**“POC ON U.S. AIRLINES DATA**​

# USING PANDAS”

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE CERTIFICATE OF

INDUSRIAL TRAINING

IN

**PYTHON, PANDAS AND MATPLOTLIB**

**Submitted By:-**

**RISHI RANJAN JHA (JIMS ENGG. MANAGEMENT TECHNICAL CAMPUS,**​

**GR. NOIDA)**

**Table of Contents**

**POC**………………………………………………………………………………………………………………………………​

…………………………………​**1**

**DECLARATION FROM STUDENT3**

**ACKNOWLEGMENT4**

**ABSTRACT5**

**REQUIREMENT SPECIFICATION6**

**INTRODUCTION7**

ABOUT THE PROJECT7

DATA ANALYSIS**7**​

PANDAS**8**​

JUPYTER NOTEBOOK**8**​

MATPLOTLIB**8**​

DATASET**8-9**​

**VISUALISATION USING IBM MATPLOTLIB9-13**

# DECLARATION FROM STUDENT

I **RISHI RANJAN JHA** hereby​ declare that project report on, “**DATA**​ **ANALYSIS ON U.S. AIRLINES SECTOR**”,​ which is being submitted in

partial fulfillment for the certificate of industrial training, is the record of authentic work

carried out by **RISHI RANJAN JHA** during​ the period **from**​ **20 MAY, 2020 to 3 JULY, 2020** under the guidance of **Ms.**​ **JYOTI MEHRA AND Mr.ROHAN KUMAR .**

We hereby declare that this project work has not been submitted before or for any other purpose.

**This is to certify that the above statement made by the candidate is correct to the best of my knowledge.**​

**TRAINER NAME:-**​ **JYOTI MEHRA AND ROHAN KUMAR**

**PROJECT GUIDE:-**​ **JYOTI MEHRA AND ROHAN KUMAR**

# ACKNOWLEDGEMENT

I **RISHI RANJAN JHA** are using this opportunity to express our gratitude to everyone who supported us throughout the course of this project. We are thankful for their aspiring guidance, invaluably constructive criticism and friendly advice during the project work. We are sincerely grateful to them for sharing their truthful and illuminating views on a number of issues related to the project.

We express our warm thanks to **Ms.**​  **JYOTI MEHRA AND Mr. ROHAN KUMAR** for​ her support and guidance.

It certainly was a big learning curve for us. Even though we did face many difficulties ultimately the careful planning and the dedication by which we were able to complete all the work. In the documentary, we try and list the complete process that went into making the text pad into complete structure.

Thank you,

**RISHI RANJAN JHA**

**ABSTRACT**

**What is PYTHON?**

Python is an interpreted, high-level, general-purpose programming language. It is created by Guido Van Rossum and first released in 1991; Its high-level built in data structures,​ combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed. ​

**Sources of PYTHON:-**​

1. **Django**​ **:-**​ **High level web framework that encourages rapid development and clean, pragmatic design.**
2. **Pandas**​ **:-**​ **Software library in computer programming, used in python to deal with Data Analysis and Manipulation.**
3. **Pipenv**​ **:-**​ **It allows us to create a “virtual” isolated python installation and install packages into the virtual installations.**
4. **Matplotlib**​ **:-**​ **A plotting library for the python programming language and its numerical mathematics extension “numpy”.**
5. **Micropython**​ **:-**​ **Is a software implementation of a programming language largely compatible with python 3, written in C, i.e. optimized to run on a microcontroller.**

**REQUIREMENT SPECIFICATIONS**

* 1. **Hardware Requirements:**

Processor : ​ AMD RYZEN 5​

Processor Speed : 2.1 GHz

RAM : 8​ GB

Hard Disk : ​ 512​  GB SSD

* 1. **Software Requirements:**

Operating System : WINDOWS7, UNIX, LINUX

Software : ANACONDA JUPYTER

* **INTRODUCTION TO PYTHON:-**​

Python is designed to be​ **simple**​ and ​**easy**​ to ​**read and write**​. Python can be utilized for a ​**wide range of applications**​ such as ​**scripting**​,​ **developing**​, and ​**testing**​. Because of its elegance and ​**simplicity**​, ​**Dropbox**,​ ​ **Google**​, ​**Quora**​, ​**Hewlett-Packard**​, and so many other top tech-companies have already implemented Python. With the growing trend of Data Science and Machine Learning, Python has equally become one of the ​**most important scripting**​ ​**languages**​ in the ​**21st century**​, because of its wide

range of libraries and framework.

* **ABOUT THE PROJECT:-**​Our project is based on “DATA​ ANALYSATION” using PANDAS working with the dataset of U.S.

AIRLINES. It shows the Customer feedback on the U.S. AIRLINES SECTOR.

