Mounica Nandipalli
11323546
INFO_5709

FINAL PROJECT-03

Flight Data Visualization

Introduction:

At present days we know that flights are the most preferred means of transport to reach their destinations. However, we see that most of us face difficulties during their journey due to unpredictable reasons. In order to provide a hassle-free journey to passengers we tried to deliver few visualizations that help in choosing right flights in right time. The following dataset contains information of domestic flights on time performance, their source and destination airports along with certain information regarding delays and cancellations for the year 2015. This dataset is published by U.S Department of Transportation monthly Air Travel Consumer Report. From this data we would like to start off by presenting the flights information from various source and destination airports and then would focus on delays followed by cancellations.

From this dataset our goal is to bring an idea to a passenger on how to choose a flight for a hassle-free journey. Our visualization particularly helps a passenger in avoiding flights which are frequently delayed. They also help in identifying the reasons for cancellation of flights.

Reason for Cancellation of flight: A - Airline/Carrier; B - Weather; C - National Aviation System; D – Security.

Source:

The source of our dataset is from
https://www.kaggle.com/usdot/flight-delays/data#flights.csv

Mounica Nandipalli
11323546
INFO_5709

Attributes:

Major attributes we used to create above dashboards and stories are total numbers of flights, total number of flights cancelled, airports, airlines, delay time and cancelled reasons.

- Airline/Carrier:
  - The cause of the cancellation or delay was due to circumstances within the airline's control (e.g. maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.)
  - B - Weather: Winter storms, lightning storms and strong winds will keep planes on the ground for the safety of everyone involved.

- National Air System:
  - Delays and cancellations attributable to the national aviation system that refer to a broad set of conditions, such as non-extreme weather conditions, airport operations, heavy traffic volume, and air traffic control.

- Security:
  - Delays or cancellations caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas.

Hypothesis:

1. What is Average Departure Delay and Average Arrival Delay of the flights.?
2. To find flight Departure Paths and the Arrival Paths.?
3. To find Arrival no. of flights with states
4. To find the Cancelled flights within a Month and within a week among all the states by airline.?
5. Reasons for cancellation by Airline and by State.?
6. What is the Most used Airlines by the people in all the states in USA.?

Mounica Nandipalli
11323546
INFO_5709

Dashboards:

1. Arrival Delay Dashboard
2. Departure Delay Dashboard

Related Work:

I was searching for a dataset which defines the problems and solutions across searching multiple sites likes Kaggle, Buzzfeed News, Socrata, Repository Machine Learning, etc. Finally, I found this dataset in Kaggle website which provided me the required information which can be used for data processing and generation of needed reports. Even lot of paid articles and need some statistics which can some statistics for calculating the appropriate data. In this dataset there are many attributes like Name of the state, Number of Flights, delayed times within the year. The data which is present in the dataset is as shown below.

Mounica Nandipalli
11323546
INFO_5709

A	B	C	D	E	F	G
IATA_CODE	AIRPORT	CITY	STATE	COUNTRY	LATITUDE	LONGITUDE
ABE	Lehigh Valley	Allentown	PA	USA	40.65236	-75.4404
ABI	Abilene Regi	Abilene	TX	USA	32.41132	-99.6819
ABQ	Albuquerque	Albuquerque	NM	USA	35.04022	-106.60919
ABR	Aberdeen Re	Aberdeen	SD	USA	45.44906	-98.42183
ABY	Southwest G	Albany	GA	USA	31.53552	-84.19447
ACK	Nantucket M	Nantucket	MA	USA	41.25305	-70.06018
ACT	Waco Regior	Waco	TX	USA	31.61129	-97.23052
ACV	Arcata Airpo	Arcata/Eurek	CA	USA	40.97812	-124.10862
ACY	Atlantic City	Atlantic City	NJ	USA	39.45758	-74.57717
ADK	Adak Airport	Adak	AK	USA	51.87796	-176.64603
ADQ	Kodiak Airpo	Kodiak	AK	USA	57.74997	-152.49386
AEX	Alexandria Ir	Alexandria	LA	USA	31.32737	-92.54856
AGS	Augusta Reg	Augusta	GA	USA	33.36996	-81.9645
AKN	King Salmon	King Salmon	AK	USA	58.6768	-156.64922
ALB	Albany Interr	Albany	NY	USA	42.74812	-73.80298
ALO	Waterloo Re	Waterloo	IA	USA	42.55708	-92.40034
AMA	Rick Husbanc	Amarillo	TX	USA	35.21937	-101.70593
ANC	Ted Stevens	Anchorage	AK	USA	61.17432	-149.99619
APN	Alpena Coun	Alpena	MI	USA	45.07807	-83.56029
ASE	Aspen-Pitkin	Aspen	CO	USA	39.22316	-106.86885
ATL	Hartsfield-Ja	Atlanta	GA	USA	33.64044	-84.42694
ATW	Appleton Int	Appleton	WI	USA	44.25741	-88.51948
AUS	Austin-Bergs	Austin	TX	USA	30.19453	-97.66987
AVL	Asheville Re	Asheville	NC	USA	35.43619	-82.54181
AVP	Wilkes-Barre	Wilkes-Barre	PA	USA	41.33815	-75.72427
AZO	Kalamazoo/I	Kalamazoo	MI	USA	42.23488	-85.55206
BDL	Bradley Inter	Windsor Locl	CT	USA	41.93887	-72.68323
BET	Bethel Airpo	Bethel	AK	USA	60.77978	-161.838

Methods:

I needed information across the year 2015, so that the information that has been extracted from the Kaggle website using tableau into Microsoft Excel. The data was cleaned and with no information that which is to be removed. The data has been cleaned and removed the unwanted data. Now information is available in excel. The information is in the table format with proper heading (rows and columns).

Mounica Nandipalli
11323546
INFO_5709

Steps:

1. Getting the information from all the individual datasets from the same repository in Kaggle.
2. Extracting the Delayed flights average time from the dataset?
3. Extracting the Arrival average time from the dataset?
4. To find the most used airlines among the people in the United States.
5. Visualization suits for the cancelled flights in a week and in a month among the states?
6. Calculating the tableau files to get the Dashboards and to find the airports and reason for cancellation?

Representation of the data in Excel using Tableau:

