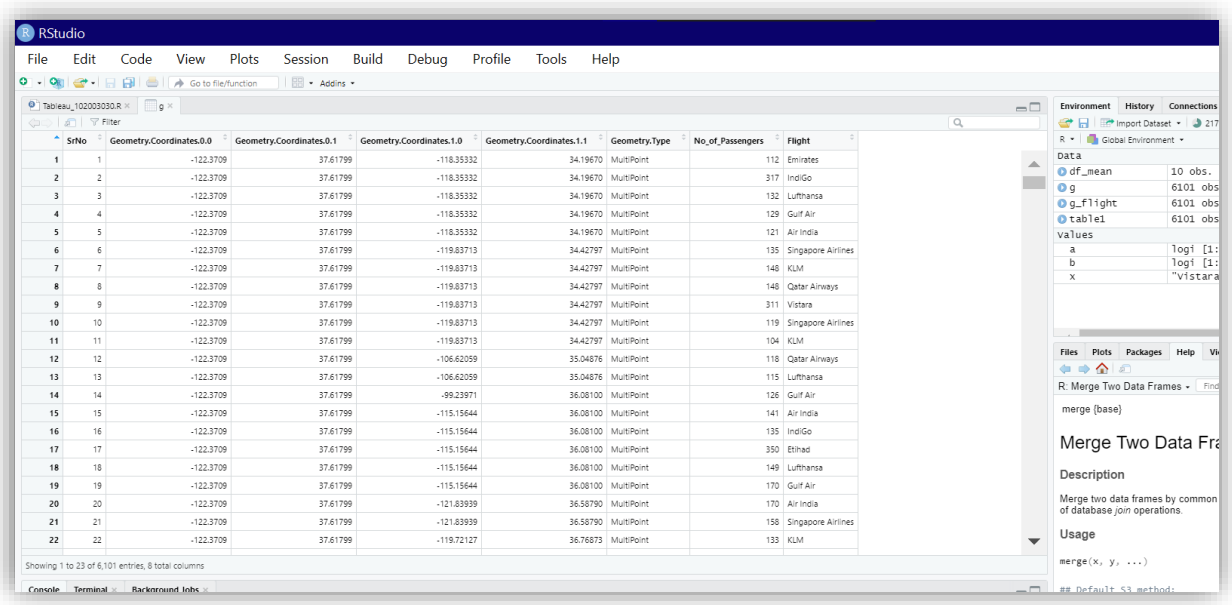
```r

#For Route Table
table3<- read_excel("C:\\Users\\MRITUNJAY\\Desktop\\Tableau Project\\Table3.xlsx")
r<-data.frame(table3)  #storing data into r as data frame
View(r)

#Spliting the Route column into two as "From" and "To" using stringr
library(stringr)
s<-str_split_fixed(r$Route, "-", 2)
colnames(s)=c("From","To")
d<-cbind.data.frame(r,s)
View(d)
```

Tables After Pre-Processing:

Table 1



	SrNo	Geometry.Coordinates.0.0	Geometry.Coordinates.0.1	Geometry.Coordinates.1.0	Geometry.Coordinates.1.1	Geometry.Type	No_of_Passengers	Flight
1	1	-122.3709	37.61799	-118.35332	34.19670	MultiPoint	112	Emirates
2	2	-122.3709	37.61799	-118.35332	34.19670	MultiPoint	317	IndiGo
3	3	-122.3709	37.61799	-118.35332	34.19670	MultiPoint	132	Lufthansa
4	4	-122.3709	37.61799	-118.35332	34.19670	MultiPoint	129	Gulf Air
5	5	-122.3709	37.61799	-118.35332	34.19670	MultiPoint	121	Air India
6	6	-122.3709	37.61799	-119.83713	34.42797	MultiPoint	135	Singapore Airlines
7	7	-122.3709	37.61799	-119.83713	34.42797	MultiPoint	148	KLM
8	8	-122.3709	37.61799	-119.83713	34.42797	MultiPoint	148	Qatar Airways
9	9	-122.3709	37.61799	-119.83713	34.42797	MultiPoint	311	Vistara
10	10	-122.3709	37.61799	-119.83713	34.42797	MultiPoint	119	Singapore Airlines
11	11	-122.3709	37.61799	-119.83713	34.42797	MultiPoint	104	KLM
12	12	-122.3709	37.61799	-106.62059	35.04876	MultiPoint	118	Qatar Airways
13	13	-122.3709	37.61799	-106.62059	35.04876	MultiPoint	115	Lufthansa
14	14	-122.3709	37.61799	-99.23971	36.08100	MultiPoint	126	Gulf Air
15	15	-122.3709	37.61799	-115.15644	36.08100	MultiPoint	141	Air India
16	16	-122.3709	37.61799	-115.15644	36.08100	MultiPoint	135	IndiGo
17	17	-122.3709	37.61799	-115.15644	36.08100	MultiPoint	350	Etihad
18	18	-122.3709	37.61799	-115.15644	36.08100	MultiPoint	149	Lufthansa
19	19	-122.3709	37.61799	-115.15644	36.08100	MultiPoint	170	Gulf Air
20	20	-122.3709	37.61799	-121.83939	36.58790	MultiPoint	170	Air India
21	21	-122.3709	37.61799	-121.83939	36.58790	MultiPoint	158	Singapore Airlines
22	22	-122.3709	37.61799	-119.72127	36.76873	MultiPoint	133	KLM

Table 2

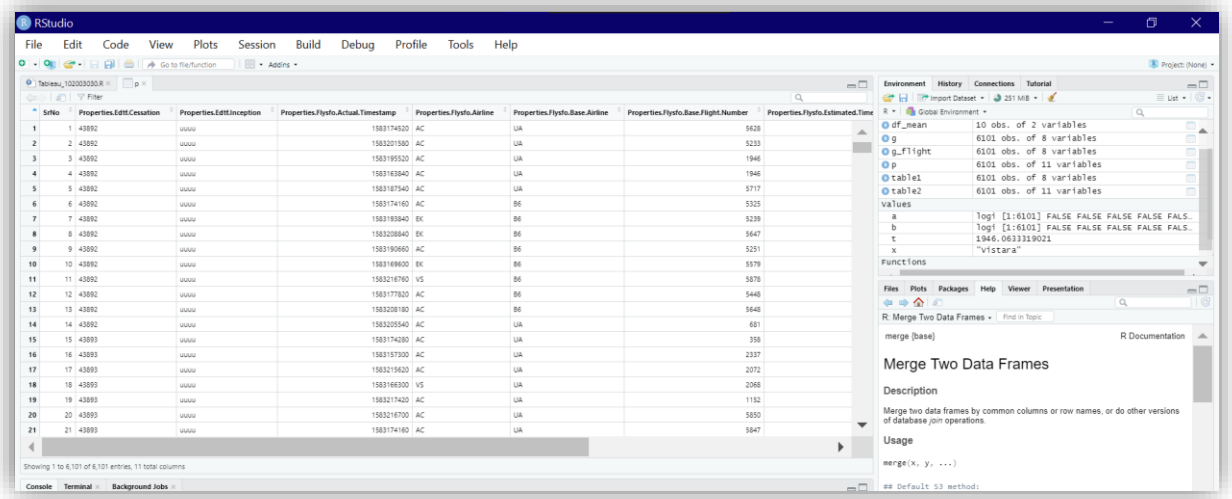
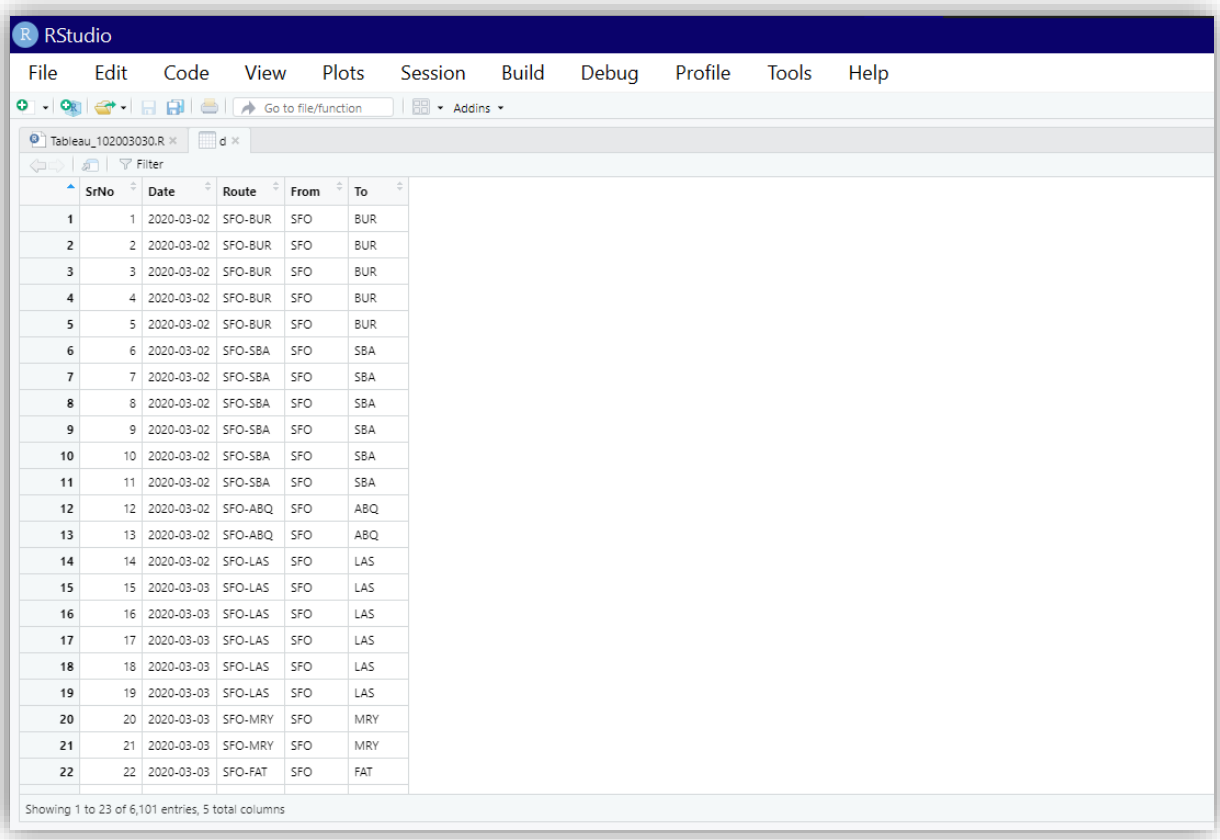


