Data Analysis on Airline On-Time Statistics and Delay Causes Using Pyspark and Spark SQL

Ahsan Mushtaq

P2688331

Abstract

An analysis of Airline On-Time Statistics and Delay Causes provided by U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics (BTS). In this assignment I will analyse the different factors and what elements causing the airline delays and which airports are affected by the delays. How many delays will happen by the specific carrier and what are the causes behind them like Whether it is due to bad weather conditions, or its due to NAS national aviation system or due to any security issues. What is number of cancelled flights and how many flights are diverted. When a flight arrives 15 minutes or later, it is called delayed Only delayed flights are given delayed minutes. When numerous reasons are assigned to a single delayed flight, each cause is divided depending on the number of delayed minutes for which it is responsible. The figures presented are rounded and may not add up to the total [2]. Finding percentages and doing visualisation. These insights are illustrating using visualization. In 2020/2021, the coronavirus pandemic and its accompanying travel restrictions had an impact on aviation industry. This report looks into the impact on reported stats of delays and shows monthly trends in 2021 by using Pyspark and Spark SQL.

Table of Contents

Table of Figure

Introduction to the Data

Data source and License

In the Air Travel Consumer Report, the US Department of Transportation publishes a monthly overview of airline on-time performance, including causes of delay. The Bureau of Transportation Statistics releases data on flight delays and on-time arrivals. It is open government data with a public information licence [1]. We may freely and flexibly utilise and re-use the Information to gain beneficial insights from it, as long as we follow a few terms and conditions. The Office of Aviation Consumer Protection of the Department of Transportation publishes the Air Travel Consumer Report every month. The purpose of the study is to provide consumers with information on the quality of airline services. On April 28, 2022, the most current report was released. We have the right to copy, publish, distribute, and transmit the Information. We can modify the data and gain important insights. We can commercially and non-commercially utilise the Info data, for example, by integrating it with other Information or putting it in your own product, system, or application.

The statistics used in this report can be accessed from(“https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp”). This analysis can be used to get useful decisions that which are the reasons for delays in the flights and make better policies for better airline management and regulations.

Data Format

The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics (BTS) released monthly Excel and txt data files on his public Gov website which can be accessed. These excel files cannot have any personal information about any passenger or any staff member that they are the reason for this happening. This excel file contain categorical data and fields like year, month, carrier, carrier name, airport, airport name, arrived flights, cancelled flights and diverted flights, weather delay, NAS delay, security delay etc.

Data Scope

The data in this report covers the whole 2021 data from January to December every month. But this data only involves which is reported by the U.S department of transportation. There should be very small number of anomalies.

Summary data on the number of on-time, late, cancelled, and diverted flights appears in the DOT's (Department of transportation) monthly Air Travel Consumer Report, which is issued roughly 30 days after the month's end, as well as summary tables available on this page. In June 2003, BTS (Bureau of Transportation Statistics) began collecting data on the causes of aircraft delays. When the Air Travel Consumer Report is produced, summary statistics and raw data are made accessible to the public.

Information of the fields

Field	Description
Year	The calendar year. This report is limited to 2021
Month	The calendar month (Jan to Dec)
Carrier	Unique airline codes
carrier_name	Unique airline full names
Airport	Unique airport code
Airport_name	Full Airport name or location
Arr_flights	Number of flights arrived on time for specific carrier
Arr_del15	Number of flights arrived 15 minutes late
Arr_cancelled	Number of flights cancelled
Arr_diverted	Number of flights diverted
Arr_delay	Arrived flights which are not on time
Carrier_delay	Flights delayed due to carriers
Weather_delay	Flights delayed due to bad weather conditions
Security_delay	Flights delayed due to security chaos
Nas_delay	Delay in flights due to National aviation Sy

Methodology

Files Uploading

One excel files Airline_Delay_Cause is added to HDFS Web console. In Jupyter new spark context is created and important libraries are imported. RDDs are created from the uploaded file using airline delay cause excel file. We are taking 2021 data here from January to December.

Data cleansing and pre prossessing

After developing rdd apply some functions like count(), first(), take() and then delete header. A new rdd is created and header were deleted.