C4 fx 1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	YEAR	MONTH	DAY	DAY_OF_WE	AIRLINE	FLIGHT_NUM	TAIL_NUMBI	ORIGIN_AIR	DESTINATIO	SCHEDULED_	DEPARTURE	DEPARTURE_	TAXI_OUT	WHEELS_OF	SCHEDULED_	ELAPSED_TI	AIR_TIME	DISTANCE	WHEELS_ON	TA
2	2015	1	1	4	AS	98	N407AS	ANC	SEA	5	2354	-11	21	15	205	194	169	1448	404	
3	2015	1	1	4	AA	2336	N3KUAA	LAX	PBI	10	2	-8	12	14	280	279	263	2330	737	
4	2015	1	1	4	US	840	N171US	SFO	CLT	20	18	-2	16	34	286	293	266	2296	800	
5	2015	1	1	4	AA	258	N3HYAA	LAX	MIA	20	15	-5	15	30	285	281	258	2342	748	
6	2015	1	1	4	AS	135	N527AS	SEA	ANC	25	24	-1	11	35	235	215	199	1448	254	
7	2015	1	1	4	DL	806	N3730B	SFO	MSP	25	20	-5	18	38	217	230	206	1589	604	
8	2015	1	1	4	NK	612	N635NK	LAS	MSP	25	19	-6	11	30	181	170	154	1299	504	
9	2015	1	1	4	US	2013	N584UW	LAX	CLT	30	44	14	13	57	273	249	228	2125	745	
10	2015	1	1	4	AA	1112	N3LAAA	SFO	DFW	30	19	-11	17	36	195	193	173	1464	529	
11	2015	1	1	4	DL	1173	N826DN	LAS	ATL	30	33	3	12	45	221	203	186	1747	651	
12	2015	1	1	4	DL	2336	N958DN	DEN	ATL	30	24	-6	12	36	173	149	133	1199	449	
13	2015	1	1	4	AA	1674	N853AA	LAS	MIA	35	27	-8	21	48	268	266	238	2174	746	
14	2015	1	1	4	DL	1434	N547US	LAX	MSP	35	35	0	18	53	214	210	188	1535	601	
15	2015	1	1	4	DL	2324	N3751B	SLC	ATL	40	34	-6	18	52	215	199	176	1590	548	
16	2015	1	1	4	DL	2440	N651DL	SEA	MSP	40	39	-1	28	107	189	198	166	1399	553	
17	2015	1	1	4	AS	108	N309AS	ANC	SEA	45	41	-4	17	58	204	194	173	1448	451	
18	2015	1	1	4	DL	1560	N3743H	ANC	SEA	45	31	-14	25	56	210	200	171	1448	447	
19	2015	1	1	4	UA	1197	N78448	SFO	IAH	48	42	-6	11	53	218	217	199	1635	612	
20	2015	1	1	4	AS	122	N413AS	ANC	PDX	50	46	-4	11	57	215	201	187	1542	504	
21	2015	1	1	4	DL	1670	N806DN	PDX	MSP	50	45	-5	9	54	193	186	171	1426	545	
22	2015	1	1	4	NK	520	N525NK	LAS	MCI	55	120	25	11	131	162	143	128	1139	539	
23	2015	1	1	4	AA	371	N3GXAA	SEA	MIA	100	52	-8	30	122	338	347	311	2724	933	
24	2015	1	1	4	NK	214	N632NK	LAS	DFW	103	102	-1	13	115	147	147	128	1055	523	
25	2015	1	1	4	AA	115	N3CTAA	LAX	MIA	105	103	-2	14	117	286	276	255	2342	832	
26	2015	1	1	4	DL	1450	N671DN	LAS	MSP	105	102	-3	11	113	183	163	150	1299	543	
27	2015	1	1	4	UA	1545	N76517	LAX	IAH	115	112	-3	11	123	183	175	156	1379	559	
28	2015	1	1	4	AS	130	N457AS	FAI	SEA	115	107	-8	25	132	213	218	186	1533	538	
29	2015	1	1	4	NK	597	N528NK	MSP	FLL	115	127	12	14	141	207	220	166	1487	527	
30	2015	1	1	4	US	413	N571UW	LAS	CLT	120	110	-10	12	122	245	224	205	1916	747	
31	2015	1	1	4	AA	2392	N3HRAA	DFW	MIA	120	141	21	12	153	227	208	188	1709	701	

Mounica Nandipalli
11323546
INFO_5709

Data Editor in tableau Desktop:

	Country	DAY	Day Of Week	Destination Airport	Flight Number	Iata Code	IATA CODE (airlines.csv)	Month	Month Names	Origin Airport	Origin-Dest	Path ID	State	Table Name	Tail
ıngeles	USA	26	1	DFW	2463	LAX	AA	1	JANUARY	LAX	LAX-DFW	1	CA	flights.csv	N5
s–Fort Worth	USA	26	1	RSW	188	DFW	AA	1	JANUARY	DFW	DFW-RSW	1	TX	flights.csv	N3
s–Fort Worth	USA	26	1	AUS	1484	DFW	AA	1	JANUARY	DFW	DFW-AUS	1	TX	flights.csv	N5
s–Fort Worth	USA	26	1	PDX	157	DFW	AA	1	JANUARY	DFW	DFW-PDX	1	TX	flights.csv	N5
ıgo	USA	26	1	MSP	1430	ORD	AA	1	JANUARY	ORD	ORD-MSP	1	IL	flights.csv	N4
s–Fort Worth	USA	26	1	STL	166	DFW	AA	1	JANUARY	DFW	DFW-STL	1	TX	flights.csv	N4
ıntonio	USA	26	1	DFW	2272	SAT	AA	1	JANUARY	SAT	SAT-DFW	1	TX	flights.csv	N5
ıngeles	USA	26	1	KOA	247	LAX	AA	1	JANUARY	LAX	LAX-KOA	1	CA	flights.csv	N5
s–Fort Worth	USA	26	1	SAN	49	DFW	AA	1	JANUARY	DFW	DFW-SAN	1	TX	flights.csv	N6
s–Fort Worth	USA	26	1	AUS	1083	DFW	AA	1	JANUARY	DFW	DFW-AUS	1	TX	flights.csv	N5
s–Fort Worth	USA	26	1	HOU	74	DFW	AA	1	JANUARY	DFW	DFW-HOU	1	TX	flights.csv	N5
s–Fort Worth	USA	26	1	ORF	294	DFW	AA	1	JANUARY	DFW	DFW-ORF	1	TX	flights.csv	N5
ıgo	USA	26	1	RSW	2317	ORD	AA	1	JANUARY	ORD	ORD-RSW	1	IL	flights.csv	N3
s–Fort Worth	USA	26	1	STL	1372	DFW	AA	1	JANUARY	DFW	DFW-STL	1	TX	flights.csv	N5
s–Fort Worth	USA	26	1	HDN	1418	DFW	AA	1	JANUARY	DFW	DFW-HDN	1	TX	flights.csv	N3
s–Fort Worth	USA	26	1	HOU	348	DFW	AA	1	JANUARY	DFW	DFW-HOU	1	TX	flights.csv	N4
s–Fort Worth	USA	26	1	MSP	227	DFW	AA	1	JANUARY	DFW	DFW-MSP	1	TX	flights.csv	N4
ıi	USA	26	1	PHX	63	MIA	AA	1	JANUARY	MIA	MIA-PHX	1	FL	flights.csv	N3
ıgo	USA	26	1	SAN	93	ORD	AA	1	JANUARY	ORD	ORD-SAN	1	IL	flights.csv	N3
ıgo	USA	26	1	SAN	936	ORD	AA	1	JANUARY	ORD	ORD-SAN	1	IL	flights.csv	N3
s–Fort Worth	USA	26	1	SFO	221	DFW	AA	1	JANUARY	DFW	DFW-SFO	1	TX	flights.csv	N3
s–Fort Worth	USA	26	1	IND	1385	DFW	AA	1	JANUARY	DFW	DFW-IND	1	TX	flights.csv	N4
s–Fort Worth	USA	26	1	SAN	1015	DFW	AA	1	JANUARY	DFW	DFW-SAN	1	TX	flights.csv	N4
s–Fort Worth	USA	26	1	STL	323	DFW	AA	1	JANUARY	DFW	DFW-STL	1	TX	flights.csv	N5
ıi	USA	26	1	STX	1293	MIA	AA	1	JANUARY	MIA	MIA-STX	1	FL	flights.csv	N3
s–Fort Worth	USA	26	1	TPA	1366	DFW	AA	1	JANUARY	DFW	DFW-TPA	1	TX	flights.csv	N3
s–Fort Worth	USA	26	1	AUS	1600	DFW	AA	1	JANUARY	DFW	DFW-AUS	1	TX	flights.csv	N4
s–Fort Worth	USA	26	1	CLT	2498	DFW	AA	1	JANUARY	DFW	DFW-CLT	1	TX	flights.csv	N3
lulu	USA	26	1	LAX	298	HNL	AA	1	JANUARY	HNL	HNL-LAX	1	HI	flights.csv	N5
ıgo	USA	26	1	PDX	99	ORD	AA	1	JANUARY	ORD	ORD-PDX	1	IL	flights.csv	N3
ıgo	USA	26	1	RSW	1375	ORD	AA	1	JANUARY	ORD	ORD-RSW	1	IL	flights.csv	N3
s–Fort Worth	USA	26	1	SAT	2418	DFW	AA	1	JANUARY	DFW	DFW-SAT	1	TX	flights.csv	N5
ıgo	USA	26	1	SFO	1459	ORD	AA	1	JANUARY	ORD	ORD-SFO	1	IL	flights.csv	N3
s–Fort Worth	USA	26	1	SJC	189	DFW	AA	1	JANUARY	DFW	DFW-SJC	1	TX	flights.csv	N5

Exploratory data analysis:

Analysis:

To start our analysis, we began by looking at arrivals and departures of different airlines from various airports. We used paths, filters, symbol maps, ranks, few calculated fields, images to create two dashboards to get a clear idea on departure delays and arrival delays of various flights. In departure delay dashboard and arrival delay dashboard we get clear idea on delay times of all the 31,200 airlines starting from 323 airports.