Table 3



	SrNo	Date	Route	From	To
1	1	2020-03-02	SFO-BUR	SFO	BUR
2	2	2020-03-02	SFO-BUR	SFO	BUR
3	3	2020-03-02	SFO-BUR	SFO	BUR
4	4	2020-03-02	SFO-BUR	SFO	BUR
5	5	2020-03-02	SFO-BUR	SFO	BUR
6	6	2020-03-02	SFO-SBA	SFO	SBA
7	7	2020-03-02	SFO-SBA	SFO	SBA
8	8	2020-03-02	SFO-SBA	SFO	SBA
9	9	2020-03-02	SFO-SBA	SFO	SBA
10	10	2020-03-02	SFO-SBA	SFO	SBA
11	11	2020-03-02	SFO-SBA	SFO	SBA
12	12	2020-03-02	SFO-ABQ	SFO	ABQ
13	13	2020-03-02	SFO-ABQ	SFO	ABQ
14	14	2020-03-02	SFO-LAS	SFO	LAS
15	15	2020-03-03	SFO-LAS	SFO	LAS
16	16	2020-03-03	SFO-LAS	SFO	LAS
17	17	2020-03-03	SFO-LAS	SFO	LAS
18	18	2020-03-03	SFO-LAS	SFO	LAS
19	19	2020-03-03	SFO-LAS	SFO	LAS
20	20	2020-03-03	SFO-MRY	SFO	MRY
21	21	2020-03-03	SFO-MRY	SFO	MRY
22	22	2020-03-03	SFO-FAT	SFO	FAT

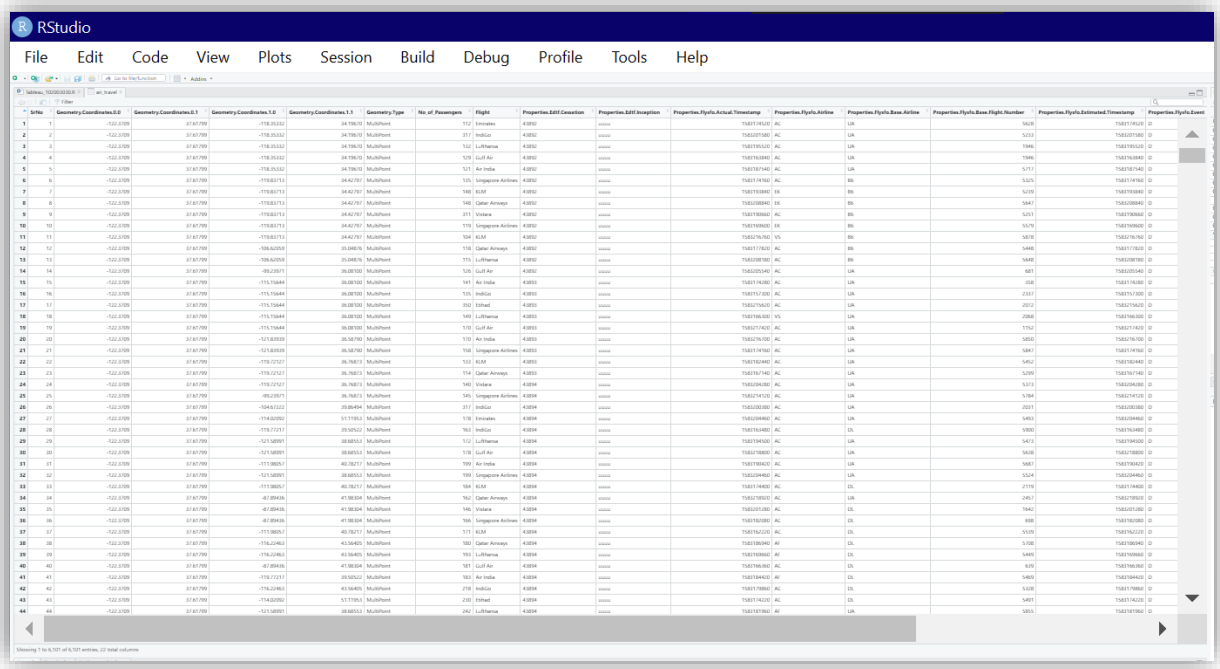
Showing 1 to 23 of 6,101 entries, 5 total columns

Merging All Three Tables and Putting in Excel:

```
#Merging of Tables
f<-merge(g,p,by="SrNo")
air_travel<-merge(f,d,by="SrNo")
View(air_travel)

#Formatting Date Column
air_travel$Date<-format(as.POSIXct(air_travel$Date), format="%d-%m-%y")
View(air_travel)

#Putting Dataframe in Excel File
install.packages("writexl")
library("writexl")
write_xlsx(air_travel,"C:\\Users\\MRITUNJAY\\Desktop\\Tableau Project\\air_travel.xlsx")
```


