

DETAILS PROJECT REPORT
ON
AIRPORT DATASET ANALYSIS

CONTENT

- ☐ Objective of this Dataset
- ☐ Benefits
- ☐ Data Sharing Agreement
- ☐ Data Validation
- ☐ Exploratory Data Analysis
- ☐ Conclusions



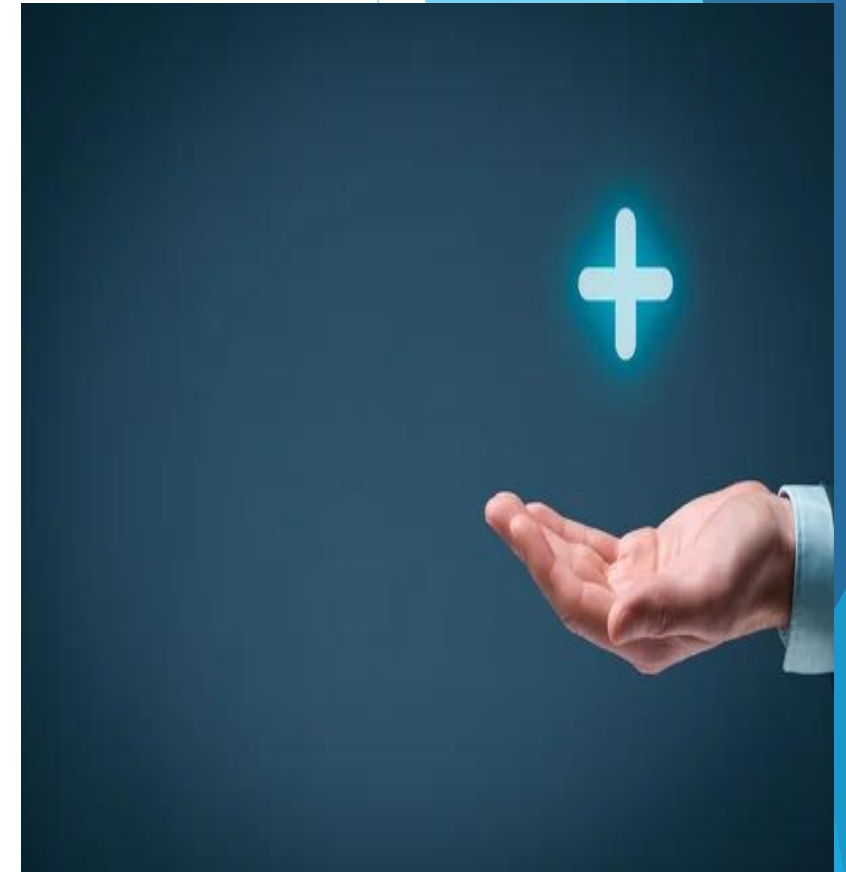
OBJECTIVE OF THIS DATASET

The airport dataset is in the form of excel file from which we can't analyse it properly. By the way we can't find some important conclusions like where the various flights are going ,what is the most lengthy and busy route etc. from this file directly

So Here We Are Going To Develop Airport dataset analysis Dashboard for identifying where the various flights are going and what is the most busy and most lengthy routes from the airport.

BENEFITS

- ❑ Gives a better insights of Flights
- ❑ Represents the complicated data in a simple way
- ❑ Finds answers of any puzzled questions related to this dataset.
- ❑ Here Gets the answers of where various flights are going and what is the most busy and most lengthy routes from the airport.



DATA SHARING AGREEMENT

DATA SUMMARY :

- ❑ File Name – “air travel dataset.xlsx - Air travel data.csv”
- ❑ Here We Have A Total Of 88532 Rows And 17 Columns.
- ❑ Some Features Are Numerical Features, And Some are Categorical .
- ❑ Geometry Type , Airline , Base Airline,Event,Flight Number, Gate and Route are categorical features and rest are Numerical Features

