Project : Airport Traffic analysis
Name : Jayakrithi Shivakumar

Problem
Covid-19 pandemic which started in late December 2019 has drastically impacted many industries affecting the global economy. One such industry which was severely impacted is the airline industry. In this project, we aim to analyse the impact of Covid-19 on the aviation industry based on the air traffic pattern.

Objective :
This is an exploratory data analysis project which analyzes the airport traffic pattern during the pandemic period across various international airports across the globe.

Technologies Requirements :
This analysis is performed on Microsoft Azure cloud. In this project, the Azure Machine Learning studio and the Azure storage services are used.
Python is a language requirement for this project as we perform data analysis on Azure ML notebooks which are python notebooks hosted on azure cloud.

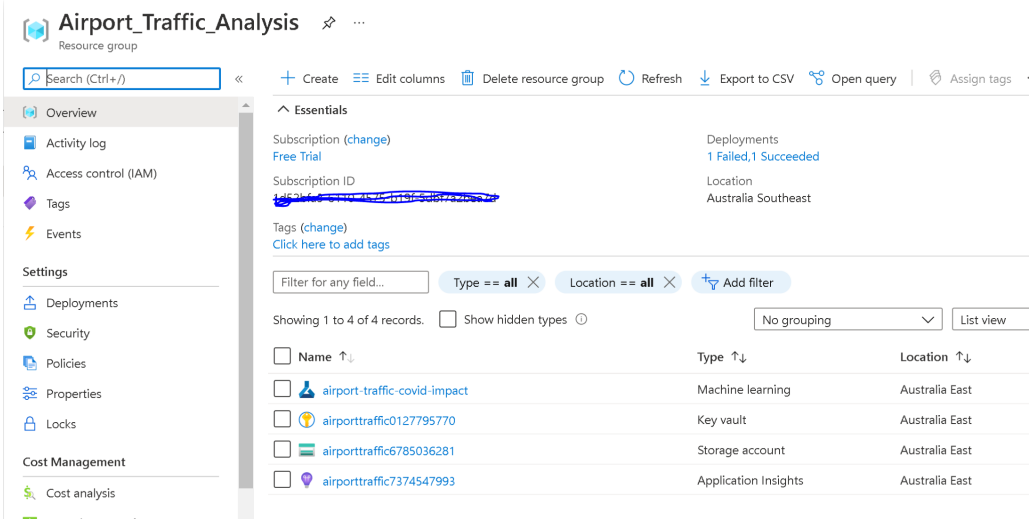
Sources :
Kaggle data : https://www.kaggle.com/terenceshin/covid19s-impact-on-airport-traffic

Implementation :

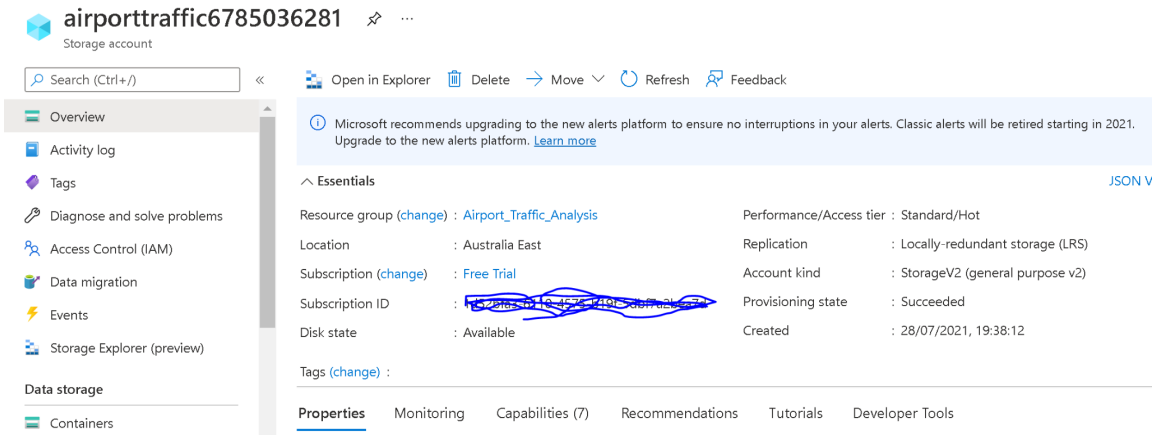
Setup:

Azure resource Group:

Creating an azure resource group is essential to group related components of the same project together. In this project we need a storage account and machine learning studio, along with other critical components such as key vault.