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help

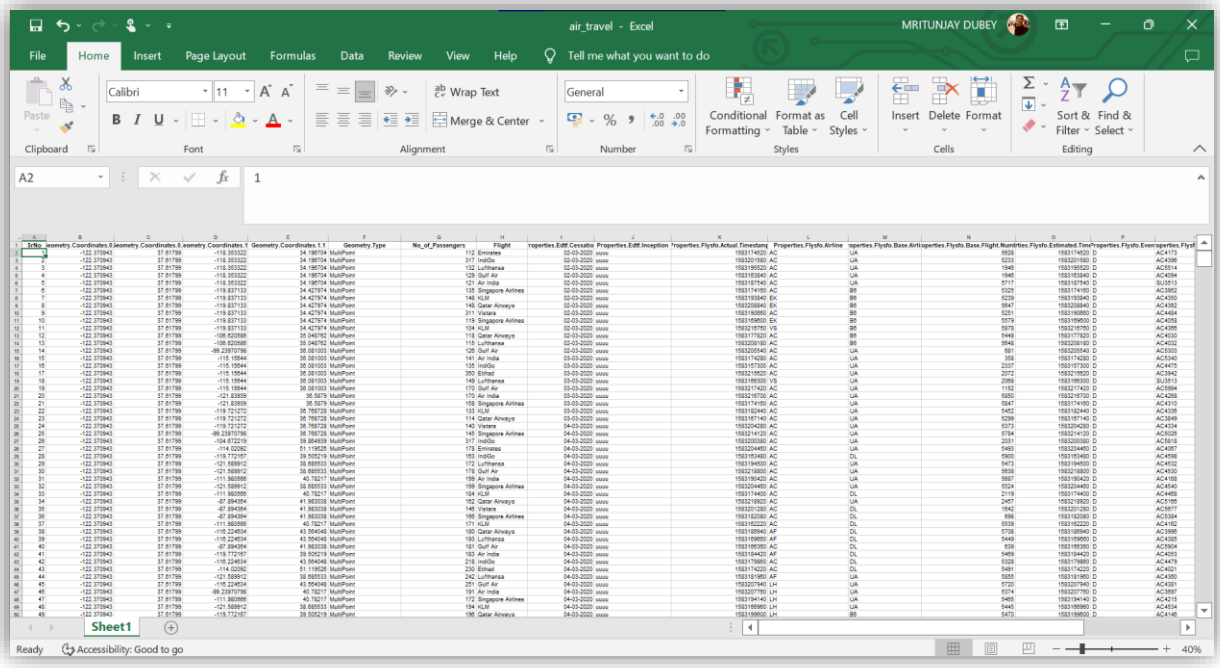
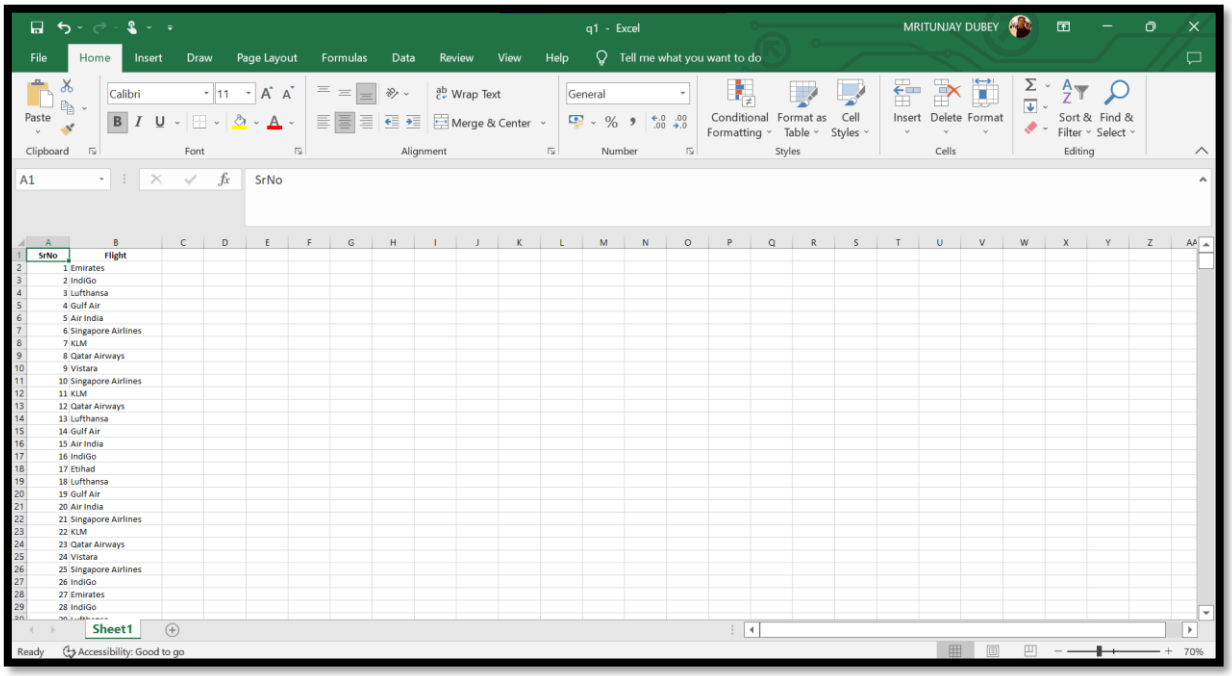

air_travel - Excel
MRITUNJAY DUBEY
File Home Insert Page Layout Formulas Data Review View Help
Tell me what you want to do
Sheet1
Ready
Accessibility: Good to go

Tableau Implementation:

Query for Sheet 1-->

Aim: What are the total number of flights in March?

```
x<-g[,c(1,8)]  #Using Table 1
write_xlsx(x,"C:\\Users\\MRITUNJAY\\Desktop\\Tableau Project\\q1.xlsx")
```

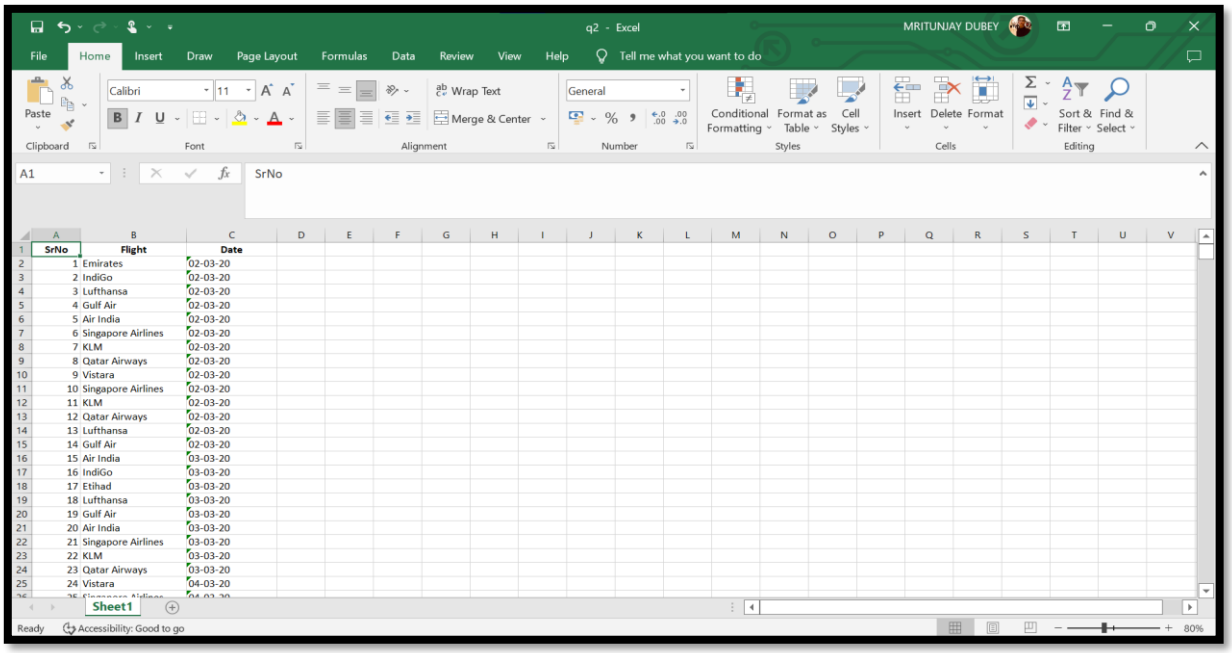


SrNo	Flight
1	Emirates
2	IndiGo
3	Lufthansa
4	Gulf Air
5	Air India
6	Singapore Airlines
7	KLM
8	Qatar Airways
9	Vistara
10	Singapore Airlines
11	KLM
12	Qatar Airways
13	Lufthansa
14	Gulf Air
15	Air India
16	IndiGo
17	Etihad
18	Lufthansa
19	Gulf Air
20	Air India
21	Singapore Airlines
22	KLM
23	Qatar Airways
24	Vistara
25	Singapore Airlines
26	IndiGo
27	Emirates
28	IndiGo

