INTRODUCTION

In this project, based on the dataset named "Flight Delay and Causes", a data analysis and data visualization study was carried out on Power BI with the 6-month USA flight data of 2019 obtained from this data set. The dataset and Power BI reports are detailed below.

1. DATASET

The "Flight Delay and Causes" dataset was obtained from Kaggle and includes many data about flights within the USA in the 6-month period between January-June 2019, such as date, airline company, airports and IATA codes, and various reasons for delays.

Source of Dataset: **www.kaggle.com/undersc0re/flight-delay-and-causes**

The columns included in the study from the columns in the data set are as follows:

Date: Scheduled flight date

DepTime: Actual departure time

ArrTime: Actual arrive time

CRSArrTime: Scheduled arrival time

UniqueCarrier: Unique carrier (vehicle, aircraft) code

Airline: Airline company

FlightNum: Flight number

TailNum: Aircraft tail number

ActualElapsedTime: Time spent in the air, including landing and take-off time (in minutes)

CRSElapsedTime: Estimated elapse time of flight (in minutes)

AirTime: Flight time (in minutes)

ArrDelay: Difference in minutes between scheduled and actual arrival time

Origin: 3-letter IATA (International Air Transport Association) code of departure airport

Org_Airport: Departure airport name

Dest: 3-letter IATA (International Air Transport Association) code of arrival airport

Dest_Airport: Arrival airport name

Distance: Distance between departure and arrival airports (mile),

TaxiIn: Minutes from the wheels down to the moment of arrival at the airport gate

TaxiOut: Departure from the origin airport gate and wheels off, in minutes

CarrierDelay: Flight delay due to carrier (maintenance or crew problems, aircraft cleaning, fueling, etc.) (in minutes)

WeatherDelay: Flight delay due to weather (in minutes)

NASDelay: Flight delay by NAS (National Aviation System) (in minutes)

SecurityDelay: Flight delay time due to security (in minutes)

LateAircraft: Flight delay time due to late arriving aircraft (in minutes).

Apart from these, the data columns added to the dataset within the scope of the study are as follows:

WeekDay: The day of the week on which the flight took place

WeekDayNumber: Number value of the day of flight according to the days of the week (Monday=1, Sunday=7)

In addition to the "Flight Delay and Causes" dataset, a dataset covering more than 10000 international airports, ferry terminals and train stations from OpenFlights.org - "OpenFlights Airports Database" was used in the study to be used in data visualizations on the map. Only the "Airport Database" was used from this data set, and the appropriate columns were selected for the study. By filtering the data, only airports located in the USA were included in the study.

Source of Dataset: **openflights.org/data.html**

The columns included in the study from the columns in the data set are as follows:

Name: Airport name

City: City where the airport is located

Country: Country or region where the airport is located

IATA: 3-letter IATA code of the airport

Latitude: The latitude of the airport,

Longitude: The longitude of the airport.

2. DATASET STATISTICS

The “Flight Delay and Causes” dataset focuses on the types and times of delays in flights. Some statistical values such as average, mode, median, kurtosis, standard deviation of the columns in which these delay types and times are given are given below.

2.1. Delay Types Fields

Statistical values of delay types and times fields in the data set are given in the tables below.

Table 1. Statistical Values of Delay Types Fields

	ArrDelay	DepDelay	CarrierDelay	Weather Delay	NASDelay	Security Delay	LateAircraft Delay
Average	60,79	57,38	17,20,	3,15	13,65	0,08	26,7
Mode	15	25	0	0	0	0	0
Median	42	40	2	0	1	0	13
Kurtosis	34,05621	36,0662	109,1286	285,2076	59,8380	5753,248	15,4653
Standard Deviation	56,77231	55,7602	39,0169	19,5018	31,5539	1,8793	40,4709

2.2. Other Numeric Fields

Statistical values of other numeric fields in the data set are given in the tables below.

Table 2. Statistical Values of Other Numerical Fields

	ActualElapsedTime	CRSElapsedTime	AirTime	TaxiIn	TaxiOut
Average	134,41	131	108,49	6,78	19,15
Mode	75	65	45	4	11
Median	116	113	89	5	15
Kurtosis	2,3435	2,3813	2,5080	76,0584	31,9833
Standard Deviation	73,8435	71,3038	69,8824	5,5554	15,3482

3. POWER BI REPORT

After the data sets were organized, data visualization was made on Power BI.

