

BDM-1024: DATA TECHNOLOGY SOLUTIONS

PROJECT REPORT

SUBMITTED TO

BHAVIK GANDHI



Lambton
College

SUBMITTED BY

GROUP E

ABISHEAK DHANABAL(C0903766)
ALWIN KANNYAKONIL SCARIA (C0894287)
ASHNA VIJI ALEX(C0901082)
JERIN THENGUMPALLIL THOMAS (C0896235)
KUNCHERIA TOM(C0900973)
PRINCE THOMAS (C0894907)
MOHAMED AFTAB (C0891945)

August 2023

CONTENT

INTRODUCTION	4
DATABASE	5
DECISION TREE	6
DATA MODEL DIAGRAM	7
SQL CODES AND INSIGHTS	8
MAPREDUCE	16
SPARK JOBS	19
VISUALIZATION	22
GITHUB REPOSITORY LINK	25
CONCLUSION	26
REFERENCES	27

Type of activity	
Research Planning	<p>Kuncheria Tom: Create a group in WhatsApp for group communication and compilation of the files needed for the presentation and report.</p> <p>Prince Thomas: Identify the scope and objectives.</p> <p>Ashna Viji Alex: Divide the topic into sub-topics that need to be covered.</p> <p>Jerin T Thomas: Research specific sub-topics.</p> <p>Abhishek Dhanabal: Set deadlines for each stage of the report process.</p>
Data Gathering	<p>Selected Flight analysis as main subject for the project together</p> <p>Prince Thomas: Determine the specific areas of flight analysis to focus on.</p> <p>Ashna Viji Alex: Gather data from various sources, such as official US Bureau of Transportation documentation, case studies, articles, etc.</p> <p>Alwin Kannyakonil Scaria: Consolidate the gathered data via WhatsApp Group</p> <p>All group members: Collaborate as a team and discuss the challenges encountered if there is any. The dataset is taken from US Bureau of Transportation website.</p>
ERD, DTD	<p>Ashna Viji Alex, Jerin T Thomas: Developed Entity Relationship Diagram based on the dataset.</p> <p>Alwin Kannyakonil Scaria, Mohamed Aftab: Developed Data Model Diagram</p>
Analysis	<p>Prince Thomas: Analyze the gathered data and extract key insights features and use cases.</p> <p>All group members: Compile the analysis and findings into a cohesive document.</p>
SQL Analysis	Alwin Kannyakonil Scaria, Ashna Viji Alex, Jerin T Thomas: Setup the dataset into PostgreSQL and analysed the dataset and created SQL queries, found out meaningful insights from the query outputs
MapReduce	Jerin T Thomas, Kuncheria Tom: Setup Hadoop cluster and ran MapReduce programs.
Spark Jobs	Abhishek Dhanabal, Ashna Viji Alex, Mohamed Aftab: Executed pyspark jobs in GCP dataproc to develop meaningful insights.
Visualization	Prince Thomas, Kuncheria: Obtain meaningful and dynamic dashboard and charts from tableau.
Report and Presentation PowerPoint Drafting	<p>All group members: Collaborated on a shared platform to ensure consistency in style, formatting, and content.</p> <p>Kuncheria Tom: Reviewed and provided feedback on each other's sections.</p>
Report Merging	Ashna Viji Alex, Prince Thomas, Mohamed Aftab: Consolidated and performed last editing.
Finalization	Jerin T Thomas, Alwin Kannyakonil Scaria: Overall check was done.
GitHub Repository	Jerin T Thomas: Created repository and updated all the files

INTRODUCTION

In 2015, the Bureau of Transportation of federal agency in the United States, provided valuable data on domestic US airlines including various flights operators, flown flights, various airports, regarding the number of delays and cancellations. This data serves as a crucial resource for analysing the performance of airlines and identifying areas for optimization.

We look at the database as much as perspectives to identify the issues associated with existing flight schedule system, we primarily focused on resolving the root causes of delays, cancellations and diversions that cause immense operational cost, time and resource wastage rather than passenger analysis.

The analysis of delay and cancellation numbers can shed light on various factors affecting airline operations. By examining the data, airlines can identify common causes of delays such as weather conditions, air traffic congestion, airline issues. Understanding these patterns have a ripple nature, by predicting these causes we can make cost effective solution and improve the industry standards to a great extent.

DATABASE

We found the data base from US Bureau of Transportation of size 658MB and has 4 tables as listed below.

1.Airports

This Table holds the data of all airports having domestic terminals, airport code in addition to this it has the location details like city, state, and coordinates.

2. Airlines

This table accommodates details like airline codes and airline carrier details.

3.Flight

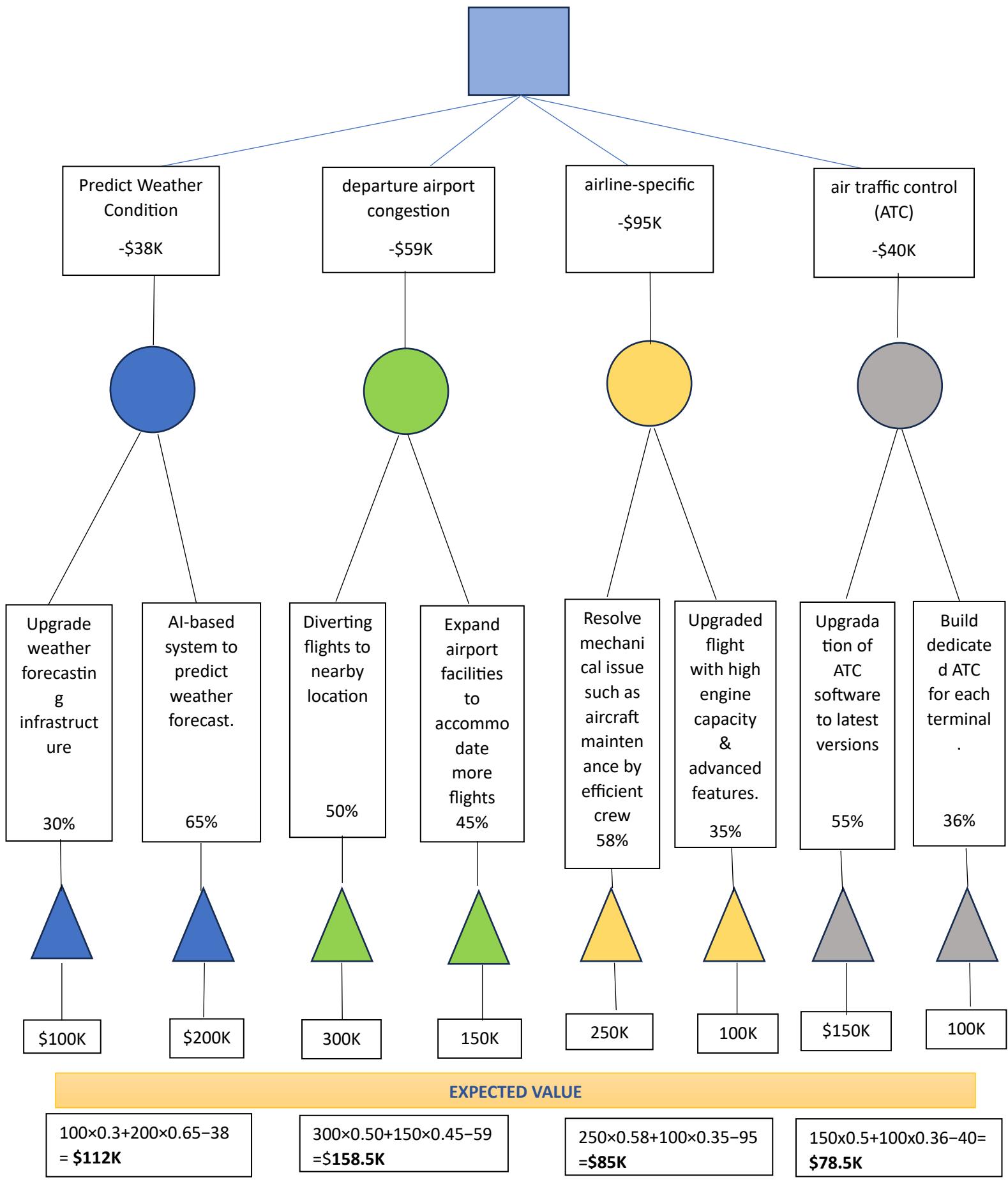
This is the main table that contains majority of data including the date of flight, source and destination airports, travel time, arrival and destination delays, cancellation code, diversion code, various factors of delay, taxi in, taxi out rates and many more.

4. Cancellation

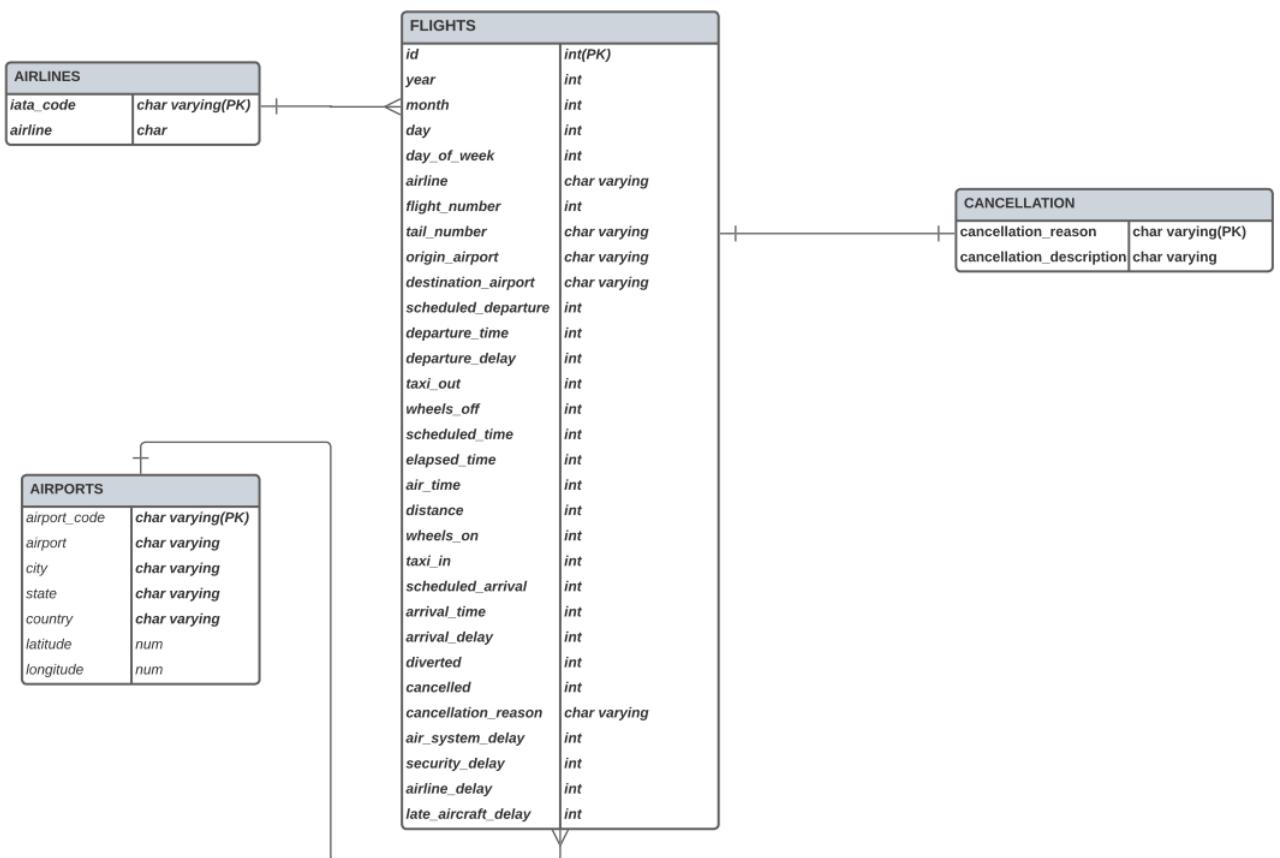
This table explains the details of each cancellation codes with description.

