MONASH UNIVERSITY

FIT5147: Data Exploration and Visualization – Sem 2 2022

11-08-2022

Programming Exercise 1: Tableau Public

Submitted By:

ISHITA KHANNA

32743823

Task 1: Checking the errors and cleaning the data

I have loaded the RFDS_Flight_July2022 data.csv into Tableau for visualizing the errors have found using Excel and seeing the table in Tableau.

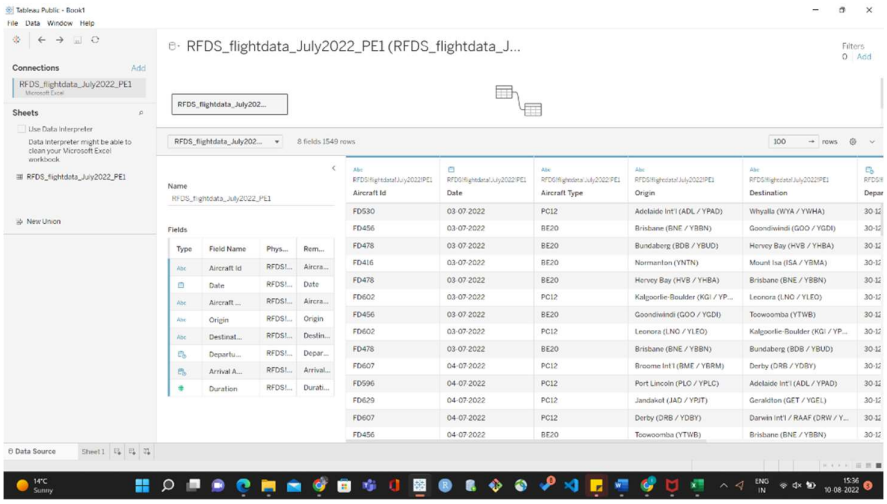

Image 1: When data is loaded in the tableau

After dragging and dropping few attributes, I was able to find three errors in the dataset and here is the explanation and visualization related to it.

Error 1: Null values

There are some flights in which the duration of the flight is not given because their arrival airport is not stated which means they have been diverted or it was a connecting flight. This created a disturbance in the data because of the null values.

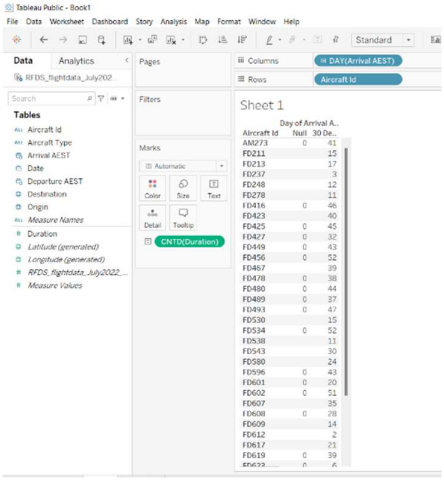

Image 2: Visualized data in a tabular form to detect the null values in Arrival time and duration

Solution: To make the adjustment, I have changed all the null values to 0 so that there will be no breakage in the visualization.

Error 2: Unknown Aircraft type

There has been a classification of aircraft model into 4 types in which 4th type is unknown and due to which few flights don't have any aircraft model. So, when we wanted to see what flights have what model, we weren't able to get the model for few flights that fell under the unknown category.

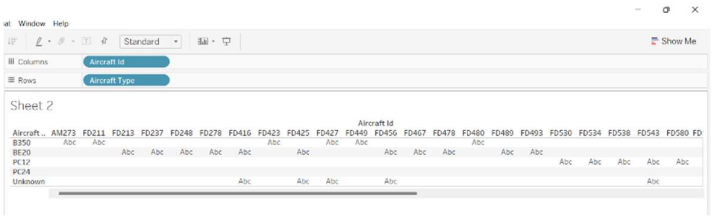

Image 3: Unknown aircraft type as an error

Solution: I think to solve this problem we just have to remove the data of flights that fall under this category because we can't really visualize and analyze the flights related to something unknown. We can still analyze the flights on another basis but just not this.

Error 3: Outliers

When we see the duration of the flights based on their origin and destination, there are some outliers because the time that flight takes to reach that destination doesn't make sense in reality which ultimately classifies them as outliers.