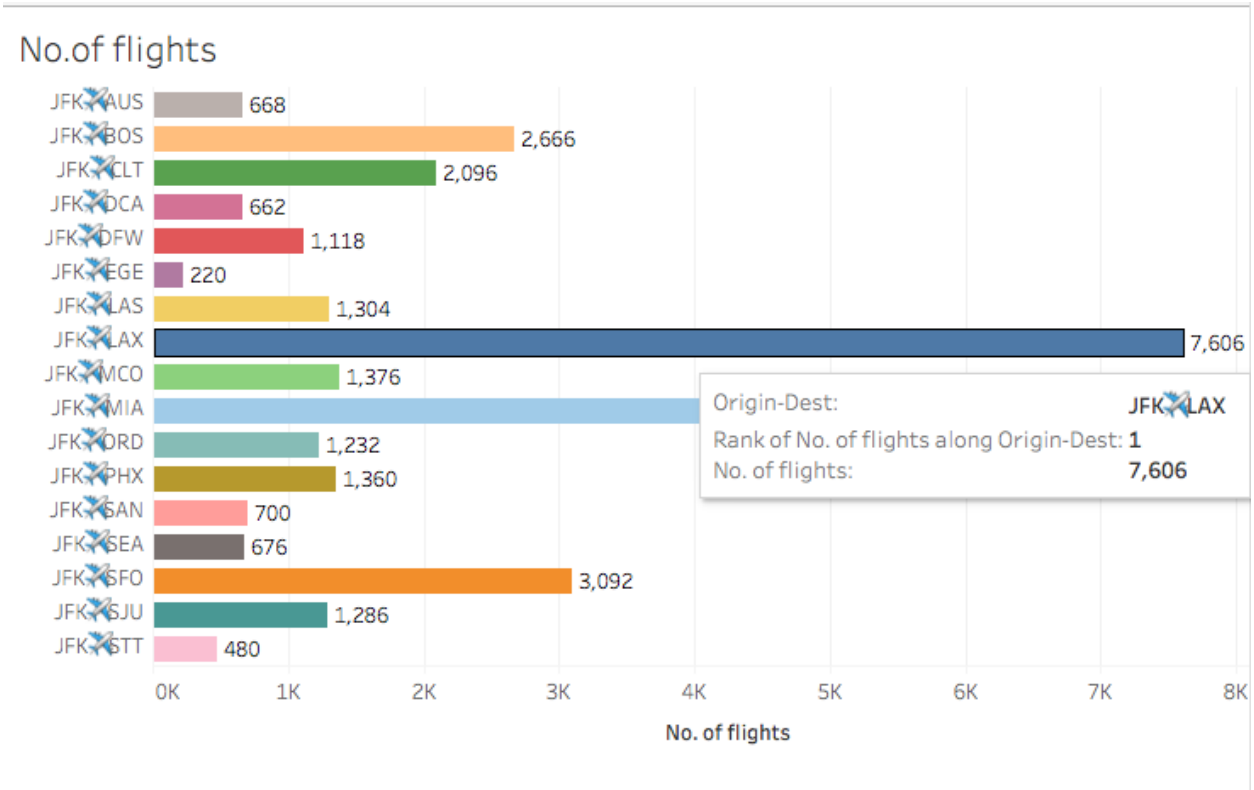
After determining delay times from dash boards, we decided to excavate further to bring a relation between timeframes and cancellations with respect to week and a year using trend lines. We also tried to develop a report on airline cancellations based on cancellation reason for each state using graphs.

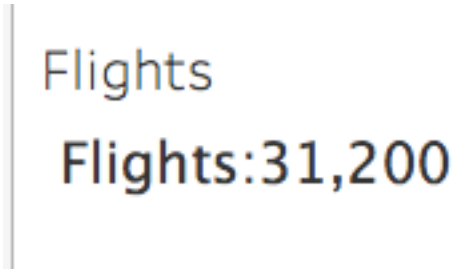
Mounica Nandipalli
11323546
INFO_5709

1. **Total no of flights in the United States 2015?**

   While looking at the data for good visualization which tells us about the increase or decrease in the number of passengers in the country. I found there is a lot of decrease in the passengers who took often flights for travelling. I used bar graphs which gives a clear understanding with detailed number of flights and ranking.

This visualization shows that JFK and LAX have highest no of flights among all Ranking Position 1

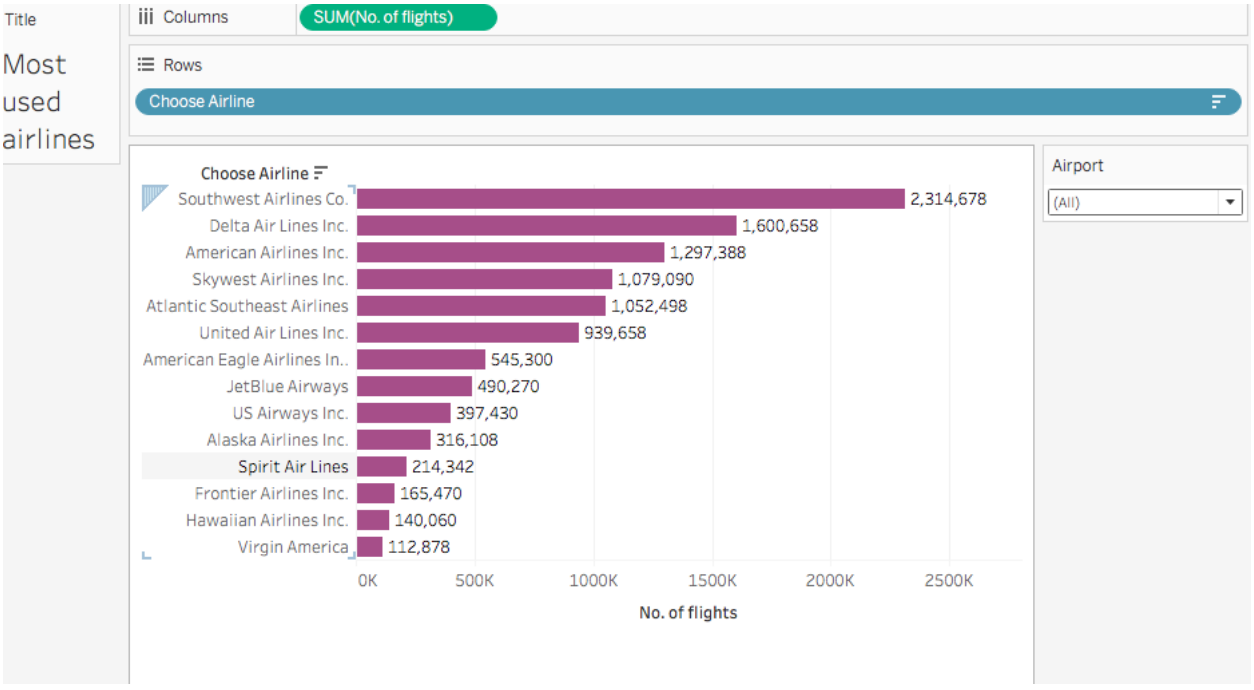




Overall no of flights

2. **Overview of flight services provided by airlines at given all the airports in the dataset.?**

This gives the visualization of over all the states in the United States and in which this shows that the Southwest Airways has the highest no of flights among all the airlines in the united states and shows that Virgin America has the least no of flights among the airlines.

