US Airlines Delay Analysis Report

Group 2

Group member:

Liweiwen Zhou 24100792,

Chen Yao 24043026,

Haoran Shi 22694199,

Yuanhao Ni 23986051,

Zhan Cheng 22547018

Abstract

The aviation industry plays a vital role in the United States of America, and the airline's punctuality is crucial to both the country's economy and transportation. There are a few key factors that contributed to airlines' delays in the US. In this article we analysed the airline delay data of the US during February of 2020 and their crossbedding airports' data. Our findings indicate that the three main delay factors in February 2020 were National Airspace System (NAS) reasons, late arrival aircraft and airline reasons. The proposed strategy is to improve the NAS by implementing upgrades. However, there may be challenges and potential delays associated with the development and maintenance of the system. Despite short-term delays, these upgrades have the potential to benefit the US aviation industry in the long run.

Keywords: US · Airlines Delay · Delay factors Analysis· Proposed

---

The detailed code, plot and data can be found in our GitHub project: https://github.com/WilliamZLee/US-Airline-2020-delay

Introduction

The aviation industry plays a vital role in the global economy and customer satisfaction is affected by the airlines' punctuality as well as accountability and transparent delay reasons. Economic growth usually led to higher average yearly distances travelled, as well as higher air traffic volumes, robustly observed among several regions worldwide until 2020. (Mitsokapas & Beck, 2021). To better solve the potential delay problems, this report analyses past datasets to evaluate the airline delay causes and different airports in February 2020.

The report considers seven delay factors: airlines caused delay named as 'carrier', weather caused delay named as 'weather', National Airspace System caused delay named as 'NAS, security issues caused delay named as 'security', late arrival aircraft caused delay named as 'late_aircraft_delay', arrival flights cancelled caused delay named as 'arr_cancelled' and arrival flights diverted caused delay named as 'arr_diverted'. We examined them from their counts and their delayed minutes. Also, to differ delayed airlines, we set up a baseline of airlines delayed by 15 minutes or more which been categorised as delayed, we named them as 'arr_del15'. The total flights were named as 'arr_flights'.

The remainder of the article is divided into five sections. We first analyzed the leading factors and got a few delay control strategies from well performed airlines and airports. The we worked on the correlation between each factor and how geographical factors affect delays, in this section we also gave a few suggestions to airports. The next part is about a few plots output, then we gave details of how we generated our datasets and the method used in conducting our analysis. Finally, we gave a potential strategy to reduce flight delays in the US and a conclusion of our report.

Leading factors analysis and best delay-controllers

Leading factors analysis

Through our analysis result, we surprisingly discovered that the most substantial delays in each state are contributed by three main factors: National Airspace System (NAS) delay, carrier caused delay and late aircraft delay. In 24 US states, carrier-caused delay is the main factor causing airline delays, 18 states suffered mainly from aircraft arriving late, while in the rest 8 states, airlines main delay is due to National Airspace System (See apemdix-1). The following figures show how these 3 factors affected airline delays in each state.