	0	1	2	3	4
Geometry Coordinates 0 0	-122.370943	-122.370943	-122.370943	-122.370943	-122.370943
Geometry Coordinates 0 1	37.61799	37.61799	37.61799	37.61799	37.61799
Geometry Coordinates 1 0	-118.353322	-118.353322	-118.353322	-118.353322	-118.353322
Geometry Coordinates 1 1	34.196704	34.196704	34.196704	34.196704	34.196704
Geometry Type	MultiPoint	MultiPoint	MultiPoint	MultiPoint	MultiPoint
Properties Edtf Cessation	3/2/2020	3/2/2020	3/2/2020	3/2/2020	3/2/2020
Properties Edtf Inception	uuuu	uuuu	uuuu	uuuu	uuuu
Properties Flysfo Actual Timestamp	1583174520	1583201580	1583195520	1583163840	1583187540
Properties Flysfo Airline	AC	AC	AC	AC	AC
Properties Flysfo Base Airline	UA	UA	UA	UA	UA
Properties Flysfo Base Flight Number	5628.0	5233.0	5318.0	5920.0	5717.0
Date	2-Mar-20	2-Mar-20	2-Mar-20	2-Mar-20	2-Mar-20
Properties Flysfo Estimated Timestamp	1583174520	1583201580	1583195520	1583163840	1583187540
Properties Flysfo Event	D	D	D	D	D
Properties Flysfo Flight Number	AC4173	AC4396	AC5514	AC4094	AC4388
Properties Flysfo Gate	F5	F15K	F15K	F15L	F15L
Route	SFO-BUR	SFO-BUR	SFO-BUR	SFO-BUR	SFO-BUR

WHAT ARE THE FEATURES :

- ❑ **Geometry Coordinates 0 0,Geometry Coordinates 0 1,Geometry Coordinates 1 0,Geometry Coordinates 1 1 -** Coordinate Of the Arrival and and departure of the flight
- ❑ **Geometry Type -** Multipoint
- ❑ **Properties Edtf Cessation -** Date of Arrival
- ❑ **Properties Edtf Inception -** Date of Departure
- ❑ **Properties Flysfo Actual Timestamp -** Actual Departure time
- ❑ **Properties Flysfo Airline -** Type of Airline
- ❑ **Properties Flysfo Base Airline -** Type of Base Airline

- ❑ **Properties Flysfo Base Flight Number** - Flight Number (Base)
- ❑ **Date** - Day on which the Flight departs
- ❑ **Properties Flysfo Estimated Timestamp** - Estimated Departure time
- ❑ **Properties Flysfo Event**- Arrival or Departure
- ❑ **Properties Flysfo Flight Number**- Flight Number
- ❑ **Properties Flysfo Gate**- Gate Through which flight starts
- ❑ **Route**- Route of the Flight i.e Path from Arrival to Departure

DATA INSIGHTS :

- ❑ There Are 33668 Null values for Airline and Flight number , 12393 Null values for Gate.
- ❑ There is no duplicated values.
- ❑ Here we have to develop a Dashboard Airport dataset analysis Dashboard for identifying where the various flights are going and what is the most busy and most lengthy routes from the airport.
- ❑ Now Let's Get Into Our Data And Do Exploratory Data Analysis, Feature Engineering, Feature Selection .