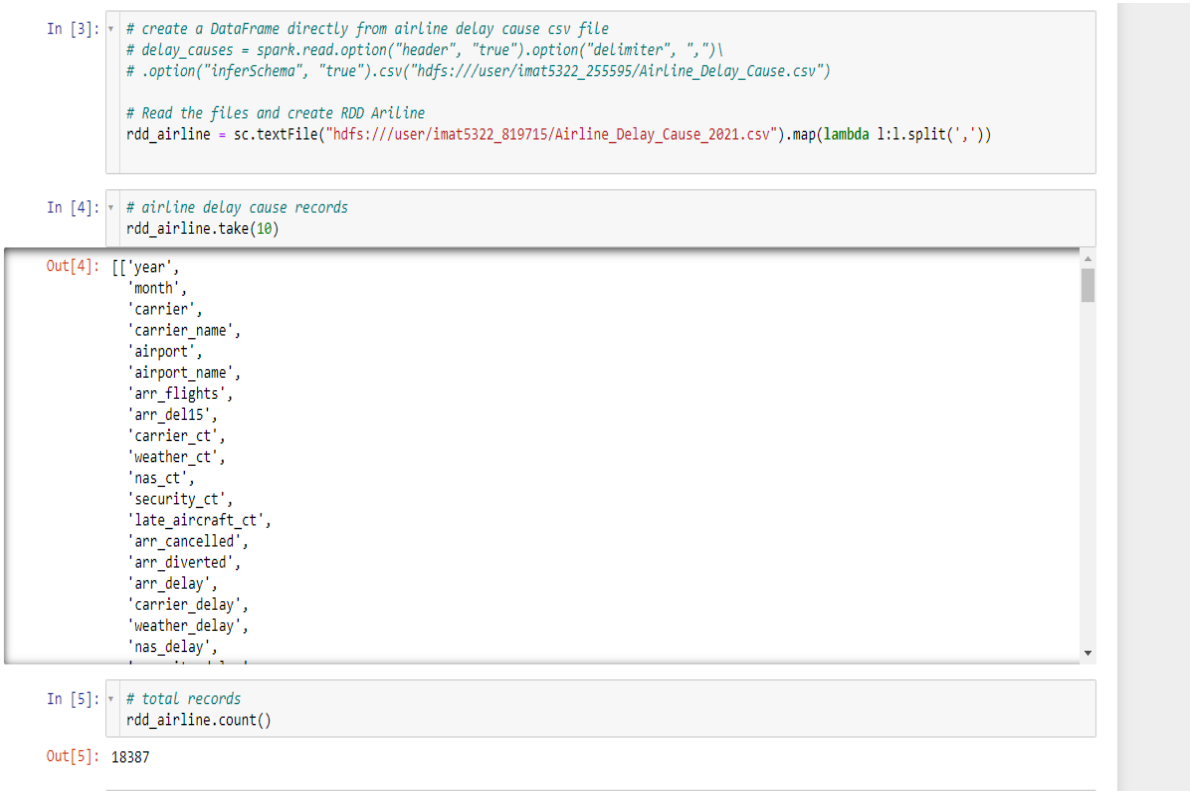

Figure 10: importing file

In figure you can see first we remove the header and then map the column. We can see from figure 1 that there 18387 records first and after removing header there is 18386 records showing in figure 4.

```
In [6]: # identifies header and store in new variable
        header = rdd_airline.first()
        display(header)

In [7]: # filter data without headers
        # new rdd here is filtered with data without the headers
        filtered_without_header = rdd_airline.filter(lambda line:line!=header)
        filtered_without_header.first()

Out[7]: ['2021',
         '12',
         '9E',
         'Endeavor Air Inc.',
         'ABE',
         '"Allentown/Bethlehem/Easton',
         ' PA: Lehigh Valley International"',
         '127.00',
         '9.00',
         '2.51',
         '0.00',
         '3.54',
         '0.00',
         '2.95',
         '0.00',
         '0.00',
         '264.00',
         '119.00',
         '0.00',

In [8]: # preprocessing the data
        # Mapping column names to row entries
        column_updates = filtered_without_header.map(lambda x:Row(year=x[0] ,month=x[1], \
                                            carrier=x[2],carrier_name=x[3],airport=x[4],airport_name=x[5],\
                                            arr_flights=x[6],  arr_del15=x[7], carrier_ct=x[8], weather_ct=x[9],\
                                            nas_ct=x[10], security_ct=x[11], late_aircraft_ct=x[12],\
                                            arr_cancelled=x[13], arr_diverted=x[14], arr_delay=x[15],\
                                               carrier_delay=x[16], weather_delay=x[17], nas_delay=x[18],\
                                               security_delay=x[19], late_aircraft_delay=x[20]))
        column_updates.take(2)

Out[8]: [Row(airport='ABE', airport_name='"Allentown/Bethlehem/Easton', arr_cancelled='2.95', arr_del15='127.00', arr_delay='0.00', arr
        _diverted='0.00', arr_flights=' PA: Lehigh Valley International"', carrier='9E', carrier_ct='9.00', carrier_delay='264.00', car
        rier_name='Endeavor Air Inc.', late_aircraft_ct='0.00', late_aircraft_delay='0.00', month='12', nas_ct='0.00', nas_delay='0.0
        0', security_ct='3.54', security_delay='68.00', weather_ct='2.51', weather_delay='119.00', year='2021'),
```

Figure 11: Removing header and mapping column

Creating Data Frame

```
In [9]: # creating dataframe using sqlContext
        airline_results = sqlContext.createDataFrame(column_updates)

        # checking schema
        airline_results.printSchema()

root
 |-- airport: string (nullable = true)
 |-- airport_name: string (nullable = true)
 |-- arr_cancelled: string (nullable = true)
 |-- arr_del15: string (nullable = true)
 |-- arr_delay: string (nullable = true)
 |-- arr_diverted: string (nullable = true)
 |-- arr_flights: string (nullable = true)
 |-- carrier: string (nullable = true)
 |-- carrier_ct: string (nullable = true)
 |-- carrier_delay: string (nullable = true)
 |-- carrier_name: string (nullable = true)
 |-- late_aircraft_ct: string (nullable = true)
 |-- late_aircraft_delay: string (nullable = true)
 |-- month: string (nullable = true)
 |-- nas_ct: string (nullable = true)
 |-- nas_delay: string (nullable = true)
 |-- security_ct: string (nullable = true)
 |-- security_delay: string (nullable = true)

In [10]: # printing the rows to check the records structure
         airline_results.show(2)

+-------+--------------------+-------------+---------+---------+------------+--------------------+-------+----------+----------
---+----------------+----------------+-------------------+-----+------+---------+-----------+--------------+----------+-------
------+----+
|airport|        airport_name|arr_cancelled|arr_del15|arr_delay|arr_diverted|         arr_flights|carrier|carrier_ct|carrier_de
lay|     carrier_name|late_aircraft_ct|late_aircraft_delay|month|nas_ct|nas_delay|security_ct|security_delay|weather_ct|weather
_delay|year|
+-------+--------------------+-------------+---------+---------+------------+--------------------+-------+----------+----------
---+----------------+----------------+-------------------+-----+------+---------+-----------+--------------+----------+-------
------+----+
|    ABE|"Allentown/Bethle...|         2.95|   127.00|     0.00|        0.00| PA: Lehigh Valle...|     9E|      9.00|       26
4.00|Endeavor Air Inc.|            0.00|               0.00|   12|  0.00|     0.00|       3.54|         68.00|      2.51|
119.00|2021|
|    ABY|            "Albany|         1.87|    73.00|     0.00|        0.00| GA: Southwest Ge...|     9E|     11.00|       34
2.00|Endeavor Air Inc.|            0.00|               0.00|   12|  0.00|     0.00|       4.11|        142.00|      5.02|
125.00|2021|
+-------+--------------------+-------------+---------+---------+------------+--------------------+-------+----------+----------
```

Figure 12: Creating Data frame

Data Cleaning using Drop, Changing data type

All data is in string so we have to change its data type so that we can use its column in visualization**.** In the next step we drop some column which are not necessary for example, carrier ct, nas ct, weather ct, late aircraft ct. Then shows schema using pandas for better data table representation (figure 5).

```
In [11]: # Convert datatype from (string to float) for the fields required.
         type_changes = airline_results.withColumn("arr_cancelled",airline_results["arr_cancelled"].cast(FloatType()))
         type_changes = type_changes.withColumn("arr_del15",type_changes["arr_del15"].cast(FloatType()))
         type_changes = type_changes.withColumn("arr_delay",type_changes["arr_delay"].cast(FloatType()))
         type_changes = type_changes.withColumn("arr_diverted",type_changes["arr_diverted"].cast(FloatType()))
         type_changes = type_changes.withColumn("carrier_ct",type_changes["carrier_ct"].cast(FloatType()))
         type_changes = type_changes.withColumn("carrier_delay",type_changes["carrier_delay"].cast(FloatType()))
         type_changes = type_changes.withColumn("late_aircraft_ct",type_changes["late_aircraft_ct"].cast(FloatType()))
         type_changes = type_changes.withColumn("late_aircraft_delay",type_changes["late_aircraft_delay"].cast(FloatType()))
         type_changes = type_changes.withColumn("nas_ct",type_changes["nas_ct"].cast(FloatType()))
         type_changes = type_changes.withColumn("nas_delay",type_changes["nas_delay"].cast(FloatType()))
         type_changes = type_changes.withColumn("security_ct",type_changes["security_ct"].cast(FloatType()))
         type_changes = type_changes.withColumn("security_delay",type_changes["security_delay"].cast(FloatType()))
         type_changes = type_changes.withColumn("weather_ct",type_changes["weather_ct"].cast(FloatType()))
         type_changes = type_changes.withColumn("weather_delay",type_changes["weather_delay"].cast(FloatType()))

         # Convert datatype from (string to int) for the fields required.
         type_changes = type_changes.withColumn("month",type_changes["month"].cast(IntegerType()))
         airline_data = type_changes.withColumn("year",type_changes["year"].cast(IntegerType()))

         # printing the new schema
         airline_data.printSchema()

In [12]: # Total airline records
         airline_data.count()

Out[12]: 18386

In [13]: # here we are droping the columns, as these are not useful.
         airline_data = airline_data.drop("carrier_ct", "late_aircraft_ct","nas_ct","security_ct","weather_ct")

         # printing the updated schema
         airline_data.printSchema()

root
 |-- airport: string (nullable = true)
 |-- airport_name: string (nullable = true)
 |-- arr_cancelled: float (nullable = true)
 |-- arr_del15: float (nullable = true)
 |-- arr_delay: float (nullable = true)
 |-- arr_diverted: float (nullable = true)
 |-- arr_flights: string (nullable = true)
 |-- carrier: string (nullable = true)
```

