# Data Analysis Project

## Data Analyst: Swapnil Chauhan

## Client/Sponsor: Public Airline

## Purpose:

*Write a brief description of why this project is happening below. Why is this project happening? What are the goals?*

The goal of this project is to predict the sentiments of the user’s tweet. This project contains the dataset which includes the tweets scrapped from twitter. The data that included in this project is of february 2015. The data has been categorized by the contributors in three categories - ‘positive’, ‘negative’ and the neutral. The categories are also added for the negative reasons e.g. ‘rude staff’, ‘delay flight’. The final deliverable will include the ML model to predict the sentiment of the tweet.

## Scope / Major Project Activities:

*What are the major parts of this project? List out the high-level steps, activities, or stages of the project, and give a brief description for each.*

| Activity | Description |
| --- | --- |
| Data cleaning | collect and clean the airline tweet data. organize and process the raw data. |
| Exploratory Data analysis | Perform the Data analysis on the tweet data. Create visualization for data. identify patterns, outliers in data. |
| Build Model | Create a model to predict the sentiment of input tweet |
| Model deployment | Deploy the model on the cloud. |

## This project does not include:

*Specify the things that this project isn’t responsible for doing (out of scope). For instance, “this project does not involve a summation of 2019 data analysis”*

* This project only includes the tweets from February 2015. All the tweets in the dataset are from the same timeline only.
* There are only three possible values for sentiment analysis outcome for this project those are - ‘positive’, ‘negative’ and ‘neutral’

## Deliverables:

*A specific list of things that your project will deliver.*

| Deliverable | Description/ Details |
| --- | --- |
| Data Analysis report | The detailed analysis report of the dataset. Identifying the patterns and outliers. |
| Predictor model | The ML model to predict the sentiment of the input tweet. |
| Final Report | The final report on the model. The model performance and other metrics like accuracy, confidence etc. |

## Schedule Overview / Major Milestones:

*The expected schedule for the project. This can be defined by milestones (e.g. “all data is cleaned and processed”), periods of time (“Week 1 / Week 2”), or other ways based on the needs of the project.*

| Milestone | Expected Completion Date | Description/Details |
| --- | --- | --- |
| *Data cleaning* | *22/11/2022* | *Includes data cleaning, processing and importing into format to be used in project* |
| *Data analysis* | *23/11/2022* | *Perform the data analysis.* |
| *Model building and deployment* | *24/11/2022* | *Creating an ML model using the dataset. Identifying the best performing model and deployment.* |

## \*Estimated date for completion:

*This is my “if all goes well and I have everything I need, this is when I’ll be done” date.*

***25/11/2022***