Aviation Performance Metrics: An in-depth analysis

Qilun Lyu
Computer Science department
Rutgers University
Piscataway, NJ, USA
Email: qilun.lyu@rutgers.edu

Abstract — **Aviation Performance Metrics provides tourists or passengers, airport authorities, and airline companies with a platform for being better aware of the aviation industry through some standard benchmarks. It answers when, where, and why the delays and cancellations happen frequently. Moreover, it bridges the gap between humans and machines through hovering, zooming, brushing, panning, and many other interaction techniques.**

The project provides users with a complete set of visualization elements and a thorough analysis behind the scenes to be better aware of different aviation industry components' performance. These include an interactive map that enables users to browse and search different airports, a chord diagram to show details of inter-airport delays, a fully correlated series of statistical modules to shed light on the reasons for airline delays and cancellations, and their corresponding time-series forecasts.

All of the above might help to reschedule the flights for airport authorities and airline companies and promote the travel experience for tourists.

I. INTRODUCTION

Delay or cancellation of flights is annoying not only for the passengers but also for the airport and airline crews. Such delay or cancellation leads to loss for all the parties engaged. The passenger loses time and money, and consequently, the airport or airline crews have to face their complaints and whining, which is not a pleasant experience for them. Furthermore, their reputation also gets a hit, in that On-Time Performance (OTP) or punctuality is regarded by the majority as the most primary Key Performance Indicators (KPI) of an airline's rating on the top of others like quality or price [1]. Delays or cancellations due to hazardous weather conditions or national security matters are unavoidable, but some other reasons can be managed and administered, thus averting those issues. What is worse, the undergoing COVID-19 pandemic shows no sign of stopping, bringing less profit to airports and airlines. The gradually formed new normal makes new forecasting exceptionally necessary after it broke out. If we could find a pattern out of it, we might help dozens of people getting accustomed to this new environment.

The first section employs webpage visualization of a liberal amount of flight history records, ranging from Jan 2011 thru Sept 2020, covering domestic flights within the United States of America, as found on the Bureau of Transportation Statistics(BTS) official website[1]. By accessible interactive tools like the map we offer, the industry's reality is vividly presented in front of our users. With the interconnected diagrams we provide, the status of operations becomes uncomplicated to monitor for the stakeholders.

In this project's analysis section, we forecast cancellation rates, on-time rates, and average delay minutes, recommended by OAG Aviation guides[2], through leveraging ARIMA models from time-series approaches. Also, we attempt to reduce the intervention from the pandemic in order for a narrower prediction interval that reflects the new normal. Hopefully, such a minor contribution would favor people if they are able to schedule their journey in the end adequately.

II. USE CASE

Tourists or passengers, airport authorities, and airline companies are the typical users of our project since their demands fits our project. The illumination of the cases is shown below.

- **Tourists/Passengers**: A person who want to travel from one destination to another in US taking domestic flights would find it useful as he/she can look at the delays of airlines, airports and cancellation rate of airlines and airports to plan his bookings. Using our application passenger can know the delay trend, delay in routes, on time rate of airports and airlines which will help him to know from which airport and airline he/she should book the flight:

Scenario 1: Passenger *A* wants to know the delay between LAX and EWR.
System input: Go to the *Route Delay* page under *Airport* Menu. In the chord diagram on the left, select the desired airports.
System output: A path will appear between them. This path shows for both directions the delay between the chosen airports.

Scenario 2: Passenger *B* wants to know the reason behind cancellations for a particular date.
System input: Go to the *Cancellation* page under *Airline* Menu and select the desired filters Finally click on any of the bars in the bottom bar chart and subsequently click on the *Reason for Cancellation* button.
System output: The various reasons that may have caused the cancellations will appear in a Google page, from where the user can select the ones he wants to delve into further.

Scenario 3: Passenger *C* wants to know how to go to a specific airport or to get more information about that

[1] Refer to https://transtats.bts.gov/Fields.asp?gnoyr_VQ=FGJ

airport.
System input: Go to *Map* page under *Airports* and either select the desired airport from the map itself or search for it using the search bar. Hover of the highlighted airport and an info box will appear. Click on *How to get there?* or *More Info*.
System output: *How to get there?* - The Google Maps with directions from current location to the desired airport. *More Info* – A Wikipedia page containing further information.

- **Airport Authorities**: Airport authorities can be informed of the OTP of the airport and also get useful and relevant statistics and plots for the airport. They can also visualize important trends of delay and cancellation for their airports.

Scenario 1: The Airport Authority of JFK airport wants to know the average delay, on-time rate and/or cancellation rate.
System input: Go to *Map* page under Airports and either select JFK from the map itself or search for it using the search bar.
System output: On the right panel, all the required information will be provided.

Scenario 2: The Airport Authority of LAX wants to know the pattern of delays or cancellations over time to better manage the airlines and reduce delay.
System input: Go to *Map* page under Airports and either select LAX from the map itself or search for it using the search bar.
System output: Various statistical data will appear on the bottom half of the page. From there, they can gather information regarding the pattern of delays, which years, months and days have higher traffic and identify potential risks behind it.

- **Airline Companies**: Airline companies can find it useful to get causes for delay within their airlines, to compare the performance with their competitors. They can also make analysis regarding the delays as well as the cancellation and also reasons and factors behind that using the functionalities integrated in our project.