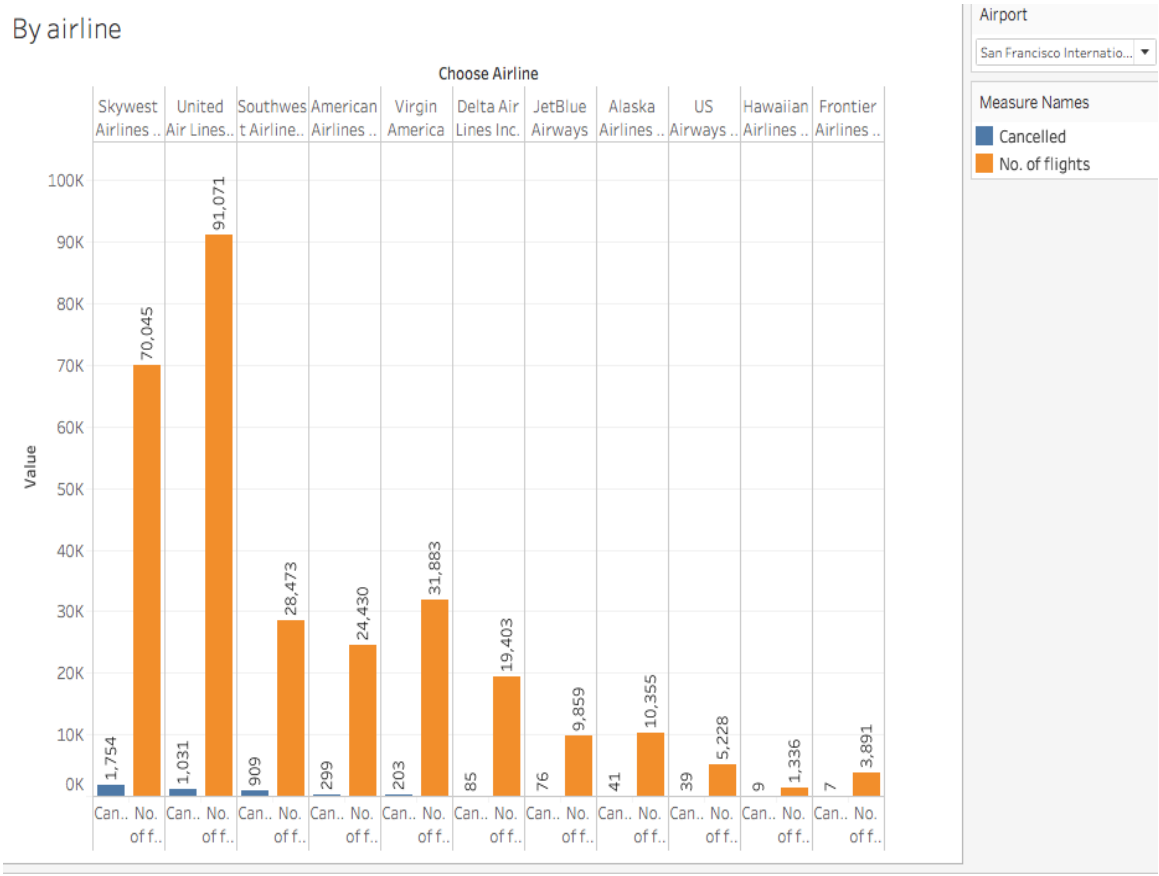


The bar chart shows that 70% of the flight services are provided by the Southwest Airlines Co. with a count of 2,314,678 flights in a year and also provides the least flight services of Virgin America by 112,878 flights in a year.

3. **Airways Cancellations**

a. We could observe number of flights, number of cancelled flights for each airline at a given airport.

This visualization of bar graph shows that the United Air. Lines has the highest no of flights among all the airlines in the united states. And shows that Skywest Airlines has the highest no of Cancelled flights in the united states in the following year of 2015.



b. The below trend lines show number of cancelled flights in a day

The above visualization shows that in a week Sunday has the highest no of cancellation followed by Monday, I have choose circles because this gives the clear understanding of the data in different unique colors which will be easy for understanding for the users.

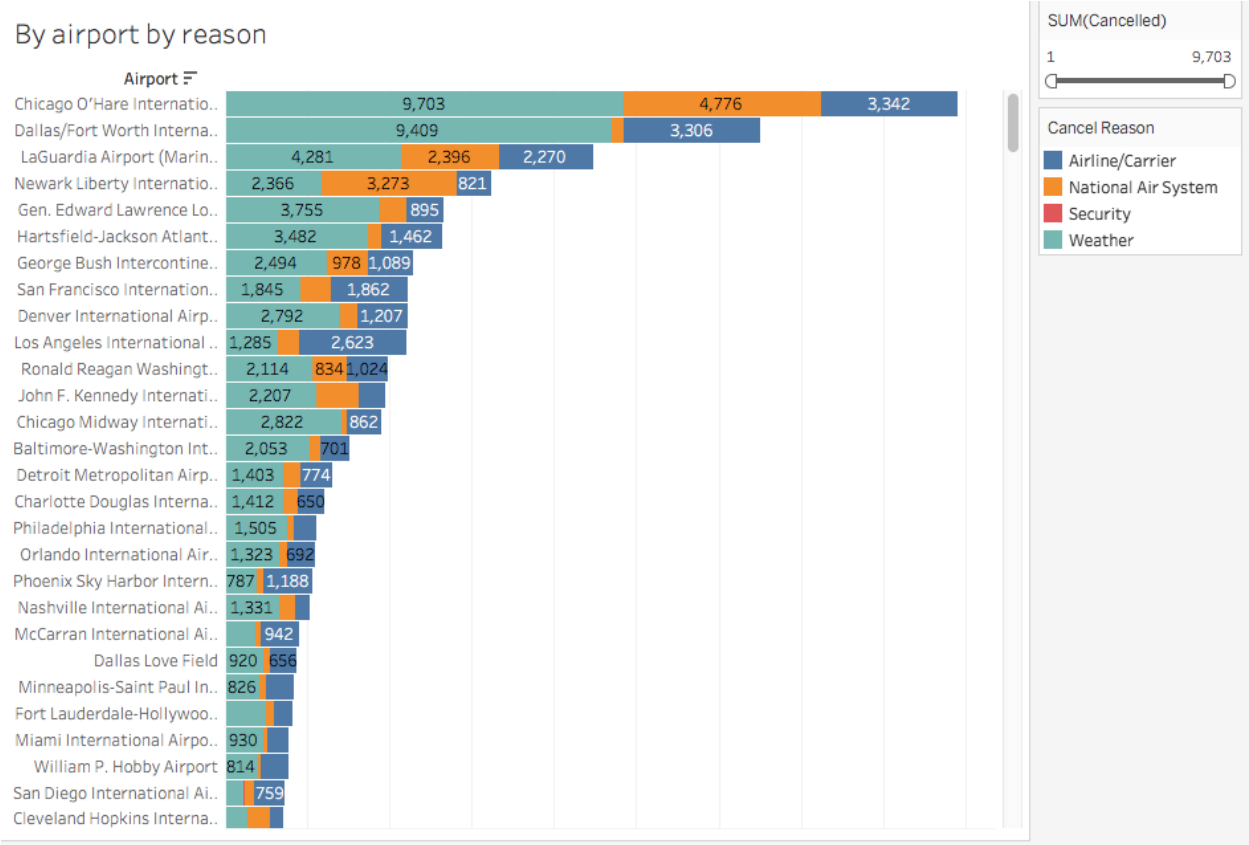
c. The below trend lines show number of cancelled flights in a month?

