Analysis of Factors in Flight Delay

Yiyang Xu[1], Luyao Liu[2], Xichen Gao[2] and Fanyu Frank Zeng[3,*]
[1]Southwest Jiaotong University, Chengdu, China
[2]Central University of Finance and Economics, Beijing, China
[3]Indiana Wesleyan University 1900 West 50th Street, Marion, Indiana 46953, USA
[*]Corresponding author

Abstract—**With the advancement of air transport industry, airline travelling is becoming increasingly prevalent in recent years. Flight delay, however, has been a headache and costly issue for both air companies and travelers, leading to inconvenience in our daily life. We can't help asking: what's the major cause for flight delay? In this paper, visualization and multiple linear regression are implemented based on Chinese flight data. Temperature, previous delay rate, month and weekday are figured out as influential factors. Further improvements are made by extracting variables to explore the optimal regression mathematical model. Then, machine learning algorithms are introduced to make future predictions.**

Keywords—flight delay; multiple linear regression; modeling, prediction; machine learning

I. Introduction

During the past decades, great changes have taken place in modern transportation system. As one of the “New Four Inventions”, airplane plays an indispensable role in our daily life. Along with the air transportation boom, however, is the potential problem that need to be addressed and eventually solved, that is, flight delay.

It is reported that the domestic flights in China arrive 21 minutes late on average. In fact, multiple factors may contribute to large scale flight delay, including typhoon, haze, air traffic control, aircraft maintenance, mechanical failure, etc. Therefore, we want to take a deeper look to find out the major causes for flight delay and predict one airline’s delay according to its historical data, which can benefit our decisions when booking flight tickets and selecting transportation in the future.

II. Objective

In this research, a number of causes to flight delays are first investigated and explored based on historical flight delay data published by Chinese Air Traffic Control Authority. First, we process the original data to identify the most effective variables and disregard irrelevant ones. Then, we used Python and statistical software program to analyze data, conduct descriptive statistical analysis, and visualize the analysis results in order to understand the result data better. For example, visualized figures about the delay frequency of Chinese airlines are generated. We then developed multiple linear regression model in SAS to analyze the significance of influential factors or variables, such as, temperature, previous delay rate, month and weekday. Based on preliminary analysis results in SAS, we improved previous analysis method and model by extracting and filtering independent variables and generated an optimal regression model. Then, machine learning algorithm with main focuses on future prediction and implementation of model[1].

III. Data Collection

Our data is composed of four different raw datasets, which are flight, airport location, city weather, and airport’s special situation.

Flight: We found flight data from Ctrip's website at https://yunhai.ctrip.com/Games/11. The data contains more than 7,000,000 historical flights within two year of period from May, 2015 to May, 2017, including IATA Code of Departure Airport, IATA Code of Departure Airport, Flight Number, Scheduled Departure Time, Scheduled Arrival Time, Actual Departure Time, Actual Arrival Time, Plane's Number, and Where the flight is canceled.

Airport Location: We found data from Baidu Wenku https://wenku.baidu.com/view/594036ac04a1b0717fd5 ddfa.html. This data includes the IATA codes and the cities of all the airports in China.

Weather: We found this data from the web of National Meteorological Information Center: http://data.cma.cn/data/cdcindex/cid/ 69be9dbf72605049.html. The data includes the weather table of historical cities within two year of period from May, 2015 to May, 2017, including cities, weather, the highest temperature, the lowest temperature and the date.

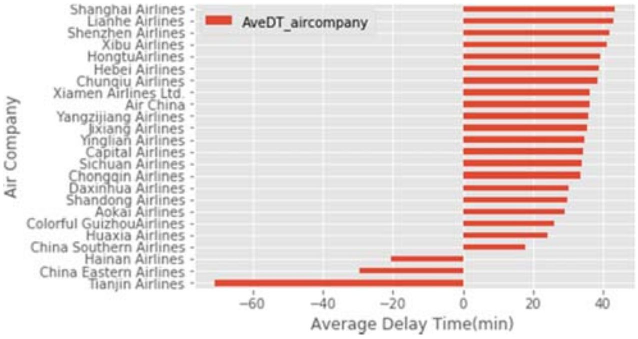
Airport's Name: As the airport information in our flight data is represented by IATA code, so we used the information of airport's code and Chinese name from Baidu Baike: https://baike.baidu.com/item/%E4%BA% 8C%E5% AD%97%E7%A0%81/8016030?fr=Aladdin

IV. Exploratory Data Analysis

To better understand the analysis results, we first used Python to visualize our data analysis results by airport, airline, and dates in order to get a picture for better understanding of those results.