Scenario 1: Alaska Airlines want to know the reason for cancellation behind their cancellation in February 2015.
System input: Go to *Cancellation* page under *Airlines* menu. Select *Alaska Airlines* from airlines bar chart, and then select *Feb* from month bar chart and select *2015* from year bar chart. Click on *Reason for Cancellation* button on the right.
System output: The various reasons that may have resulted in the cancellations will appear in a Google page, from where they can select the ones they want to investigate.

Scenario 2: American Airlines want to know which month and year had the highest delay for *Late Aircraft Delay*.
System input: Go to *Overall Comparison* page under *Airlines* menu. Select *American Airlines* from airlines bar chart and *Late Aircraft Delay* from the pie chart.
System output: The bar chart at the bottom will show the distribution of delays and the user can view which months and years had higher delays compared to the rest.

III. DATA

The Reporting Carrier OTP Datasets are provided by BTS, affiliated by US Department of Transportation. The data covers essential details of US domestic flights since 1987, with 109 attributes in total. Here are the main attributes we have selected and their brief description. Note that CRS here stands for Computer Reservation System.

ATTRIBUTE	**DESCRIPTION**
FlightDate	Flight Date(in *yyyymmdd* format)
IATA_CODE_Reporting_Airline	Code assigned by IATA and commonly used to identify a carrier
Origin	Origin Airport
Dest	Destination Airport
DepDelayMinutes	Difference in minutes between scheduled and actual departure time. Early departures set to 0
DepTimeBlk	CRS Departure Time Block, Hourly Intervals
ArrDelayMinutes	Difference in minutes between scheduled and actual arrival time. Early arrivals set to 0
ArrTimeBlk	CRS Arrival Time Block, Hourly Intervals
Cancelled	Cancelled Flight Indicator(1=Yes)
Diverted	Diverted Flight Indicator (1=Yes)
Distance	Distance between airports (miles)

And the following is the data snippet we screenshotted from the database.

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content: | Fetch rows:

ID	FlightDate	IATA_Airline	Tail_Number	Flight_Number	Origin	Dest	CRSDepTime	CRSArrTime	CRSElapsedTime	Distance	Canc
1	2015-01-01	AA	N787AA	1	JFK	LAX	0900	1230	390	2475	0
2	2015-01-02	AA	N795AA	1	JFK	LAX	0900	1230	390	2475	0
3	2015-01-03	AA	N788AA	1	JFK	LAX	0900	1230	390	2475	0
4	2015-01-04	AA	N791AA	1	JFK	LAX	0900	1230	390	2475	0
5	2015-01-05	AA	N783AA	1	JFK	LAX	0900	1230	390	2475	0
6	2015-01-06	AA	N799AA	1	JFK	LAX	0900	1235	395	2475	0
7	2015-01-07	AA	N784AA	1	JFK	LAX	0900	1235	395	2475	0
8	2015-01-08	AA	N787AA	1	JFK	LAX	0900	1235	395	2475	0
9	2015-01-09	AA	N795AA	1	JFK	LAX	0900	1235	395	2475	0
10	2015-01-10	AA	N790AA	1	JFK	LAX	0900	1235	395	2475	0
11	2015-01-11	AA	N786AA	1	JFK	LAX	0900	1235	395	2475	0
12	2015-01-12	AA	N790AA	1	JFK	LAX	0900	1235	395	2475	0
13	2015-01-13	AA	N796AA	1	JFK	LAX	0900	1235	395	2475	0
14	2015-01-14	AA	N793AA	1	JFK	LAX	0900	1235	395	2475	0
15	2015-01-15	AA	N790AA	1	JFK	LAX	0900	1235	395	2475	0
16	2015-01-16	AA	N797AA	1	JFK	LAX	0900	1235	395	2475	0
17	2015-01-17	AA	N795AA	1	JFK	LAX	0900	1235	395	2475	0
18	2015-01-18	AA	N786AA	1	JFK	LAX	0900	1235	395	2475	0
19	2015-01-19	AA	N785AA	1	JFK	LAX	0900	1235	395	2475	0
20	2015-01-20	AA	N797AA	1	JFK	LAX	0900	1235	395	2475	0
21	2015-01-21	AA	N789AA	1	JFK	LAX	0900	1235	395	2475	0
22	2015-01-22	AA	N790AA	1	JFK	LAX	0900	1235	395	2475	0
23	2015-01-23	AA	N790AA	1	JFK	LAX	0900	1235	395	2475	0
24	2015-01-24	AA	N787AA	1	JFK	LAX	0900	1235	395	2475	0
25	2015-01-25	AA	N791AA	1	JFK	LAX	0900	1235	395	2475	0

Fig. 1 Data Snippet

To better evaluate airports and airlines' overall performance, we rounded the period into the nearest decade from Jan 2011 thru Sep 2020 to eliminate the impact of great changes of airports, airlines, and air routes. Thus, we have 117 CSV-format files indicating 117 months with over 60 million records.