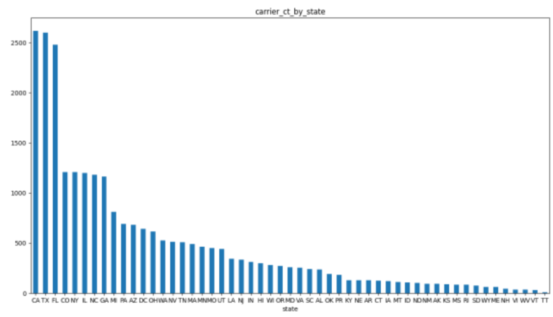

Figure1: Carriers delay

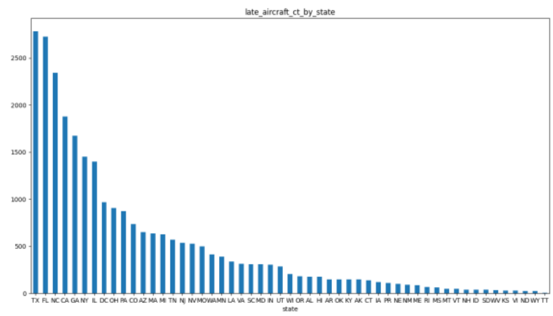

Figure2: Late aircraft delay

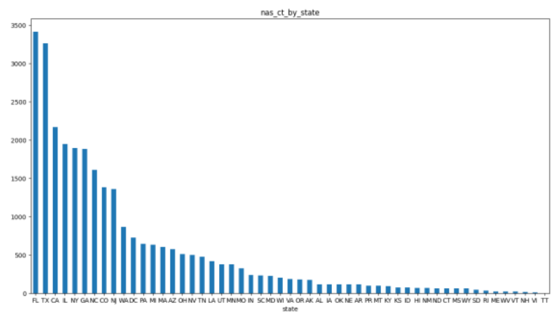

Figure3: NAS caused delay

To examine the highest delay severity state, we chose to examine the severity by delay-punctual rate, we generated a new data frame with state sorted from highest to lowest, and found that the most serious delay during February 2020 in US happened in West Virginia by 21.5% while the best performed state in delay control was Hawaii by 5.9%. Additionally, except for Hawaii, the delay-punctual rate was above 10% in the rest 49 states.

As airlines are also victims of delayed flights, we discovered through our analysis that the most delays caused by carriers were in California, Texas, Florida, Colorado, New York state, Illinois, North Carolina and Georgia were all above 1100 times delayed during February

2020. In our data set, there were 17 different airlines, we sorted them by ‘carrier_ct’, which resulted that there were 8 airlines that got more than 1000 delayed flights in February. However, as the number of delayed aircraft is linearly related to total flights, we used the delay-punctual rate to represent the delay control level. The following bar charts represent these 17 airlines' delay control performances and a combined plot of ranked airlines' delay count by 7 delay factors.

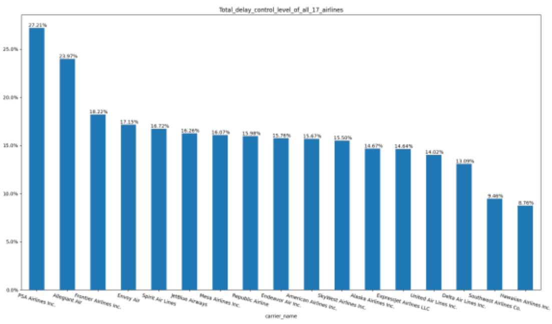

Figure4: Control level of 17 Airlines

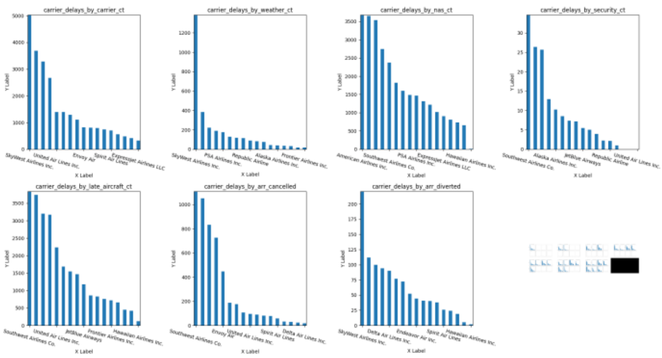

Figure5: Airlines affected by 7 delay factors

Figure 4 shows the ranked control level of all 17 Airlines in US during this month, we can see that only two Airlines controlled their delay-punctual rate under 10% which were: Hawaiian Airlines Inc. by 8.76% and Southwest Airlines Co. by 9.46% while PSA Airlines Inc. got 27.21% of delays in their February’s total arrived flights. But as mentioned earlier, to make this conclusion more statistically significant, we’ll introduce a total number of flights as an additional judgement later, but before that, we’ll analyse the relation between 7 delay factors and their impact on the airline's delay control performances.

From Figure5 we can see that SkyWest Airlines Inc. suffered from carrier-caused delays, weather-caused delays and diverted arrivals while Southwest Airlines Inc. underperformed in controlling delays caused by security, late arrival and arrival flights cancelled. The airline that suffered the most from delays caused by NAS was American Airlines Inc.

Best delay-controllers

After analysis of the main delay factors, we tried to find who performed the best in controlling airlines' delays. Here we introduce flights that arrived to represent the size of airports to avoid small airports which got non-statistically significant value. From airlines and airports, we both chose the top 6 best delay controllers to present the next step analysis.

By setting a filter of total arrived flights upper than 8000 when examining airports, we got a list of top-size airports in the US. The bar charts below show the name and their control levels.