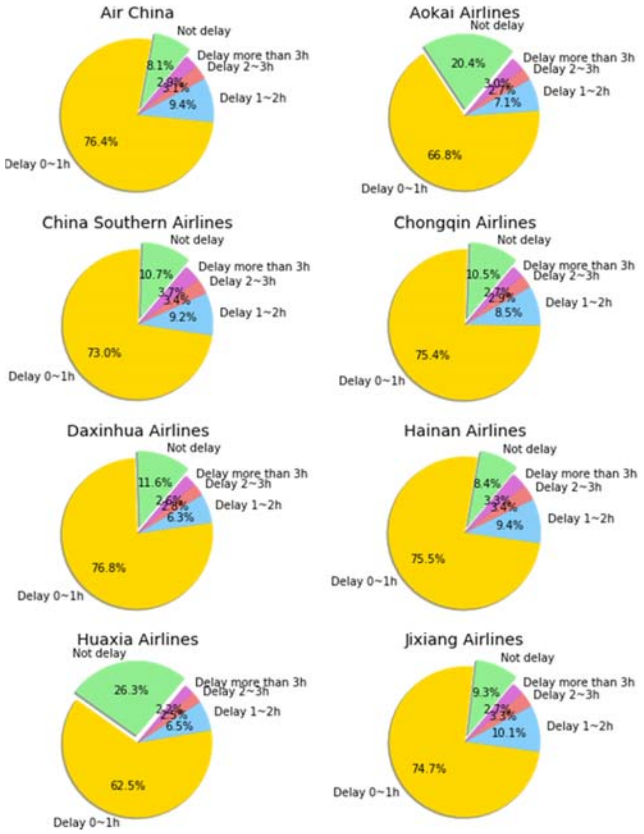
A. By Airline Company

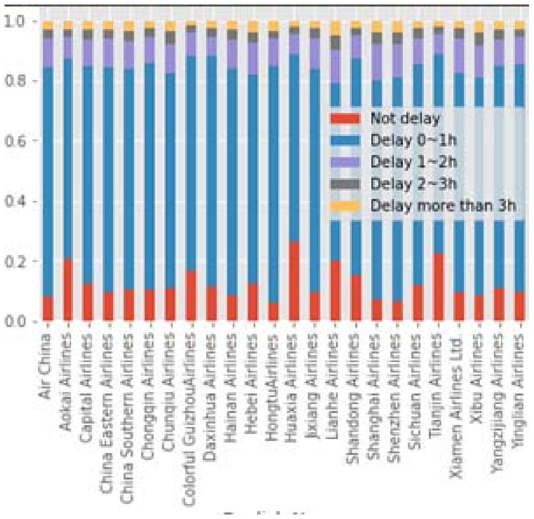
We grouped the flight data by airline and used statistic analysis result, averages to demonstrate each airline’s average delay time during the two year of period.



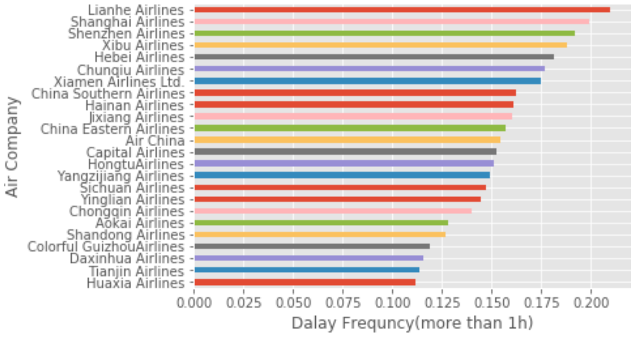
As we can see that the average delay time among those companies are quite different, Shanghai Airlines has the longest average delay time, 41minutes.

In order to see the performance of different companies better, for each airline we calculated frequencies of average delay time in 4 categories, which are not delay, delay within 1 hour, delay from 1 hour to 2 hours, delay from 2 hours to 3 hours and delay more than 3 hours. Then we draw pie charts to show each group's rate like this:





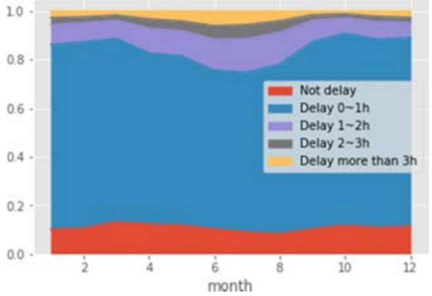
As we find that it is common that among all the airlines we studied, more than 80% of the flight will delay and around 70% of the flight will delay within 1 hour, so we can conclude that most of flights will likely be delayed within 1 hour.