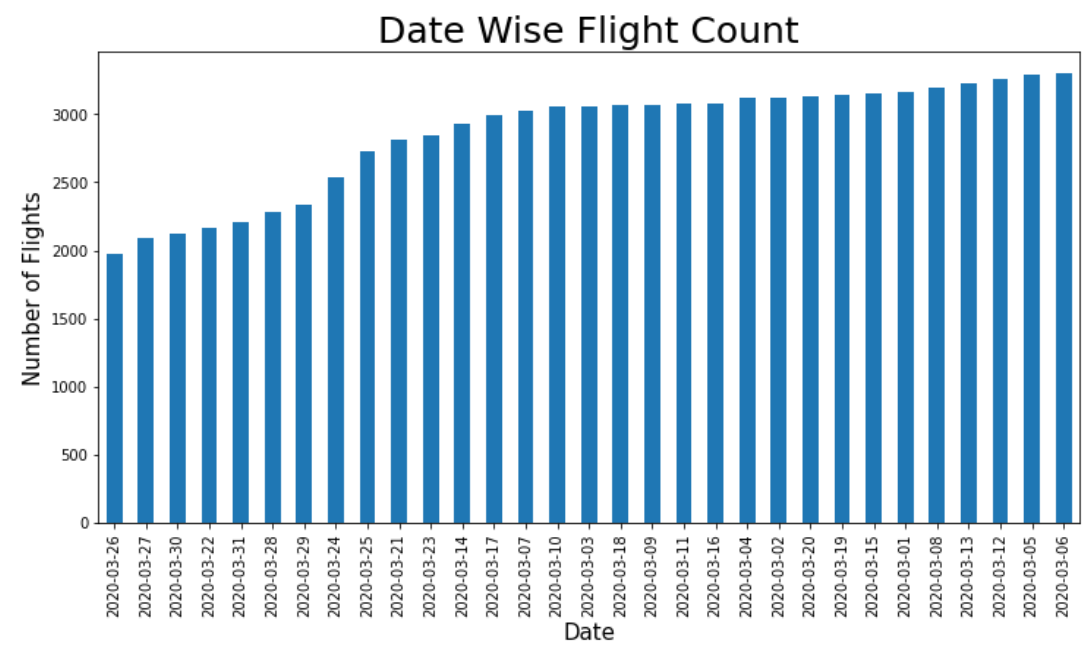
EXPLORATORY DATA ANALYSIS



DATE WISE FLIGHT :

Taking Dataframe of “Date” in X-axis and its Value Counts on Y-axis We Plot the Bar Graph and Get Some Information about the Flights ;

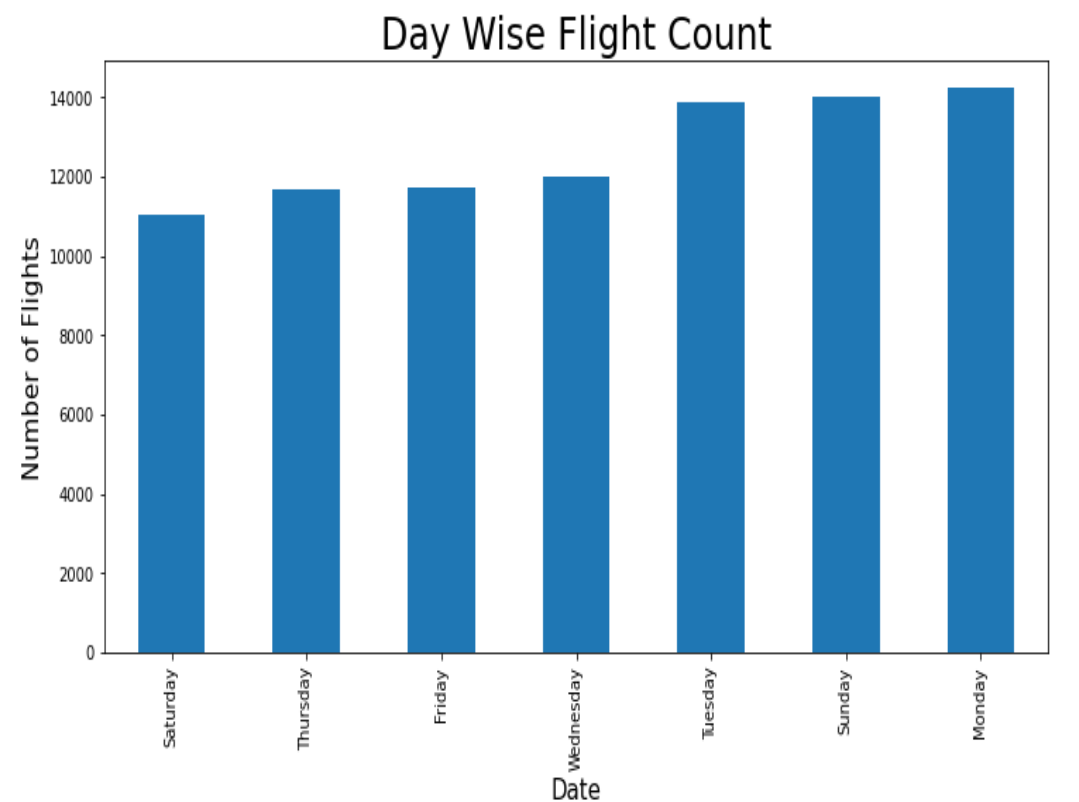
- ❑ This data contains 1 month Flight information.
- ❑ Here We Observed that Maximum no. of flights is in March 6,2020.
- ❑ Minimum No. of Flights in March 26,2020