DECISION TREE

AIR TRAVEL OPTIMISATION



DATA MODEL DIAGRAM



SQL CODES & INSIGHTS

1. TIME VARIATION OF ALL FLIGHTS AND STATUS OF ARRIVAL

```

4 --TIME VARIATION OF ALL FLIGHTS AND STATUS OF ARRIVAL
5 select
6 ap.city as FROM_CITY,
7 ap.airport as from_airport,
8 fl.tail_number,
9 aps.city as destination_city,
10 aps.airport as destination_airport,
11 aline.airline,
12 fl.day_of_week,
13 (fl.departure_delay+fl.arrival_delay) TIME_VARIATION,
14 fl.arrival_delay,
15 fl.departure_delay,
16 CASE WHEN fl.arrival_delay<0 THEN 'DELAYED' WHEN fl.arrival_delay = 0 THEN 'ON TIME' WHEN fl.arrival_delay > 0 THEN 'EARLY ARRIVAL'ELSE 'NA'
17 from
18 flight fl
19 inner join airports ap on ap.airport_code = fl.origin_airport
20 inner join airports aps on aps.airport_code = fl.destination_airport
21 inner join airlines aline on aline.iata_code = fl.airline
22 order by day_of_week;
23

```

Data Output Messages Notifications

	from_city	from_airport	tail_number	destination_city	destination_airport	airline	day_of_week	time_variation	arrival_delay	departure_delay	status_completed
1	Detroit	Detroit Metropolitan Airport	N910EV	Grand Rapids	Gerald R. Ford International Airport	Skywest Airlines Inc.	1	-19	-14	-5	EARLY ARRIVAL
2	Pasco	Tri-Cities Airport	N616QX	Seattle	Seattle-Tacoma International Airport	Skywest Airlines Inc.	1	63	42	21	DELAYED
3	Las Vegas	McCarran International Airport	N689CA	Seattle	Seattle-Tacoma International Airport	Skywest Airlines Inc.	1	-27	-24	-3	EARLY ARRIVAL
4	Paducah	Barkley Regional Airport	N963SW	Chicago	Chicago O'Hare International Airport	Skywest Airlines Inc.	1	-13	-10	-3	EARLY ARRIVAL
5	Minneapolis	Minneapolis-Saint Paul International Airport	N151SY	Chicago	Chicago O'Hare International Airport	Skywest Airlines Inc.	1	27	26	1	DELAYED
6	San Francisco	San Francisco International Airport	N471UA	Las Vegas	McCarran International Airport	United Air Lines Inc.	1	-12	-11	-1	EARLY ARRIVAL
7	Chantilly	Washington Dulles International Airport	N8751Z	Denver	Denver International Airport	United Air Lines Inc.	1	-12	-12	0	EARLY ARRIVAL
8	Los Angeles	Los Angeles International Airport	N3413I	Newark	Newark Liberty International Airport	United Air Lines Inc.	1	-50	-41	-9	EARLY ARRIVAL
9	St Louis	St. Louis International Airport at Lambert Field	N833AW	Charlotte	Charlotte Douglas International Airport	American Airlines Inc.	1	-2	0	-2	ON TIME
10	Atlanta	Hartsfield-Jackson Atlanta International Airport	N593NW	Seattle	Seattle-Tacoma International Airport	Delta Air Lines Inc.	1	-53	-50	-3	EARLY ARRIVAL
11	Seattle	Seattle-Tacoma International Airport	N815DN	Salt Lake City	Salt Lake City International Airport	Delta Air Lines Inc.	1	-2	0	-2	ON TIME

Total rows: 1000 of 5332914 Query complete 00:00:19.932 Ln 4, Col 1

Analyse the general pattern of each flight on its delay by comparing the arrival & destination delays.

Insight: Majority of flights fail to departure/ arrive at scheduled time. Need to investigate more on this to identify the cause and need to look by airlines and airports statistics. Identified as serious issue in airline industry.

2. LIST OF ALL FLIGHTS WITH CANCELLATION

```

24 -- LIST OF ALL FLIGHTS WITH CANCELLATION
25 select
26 ap.city as from_city,
27 ap.airport as from_airport,
28 fl.tail_number,
29 aps.city as destination_city,
30 aps.airport as destination_airport,
31 aline.airline as Airline,
32 fl.day_of_week as Day_OF_WEEK,
33 fl.month as MONTH,
34 cancellation_description AS CANCELLATION_DESCRIPTION
35 from
36 flight fl
37 inner join airports ap on ap.airport_code = fl.origin_airport
38 inner join airports aps on aps.airport_code = fl.destination_airport
39 inner join airlines aline on aline.iata_code = fl.airline
40 inner join cancellation c on c.CANCELLATION_REASON = fl.CANCELLATION_REASON
41 where
42 fl.CANCELLED = 1
43 order by fl.day_of_week,fl.month;
44

```

Data Output Messages Notifications

	from_city	from_airport	tail_number	destination_city	destination_airport	airline	day_of_week	month	cancel char
1	San Francisco	San Francisco International Airport	[null]	Philadelphia	Philadelphia International Airport	United Air Lines Inc.	1	1	Weath
2	Seattle	Seattle-Tacoma International Airport	N371DA	New York	John F. Kennedy International Airport (New York International Airport)	Delta Air Lines Inc.	1	1	Weath
3	Phoenix	Phoenix Sky Harbor International Airport	N415WN	Philadelphia	Philadelphia International Airport	Southwest Airlines Co.	1	1	Weath
4	Philadelphia	Philadelphia International Airport	[null]	Las Vegas	McCarran International Airport	US Airways Inc.	1	1	Weath
5	Rochester	Greater Rochester International Airport	N516JB	New York	John F. Kennedy International Airport (New York International Airport)	JetBlue Airways	1	1	Weath
6	Covington	Cincinnati/Northern Kentucky International Airport	N665MQ	New York	John F. Kennedy International Airport (New York International Airport)	American Eagle Airlines Inc.	1	1	Weath
7	New York	LaGuardia Airport (Marine Air Terminal)	N371CA	Covington	Cincinnati/Northern Kentucky International Airport	Atlantic Southeast Airlines	1	1	Airline
8	San Francisco	San Francisco International Airport	N788AA	New York	John F. Kennedy International Airport (New York International Airport)	American Airlines Inc.	1	1	Weath
9	Grand Rapids	Gerald R. Ford International Airport	N607SW	Orlando	Orlando International Airport	Southwest Airlines Co.	1	1	Weath
10	Atlanta	Hartsfield-Jackson Atlanta International Airport	[null]	Philadelphia	Philadelphia International Airport	US Airways Inc.	1	1	Weath

Total rows: 1000 of 87430 Query complete 00:00:01.272 Ln 43, Col 34

Find the list of all flights cancelled and matched it with proper reasons to check the depth of the issue.

Insight: A sizable share of flights got cancelled due to weather and airline issues along with other reasons. Since it is a serious issue in Industry, need to find the patterns to cut down the numbers.

3. ANALYSIS OF DELAY ON DAILY BASIS

```

45 --ANALYSIS OF DELAY ON DAILY BASIS
46
47 SELECT
48   AP.AIRPORT,
49   aline.airline AS AIRLINE,
50   fl.DAY_OF_WEEK AS day_of_week,
51   fl.DAY AS day,
52   fl.MONTH AS month,
53   COUNT(*) AS num_delays,
54   ( case when fl.day_of_Week <=5 then'Weekday' else 'Weekend' end)as weektype
55
56 FROM
57 flight fl
58 inner join airports ap on ap.airport_code = fl.origin_airport
59 inner join airlines aline on aline.iata_code = fl.airline
60 WHERE
61   departure_delay > 0
62 GROUP BY
63   AP.AIRPORT,
64   aline.airline,
65   fl.DAY_OF_WEEK,
66   fl.DAY,
67   fl.MONTH
68 ORDER BY
69   num_delays DESC;
70
71 Data Output Messages Notifications
72
73
74 i | airport character varying | airline character varying | day_of_week integer | day integer | month integer | num_delays bigint | weektype text |
75 c | character varying | character varying | integer | integer | integer | bigint | text |
76 1 | Hartsfield-Jackson Atlanta International Airport.. Delta Air Lines Inc. | 7 | 23 | 8 | 577 | Weekend |
77 2 | Hartsfield-Jackson Atlanta International Airport.. Delta Air Lines Inc. | 3 | 30 | 12 | 574 | Weekday |
78 3 | Hartsfield-Jackson Atlanta International Airport.. Delta Air Lines Inc. | 5 | 10 | 4 | 552 | Weekday |
79 4 | Hartsfield-Jackson Atlanta International Airport.. Delta Air Lines Inc. | 2 | 24 | 2 | 488 | Weekday |
80 5 | Hartsfield-Jackson Atlanta International Airport.. Delta Air Lines Inc. | 4 | 5 | 3 | 484 | Weekday |
81 6 | Hartsfield-Jackson Atlanta International Airport.. Delta Air Lines Inc. | 2 | 18 | 8 | 472 | Weekday |
82
83
Total rows: 1000 of 248172 Query complete 00:00:06.113 Ln 45, Col 1

```

Attempt to identify the patterns in delay on weekday and weekends.

Insight: Major share of delays are shown in weekdays except few cases. Need to use larger carrier to accommodate more passengers and engine power on relevant days and reduce the number of smaller flights to reduce the traffic.

4. ANALYSIS by AIRLINE ON ARIVAL-DEPARTURE DELAY

```

71 -- ANALYSIS by AIRLINE ON ARIVAL-DEPARTURE DELAY
72
73 select
74   aline.airline AIRLINE,
75   COUNT(*) Number_of_Flights,
76   SUM(CASE WHEN fl.departure_delay > 0 THEN 1 ELSE 0 END) Number_Of_Delays_At_Departure,
77   SUM(CASE WHEN fl.arrival_delay > 0 THEN 1 ELSE 0 END) Number_Of_Delays_At_Arrival,
78   SUM(CASE WHEN fl.departure_delay < 0 THEN 1 ELSE 0 END) Number_of_Early_Departure,
79   SUM(CASE WHEN fl.arrival_delay < 0 THEN 1 ELSE 0 END) Number_of_Early_Arrival,
80   SUM(CASE WHEN fl.departure_delay = 0 and fl.arrival_delay = 0 THEN 1 ELSE 0 END) On_time_flights
81
82 from
83   airlines aline
84   inner join flight fl on aline.iata_code = fl.airline
85
86 group by aline.airline
87
88 Data Output Messages Notifications
89
90
91 i | airline character varying | number_of_flights bigint | number_of_delays_at_departure bigint | number_of_delays_at_arrival bigint | number_of_early_departure bigint | number_of_early_arrival bigint | on_time_flights bigint |
92 c | character varying | bigint | bigint | bigint | bigint | bigint | bigint |
93 1 | Alaska Airlines Inc. | 172521 | 43566 | 56953 | 121253 | 110162 | 306 |
94 2 | American Airlines Inc. | 725984 | 245904 | 252191 | 436039 | 446655 | 916 |
95 3 | American Eagle Airlines Inc. | 294632 | 93726 | 103505 | 171512 | 170102 | 522 |
96 4 | Atlantic Southeast Airlines | 571977 | 169897 | 213217 | 365593 | 328301 | 924 |
97 5 | Delta Air Lines Inc. | 875881 | 282463 | 250840 | 522581 | 601859 | 1697 |
98 6 | Frontier Airlines Inc. | 90836 | 34893 | 41232 | 52144 | 46673 | 114 |
99 7 | Hawaiian Airlines Inc. | 76272 | 20146 | 30179 | 52841 | 42545 | 272 |
100 8 | JetBlue Airways | 267048 | 102061 | 101998 | 147160 | 155170 | 403 |
101 9 | Skywest Airlines Inc. | 588353 | 171572 | 222435 | 380570 | 339967 | 1094 |
102 10 | Southwest Airlines Co. | 1261855 | 566807 | 470767 | 587783 | 742332 | 3093 |
103 11 | Spirit Air Lines | 117379 | 52089 | 56887 | 58663 | 55893 | 154 |
104 12 | United Air Lines Inc. | 515723 | 256550 | 186227 | 224854 | 312159 | 644 |
105 13 | US Airways Inc. | 198715 | 62565 | 76285 | 123032 | 113399 | 314 |
106 14 | Virgin America | 41000 | 20000 | 20100 | 20100 | 20000 | 100 |
107
108
109
Total rows: 14 of 14 Query complete 00:00:01.704 Ln 79, Col 36

```

Attempt to find time management of each airline in whole year.