This below circle shows the visualization part of flights that are cancelled in a month and I have besides I have given the reasons of which way the flights are cancelled in the following year of 2015. And I have did small calculation to find out the exact count pf all the flights in a month which the stats show that February has the highest no of cancelled flights whereas September has the least count of cancelled flights. This circle graph gives the good visuals of clear understanding.

Reasons of Cancellation:

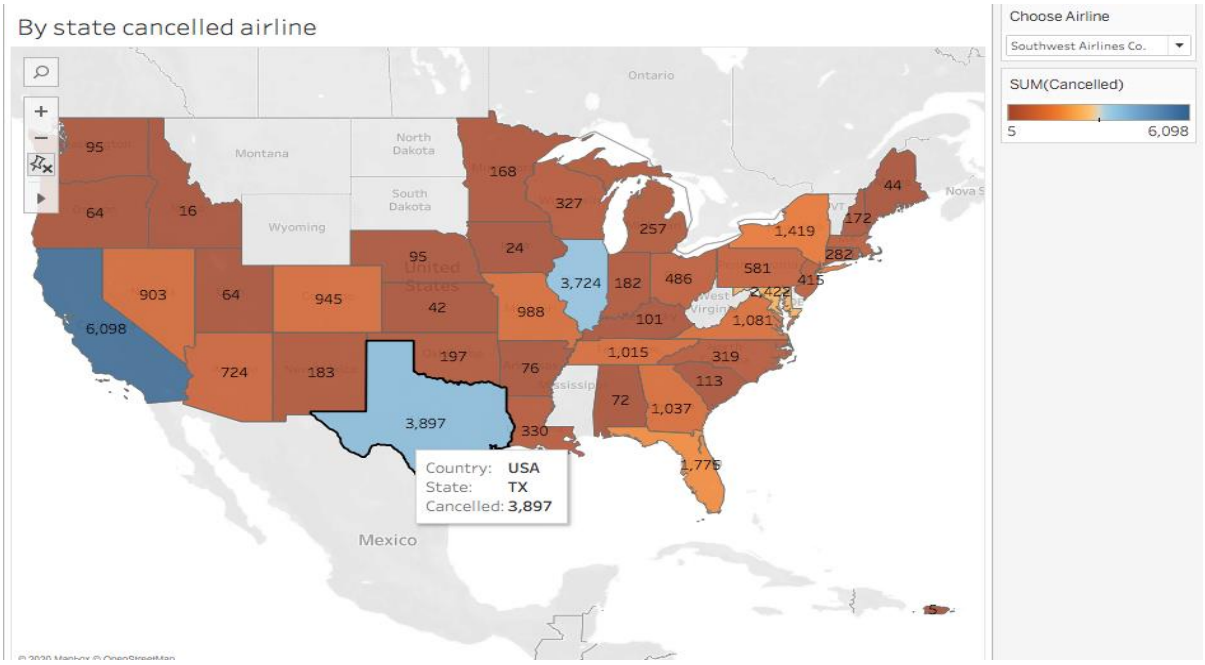
d. By airport cancellation

The below bar graph shows that the Chicago O'Hare International Airport has the highest no of cancellation among all the airports in the united states. This shows that the weather is the main cause of reason where the flights are cancelled at Chicago airport. That which is followed by the Cleveland airport. There by it shows that 9703 flights have been cancelled at Chicago Airport due to weather and 4776 by National Air Security reasons and followed by 3342 by Airline/Carrier. This shows that the weather is the main reason for all the states in the united states irrespective of the other reasons.



e. By state Cancellation:

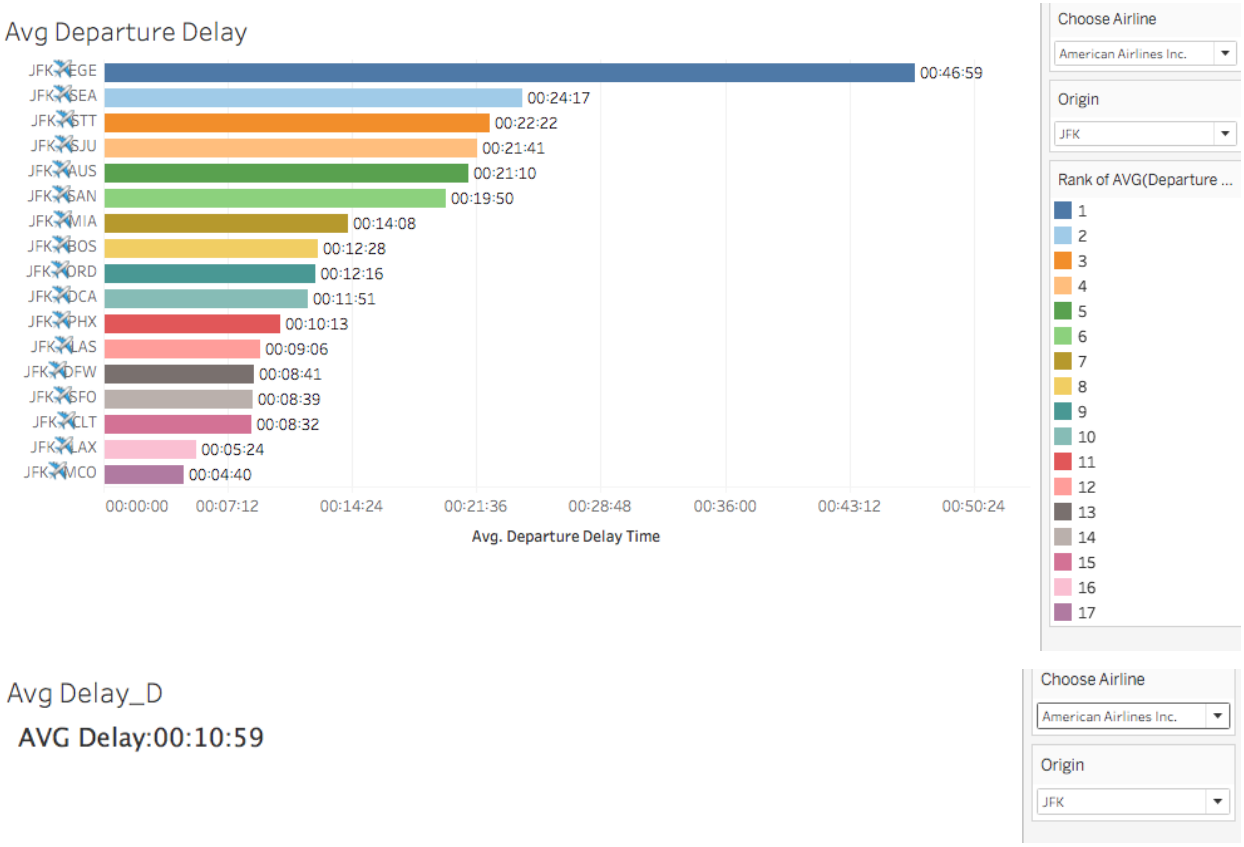
The map of the united states shows that the Texas has the highest number of flights cancelled among all the other states. This depends on the flight schedule from the other states as well due to weather and the National Air Security reasons. The visualization gives a clear understanding of all the states which number of flights cancelled in a year and which makes a proper understanding.