3.1. Weekday/End Flight Intensities

On the Weekday/Weekend Flight Intensities page, the total number of flights on weekdays and weekends in the 6-month period covered by the data set was shown on the bar chart according to the days of the week. The ratio of the total count of weekday and weekend flights was shown on the pie chart.

The intensity of the weekday and weekend flights according to the destinations was shown on the map. From this section, it is possible to examine which destinations are mostly flown in which part of the week.

With the slicers added to the report, these charts can be viewed by filtering them by date and airline companies.

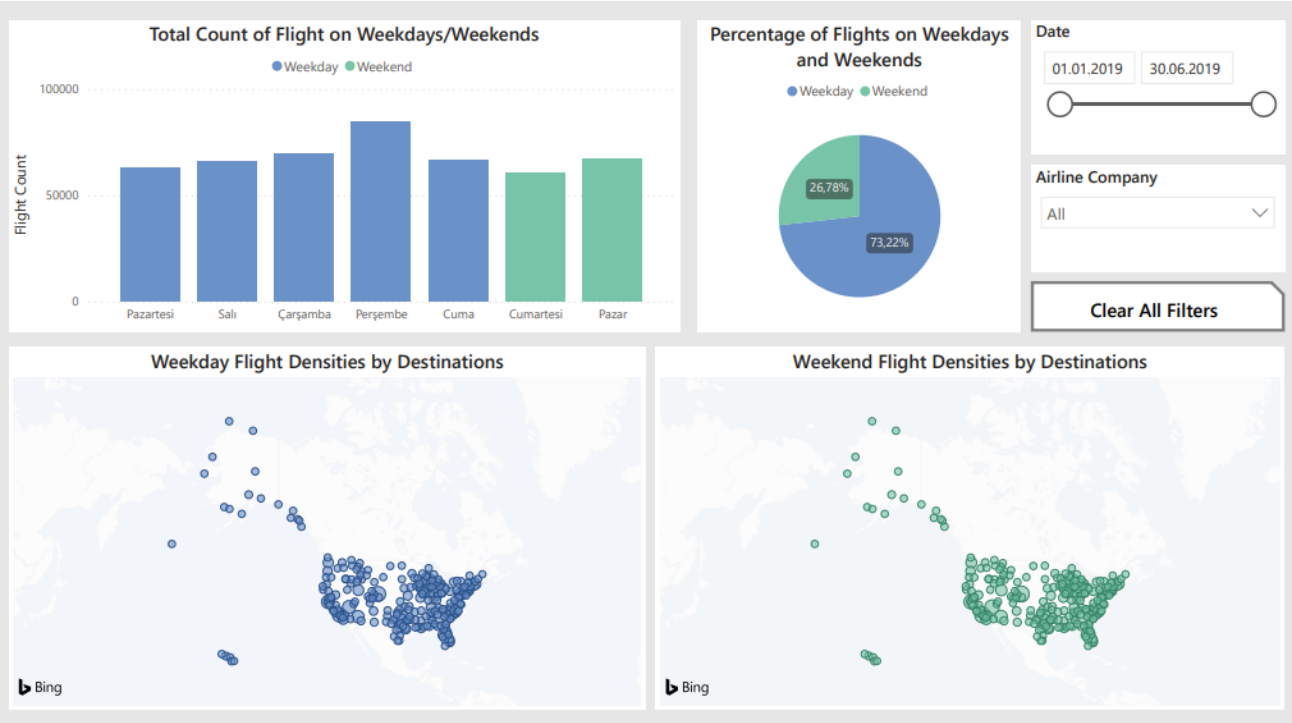

Figure 1. Weekday/Weekend Flight Intensities Page

3.2. Flight Intensities with Hourly Intervals

On this page, the departure and arrival times of all 6-month flight data are grouped in 6-hour periods and analyzed in this way. These time periods are as follows: 00:00-05:59, 06:00-11:59, 12:00-17:59, 18:00-23:59.

The departure time of the aircraft, departure point, destination and the number of flights on this route were listed and shown on the table. With this table, it is aimed to obtain information about which route has more flights in which time period, and in which time period this route is more intense. In addition, it is possible to list with the time period filter with the slicer added only depending on this table.

At the bottom of the page, the number of flights according to the time period of the departure time of the flights and the number of flights according to the time period of the arrival time are given with column graphics. In this way, it can be observed which time period is more intense for departure and arrival.