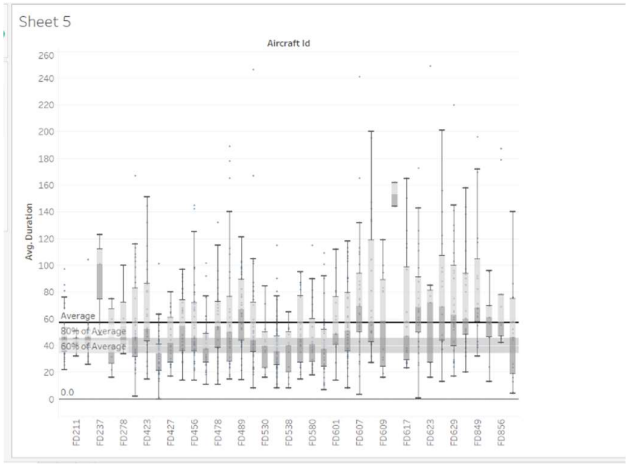

Image 4: Outliers displayed when visualize the average duration of each flight

Error 4: Destination and Origin airport is same

As seen in Image 7 and 8 that there are two flights which have same airport as destination and origin which is not possible in reality.

Image 5: Adelaide have same origin and destination of PC12 flight

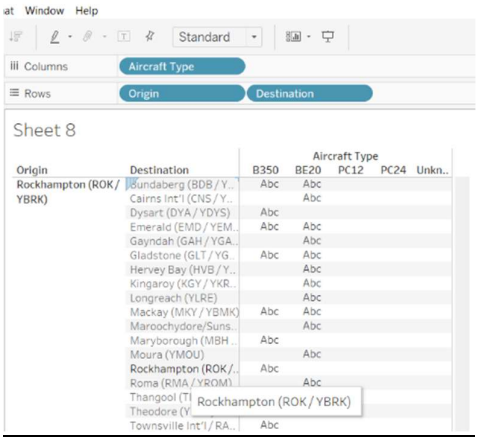

Image 6: Rockhampton have same origin and destination for B350

Task 2: Data Exploration

Q1: Compare and contrast when different types of aircraft flew. In what ways does this show that the usage of the aircraft types is similar and/or different to each other? Consider this on both an hourly and day-of-the-week timescale. Does this data show that the RFDS is in the air 24 hours a day?

Answer: After reading this question, I have dragged and dropped Date into column along side weekday to see at what day what aircraft types are flying and to compare how much each aircraft type is used.

Findings:

- Usage of aircraft type is not at all similar, it differentiates on a daily plus hourly basis.
- Seeing the line graph for PC24 shows that it has been flat related to the departure time.

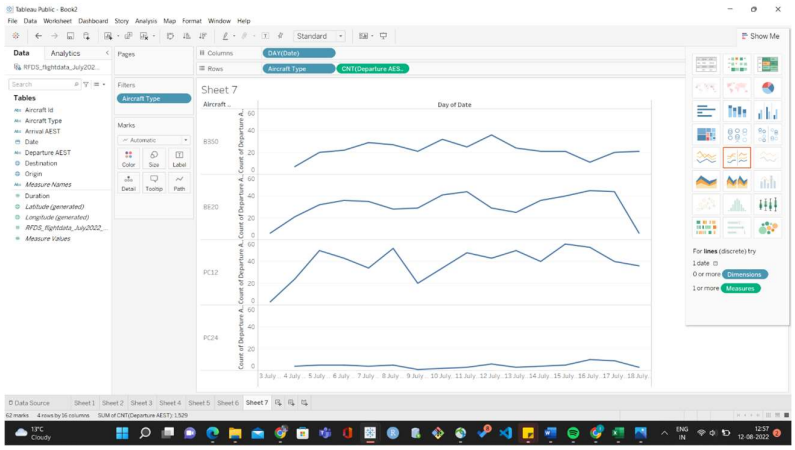

Image 7: Comparing aircraft type on days and hourly basis

This data doesn't show that the RFDS is in the air 24 hours a day because for that we do have to compare the origin and destination AEST in terms of aircraft type which can't be done because of the ungrouped data and latitude and longitude not specified in the origin and destination due to which we can't state that duration is correct in terms of the route.

Q2: Compare and contrast the duration of flights for different aircraft types. What does this tell you about the use of the different types of aircraft by the RFDS? In what ways does this support, challenge or change your conclusions for the first question?

Answer: Aircraft type dropped in the color mark to differentiate it on the basis of color in the stacked bar chart for analysis.

Findings:

- Image 10 shows that PC12 has been used the most and PC24 being the least because it is an overall sum of the duration the aircraft type is been in the air.
- If we see an average of a aircraft type in terms of duration then we can see that PC24 has the greatest average as compared to every other type because it is not used frequently but whenever it is used, it is used for longer duration.
-

Another way to check the data is to make a line graph on the basis of usage i.e., sum of duration of flights on each day which is also showing that PC12 has been used to flew the most and PC24 being the least.

