**Complete Flight Passenger Analysis using Python**

**Aim**

There is a lot of flights going to and fro now-a-days. As a data analyst you have been hired to assist companies, increasing their sales and business.

You have been shared with customer satisfactions survey on various parameters. Identify the Key important parameters that can bring in more customers and have a positive impact on their journey.

**Problem Statement**

In the effort to provide a seamless and personalized passenger journey to air travellers, the travel industry must continuously adapt to market changes and new technology.

In this context, company has carried out the Global Passenger Survey since 2012, which has been designed to provide objective and in-depth insights into the preferences and behaviours of air travellers all around the world.​​

Passengers are considering a variety of factors when planning their travels, such as:

* Ticket prices
* Airline awareness and reputation
* Brand loyalty
* Cleanliness
* Flying history (i.e. frequent fliers)
* Travel time (i.e. nonstop route availability)​​​​​​​​​

An airline passenger satisfaction survey refers to a type of market research that gathers feedback and information from passengers who have completed a trip with your airline or at your airport.

Typically, a survey is sent a short while after the trip in an effort to learn more about the passenger’s flying experience.

As with any market research that is conducted, defining the goals and objectives of the study is the number one priority.

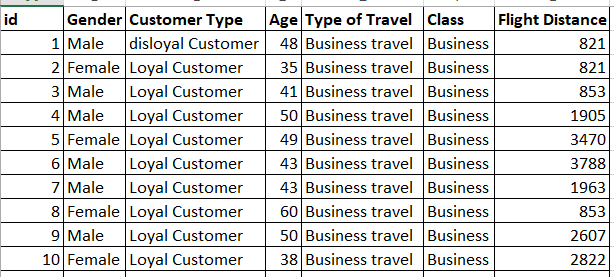
**Learning Outcome**

* Pandas Joins and Merge
* Data Manipulation
* Data cleaning
* Creating charts and bars
* Perform wrangling operations to draw more insights
* Evaluation of bi-columns on the basis of next attribute
* Creating new columns to drill down on analysis

**Data Information**

There are 2 csv files that are shared here.

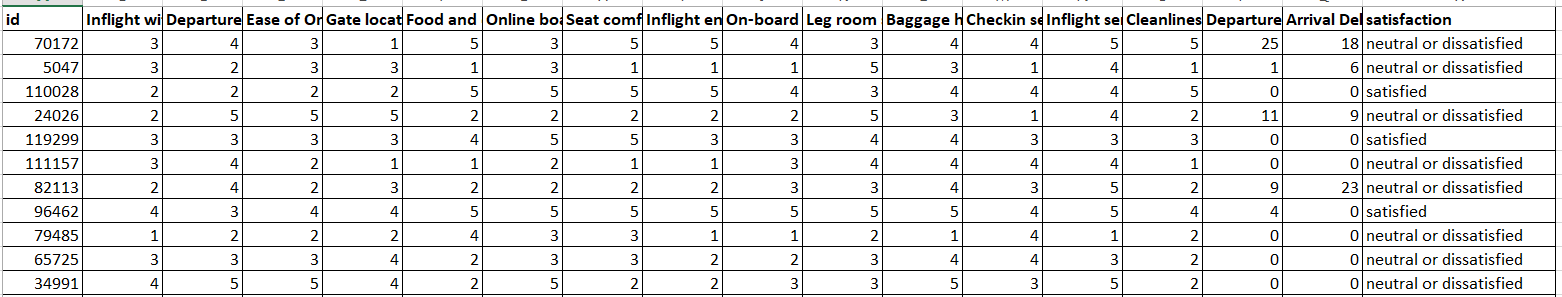
1. passenger data



**Attributes:**

| **Name** | **Description** |
| --- | --- |
| Gender | Gender of the passengers (Female, Male) |
| Customer Type | The customer type (Loyal customer, disloyal customer) |
| Age | The actual age of the passengers |
| Type of Travel | Purpose of the flight of the passengers (Personal Travel, Business Travel) |
| Class | Travel class in the plane of the passengers (Business, Eco, Eco Plus) |
| Flight distance | The flight distance of this journey |

1. survery data



**Attributes:**

| **Name** | **Description** |
| --- | --- |
| Inflight wifi service | Satisfaction level of the inflight wifi service (0 Not Applicable ; 1-5) |
| Departure/Arrival time convenient | Satisfaction level of Departure/Arrival time convenient |
| Ease of Online booking | Satisfaction level of online booking |
| Gate location | Satisfaction level of Gate location |
| Food and drink | Satisfaction level of Food and drink |
| Online boarding | Satisfaction level of online boarding |
| Seat comfort | Satisfaction level of Seat comfort |
| Inflight entertainment | Satisfaction level of inflight entertainment |
| On-board service | Satisfaction level of On-board service |
| Leg room service | Satisfaction level of Leg room service |
| Baggage handling | Satisfaction level of baggage handling |
| Check-in service | Satisfaction level of Check-in service |
| Inflight service | Satisfaction level of inflight service |
| Cleanliness | Satisfaction level of Cleanliness |
| Departure Delay in Minutes | Minutes delayed when departure |
| Arrival Delay in Minutes | Minutes delayed when Arrival |
| Satisfaction | Airline satisfaction level(Satisfaction, neutral or dissatisfaction) |

**Skill Requirement**

* numpy
* pandas
* matplotlib
* seaborn

# **Phase1**

You have been provided with 2 datasets. You will be learning here how to create the dataframe from 2 datasets and make some minor changes as required.

Recognize the attributes carefully and make sure they are aligned in proper format.

## Task1

1. Import all the relevant packages (Eg: Numpy, Seaborn...)
2. Import the datasets into the python environment.
3. Check the structure, statistics and other important functions. (Only observe the changes)
4. Create a new dataframe “df” by joining the 2 datasets

## Task2

1. Deal for missing values
2. Drop the duplicate data
3. Rename the columns with string title format
4. Set “ID” as Index columns
5. Replace the Column values in “Satisfaction” column with Logic;

| **Old Names** | **New Names** |
| --- | --- |
| neutral or dissatisfied | No |
| satisfied | Yes |

# **Phase2**

Congratulations on completing the necessary activities and get ready with the dataset. Now you will be starting to learn a little more about the dataset by performing data wrangling.

This is the most time consuming task. You have to understand the distribution of records with Pandas and visualizations with Matplotlib or Seaborn.

Make sure you create charts for almost all the analysis you are performing with pandas.

Expected dimension of the dataset: 120634 rows, 23 columns

## Task

1. Univariate analysis of each variable
2. Bivariate Analysis of categorical to numerical variables
3. Multivariate Analysis among categorical and numerical variables
4. Check distribution of variables

# **Phase3**

Well this is the time to apply your understanding of the dataset and find out the records that gave some interesting responses.

This phase of the work involves creating queries or groups. You have to answer the questions in order to reach the conclusions.

Make sure you create charts for almost all the analysis you are performing with pandas.

Expected dimension of the dataset: 120634 rows, 23 columns

## Task

1. Perform data manipulations and slicing’s to find the hidden insights on the dataframe “df”.
2. Create a new Dataframe “newdata” which contains all numerical columns as well as dummy values of the categorical columns. Perform Correlation on this dataframe