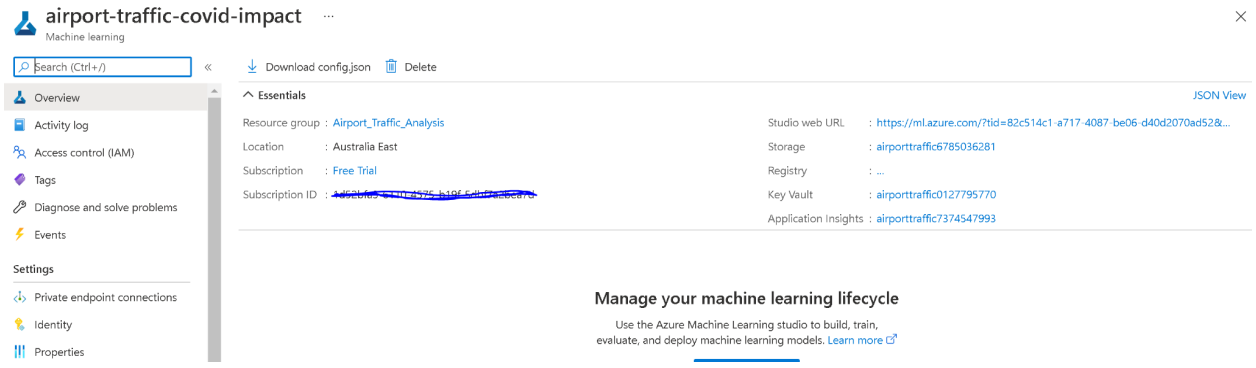


Azure Storage Account :



- The dataset can be stored as a blob under containers in the azure storage account. This blob storage will be later accessed by the machine learning tool by azure via azure ML datastore.

Azure Machine Learning Studio:



- The azure machine learning studio is a GUI for building, training and deploying machine learning models. For this project, I have used the Azure ML Notebooks and Azure Datastore.
- The Azure Datastore enables us to connect the azure storage with the azure Machine learning tool to access data.

Azure ML Notebooks:

- A new compute engine called *basic_compute* (Standard_DS2_v2 (2 cores, 7 GB RAM, 14 GB disk)) is created.

Code Implementation

Libraries Used:

- Azure essential libraries such as azure.core for importing Workspace and Datastore.
- Importing Pandas for dataframe
- Importing seaborn and matplot for data visualisation.

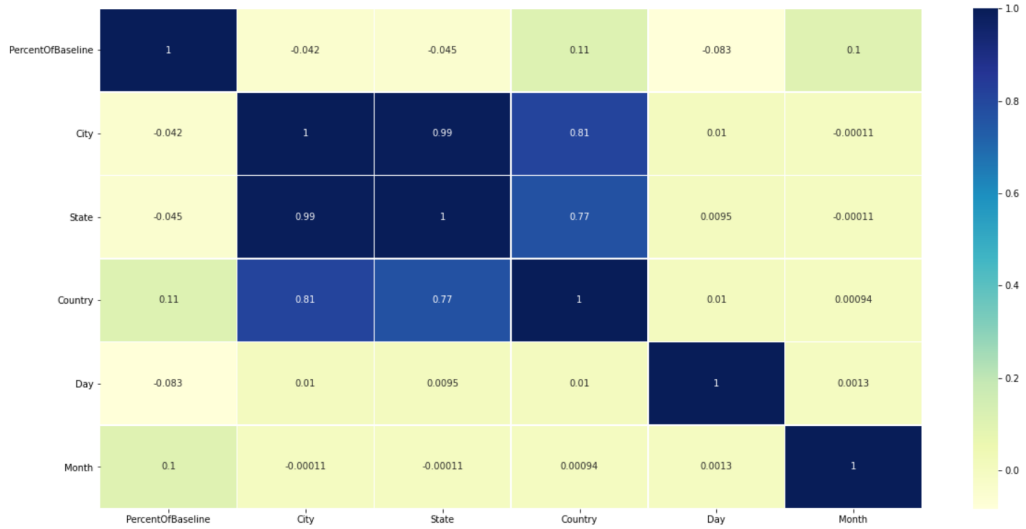
Results :

Techniques used :

