NOVA SCHOOL OF BUSINESS & ECONOMICS

Research Report

```
print('Flights analysis in the US')
```

```
Professor = ['Pedro Garcia']
TA = ['João Fonseca']
Course = ['Introduction to Programming']
Semester = ['Fall 2018']


Group = { 'Bruno Carvalho' : 13674, 'Catarina Antunes' : 14318,
'João Santos' : 14099}
```

Index

● Introduction

The Air transportation sector plays a major role in our life, not only allowing millions of people to travel, but also contributing to the creation of employment, and therefore generating a richer welfare and stimulating the overall economy, representing 3.5% of the world GDP. This sector is in constant increasing **(Annex. 1)**, and the proof is, the fact, that the number of people flying increased by 6.6% in 2017. The reasoning behind this situation, can be explained by a combination of the process of globalization, creating new destinations and therefore more supply, and the progressively technical improvements, granting a reduction of costs, which might be associated with the increasing of the number of airplanes. Despite the overall scenario might be extremely positive, this sector also has its problems. One of them is, in fact, associated with the delays of passenger flights. These delays can be explained by several factors:

a) Extreme weather conditions, for example snow storms and thunderstorms, which can temporarily create major disruptions;

b) Air carrier delays associated with aircraft maintenance and fuelling;

c) Heavy traffic volume and air traffic control, which can aggravate the overall situation of total delays.

Thus, the purpose of this paper is to understand the delays in domestic flights relatively to the provided data, which concerns only the USA. Finally, the group is going to construct a model to predict the number of minutes a flight will be delayed.

- ## Data Configuration

From implementing the data, regarding the number of flights (flights.csv), in the visualization window of Jupyter Notebook, the group was able to conclude that the original file has 570118 rows, but also has 570118 rows that contain unfilled attributes. The reasoning behind this is because, the column "Unnamed: 23", had only null values, therefore all rows had at least one null value. After researching about the variables, we concluded that there were some problems associated with the Data Frame. Firstly, the last column, "Unnamed: 23", had only null values, therefore their explicative value was absolutely none, leading the group, to remove it. After the removal, the number of rows is still the same, 570118, but now the total number of rows containing null values is, in fact, 19096.

Complementing the Data Frame with the file containing Carriers (carriers.csv), the group was able to associate each code of the carrier to its description (i.e "UA" stands for United Airlines Inc.), allowing for a better understanding of the reality of the current framework, relatively to the situation of carriers.

Regarding the least explanatory columns the group decided to eliminate 'TAIL_NUM' , 'OP_CARRIER_FL_NUM' and 'ACTUAL_ELAPSED_TIME' due to their lack of informative value. We substituted the missing values present in the columns 'ARR_TIME', 'ARR_DELAY', 'DEP_DELAY' and 'AIR_TIME' by their median. However, we did not fill the null values in 'DEP_TIME' since the group considered that this variable was in another unit (HH:mm). After all these changes, the total number of rows is 553.295, and the total number of rows containing null values is 0, as expected.

- ## Data Analysis

Regarding the number of flights per day, the group decided to represent the plot with the graphing library Plotly, to display a more visual representation of the graph. The number of flights reached its highest level in the second day of the month and had its lowest value in day 13th **(Annex. 2)**. The average number of flights in the month of January 2018, was, approximately, 17848 flights. Furthermore, another important consideration about this data, is the following: all of the Saturdays in this month correspond to lowest values of flights, which can be explained by the difference in prices, but also by the psychological mindset that our society shares, associated with the fact, that individuals prefer to fly on a Friday, related with business trips and short-term vacations. In the first week of the month, currently in the New

Year Holidays, the number of flights is way more consistent, than the other weeks, and this, also, can explain the fact, the number of flights hit their highest, exactly in this week.