Insight: Only a minority (less than 1%) could fly on scheduled time and rest either flew late or early. Checking further on next query.

5. ANALYSIS BY AIRLINE BY DELAY CAUSE

```

88 --ANALYSIS BY AIRLINE BY DELAY CAUSE
89
90 select
91 aline.airline AIRLINE,
92 count(*) Number_of_Flights,
93 SUM(CASE WHEN fl.AIR_SYSTEM_DELAY > 0 THEN 1 ELSE 0 END) Number_of_Delays_by_AIR_SYSTEM_DELAY,
94 SUM(CASE WHEN fl.SECURITY_DELAY > 0 THEN 1 ELSE 0 END) Number_of_Delays_by_SECURITY_DELAY,
95 SUM(CASE WHEN fl.AIRLINE_DELAY > 0 THEN 1 ELSE 0 END) Number_of_Delays_by_AIRLINE_DELAY,
96 SUM(CASE WHEN fl.LATE_AIRCRAFT_DELAY > 0 THEN 1 ELSE 0 END) Number_of_Delays_by_LATE_AIRCRAFT_DELAY,
97 SUM(CASE WHEN fl.WEATHER_DELAY > 0 THEN 1 ELSE 0 END) Number_of_Delays_by_WEATHER_DELAY
98 from
99 airlines aline
100 inner join flight fl on aline.iata_code = fl.airline
101 group by aline.airline
102
103 Data Output Messages Notifications
104
105
106 airline character varying | number_of_flights bigint | number_of_delays_by_air_system_delay bigint | number_of_delays_by_security_delay bigint | number_of_delays_by_airline_delay bigint | number_of_delays_by_late_aircraft_delay bigint | number_of_delays_by_weather_delay bigint
107
108 1 Alaska Airlines Inc. | 172521 | 14897 | 250 | 8363 | 7866 | 867
109 2 American Airlines Inc. | 725984 | 70383 | 731 | 68656 | 57590 | 9967
110 3 American Eagle Airlines Inc. | 294632 | 35806 | 308 | 27236 | 32125 | 9105
111 4 Atlantic Southeast Airlines | 571977 | 63586 | 0 | 53943 | 54904 | 2852
112 5 Delta Air Lines Inc. | 875881 | 64230 | 58 | 63128 | 50112 | 11838
113 6 Frontier Airlines Inc. | 90836 | 18066 | 0 | 11517 | 11638 | 580
114 7 Hawaiian Airlines Inc. | 76272 | 325 | 29 | 7395 | 4877 | 572
115 8 JetBlue Airways | 267048 | 33913 | 433 | 38665 | 31684 | 2173
116 9 Skywest Airlines Inc. | 588353 | 58509 | 345 | 39127 | 62432 | 4426
117 10 Southwest Airlines Co. | 1261855 | 97772 | 554 | 144524 | 163188 | 10086
118 11 Spirit Airlines | 117379 | 28378 | 299 | 17307 | 11722 | 925
119 12 United Air Lines Inc. | 515723 | 49858 | 26 | 65999 | 49865 | 7531
120 13 US Airways Inc. | 198715 | 23708 | 372 | 19717 | 13654 | 1722
121 14 Virgin America | 61903 | 5395 | 79 | 4445 | 5396 | 2072
122
123
124
125
126
127
Total rows: 14 of 14 Query complete 00:00:01.616

```

Checking on each airline on total number of delays and the share of various causes in total.

Insight: Least issues are reported by security issues and majority of airlines are affected by either flight issues or air traffic issues. Weather also has a share in delay.

6. ANALYSIS by AIRLINE ON CANCELLATION and REASONS

```

Query Query History
103 -- ANALYSIS by AIRLINE ON CANCELLATION and REASONS
104
105 select
106 (SELECT Count(*) from Flight f where aline.iata_code = f.airline group by f.airline) Total_number_of_Flights,
107 SUM(CASE WHEN fl.CANCELLATION_REASON = 'A' THEN 1 else 0 END) total_Flights_Cancelled,
108 SUM(CASE WHEN fl.CANCELLATION_REASON = 'B' THEN 1 else 0 END) CANCELLED_BY_ISSUE_AirlineOrCarrier,
109 SUM(CASE WHEN fl.CANCELLATION_REASON = 'C' THEN 1 else 0 END) CANCELLED_BY_ISSUE_Weather,
110 SUM(CASE WHEN fl.CANCELLATION_REASON = 'D' THEN 1 else 0 END) CANCELLED_BY_ISSUE_National_Air_System,
111 SUM(CASE WHEN fl.CANCELLATION_REASON = 'E' THEN 1 else 0 END) CANCELLED_BY_ISSUE_Security
112 from
113 airlines aline
114 inner join flight fl on aline.iata_code = fl.airline
115 inner join cancellation c on c.CANCELLATION_REASON = fl.CANCELLATION_REASON
116 group by aline.airline,aline.iata_code;
117
118 Data Output Messages Notifications
119
120
121 airline character varying | total_number_of_flights bigint | total_flights_cancelled bigint | cancelled_by_issue_airlinecarrier bigint | cancelled_by_issue_weather bigint | cancelled_by_issue_national_air_system bigint | cancelled_by_issue_security bigint
122
123 1 American Airlines Inc. | 725984 | 10919 | 2879 | 7306 | 730 | 4
124 2 Alaska Airlines Inc. | 172521 | 669 | 334 | 317 | 18 | 0
125 3 JetBlue Airways | 267048 | 4276 | 883 | 2464 | 928 | 1
126 4 Delta Air Lines Inc. | 875881 | 3824 | 594 | 2973 | 257 | 0
127 5 Atlantic Southeast Airlines | 571977 | 15231 | 3604 | 5082 | 6544 | 1
128 6 Frontier Airlines Inc. | 90836 | 588 | 308 | 280 | 0 | 0
129 7 Hawaiian Airlines Inc. | 76272 | 171 | 170 | 1 | 0 | 0
130 8 American Eagle Airlines Inc. | 294632 | 15025 | 2475 | 9164 | 3385 | 1
131 9 Spirit Airlines | 117379 | 2004 | 654 | 1068 | 279 | 3
132 10 Skywest Airlines Inc. | 588353 | 9960 | 3205 | 5539 | 1216 | 0
133 11 United Air Lines Inc. | 515723 | 6573 | 2870 | 3312 | 391 | 0
134 12 US Airways Inc. | 198715 | 4067 | 1007 | 2490 | 570 | 0
135 13 Virgin America | 61903 | 534 | 157 | 12 | 365 | 0
136 14 Southwest Airlines Co. | 1261855 | 16043 | 6122 | 8843 | 1066 | 12
137
138
139
140
141
142
Total rows: 14 of 14 Query complete 00:00:34.304

```

Checking on each airline on total number of cancellation and the share of various causes in total.

Insight: Airlines are primarily affected by either weather issues or carrier/traffic issues.

7. TOP 10 DELAYED ROUTES

Query History

```
--Top 10 delayed routes
SELECT
    ap.airport || ' to ' || aps.airport AS route,
    COUNT(*) AS no_of_delayed_flights,
    SUM(CASE WHEN fl.AIR_SYSTEM_DELAY > 0 THEN 1 ELSE 0 END) Number_of_Delay_by_AIR_SYSTEM_DELAY,
    SUM(CASE WHEN fl.SECURITY_DELAY > 0 THEN 1 ELSE 0 END) Number_of_Delays_by_SECURITY_DELAY,
    SUM(CASE WHEN fl.AIRLINE_DELAY > 0 THEN 1 ELSE 0 END) Number_of_Delays_by_AIRLINE_DELAY,
    SUM(CASE WHEN fl.LATE_AIRCRAFT_DELAY > 0 THEN 1 ELSE 0 END) Number_of_Delays_by_LATE_AIRCRAFT_DELAY,
    SUM(CASE WHEN fl.WEATHER_DELAY > 0 THEN 1 ELSE 0 END) Number_of_Delays_by_WEATHER_DELAY
FROM flight fl
INNER JOIN airports ap ON ap.airport_code = fl.origin_airport
INNER JOIN airports aps ON aps.airport_code = fl.destination_airport
WHERE departure_delay > 0
GROUP BY ap.airport, aps.airport
ORDER BY no_of_delayed_flights DESC
LIMIT 10;
```

Data Output Messages Notifications

route	no_of_delayed_flights	number_of_d	number_of_di	number_of_dx	number_of_delays_by_late_aircraft_delay	number_of_delays_by_weather_delay
Los Angeles International Airport to San Francisco International Airport	6004	1780	6	1367	1895	119
San Francisco International Airport to Los Angeles International Airport	5552	1889	8	1353	2111	88
Los Angeles International Airport to John F. Kennedy International Airport (New York International Airp...	4497	1026	9	988	718	80
Chicago O'Hare International Airport to Los Angeles International Airport	4415	985	10	1320	899	274
Chicago O'Hare International Airport to San Francisco International Airport	4282	937	2	1106	808	228
Los Angeles International Airport to McCarran International Airport	4058	955	4	1150	1501	54
Chicago O'Hare International Airport to LaGuardia Airport (Marine Air Terminal)	4048	1312	3	912	916	265
McCarran International Airport to Los Angeles International Airport	4002	1529	3	1189	1469	76
John F. Kennedy International Airport (New York International Airport) to Los Angeles International Air...	3943	887	20	1027	467	228
McCarran International Airport to San Francisco International Airport	3660	977	2	735	1081	127

Total rows: 10 of 10 Query complete 00:00:05.072 Ln 147, Col 10

Checking the highly delayed top 10 routes in U.S

Insight: Los Angeles to and from San Francisco is the highly delayed route, as well as routes involving airports like John F Kennedy, Chicago and McCarran are reported to be most delayed.

8. TOP 10 CANCELLED ROUTES

--Top 10 Cancelled routes

```
SELECT
    ap.airport || ' to ' || aps.airport AS route,
    COUNT(*) AS no_of_cancelled_flights,
    SUM(CASE WHEN fl.CANCELLATION_REASON = 'A' THEN 1 ELSE 0 END) CANCELLED_BY_ISSUE_AirlineORCarrier,
    SUM(CASE WHEN fl.CANCELLATION_REASON = 'B' THEN 1 ELSE 0 END) CANCELLED_BY_ISSUE_Weather,
    SUM(CASE WHEN fl.CANCELLATION_REASON = 'C' THEN 1 ELSE 0 END) CANCELLED_BY_ISSUE_National_Air_System,
    SUM(CASE WHEN fl.CANCELLATION_REASON = 'D' THEN 1 ELSE 0 END) CANCELLED_BY_ISSUE_Security
FROM flight fl
INNER JOIN airports ap ON ap.airport_code = fl.origin_airport
INNER JOIN airports aps ON aps.airport_code = fl.destination_airport
WHERE cancelled = 1
GROUP BY ap.airport, aps.airport
ORDER BY no_of_cancelled_flights DESC
LIMIT 10;
```

Data Output Messages Notifications

route	no_of_delayed_flights	number_of_d	number_of_di	number_of_dx	number_of_delays_by_late_aircraft_delay	number_of_delays_by_weather_delay
Los Angeles International Airport to San Francisco International Airport	6004	1780	6	1367	1895	119
San Francisco International Airport to Los Angeles International Airport	5552	1889	8	1353	2111	88
Los Angeles International Airport to John F. Kennedy International Airport (New York International Airp...	4497	1026	9	988	718	80
Chicago O'Hare International Airport to Los Angeles International Airport	4415	985	10	1320	899	274
Chicago O'Hare International Airport to San Francisco International Airport	4282	937	2	1106	808	228
Los Angeles International Airport to McCarran International Airport	4058	955	4	1150	1501	54
Chicago O'Hare International Airport to LaGuardia Airport (Marine Air Terminal)	4048	1312	3	912	916	265
McCarran International Airport to Los Angeles International Airport	4002	1529	3	1189	1469	76
John F. Kennedy International Airport (New York International Airport) to Los Angeles International Air...	3943	887	20	1027	467	228
McCarran International Airport to San Francisco International Airport	3660	977	2	735	1081	127

To identify the top 10 cancelled routes

Insight: The Los Angeles to San Francisco route has been reported as one of the routes with significant cancellation. Additionally, routes involving airports like John F Kennedy, Chicago, and McCarran have also been noted for experiencing a higher number of cancellations.