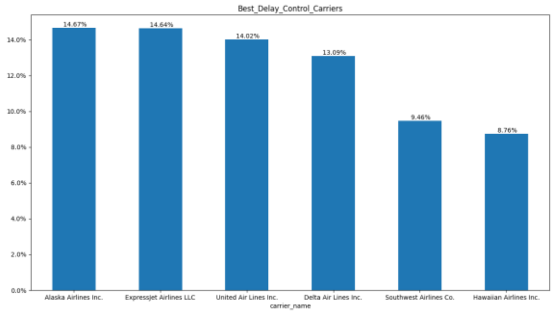

Figure6: Top6 delay control Airlines

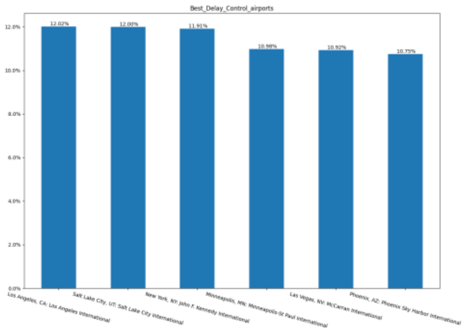

Figure7: Top6 delay control Airports

From these 6 airlines, we found that Southwest Airlines Co. performed the best as they got 101,581 flights this month and only delayed by 9.46%, and their best-controlled factor was security. The second-best airline was Hawaiian Airlines Inc., during the whole of February they only got 2 flights diverted (See appendix-2, appendix-3). As for airports, the 6 best delay controllers were: Phoenix Sky Harbor International, McCarran International, Minneapolis-St Paul International, John F. Kennedy International, Salt Lake City International and Los Angeles International Airports. Among them, Phoenix Sky Harbor International Airport got 14557 flights and only 10.7% of them got delayed. During the whole month, this airport only got 4 flights delayed by security, 7 flights delayed by diverted arrival and below 30 flights delayed by weather. Convincingly, these airports all performed well in controlling security factors, which caused delays (See appendix-4, appendix-5).

Correlation between factors and Geographical Factors

Correlation analysis

Here we discuss the correlation between these seven main factors and the phenomenon of aircraft delays caused by geographical factors. Based on the correlation matrix (see appendix-6), we reached the result that the weather factor is strongly correlated with flight cancellations, and delays due to security, routing, and NAS. Safety reasons and flight cancellations show a strong correlation. The relationship between route delays and safety and flight cancellations also deserves to be looked at. By generating a heatmap of the correlation matrix (Figure 8), the same result is also shown in the visualization.

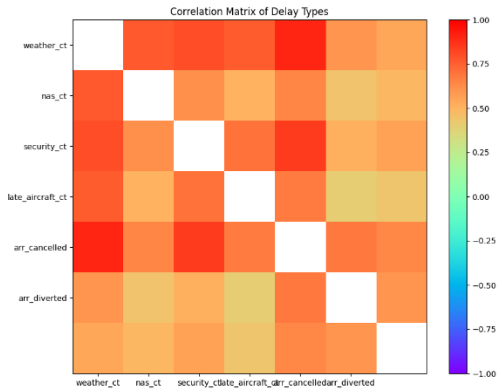

Figure 8: Correlation matrix heatmap

From Figure 8 we can see that weather factors are inverted to be the main factor for all kinds of flight delays and cancellations. There is also a large reason why flights are cancelled due to security reasons. Delays in flights are usually associated with safety, flight cancellations and NAS in no small way. Rerouted arrivals have a very low probability of causing flight delays, but a slight probability of causing flight cancellations.

Geographical factors analysis

In our further analysis of geographical factors, which to be specific is to sort airports by state, we found that in the top 10 most delayed airports, 7 of them are located in the eastern part of the US, while the rest 3 airports are all located around the middle part of the US. The most delayed airport was Charlotte Douglas International located in Charlotte, North Carolina with a delay rate of 20.4%, while in Newark Liberty International, New Jersey the rate was 20.2% where this airport was also located in the eastern part of the US.