Figure 13: Data Cleansing

Data Analysis by using Spark SQL and Visualization

Spark SQL:

Register a data frame airline_data as a table and apply select Query to fetch all the data from excel file. Using pandas library for table representation.

```
In [14]: # Register the DataFrame as a table
         sqlContext.registerDataFrameAsTable(airline_data, "airline_table")
```

```
In [37]: # checking if the data loads for sql query
         sqlContext.sql("select * from airline_table").take(5)
         pdpyr= airline_data.toPandas()
         pdpyr
```

Out[37]:

	airport	airport_name	arr_cancelled	arr_del15	arr_delay	arr_diverted	arr_flights	carrier	carrier_delay	carrier_name	late_aircraft_del
0	ABE	"Allentown/Bethlehem/Easton	2.95	127.0	0.0	0.0	PA: Lehigh Valley International"	9E	264.0	Endeavor Air Inc.	0
1	ABY	"Albany	1.87	73.0	0.0	0.0	GA: Southwest Georgia Regional"	9E	342.0	Endeavor Air Inc.	0
2	AEX	"Alexandria	1.15	62.0	0.0	0.0	LA: Alexandria International"	9E	439.0	Endeavor Air Inc.	0
3	AGS	"Augusta	2.74	166.0	1.0	0.0	GA: Augusta Regional at Bush Field"	9E	1266.0	Endeavor Air Inc.	0
4	ALB	"Albany	0.82	52.0	0.0	0.0	NY: Albany International"	9E	497.0	Endeavor Air Inc.	0

Figure 14: Registering the data frame as table and showing data in PANDA

```
In [36]: # Total types of unique carriers
         sqlContext.sql("select distinct carrier_name from airline_table").count()
Out[36]: 17

In [17]: # Total unique airports
         sqlContext.sql("select distinct airport from airline_table").count()
Out[17]: 371

In [18]: # Total operation performed by certain carrier, presented in descending order
         sqlContext.sql("select carrier, count(carrier) as carrier_op from airline_table group by carrier order by \
         carrier_op desc \
         ").show()
```

```
+-------+----------+
|carrier|carrier_op|
+-------+----------+
|     OO|      2531|
|     MQ|      1555|
|     G4|      1382|
|     DL|      1313|
|     9E|      1232|
|     AA|      1134|
|     WN|      1132|
|     F9|      1093|
|     YV|      1082|
|     UA|      1076|
|     OH|      1023|
|     YX|       989|
|     AS|       829|
|     B6|       673|
|     NK|       569|
|     QX|       542|
|     HA|       231|
+-------+----------+
```

```
In [19]: # Total flight operations
         sqlContext.sql("select * from airline_table").count()
Out[19]: 18386
```

Figure 15: Select query for fetch data

Average delay by each carrier:

```
In [20]: # Average carrier delay
sqlContext.sql("select carrier, avg(carrier_delay) as delays from airline_table \
where carrier_delay > 0.0 group by carrier order by delays desc").show(20)
```

```
+-------+------------------+
|carrier|            delays|
+-------+------------------+
|     WN|10295.718197879858|
|     AA| 8072.110412926391|
|     B6| 5588.079510703364|
|     UA|  4684.64320625611|
|     NK|  4349.02495543672|
|     DL| 4205.476190476191|
|     OO| 3849.486475409836|
|     YX| 3315.818994413408|
|     YV|  2563.53227408143|
|     OH| 2084.424120603015|
|     F9|          2007.349|
|     AS|1845.4074074074074|
|     G4|1630.4778694673669|
|     MQ|1578.0712328767124|
|     9E|1289.0008561643835|
|     QX|           1262.76|
```

Figure 16:Average delay by each carrier

```
display(airline_data)
```

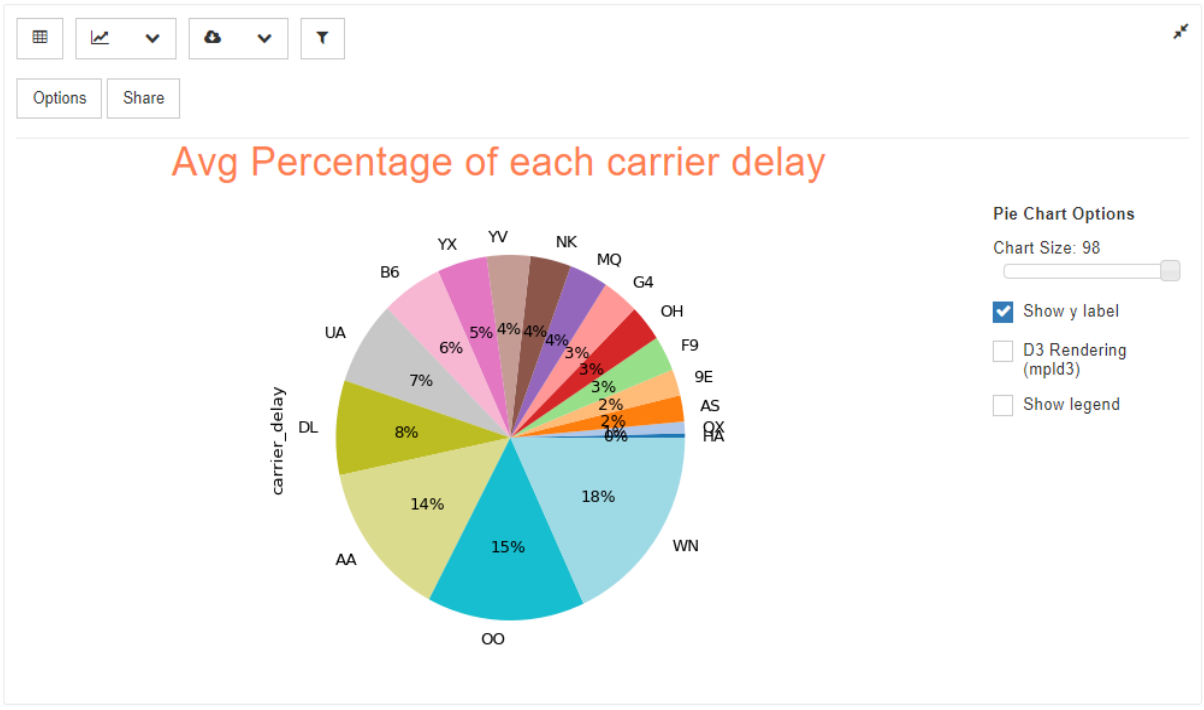

Figure 17: Avg % of flight delay by carrier

Above pie chart illustrates that SkyWest airline (OO) has 15 percent delays in their flights, 18 percent flights of Southwest Airlines (WN) are delayed, 14 percent flights of American airline (AA) are delayed, 8 percent flights of delta airline (DL) are delayed. Percentages of others flights can be seen from pie chart.