A. Data Processing

We have identified the following entities, attributes and relationships for given data as shown in Fig. 2 in the ideal Entity Relationship Diagram (ERD):

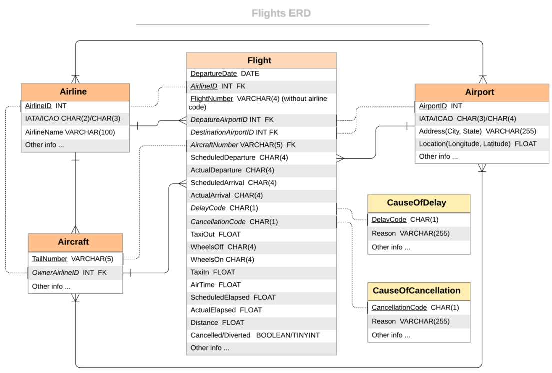


Fig. 2. ERD

In the given figure, it clarifies the entities and the relations in between. Airline, airport, aircraft, flight, cause of delay and cause of cancellation are evident entities. For relations, Flight presents *many-to-one* with others, which is the same as that between aircraft with airline. Also, airport indicates *one-to-many* with aircraft and *many-to-many* with airline, and similar in aircraft's case. Every chart contains

- No anatomic domain,
- At least one set of unique identifiers as a primary key AND
- No dependency between any other non-primary attributes.

Thus, the whole diagram satisfies Boyce-Codd NF or 3.5NF.

The challenge here is if we can arithmetically derive an quantity of our interest. For instance, suppose we want to calculate the duration of a flight from the given departure and arrival timepoints. We will soon encounter difficulties such as lack of time zone details and Daylight Saving Time (DST) availability information. Hence, we must reserve the duration as an attribution.

Empirically, however, we determined to lower down the paradigm level (i.e. Normal Form) to 3NF for fewer calculations and faster retrievals, in the sense that the scale of data to display is significantly large, which in other words we kept some dependent non-primary attributes that have already appeared in our raw data. Note that we only focus on punctual flights, delayed flights and cancelled flights, regardless of diverted ones.

Later on, we will use common statistics for visual analytics, such as average delay of airlines, on-time rate of airports, frequency of flights of the two, and density of airports within a region, and so on so forth.

B. Context Diagram

Fig. 3 represents the Context Diagram at level 0 of our project. It includes the four major kinds of users i.e. *Developers* which maintains the web application while *Airline Companies*, *Travelers* and *Airport Authorities* can perform interaction and queries to grasp suitable information from data.

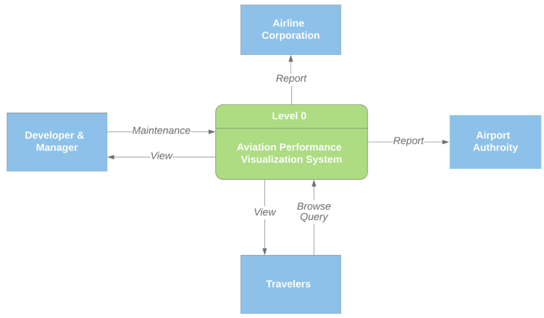


Fig. 3. Context Diagram

C. System Flow Diagram

Fig. 4 shows the System Flow diagram of our project, which summarizes the process of a user's operation on the webpage. The back-end server first receives requests from the user's actions, then grabs and tidies the data from the database and responds to the front-end GUI. The data is stored in our MySQL database using ETL process (Extraction, Transformation and Loading).

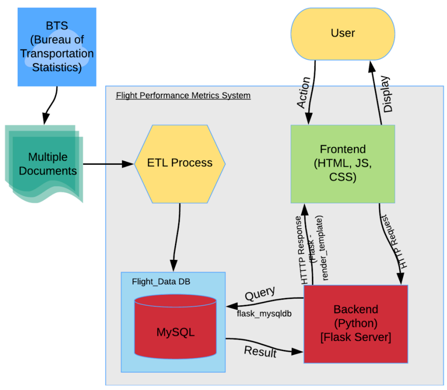


Fig. 4. System Flow Diagram

IV. VISUALIZATION

First, let us get down to the *Time Series* module.

A. Time Series Module

It pops up the following line chart as shown in Fig. 5. It allows selecting and hovering over lines for the exact values. While sometimes the percentages of cancellation and delay on certain days like Mar 14, 2017 (Fig. 6) are approximate, most of the differences between them are evidently large. In general, they are negatively correlated, as we expected.

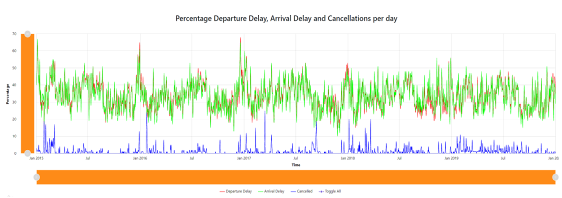


Fig. 5. A macro-view of delay and cancellation rates

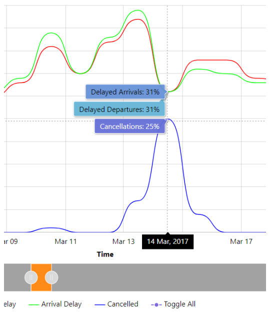


Fig. 6. An anomaly discovered while sliding the series plot

Now, let us take a look at the *Map* module.

B. Map Module

As is shown in Fig. 7, the darker the dot is, the poorer its punctuality is. Moreover, the size indicates the scale, measured by the number of visible destinations after clicking on an airport (Fig. 8). Besides, regardless of the size, Flight Route Mode offers discernible paths in a darker environment(Fig. 9), which is appropriate for specialists to recognize its traffic pattern of the network.