Taking into account the number airlines, in the current Data, there's 18 unique carriers. The firms that have the highest numbers are: Southwest Airlines Co., American Airlines Inc. and Delta Air Inc., with 107.240, 72.157 and 69.754 flights, respectively. On the other hand, the companies that have the least flights are: Allegiant Air, Hawaiian Airlines Inc. and Virgin America with 6.761, 6.617 and 5.722 flights, respectively. **(Annex. 3)[1].**

The Southwest Airlines Co. firm, is one of the most influential airlines in the world and their success is due to, not only charging the lowest possible fare, but also proportioning a new level of content that no other airline is bringing to the table: fun, associated, for example with love potions theme (on board drinks). Contrarily, we have Allegiant Air, and their unsuccess regards their lack of safety, in terms of aircrafts. According to *The Times* , the famous magazine found that "the budget carrier's planes are four times as likely to fail during flight as those operated by other major U.S. airlines.", and also stated that "Forty-two of Allegiant's 86 planes broke down in mid-flight at least once in 2015."[2] This aspect combined with a lower destination possibility that Allegiant Air offers, explains the lowest number of flights in January, 2018. Concerning, the variation of the number of flights per week **(Annex. 4)**, the graphical representation was conducted, again, with the help of the library Plotly, which allowed us to represent all the Airlines flights per week. In structural terms, the code was possible due to a *"for"*, which loop the entire list of unique Airlines and representing each line, in just one plot. Secondly, we conceived a Dictionary, having as keys, the name of the Airline, and in their respective values, the list containing the five values of Number of flights, one for each week. After analysing the graph, we conclude that the last week of January, is the one that has the lowest flights recorded. The tendency that almost Airlines demonstrate is that, the first week of January displays the highest number of flights, again, explained by the New Year Holidays, while the last week is the one that has the lowest, due to only having three days (from 28th to 31th). Furthermore, it's relevant to consider that the firm Southwest Airlines Co. is the one that has the highest number of flights in all five weeks, and also is the one that exhibit a higher variation, ranging from 24.912 flights in week one, to 10.886 to week five, leading to an overall standard deviation of 5.522,5. On the other side, Virgin America is the Airline that shows the least variation, and so being the more consistent firm, ranging from 1.272 in week one to 566 flights in week five, which corresponds to an overall standard deviation of 298,1. [3] The group

[1] Annex.2 includes all the number of flights for each Airline.
[2] https://www.smartertravel.com/allegiant-airs-planes-fail/
[3] Standard deviation takes into account the five values for each week of the month.

considered that this variation is significative only in a few Airlines, such has the already mentioned, Southwest Airlines Co., but also the American Airlines Co., Delta Airlines Co. and SkyWest Airlines Inc., in which have a standard deviation higher than 3.000 flights, indicating that the data points are spread out over a wider range of values. Compared for example with Virgin America, Hawaiian Airlines Inc., Frontier Airlines Inc. and Allegiant Air, the group recognized that these variations were not much significative, because it had a standard deviation below the value of 600, corresponding to a closer relationship between the data points and the mean.

Moving forward to the total number of flights in the US Airports, we decided to plot a stack bar chart, in order to visualise the interaction between the combinations of flights per day, according to the airports.

Analysing the five largest Airports, these are: ATL, ORD, DFW, CLT and DEN **(Annex. 5).** In all thirty-one days of January, the Airport ATL, in Atlanta, leads the number of flights, reaching its highest in the day 29 of January with 1083 flights, which is followed by ORD, in Chicago, achieving their peak in 2nd of January, with 880 flights. It's worth mentioning that, the average values are: 988.0, 828.0, 721.0, 592.0 and 588.9, for ATL, ORD, DFW, CLT and DEN, respectively and the days that these Airports have the least number of flights, corresponds to Saturdays.

According to the current data, the worst performing Airports are: YNG, BFF, ADK, HGR and OWB, this is associated to the fact that these Airports are regional ones, therefore the number of flights will be, obviously, lower **(Annex. 6)**. In all thirty-one days of January, both YNG and BFF only have two flights, while YNG has these flights in the first week, which might be associated with the Holidays season, BFF only has its flights at the end of the month. A more concrete analysis shows that in these airports, there's no flights on Sundays, and the reasoning behind this, is simple, since there's already a lower total number of flights on Sundays, the one's that exist, are on major Airports. When choosing one of the best performing Airports, with the help of our script **(Annex. 7)**, we were able to get a more detailed information regarding the number of flights per day and also the growth rate. The script allows for the user to choose one of the Top 5 largest Airports and analyse their respective number of flights and growth, for example, we'll choose the Airport "ATL". Regarding "ATL", Hartsfield-Jackson Atlanta International Airport, the patter of flights is clear to see. On the other hand, ATL, has the most flights on Fridays, which represents the earlier discussion about the society mindset, in terms of preferring to travel in a Friday, which can be associated with short-term vacations and also business trips. On 6th, 13th, 20th and 27th, the busiest Airport has the least flights in January, and curiously, this corresponds, again, to Saturdays, the explanation is associated with the one presented aforementioned. Concluding, it is on Saturdays that the growth rate is the lowest, because of the negative evolution from Friday to Saturday, reaching most of the times a growth of -30%. Generally speaking, all the largest airports behave this way, thus, when moving from Friday to Saturday, according to the data provided, it's likely that the number of flights will decrease.