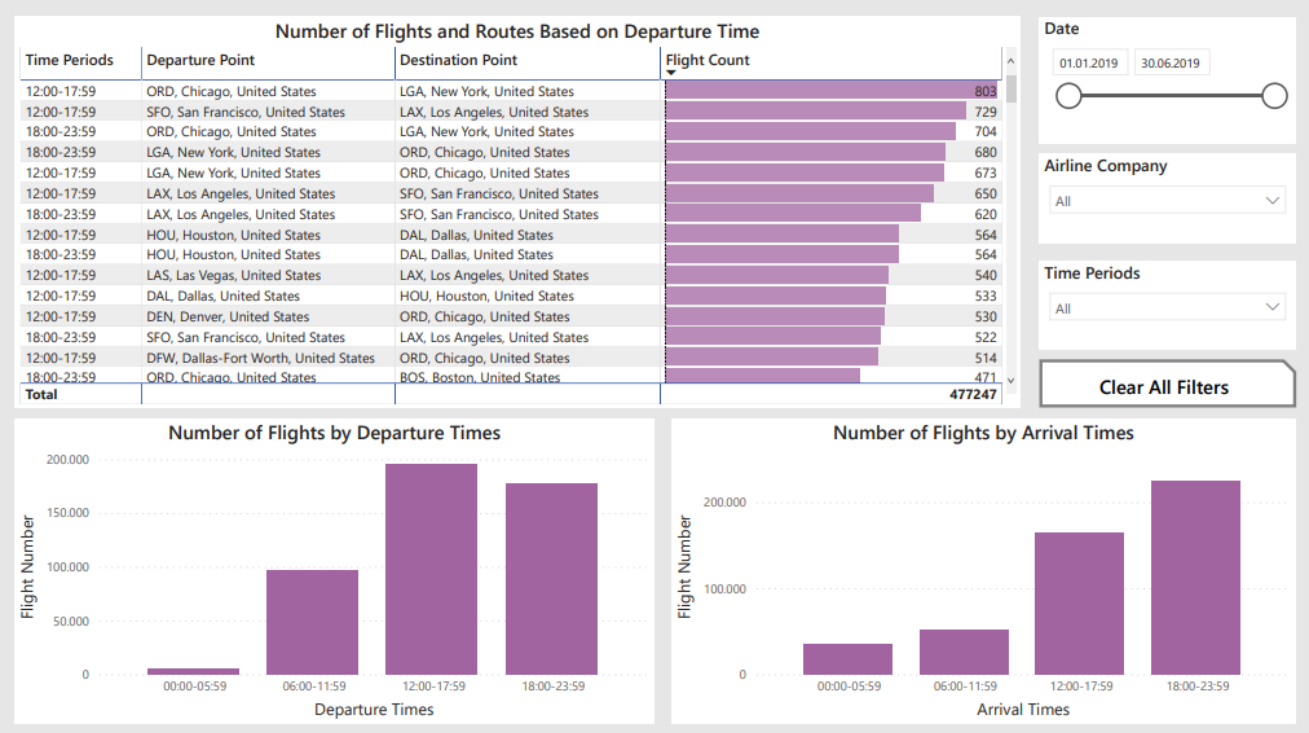

Figure 2. Flight Intensities with Hourly Intervals Page

3.3. Total Delay Times Analysis

On this page of the report, the total delay times are examined according to different criteria.

The line chart titled “Total Delay Times by Dates” is shown departure time delays and arrival time delays relative to the scheduled arrival time for each month in a 6-month period. Similarly, the total delay times according to the days of the week for the same fields are shown in the line chart titled "Total Delay Times by Days of the Week". With these graphs, it is understood in which months and on which days of the week there are more delays, and how many delays occur. In addition, the relationship between the delay times at the time of departure and the delay times at the time of arrival can also be observed.

In the last chart, the total delay times of all flights of the companies are shown according to the airline companies.

Figure 3. Total Delay Times Analysis Page

3.4. Delay Reasons and Durations

In the dataset, the reasons that cause the most delays and how many minutes they cause delays for each flight are included. On this page, flights are examined from this perspective.

According to the data in the data set in the report, the ratios of the total delay times caused by five delay reasons to each other are given as percentages on the donut chart. The "Security Latency" in this chart has a very small share of 0.13% compared to other reasons. Therefore, this delay reason was not included in the values in the other graphs.