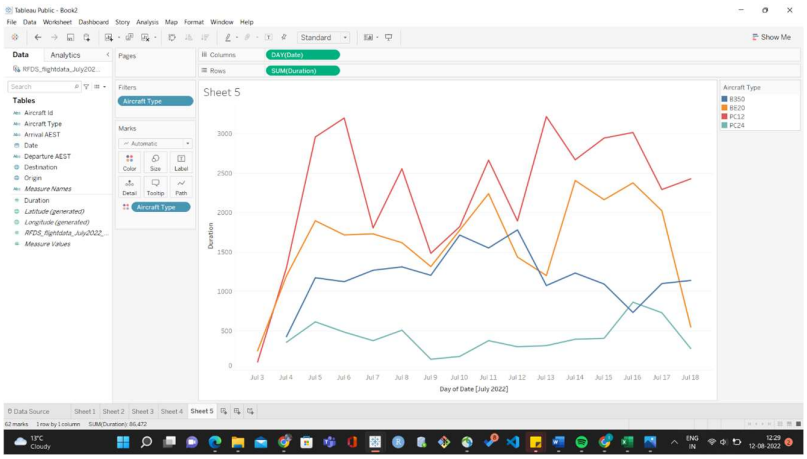

Image 8: Line graph to show usage of aircraft type each day

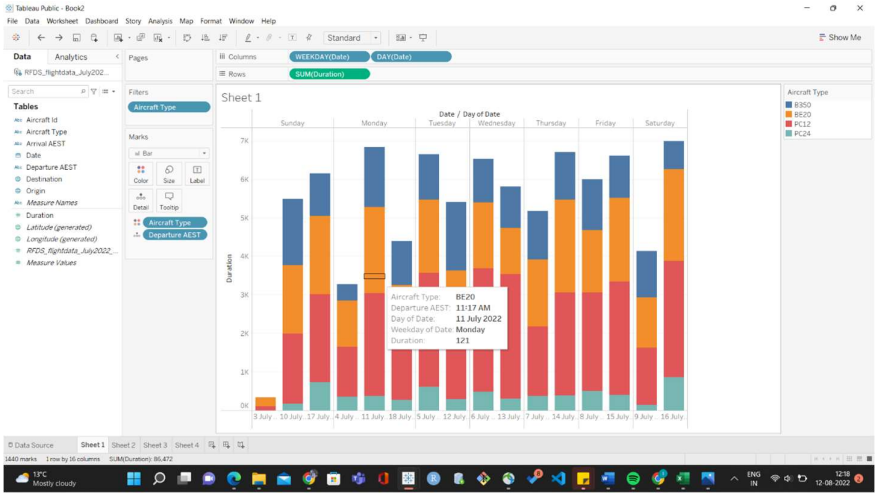

Image 9: Stacked bar chart for each aircraft type.

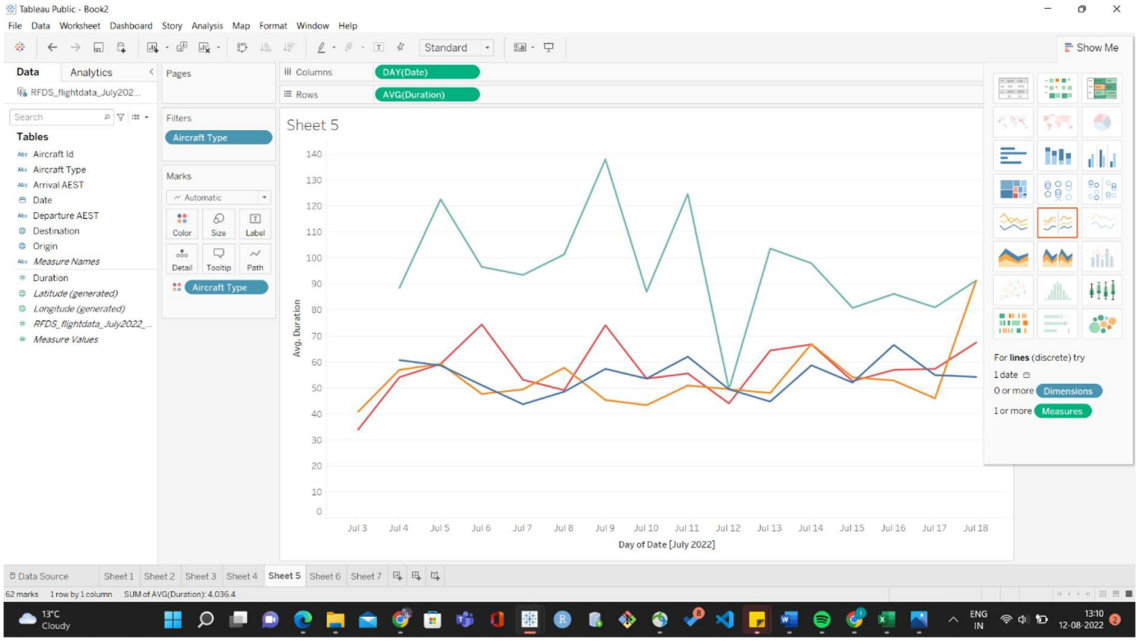

Image 10: Shows the aviation of aircraft type on an average basis