After closely investigating the relationship between the Airports, the group elaborated two functions to analyse the existing connections between airports. The difference between these functions, is the fact, that the user can sort the carrier to fly with.

- **Function Name:** *ComCarrier(origin, carrier)*

To the first one, the function inputs are: Place of take-off ("origin") and the Airline ("carrier"). The group submitted a query regarding the "origin", and then on the new DataFrame already containing the query, we divided the data again, but now subject to the "carrier". Next, we grouped the data according to "DEST", and solved some problems related with the visualization of the data, namely changing the name of the column to "Number of Flights". Still associated with this function, the group did exactly the same thing, but now considering our "origin" as our destiny, getting all connections from other airports that land in "origin".

- **Function Name:** *SemCarrier(origin)*

To the second one, the function input is: Place of take-off ("origin"). The group submitted a query regarding the "origin". Next, we grouped the data according to "DEST", and solved some problems related with the visualization of the data Frame, namely changing the name of the column to "Number of Flights" and, also, changing their position. Still associated with this function, the group did exactly the same thing, but now considering our "origin" as our destiny, getting all connections from other airports that land in "origin".

For example, if we choose "LAX" as our place to take off, and select the carrier "UA", we will get 42 connections and two tables representing the possible destinations and the origins that go to "LAX" **(Annex. 8).**

The average number of flights per hour is: 766.0. The method to reach it was, simply, get all values of the number of flights per hour according the to the day, and then the average of these values. Regarding to the code that allowed to represent graphically the number of flights per hour according to the day, we concluded that most flights are, either, at 00:00h or at 04:00h. So, answering the second part of this question, the number of flights decreases drastically in the other hours. For example, in day 1 **(Annex. 9)**.: the lowest number of flights occurred at 13:00h with 650 flights, and the highest value of flights occurred at 00:00 with 1018 flights. We can conclude that the number of flights is clearly influenced by the hour of the day.

Considering the previous analysis, the distribution of the number of flights per hour is similar regarding the different seven days of the week. After analysing the correspondent graphs, we observed that the majority of flights continue to be scheduled, either, at 00:00h or at 04:00h and then the number of flights decreases severely in the rest of the day. Nevertheless, on Sundays **(Annex. 10)** there are even more hours with low numbers of flights than in the remaining days of the week.

In relation to the average duration of a flight the mean is 111,59 minutes. Taking into consideration that flight duration can be categorized into: short-haul flights (less than 180 minutes), medium-haul flights (180-360 minutes), long-haul flights (361-720 minutes) and ultra-long-haul flights (more than 720 minutes), the percentage of each category regarding the flights.csv dataset is the following: short-haul flight 82,98%, medium-haul flight 13,38%, long-haul flight 0,41% and ultra-long-haul flight 3,23%. So, the average duration of a flight in this dataset being included in the short-haul flights category was expected by the group. In fact, this is because all flights are only made between United States cities, so the American airlines fly point-to-point destinations domestically. Therefore, flights that take less than 360 minutes are the overwhelming majority accounting for 96,36% of total flights.

- ## Delays

Regarding the delays, we decided to create a dataframe where we could observe how many flights were cancelled, how many flights were diverted and how many flights reached the assigned destination. From our dataframe we can also observe the average air time per flight, the average distance and the average arrival delay, besides we can observe all these variables by the different carriers. Analysing the table below, the carrier with the most cancelled flights was WN with 2467 flights. However, in percentage was OH with 6.99% of all flights delayed. On the diverted flights the carrier with the most was OO with 291 flights diverted. To conclude the number of flights with a correct destination was 568.869.

In January 2018 there was a total of 183.723 flights delayed. The percentage of flights departing late is 33.24 %. The main reasons for delayed departures are the three ones stated previously in the introduction. In regard to the percentage of flights arriving ahead it accounts for about 65.72 %. This can be a result of several factors such as wind in favour and departure made ahead of time. Furthermore, there are some reports that claim that airlines are extending flight times to avoid pay-outs for delays.