B. By Travel Date

It is always interesting to find out if there exists any pattern from other factors, such as certain date (the days around holiday). Hence, we also process some data to reveal the delay pattern by date.

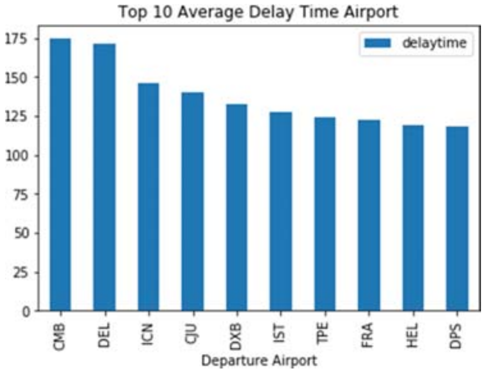
We are interested to discover which month is the busiest month in an entire year. We assume the busiest month is in winter or summer because there is more demand for flights from vacationers during the period of time in these periods and generated graph confirms that summer is the busiest month.



As you can see, vacation times are indeed busier than other time. The red area, which represents not delay, is smaller between June and August, while the purple area, which represents delay between 1hour and 2 hours area is larger. Therefore, we think that it's important to add the factor of time into our prediction model.

C. By Airport

Different airports perform differently, due to different sizes, flight numbers and so on. In this part, we want to see which airports perform well and which airports perform bad. Therefore, we calculate the average delay time of different airport and draw them into a bar graph. The longest Average delay time is about 175 minutes.



V. MULTIPLE LINEAR REGRESSION WITH SAS

In the previous part, we find that the average delay time of different companies, different airport, different airlines are greatly different, so we think the previous average delay time have great relationship with the delay time we want to predict. Therefore, in this part, we use multiple linear regression on SAS based on the independent variables of Average Delay Time of airlines, Average Delay Time of air companies, Average Delay Time of airports, and the dependent variable is Delay Time,

A. Model Construction

1) Linear Regression Models

In order to relate the response variable y to the predictor variables $x_1, x_2, \dots, x_k$, linear regression model can be used. The general form of the model is:

$\mathrm{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$, where $\varepsilon \sim N(0, \alpha^2)$, and α^2 is a constant.

If $x_1 = x, x_2 = x^2, x_3 = x^3, \dots, x_k = x^k$, then the model will be $\mathrm{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k + \varepsilon$. It is called a polynomial regression model which is relating the response variable y to a single predictor variable x.

Actually, if we want to relate the response variable y to more than one predictor variable, there are many different models. For example, the predictor variables are x_1 and x_2 :

$$\mathrm{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon,$$

$$\mathrm{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2^2 + \beta_5 x_2^3 + \varepsilon,$$

$$\mathrm{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon,$$

and so on.

In the model $\mathrm{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$, we can notice that there is a cross-product term $\beta_3 x_1 x_2$, which means that if x_2 is held fixed, the change in E(y) for a 1-unit change in x_1 is dependent on the value of x_2. Because of this, x_1 and x_2 are said to interact, and this model is called an interaction model.

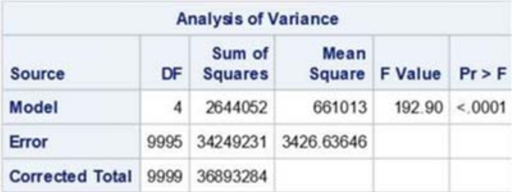
2) Construct the Original Model

Initially, due to the three variables: the average delay time of airline company (x1), the average delay time of airline (x2), the average delay time of airport (x3), we can assume the model is:

$$\begin{aligned}\mathrm{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 \\ + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \beta_7 x_1^2 + \beta_8 x_2^2 \\ + \beta_9 x_3^2 + \varepsilon\end{aligned}$$

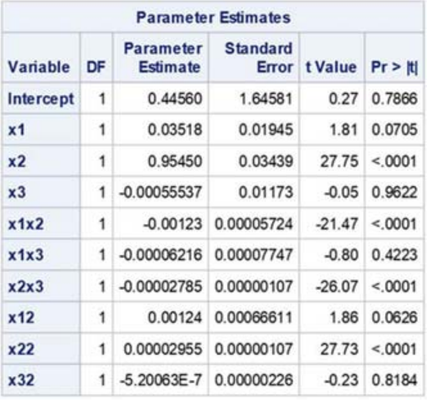
Then, it is necessary to conduct a test for the linear regression model at the 0.05 level of significance.