If a flight arrives at (or departs) the gate, 15 minutes or more after the scheduled arrival (departure) time as recorded in the Computerized Reservation System, it is considered as delayed.

The Air Carrier On-Time Reporting Advisory Committee defined broad categories for airlines to disclose delay causes. Air Carrier, delay due to National Aviation System, delay due to Weather, Late-Arriving Aircraft, and delay due to Security threats are the categories. The reasons for cancellation are the same, with the exception that there is no category for late-arriving aircraft. We will discuss these reasons or causes in our analysis below. Ground staff being late with aircraft refuelling, the airport trying to handle with the sheer amount of passengers or needing to wait for the pilots, or inclement weather are all possible causes of delays.

Security delay with respect to airport and carrier

Delays or cancellations caused by a terminal or concourse evacuation, re-boarding of aircraft due to a security breach, malfunctioning screening equipment, and/or large line-ups of more than 29 minutes at screening locations.

Figure below illustrates that DEN Airport has delays due to security i.e. 6859.93 and DFW has 6423.85 so on. It presents in descending order.

Following the terrorist attacks on September 11, 2001, in USA airport security systems underwent significant improvements. Airlines, for example, advised travellers to arrive at airports up to two hours before a domestic flight's departure. Passengers were randomly picked for further screening in the boarding area after passing through security checks, which included a hand search of their carry-on luggage. After a passenger attempted to ignite a bomb in his shoe during a flight in December 2001, security screeners instructed passengers to remove their shoes when going through checkpoints [4].

```
In [24]: # Average security delay with respect to airports
         sqlContext.sql("select airport, avg(security_delay) as delays from airline_table \
         where security_delay > 0.0 group by airport order by delays desc").show(20)
```

```
+-------+------------------+
|airport|            delays|
+-------+------------------+
|    DEN| 6859.934579439252|
|    DFW| 6423.850393700787|
|    ORD| 5018.810218978102|
|    DAL| 4947.260869565217|
|    MCO|          4164.375|
|    EWR|3684.2678571428573|
|    FLL|3210.9887640449438|
|    SFB| 2984.909090909091|
|    JFK|2730.2368421052633|
|    LAS|2601.9327731092435|
|    IAH|       2572.703125|
|    ATL| 2502.909090909091|
|    PIE| 2473.181818181818|
|    SEA|2371.5363636363636|
|    PGD| 2287.181818181818|
|    MIA|2193.0315789473684|
|    CLT|  2104.85593220339|
|    LAX|            1985.7|
|    HOU|1938.6888888888889|
|    MDW|1840.1914893617022|
+-------+------------------+
only showing top 20 rows
```

Figure 18:Security delay with respect to airport

In the below bar chart we can see that the carrier Southwest Airlines (WN)has the most delays regarding security and American Airlines (AA) has second most delays due to security. Besides Hawaiian Airlines (HA) and Horizon Air (QX) have least delays due to security.

Figure 19: Security delay with respect to carrier

Average NAS delay with respect to airports

Delays and cancellations caused by the national aviation system, which include non-extreme weather, airport operations, large traffic volume, and air traffic management. There are four main reasons that flights are delayed due to National aviation system first of all volume that flights are in huge volume that making arrangement for them is getting difficult and causing lateness in flights. Approximately 45.96% flights are delayed due to huge volume. And when the weather is not suitable then there is more strict instruction to flight pilot and other staff and more measurement can be taken for example they have to be wait till the weather becomes normal NAS instructed the pilot and other staff that you have to wait until the weather becomes normal. Almost 42.71% flights are delayed because of this weather conditions. If there is any fault in equipment then its also considered as NAS delay and flights are also late because of closed runway 4.63% flights are late due to close Runway [5]. The figure below illustrates that the airport DIK (Dickinson-Theodore Roosevelt Regional Airport) has more delayed flights that is 3116 following DFW (Dallas/Fort Worth) 3066 and so on. Because of National aviation system reason. For instance, it can be more volume, Equipment, weather or runway. Pie chart is also shown the percentages of delayed flights with respect to airport.

```
In [25]: # Average nas delay with respect to airports
         sqlContext.sql("select airport, avg(nas_delay) as delays from airline_table \
         where nas_delay > 0.0 group by airport order by delays desc").show(20)
```

```
+-------+-----------------+
|airport|           delays|
+-------+-----------------+
|    DIK|           3116.0|
|    DFW|3066.8396226415093|
|    DEN|1991.1935483870968|
|    SLC|1784.6052631578948|
|    ASE|           1764.3|
|    JMS|           1763.5|
|    DVL|1727.6666666666667|
|    ORD| 1714.061403508772|
|    IAH|1550.3333333333333|
|    MSP|1423.6605504587155|
|    XWA| 1386.888888888889|
|    INL|          1241.25|
|    CLT| 1158.159574468085|
|    DTW| 1128.989898989899|
|    LAX|1097.1359223300972|
|    JFK|1001.4137931034483|
```