According to research done by OAG, the aviation analyst, despite new technology has been introduced over the last years, scheduled flight times have actually increased by more than 50%. Air Transportation industry names this practice “schedule padding” and defends that it aims to improve punctuality. However, some argue that “airlines just have adjusted their flight arrival times, so they can have a better record of on-time arrivals, so they might say a flight takes two hours when it really takes an hour and 45 minutes.”

For the question 3 we decided to use a scatter plot where we could observe the frequency of the delays (both departure and arrival) considering the number of minutes **(Annex 13, Annex 14)**. Derived from curiosity we calculated the average arrival delay (2.83 min) and the average departure delay (9.5 min). We also decided to add the opportunity to let the user choose which company wants to know how many flights delayed it had during the month. Most of the arrival delays are often connected with congestion.

Infrastructure congestion is also used as a competitive advantage, since some individual airlines can benefit by limiting access to its competitors allowing to practice higher prices.

From our analysis during January of 2018 the day with the highest percentage of departure delays was 1 ,with a percentage of 51.8% , with the lowest percentage value of delays was the day 27 with a percentage of 15.8%. To calculate the difference in flight volume we did the difference among all the flights with the day that had an higher percentage of flights delayed with the day with the lowest percentage of delays. We reached a total value of 2770 flights.

Plotting the number of delays through the different days of the month allows to conclude that in average the total number of delays is decreasing as the end of the month is approaching. However we can´t conclude if it would be a cycle through the different months of the year or if it would be always decreasing **(Annex 11 and Annex 12)**.

- **Selection of variables**

In this topic, we are going to transform the current variables, into ones that can improve our model. After understanding the phenomenon of delays, either by data manipulating through Jupyter Notebook, or by researching more information online, the group was able to create two meaningful variables. The first one is “Velocity km/h”, and we were able to achieve this variable, by dividing the column “DISTANCE” by “AIR_TIME” of the current data Frame, and later on, divided by 60. This variable will proportionate us the ability to, check the correlation between this new variable(“Velocity km/h”) and the other ones, particularly the departure delays and arrival delays. The second one is a binary/dummy variable called "Dummy". This variable translates the criteria set by the group relatively to the percentage of flights that are going to have delays, for each airport. Hence, if this ratio is higher than 50%, then it's more likely that it will run into a delay, and therefore it will be allocated the number 1. On the other hand, if this ratio is lower than 50% then it's more likely that it will not run into a delay, and therefore, it will be allocated the number 0. For example, the airport BQN from Aguadilla, Puerto Rico, has a percentage of delayed flights higher than 50%, hence according to our criteria will be corresponded to the number 1, while FLL, the Fort Lauderdale–Hollywood International Airport, is the other way around, therefore will be assigned the number 0. Regarding the correlation matrix **(Annex. 15)**, the group decided not to focus in great depth, in the other variables, as they can be analysed in the report generated by the PandasProfiling, hence we are going to study the correlation between our new variables. In fact, there's a tremendous positively correlation between ARR_DELAYS and DEP_DELAYS, which makes sense, because if our flight is going to be delayed due to issues/problems in the departure, then it's more likely to have delays on arrivals. Secondly, our new variable "Velocity km/h", was a huge success. It is positively correlated with "AIR_TIME" and also "DISTANCE", which also is logical, due to, for example, more distance allow, the aircraft to reach an higher level of speed. Associated with the air time, which it's also positively correlated, we can think about this, as in the following view: less air time represents less time for the aircraft to have enough power to reach a higher level of speed, therefore more air time, according to our data, will lead to higher "Velocity km/h". Thirdly, our dummy variable is not correlated as much, as the group thought it was, nevertheless we decided to include it, in our Multiple Linear regression model, just to verify whether, it has statistical significance.

- ## Regression

In this topic, the group selected the variables 'Velocity km/h', 'Dummy', 'DISTANCE', 'DAY_OF_MONTH' and 'DEP_DELAY' since we considered that they were directly linked with 'ARR_DELAY'. We ensured that the selected variables were not highly correlated between themselves. The regression model looks like this:

$$ARR\_DELAY = \beta_1\ VELOCITY\ KM/H + \beta_2\ DUMMY + \beta_3\ DISTANCE +$$
$$\beta_4\ DAY\_OF\_MONTH + \beta_5\ DEP\_DELAY + \mu$$

Regarding the Multiple Linear Regression assumptions:

[MLR.1] - Linearity in Parameters: each of the parameters in the model is linear, so the assumption is verified.