**1. DATA ANALYSIS**​ **:-**​ **Data analysis**​ ​ ​is a process of

inspecting, ​[cleansing](https://en.wikipedia.org/wiki/Data_cleansing)​, ​[transforming](https://en.wikipedia.org/wiki/Data_transformation)​ and ​[modeling](https://en.wikipedia.org/wiki/Data_modeling)​​[data](https://en.wikipedia.org/wiki/Data)​​with the goal of discovering useful information, informing conclusions and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains. In today's business world, data analysis plays a role in making decisions more scientific and helping businesses operate more effectively.

**2.** ​ **PANDAS**​ **:-**​Pandas is mainly used for ​ ​[data analysis](https://en.wikipedia.org/wiki/Data_analysis)​[.](https://en.wikipedia.org/wiki/Data_analysis) Pandas a​llow importing data from various file format such as ​comma separated values​, ​[JSON](https://en.wikipedia.org/wiki/JSON)​[,](https://en.wikipedia.org/wiki/JSON)​[SQL](https://en.wikipedia.org/wiki/SQL)​, ​[Microsoft Excel](https://en.wikipedia.org/wiki/Microsoft_Excel)​. Pandas allow various data manipulation operations such as merging, reshaping, selecting, as well as ​[data cleaning](https://en.wikipedia.org/wiki/Data_cleaning)​, and ​[data wrangling](https://en.wikipedia.org/wiki/Data_wrangling) features.​

**3.** ​ **JUPYTER NOTEBOOK**​ **:-**​Jupyter ​ ​[Notebook](https://en.wikipedia.org/wiki/Notebook_interface)​ is a ​[web-based interactive](https://en.wikipedia.org/wiki/Rich_Internet_application)​computational environment for creating Jupyter notebook documents. provides a browser-based ​[REPL](https://en.wikipedia.org/wiki/Read%E2%80%93eval%E2%80%93print_loop)​built upon a number of popular [open-sourc](https://en.wikipedia.org/wiki/Open-source_software)​ [e](https://en.wikipedia.org/wiki/Open-source_software)​ libraries:

* [IPython](https://en.wikipedia.org/wiki/IPython)
* [Tornado (web server)](https://en.wikipedia.org/wiki/Tornado_(web_server))
* [jQuery](https://en.wikipedia.org/wiki/JQuery)
* [Bootstrap (front-end framework)](https://en.wikipedia.org/wiki/Bootstrap_(front-end_framework))
* [MathJax](https://en.wikipedia.org/wiki/MathJax)

**4.** ​ **MATPLOTLIB**​ :-​ ​ Matplotlib is a [plottin](https://en.wikipedia.org/wiki/Plotter)​ [g](https://en.wikipedia.org/wiki/Plotter)​ ​[library](https://en.wikipedia.org/wiki/Library_(computer_science))​for the ​[Python](https://en.wikipedia.org/wiki/Python_(programming_language)) programming​ language and its numerical mathematics extension ​[Numpy](https://en.wikipedia.org/wiki/NumPy)​. Pyplot is a Matplotlib module which provides a MATLAB-like interface. Matplotlib is designed to be as usable as MATLAB, with the ability to use Python, and the advantage of being free and open-source.

**4.** ​ **DATA SET**​ **:-**​Data set is the set of data which we want to use for our​ analyzation. We can analyze the data by plotting graphs, making heat maps etc. and can study a smaller or comparatively larger portion of data anyhow.

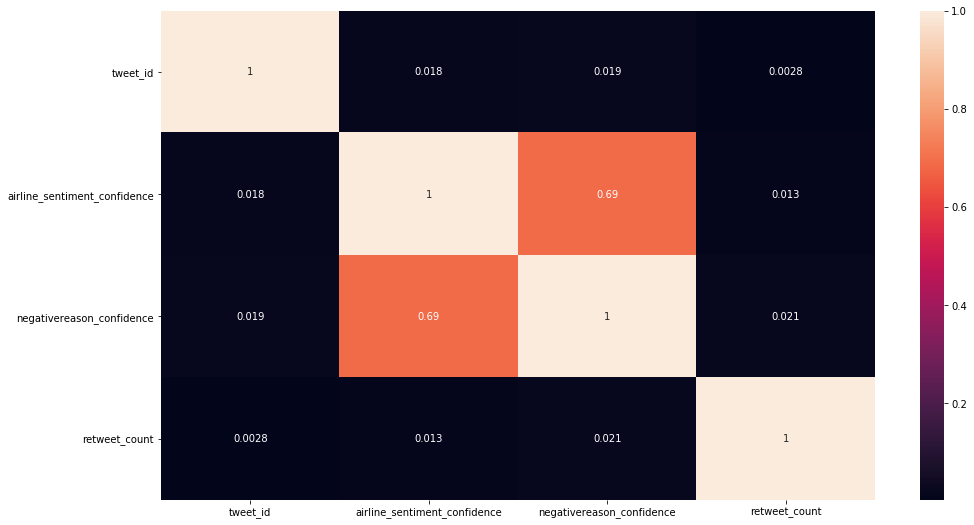
Here we have the data set of “**U.S. AIRLINES SECTOR**​ ” based on​ their **feedback of travelling from 6 different types of airlines and**​  **the problems they faced during their travel**. This data set has been​ picked up from Kaggle website and has analyzed accordingly through maps and graphs. Also, dropping down the null values to make it more appropriate and feasible.​

