



Prediction of Best Airlines

Group 01

Tomasz Olewicz, Alex Lamp, Irene Kang, Joe Chun

TABLE OF CONTENTS

- 01 Topic for the Project**
- 02 Source of Dataset**
- 03 Tools for the Project**
- 04 Data Exploration and Analysis**



“The World is one big data problem.”

— **Andrew McAfee, co-director of the MIT Initiative**

01 Our Topic

Our project's goal is :

- Classification of each airline that traffics through San Francisco International Airport(SFO)
- Prediction of each airline's accuracy based on their performances
- Provide potential future customers with precise information of each airline's quality and how to choose one for their own benefit

We chose this because



Popularity

A lot of people in the world travel for business, vacation, reunion, etc.

San Francisco International Airport(SFO) is the point of intersection and the major gateway to the rest of the world

We plan to achieve our goal by



Machine Learning Model

Build a machine learning model to predict which airline has better quality so more potential customers in the future can have better experience

A wide-angle, high-angle photograph of the San Francisco International Airport terminal at dusk. The terminal's iconic glass and steel structure is illuminated from within, with the words "SAN FRANCISCO INTERNATIONAL" prominently displayed in large, dark letters across the upper glass facade. The roof is a complex, arched steel truss system. Below the terminal, a multi-level highway interchange is visible, with several vehicles including cars, buses, and a train moving through the scene. The foreground shows a parking lot with yellow markings and some trees. The sky is a deep blue, indicating twilight.

SAN FRANCISCO INTERNATIONAL

Which airline would you choose ?



02

Source of Dataset

Which Dataset did we use to
achieve our goal?

SFO Air Traffic Passenger and Landings Statistics

<https://www.kaggle.com/san-francisco/sf-air-traffic-passenger-and-landings-statistics?select=air-traffic-landings-statistics.csv>

Top 100 Airline Fleets

<https://www.kaggle.com/tracyvanp/airlinefleet>

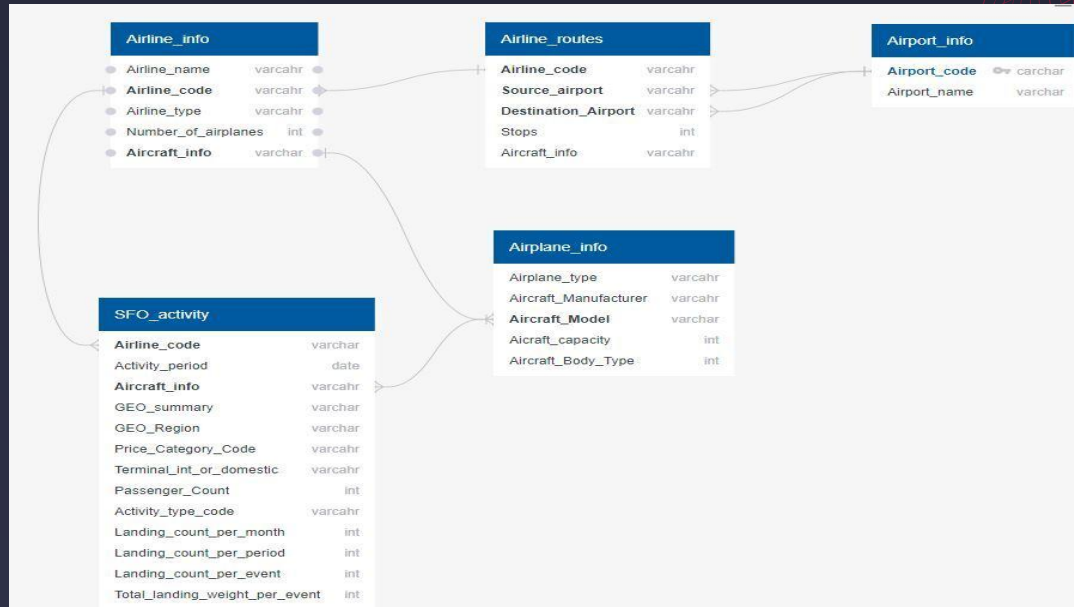
Airline Performance

airline_performance.csv

includes information of each airline's traffic details such as

- Passenger count
- Landing Count
- Number of Airplanes
- Total Cost
- Average Fleet Age
- Each Published Airlines

Our initial **ERD**





03

Tools for the Project

Which tools did we use on the project?



Pandas for

- Encoding
- Trimming out unnecessary values
- Merging different frames together
- Tables in Postgres
- Further analysis



Tableau for

- Visualization
- Storytelling
- Presentation



Machine Learning Model

- Logistic Regression
- Support Vector Machine(SVM)
- Python
- Scikit Library

04

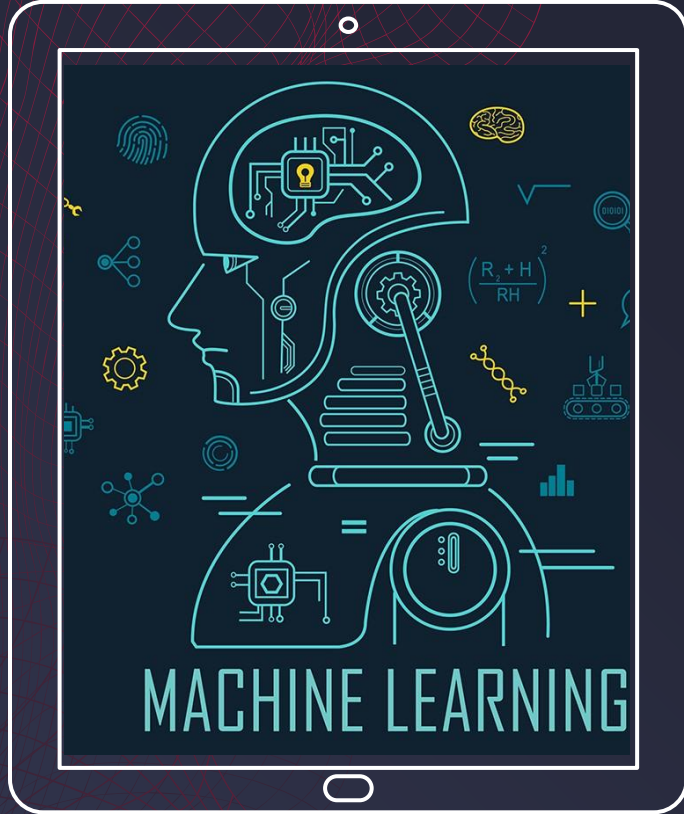
Data Exploration and Analysis

Three different datasets were analyzed and prepared for execution in the ML model

Using Python-pandas :
cleaned data frames and
further analysis was done

Dropped columns that were
not useful and merged into
one to prepare for the ML
model





Continued Analysis

Scikit Library will be used for our ML model for classification. We plan to deploy two different models; that are Logistic Regression and Support Vector Machine

Dashboard

We plan to use Tableau for our Dashboard which will be great for presentation including analysis and visualized datasets

Interactive Elements

Accuracy
Tables
Charts
Features
Priorities
Preferences
Scale(Best to worst)