Fig. 7. Airport distribution & performance

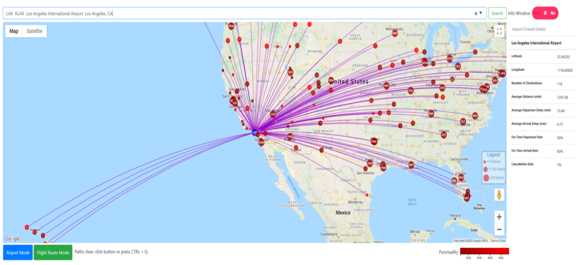


Fig. 8. Connectivity of LAX

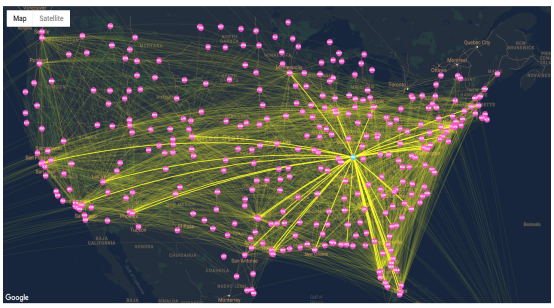


Fig. 9. Interconnectivity between airports

C. Statistical Diagram Module

Fig. 10 displays inter-airport delays between major airports, making it possible to analyze each direction of a specified flight path. To fetch the delays or cancellations of specific airlines, it offers filtering, which plays a crucial role in getting their insights. Add/remove airlines, months, years, and delay types to narrow down the search (Fig. 11). There are other modules in our project which (due to lack of space) were not incorporated in summary.

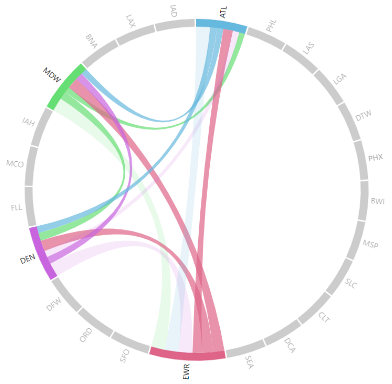


Fig. 10. Inter airport delays

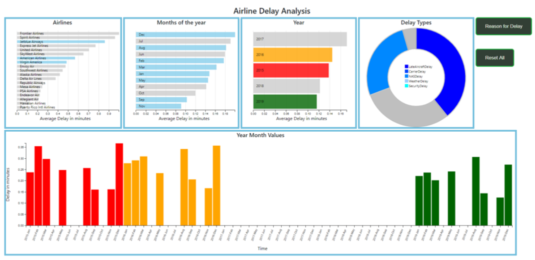


Fig. 11. Airline analysis at a micro level

V. ANALYSIS

A. Exploration Steps

The quantities of time series of our interest are cancellation rate, on-time rate, and average delay minutes after reading the report with regard to guide of performance measures [3]. By definition, every flight is either on-time, delay, canceled, or diverted. Furthermore, the portion of diverted flights can be ignored if compared to the others. Thus, we learned that cancellation and execution (normal or delayed) are strongly negatively correlated, which matches the results we have found in the previous visual plots. The remaining work is model fitting with residual analysis, as well as forecasting.

Nevertheless, before modeling, we have to check each series' stationarity by both the qualification (e.g., ACF, PACF, and EACF plots) and the quantification (e.g., *Ljung-Box Test* and *Augmented Dickey-Fuller Test*) methods. We can then fit them with ARIMA models by maximum likelihood method after narrowing down the range of possible orders p, d, q by AIC and BIC. Meanwhile, coefficients are estimated by the models, making it easier for diagnostics. Thanks to Box and Jenkins' airlines model (1976) [3], I treated my model as a seasonal one, and the tuition was eventually proved correct.

Once we select a model, diagnosing begins. The residual analysis helps us better determine a model by examining whether assumptions, such as zero mean, constant variance, and normality, hold. Also, causality, invertibility, overfitting, and identifiability have a chance to be checked during the process. Besides, transformation is applied when necessary.

Finally, we tried to forecast the future series to minimize mean squared error (MSE). To better explain the test results, I unified the test size and set it to be 5%.

B. Interesting Findings

There are many questions that can be answered using the web application. From that some of the interesting findings were as follows:

- Airports in the east coast have more connections compared to west coast (Fig. 9).
- There is a positive correlation between Departure Delay and Arrival Delay, which is expected.
- There is a negative correlation between Cancellation and Delay, which is expected (Fig. 6).
- Late Aircraft Delay, Carrier Delay and NAS delay are the main reasons of delay.
- Among the big-name airlines, Alaska Airlines performs the best.

We also found some interesting findings when analyzing the data. According to the Pareto principle, we found top 10% of airports with the most significant annual number of scheduled flights own more than 94.8% of flights. Ridiculously, the top 10% of airlines take up over 97.7% of all flights. The half-normal plot (see Fig. 12) distinguishes these airports and airlines from others.

