

Wrangling & Exploration – Report

Table of Contents

1. Data checking and cleaning	1
<i>i) Outlier of flight duration</i>	1
<i>ii) Unknown aircraft type</i>	2
<i>iii) Missing values in duration variable for all aircraft types</i>	3
2. Data exploration	4
<i>i) Answer to Q1:</i>	4
<i>ii) Answer to Q2:</i>	5

1. Data checking and cleaning

Figure 1 shows the screenshot of Tableau after the csv file of RFDS_flightdata_July2022_PE1.csv file has been loaded in Tableau.

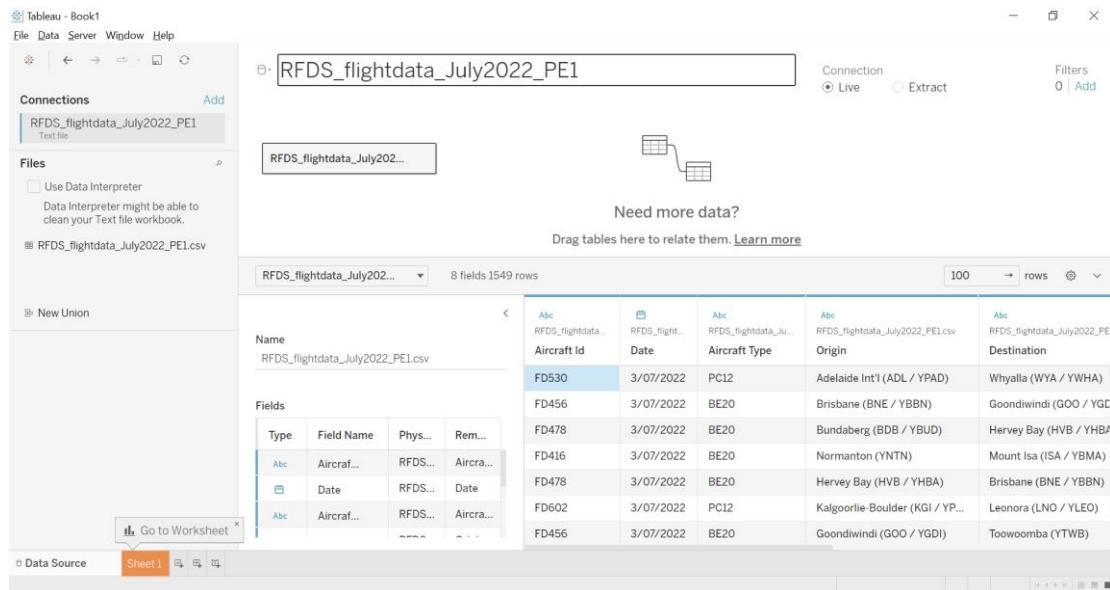
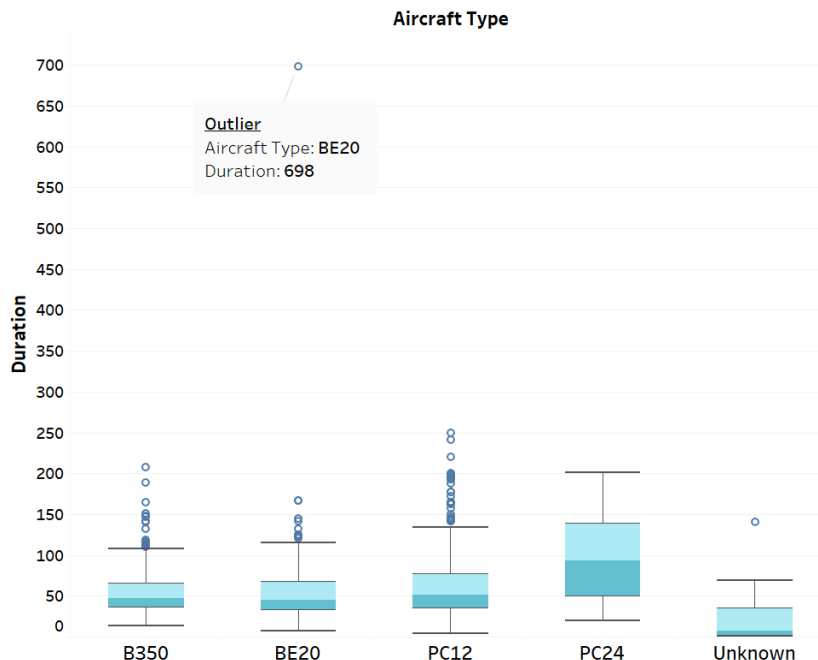


Figure 1. Data in csv format loaded in Tableau.