[MLR.2] - Random Sampling: the data collected to estimate our model comes from the all domestic flights on USA, hence the entire population. Consequently, this assumption can be ignored since no sampling is made.

[MLR.3] - No Perfect Collinearity: already mentioned before, we ensured that the chosen variables were not highly correlated. The correlation between any two different regressors is different from 1, thus the assumption holds.

[MLR.4] - Zero Conditional Mean: the assumption holds if the expected value of the error term given any independent variable is zero, $E(\mu|x)=0$. To verify it, we would need to make sure that there are no omitted variables in our model. However, the model may in fact be missing some variables that we did not account for, and, therefore, create biased coefficients. We will assume that this assumption holds in order to infer results from unbiased and consistent coefficients. Given that assumptions MLR.1 to MLR.4 are verified, our OLS estimators are unbiased and consistent. Thus, we are able to see the real impact on the population.

Knowing that "statsmodels" library allow us to fit statistical models using R-style formulas, we started by importing statsmodels.formula.api and statsmodels.stats.api. Then we fit the regression model by doing *sm.OLS(Y,X).fit()* and we examined the results by doing *results.sumary()*. Automatically, a table of OLS Regression Results was generated and several statistics results were provided, namely the R-squared with a value of 0.907. Since it is close to 100%, it indicates that the model explains almost all the variability of the response data around its mean. Hence, 90,7% of the variation of the dependent variable (ARR_DELAY) is explained by the variation of the independent variables.

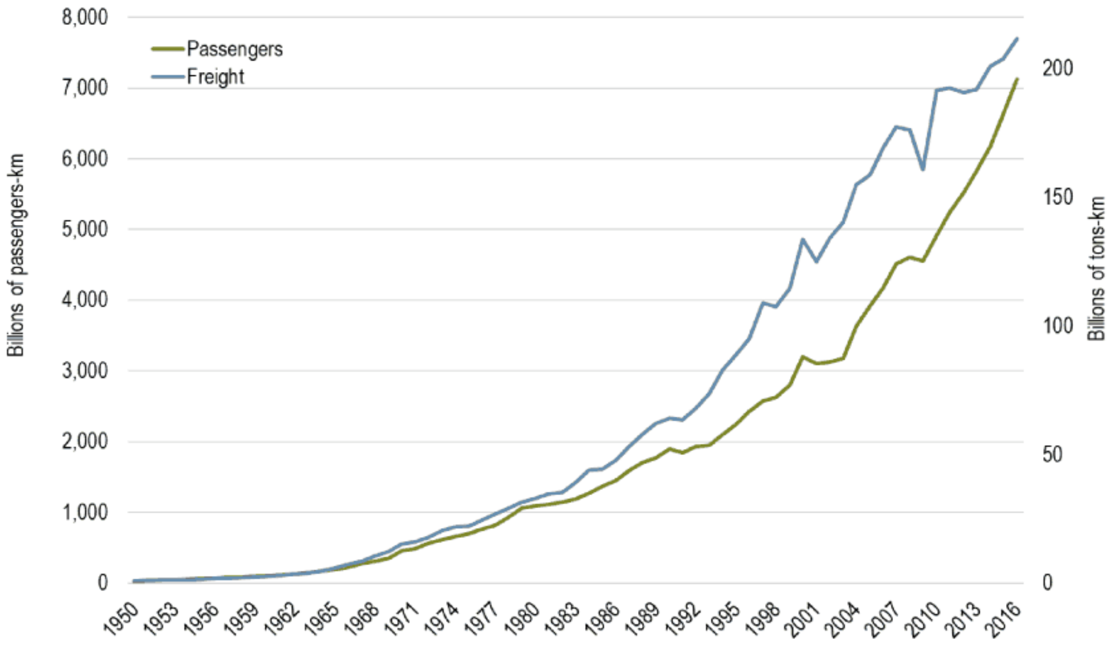
$$ARR\_DELAY = (-0.0154)\ VELOCITY\ KM/H + (-0.2512)\ DUMMY + (-0.0015)\ DISTANCE + (0.0225)\ DAY\_OF\_MONTH + (0.9959)\ DEP\_DELAY + \mu$$