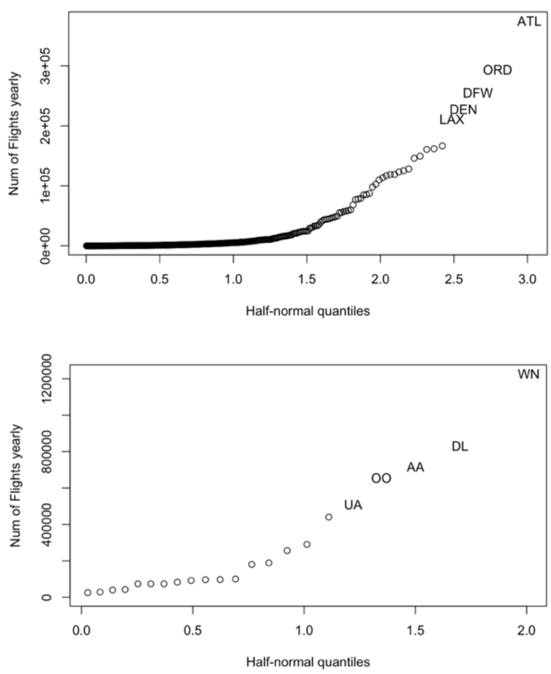


Fig. 12. Half-normal quantile plot for identifying major airports and airlines. According to the average of annual number of scheduled flights, Atlanta International Airport (ATL), Chicago O'Hare International Airport (ORD) and Dallas/Fort Worth International Airport (DFW) rank the top 3. And Southwest Airlines (WN), Delta Air Lines (DL) and American Airlines (AA) own the most scheduled flights yearly.

As for the average hourly number of flights, we discovered that the minimum average number of arrival flights in a day during the 06:00-06:59 period comes with the maximum average number of departure ones (Fig. 13).

Take cancellation rate as an example. The cancellation rate is highly affected by extreme values that appeared in early 2020, making the stationarity obscure (Fig. 14(a)). After removing this part (i.e., outliers), I got stationary series for model fitting (Fig. 14(b)), which enables me to carry out the following work [4].

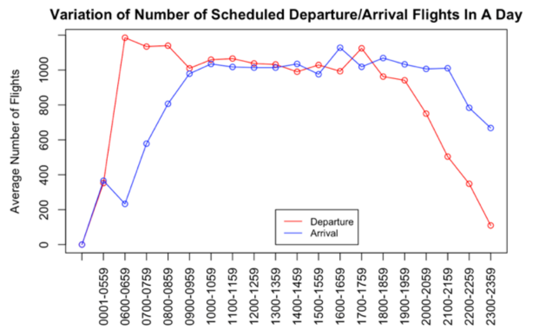
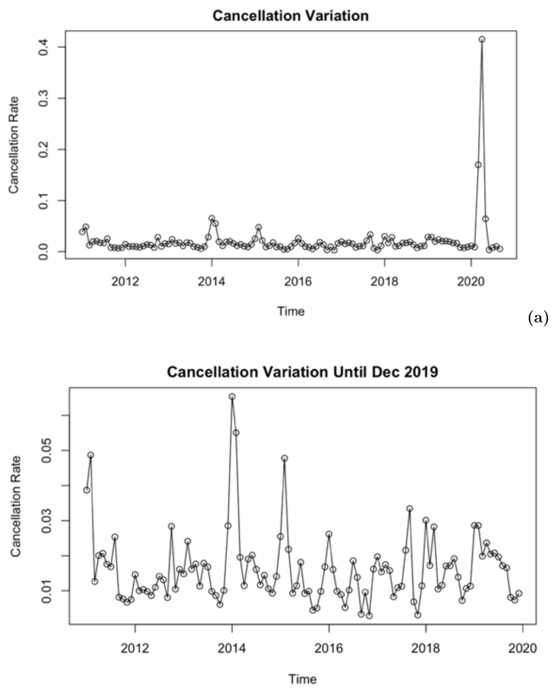


Fig. 13. Variations of the average number of scheduled flights within one day. The label of x-axis is a set of hourly time blocks except 00:01-05:59. Red line refers to the changes of the number of departure flights and the blue one to the variation of the number of arrival flights.



(b)

Fig. 14. Time series plots of cancellation rates. Plot (a) shows extreme values in the early 2020 as the pandemic broke out. Plot (b) focus on the series before 2020.

C. Modelling

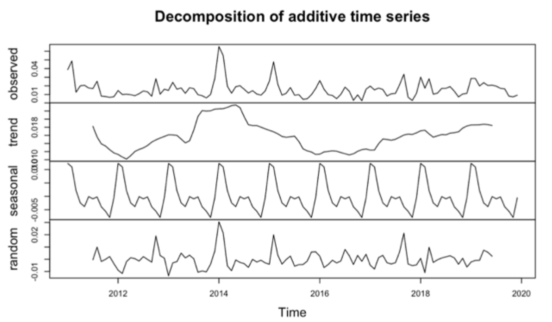
At first sight, we decomposed the series in order to find out valuable factors. As the graph Fig. 15(a) shows, the series' trend is not evident; however, the seasonal component tells a strong seasonal pattern of the data. For the stationarity test, we adopted the *Augmented Dickey-Fuller*

Test result with the help of ACF and PACF plots. The following is the test result from R.