4. **Delay in Airways**

a. Average Departure Delay

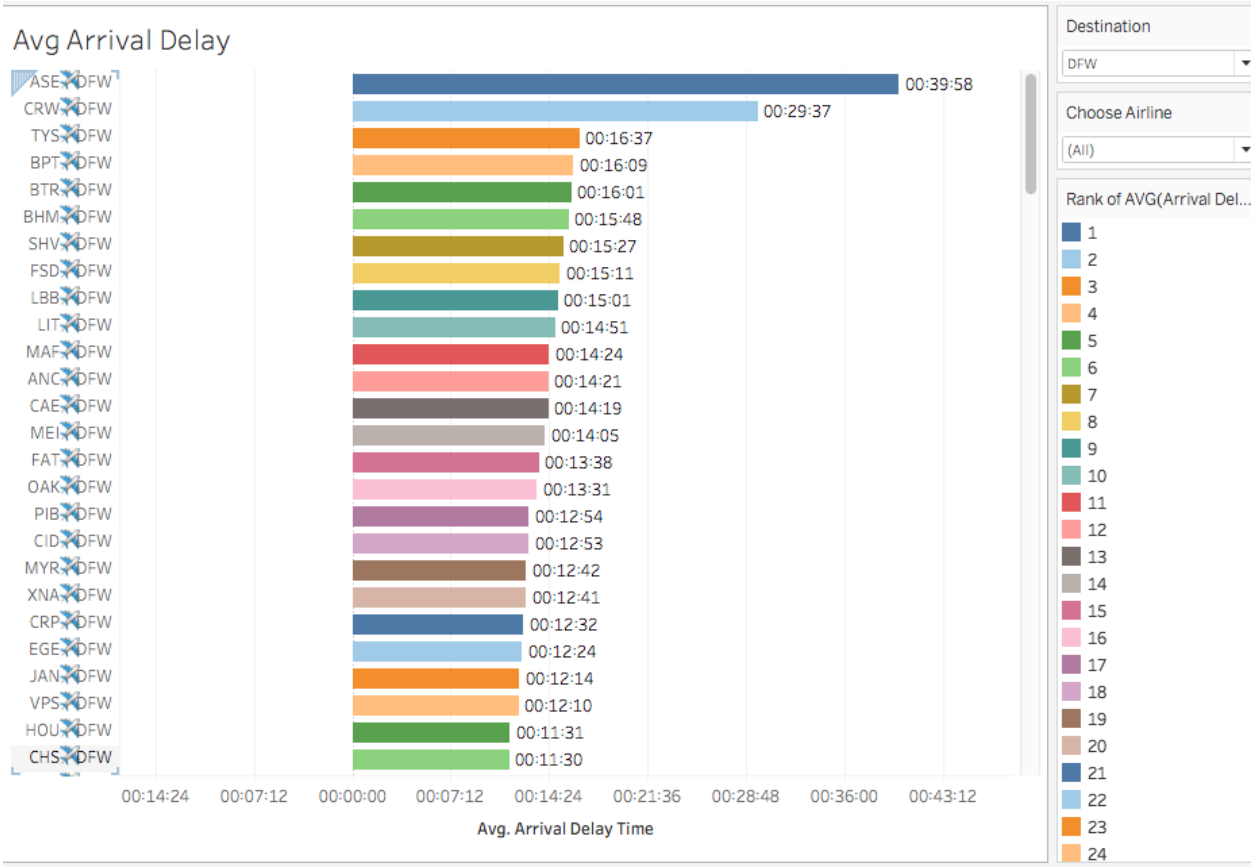
This bar graph gives the avg departure delay of American Airlines flights in among the states in the United States. Shows that JFK-EGE have the height delay of 46:59 secs at JFK airport



This is the Ave Departure Delay of all the American Airlines flights at JFK in 2015.

b. Average Arrival Delay:

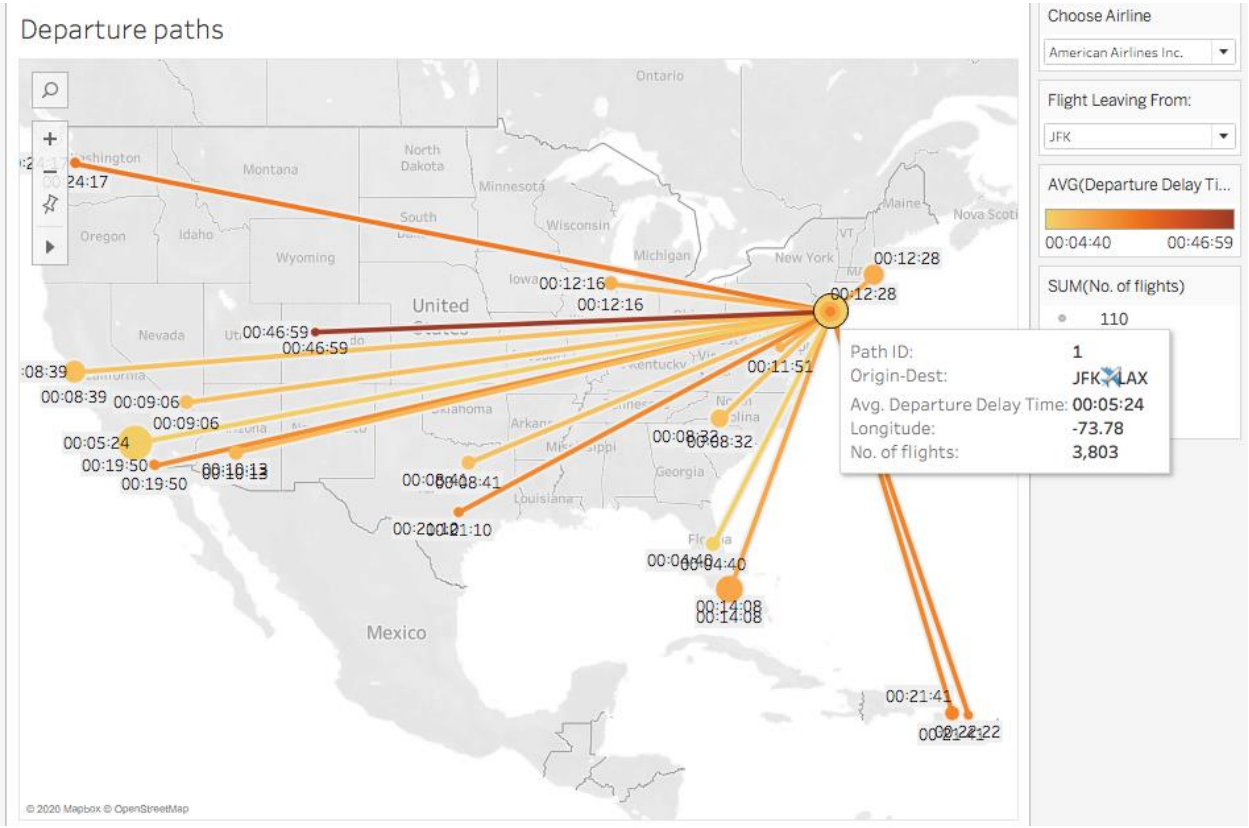
This bar graph shows that the ASR-DFW airport have the highest no of arrival delay in the flights with 39:58 secs. This shows the highest no of flights among all the other state connecting flights at DFW Airport.



This shows the Avg Delay in Arrival of all the flights at DFW in 2015.

c. Departure Paths:

The below graph of lines shows the Departure paths of all the American Airlines flights in 2015 from origin to destination. This inturn, shows the longitude of the flight and the path id for easy recognition of the flights at JFK Airport. This gives a detailed description of all the flights of AA with delay in time from origin to destination at JFK.



D. Arrival Paths:

The below graph of lines shows the Departure paths of all the Airlines flights in 2015 from origin to destination at DFW. This in turn, shows the longitude and latitudes of the flight and the path id for easy recognition of the flights at DFW Airport. This gives a detailed description of all the flights with delay in time from origin to destination at DFW.