9. TOP 10 DIVERTED ROUTE

```
--Top 10 diverted route
169
170 SELECT
171 ap.airport || ' to ' || aps.airport AS route,
172 COUNT(*) AS no_of_diverted_flights
173 FROM flight fl
174 INNER JOIN airports ap ON ap.airport_code = fl.origin_airport
175 INNER JOIN airports aps ON aps.airport_code = fl.destination_airport
176 WHERE diverted = 1
177 GROUP BY ap.airport, aps.airport
178 ORDER BY no_of_diverted_flights DESC
179 LIMIT 10;
180
181 Data Output Messages Notifications
182
183
184 route text no_of_diverted_flights
185 1 Miami International Airport to LaGuardia Airport (Marine Air Terminal) 76
186 2 Salt Lake City International Airport to Friedman Memorial Airport 72
187 3 Chicago O'Hare International Airport to LaGuardia Airport (Marine Air Terminal) 66
188 4 Gen. Edward Lawrence Logan International Airport to Los Angeles International Airport 63
189 5 Gen. Edward Lawrence Logan International Airport to San Francisco International Airport 62
190 6 Los Angeles International Airport to Aspen-Pitkin County Airport 54
191 7 Cyril E. King Airport to Philadelphia International Airport 52
192 8 Hartsfield-Jackson Atlanta International Airport to LaGuardia Airport (Marine Air Terminal) 48
193 9 Dallas/Fort Worth International Airport to LaGuardia Airport (Marine Air Terminal) 47
194 10 Chicago O'Hare International Airport to Dallas/Fort Worth International Airport 47
195
196
197
198
199
200
201
202
203
204
205
206
207
e Total rows: 10 of 10 Query complete 00:00:00.780 Ln 170, Col 1
```

Identifying the top 10 diverted routes

Insight: The highest number of flights are diverted on the route to LaGuardia Airport.

10. TOP 10 FLOWN ROUTES

```
--Top 10 flown routes
186
187 SELECT
188 ap.airport || ' to ' || aps.airport AS route,
189 COUNT(*) AS no_of_flights
190 FROM flight fl
191 INNER JOIN airports ap ON ap.airport_code = fl.origin_airport
192 INNER JOIN airports aps ON aps.airport_code = fl.destination_airport
193 GROUP BY ap.airport, aps.airport
194 ORDER BY no_of_flights DESC
195 LIMIT 10;
196
197 Data Output Messages Notifications
198
199
200 route text no_of_flights
201 1 San Francisco International Airport to Los Angeles International Airport 13744
202 2 Los Angeles International Airport to San Francisco International Airport 13457
203 3 John F. Kennedy International Airport (New York International Airport) to Los Angeles International Airport 12016
204 4 Los Angeles International Airport to John F. Kennedy International Airport (New York International Airport) 12015
205 5 McCarran International Airport to Los Angeles International Airport 9715
206 6 LaGuardia Airport (Marine Air Terminal) to Chicago O'Hare International Airport 9639
207 7 Los Angeles International Airport to McCarran International Airport 9594
208 8 Chicago O'Hare International Airport to LaGuardia Airport (Marine Air Terminal) 9575
209 9 San Francisco International Airport to John F. Kennedy International Airport (New York International Airport) 8440
210 10 John F. Kennedy International Airport (New York International Airport) to San Francisco International Airport 8437
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
579
580
581
582
583
584
585
586
587
588
589
589
590
591
592
593
594
595
596
597
598
599
599
600
601
602
603
604
605
606
607
608
609
609
610
611
612
613
614
615
616
617
618
619
619
620
621
622
623
624
625
626
627
628
629
629
630
631
632
633
634
635
636
637
638
639
639
640
641
642
643
644
645
646
647
648
649
649
650
651
652
653
654
655
656
657
658
659
659
660
661
662
663
664
665
666
667
668
669
669
670
671
672
673
674
675
676
677
678
679
679
680
681
682
683
684
685
686
687
688
689
689
690
691
692
693
694
695
696
697
698
699
699
700
701
702
703
704
705
706
707
708
709
709
710
711
712
713
714
715
716
717
718
719
719
720
721
722
723
724
725
726
727
728
729
729
730
731
732
733
734
735
736
737
738
739
739
740
741
742
743
744
745
746
747
748
749
749
750
751
752
753
754
755
756
757
758
759
759
760
761
762
763
764
765
766
767
768
769
769
770
771
772
773
774
775
776
777
778
779
779
780
781
782
783
784
785
786
787
787
788
789
789
790
791
792
793
794
795
796
797
798
799
799
800
801
802
803
804
805
806
807
808
809
809
810
811
812
813
814
815
816
817
818
819
819
820
821
822
823
824
825
826
827
828
829
829
830
831
832
833
834
835
836
837
838
839
839
840
841
842
843
844
845
846
847
848
849
849
850
851
852
853
854
855
856
857
858
859
859
860
861
862
863
864
865
866
867
868
869
869
870
871
872
873
874
875
876
877
878
879
879
880
881
882
883
884
885
886
887
888
889
889
890
891
892
893
894
895
896
897
898
899
899
900
901
902
903
904
905
906
907
908
909
909
910
911
912
913
914
915
916
917
918
919
919
920
921
922
923
924
925
926
927
928
929
929
930
931
932
933
934
935
936
937
938
939
939
940
941
942
943
944
945
946
947
948
949
949
950
951
952
953
954
955
956
957
958
959
959
960
961
962
963
964
965
966
967
968
969
969
970
971
972
973
974
975
976
977
978
979
979
980
981
982
983
984
985
986
987
988
989
989
990
991
992
993
994
995
996
997
998
999
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1088
1089
1090
1091
1092
1093
1094
1095
1095
1096
1097
1098
1099
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1188
1189
1190
1191
1192
1193
1194
1195
1195
1196
1197
1198
1199
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1297
1298
1299
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1388
1389
1390
1391
1392
1393
1394
1395
1396
1396
1397
1398
1399
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1497
1498
1499
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1597
1598
1599
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1697
1698
1699
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1797
1798
1799
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1988
1989
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2088
2089
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2098
2099
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2188
2189
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2198
2199
2199
2200
2201
2202
2203
2204
2205
2206
2207
```

11. MOST DELAYED MONTH

```
--Most Delayed Month
SELECT month, COUNT(<>) AS count
FROM flight
WHERE arrival_delay > 0
GROUP BY month
ORDER BY count DESC
```

month	count
1	206989
2	199717
3	190133
4	188310
5	180891
6	179494
7	175443
8	175178
9	171820
10	149439
11	141250
12	133432

To find the month which is having the highest number of delayed flights.

Insight: June and July months are having the highest number of delays.

12. MOST ON-TIME MONTH

```
--Most On-Time Month
SELECT month, COUNT(<>) AS count
FROM flight
WHERE arrival_delay = 0
GROUP BY month
ORDER BY count DESC
```

month	count
1	11477
2	11324
3	11319
4	11158
5	10955
6	10832
7	10597
8	10126
9	9956
10	9745
11	9660
12	9084

Identify the month having the highest number of flights arrived on time.

Insight: The month of march and April is having the highest number of flights arrived on time

13. MOST TRAVELED DAYS IN A WEEK

```

209
210 -- Most Travelled Days in a Week in US
211 SELECT day_of_week, Sum(distance) as Total_distance FROM flight
212 GROUP BY day_of_week
213 ORDER BY Total_distance DESC;
214

```

Data Output Messages Notifications

	day_of_week	total_distance
	integer	bigint
1	4	714260507
2	1	706583476
3	5	705225205
4	3	694803803
5	2	683362562
6	7	680372853
7	6	600749003

To find which day of the week is having the most distance travelled.

Insight: Thursday followed by Monday and then Friday is having the highest

14. NUMBER OF MORNINGS, EVENING ARRIVALS AND DELAYS ANALYSIS BY AIRPORT MONTH WISE

```

--number of morning, evening arrival and delay analysis by airport monthwise
SELECT
ap.airport,month,
SUM(CASE WHEN arrival_time < 1200 THEN 1 ELSE 0 END) AS morning_arrival,
SUM(CASE WHEN arrival_time >= 1200 THEN 1 ELSE 0 END) AS evening_arrival,
SUM(CASE WHEN arrival_time < 1200 THEN 1 ELSE 0 END) + SUM(CASE WHEN arrival_time >= 1200 THEN 1 ELSE 0 END) AS total_arrivals,
CASE
WHEN SUM(CASE WHEN arrival_time < 1200 THEN 1 ELSE 0 END) < SUM(CASE WHEN arrival_time >= 1200 THEN 1 ELSE 0 END) THEN 'Evening'
WHEN SUM(CASE WHEN arrival_time < 1200 THEN 1 ELSE 0 END) > SUM(CASE WHEN arrival_time >= 1200 THEN 1 ELSE 0 END) THEN 'Morning'
ELSE 'Equal' END AS higher_arrival_time,
SUM(CASE WHEN arrival_delay > 0 THEN 1 ELSE 0 END) AS delayed,
SUM(CASE WHEN arrival_delay = 0 THEN 1 ELSE 0 END) AS on_time,
SUM(CASE WHEN arrival_delay < 0 THEN 1 ELSE 0 END) AS early
FROM flight inner join airports ap on ap.airport_code = destination_airport
group by airport,month
order by airport,month ;

```

Data Output Messages Notifications

airport	month	morning_arrival	evening_arrival	total_arrivals	higher_arrival_time	delayed	on_time	early
	character varying	integer	bigint	bigint	text	bigint	bigint	bigint
Aberdeen Regional Airport	1	3	57	60	Evening	15	0	45
Aberdeen Regional Airport	2	0	55	55	Evening	11	1	41
Aberdeen Regional Airport	3	0	61	61	Evening	8	1	52
Aberdeen Regional Airport	4	1	58	59	Evening	6	1	52
Aberdeen Regional Airport	5	0	62	62	Evening	9	3	49
Aberdeen Regional Airport	6	1	58	59	Evening	18	1	40
Aberdeen Regional Airport	7	1	61	62	Evening	25	2	34
Aberdeen Regional Airport	8	0	61	61	Evening	17	3	41
Aberdeen Regional Airport	9	0	60	60	Evening	17	1	42
Aberdeen Regional Airport	10	4	58	62	Evening	23	2	36
Aberdeen Regional Airport	11	3	58	61	Evening	14	1	46
	12

rows: 1000 of 3417 Query complete 00:00:01.654

Ln 216, Col 1

Insight: The highest number of airplanes arriving at airports is consistently more in the second half of the day and the on-time arrivals are significantly low in number.