From the output of codes (See appendix-7) we can see that the main reason which contributed to the delay was late aircraft arrival and NAS. As in the eastern US, particularly in these mentioned states, significant and sustained snowfall could occur in February due to extreme winter weather caused by global warming. Surprisingly, however, over the past two to three decades, the increase in extreme weather has included more (not fewer) severe cold-air outbreaks and heavy snowfalls observed both in North America and Eurasia we reached the result that such location affected a lot in weather which caused the airlines' delay (Cohen et al., 2018). As it's hard for humans to change persistent extreme weather, the suggestions we came up with could only focus on airport infrastructures. One is the construction of more meteorological warning facilities, by using more atmospheric sensors and remote sensing satellites to predict widespread snowfall in winter. The other potential solution is airports in the eastern and middle region of the US should construct more runway de-icing facilities, temperature-controlled aviation fuel storage facilities, temperature-controlled maintenance aprons and one or two landing strips to prevent icing or waterlogging of the main runway. The last suggestion is that during the winter months, airlines should deploy additional ground staff, maintenance engineers and tower liaison officers at airports to ensure the safety and successful take-off and landing of standby aircraft.

Heatmap and Pie charts

In this section, we went further into geographical factors by filtering the airports within the region of the latitude range of 32° to 37° and the longitude range of -100° to -80°. The weather in this region is slightly different, the northern and mid-western part of this region constantly shows snowy and frozen weather in February while in the southern part of this region, the weather is warmer. As this also indicated the conclusion we reached in the last section, we generated two figures for airports within this region.

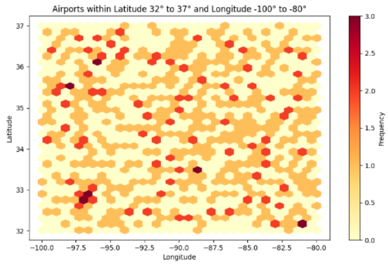

Figure 9: Hexagonal binning plots of filter airports

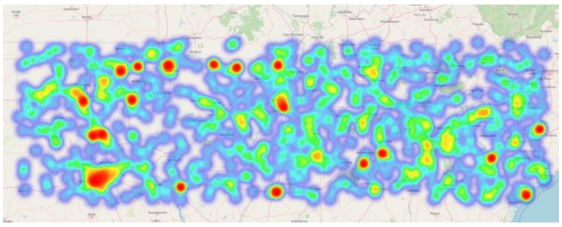

Figure 10: Geographical heatmap of filtered airports

Figure 9 and figure 10 are two different types of the heatmap, both show the numbers of airports in this region for the deeper the colour, the larger the number of airports exist.

Finally, before we go to the data and method section, two pie charts should be introduced. They represented the delay-punctual rate and the proportion of delay factors (see Figure 11,12). A highlighted section was deployed in both charts. The variables used in these two plots are all total delay / punctual arrived flight counts.

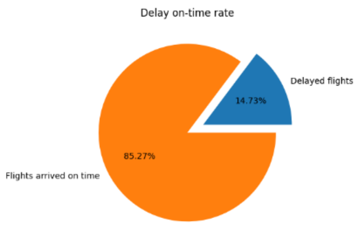

Figure 11: pie charts of delay-punctual rate

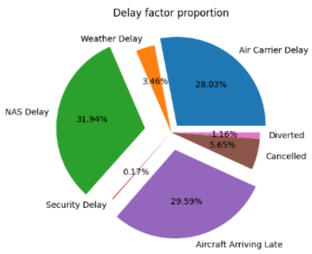

Figure 12: Pie chart of delay factor proportion

Data and Methods

Data

In this section, we present what data we used, the way we operate with the data's values, our algorithm structure and the specific methods or Python packages we used in code writing.

The datasets used in this report were "airline_delay_causes_Feb2020.csv" and the corresponding data named "airports.csv". To insight into the state, we added a new column 'state' in the original data.

Method and algorithm structure

To analyse the delay factors and find out a potential strategy, we first cleansed and reclassified the data, then we examined the leading factors in each state, by ranking the severity of delay from highest to the lowest by state, we then defined a new parameter: delay-punctual rate, using this parameter we sorted all the airlines in data and their performance towards controlling each delay factors. Then we selected and evaluated the top affected airports as well as 6 best delay-controlled airlines and airports in February 2020, US. By calculating the correlation between delay factors, we then investigate possible geographical factors correlated with some particular delay types. A filtered airport heatmap was generated with a latitude range of 32° to 37° and a longitude range of -100° to -80°, we generated two pie charts to visualize the delay-punctual rate and the proportion of each delay factor in total delay during February.

The Algorithm structure of our codes, with respect to code elegance, we chose to separate functional analysing functions as subfunctions defined in the following .py files. By importing them in main.py and calling these subfunctions, we successfully made our code easy to read as every code block is modular. In addition, in the process of writing code, in order to carry out code block debugging to prevent reporting errors, we also added some breakpoints to output after several complicated code blocks to check whether the code can be run properly, these breakpoints codes in the form of comments still exist in the formal project code, which allows us to retain the ability to edit our codes in the future!