Figure 20:Average NAS delay w.r.t Airport

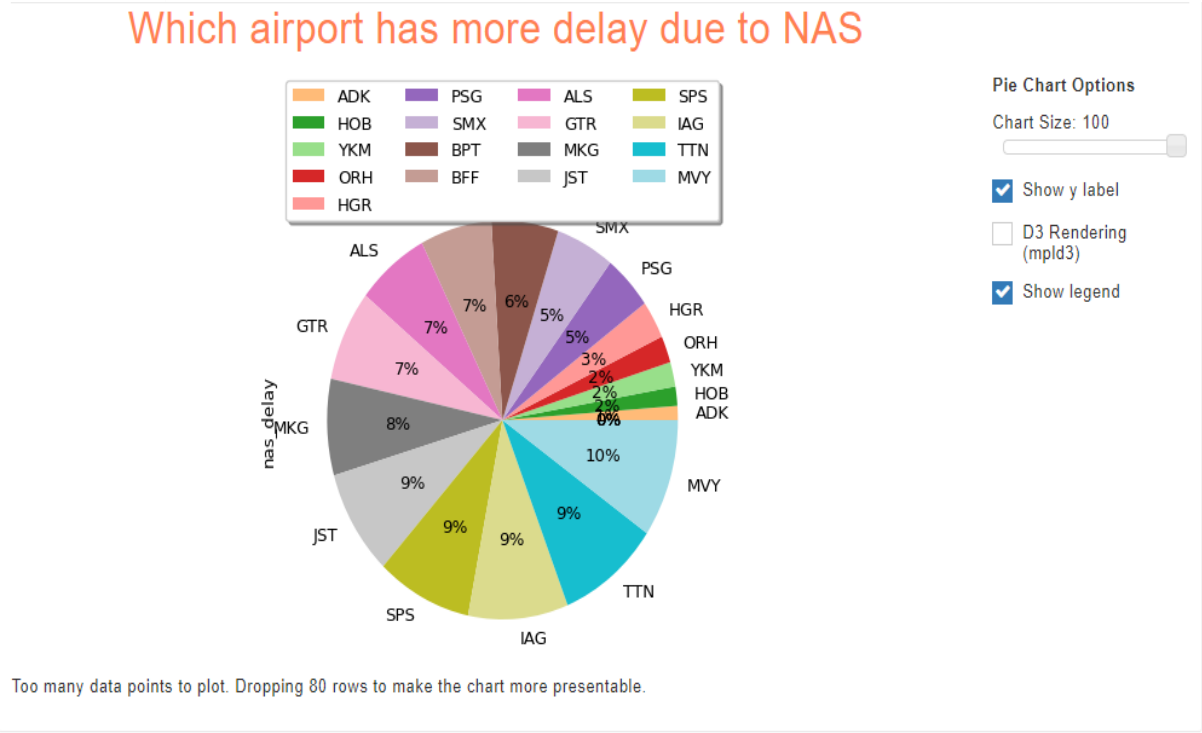

Figure 21: Average NAS delay w.r.t Airport pie chart

Flight diverted for certain carrier

Below figure illustrate that carrier OO (SkyWest Airlines) has 1405 flights are diverted in 2021. And then WN (Southwest Airlines) carrier has 932 diverted flights and then AA American Airline. We can see these results from bar chart as well. Flights are diverted due to Insufficient fuel owing to airspace congestion or a shortage of landing space at the target airport, as well as bird strikes in the air and disruptive passenger conduct on board. Some other that why flights are diverted is weather, technical issues with the aircraft, Medical issues, Passenger disruption or safety or other concerns. The aircraft may continue to its destination after a little delay in the most straightforward instances. The location where an aircraft lands has a significant influence on handling the diversion. Airlines will strive to plan diversions to those airports where they have a carrier or their network presence, or at least where they easily find and use adequate services. There are more number of diverted flights in the month of July and then June. [6]. And from figure it will be shown that more flights are diverted in the mointh of July and June and April. See the figures that showing Which carrier has more diverted flights in each month.

```
# Flight diverted for certain carrier
sqlContext.sql("select carrier, count(arr_diverted) as total_diverted from airline_table \
where arr_diverted > 0.0 group by carrier order by total_diverted desc").show()
```

```
+-------+--------------+
|carrier|total_diverted|
+-------+--------------+
|     OO|          1405|
|     WN|           932|
|     AA|           873|
|     MQ|           784|
|     G4|           725|
|     YV|           612|
|     UA|           531|
|     YX|           522|
|     F9|           436|
|     OH|           418|
|     AS|           415|
|     NK|           385|
|     B6|           379|
|     QX|           317|
|     DL|           313|
|     9E|           192|
|     HA|            42|
+-------+--------------+
```