15. TOP 10 DELAYED ROUTES BY AIRCRAFT ISSUES

```

235
236 SELECT
237 ap.airport || ' to ' || aps.airport AS route,
238 al.airline,
239 COUNT(*) AS no_of_delayed_flights,
240 SUM(CASE WHEN fl.AIRLINE_DELAY > 0 THEN 1 ELSE 0 END) Number_Of_Delays_by_AIRLINE_DELAY,
241 SUM(CASE WHEN fl.LATE_AIRCRAFT_DELAY > 0 THEN 1 ELSE 0 END) Number_of_Delays_by_LATE_AIRCRAFT_DELAY
242 FROM flight fl
243 INNER JOIN airports ap ON ap.airport_code = fl.origin_airport
244 INNER JOIN airports aps ON aps.airport_code = fl.destination_airport
245 INNER JOIN airlines al ON al.iata_code = fl.airline
246 WHERE departure_delay > 0 AND fl.AIRLINE_DELAY > 0 AND fl.AIRLINE_DELAY > 0
247 GROUP BY ap.airport, aps.airport, al.airline
248 ORDER BY no_of_delayed_flights DESC
249 LIMIT 10
250

```

Data Output Messages Notifications

route	airline	no_of_delayed_flights	number_of_delays_by_airline_delay	number_of_delays_by_late_aircraft_delay
1 Kahului Airport to Honolulu International Airport	Hawaiian Airlines Inc.	886	886	806
2 Dallas Love Field to William P Hobby Airport	Southwest Airlines Co.	869	869	545
3 George Bush Intercontinental Airport to Los Angeles International Airport	United Air Lines Inc.	820	820	399
4 William P Hobby Airport to Dallas Love Field	Southwest Airlines Co.	804	804	502
5 Dallas/Fort Worth International Airport to Los Angeles International Airport	American Airlines Inc.	773	773	320
6 Chicago O'Hare International Airport to San Francisco International Airport	United Air Lines Inc.	749	749	366
7 Chicago O'Hare International Airport to Los Angeles International Airport	United Air Lines Inc.	723	723	362
8 Newark Liberty International Airport to San Francisco International Airport	United Air Lines Inc.	721	721	350
9 McCarran International Airport to Los Angeles International Airport	Southwest Airlines Co.	714	714	484
10 San Francisco International Airport to Chicago O'Hare International Airport	United Air Lines Inc.	708	708	396

Total rows: 10 of 10 Query complete 00:00:01.810

Ln 235, Col 1

Insight: Hawaiian Airlines is the aircraft causing the highest number of delays in the Kahului to Honolulu airport and United Air Lines is the aircraft having the highest number of delays

16. ACTUAL NUMBER OF FLIGHTS FLOWN BY EACH AIRLINE ON EACH MONTH

```

252 -- Actual number of flights flown by each airlines on each month
253 select
254 aline.airline AIRLINE,
255 count(*) Number_of_Flights,
256 SUM(CASE WHEN fl.cancelled = 1 THEN 1 ELSE 0 END) no_of_cancelled,
257 (count(*) - SUM(CASE WHEN fl.cancelled = 1 THEN 1 ELSE 0 END)) ACTUAL_FLIGHTS
258 from
259 airlines aline
260 inner join flight fl on aline.iata_code = fl.airline
261 group by aline.airline;

```

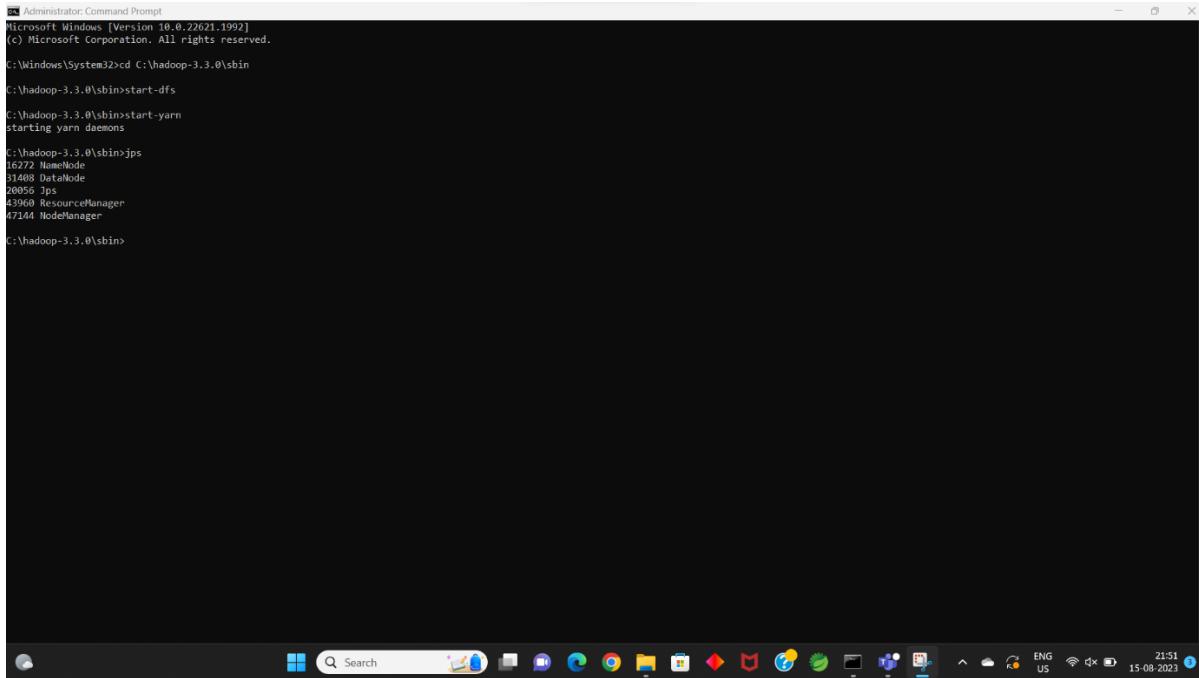
Data Output Messages Notifications

airline	number_of_flights	no_of_cancelled	actual_flights
1 Alaska Airlines Inc.	172521	669	171852
2 American Airlines Inc.	725984	10919	715065
3 American Eagle Airlines Inc.	294632	15025	279607
4 Atlantic Southeast Airlines	571977	15231	556746
5 Delta Air Lines Inc.	875881	3824	872057
6 Frontier Airlines Inc.	90836	588	90248
7 Hawaiian Airlines Inc.	76272	171	76101
8 JetBlue Airways	267048	4276	262772
9 Skywest Airlines Inc.	588353	9960	578393
10 Southwest Airlines Co.	1261855	16043	1245812
11 Spirit Airlines	117379	2004	115375
12 United Air Lines Inc.	515723	6573	509150
13 US Airways Inc.	198715	4067	194648
14 Virgin America	61903	534	61369

Insight: On observing the above data we could find that after reducing the cancelled flights from scheduled trips we can find a significant reduction in the actual flights flown.

MAPREDUCE JOBS

Hadoop cluster set-up was done. All nodes and daemons are up and running.



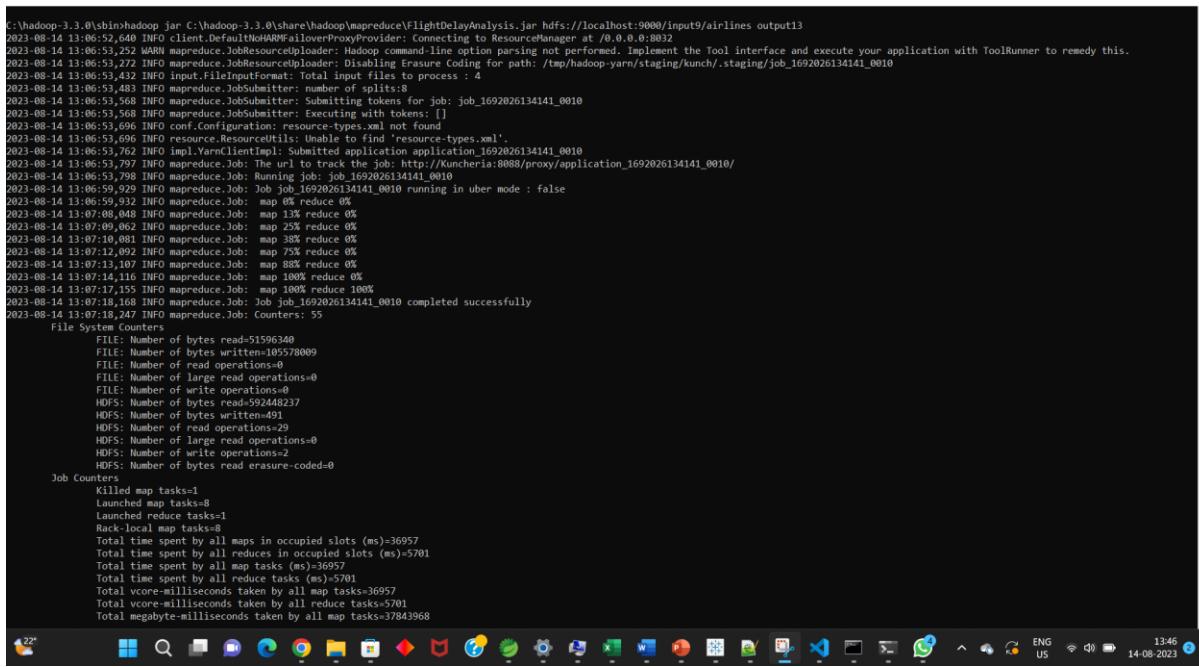
```
C:\Administrator: Command Prompt
Microsoft Windows [Version 10.0.22621.1992]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\System32>cd C:\hadoop-3.3.0\sbin
C:\hadoop-3.3.0\sbin>start-yarn
starting yarn daemons

C:\hadoop-3.3.0\sbin>jps
16722 NameNode
31408 DataNode
20056 Jps
43969 ResourceManager
47344 NodeManager
C:\hadoop-3.3.0\sbin>
```

Few MapReduce jobs was done on the running Hadoop cluster and found some insights. Below are some of the insights got while running the MapReduce jobs:

1. Average delay by airline



```
C:\hadoop-3.3.0\sbin>hadoop jar C:\hadoop-3.3.0\share\hadoop\mapreduce\flightDelayAnalysis.jar hdfs://localhost:9000/input9/airlines output13
2023-08-14 13:06:52,640 INFO client.DefaultHttpAioOverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8082
2023-08-14 13:06:53,252 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-08-14 13:06:53,252 INFO mapreduce.JobResourceUploader: Uploading Erasure Coding for path: /tmp/hadoop-yarn/staging/kunchi/.staging/job_1692026134141_0010
2023-08-14 13:06:53,432 INFO InputFormat: Total input files in process : 4
2023-08-14 13:06:53,483 INFO mapreduce.JobSubmitter: Number of splits:8
2023-08-14 13:06:53,568 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1692026134141_0010
2023-08-14 13:06:53,568 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-08-14 13:06:53,696 INFO conf.Configuration: resource-types.xml not found
2023-08-14 13:06:53,696 INFO resource.ResourceUtil: Unable to find 'resource-types.xml'.
2023-08-14 13:06:53,762 INFO impl.YarnClientImpl: Submitted application application_1692026134141_0010
2023-08-14 13:06:53,797 INFO mapreduce.Job: The url to track the job: http://Kuncheria:8088/proxy/application_1692026134141_0010/
2023-08-14 13:06:55,591 INFO mapreduce.Job: Running job: job_1692026134141_0010
2023-08-14 13:06:59,931 INFO mapreduce.Job: map 0% reduce 0%
2023-08-14 13:07:00,049 INFO mapreduce.Job: map 13% reduce 0%
2023-08-14 13:07:00,049 INFO mapreduce.Job: map 25% reduce 0%
2023-08-14 13:07:10,081 INFO mapreduce.Job: map 38% reduce 0%
2023-08-14 13:07:12,092 INFO mapreduce.Job: map 75% reduce 0%
2023-08-14 13:07:13,107 INFO mapreduce.Job: map 88% reduce 0%
2023-08-14 13:07:14,116 INFO mapreduce.Job: map 100% reduce 0%
2023-08-14 13:07:17,155 INFO mapreduce.Job: map 100% reduce 100%
2023-08-14 13:07:21,938 INFO mapreduce.Job: Job job_1692026134141_0010 completed successfully
2023-08-14 13:07:40,241 INFO mapreduce.Job: Counters
File System Counters
FILE: Number of bytes read=51596340
FILE: Number of bytes written=105578009
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=592448237
HDFS: Number of bytes written=491
HDFS: Number of read operations=29
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
Job Counters
Killed map tasks=1
Launched map tasks=8
Launched reduce tasks=1
Rack-local map tasks=8
Total time spent by all maps in occupied slots (ms)=36957
Total time spent by all reducers in occupied slots (ms)=5701
Total time spent by all map tasks (ms)=36957
Total time spent by all reduce tasks (ms)=5701
Total vcore-milliseconds taken by all map tasks=36957
Total vcore-milliseconds taken by all reduce tasks=5701
Total megabyte-milliseconds taken by all map tasks=37843968
```