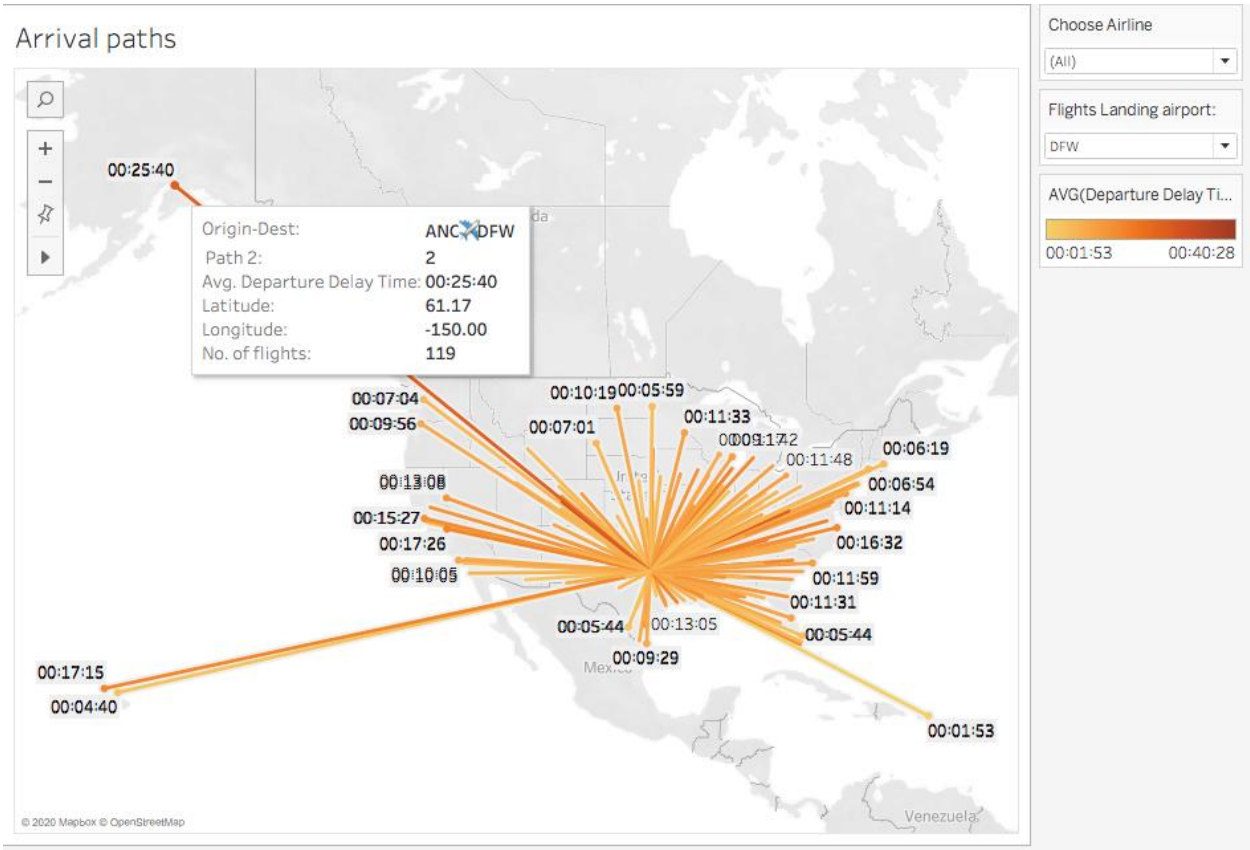


Tableau Functionalities:

Calculated fields: Following are the various calculated fields we used to create dashboards and story.

1. Arrival delay time:

   Arrival delay time/1440

2. Cancelled Reason:

   CASE [Cancellation Reason]

   WHEN "A" THEN "Airline/Carrier"

   WHEN "B" THEN "Weather"

   WHEN "C" THEN "National Air System"

   WHEN "D" THEN "Security"

   END

3. Orgin - Dest:

   [Origin Airport] +"✈"+ [Destination Airport]

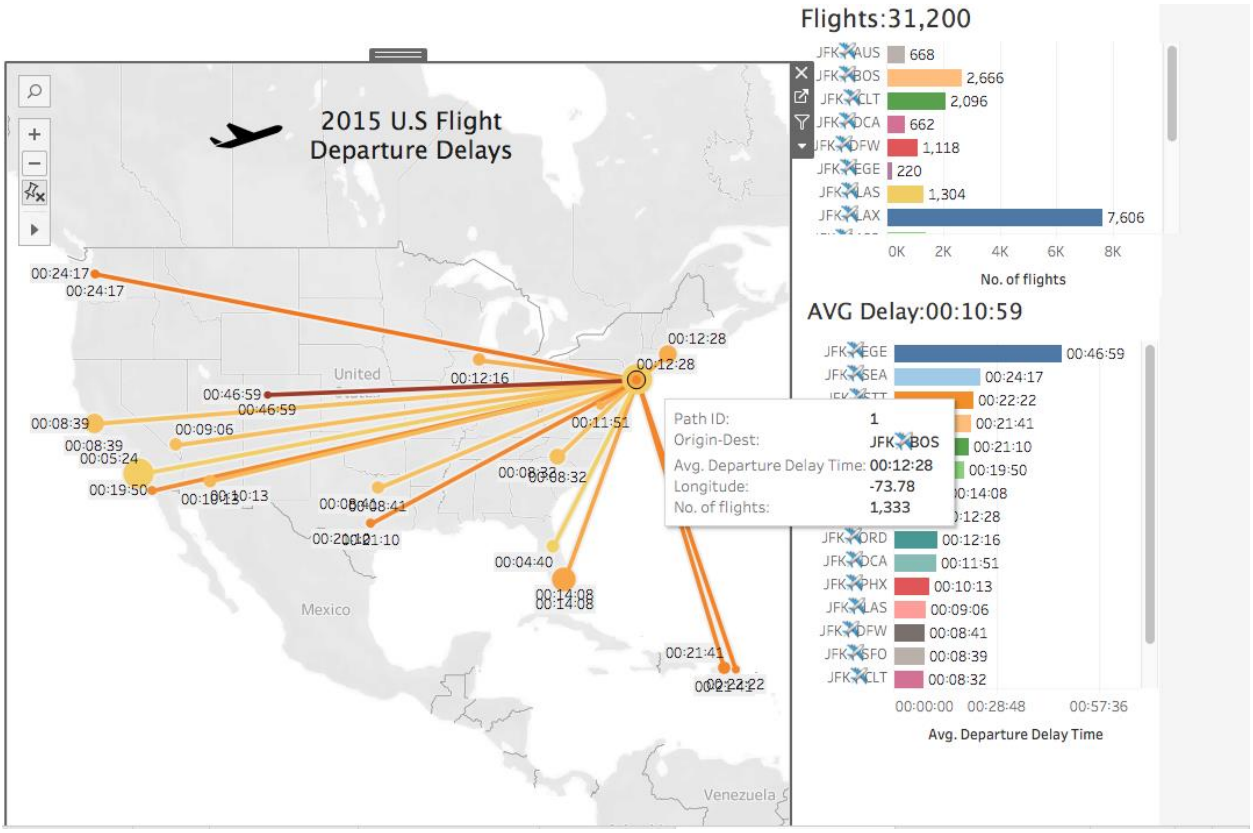
4. Path ID:

   if RIGHT ([Table Name],1) == "1" THEN 2 ELSE 1 END

Dashboards:

1. **2015 U.S Flights Departure Delays**

The following Dashboard gives the lines graph with clear understanding of all the departure AA flights in 2015 at JFK. With total no. of flights of 31,200 with avg delay of 10:59 secs.