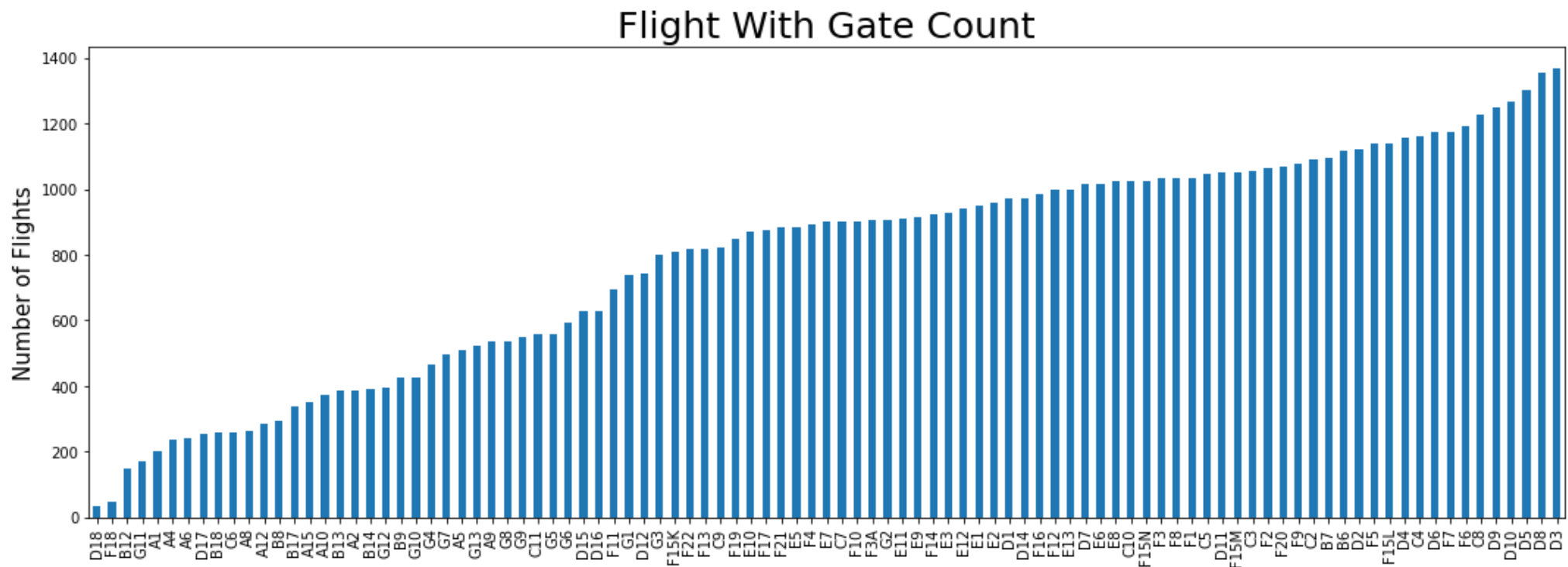
DAY WISE FLIGHT :

Creating Dataframe Of “Day” using the Date ,Taking Day on X-axis and its Counts on Y-axis we Plot the Bar chart.Here We found ;

- ❑ Maximum Flights on Monday
- ❑ Minimum Flights on Saturday



GATE WISE FLIGHT :