OUTPUT:

```
C:\hadoop-3.3.0\sbin>hadoop fs -cat output13/part-r-00000
American Airlines     8.900856346719886
Alaska Airlines       1.7858007096736666
JetBlue Airways      11.5143526744102
Delta Air Lines       1.369254176813222
ExpressJet Airlines   8.1593449776958
Frontier Airlines    13.358958345331709
Hawaiian Airlines    0.48571315905796407
Delta Air Lines       10.125188203309524
Spirit Airlines       7.801103880415331
SkyWest Airlines     15.9447658807873688
United Airlines      14.435441010805953
US Airways Inc.      6.141136917746696
Virgin America       9.0225950896521952
Southwest Airlines   10.581986295158847

C:\hadoop-3.3.0\sbin>
```

2. The number of flights and distance travelled based on category of HAUL (flight distance)

```
E:\hadoop-3.3.0\bin>hadoop jar E:\hadoop-3.3.0\share\hadoop\mapreduce\FlightDistanceAnalysis.jar hdfs://localhost:9000/input9/airlines output27
2023-08-14 16:15:29 942 INFO Client:DefaultYARNFollowerProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2023-08-14 16:15:31,500 INFO mapreduce.JobResourceUploader: Uploading local file resources to remote application master via local resource manager
2023-08-14 16:15:23,520 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/kunch-staging/job_1692026134141_0025
2023-08-14 16:15:21,694 INFO InputFormat: Total input files to process : 1
2023-08-14 16:15:22,159 INFO mapreduce.JobSubmitter: number of splits:5
2023-08-14 16:15:22,247 INFO mapreduce.JobSubmitter: Submitting tasks for job: job_1692026134141_0025
2023-08-14 16:15:22,247 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-08-14 16:15:22,370 INFO conf.Configuration: resource-types.xml not found
2023-08-14 16:15:22,429 INFO resource.ResourceTills: Unable to find 'resource-types.xml'.
2023-08-14 16:15:22,429 INFO impl.YarnClientImpl: Submitted application application_1692026134141_0025
2023-08-14 16:15:22,429 INFO mapreduce.Job: The url to track the job: http://kuncherla:8088/proxy/application_1692026134141_0025/
2023-08-14 16:15:22,454 INFO mapreduce.Job: Running job: job_1692026134141_0025
2023-08-14 16:15:29,571 INFO mapreduce.Job: Job job_1692026134141_0025 running in uber mode : false
2023-08-14 16:15:29,571 INFO mapreduce.Job: map 0% reduce 0%
2023-08-14 16:15:37,678 INFO mapreduce.Job: map 20% reduce 0%
2023-08-14 16:15:38,693 INFO mapreduce.Job: map 40% reduce 0%
2023-08-14 16:15:39,704 INFO mapreduce.Job: map 60% reduce 0%
2023-08-14 16:15:40,717 INFO mapreduce.Job: map 100% reduce 0%
2023-08-14 16:15:46,789 INFO mapreduce.Job: map 100% reduce 100%
2023-08-14 16:15:46,802 INFO mapreduce.Job: Job job_1692026134141_0025 completed successfully
2023-08-14 16:15:46,803 INFO mapreduce.Job: Counters: 55
File System Counters
FILE: Number of bytes read=101050533
FILE: Number of bytes written=203691425
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=59242345
HDFS: Number of bytes written=3100
HDFS: Number of read operations=20
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read crasure-coded=0
Job Counters
Killed map tasks=1
Launched map tasks=5
Launched reduce tasks=1
Rack-local map tasks=3
Total time spent by all maps in occupied slots (ms)=25664
Total time spent by all reducers in occupied slots (ms)=5827
Total time spent by all map tasks (ms)=25664
Total time spent by all reduce tasks (ms)=5827
Total vcore-milliseconds taken by all map tasks=25664
Total vcore-milliseconds taken by all reduce tasks=5827
Total megabyte-milliseconds taken by all map tasks=2627936
Total megabyte-milliseconds taken by all reduce tasks=5966848
Map-Reduce Framework
2023-08-14 16:15:46,803 INFO mapreduce.Job: Counters: 55
File System Counters
FILE: Number of bytes read=101050533
FILE: Number of bytes written=203691425
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=59242345
HDFS: Number of bytes written=3100
HDFS: Number of read operations=20
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read crasure-coded=0
Job Counters
Killed map tasks=1
Launched map tasks=5
Launched reduce tasks=1
Rack-local map tasks=3
Total time spent by all maps in occupied slots (ms)=25664
Total time spent by all reducers in occupied slots (ms)=5827
Total time spent by all map tasks (ms)=25664
Total time spent by all reduce tasks (ms)=5827
Total vcore-milliseconds taken by all map tasks=25664
Total vcore-milliseconds taken by all reduce tasks=5827
Total megabyte-milliseconds taken by all map tasks=2627936
Total megabyte-milliseconds taken by all reduce tasks=5966848
Map-Reduce Framework
```

Map-Reduce Framework

```
C:\hadoop-3.3.0\bin>hadoop fs -cat output27/part-r-00000
Long Haul    Total Flights: 786370, Total Distance: 1617872089, Max Distance: 4983, Avg Distance: 2857.44
Medium Haul  Total Flights: 291254, Total Distance: 2524130631, Max Distance: 1590, Avg Distance: 865.64
Short Haul   Total Flights: 2126155, Total Distance: 0433424893, Max Distance: 497, Avg Distance: 383.44

C:\hadoop-3.3.0\bin>
```

3. Average delay on different days of a week

```
C:\> hadoop-3.3.0\bin\hadoop jar C:\hadoop-3.3.0\share\hadoop\mapreduce\DayofWeekAnalysis.jar hdfs://localhost:9000/input9/airlines output34
2023-08-15 22:23:14,184 INFO client.DefaultNWARNFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2023-08-15 22:23:14,762 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-08-15 22:23:14,783 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/kunchi-staging/job_1692150679872_0004
2023-08-15 22:23:14,952 INFO input.FileInputFormat: Total input files to process : 4
2023-08-15 22:23:14,999 INFO mapreduce.JobSubmitter: number of splits:4
2023-08-15 22:23:15,001 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1692150679872_0004
2023-08-15 22:23:15,089 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-08-15 22:23:15,212 INFO configuration: resource-types.xml not found
2023-08-15 22:23:15,213 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-08-15 22:23:15,270 INFO impl.YarnClientImpl: Submitted application application_1692150679872_0004
2023-08-15 22:23:15,299 INFO mapreduce.Job: The url to track the job: http://Kuncheria:8088/proxy/application_1692150679872_0004
2023-08-15 22:23:15,306 INFO mapreduce.Job: Running job: job_1692150679872_0004
2023-08-15 22:23:22,442 INFO mapreduce.Job: Job job_1692150679872_0004 running in uber mode : false
2023-08-15 22:23:22,451 INFO mapreduce.Job: map 0% reduce 0%
2023-08-15 22:23:22,452 INFO mapreduce.Job: map 1% reduce 0%
2023-08-15 22:23:22,630 INFO mapreduce.Job: map 10% reduce 0%
2023-08-15 22:23:22,631 INFO mapreduce.Job: map 63% reduce 0%
2023-08-15 22:23:34,657 INFO mapreduce.Job: map 75% reduce 0%
2023-08-15 22:23:36,675 INFO mapreduce.Job: map 88% reduce 0%
2023-08-15 22:23:37,681 INFO mapreduce.Job: map 100% reduce 0%
2023-08-15 22:23:37,723 INFO mapreduce.Job: map 100% reduce 100%
2023-08-15 22:23:41,768 INFO mapreduce.Job: Job job_1692150679872_0004 completed successfully
2023-08-15 22:23:41,839 INFO mapreduce.Job: Counters:
File System Counters
FILE: Number of bytes read=5730926
FILE: Number of bytes written=116985981
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=592448237
HDFS: Number of bytes written=278
HDFS: Number of read operations=29
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
Job Counters
Killed map tasks=1
Launched map tasks=8
Launched reduce tasks=1
Rack-local map tasks=8
Total time spent by all maps in occupied slots (ms)=43275
Total time spent by all reduces in occupied slots (ms)=5674
Total time spent by all map tasks (ms)=43275
Total time spent by all reduce tasks (ms)=5674
Total vcore-milliseconds taken by all map tasks=43275
Total vcore-milliseconds taken by all reduce tasks=5674
Total megabyte-milliseconds taken by all map tasks=44313600
22:24 15-08-2023
```

OUTPUT:

```
C:\> hadoop-3.3.0\bin\hadoop fs -cat output34/part-r-00000
Day of Week: 1 Average Delay: 16.178976
Day of Week: 2 Average Delay: 16.07923
Day of Week: 3 Average Delay: 16.232548
Day of Week: 4 Average Delay: 16.555765
Day of Week: 5 Average Delay: 16.721626
Day of Week: 6 Average Delay: 15.290664
Day of Week: 7 Average Delay: 15.7076
C:\> hadoop-3.3.0\bin\
```

SPARK JOBS

Some spark jobs were run on GCP DATAPROC to obtain meaningful insights. Mentioned below are some of them:

1. Average departure and arrival delays for each airline based on year and month.

Job details CLONE DELETE STOP REFRESH

Type: Dataproc Job Status: Succeeded

Output: LINE WRAP: OFF

Spark jobs take ~60 seconds to initialize resources. DISMISS

```
|year|month|airline|avg_departure_delay| avg_arrival_delay|
+---+---+---+---+---+
|2015| 1 | AA | 10.593542260208928 | 6.955843432232982 |
|2015| 1 | AS | 3.1782088195181086 | -0.3208881453881834 |
|2015| 1 | B6 | 10.035555988505187 | 7.347280539099862 |
|2015| 1 | DL | 5.984238298406325 | -2.04384679753125195 |
|2015| 1 | EV | 9.752522427331304 | 8.537496880459196 |
|2015| 1 | F9 | 17.98443291326909 | 18.357238307349665 |
|2015| 1 | HA | 1.210654409473356 | 3.5126404494382024 |
|2015| 1 | MQ | 16.081267083483918 | 18.164973882762624 |
|2015| 1 | NK | 13.146293512200764 | 11.39805375347544 |
|2015| 1 | OO | 12.155156988868642 | 10.889893962046941 |
|2015| 1 | UA | 14.01039374165644 | 6.35272035280893 |
|2015| 1 | US | 5.197315865126567 | 3.10745735548587 |
|2015| 1 | VX | 6.910771877015693 | 1.4207015278674413 |
|2015| 1 | WN | 9.514469976705627 | 3.3984656332857434 |
|2015| 2 | AA | 11.49939442874449 | 8.938527176134042 |
|2015| 2 | AS | 7.20868415851313 | 5.549828178694158 |
|2015| 2 | B6 | 25.272304832713754 | 24.42070275403609 |
|2015| 2 | DL | 28.94596670934699 | 15.992504523132592 |
|2015| 2 | EV | 16.220496894409937 | 15.669058295964126 |
|2015| 2 | F9 | 45.60138248847926 | 48.88425925925926 |
+---+---+---+---+---+
only showing top 20 rows
```

DataFrame[year: int, month: int, airline: string, avg_departure_delay: double, avg_arrival_delay: double]

Search 225 PM 14-Aug-23

2. Flights with highest arrival delay

Job details CLONE DELETE STOP REFRESH

Job ID: job-240f66cc Job UUID: e79772d0-3c97-4675-ac3d-b7fc2bfe7b3d

Type: Dataproc Job Status: Succeeded

Output: LINE WRAP: OFF

Spark jobs take ~60 seconds to initialize resources. DISMISS

```
|airline|flight_number|arrival_delay|departure_delay|
+---+---+---+---+
|Frontier Airlines...| 1274 | 324 | -6 |
|Delta Air Lines Inc.| 1156 | 237 | 0 |
|American Airlines...| 1307 | 227 | -4 |
|American Airlines...| 291 | 225 | -5 |
|American Airlines...| 126 | 226 | -2 |
|Spirit Air Lines| 718 | 243 | 16 |
|American Airlines...| 1283 | 216 | -4 |
|American Airlines...| 2346 | 438 | 219 |
|Atlantic Southeas...| 4169 | 206 | -2 |
|American Airlines...| 1120 | 353 | 145 |
|United Air Lines ...| 1106 | 388 | 183 |
|American Airlines...| 2335 | 339 | 136 |
|American Airlines...| 125 | 337 | 135 |
|Skywest Airlines ...| 4739 | 195 | -4 |
|Atlantic Southeas...| 6017 | 403 | 205 |
|American Airlines...| 643 | 188 | -9 |
|American Eagle Ai...| 3418 | 191 | -4 |
|Delta Air Lines Inc.| 455 | 191 | -3 |
|JetBlue Airways| 350 | 188 | -6 |
|JetBlue Airways| 643 | 183 | -8 |
+---+---+---+---+
```