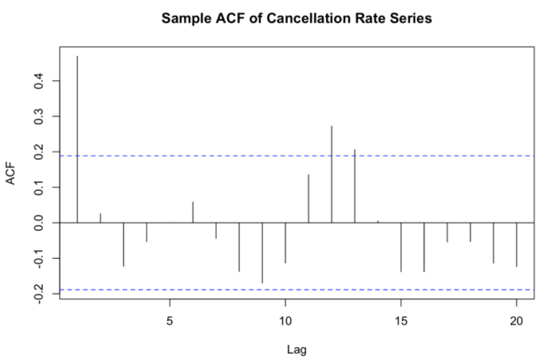
```
Augmented Dickey-Fuller Test

data:  part.cancellation.series
Dickey-Fuller = -6.2738, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
```

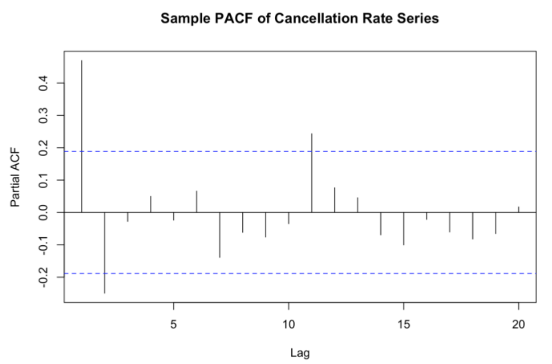
From the result above, we retain the stationary hypothesis. Also, in the ACF and PACF plots, the decay trend can be easily explored: ACF plot (Fig. 15(b)) of this partial series of cancellation rates cuts off at lag 4 while PACF plot((Fig. 15(c))) cuts off at lag 3. we can therefore limit the choices of p from 0 to 3 and q from 0 to 4. The seasonal component is so stationary that we only need to set p=1 and d=q=0 in the seasonal model. Thus, we constructed an ARIMA(p,0,q) $\times$(1,0,0)$_{12}$ model.



(a)



(b)



(c)

Fig. 15. (a) The decomposition of series of cancellation rates. (b) sample ACF of series of cancellation rates. (c) sample PACF of series of cancellation rates. In (b) and (c) the frequency of the series is 1 in lieu of 12.

Then by recursively running ARIMA models with different p, q, I found the model ARIMA(2,0,0) $\times$(1,0,0)$_{12}$ has the lowest few AICs and BICs(Tab. 1), since the first common smallest position is at p=2 and q=0. And the coefficients ϕ_1=0.5457, ϕ_2=−0.1958 as well as the seasonal model coefficients ϕ=0.1842. Here is the result from R compared to that of **auto.arima()** function.

```
ARIMA(2,0,0)(1,0,0)[12] with non-zero mean

Coefficients:
         ar1      ar2    sar1    mean
      0.5457  -0.1958  0.1842  0.0164
s.e.  0.0984   0.1042  0.1074  0.0015

sigma^2 estimated as 7.624e-05:  log likelihood=360.44
AIC=-710.88   AICc=-710.29   BIC=-697.47
```

Tab. 1(a) AIC of the model for cancellation rates series

q \ p	0	1	2	3	4
0	-687.3586	-709.6518	**-711.0676**	-710.1408	-708.3429
1	-709.4148	-709.7693	**-711.4321**	-709.5554	-707.6951
2	-710.8785	-709.0180	-709.5976	-710.5079	-708.5079
3	-709.1491	-708.1290	-709.3919	-708.3933	**-711.3695**

(b) BIC of the model for cancellation rates series

q \ p	0	1	2	3	4
0	-679.3122	**-698.9232**	**-697.6569**	-694.0480	-689.5680
1	**-698.6863**	-696.3586	-695.3393	-690.7804	-686.2380
2	-697.4678	-692.9252	-690.8227	-689.0508	-684.3687
3	-693.0563	-689.3541	-687.9349	-684.2541	-684.5482

The color above represents the ranking of values: 1st red, 2nd green, 3rd blue, 4th orange.

In addition, by raising the power d, both AIC and BIC of the fitted model increase. So, d=0 is the by far best bet.

D. Residual Analysis

The result unit root test (see Fig. 16) indicates that the stationarity of residual series holds under lag 4 (if treated monthly). Moreover, the result of following *Ljung-Box Test* of the residuals guarantees the statement.

```
Box-Ljung test

data:  residuals from  cancellation.model
X-squared = 7.276, df = 9, p-value = 0.6084
```

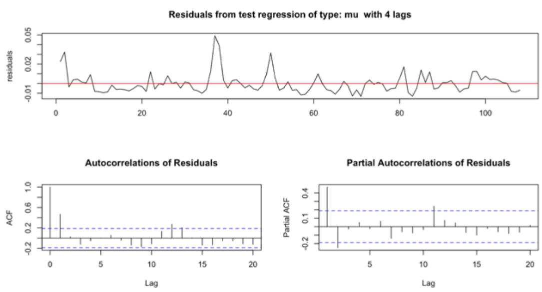
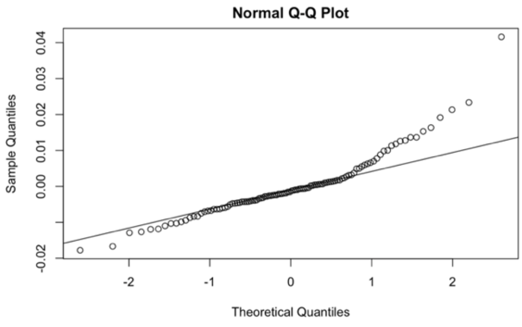


Fig. 16. Residual plots, ACF and PACF of residuals

However, when testing the residual normality, we were shocked by the inconsistency of the QQ-plot and Shapiro-Wilk normality test results, as shown in Fig. 17. SW Test turns out a small p-value and rejects the normality assumption. But in QQ-plot, most of the data points lie on a 45-degree line with a limited number of exceptions.