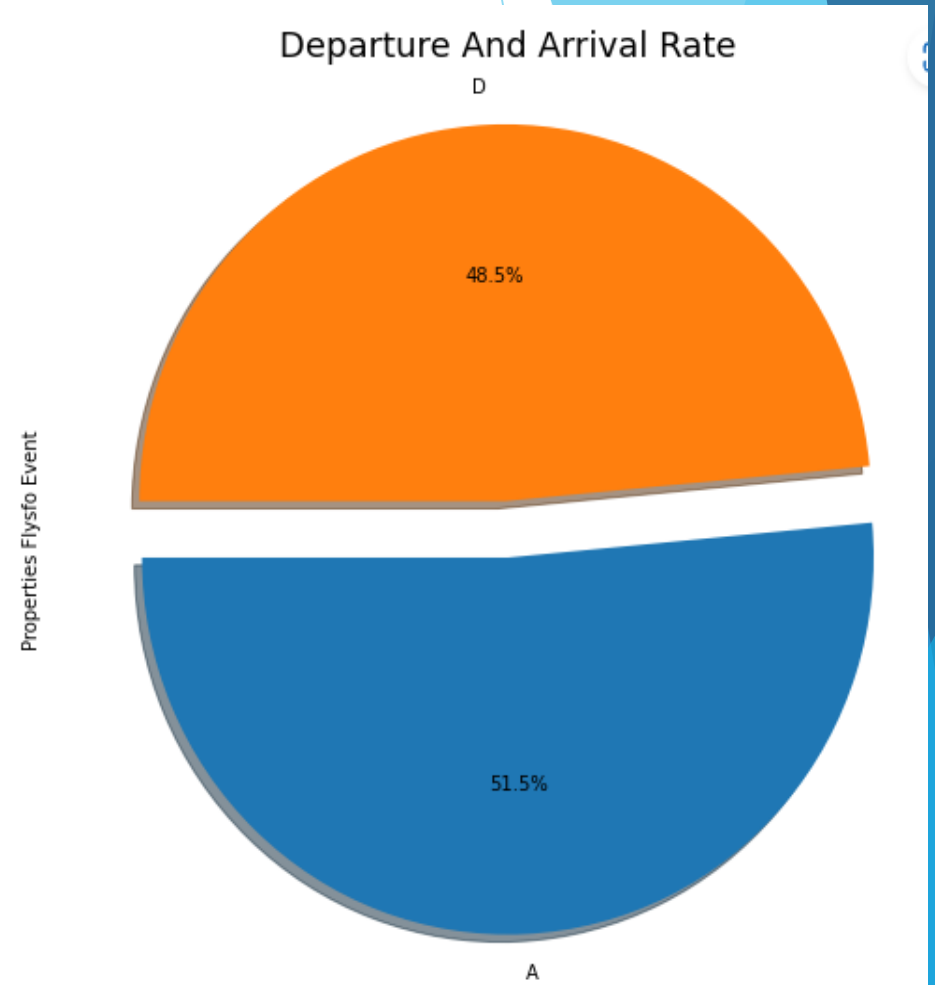
There are lot of Gates through which Flights start their Journey. So By Bar Plot of Gate and No. of Flights we Found ;

- ❑ Maximum Flights are through Gate “D3”.
- ❑ Minimum Flights are through Gate “D18”.

ARRIVAL OR DEPARTURE :

Since There are two types of event arrival and departure Pie Chart is very much Suitable for this type of Data .So Plotting Pie Chart of these two events we got ;

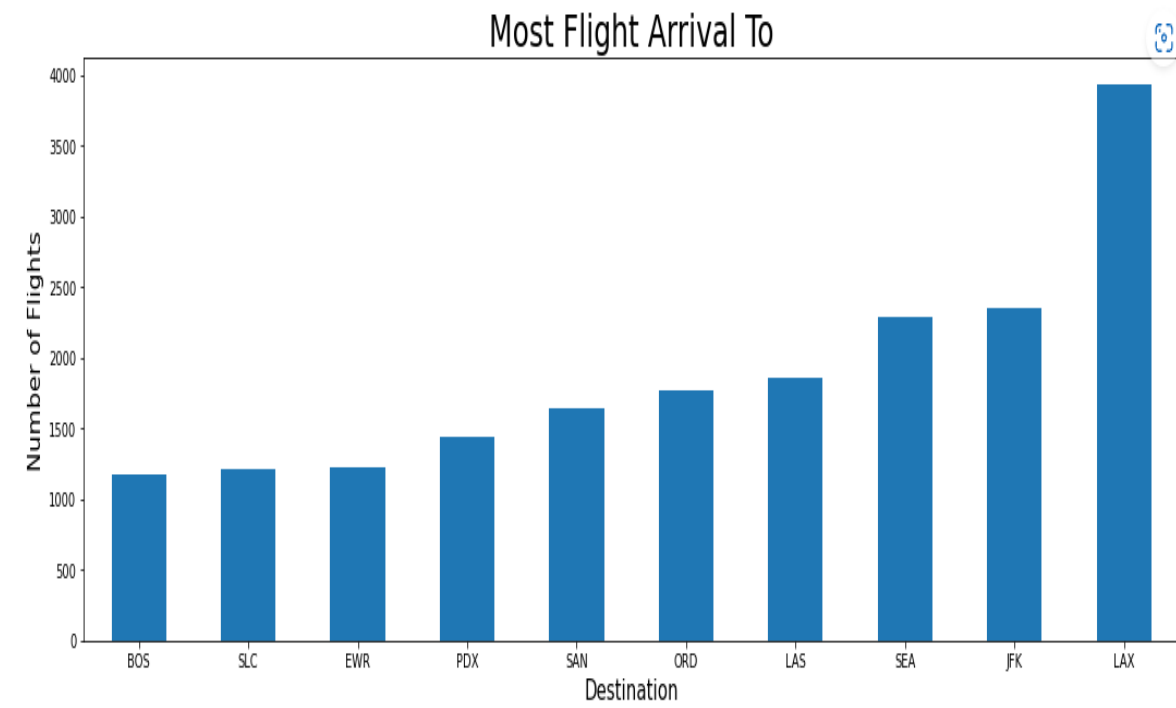
- ❑ There is Not Huge difference between Arrival and Departure Rate .
- ❑ There is a slight difference