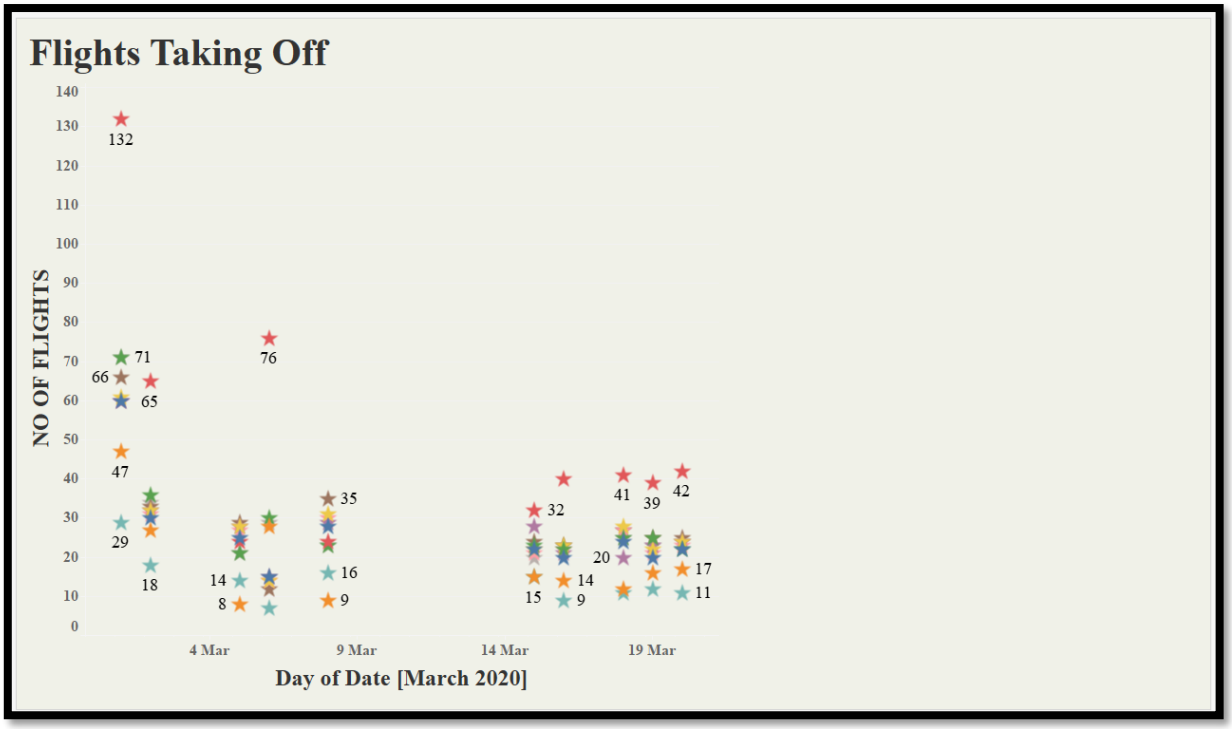
Query for sheet 2-->

Aim: Busiest Day at SFO Airport based on different Airlines.

```
y<-g[,c(1,8)]   #Using Table 1 and Table 3
z<-d[,c(1,2)]
f<-merge(y,z,by="SrNo")
f$Date<-format(as.POSIXct(f$Date), format="%d-%m-%y")
write_xlsx(f,"C:\\Users\\MRITUNJAY\\Desktop\\Tableau Project\\q2.xlsx")
```



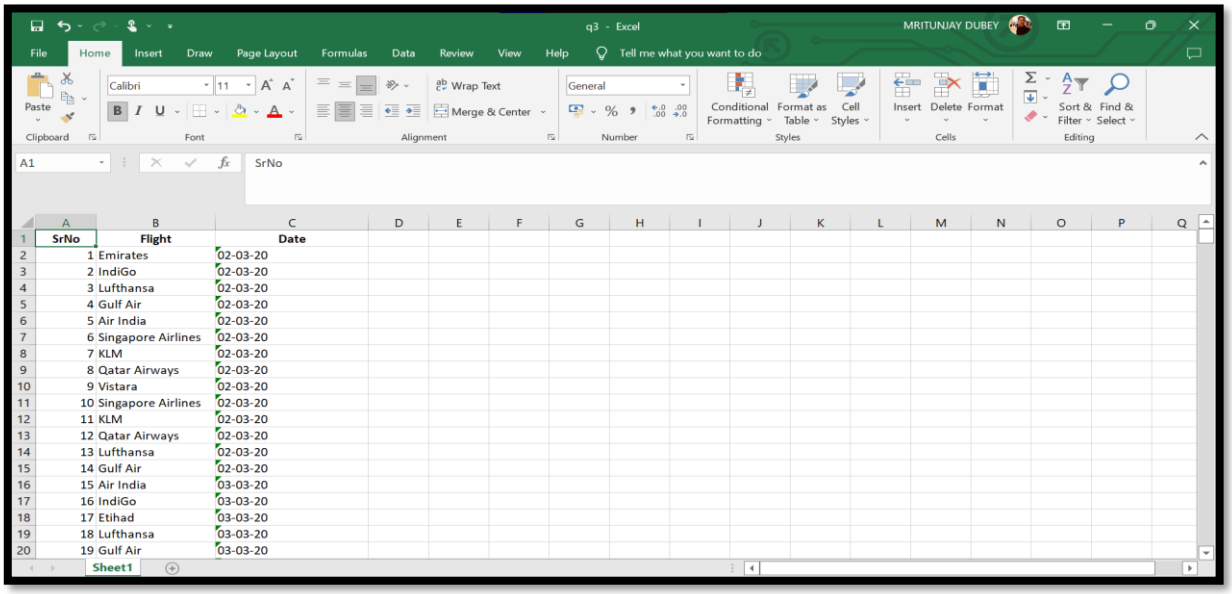
SrNo	Flight	Date
1	Emirates	02-03-20
2	IndiGo	02-03-20
3	Lufthansa	02-03-20
4	Gulf Air	02-03-20
5	Air India	02-03-20
6	Singapore Airlines	02-03-20
7	KLM	02-03-20
8	Qatar Airways	02-03-20
9	Vistara	02-03-20
10	Singapore Airlines	02-03-20
11	KLM	02-03-20
12	Qatar Airways	02-03-20
13	Lufthansa	02-03-20
14	Gulf Air	02-03-20
15	Air India	03-03-20
16	IndiGo	03-03-20
17	Etihad	03-03-20
18	Lufthansa	03-03-20
19	Gulf Air	03-03-20
20	Air India	03-03-20
21	Singapore Airlines	03-03-20
22	KLM	03-03-20
23	Qatar Airways	03-03-20
24	Vistara	04-03-20



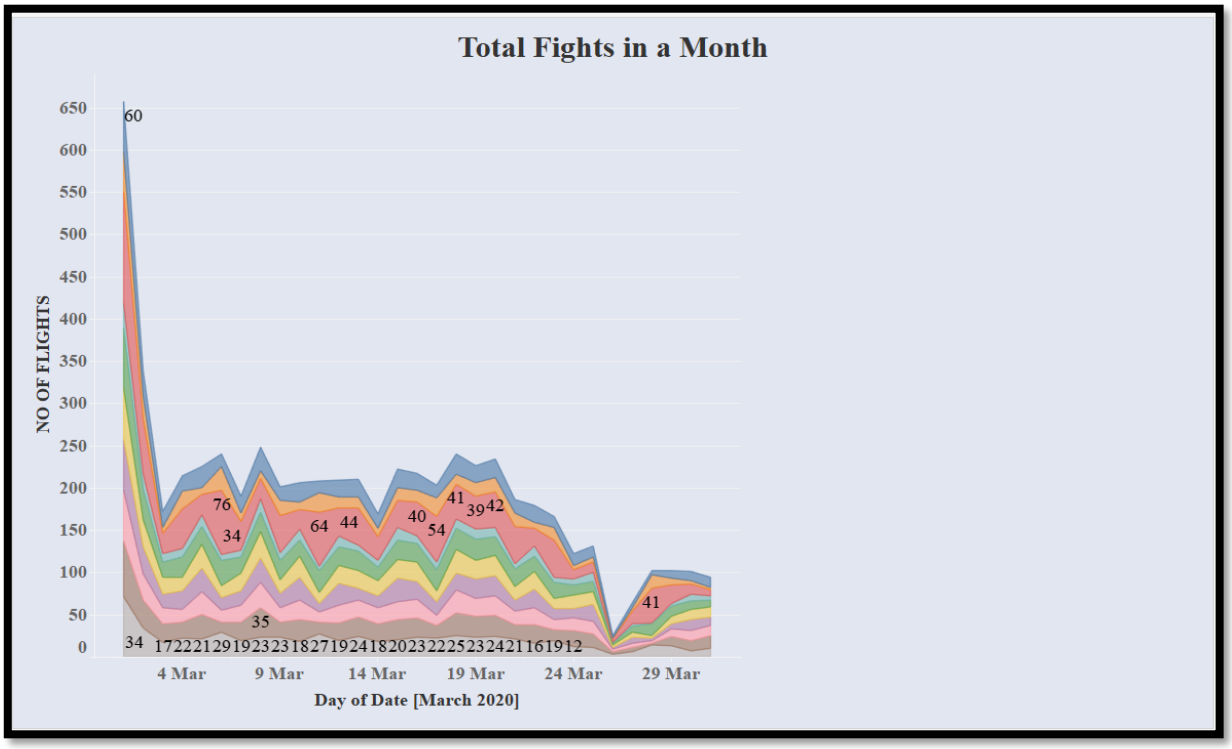
Query for sheet 3-->

Aim: What are the total number of flights by each airline per day in a month?

```
y<-g[,c(1,8)]
z<-d[,c(1,2)]
f<-merge(y,z,by="SrNo")
f$Date<-format(as.POSIXct(f$Date), format="%d-%m-%y")
write_xlsx(f,"C:\\Users\\MRITUNJAY\\Desktop\\Tableau Project\\q3.xlsx")
```