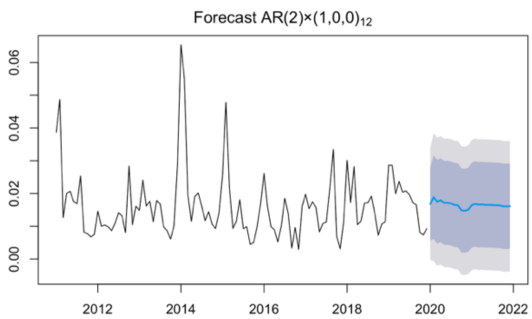

```
Shapiro-Wilk normality test

data:  cancel.residual
W = 0.90594, p-value = 1.24e-06
```

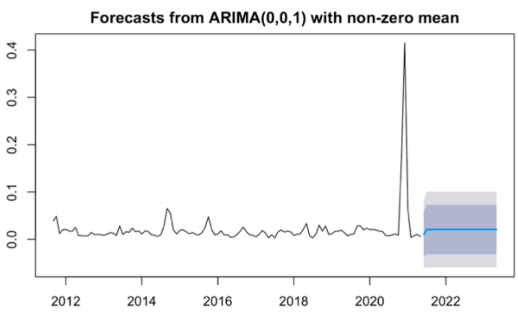
Fig. 17. QQ-plot and Shapiro-Wilk test results

VI. FORCASTING

Having fitted the model, we forecast the future series based on the factors we have extracted. As the forecast plot (Fig. 18(a)) suggests, the blue line represents the mean value, and the shaded purple area forms the confidence interval of the mean. Furthermore, the confidence intervals infer that the intervals would cover 95% of the future cancellation rate if the pandemic did not happen. Compared to Fig. 18(b), which we consider the impact caused by COVID-19, the interval is much narrower. Since the cancellation rate usually is less than 2%, wider intervals mean less precision on the forecasting.



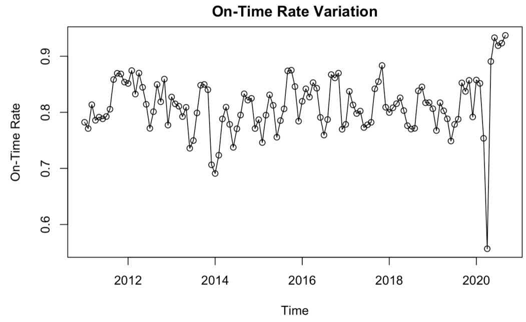
(a)

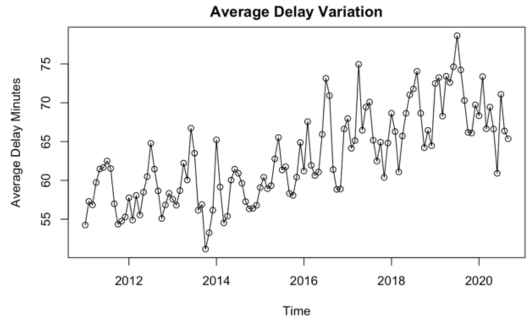


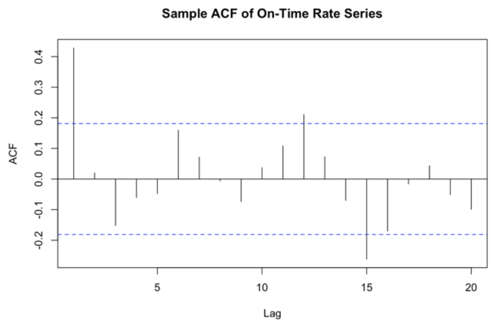
(b)

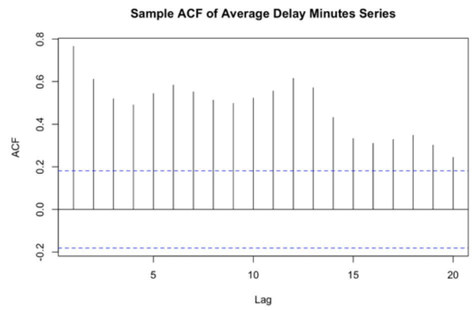
Fig. 18. (a) Forecasts of cancellation rate without the impact of COVID-19. (b) Forecasts of cancellation rate with the impact of COVID-19.

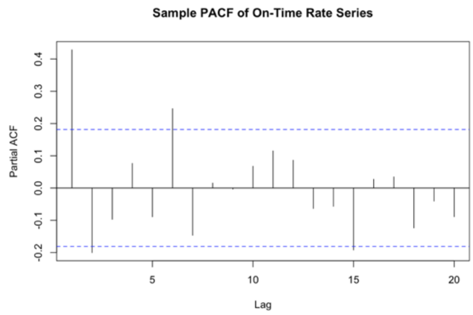
We repeat the same process (see Fig. 19) to study on-time rate and average delay minutes. Similarly, we got ARIMA(1,0,2)×(0,0,1)$_{12}$ model with ϕ_1= -0.7828, θ_1=1.3830, θ_2=0.6087 and seasonal model coefficient θ=0.4203. Plus, the forecasts are shown in Fig. 20.