DESTINATION WISE FLIGHT ARRIVAL:

To find the Most Arrival Flight on Destination wise we Took Destination on X-axis and Number of Flights on the basis of Arrival we made a Bar Plot and Observed that;

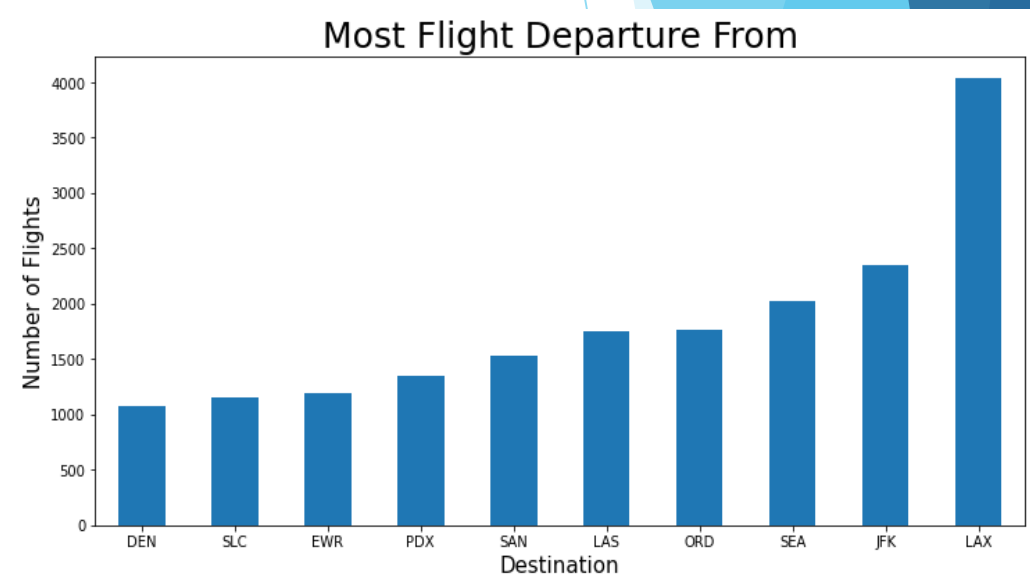
- ❑ Most Flight Arrives to “LAX”, followed by “JFK”
- ❑ Least No. Of Flights Arrives to “BOS”



DESTINATION WISE FLIGHT DEPARTURE:

Similarly ,To find the Most Departure Flight on Destination wise we Took Destination and Number of Flights on the basis of Departure we made a Bar Plot and Observed that ;

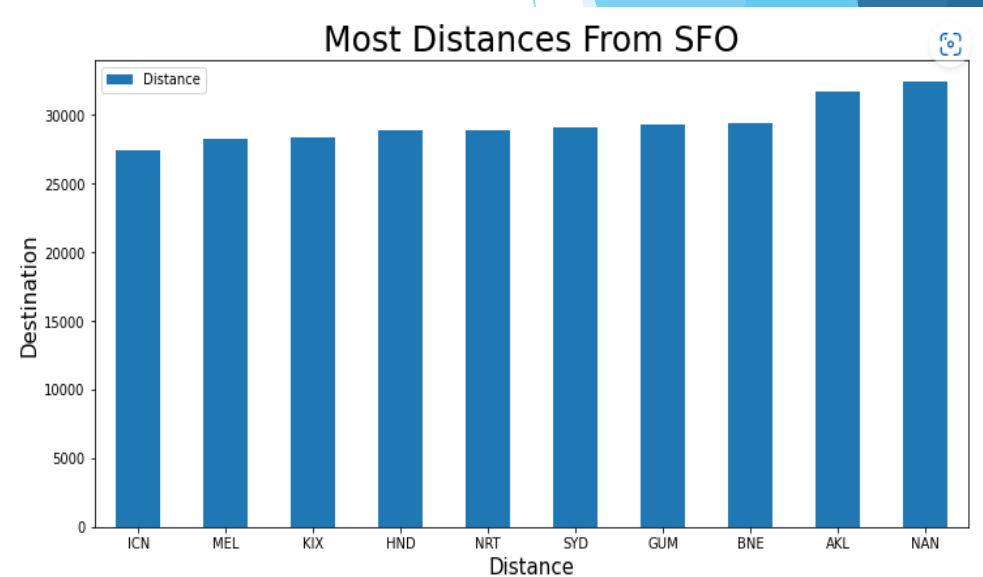
- ❑ Most Flights Departs from “LAX”,followed by “JFK”
- ❑ Least Flights Departs from “DEN”