Test $H_0: \beta_1 = \beta_2 = \cdots = \beta_9 = 0$ against $H_1: at\ least\ one\ of\ \beta_i\ is\ nonzero$



Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2644052	661013	192.90	<.0001
Error	9995	34249231	3426.63646		
Corrected Total	9999	36893284			

From the ANOVA table, we can see that p-value<0.0001. Therefore, H_0 is rejected at the 0.05 level of significance and we can conclude that the model makes sense to some degree.



Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.44560	1.64581	0.27	0.7866
x1	1	0.03518	0.01945	1.81	0.0705
x2	1	0.95450	0.03439	27.75	<.0001
x3	1	-0.00055537	0.01173	-0.05	0.9622
x1x2	1	-0.00123	0.00005724	-21.47	<.0001
x1x3	1	-0.00006216	0.00007747	-0.80	0.4223
x2x3	1	-0.00002785	0.00000107	-26.07	<.0001
x12	1	0.00124	0.00066611	1.86	0.0626
x22	1	0.00002955	0.00000107	27.73	<.0001
x32	1	-5.20063E-7	0.00000226	-0.23	0.8184

B. Model Improvement

1) Comparing two Nested Models

If a model contains all terms of the other model and at least one additional term, we can say that the two models are nested.

Consider two nested models as follows:

Completed model: $\mathrm{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \beta_{p+1} x_{p+1} + \cdots + \beta_k x_k + \varepsilon$

Reduced model: $\mathrm{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon$

We definitely want to choose the more appropriate model to describe the dataset. To be more specific, it is equivalent to test $H_0: \beta_{p+1} = \cdots = \beta_k = 0$ against $H_1: at\ least\ one\ of\ \beta_{p+1}, \ldots, \beta_k\ is\ nonzero.$

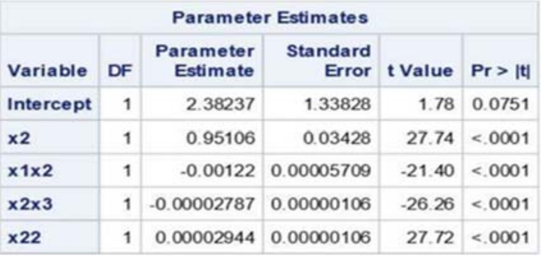
Obviously, under H_0, the complete model becomes the reduced model with fewer parameters, which means that the reduced model is more parsimonious. When the two competing models have essentially the same predictive power, we should definitely choose the more parsimonious of the two. Then we use SSE_C and SSE_R to denote the residual sums of squares for the complete model and the reduced model respectively. The difference between them is called the drop in the unexplained variation attributable to the predictor variables $x_{p+1}, \ldots, x_k$. Therefore, we obtain the ratio:

$$F = \frac{SSE_D/(k-p)}{SSE_C/[n-(k+1)]}$$

which is called the partial F statistic, and H_0 is rejected at the α level of significance if $F > F_\alpha(k-p, n-k-1)$ or $\mathrm{p-value} < \alpha$.

2) Construct the Reduced Model

We can get the SAS result like this (part of the picture):



Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.38237	1.33828	1.78	0.0751
x2	1	0.95106	0.03428	27.74	<.0001
x1x2	1	-0.00122	0.00005709	-21.40	<.0001
x2x3	1	-0.00002787	0.00000106	-26.26	<.0001
x22	1	0.00002944	0.00000106	27.72	<.0001

Since the p-value of x1 (the average delay time of aircompany) is 0.0705, therefore the effects of x1 is not significant.

Since the p-value of x2 (the average delay time of airline) is less than 0.0001, therefore the effects of x2 is significant.

Since the p-value of x3 (the average delay time of airport) is 0.9622, therefore the effects of x3 is not significant.

Since the p-value of x1x2 (x1*x2) is less than 0.0001, therefore the effects of x1x2 is significant.

Since the p-value of x1x3 (x1*x3) is 0.4223, therefore the effects of x1x3 is not significant.