3. Flights that departed earlier than their scheduled departure time

[Job details](#) [CLONE](#) [DELETE](#) [STOP](#) [REFRESH](#)

Job ID	job-28d79b1e
Job UUID	a19dcdc4-cc72-4658-b5bb-98eafe4cccd7
Type	Dataproc job
Status	Succeeded

Output [LINE WRAP: OFF](#)

! Spark jobs take ~60 seconds to initialise resources. [DISMISS](#)

```
23/08/15 01:00:18 INFO GoogleCloudStorageImpl: Ignoring exception of type GoogleJsonResponseException; verified object already exists with desired state.
23/08/15 01:00:18 WARN GfsStorageStatistics: Detected potential high latency for operation op_mkdirs. latencyMs=184; previousMaxLatencyMs=0; operationCount=1; conte
+-----+-----+-----+-----+
| airline|flight_number|scheduled_departure|departure_delay|
+-----+-----+-----+-----+
| Spirit Air Lines| 298| 0003| -1|
| American Airlines...| 258| 0001| -3|
| Spirit Air Lines| 298| 0003| -1|
| United Air Lines ...| 460| 0003| -1|
| United Air Lines ...| 1550| 0002| -3|
| Delta Air Lines Inc.| 1359| 0001| -4|
| Spirit Air Lines| 298| 0003| -2|
+-----+-----+-----+-----+
23/08/15 01:00:58 WARN GfsStorageStatistics: Detected potential high latency for operation op_rename. latencyMs=247; previousMaxLatencyMs=0; operationCount=1; conte

```

Output is complete [Stop sharing](#) [Hide](#)

[EQUIVALENT COMMAND LINE](#)



9:02 PM
8/14/2023

4. Analysing Airport Distances within a Threshold

[Job details](#) [CLONE](#) [DELETE](#) [STOP](#) [REFRESH](#)

Job ID	job-7cbd9f00
Job UUID	2e48c086-bd33-4c5b-a792-5d87eb3144b0
Type	Dataproc Job
Status	Succeeded

Output [LINE WRAP: OFF](#)

! Spark jobs take ~60 seconds to initialize resources. [DISMISS](#)

```
|airport1|airport2| distance_in_km|
+-----+-----+-----+
| ABE| AVP| 79.8911541444799|
| ABE| PHL| 88.40934907058318|
| ABE| TTN| 67.51170846955036|
| ABQ| SAF| 79.65206030968065|
| ACK| HYA| 49.79288674137177|
| ACK| MVY| 48.8230058001908|
| ACT| GRK| 83.1203377785118|
| ACV| CEC| 89.81902880554524|
| ACY| ILG| 91.59493341853448|
| ACY| PHL| 73.16426735892473|
| ACY| TTN| 93.28676524694741|
| ALO| CID| 93.88812139885276|
| ASE| EGE| 46.8245915886323|
| ASE| GUC| 76.83758560497205|
| ATW| GRB| 40.015298381501516|
| AUS| GRK| 97.95324687691345|
| AVL| GSP| 66.88913528874063|
| AVP| ABE| 79.8911541444799|
| AVP| BGM| 99.06557851546971|
| AZO| GRR| 71.86530267371367|

```



ENG IN 03:39 PM 16-08-2023

5. Analyzing High Delay Flights

[Job details](#) [CLONE](#) [DELETE](#) [STOP](#) [REFRESH](#)

Job ID	job-0b8c5f10
Job UUID	c30e9326-e59f-4e36-aa56-369e9164dbeb
Type	Dataproc Job
Status	Succeeded
Output	LINE WRAP: OFF

! Spark jobs take ~60 seconds to initialize resources. [DISMISS](#)

```
+-----+-----+-----+
|airline|flight_number|departure_delay|arrival_delay|
+-----+-----+-----+
| DL | 2478 | 999 | 995 |
| AA | 301 | 997 | 1035 |
| F9 | 333 | 996 | 1005 |
| OO | 4443 | 995 | 996 |
| UA | 200 | 994 | 979 |
| EV | 4912 | 993 | 1005 |
| DL | 1977 | 991 | 998 |
| AA | 1242 | 991 | 958 |
| DL | 1935 | 990 | 985 |
| NK | 293 | 99 | 99 |
+-----+-----+-----+
```

23/08/16 20:21:37 WARN GhfsStorageStatistics: Detected potential high latency for operation stream_write_close_operations. latencyMs=108; previousMaxLatencyMs=0; operationCount=1; contextX=23/08/16 20:21:38 WARN GhfsStorageStatistics: Detected potential high latency for operation op_rename. latencyMs=233; previousMaxLatencyMs=0; operationCount=1; contextX=



6. Analyzing Airport Distances within a Threshold of 100Kms (Finding Neighbouring Airports)

[Job details](#) [CLONE](#) [DELETE](#) [STOP](#) [REFRESH](#)

Job ID	job-7cbd9f00
Job UUID	2e48c086-bd33-4c5b-a792-5d87eb3144b0
Type	Dataproc Job
Status	Succeeded
Output	LINE WRAP: OFF

! Spark jobs take ~60 seconds to initialize resources. [DISMISS](#)

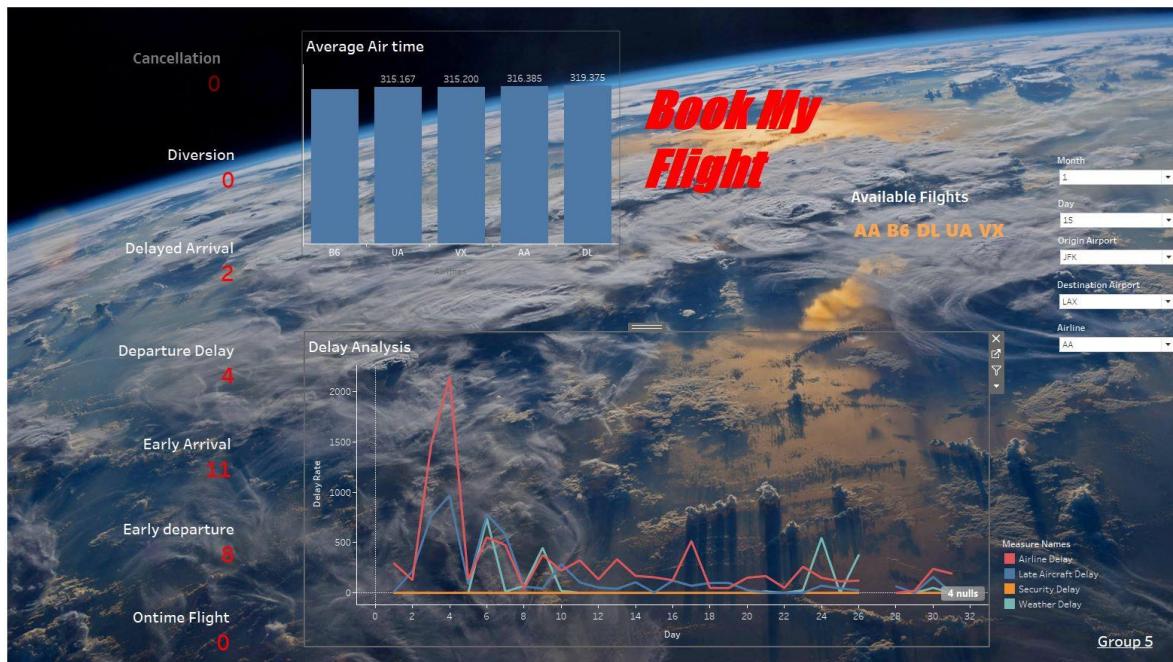
```
+-----+-----+
|airport1|airport2| distance_in_km|
+-----+-----+
| ABE | AVP | 79.891154144279 |
| ABE | PHL | 88.40934907058318 |
| ABE | TTN | 67.51170846955036 |
| ABQ | SAF | 79.652063030968065 |
| ACK | HYA | 49.79288674137177 |
| ACK | MVY | 48.8230058001908 |
| ACT | GRK | 83.1203377785118 |
| ACV | CEC | 88.81902880554524 |
| ACY | TLG | 91.59493341853448 |
| ACY | PHL | 73.16426735892473 |
| ACY | TTN | 93.28676524694741 |
| ALO | CID | 93.88812139885276 |
| ASE | EGE | 46.82459159886323 |
| ASE | GUC | 76.83758560497205 |
| ATW | GRB | 40.015298381501516 |
| AUS | GRK | 97.95324687691345 |
| AVL | GSP | 66.88913528874063 |
| AVP | ABE | 79.891154144279 |
| AVP | BGM | 99.06557851546971 |
| AZO | GRR | 71.86530267371367 |
+-----+-----+
```



VISUALIZATION

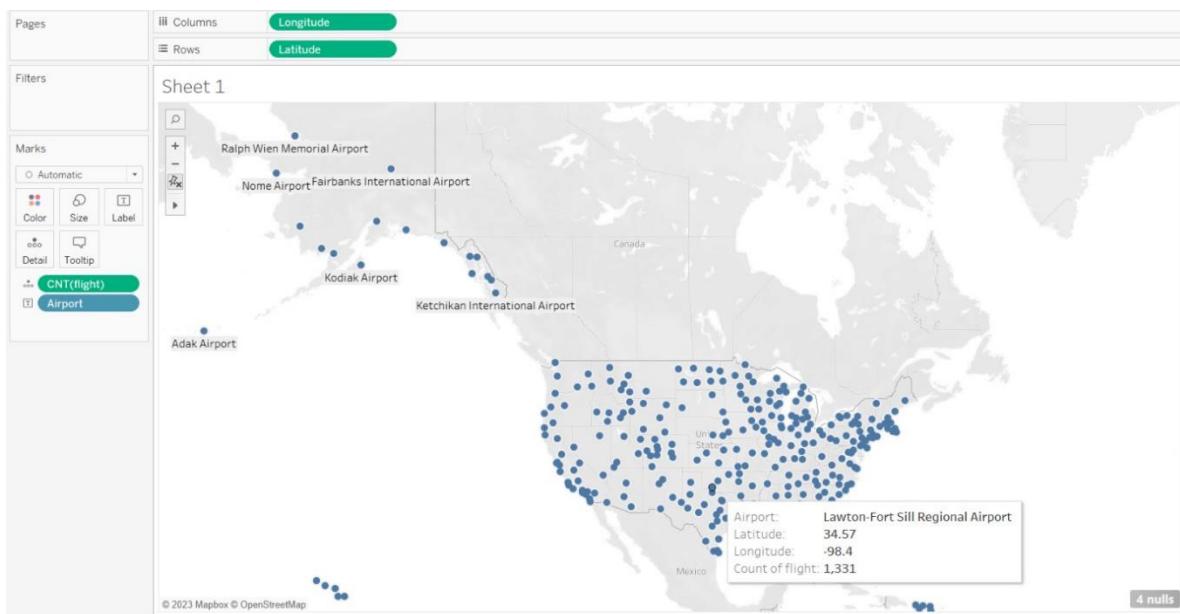
Visualization was done using tableau. A dynamic dashboard for airline and flight analysis is done. In which the users will get to know real-time information about any flight airtime, delays, and cancellations due to specific reasons.

REAL-TIME DASHBOARD FOR FLIGHT BASED ON MONTH AND DAY.