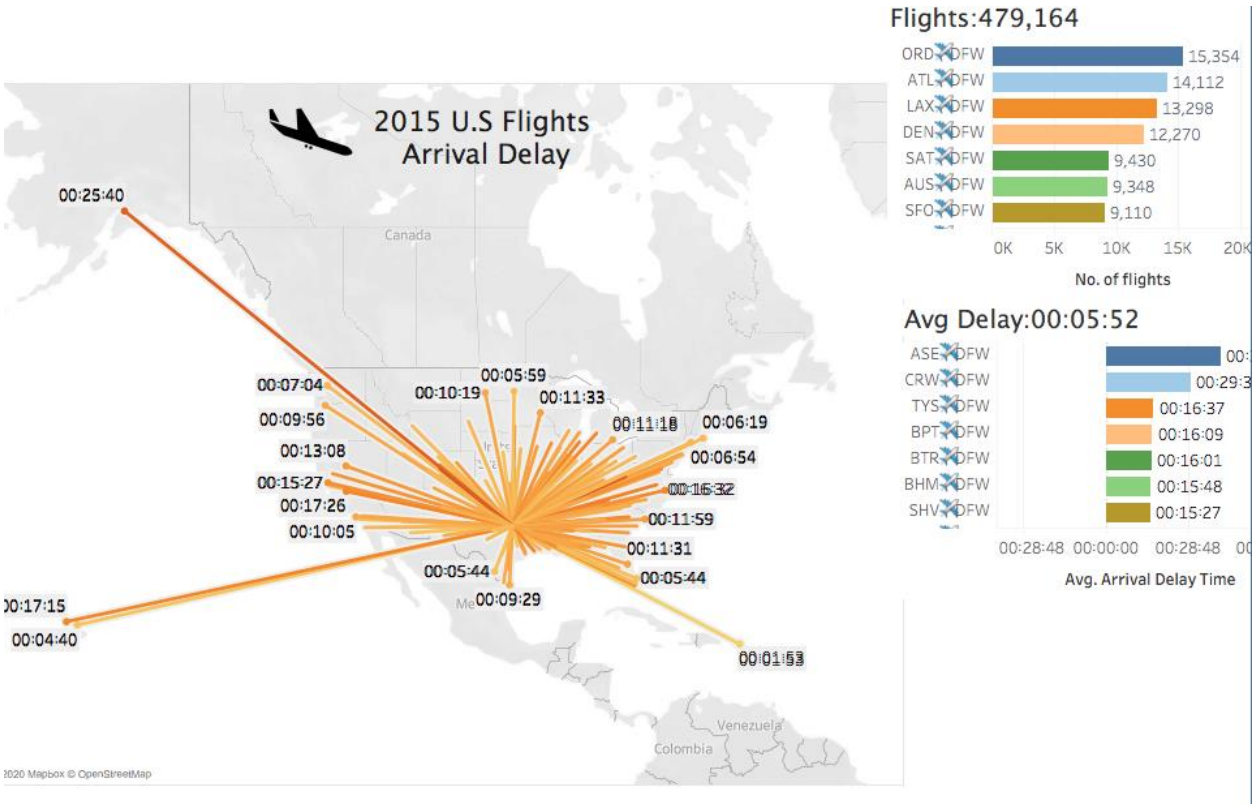


Mounica Nandipalli
11323546
INFO_5709

Here in this dashboard we can see all the flights that are originated at JFK to different states and with the path id as well the delay in all the states with times. Ave time is given as 10:59 secs for all the American Airway flights at 2015 in JKF Airport.

2. 2015 U.S Flights Arrival Delays

The following Dashboard gives the lines graph with clear understanding of all the departure flights at 2015 in DFW. With total no. of flights of 479,164 wit avg delay of 5:52 secs.

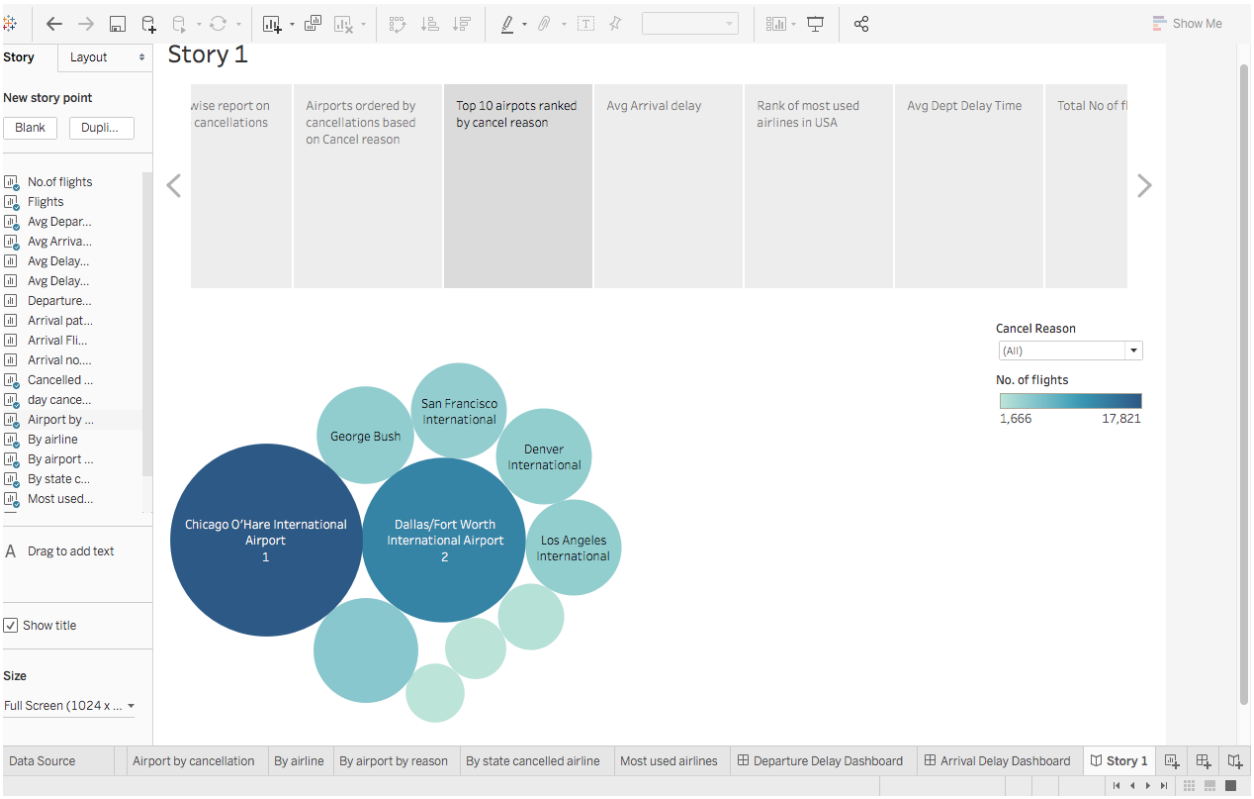


Mounica Nandipalli
11323546
INFO_5709

Here in this dashboard we can see all the flights that are originated at DFW to different states and with the path id as well the delay in all the states with times. Ave time is given as 5:52 secs for all the airlines flights at 2015 in DFW Airport.

3. Story Dashboard:

This dashboard gives all the information in the dataset from the information which need to be predicted to get the desired results.



Discussion:

Using these charts, we can learn about the number of flights in the United States at 2015. And how many flights have been cancelled and delayed based on the reasons that are given in the dataset from various reasons. These reasons of the airlines and originated from different states from the locations. These give a clear understanding from all the airlines from the different positions in the united states. These allows the passengers to choose the flights accordingly and then the correct airways for the destinated travels.

Future Work:

As this entire process is automated it can be easily be extended, and charts can be generated and processed easily. This work can be extended for every year and generate charts easily for human use. Similar type of reports can be generated for all the half-yearly data also to have better analysis and compare between them. These charts can have been better impact, and it shows impacts from the locations and the origin states in the data.

Conclusion:

The project primarily focuses on developing a decision-making application for passengers to choose a particular airline of their choice with respect to delays and cancellations. From our project we were able to show no of flights for a particular airline along with number of flights cancelled respectively. we observed that weather conditions are one of the major reasons for flight cancellations.

References:

https://www.kaggle.com/usdot/flight-delays/data#flights.csv

https://konstantingreger.net/connecting-the-dots-visualizing-paths-in-tableau/

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