As for specific methods or Python packages used in our code, in our main.py, to avoid recurring call-in output function, we defined the factors list inside the output function. Also, to solve the problem of RAM overflow that occurred when generating plots, we added **matplotlib.use('Qt5Agg')** to instruct matplotlib to use **Qt5Agg** backend for subsequent plotting operations (Matplotlib.use('Qt5Agg) what backend means, 2019), which enables graphics rendering and display. Another interesting method we used was when generating pie charts, we used **explode = ( )** to set the highlighted section. The specific Python package we used was **shutil**. This enabled us to set a separator when printing results. The structure of our algorithm is as follows.

Algorithm of analyzing airline delays:

Input:

.csv file: airline_delay_causes_Feb2020.csv

.csv file: airports

Process:

Regenerate data frame by different column names

Define a function in main.py to store the output of results

Define subfunctions in different .py files to analysis delay factors

Define output function to call subfunctions in followed .py files in main.py and output results

Output:

The delay severity sort by state

The total delay count sort by state

The delay-punctual rate sort by state

The greatest delay factor in each state

The severity of delays caused by carrier sort by state

Bar charts of 7 factors sort by state

define factors list in main.py, elements in list are column names of original data frame
for $each\ factor \in factors$ **do**
    df_sort = $\sum$ df[$factor$].sort_valuse
    use df_sort to plot bar charts
    save plot in folder ‘plot/’ named as ‘{factor}_by_state.png’
    closing the canvas

Number of delays caused by airlines sort by airlines

Combined bar charts of 7 factors effect on airlines (for)

define combined plot size using **fig, axes**
create list to store images
for $i, factor \in enumerate(factors)$ **do**
    create ax to generate labels
    set title as ‘carrier_delays_by_{factor}’
    df_sort = $\sum$ df[$factor$].sort_valuse
    use df_sort to plot bar charts
    use **np.arange(len(df_sort.index)** to set subgraph’s xlab length
    use **every_nth** and **rotation** to avoid labes overlaping
    append subgraphs to integrated graph
define integrated graph’s name
set storage paths to folder ‘plot/’
closing the canvas

The delay control level sort by airlines

Best 6 delay-controlled airlines and details

Bar chart of top 6 delay-controlled airlines in February

Best 6 delay-controlled airports’ delay-punctual rate and their details

Bar chart of top 6 delay-controlled airports in February

Top 10 most delayed airports (filtered $'arr\_flights' >= 8000$) names and details

Correlation matrix of 7 factors and their heatmap

Scatter plots, hexagonal binning plots and geographical heatmap of filtered airports within given region.

Pie charts of delay-punctual rate and proportion of delay factors

Our Suggestion, restrictions and conclusion

Finally, we reached the last section of this report. In this section, we first give a potential strategy to reduce flight delays and improve the passenger experience. Then by pointing out some restrictions of our analysis, we will give a conclusion to this report.

Potential suggestion

The strategy we proposed is to improve the National Airspace System (NAS) and to be specific to develop an even better flight route planning algorithm. According to the airline's delay data for February 2020, delays due to NAS took the largest proportion through the whole of America by 31.94% with around 30,000 flights. Also, NAS includes air traffic control, aeronautical communications, navigational facilities and related regulatory and supervisory agencies. The improvement of NAS should be taken in multiple ways and through software and hard drives. As weather factor is unavoidable, what we can do is improve flight path planning as flight path and route restrictions may lead to other key delay factors such as late arrival aircraft or arrivals diverted.

Developing better flight route planning algorithms to navigate the optimized routes could save fuel, arrival time and aircraft maintenance costs. Also, taking the quickest or most comfortable routes will also enhance passengers’ satisfaction.

However, it is important to acknowledge potential challenges. Firstly, developing such algorithms requires substantial commercial budgets to hire mathematics, statistics, and data scientists. Additionally, airlines or the Federal Aviation Administration (FAA) would need to collect extensive and cumbersome data to build robust datasets. Also, the maintenance cost and the impact of delays caused by the process of upgrading the system could not be ignored. It is foreseeable that the initial implementation of such measures could lead to short-term chaos or even disasters. However, once the algorithm has been fully upgraded, it will significantly benefit the overall efficiency and reliability of air travel in the US in the long term, benefiting both passengers and the US aviation industry.