Figure 22: Diverted flights of unique carrier

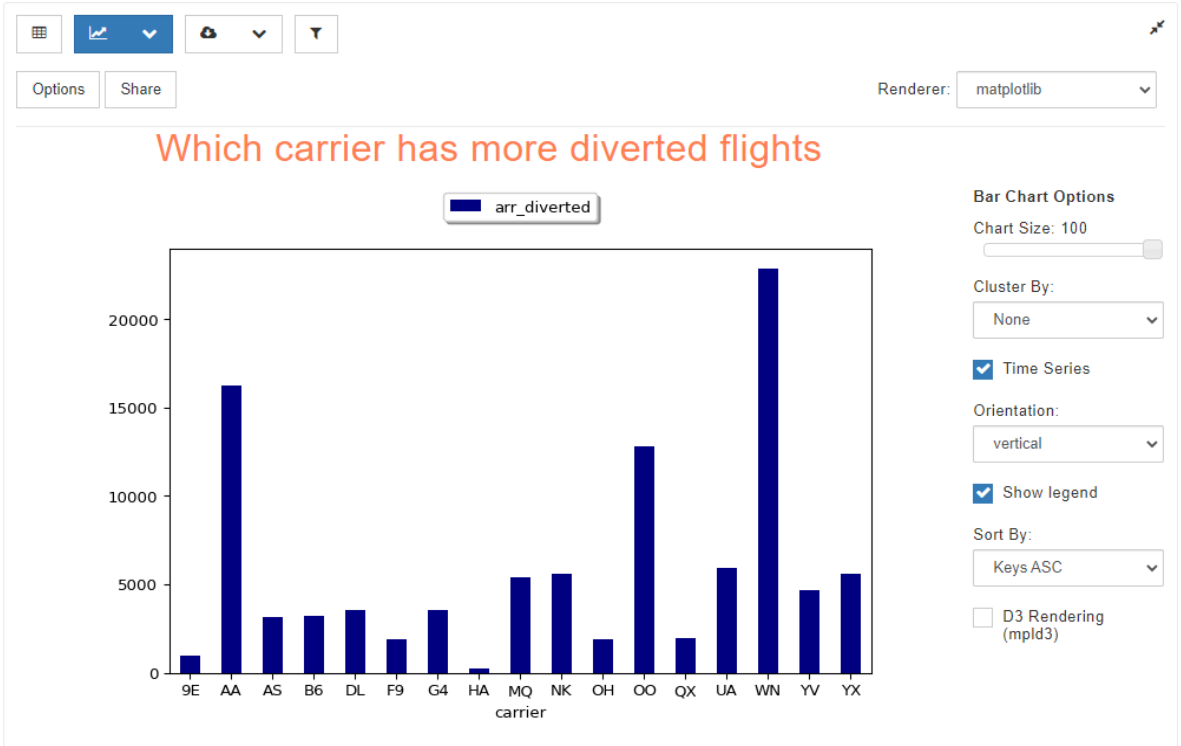

Figure 23: Bar chart of carrier diverted flights

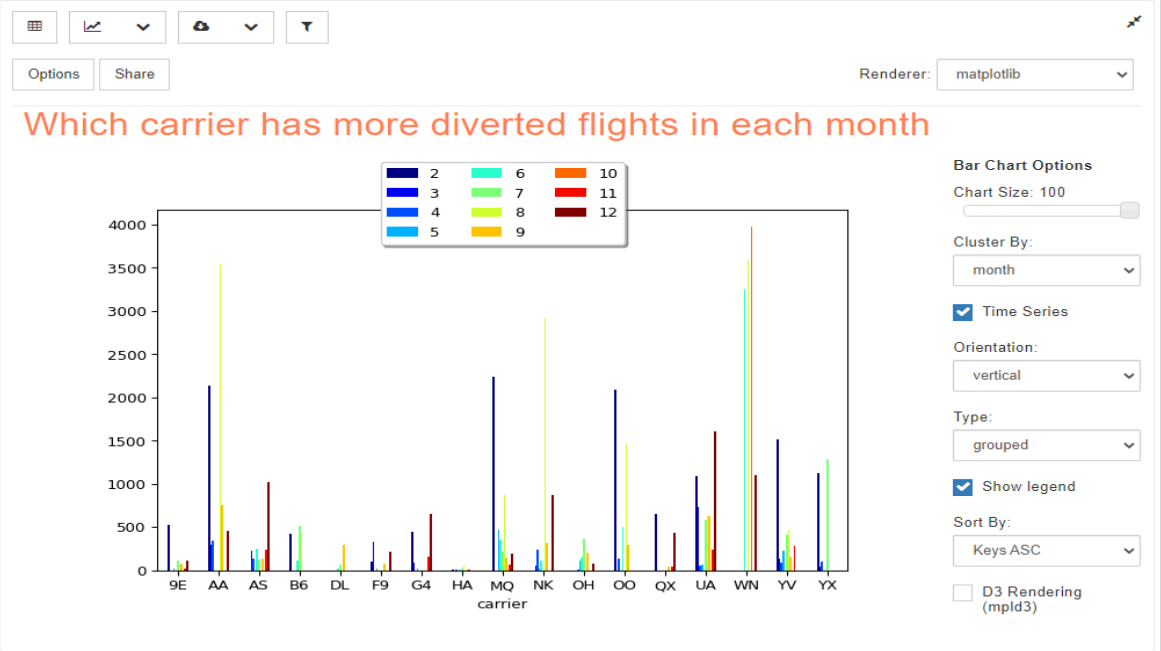

Figure 24:Which carrier has more diverted flights in each month

Flights cancelled from the airport or carrier

As we are taking the data of 2021 amidst covid-19 so can imagine a large number of flights being cancel due to corona virus. I analyse the data that which airport has most cancel flights BNA, ATL, PIT airport has most cancel flights 151, 144, 137 respectively. Pie chart is also showing the percentage of cancel flights. At least 4,500 trips all over the planet dropped/cancel as Covid-19 causes travel tumult. United Airlines and Delta Airline cancelled 280 flights between them on the Eve Christmas, according to online tracker (FlightAware.com). As a result of staff shortages, 2,401 flights were cancelled and 10,000 were delayed on Christmas Eve throughout the world. According to FlightAware.com, 1,779 itineraries were cancelled on Christmas Day, with 409 more cancelled on Boxing Day. United Airlines and Delta were among the worst-affected airlines, cancelling a total of 280 flights on Friday [7]. But the main reason for the cancellation of these flights is covid-19 in 2021.

```
# Flights cancelled from the airports
sqlContext.sql("select airport, count(arr_cancelled) as total_cancel from airline_table \
where arr_cancelled  > 0.0 \
group by airport order by total_cancel desc").show()
```

```
+-------+------------+
|airport|total_cancel|
+-------+------------+
|    BNA|         151|
|    ATL|         144|
|    PIT|         137|
|    DTW|         137|
|    CLE|         137|
|    AUS|         137|
|    RDU|         134|
|    MCI|         133|
|    ORD|         132|
|    PHX|         132|
|    MSP|         132|
|    CVG|         131|
|    CHS|         131|
|    IND|         131|
|    MSY|         128|
|    LAS|         127|
|    LAX|         126|
|    STL|         124|
|    IAH|         123|
|    SAN|         119|
+-------+------------+
only showing top 20 rows
```