DISTANCE FROM SFO :

Here we tried to calculate the Distance from two place using the Dataframe of Geometry Coordinates and made a Plot of Top 10 most distances and Get ;

- ❑ The Distance between SFO and NAN is maximum.
- ❑ The Distance between SFO and ICN is Minimum.

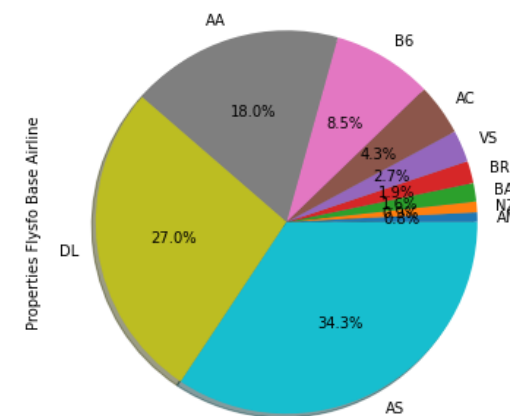


TYPES OF AIRLINE AND BASE AIRLINE :

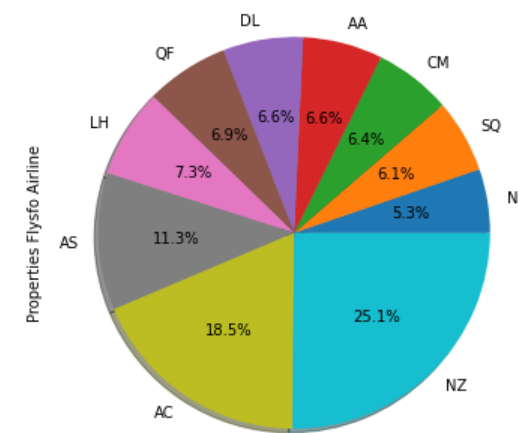
On the basis of Airline and Base Airline we made Pie Plot taking Value counts using Seaborn Library and got the following Results ;

- ❑ Maximum Flights are from Airline “NZ” and from Base Airline “AS”.
- ❑ Least No. of Flights are from Airline “NH” and from Base Airline “AM”