Using data visualisation, the three types of data irregularities or errors in the dataset are identified as follows:

i) Outlier of flight duration

Outlier of Flight Duration



Duration for each Aircraft Type. The data is filtered on Duration, which keeps non-Null values only.

Figure 2. Outlier data in flight duration.

As illustrated in **Figure 2**, an outlier was found in the flight duration data for aircraft type BE20. An exceptionally long duration of 698 minutes (11 hours 38 min) was recorded for one of the 497 records for the aircraft type. This is probably a time calculation error. Recalculating the duration using the recorded departure and arrival times gives a flight duration of 22 minutes. Other flights by the same aircraft type for the same flight route (same origin and destination) have recorded flight durations, ranging from 7 to 22 minutes. Hence, the duration for this particular entry is corrected to 22 minutes.

Before data correction:

46	FD849	18/07/2022	PC12	Darwin Int'l / RAAF (DRW / YPDN)	Alice Springs (ASP / YBAS)	12:56 PM	Diverted	NA	
47	FD493	14/07/2022	BE20	Kingaroy (KGY / YKRY)	Toowoomba (YTWB)	11:42 PM	12:04 AM	698	
48	FD623	14/07/2022	PC12	Jandakot (JAD / YPJT)	Broome Int'l (BME / YBRM)	7:07 PM	11:16 PM	249	

After data correction:

46	FD849	18/07/2022	PC12	Darwin Int'l / RAAF (DRW / YPDN)	Alice Springs (ASP / YBAS)	12:56 PM	Diverted	NA	
47	FD493	14/07/2022	BE20	Kingaroy (KGY / YKRY)	Toowoomba (YTWB)	11:42 PM	12:04 AM	22	
48	FD623	14/07/2022	PC12	Jandakot (JAD / YPJT)	Broome Int'l (BME / YBRM)	7:07 PM	11:16 PM	249	

