TechTronix Round 2 Task

There have been many incidents of flight delays in the US, affecting the country's rank and reputation on a global level. The federal government has taken this downfall seriously and wants to know the reasons for the same from the National Airport Committee of US. The Chief of National Airport Committee has asked you, **The Chief of Insights** to dig deeper into problem and find patterns and insights which can help the committee address the issue.

You have been given a dataset consisting of 36 files (.csv format). Each file is the data for one calendar month. The data are the records of all flights taking off from and landing at all US airports. It covers a period of 3 years. Each file has approximately half a million rows, the total dataset consisting of close to 1.7 crore rows.

There are 44 fields. Out of these please ignore the last 6 fields (DISTANCE_GROUP, CARRIER_DELAY, WEATHER_DELAY, NAS_DELAY, SECURITY_DELAY, LATE_AIRCRAFT_DELAY) for this exercise. The description of rest of the fields has been provided in the data description file.

You need to create dashboards, visualizations, charts or any other thing that you find relevant, generate insights and provide recommendations.

For example, you can create dashboards to visualize delays at various airports, insight here can be identifying top airports that have had the most proportion of flights (taking off and landing there) delayed at (a.) take-off and (b.) landing during the 3 year period. Further you can create visualizations to identify if there is any seasonality associated with airport delays. For example, airports A, B & C may be located in high snowfall areas and have more delays in winter, while airports X, Y & Z may be located close to the sea and have more delays in summer (and / or similarly for other season(s)). You can identify airports that have the seasonality component. Another can be dashboards regarding performance of various airlines and identify the top and worst performing airlines.

Please note this is not an exhaustive list, these are just a few examples, you can create dashboards regarding all the things you find relevant and you can generate as may insights as you can.

For airport delays consider only those cases (days / airports) that have had at least 10 take-offs and 10 landings per day (on an average during the 3 year period)

A seasonal delay will be recognized if a high proportion of delays (compared to the average proportion of delays at that airport) have occurred in the same set of consecutive days / weeks / months in at least 2 of the 3 years. Consider only those cases (days / airports) that have had at least 10 take-offs and 10 landings on a given day.

All delays may not be attributable to the airlines. For example, there may be a baggage handling system fault in an airport on a given day. All flights going out of that airport will be delated that day. Or, there may be a security alert / heavy snow in an airport one day. All flights coming in or going out of that airport will be delayed on that day. These types of delays are called 'systemic delay'. For evaluating airline performance, all systemic delays will need to be ignored. Also, consider only those airlines that have operated at least an average of 10 flights per day for at least

2 years (not necessarily contiguous) during the 3-year period.

What is a delay? A flight will be considered delayed if (actual arrival time – scheduled arrival time) / (Scheduled flying time) > 10%. A landing will be considered delayed if (actual arrival time – actual departure time – scheduled flying time) / (scheduled flying time) > 10%. A take-off will be considered delayed if (actual departure time – scheduled departure time) / (scheduled flying time) > 10%.

Note: Time of flights are local times at the respective airports. Adjust for time-zones where required.

DELIVERABLES

- Prepare a presentation having snippets of your dashboard, visualizations, charts, etc and mention your insights and recommendation in that
- Send a zip file to infinit-I's mail id containing the presentation, excel sheets or any other files that are necessary

You can use any software to create visualizations and analyze data.