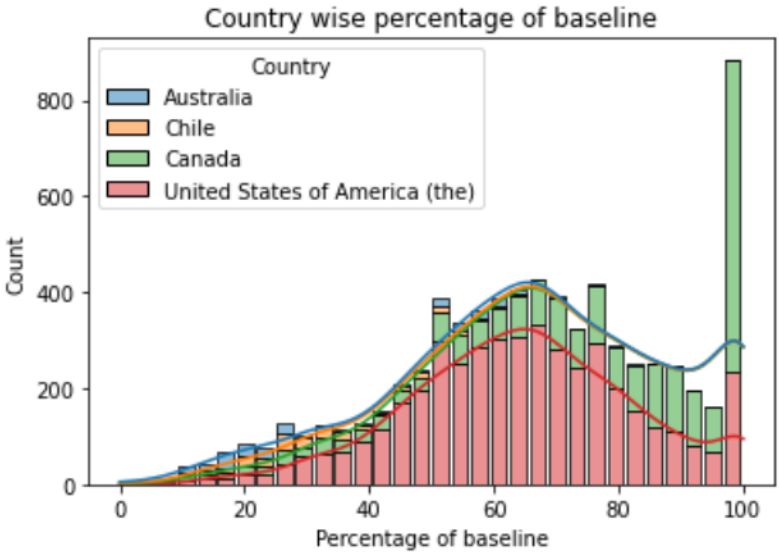
As this is mainly an EDA project, we have incorporated various visualisation techniques using seaborn and matplot libraries to analyse the pattern in the given time frame.

Correlation Matrix :

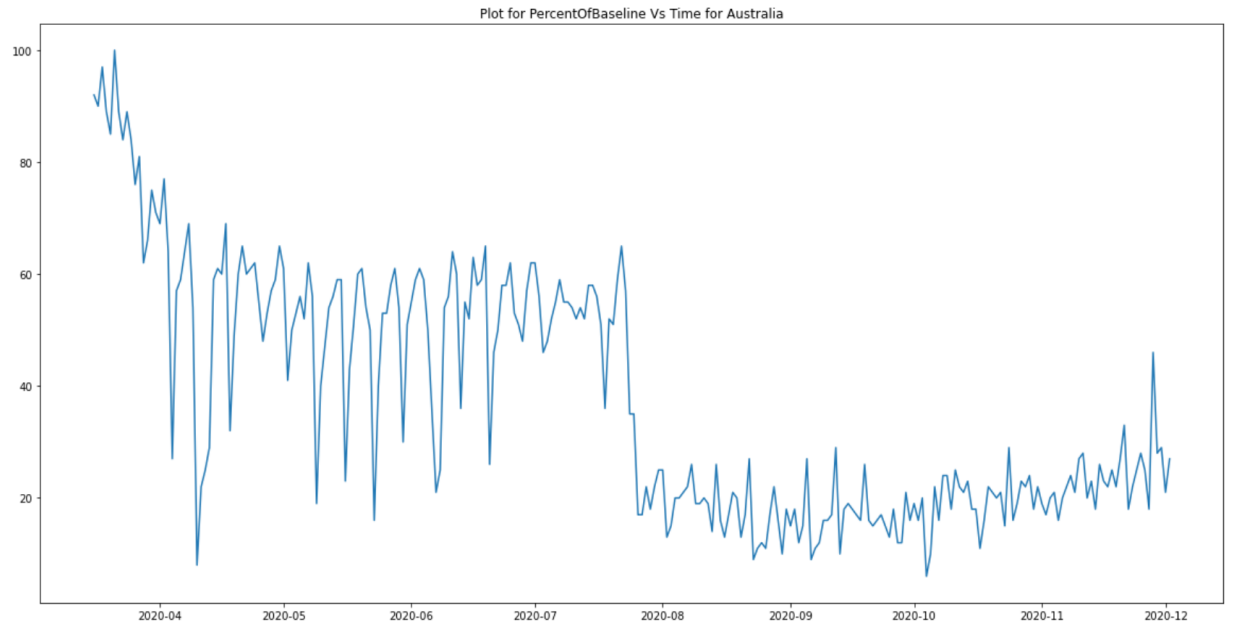
- We first label encode all the values.
- Using seaborn library, we plotted a correlation matrix to understand the dependency factors between the different attributes.
- But the main pattern we would like to analyse is the dependency of percentage Baseline with the other attributes.
- We can clearly conclude that there is a very slight positive correlation between percentage baseline - Month and country. Which means that the percentage baseline which is the average number of trips made in that airport was influenced by country and the month.



Plotting percentage of baseline based on country



Analysis for Australia :



We can see that there is a drastic drop in the percent of Baseline over the period of 9 months from April to December 2020.

Reference :

1. Github : https://github.com/jayakrithi/airport-traffic-pattern/blob/main/airport_traffic_analysis.ipynb
2. Code reference : https://www.kaggle.com/palanjali007/time-series-eda-forecasting