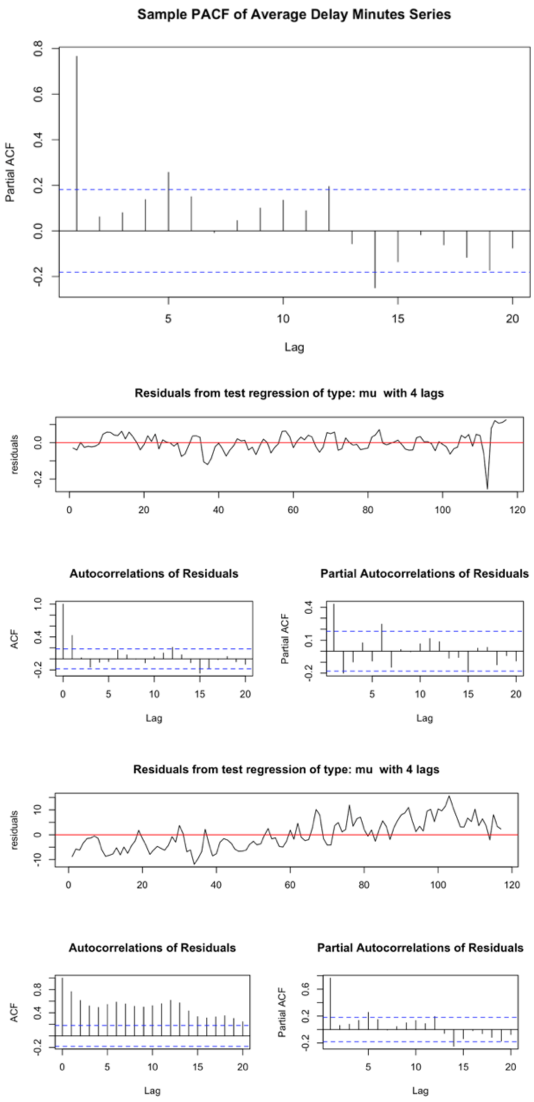
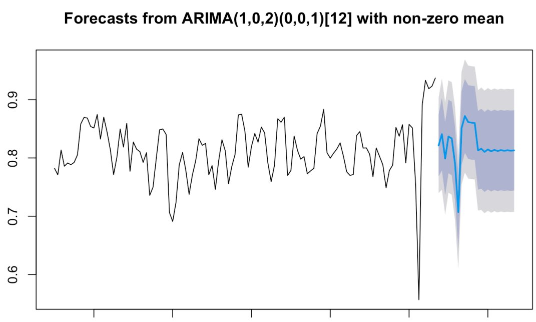
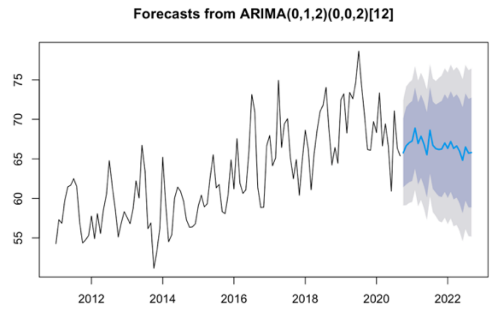


Fig. 19. Plots of series of on-time rate and average delay minutes.



(a)



(b)

Fig. 20. Forecasts of series of on-time rate(a) and average delay minutes(b).

Acknowledgements

We would like to express our appreciation to Professor James Abello for guiding and helping our team to a successful completion of the project. Assistance given by Spring 2020 CS526 DIVA course TA Nishant Kumar for JavaScript Libraries is greatly appreciated. We would like to offer our special thanks to TA Fangda Han for important suggestions during different phases of our project. We would also like to thank Enkai Ji for his input in different UI elements of the front-end.

We would even like to give special thanks to Professor Ruobin Gong for her dedication in teaching MSDS 596 Regression and Time Series course which inspired us quite a lot to fulfill this project both theoretically and practically. Without the comprehensive analysis part, the project would never be wonderful like what you have seen so far.

Reference

[1] B. Walther, "6 Most Important KPIs For Airline Operations," Information Design, 3 2021. [Online]. Available: https://www.id1.de/2019/10/25/6-most-important-kpis-for-airline-operations-and-performance-analysis/.

[2] "ON-TIME PERFORMANCE OTP IS BECOMING INCREASINGLY IMPORTANT TO AN AIRLINES AND AIRPORTS," OAG Aviation Worldwide Limited, 2021. [Online]. Available: https://www.oag.com/on-time-performance-airlines-airports.

[3] G. D. Gosling, "Aviation System Performance Measures," UC Berkeley, 1999. [Online]. Available: https://escholarship.org/uc/item/2xw9204x.

[4] G. a. J. G. Box, Time Series Analysis: Forecasting and Control, San Francisco: Holden-Day, 1970.

[5] S. Chatterjee, "Time Series Analysis Using ARIMA Model In R," 05 02 2018. [Online]. Available: https://datascienceplus.com/time-series-analysis-using-arima-model-in-r/.