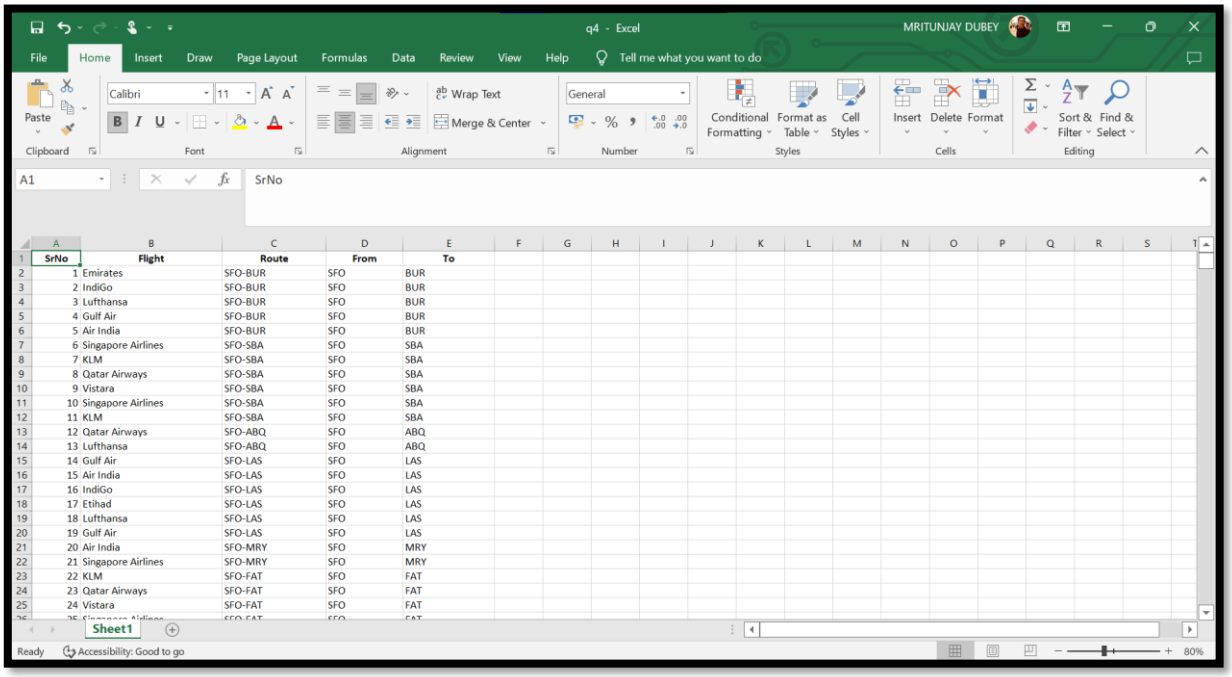
SrNo	Flight	Date
1	Emirates	02-03-20
2	IndiGo	02-03-20
3	Lufthansa	02-03-20
4	Gulf Air	02-03-20
5	Air India	02-03-20
6	Singapore Airlines	02-03-20
7	KLM	02-03-20
8	Qatar Airways	02-03-20
9	Vistara	02-03-20
10	Singapore Airlines	02-03-20
11	KLM	02-03-20
12	Qatar Airways	02-03-20
13	Lufthansa	02-03-20
14	Gulf Air	02-03-20
15	Air India	03-03-20
16	IndiGo	03-03-20
17	Etihad	03-03-20
18	Lufthansa	03-03-20
19	Gulf Air	03-03-20



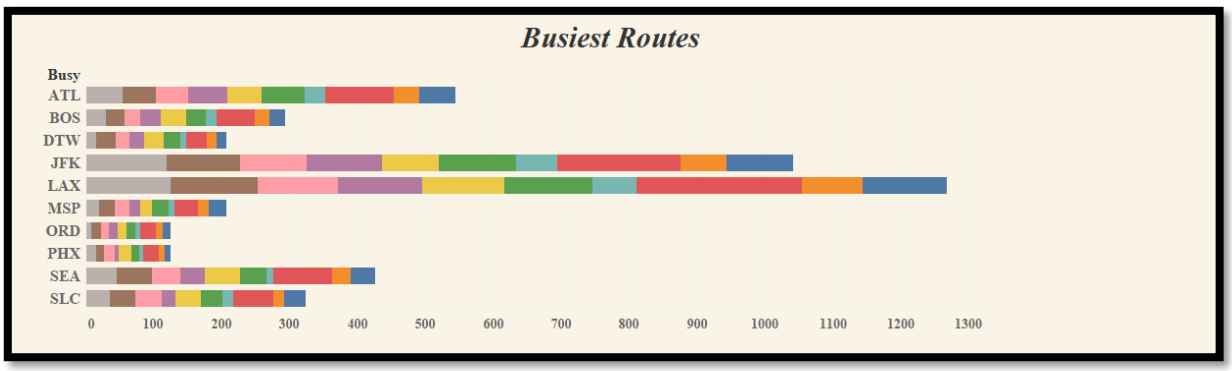
Query for sheet 4-->

Aim: What are the top 10 busiest flight routes?

```
y<-g[,c(1,8)]
busy<-d[,c(1,3,4,5)]
f<-merge(y,busy,by="SrNo")
write_xlsx(f,"C:\\Users\\MRITUNJAY\\Desktop\\Tableau Project\\q4.xlsx")
```



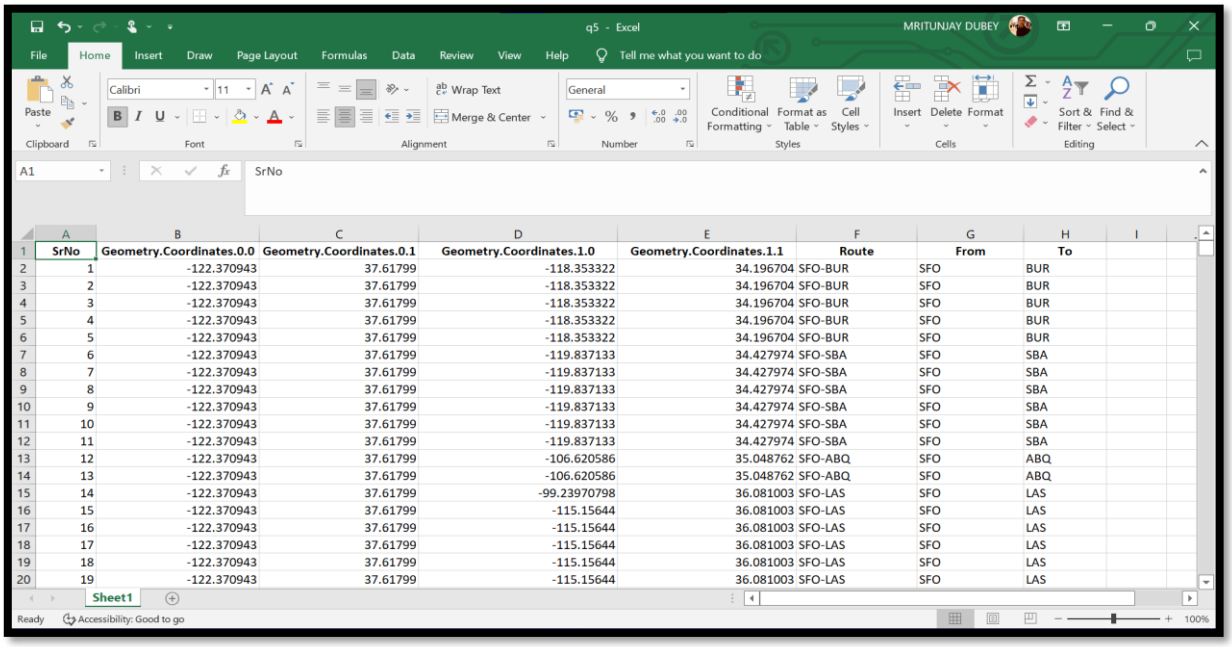
SrNo	Flight	Route	From	To
1	Emirates	SFO-BUR	SFO	BUR
2	IndiGo	SFO-BUR	SFO	BUR
3	Lufthansa	SFO-BUR	SFO	BUR
4	Gulf Air	SFO-BUR	SFO	BUR
5	Air India	SFO-BUR	SFO	BUR
6	Singapore Airlines	SFO-SBA	SFO	SBA
7	KLM	SFO-SBA	SFO	SBA
8	Qatar Airways	SFO-SBA	SFO	SBA
9	Vistara	SFO-SBA	SFO	SBA
10	Singapore Airlines	SFO-SBA	SFO	SBA
11	KLM	SFO-SBA	SFO	SBA
12	Qatar Airways	SFO-ABQ	SFO	ABQ
13	Lufthansa	SFO-ABQ	SFO	ABQ
14	Gulf Air	SFO-LAS	SFO	LAS
15	Air India	SFO-LAS	SFO	LAS
16	IndiGo	SFO-LAS	SFO	LAS
17	Etihad	SFO-LAS	SFO	LAS
18	Lufthansa	SFO-LAS	SFO	LAS
19	Gulf Air	SFO-LAS	SFO	LAS
20	Air India	SFO-MRY	SFO	MRY
21	Singapore Airlines	SFO-MRY	SFO	MRY
22	KLM	SFO-FAT	SFO	FAT
23	Qatar Airways	SFO-FAT	SFO	FAT
24	Vistara	SFO-FAT	SFO	FAT