ii) Unknown aircraft type

Unknown Aircraft Type

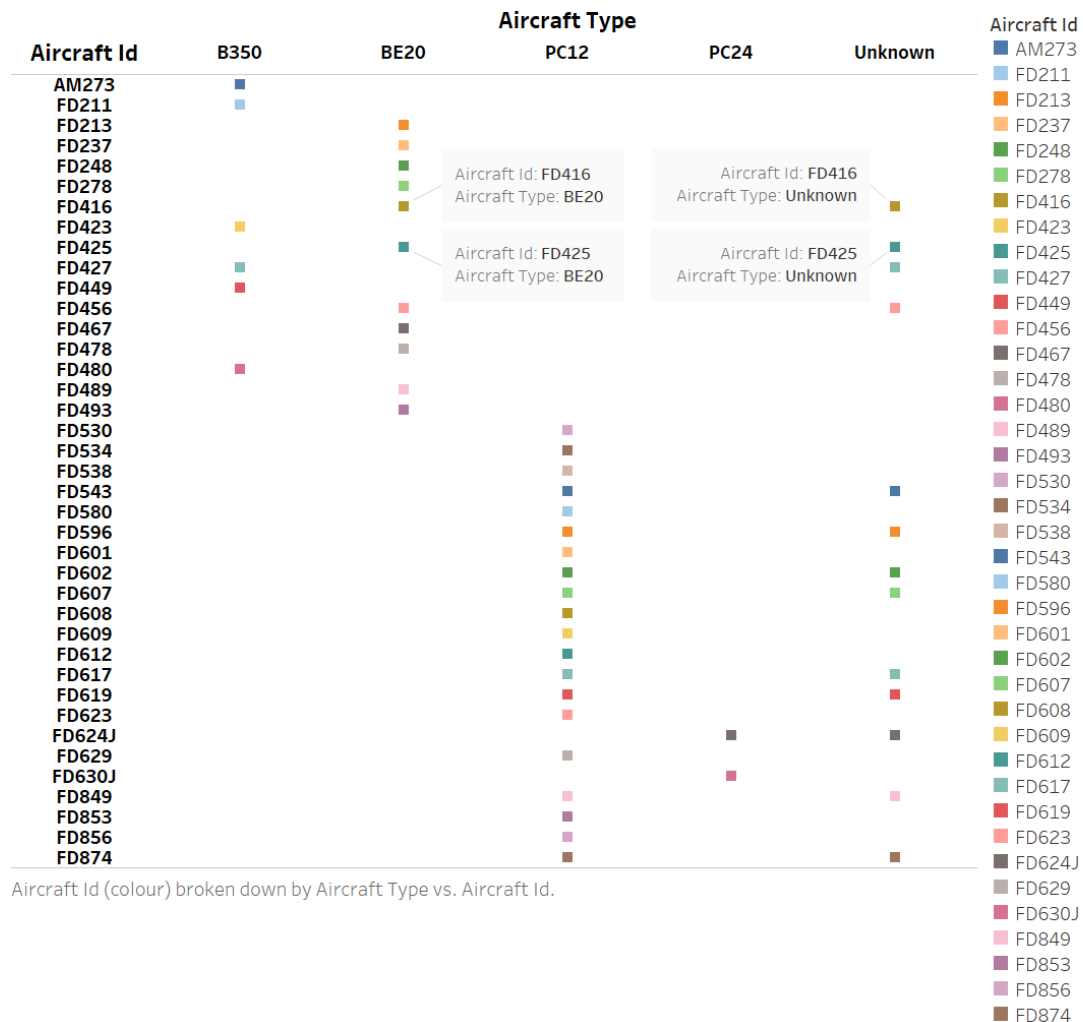


Figure 3. Aircraft of unknown type. The colour represents individual aircraft.

Figure 3 shows that thirteen out of the thirty nine aircrafts in the dataset have two aircraft types: the correct type and unknown type. Since each of the thirteen aircrafts has multiple entries in which their correct aircraft types have been recorded, the aircraft types of these aircrafts can be corrected by matching the correct aircraft types.

Before data correction:

209	FD416	17/07/2022	BE20	Mount Isa (ISA / YBMA)	Cloncurry (CNJ / YCCY)	6:54 PM	7:12 PM	18	
210	FD416	11/07/2022	Unknown	Mount Isa (ISA / YBMA)	Mount Isa (ISA / YBMA)	10:39 AM	10:41 AM	2	
211	FD423	10/07/2022	B350	Brisbane (BNE / YBBN)	Townsville Int'l / RAAF (TSV / YBTL)	2:12 PM	4:43 PM	151	

After data correction:

209	FD416	17/07/2022	BE20	Mount Isa (ISA / YBMA)	Cloncurry (CNJ / YCCY)	6:54 PM	7:12 PM	18	
210	FD416	11/07/2022	BE20	Mount Isa (ISA / YBMA)	Mount Isa (ISA / YBMA)	10:39 AM	10:41 AM	2	
211	FD423	10/07/2022	B350	Brisbane (BNE / YBBN)	Townsville Int'l / RAAF (TSV / YBTL)	2:12 PM	4:43 PM	151	