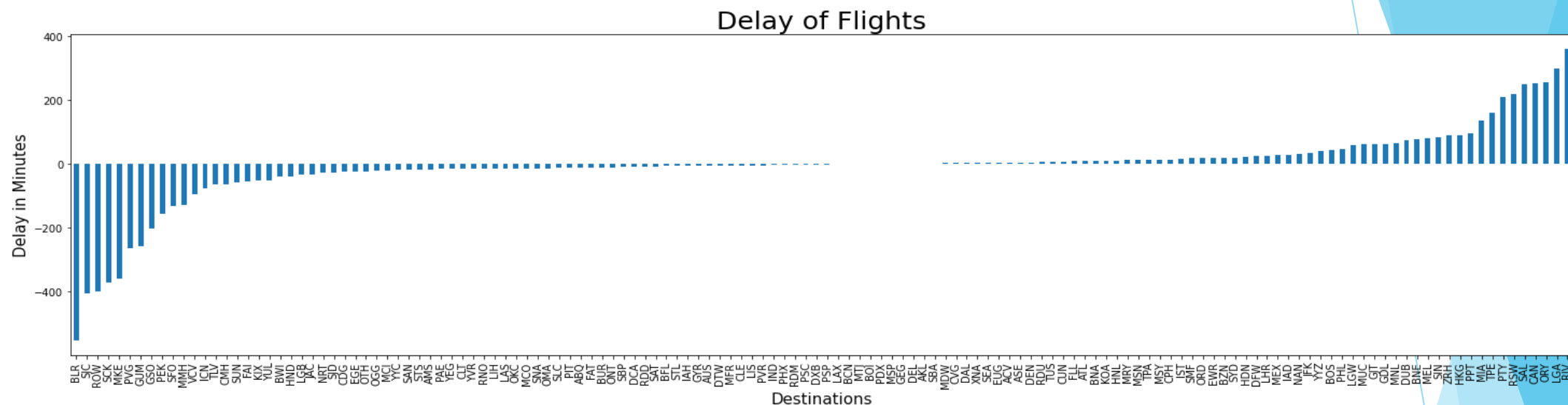
Different Base Airline Count



Different Airline Count



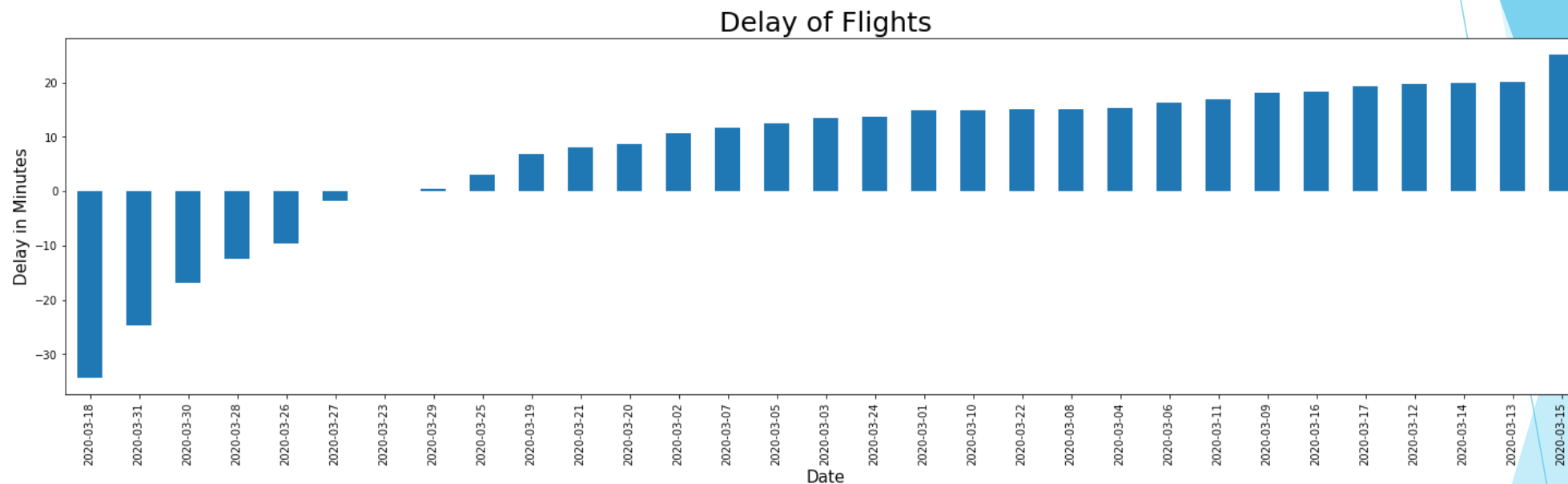
TIME DELAY OF FLIGHTS :



We calculated the Dataframe delay by subtracting estimated and actual arrival time. Then Plotted the graph using Group by method taking Dataframe 'Delay' and 'City'. Here we Observe that '

- ❑ Most Flights are on time.
- ❑ Flight to 'BLR 'is so Punctual and came before ontime.
- ❑ The most Delay Flight is to 'RIV' which is more than 6 hours late.

DAY WISE DELAY(AVG) :



Similarly we represented Day wise average Delay of Flights . Here we Observe that ;

- ❑ Delay is maximum on March 15.
- ❑ The Flights are before ontime on March 18,26,27,28,30,31.

From the EDA we Get the following Conclusions:

- ❑ Maximum Flights on March 6, 2020 .
- ❑ Most Flight Arrival and Departure at “LAX”.
- ❑ Maximum Distance between ‘SFO’ and “NAN”.
- ❑ Most Flights from Airline “NZ”.
- ❑ The most Delay Flight is to “RIV” .

Questions and Answers

1) Where the various Flights are Going ?

Ans. Various Flights are going to 'DEN' , 'SLC' , 'JFX' , 'LAX' and many more

2) What is the most busy Airport?

Ans. "LAX" is the most busy Airport.

3) What is the most lengthy Route?

Ans. "SFO-NAN" is the most lengthy Route.

THANK YOU