The reasons for the delays in the flights in the 6-month period and the total number of minutes of delay are shown in the bar chart. Based on these graphs, it can be observed which delay occurs more frequently and how long it lasts.

In addition, with the line graph showing the distribution of delay times according to the days of the week, it is possible to track how much delays occur on which days of the week.

A screenshot of the page is given in Figure 4. It is seen in the graphs that “Delay Due to Late Arriving Aircraft” has the highest value. In this context, this reason for delay is examined in detail on the next page.

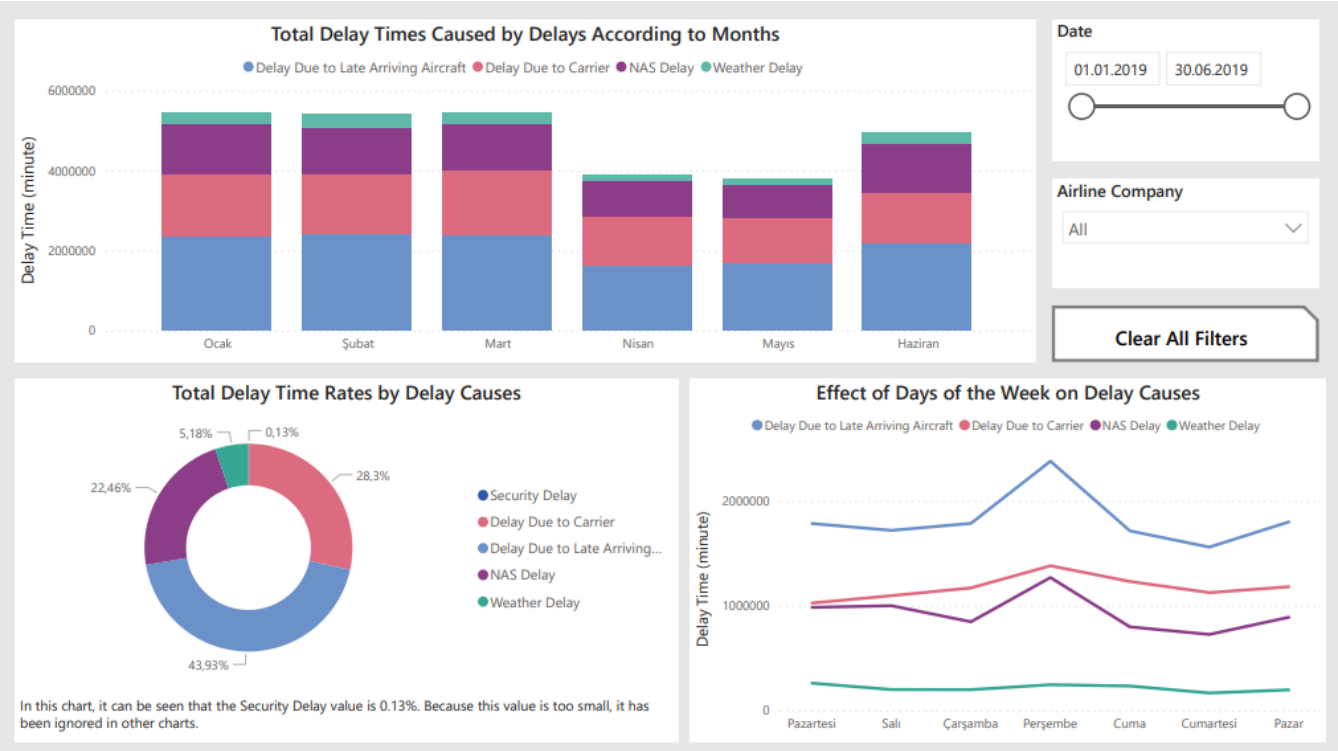

Figure 4. Delay Reasons and Durations Page

3.6. Delay Analysis Due to Late Arriving Aircraft

From the previous analyzes, it was seen that “Delay Due to Late Arriving Aircraft” had high values. This section analyzes this reason.

The total delay times according to the departure times are given on the line graph. In this way, it can be observed at which hours there are more delays. It is also shown on the map to which destinations there are more delays in flights.

With the added slicers, a more detailed examination can be made by filtering the date and airline companies.

Figure 5. Delay Analysis Due to Late Arriving Aircraft Page