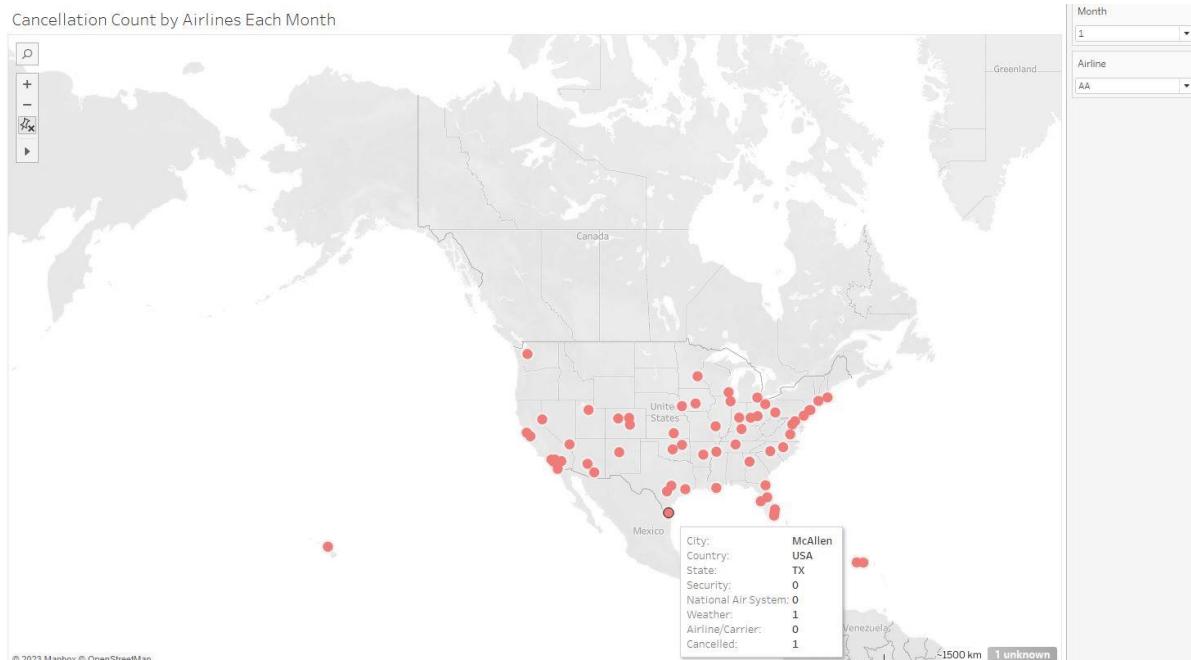


The dashboard is dynamic and provides results regarding a flight delay due to airline-delay, late aircraft delay, security delay, weather delay. Average airtime for each aircraft based on the user mentioned airport is obtained. The dashboard also provides better understanding of how many cancellations was happened to the airline, diversion taken, delayed arrival count, departure delay count, early arrival, early departure and on time flight is gained. Based on the inputs provided by the user, the dashboard updates on the available flights from the origin airport.

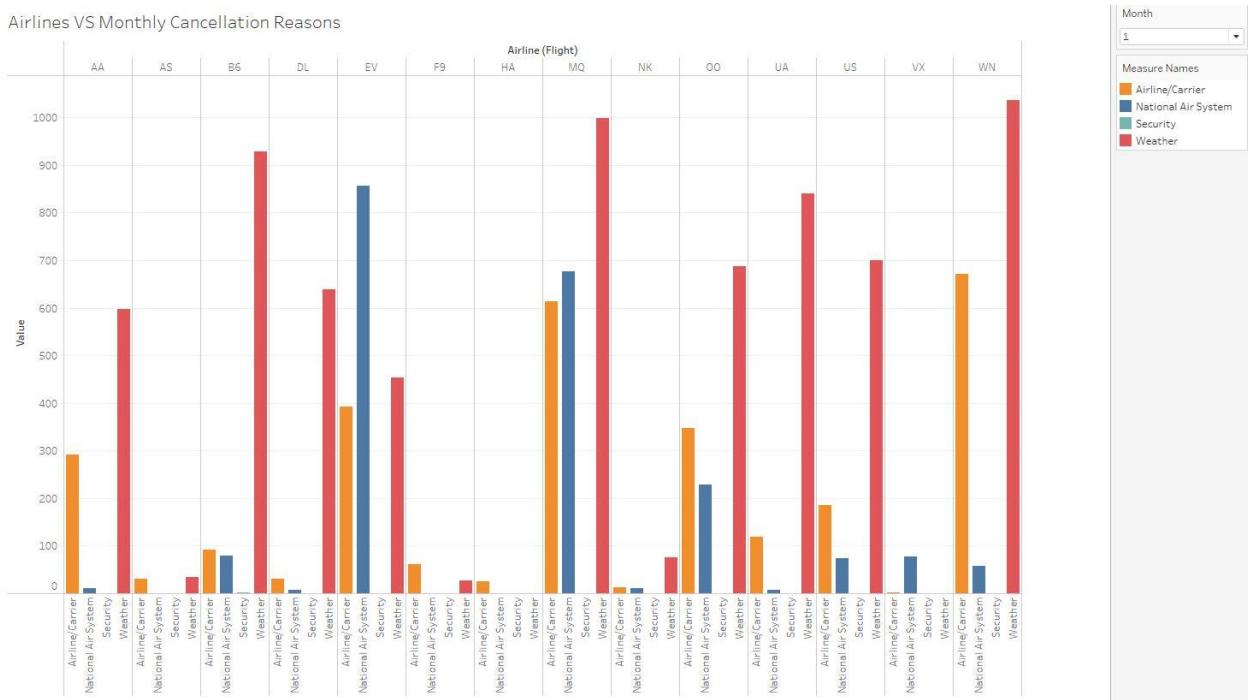
1. Total number of flights from airports



2. Cancellation count by airlines for each month

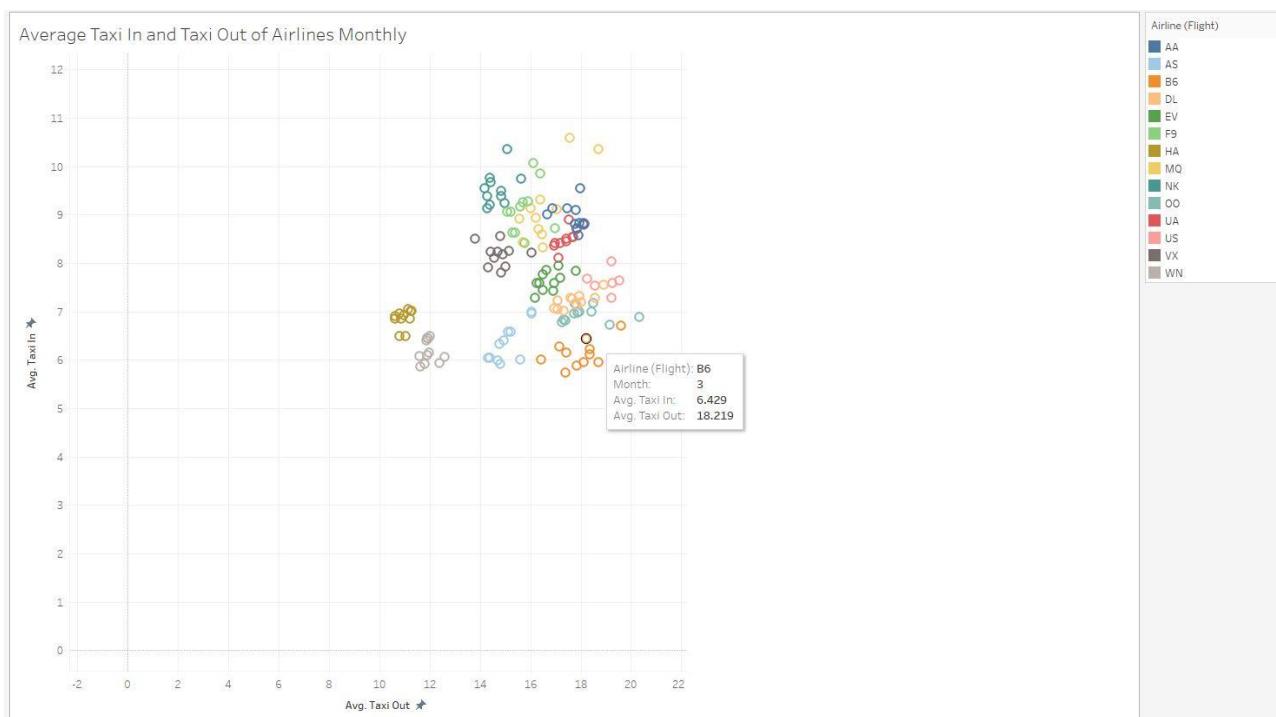


3. Based on monthly updates on cancellation reasons for airline



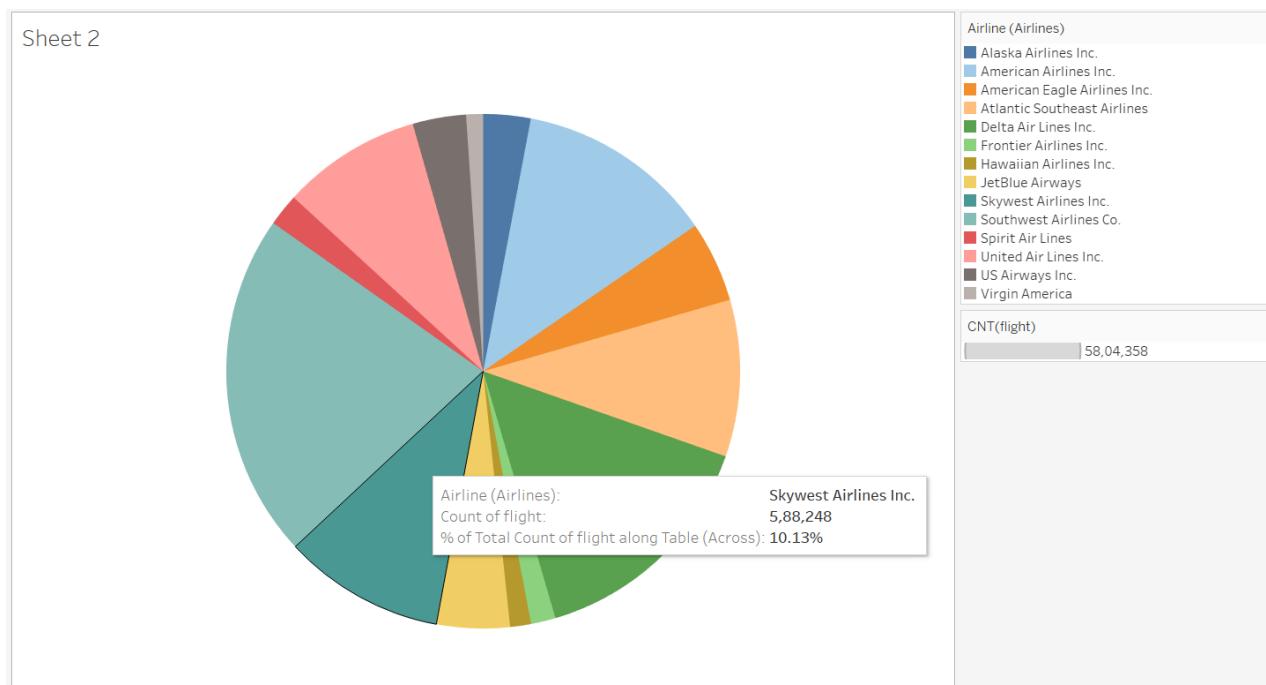
This chart helps to understand the cancellation rate for airlines(flight) which were cancelled.

4. Monthly average Taxi-in and Taxi-out for airlines



The chart shows the average rate of taxi-in and taxi-out for the mentioned airlines during the specific months.

5. Market share of each airline



GITHUB REPOSITORY LINK

<https://github.com/Jerin-T/airline-analysis>

CONCLUSION

On the analysis of the dataset, we could find that delay, cancellations, and diversion are the main challenges and cause serious resource wastages that leads to performance issue in the domestic airlines industry.

We could find a common pattern by airports, airlines and season that are usually delayed. Weather, airline issues and air traffic in each entity are the primary cause of this.

By addressing these issues by identifying the patterns from historic data we could minimize the root cause of the issues and enhance the industry performance up to a great extent.

REFERENCES

1. <https://www.bts.gov/newsroom/2015-annual-and-december-us-airline-traffic-data>
2. <https://www.bts.gov/newsroom/2015-us-based-airline-traffic-data>