Restriction of our analysis

As for restrictions of our analysis, the sample size of our dataset is not enough to deploy machine learning methods as it only contains airlines delay data within February 2020 in the US. Also, the data only contains delay data within one month and by having no other time factors, the only prediction could be made is through regression but not time series, which means we cannot generate further delay prediction models through this data.

Conclusion

To summarize this report, we notice the performance of airlines and airports controlling airlines' delays in the US during February 2020 was flawed. Through our analysis, we discovered that the geographical factor contributed a lot to weather which could cause airline delays, and the main factor that caused airline delays among 7 was NAS issue. To address the issue, we analysed the data and proposed one potential strategy, which is to develop a better flight route planning algorithm aiming to improve punctuality. Our analysis results demonstrate our conclusions and for further prediction, more data should be provided.

Appendix

Appendix-1: table of greatest delay factor in each state with their count:

greatest delay factor in each state with their count:

state	carrier_ct	weather_ct	nas_ct	security_ct	late_aircraft_ct	arr_cancelled	arr_diverted	max	max_idx
AK	91.45	13.25	176.68	1.90	144.73	37	16	176.68	nas_ct
AL	238.49	39.97	117.21	2.00	175.34	35	8	238.49	carrier_ct
AR	127.77	13.94	114.30	0.00	148.97	20	1	148.97	late_aircraft_ct
AZ	683.83	52.68	576.79	4.92	651.79	88	7	683.83	carrier_ct
CA	2615.93	222.56	2170.83	30.91	1878.77	480	137	2615.93	carrier_ct
CO	1208.00	145.56	1385.22	6.95	733.28	295	72	1385.22	nas_ct
CT	122.85	8.90	66.22	1.62	135.41	17	6	135.41	late_aircraft_ct
DC	639.88	52.91	727.19	2.49	965.50	277	58	965.50	late_aircraft_ct
FL	2477.12	232.80	3411.47	18.38	2726.24	285	102	3411.47	nas_ct
GA	1163.15	242.67	1881.48	7.65	1673.08	129	67	1881.48	nas_ct
HI	300.66	25.21	71.56	0.87	174.72	37	7	300.66	carrier_ct
IA	120.20	17.15	117.19	0.56	120.91	32	4	120.91	late_aircraft_ct
ID	106.85	28.33	73.72	0.00	37.09	10	7	106.85	carrier_ct
IL	1197.68	204.71	1945.68	7.07	1397.86	569	23	1945.68	nas_ct
IN	313.07	42.83	236.30	1.32	302.50	55	3	313.07	carrier_ct
KS	90.23	12.04	75.39	0.00	27.30	7	3	90.23	carrier_ct
KY	130.54	17.73	90.85	0.00	146.87	20	6	146.87	late_aircraft_ct
LA	342.40	41.23	421.03	1.79	335.52	35	61	421.03	nas_ct
MA	490.27	34.70	602.86	1.35	635.80	44	10	635.80	late_aircraft_ct
MD	257.90	11.99	226.19	0.99	307.92	60	6	307.92	late_aircraft_ct
ME	60.23	2.81	24.50	0.00	85.45	19	2	85.45	late_aircraft_ct
MI	811.71	195.14	631.75	1.50	624.88	92	19	811.71	carrier_ct
MN	464.30	124.10	375.14	2.13	390.33	31	9	464.30	carrier_ct
MO	450.72	34.89	323.49	1.96	497.89	109	8	497.89	late_aircraft_ct
MS	84.59	15.91	65.05	0.57	63.87	20	7	84.59	carrier_ct
MT	112.01	32.03	96.03	0.62	46.29	2	12	112.01	carrier_ct
NC	1181.15	167.42	1608.25	5.55	2343.60	695	114	2343.60	late_aircraft_ct
ND	104.77	22.70	66.78	0.00	25.76	18	4	104.77	carrier_ct
NE	128.96	22.38	114.94	1.00	100.75	22	1	128.96	carrier_ct
NH	45.43	4.16	19.56	1.00	38.84	10	3	45.43	carrier_ct
NJ	336.32	27.70	1362.85	3.42	535.74	41	15	1362.85	nas_ct
NM	95.17	5.92	69.83	0.00	92.08	24	1	95.17	carrier_ct
NV	511.72	49.77	500.99	7.22	524.26	58	7	524.26	late_aircraft_ct
NY	1206.30	152.30	1893.54	3.57	1451.37	280	50	1893.54	nas_ct
OH	615.78	57.35	510.65	2.64	903.58	92	5	903.58	late_aircraft_ct
OK	192.35	19.61	116.49	0.00	148.53	41	3	192.35	carrier_ct
OR	273.56	31.73	181.51	0.30	178.87	20	6	273.56	carrier_ct
PA	691.67	64.50	645.23	3.98	872.65	122	15	872.65	late_aircraft_ct
PR	180.73	3.19	100.82	1.60	108.67	2	1	180.73	carrier_ct
RI	83.82	3.65	37.65	0.00	67.87	26	2	83.82	carrier_ct
SC	242.94	36.52	230.12	0.24	308.17	64	10	308.17	late_aircraft_ct
SD	75.36	9.17	48.99	0.00	36.51	7	2	75.36	carrier_ct
TN	508.98	62.72	476.52	2.35	569.49	79	20	569.49	late_aircraft_ct
TT	8.12	0.00	1.39	0.00	4.49	1	0	8.12	carrier_ct
TX	2600.08	234.53	3260.48	14.47	2782.41	466	69	3260.48	nas_ct
UT	441.01	102.21	379.97	1.39	283.40	31	7	441.01	carrier_ct
VA	253.97	34.03	186.60	1.07	315.34	63	11	315.34	late_aircraft_ct
VI	37.51	0.34	11.82	0.05	26.27	1	0	37.51	carrier_ct
VT	32.77	3.31	21.32	0.00	45.60	27	2	45.60	late_aircraft_ct
WA	527.06	85.19	865.33	3.72	413.69	63	21	865.33	nas_ct
WI	282.03	40.97	200.99	1.57	202.47	39	7	282.03	carrier_ct
WV	33.63	13.96	22.56	0.00	33.86	18	4	33.86	late_aircraft_ct
WY	61.40	17.69	62.99	0.00	24.92	20	16	62.99	nas_ct