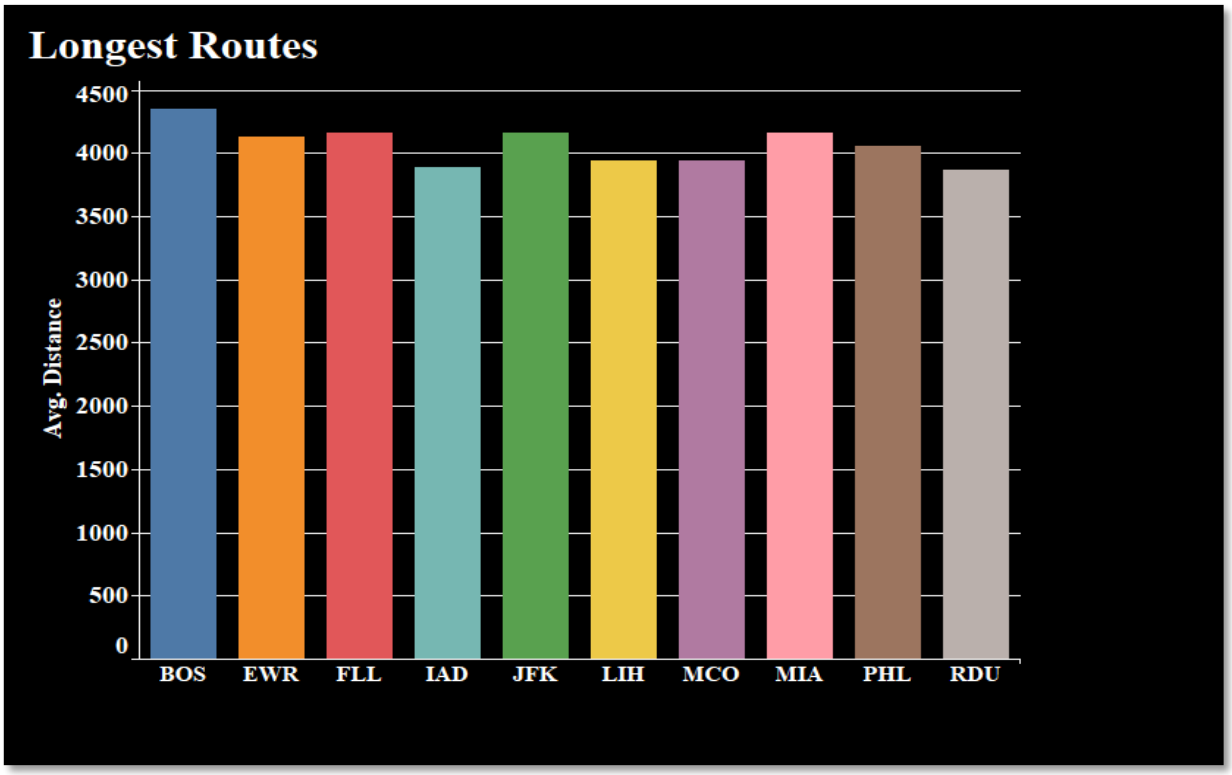
Query for sheet 5-->

Aim: What are the top 10 longest flight routes?

```
distance<-g[,c(1:5)]
busy<-d[,c(1,3,4,5)]
f<-merge(distance,busy,by="SrNo")
write_xlsx(f,"C:\\Users\\MRITUNJAY\\Desktop\\Tableau Project\\q5.xlsx")
```



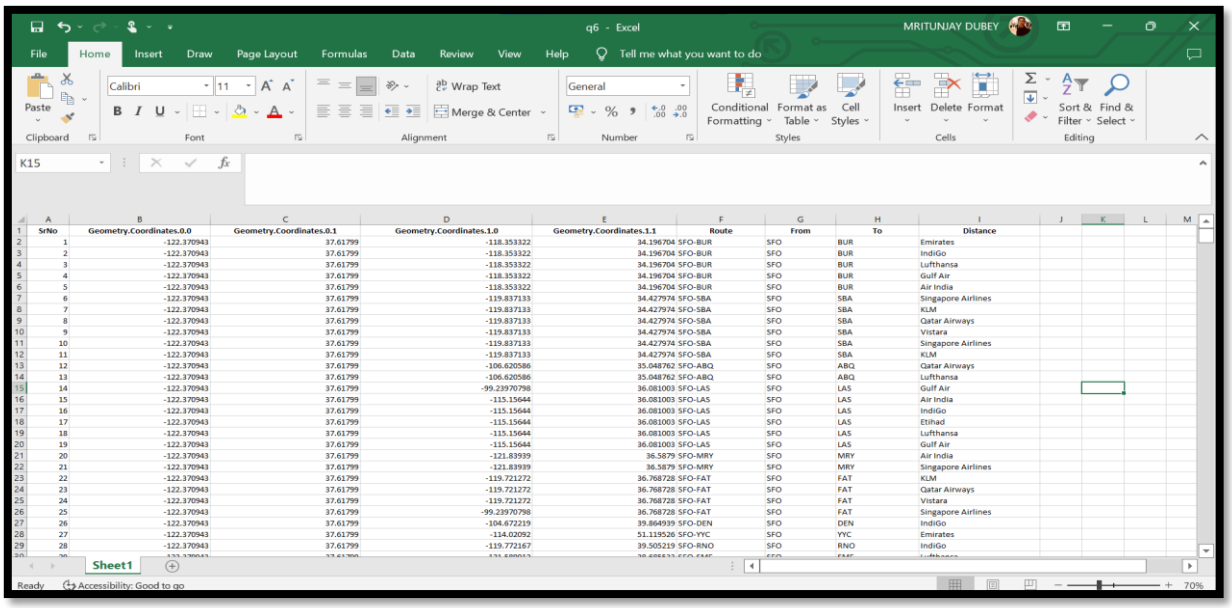
SrNo	Geometry.Coordinates.0.0	Geometry.Coordinates.0.1	Geometry.Coordinates.1.0	Geometry.Coordinates.1.1	Route	From	To
1	-122.370943	37.61799	-118.353322	34.196704	SFO-BUR	SFO	BUR
2	-122.370943	37.61799	-118.353322	34.196704	SFO-BUR	SFO	BUR
3	-122.370943	37.61799	-118.353322	34.196704	SFO-BUR	SFO	BUR
4	-122.370943	37.61799	-118.353322	34.196704	SFO-BUR	SFO	BUR
5	-122.370943	37.61799	-118.353322	34.196704	SFO-BUR	SFO	BUR
6	-122.370943	37.61799	-119.837133	34.427974	SFO-SBA	SFO	SBA
7	-122.370943	37.61799	-119.837133	34.427974	SFO-SBA	SFO	SBA
8	-122.370943	37.61799	-119.837133	34.427974	SFO-SBA	SFO	SBA
9	-122.370943	37.61799	-119.837133	34.427974	SFO-SBA	SFO	SBA
10	-122.370943	37.61799	-119.837133	34.427974	SFO-SBA	SFO	SBA
11	-122.370943	37.61799	-119.837133	34.427974	SFO-SBA	SFO	SBA
12	-122.370943	37.61799	-106.620586	35.048762	SFO-ABQ	SFO	ABQ
13	-122.370943	37.61799	-106.620586	35.048762	SFO-ABQ	SFO	ABQ
14	-122.370943	37.61799	-99.23970798	36.081003	SFO-LAS	SFO	LAS
15	-122.370943	37.61799	-115.15644	36.081003	SFO-LAS	SFO	LAS
16	-122.370943	37.61799	-115.15644	36.081003	SFO-LAS	SFO	LAS
17	-122.370943	37.61799	-115.15644	36.081003	SFO-LAS	SFO	LAS
18	-122.370943	37.61799	-115.15644	36.081003	SFO-LAS	SFO	LAS
19	-122.370943	37.61799	-115.15644	36.081003	SFO-LAS	SFO	LAS