Since the p-value of x2x3 (x2*x3) is less than 0.0001, therefore the effects of x2x3 is significant.

Since the p-value of x12 (x1*x1) is 0.0626, therefore the effects of x1 is not significant.

Since the p-value of x22 (x2*x2) is less than 0.0001, therefore the effects of x22 is significant.

Since the p-value of x32 (x3*x3) is 0.8184, therefore the effects of x32 is not significant.

Therefore, x2, x1x2, x2x3, x22 affect the delay time more.

Thus, we assume the reduced model is:

$$y = \beta_0 + \beta_1 x_2 + \beta_2 x_1 x_2 + \beta_3 x_2 x_3 + \beta_4 x_2^2 + \varepsilon$$

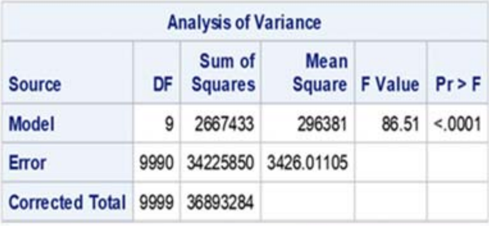
3) Model Improvement Validation

The two models are as follows:

Complete model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \beta_7 x_1^2 + \beta_8 x_2^2 + \beta_9 x_3^2 + \varepsilon$

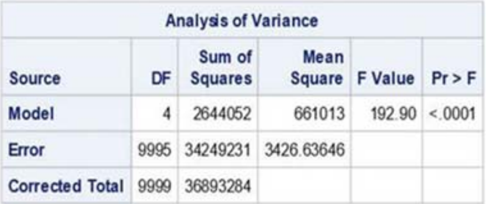
Reduced model: $y = \beta_0 + \beta_1 x_2 + \beta_2 x_1 x_2 + \beta_3 x_2 x_3 + \beta_4 x_2^2 + \varepsilon$

It is equivalent to test $H_0: \beta_5 = \beta_6 = \cdots = \beta_9 = 0$ against $H_1: not\ both\ \beta_5, \beta_6, \beta_7, \beta_8\ and\ \beta_9\ are\ zero.$



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	2667433	296381	86.51	<.0001
Error	9990	34225850	3426.01105		
Corrected Total	9999	36893284			



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2644052	661013	192.90	<.0001
Error	9995	34249231	3426.63646		
Corrected Total	9999	36893284			

From the SAS printouts for the complete model and the reduced model, we can see that

$SSE_C = 34225850$ and $SSE_R = 34249231$. Thus,

$$F = \frac{\frac{SSE_D}{k-p}}{\frac{SSE_C}{[n-(k+1)]}} = \frac{\frac{34249231 - 34225850}{9-4}}{\frac{34225850}{10000-10}} = 1.3649$$

$$< 2.21 = F_{0.05}(5{,}10000)$$

$$\text{or} \ < 3.02 = F_{0.01}(5{,}10000)$$

Hence, H_0 is not rejected at the 0.05 level of significance and we conclude that the reduced model is appropriate (the quadratic terms should not be retained in the model).

C. Parameter Estimation

1) Theory

The parameter of the linear regression model can be written in the vector form: $\beta' = [\beta_0, \beta_1, \beta_2, \ldots, \beta_p]$, and the dataset $\{(x_{j1}, x_{j2}, \ldots, x_{jp}; y_j), j = 1,2, \ldots, n\}$ can be organized into two matrices:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

Therefore, the general linear regression model can also be written in the form $Y = X\beta + \varepsilon$, where ε is a column vector.

The least squares estimator vector $b' = [b_0, b_1, b_2, \dots, b_p]$ for the parameter β is the minimizer of the quadratic loss function (SSE):

$$L(b_0, b_1, b_2, \dots, b_p) := SSE = \sum_{i=1}^{n}\left(y_i - (b_0 + b_1x_{i1} + b_2x_{i2} + \cdots + b_px_{ip})\right)^2$$

$$= (Y - Xb)'(Y - Xb) = b'(X'X) - 2b'X'Y + Y'Y$$

The solution of SSE should satisfy the first order optimality condition: $\frac{\partial L}{\partial b_i} = 0, for\ all\ i = 0,1,\dots,p$

Through this, we can obtain the linear equations $(X'X)b = X'Y$. Assume that the matrix X is full column rank, which is true in most cases for $n \gg p + 1$. Then, we can learn that:

(1) The least squares estimator vector is determined by $b = (X'X)^{-1}X'Y$

(2) The optimal values of L(SSE) is equal to $b'(X'X)b - 2b'X'Y + Y'Y = -b'X'Y + Y'Y$. Due to this, the parameter σ^2 estimated by $s^2 = \frac{SSE}{n-(p+1)} = \frac{Y'Y - b'X'Y}{n-(p+1)}$

(3) Because $E(Y) = X\beta, V(Y) = I$, so $V(b) = (X'X)^{-1}X'V(Y)X(X'X)^{-1} = \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}$. Therefore, $b \sim N^{p+1}(\beta, \sigma^2(X'X)^{-1})$.

2) *Application*

From the coefficients table, We obtain the following estimated regression equation for the linear regression model:

$$\hat{y} = 2.38237 + 0.95106x_2 - 0.00122x_1x_2 - 0.00002787x_2x_3 + 0.00002944x_2^2$$

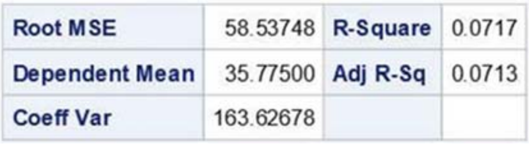
3) *Multiple Coefficient*

1. The multiple coefficient of determination R^2 is defined as $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$, where $SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 = Y'Y - n\bar{y}^2$, $SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = b'X'Y - n\bar{y}^2$, $SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = Y'Y - b'X'Y$. It represents the proportion of the total variation in y that can be explained by $x_1, x_2, \dots, x_p$ through the multiple regression model. At the same time, the value $R = \sqrt{R^2}$ is called the multiple correlation coefficient.

2. To test $H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$ against $H_1: at\ least\ one\ of\ \beta_i\ is\ nonzero$

Test statistic: $= \frac{SSR/p}{SSE/(n-p-1)}$, and H_0 is rejected at the α level of significance if $F > F_\alpha(p, n-p-1)$ or $p - value < \alpha$.

4) *Results*



Root MSE	58.53748	R-Square	0.0717
Dependent Mean	35.77500	Adj R-Sq	0.0713
Coeff Var	163.62678		

Since $R^2 = 0.0717$, about 7.17% of the variation in y (delay time) can be explained by x2, x1*x2, x2*x3 and x2*x2 through the estimated equation. Unfortunately, the result is not very well.

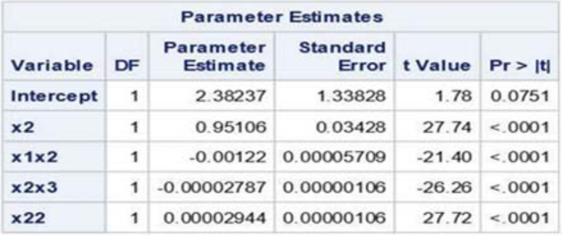
VI. Delay Classfication Prediction Based on Machine Learning with Python

After the first try, we found that the R, of our multiple linear regression model is very small, which means it's not a good model for use. We think it's because that the specific delay time is so complicated that we can't just use the previous average delay time to predict. It is greatly influenced by the random situation, even the emotion of the captain.

Therefore, in this part, we changed our object to predict whether the flight will delay more than 1 hour, which is a classification prediction problem. Also, we explored more on our data and use the information of weather and airport's special situation. And to realize these purposes, we use the tool of Python.

A. *Feature Engineering*

We created those variables using python as our dependent variables



Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.38237	1.33828	1.78	0.0751
x2	1	0.95106	0.03428	27.74	<.0001
x1x2	1	-0.00122	0.00005709	-21.40	<.0001
x2x3	1	-0.00002787	0.00000106	-26.26	<.0001
x22	1	0.00002944	0.00000106	27.72	<.0001

While our independent variable is also a boolean variable which means whether the flight delay more than one hour.

Variable	Type	Meaning
Delay1h	Boolean	Whether the flight delay more than 1 h

We can see the variables' correlations first:

	Whether Change	*Whether Rain*	*Whether Snow*	*AVG Rate*	*y*
Whether Change	1.00	0.36	0.10	0.00	0.02
Whether Rain	0.36	1.00	(0.06)	0.03	0.08
Whether Snow	0.10	(0.06)	1.00	(0.01)	0.01
AVG Rate	0.00	0.03	(0.01)	1.00	0.27
y	0.02	0.08	0.01	0.27	1.00

B. Methods

1) Logistic Regression

The logistic regression can be understood simply as finding the β parameters that best fit:

$$y = \begin{cases} 1 & \beta_0 + \beta_1 x + \varepsilon > 0 \\ 0 & else \end{cases}$$

where ε is an error distributed by the standard logistic distribution. (If the standard normal distribution is used instead, it is a profit model.) [4].

