Eric Evans

Irene Zou

Data Analytics Boot Camp

09 February 2021

Covid-19 impact on airport traffic - ETL Project

Extraction

We found 11 files of data about covid-19 cases, its impact on airport and the vaccine progress around the world on a Kaggle. Subgroups of these files focused on elements such as location of the data including the country, state and city, daily vaccine distribution, percent of base line for airport traffic, and other topics. In this project, we used the data in the country\_vaccinations.csv, covid\_19\_clean\_complete.csv, and covid\_impact\_on\_airport\_csv files from Kaggle.

Transform

The goal was to align the date of covid-19 cases data and the data for the airport traffic by date for users to see the potential connection between the two if any; and to check the numbers of vaccine from different providers. In order to align the data, we had to remove changed the date format and match the countries’ names. We also update the vaccine providers columns to make it better to check the popularity of the vaccines in different continents.

For the data from country\_vaccinations.csv, we focused on the vaccine providers and countries they provided for. We trimmed down columns with a lot of missing data and was able to find one column that contains data that is important to use. From the same file, we used the loc function to create new data frames to make further analysis available.

For covid\_19\_clean\_complete.csv and covid\_impact\_on\_airport\_csv, we changed the column names that make the join possible for loading and we also tried to group the data by dates so that it would it possible to match the two and have the right date.

Load

We load the covid-19 case date frame and the airport traffic data frame as a joined table because it might be useful for potential users to monitor and correlation between the two and make business decision for the future base on the situation if there is similar situation. In addition, we load the vaccine data frame sedately because its date won’t match with the other two and we are focus on the vaccine providers for that data.