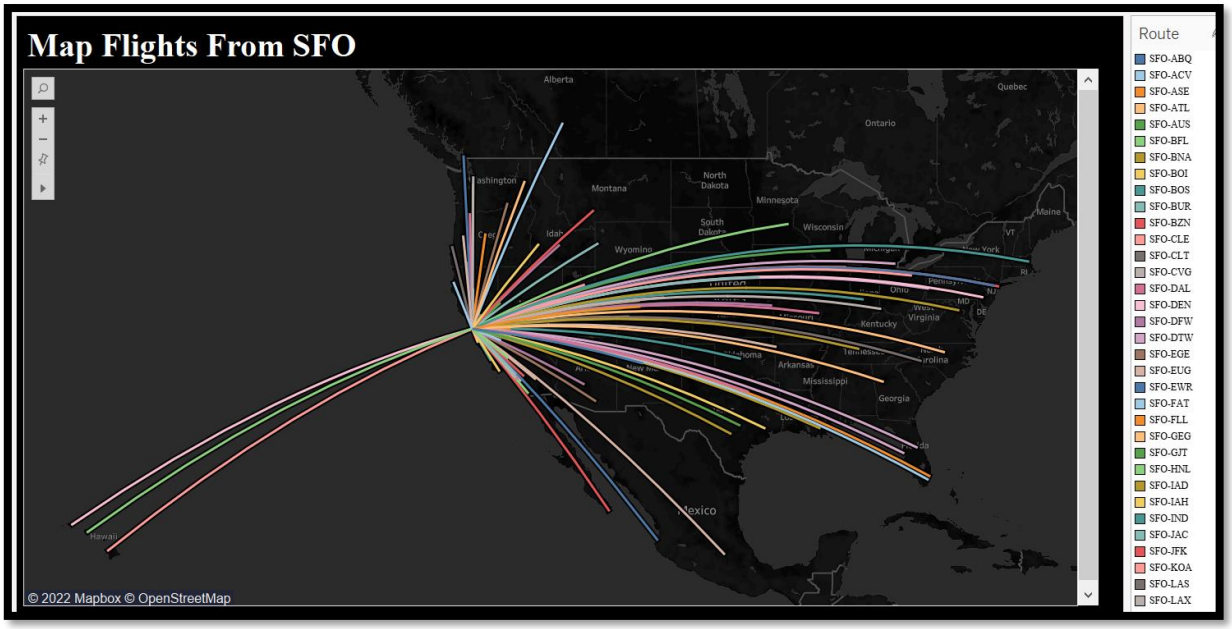


Query for sheet 6-->

Aim: Represent the flights flying outside SFO towards different parts of the world.

```
distance<-g[,c(1:5)]
busy<-d[,c(1,3,4,5)]
x<-g[,8]
names(x)=c("Flight")
b<-cbind(busy,x)
f<-merge(distance,b,by="SrNo")
write_xlsx(f,"C:\\Users\\MRITUNJAY\\Desktop\\Tableau Project\\q6.xlsx")
```

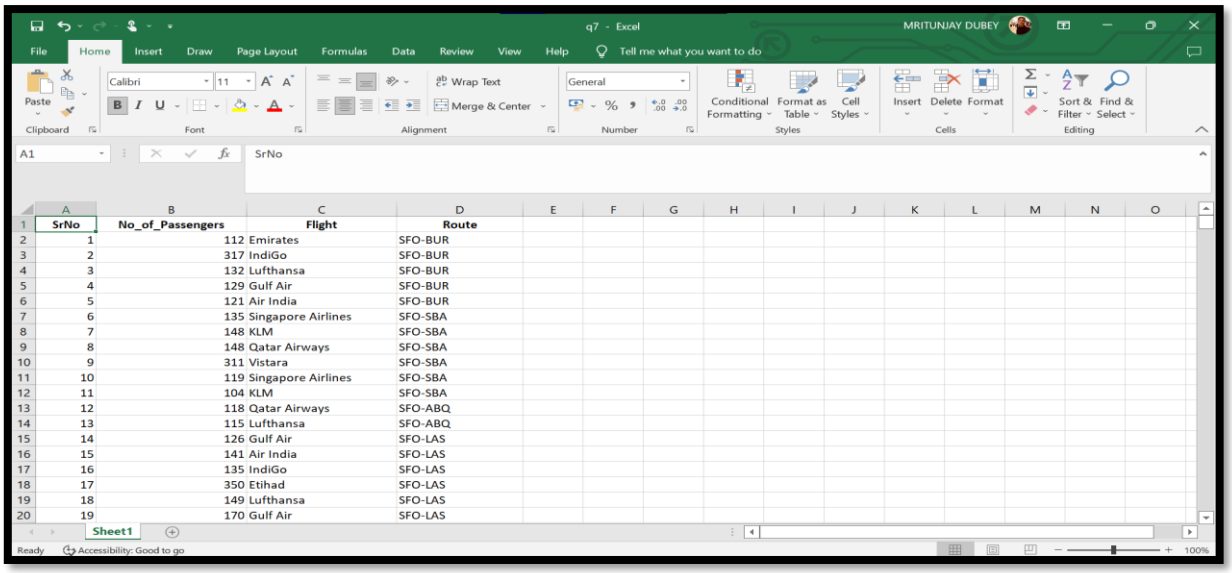




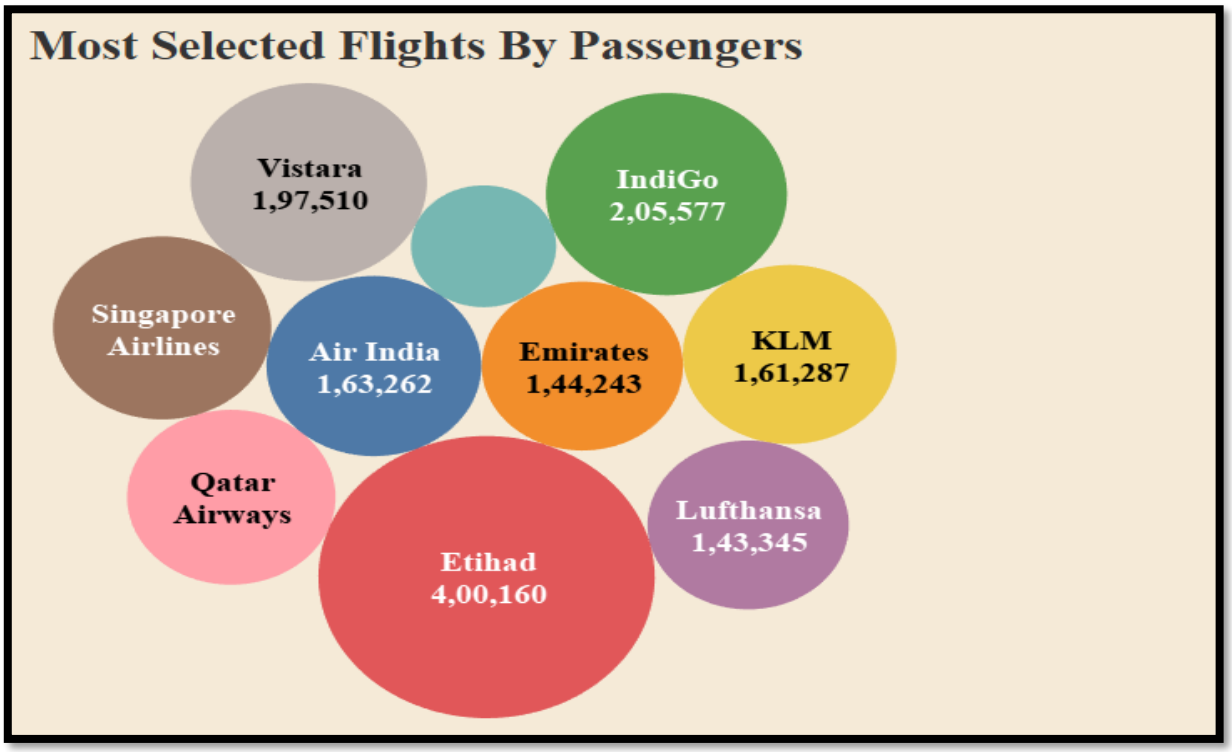
Query for sheet 7-->

Aim: Which airlines are mostly selected by passengers from SFO Airport?

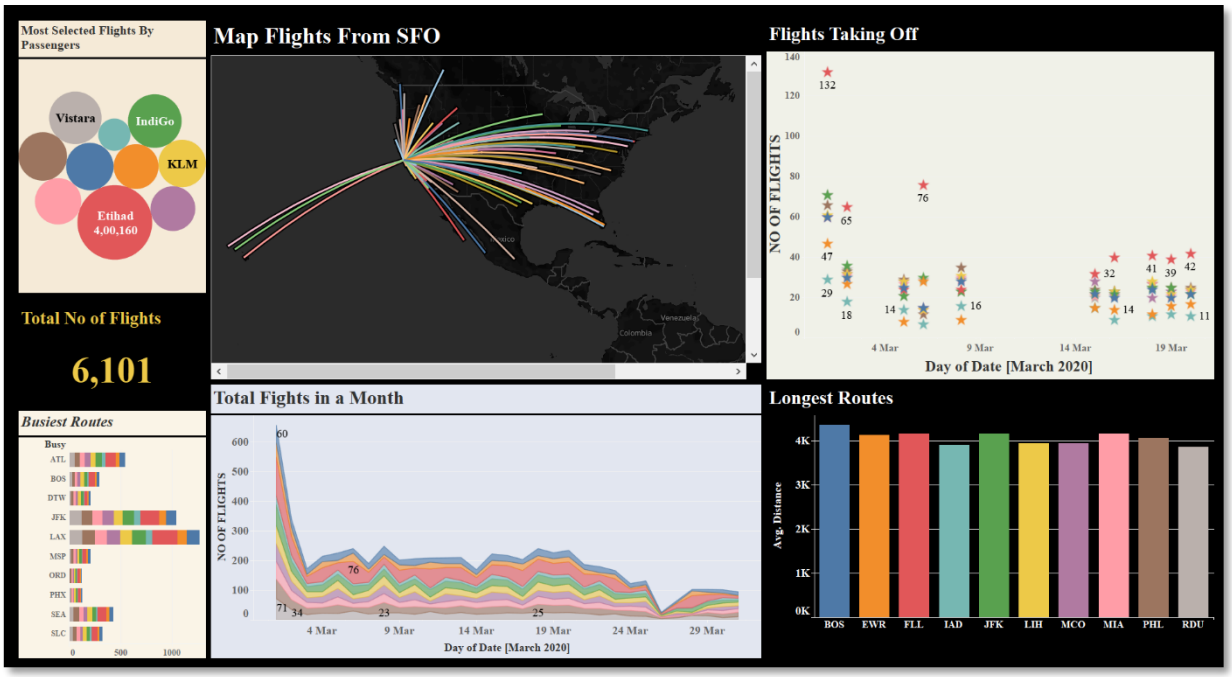
```
passengers<-g[,c(1,7,8)]
route<-d[,c(1,3)]
f<-merge(passengers,route,by="SrNo")
write_xlsx(f,"C:\\Users\\MRITUNJAY\\Desktop\\Tableau Project\\q7.xlsx")
```