	Positive	Negative
Positive	27573	266
Negative	4870	291

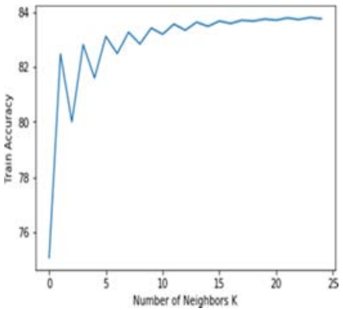
2) K-Nearest Neighborhood

The k-nearest neighbor algorithm (k-NN) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space[2]. The output depends on whether k-NN is used for classification or regression:

In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.

In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbors. The best choice of K depends on the data; Generally, a larger K value reduces the impact of noise on classification but makes the boundaries between classes less obvious. K can be selected by various heuristic techniques (see hyperparametric optimization). The special case of predicting the closest class of training samples (i.e., when k = 1) is called the nearest neighbor algorithm[2].

In our case, we changed our objective from predicting the delay time to predict whether it could delay more than one hour, so we use KNN to perform a classification model and based on the accuracy, the optimal number of neighbors is 23 with 83.8%.



	Positive	Negative
Positive	27516	323
Negative	4879	282

3) Decision Tree

The decision tree can be linearized into decision rules, where the outcome is the contents of the leaf node, and the conditions along the path form a conjunction in the if clause. In general, the rules have the form[3]:

if condition1 and condition2 and condition3

then outcome.

Decision rules can be generated by constructing association rules with the target variable on the right. They can also denote temporal or causal relations.

	Positive	Negative
Positive	25946	1893
Negative	4488	673

4) Results

Except for those 3 methods above, we also use some machine learning packages in the sklearn package of Python. After running, we get seven confusion matrix and calculate the accuracy:

Models	Accuracy
Gradient Boosting	0.845061
Logistic Model	0.844364
K-Near Neighbors	0.842364
Gausian NB	0.833909
Support Vector Machine	0.830939
DecisionTreeClassifier	0.806636
Random Forest Classifier	0.799242

As shown above, the highest accuracy of our model is 84.5%, which is much better than our previous work.

VII. Future Work

In this research, we selected our 4 raw data sets for 2 year period of time from 4 different reliable sources, Ctrip(a Chinese travel product website), Baidu Baike, National Meteorological Information Center and Baidu Wenku separately. In the future, we plan to select more data other sources in order to improve the accuracy of our mathematical model and results.

Also, in the sixth parts, we use three machine learning models and four other models which are included in the sklearn packages of Python. And we plan to change the parameters in each model in order to discover the best results.

VIII. Conclusion

In this paper, we try different methods to reveal the patterns of flight delay by using a number of mathematical and statistical tools to analyze historical data from 4 reliable sources. Since visualization is a good tool to help us better understand our analysis and results and to further reveal real cause and patterns. In addition to simple statistical analysis method and tools, multiple linear regression is used in analysis before machine learning methods are used to better predict a classification problem. At last, we found that classification prediction performs much better than the specific number prediction, and we think this may be caused by huge random factors. Besides our research have discovered certain findings and patterns, we also believe future work is necessary and should focus on improvement of accuracy and performance of our prediction model.

REFERENCES

[1] Martinez V. Flight Delay Prediction[D]. ETH Zürich, Department of Computer Science, 2012.

[2] Yao R, Jiandong W, Jianli D. RIA-based visualization platform of flight delay intelligent prediction[C]//Computing, Communication, Control, and Management, 2009. CCCM 2009. ISECS International Colloquium on. IEEE, 2009, 2: 94-97

[3] Kim Y J, Choi S, Briceno S, et al. A deep learning approach to flight delay prediction[C]//Digital Avionics Systems Conference (DASC), 2016 IEEE/AIAA 35th. IEEE, 2016: 1-6.

[4] Rebollo J J, Balakrishnan H. Characterization and prediction of air traffic delays[J]. Transportation research part C: Emerging technologies, 2014, 44: 231-241.