**VISUALISATION USING IBM MATPLOTLIB:-**​



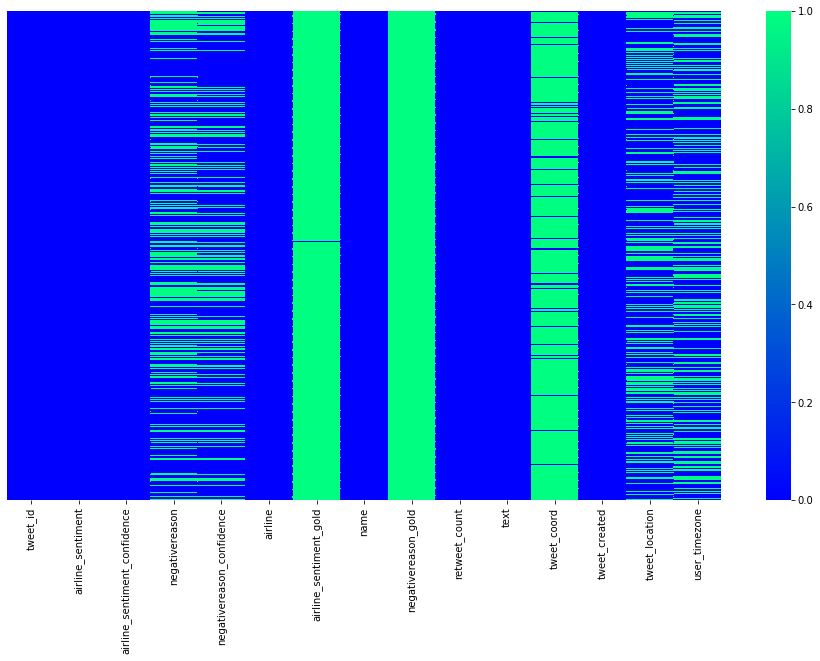
## GRAPH 1

This is the first graph used in our project which shows customers feedback as “positive”, ”negative” or “neutral” according to the Airlines in which they have travelled and the conditions being there. Also, the problems faced by them accordingly resulted in this type of graph.



## HEAT MAP 1

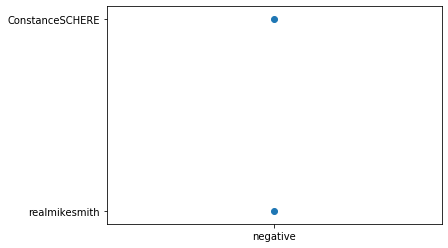
A heatmap is a two-dimensional graphical representation of data where the individual values that are contained in a matrix are represented as colors. Here also, all the data present in data set in different rows and columns has been shown in a graphical format. Which is done using seaborn.



## HEAT MAP 2

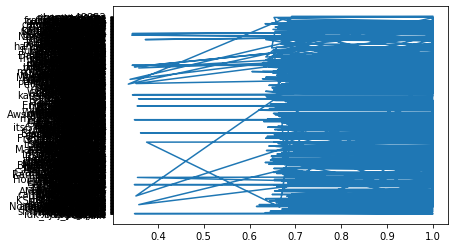
This is a heatmap which is used to check if there is any null values present in the given data set or not. Here, we got a number of null values as spaces between lines which represent the null values.

This makes us aware that if we want to analyze the data perfectly, we need to first remove or drop down these null values.



## SCATTER PLOT USING MATPLOTLIB

This scatter plot is based on the feedback from customers who have travelled from airline named as ‘AIRLINE SENTIMENT’. A limited amount of data has been taken from data set to make a clear scatter plot using Matplotlib. Here, we have taken data of 1000 customers who have travelled in the particular airline and plotted its scatter diagram.



## GRAPH 2

This graph is based on the feedback from customers who have travelled from airline named as ‘AIRLINE SENTIMENT CONFIDENCE’. A limited amount of data has been taken from data set to make a clear scatter plot using Matplotlib. Here, we have taken data of 1000 customers who have travelled in the particular airline and plotted a graph between their names and that type of airline.

**THIS IS ALL ABOUT THE PROJECT BEING SUBMITTED.**