iii) Missing values in duration variable for all aircraft types

Missing Values In Duration Variable

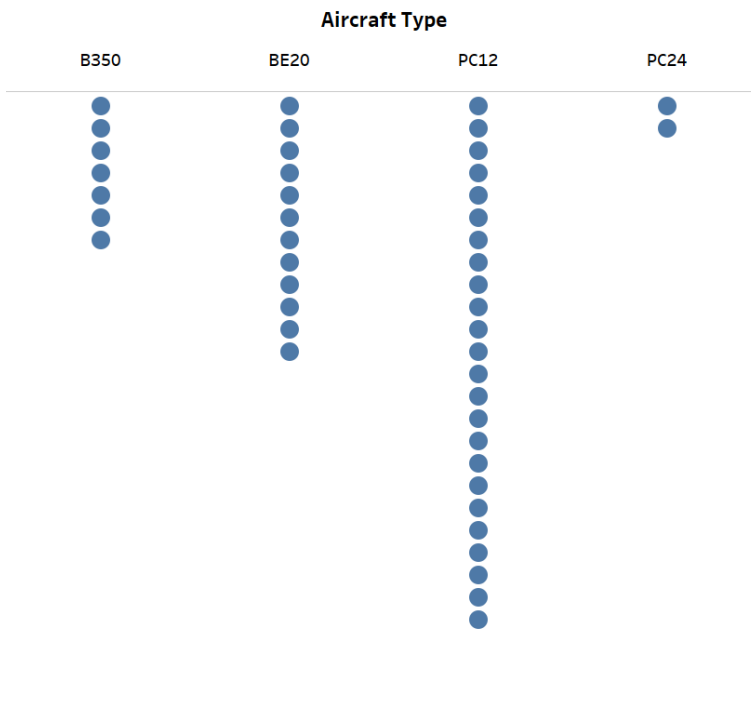


Figure 4 demonstrates that the duration field of certain entries are empty because there were no record of arrival time due to the flights being diverted. As there is no means to recover the data on arrival time, it is not possible to correct these null values. Hence, a better way would be to filter out these entries by deleting them.

The view is broken down by Aircraft Type. The data is filtered on Duration, which keeps Null.

Figure 4. Missing values in the duration variable.

Before deletion:

44	FD601	17/07/2022	PC12	Kalgoorlie-Boulder (KGI / YPKG)	Esperance (EPR / YESP)	9:58 AM	Diverted	NA	
45	FD493	17/07/2022	BE20	Toowoomba (YTWB)	Charleville (CTL / YBCV)	11:26 AM	Diverted	NA	
46	FD849	18/07/2022	PC12	Darwin Int'l / RAAF (DRW / YPDN)	Alice Springs (ASP / YBAS)	12:56 PM	Diverted	NA	
47	FD493	14/07/2022	BE20	Kingaroy (KGY / YKRY)	Toowoomba (YTWB)	11:42 PM	12:04 AM	22	
48	FD623	14/07/2022	PC12	Jandakot (JAD / YPJT)	Broome Int'l (BME / YBRM)	7:07 PM	11:16 PM	249	

After deletion:

2	FD493	14/07/2022	BE20	Kingaroy (KGY / YKRY)	Toowoomba (YTWB)	11:42 PM	12:04 AM	22	
3	FD623	14/07/2022	PC12	Jandakot (JAD / YPJT)	Broome Int'l (BME / YBRM)	7:07 PM	11:16 PM	249	
4	FD607	13/07/2022	PC12	Broome Int'l (BME / YBRM)	Jandakot (JAD / YPJT)	9:01 PM	1:02 AM	241	
5	FD629	15/07/2022	PC12	Jandakot (JAD / YPJT)	Broome Int'l (BME / YBRM)	6:25 PM	10:05 PM	220	
6	FD480	5/07/2022	B350	Brisbane (BNE / YBBN)	Townsville Int'l / RAAF (TSV / YBTL)	5:24 PM	8:52 PM	208	

2. Data exploration

i) Answer to Q1:

Usage Of Aircraft In July 2022

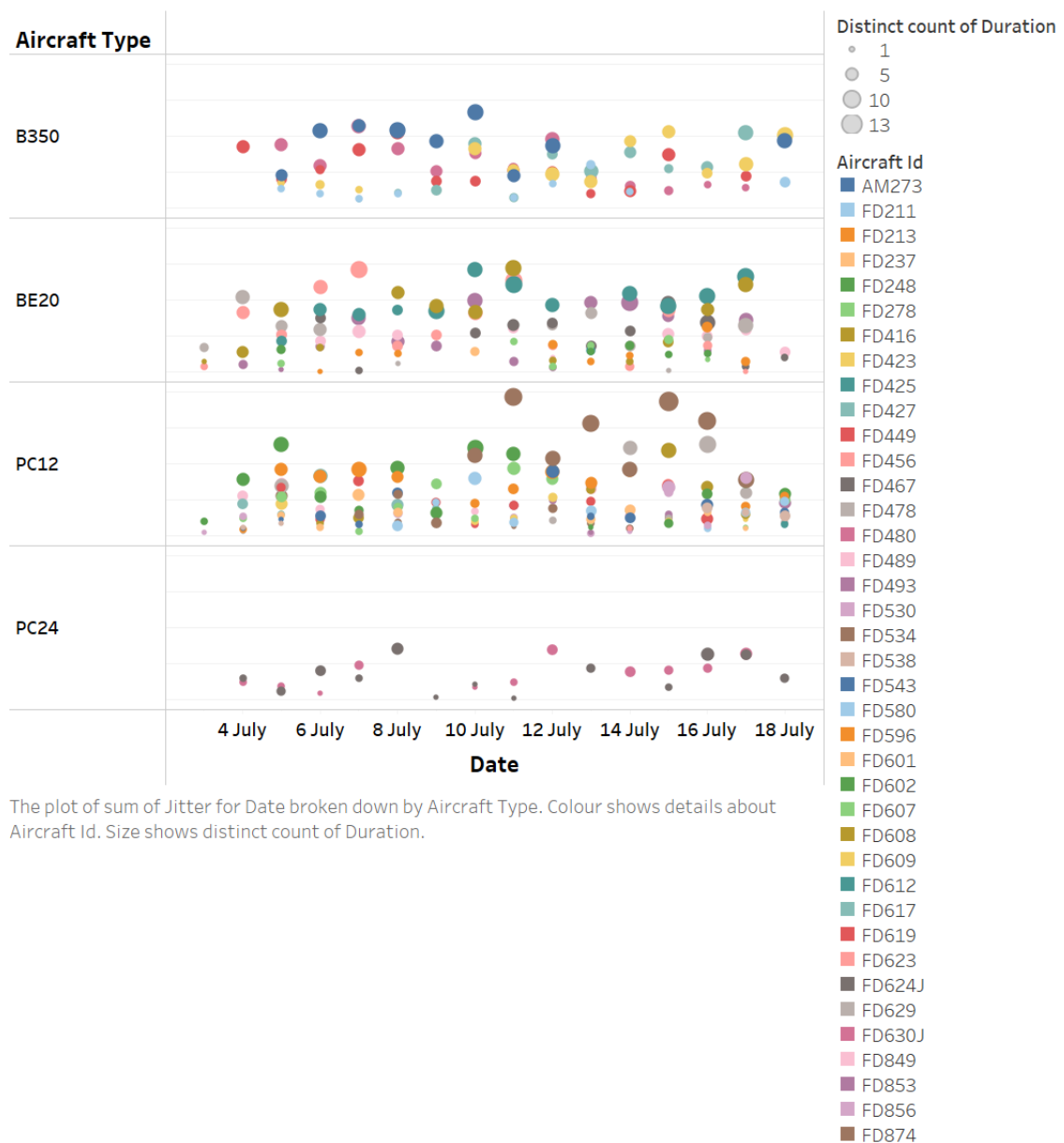


Figure 5. Frequency of usage of aircraft in July 2022. The size of filled circle represents the number of flights (count of duration) and the colour represents individual aircraft.

As shown in **Figure 5**, in July 2022, aircrafts of type PC12 and PC24 were the most and least used aircraft types, respectively, in terms of number of flights flown. Heavy usage occurred between July 10 and July 16 for PC12. B350 were flown more frequently between July 6 and July 12. As for BE20, the frequency of usage was similar through the period, except for July 3 and July 18 when the usage was much lower.

This type of visualisation was chosen to answer Q1 because the usage frequency of each aircraft type on each day could be clearly visualised by the size of the filled circle across the time period of interest.

Visualising this data on both an hourly and day-of-the week timescale will certainly show that the RFDS is in the air 24/7 because these timescale are cyclic within the time period that we are looking at. For instance, it is highly likely that RFDS is in the air, for instance at 1pm across this 16-day period.

ii) Answer to Q2:

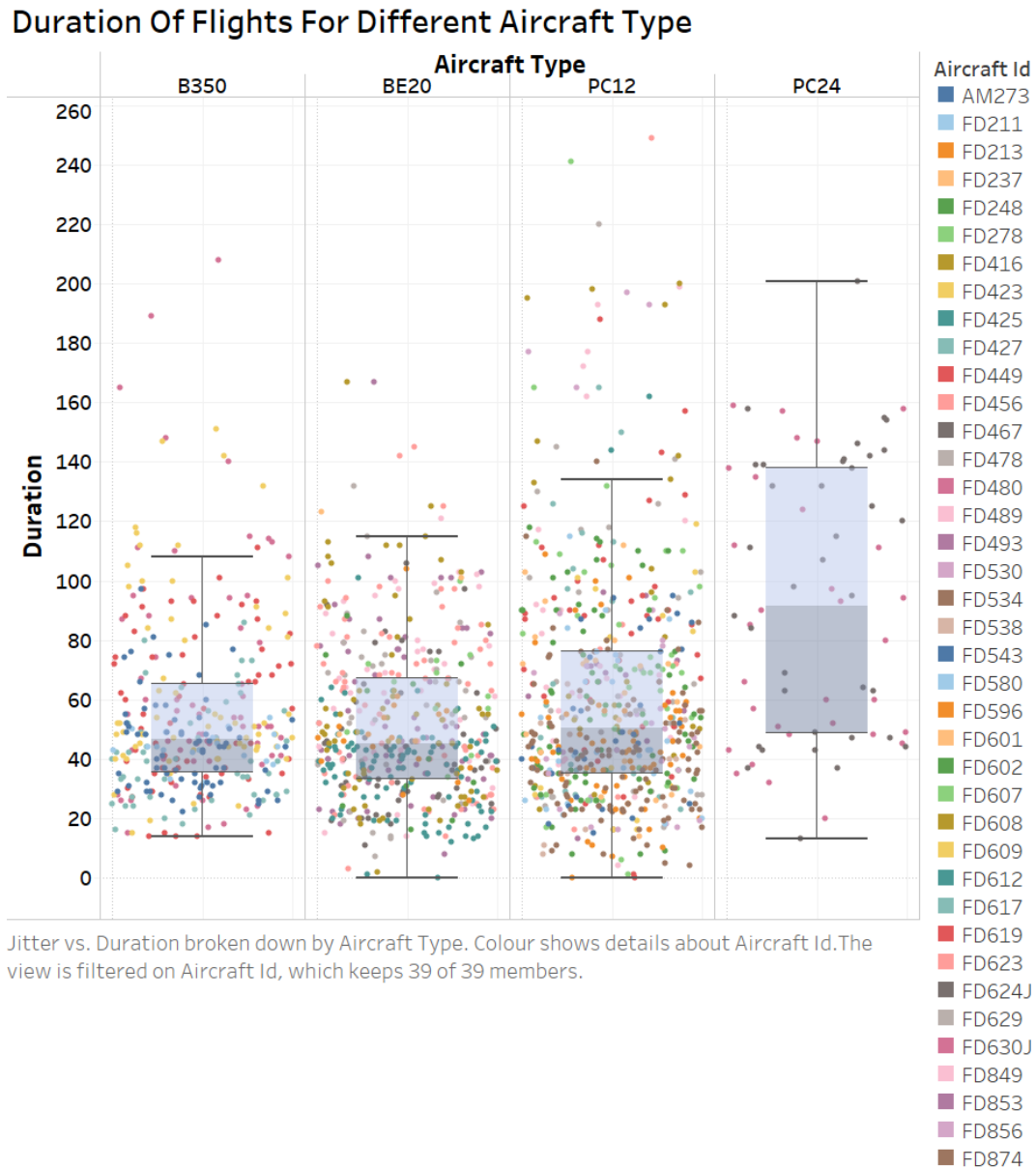


Figure 6. Duration of flight for different aircraft types. The colour represents individual aircraft.

Figure 6 shows that aircrafts of types B350, BE20, and PC12 have similar flight duration medians, whereas PC24 has relatively longer flight duration median. In term of number of flights, the visual observation of the highest density of coloured circles for PC12 among all aircraft types showed that PC12 was flown the highest number of times followed by BE20, B350, and lastly PC24. Although aircrafts of type PC24 may be doing less number of flights (as shown in **Figure 5**), that does not mean that they have lower usages as they were spending longer time in the air per flight. This shows that Figure 5 alone is inadequate to conclude that PC24 was the least used aircraft type.

Box plot was chosen to answer Q2 because of its ability to provide summary statistics about the flight duration for each aircraft type. Also, the flight duration of each aircraft of each aircraft type can be clearly visualised through this type of visualisation.