Figure 25: cancel flights in airports

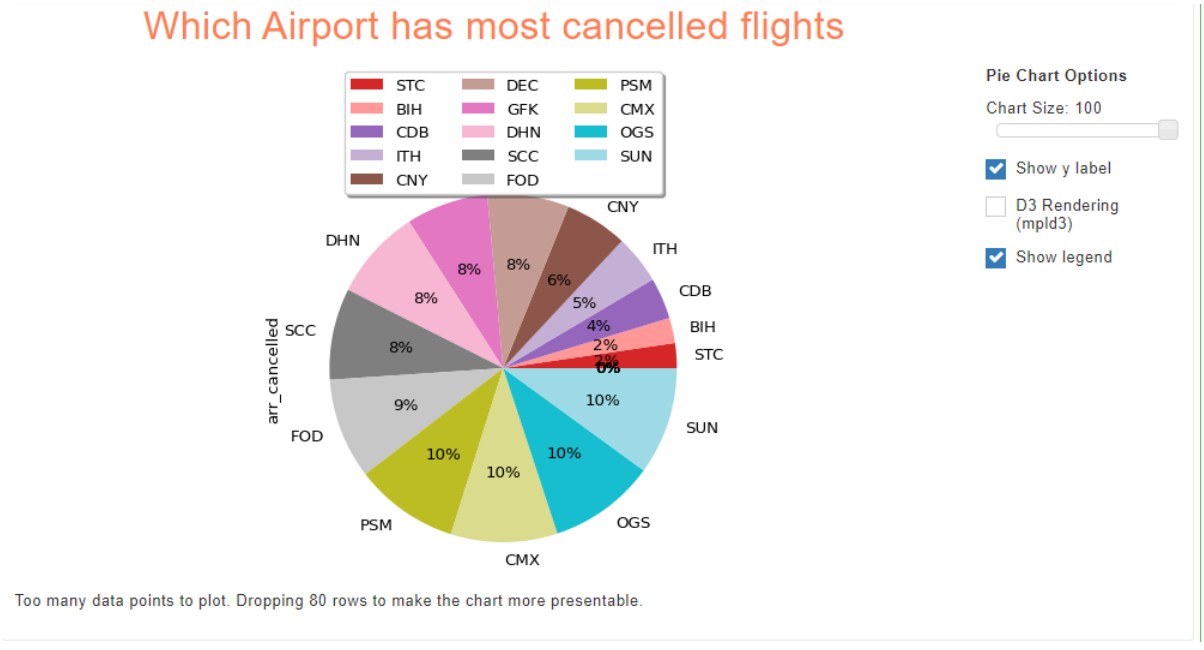

Figure 26: pie chart cancel flight in airport

Which airport has more weather delay

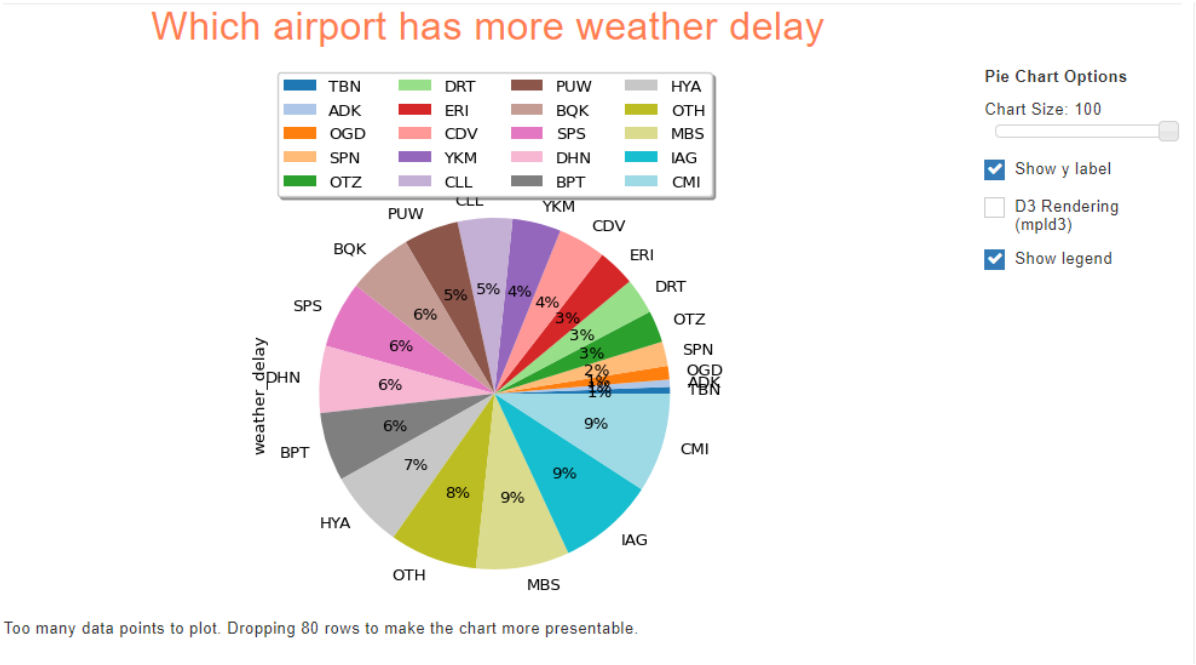

Figure 27:Weather delay

Airports with code CMI, IAG, MBS has 9 percent flights delay due to weather OTH has 8 percent flights delayed due to weather and HYA has & percent delay due to weather. DHN SPS BQK has 6 percent weather delay. Other effective airports are can be seen from the pie chart.

Significant meteorological circumstances (actual or predicted) that, in the carrier's opinion, cause a flight to be delayed or cancelled, such as a tornado, snowstorm, or hurricane.

Extreme weather that forbids flying falls under this category, Within the NAS category, there is another weather category. This weather hinders the system's functioning but does not block flights. The sort of weather delays or cancellations labelled "NAS" might be decreased with corrective action by airports or the Federal Aviation Administration. Weather was responsible for 45.8% of NAS delays in 2021. In 2021, NAS delays accounted for 33.4 percent of all delays [3].

Each month late aircraft delay by carrier

Analysis of carrier delay with respect to month. For the bar chart it can be seen that in April there more carrier delays and then in month of July and then June. For most fliers, this is an all-too-familiar scene. Your flight time has passed, but you're still waiting for your plane to arrive from its prior destination.

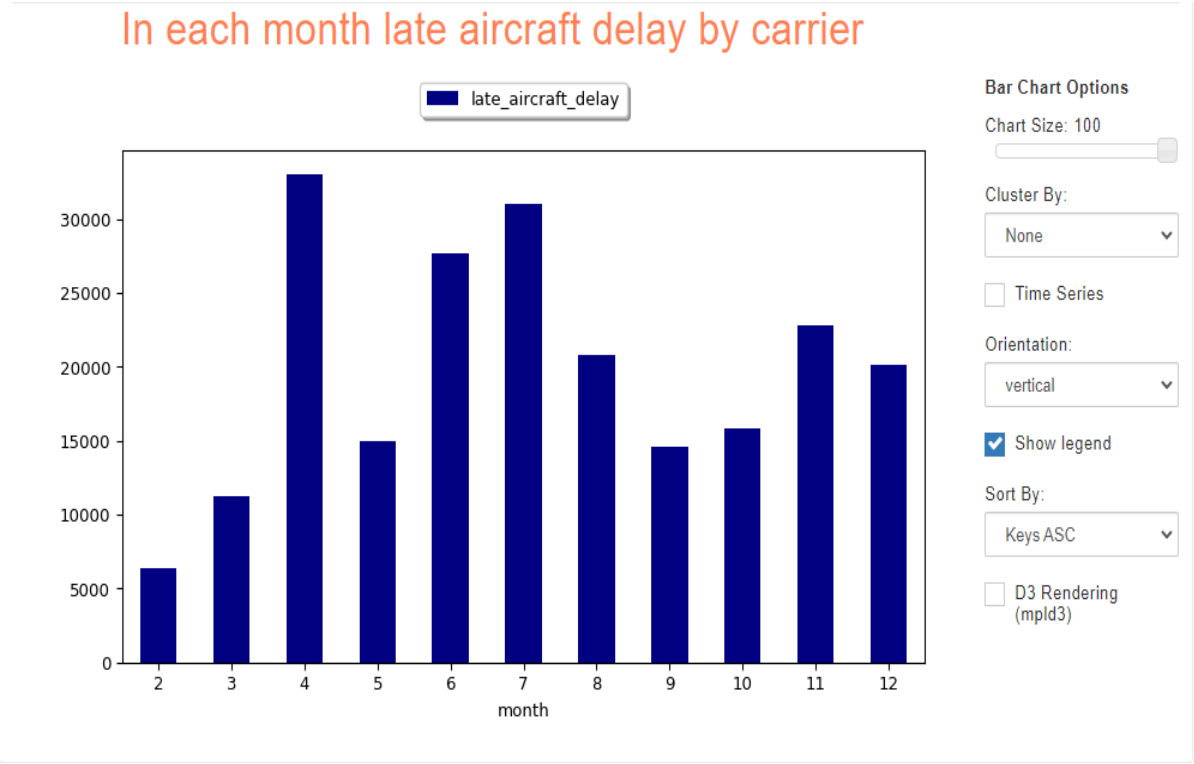

Figure 28: late aircraft delay by carrier

Overall delay of flights with respect to months

From the below bar charts delay of flights can be seen and shows that most delays are in the month of July, August and June respectively. And for the verification I attached the data from Bureau of Transportation figure. It says that 24.54% delayed in July, 23.36% in June and 21.99% in August which is matching with our analysis with respect to month wise.