SrNo	No_of_Passengers	Flight	Route
1	112	Emirates	SFO-BUR
2	317	IndiGo	SFO-BUR
3	132	Lufthansa	SFO-BUR
4	129	Gulf Air	SFO-BUR
5	121	Air India	SFO-BUR
6	135	Singapore Airlines	SFO-SBA
7	148	KLM	SFO-SBA
8	148	Qatar Airways	SFO-SBA
9	311	Vistara	SFO-SBA
10	119	Singapore Airlines	SFO-SBA
11	104	KLM	SFO-SBA
12	118	Qatar Airways	SFO-ABQ
13	115	Lufthansa	SFO-ABQ
14	126	Gulf Air	SFO-LAS
15	141	Air India	SFO-LAS
16	135	IndiGo	SFO-LAS
17	350	Etihad	SFO-LAS
18	149	Lufthansa	SFO-LAS
19	170	Gulf Air	SFO-LAS

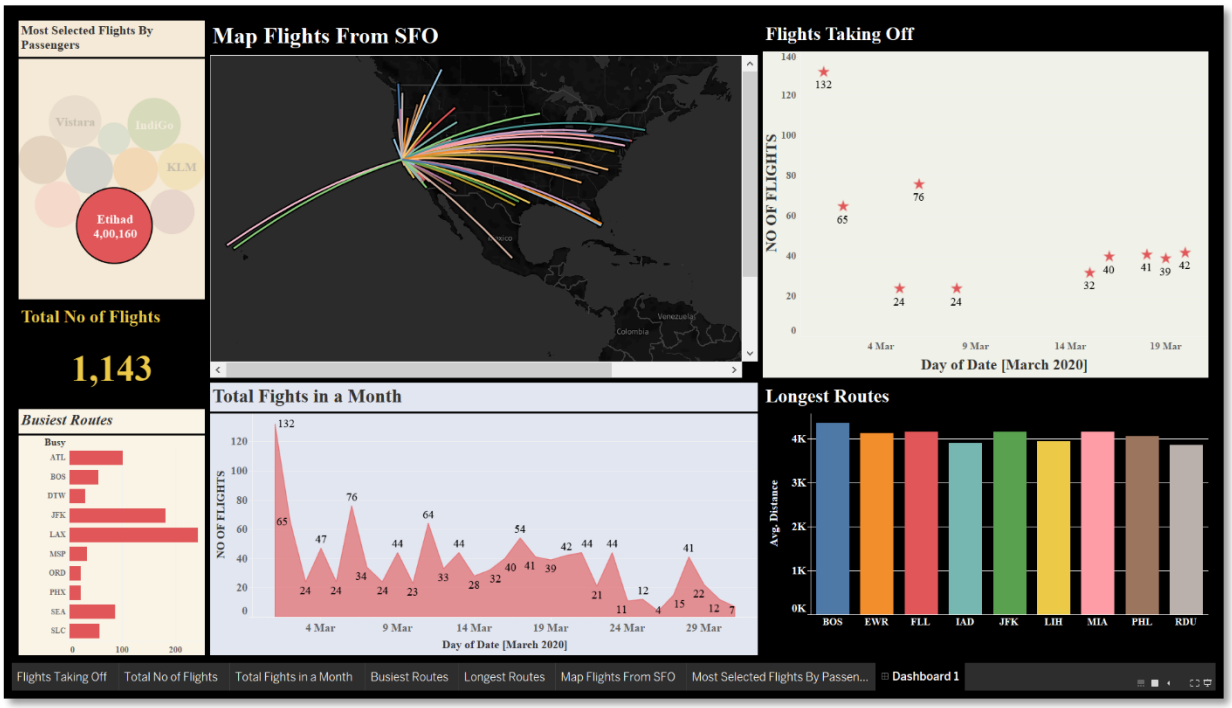


Final Dashboard:

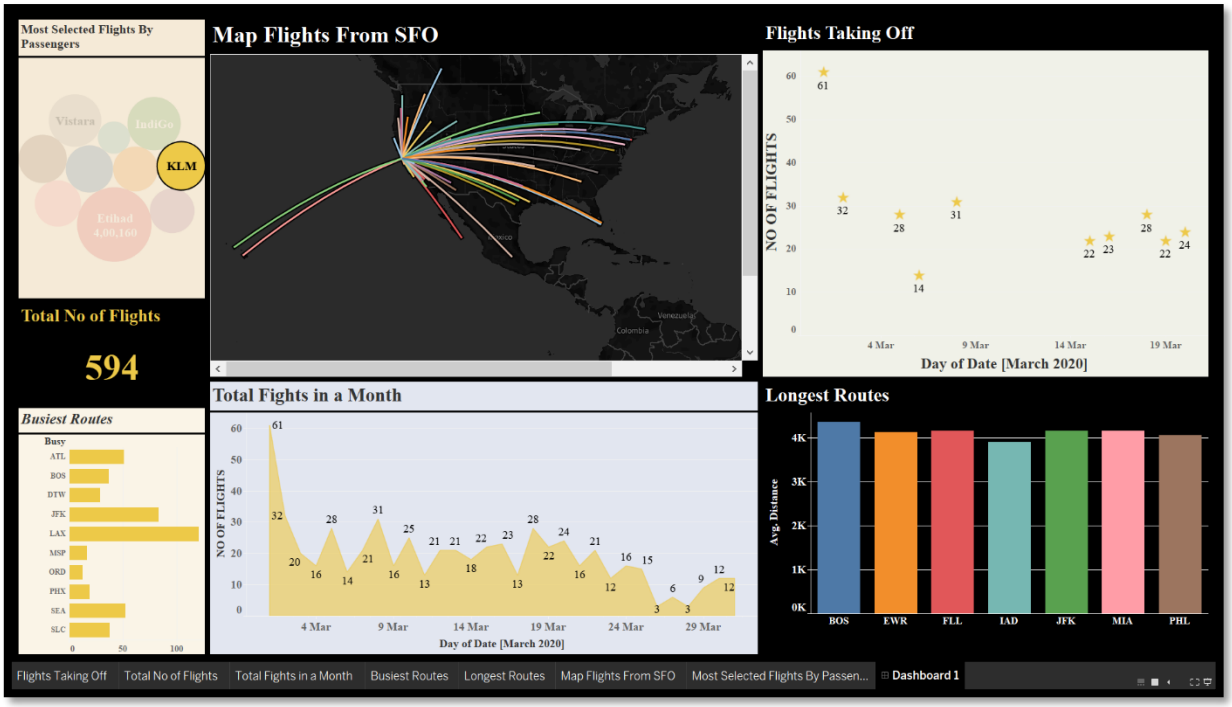


Visualization by selecting particular Airline:

For Etihad



For KLM


Most Selected Flights By Passengers
Vistara
IndiGo
KLM
Etihad
4,00,160
Total No of Flights
594
Busiest Routes
Busy
ATL
BOS
DTW
JFK
LAX
MSP
ORD
PHX
SEA
SLC
0
50
100
Map Flights From SFO
Flights Taking Off
NO OF FLIGHTS
Day of Date [March 2020]
4 Mar
9 Mar
14 Mar
19 Mar
Total Fights in a Month
29 Mar
Longest Routes
Avg. Distance
4K
3K
2K
1K
0K
BOS
EWR
FLL
IAD
JFK
MIA
PHL
Flights Taking Off
Total No of Flights
Total Fights in a Month
Busiest Routes
Longest Routes
Map Flights From SFO
Most Selected Flights By Passen...
Dashboard 1