Appendix-2: table of ranked airlines

```
delay control level sort by carrier:

carrier_name
PSA Airlines Inc.             0.272072
Allegiant Air                 0.239729
Frontier Airlines Inc.        0.182180
Envoy Air                     0.171537
Spirit Air Lines              0.167243
JetBlue Airways               0.162622
Mesa Airlines Inc.            0.160741
Republic Airline              0.159797
Endeavor Air Inc.             0.157610
American Airlines Inc.        0.156722
SkyWest Airlines Inc.         0.155035
Alaska Airlines Inc.          0.146656
ExpressJet Airlines LLC       0.146440
United Air Lines Inc.         0.140191
Delta Air Lines Inc.          0.130928
Southwest Airlines Co.        0.094634
Hawaiian Airlines Inc.        0.087571
```

Appendix-3: table of best6 delay control airlines details

how these carriers control delay:

carrier_name	arr_flights	arr_del15	carrier_ct	weather_ct	nas_ct	security_ct	late_aircraft_ct	arr_cancelled	arr_diverted	min_index	min_value
Alaska Airlines Inc.	19515	2862.0	736.14	40.01	1215.37	10.24	860.34	176	38	security_ct	10.24
Delta Air Lines Inc.	75446	9878.0	2662.67	384.32	3652.94	3.89	3174.12	18	90	security_ct	3.89
ExpressJet Airlines LLC	11404	1670.0	319.99	32.96	900.15	0.00	416.87	78	19	security_ct	0.00
Hawaiian Airlines Inc.	6509	570.0	414.71	18.73	9.36	7.39	119.81	32	2	arr_diverted	2.00
Southwest Airlines Co.	101581	9613.0	3281.77	79.91	2374.08	34.63	3842.55	1109	112	security_ct	34.63
United Air Lines Inc.	46244	6483.0	1389.77	113.00	2742.50	0.00	2237.66	94	77	security_ct	0.00

Appendix-4: table of ranked best 6 delay control airports

```
 best 6 delay control airports sort by delay ratio:

airport_name
Los Angeles, CA: Los Angeles International            0.120171
Salt Lake City, UT: Salt Lake City International      0.119996
New York, NY: John F. Kennedy International           0.119111
Minneapolis, MN: Minneapolis-St Paul International    0.109806
Las Vegas, NV: McCarran International                 0.109245
Phoenix, AZ: Phoenix Sky Harbor International         0.107508
dtype: float64
```