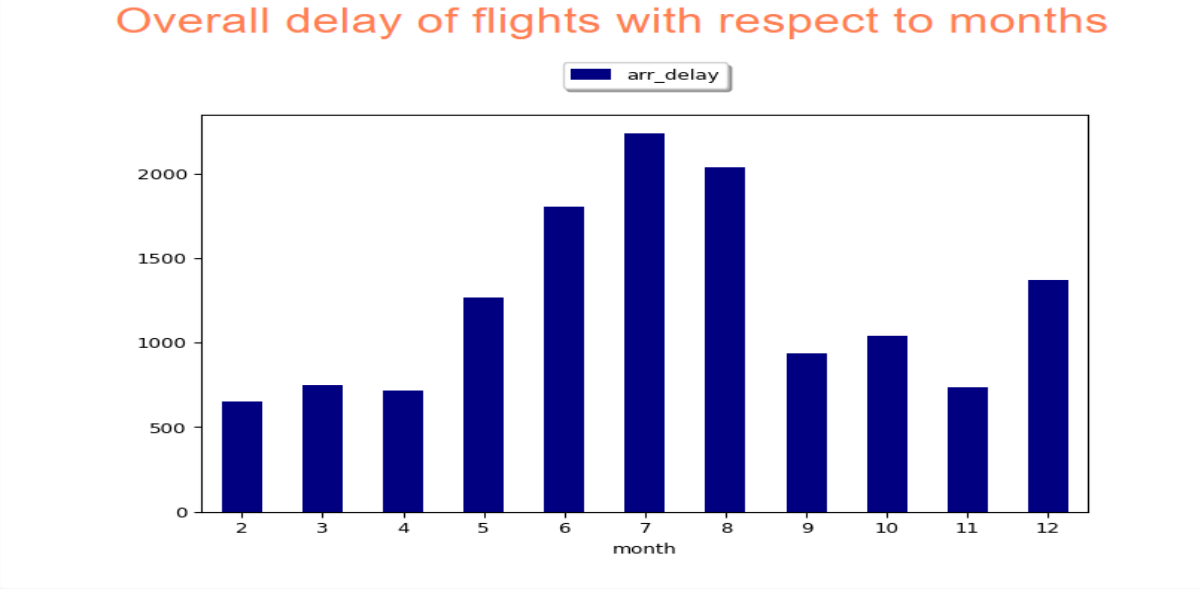

Figure 29: Overall flight delay w.r.t months

.

Month	Ontime Arrivals	Ontime (%)	Arrival Delays	Delayed (%)	Flights Cancelled	Cancelled (%)	Diverted	Flight Operations
January	322,243	89.16%	34,961	9.67%	3,647	1.01%	577	361,428
February	267,795	80.55%	45,593	13.71%	18,430	5.54%	650	332,468
March	393,322	88.49%	44,643	10.04%	5,766	1.30%	745	444,476
April	399,743	88.71%	47,682	10.58%	2,493	0.55%	719	450,637
May	426,391	86.05%	65,606	13.24%	2,283	0.46%	1,264	495,544
June	407,903	74.69%	127,565	23.36%	8,850	1.62%	1,806	546,124
July	428,758	73.51%	143,123	24.54%	9,142	1.57%	2,235	583,258
August	432,560	74.69%	127,387	21.99%	17,192	2.97%	2,040	579,179
September	457,014	84.94%	73,363	13.63%	6,740	1.25%	934	538,051
October	452,398	80.10%	99,586	17.63%	11,764	2.08%	1,040	564,788
November	461,728	84.32%	81,673	14.92%	3,424	0.63%	734	547,559
December	418,033	75.75%	119,150	21.59%	13,329	2.42%	1,373	551,885
--------	--------	--------	--------	--------	--------	--------	--------	--------
2021 (Annual)	4,867,888	81.19%	1,010,332	16.85%	103,060	1.72%	14,117	5,995,397

Figure 30: Data provided by BTS

Future recommendations or work for better operations for flight

1. **Better use of Technology**

Increased use of high-tech monitoring technologies to assist airports and airlines function more effectively is one method to decrease delays. The majority of departure delays are caused by so-called turnaround services, which are out of an airline's control. IntellAct's System artificial intelligence (AI) software technology can automatically detect delays in these turnaround services using existing security and surveillance cameras at airports and airlines.It can then alert airport personnel or ground crew to the issue and provide a real-time remediation strategy [10].

2. **Advanced Weather prediction and Forecast**

There should be advanced forecasting system for Terminal, En route, & Irregular Operations. The Weather Company should provides a variety of services to assist operators in proactively planning operations and anticipating and minimising disruptive weather impacts on terminal operations. The forecast should Provides hour-by-hour predictions and alerts for essential metrics affecting capacity and demand, as well as winter operations at major ports. Advanced predictions are now produced by The Weather Company at airports around Asia, Europe, and the United States, including Alaska [11].

3. **Allocating the airport more space and by increasing its capacity**

In the Aviation or airline industry, Engineering and technology variables such as airport design (e.g. runway configuration) and air traffic control needs (e.g. ideal aircraft spacing) significantly dictate the amount of production an airport can safely sustain, given weather conditions. Space capacity should be enough for every airport so that run way is available every time for the flights[12].

4. **Better and Quick Security**

Advanced and fast checking equipment's should be used so that passengers are check out more quickly and efficiently.

References:

1. Bts.gov. (2020). *OST_R | BTS | Title from h2*. [online] Available at: https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp.

2. www.bts.gov. (n.d.). *Bureau of Transportation Statistics*. [online] Available at: https://www.bts.gov/browse-statistical-products-and-data/bts-publications/airline-service-

3. https://www.bts.gov/topics/airlines-and-airports/understanding-reporting-causes-flight-delays-and-cancellations

4. Blalock, G., Kadiyali, V. and Simon, Daniel H. (2007). The Impact of Post-9/11 Airport Security Measures on the Demand for Air Travel. *The Journal of Law and Economics*, 50(4), pp.731–755. doi:10.1086/519816.

5. Bts.gov. (2017). *OST_R | BTS | Title from h2*. [online] Available at: https://www.transtats.bts.gov/OT_Delay/ot_delaycause1.asp?6B2r=I&20=E [Accessed 17 May 2022].

6. Simple Flying. (2021). *What Happens When An Aircraft Diverts?* [online] Available at: https://simpleflying.com/what-happens-when-an-aircraft-diverts/ [Accessed 17 May 2022].

7. Sky News. (n.d.). *COVID-19: At least 4,500 flights around the world cancelled as coronavirus causes travel chaos*. [online] Available at: https://news.sky.com/story/covid-19-at-least-4-500-flights-around-the-world-cancelled-as-coronavirus-causes-travel-chaos-12503644 [Accessed 17 May 2022].

8. SmarterTravel. (2019). *Flight Delays: What to Do and How to Prevent Them*. [online] Available at: https://www.smartertravel.com/avoid-flight-delays/.

9. Justice.gov. (2015). *Proposal For A Market-Based Solution To Airport Delays*. [online] Available at: https://www.justice.gov/atr/proposal-market-based-solution-airport-delays.

10. The airport tech helping to prevent delayed flights. (2022). *BBC News*. [online] 7 Feb. Available at: https://www.bbc.co.uk/news/business-60228430.

11. www.ibm.com. (n.d.). *Aviation Weather Forecast Solutions | The Weather Company, an IBM Business*. [online] Available at: https://www.ibm.com/weather/industries/aviation.