Appendix-5: table of best delay control airlines details

how these airports control delay:

airport_name	arr_flights	arr_del15	carrier_ct	weather_ct	nas_ct	security_ct	late_aircraft_ct	arr_cancelled	arr_diverted	min	min_value
Las Vegas, NV: McCarran International	13099	1431.0	436.69	43.09	469.13	7.22	474.84	55	4	arr_diverted	4.00
Los Angeles, CA: Los Angeles International	16643	2000.0	750.35	69.83	666.43	7.98	505.39	106	28	security_ct	7.98
Minneapolis, MN: Minneapolis-St Paul International	11575	1271.0	428.77	110.06	359.91	2.13	370.14	26	8	security_ct	2.13
New York, NY: John F. Kennedy International	9764	1163.0	401.03	16.49	325.80	2.65	417.05	23	3	security_ct	2.65
Phoenix, AZ: Phoenix Sky Harbor International	14557	1565.0	526.47	29.58	470.64	4.63	533.69	79	7	security_ct	4.63
Salt Lake City, UT: Salt Lake City International	9467	1136.0	404.00	97.54	355.59	0.39	278.46	27	4	security_ct	0.39

Appendix-6: table of correlation matrix

output of correlation matrix:

	carrier_ct	weather_ct	nas_ct	security_ct	late_aircraft_ct	arr_cancelled	arr_diverted
carrier_ct	NaN	0.766672	0.802605	0.758725	0.898456	0.596191	0.546648
weather_ct	0.766672	NaN	0.614960	0.502162	0.647010	0.444804	0.490560
nas_ct	0.802605	0.614960	NaN	0.699045	0.845741	0.514642	0.558181
security_ct	0.758725	0.502162	0.699045	NaN	0.677638	0.400576	0.436363
late_aircraft_ct	0.898456	0.647010	0.845741	0.677638	NaN	0.684362	0.635885
arr_cancelled	0.596191	0.444804	0.514642	0.400576	0.684362	NaN	0.598068
arr_diverted	0.546648	0.490560	0.558181	0.436363	0.635885	0.598068	NaN

Appendix-7: table of worst 10 airports in February

how these airports been affected by delay factors:

airport_name	carrier_ct	weather_ct	nas_ct	security_ct	late_aircraft_ct	arr_cancelled	arr_diverted	max	max_value
Atlanta, GA: Hartsfield-Jackson Atlanta Interna...	1052.49	219.85	1797.66	6.23	1532.78	99	63	nas_ct	1797.66
Boston, MA: Logan International	471.65	34.70	600.39	1.35	630.88	44	9	late_aircraft_ct	630.88
Charlotte, NC: Charlotte Douglas International	727.17	100.43	1261.52	3.13	1739.73	577	91	late_aircraft_ct	1739.73
Chicago, IL: Chicago O'Hare International	947.96	168.35	1759.86	4.35	1135.50	420	15	nas_ct	1759.86
Dallas/Fort Worth, TX: Dallas/Fort Worth Intern...	976.71	104.86	1535.73	8.81	1211.88	157	35	nas_ct	1535.73
New York, NY: LaGuardia	402.42	80.89	1301.31	0.00	665.40	135	23	nas_ct	1301.31
Newark, NJ: Newark Liberty International	302.58	26.73	1335.83	1.42	505.47	35	14	nas_ct	1335.83
Orlando, FL: Orlando International	565.29	32.41	910.04	2.65	597.58	60	15	nas_ct	910.04
Seattle, WA: Seattle/Tacoma International	433.57	61.00	767.55	3.72	378.16	60	17	nas_ct	767.55
Washington, DC: Ronald Reagan Washington National	445.91	26.32	527.91	1.38	736.44	242	56	late_aircraft_ct	736.44

References

Mitsokapas, E., Schaefer, B., Harris, R. J., & Beck, C. (2021). Statistical characterization of airplane delays. Scientific Reports, 11(1), 7855–7855. https://doi.org/10.1038/s41598-021-87279-8

Cohen, J., Pfeiffer, K., & Francis, J. A. (2018). Warm Arctic episodes linked with increased frequency of extreme winter weather in the United States. Nature Communications, 9(1), 869–12. https://doi.org/10.1038/s41467-018-02992-9

Matplotlib.use(‘Qt5Agg) what backend means. (2019, April 15). In CSDN. https://blog.csdn.net/The_Time_Runner/article/details/89313599