Relatively to the coefficients, since the group aimed to construct a simple regression model, but effective, we conducted a Level-Level model, which means we did not implement any natural log in our variables. Firstly, "Velocity km/h has a negative coefficient, therefore if we increase Velocity hm/h by one unit, arrival delays will decrease 0.0154 minutes, on average, *ceteris paribus*. This variable, also, has statistical significance because its *p-value* is 0. Secondly, "DISTANCE", also has a negative coefficient and a small *p-value*, meaning, not only that it's a relevant variable, but also that if we increase it by one unit, arrival delays will decrease by 0.0015 minutes, on average, *ceteris paribus*. Thirdly, "DAY_OF_MONTH" has a positive coefficient, showing that if we move forward one more day, the arrival delays are expected to increase by 0.0225, on average, *ceteris paribus*. Again, this variable reflects a small *p-value*, therefore we can conclude that this variable is relevant to explanation of the dependent variable. Next, we have DEP_DELAY which has a huge coefficient, due to its relationship with the independent variable, therefore an increase in the departure delays by 1 minute, is expected to increase the arrival delays by 0.9959 minutes, on average, *ceteris paribus*. This variable also has a *p-value* close to 0, therefore it has statistical significance. Finally, "Dummy" has a *p-value* of 0.789 meaning that is higher than the standard unit of 0.05 (5%), leading the group to conclude that we cannot conclude, for sure, whether this, variable has or not statistical significance, nevertheless since the probability associated with the p-value is definitely, extremely high, we can conclude that a ratio higher than 50%, most likely does not bring much explanatory value to predict the number of minutes concerning the arrival delays. Regarding the confidence interval for the dummy variable, considering a significance level of 5%, contains the value zero, which means that there is statistical evidence that this variable is not relevant for the interpretation of the dependent variable. For the other variables there is statistical support that they are relevant to explain the behaviour of the dependent variable.

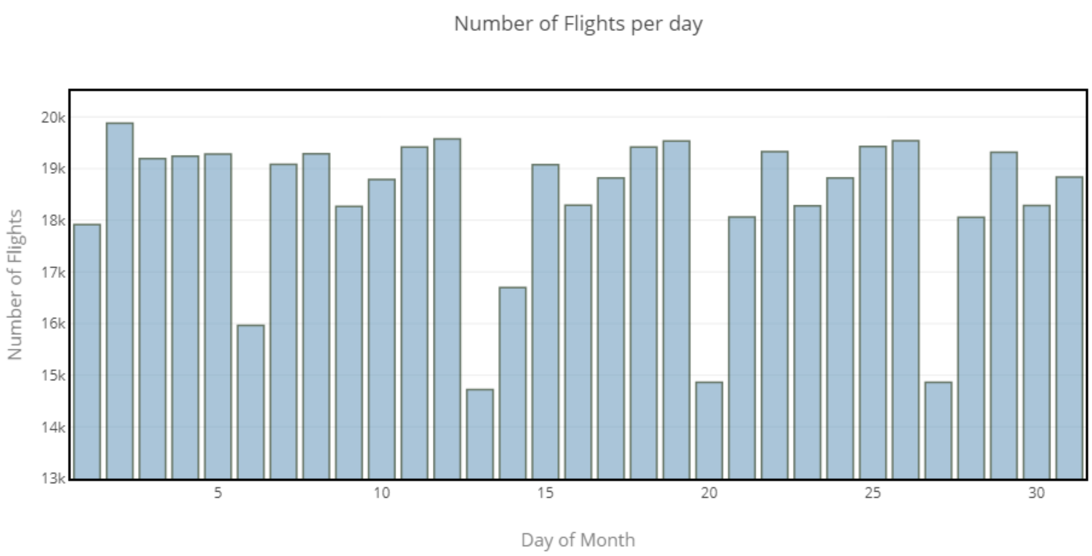
Despite our unsuccessful attempt of creating a meaningful variable ("Dummy"), our model was able to show some significant variables which allow us to verify why this issue of delays is happening.

- **Annex**

Annex. 1 : World Air Travel and World Air Freight Carried, 1950-2016



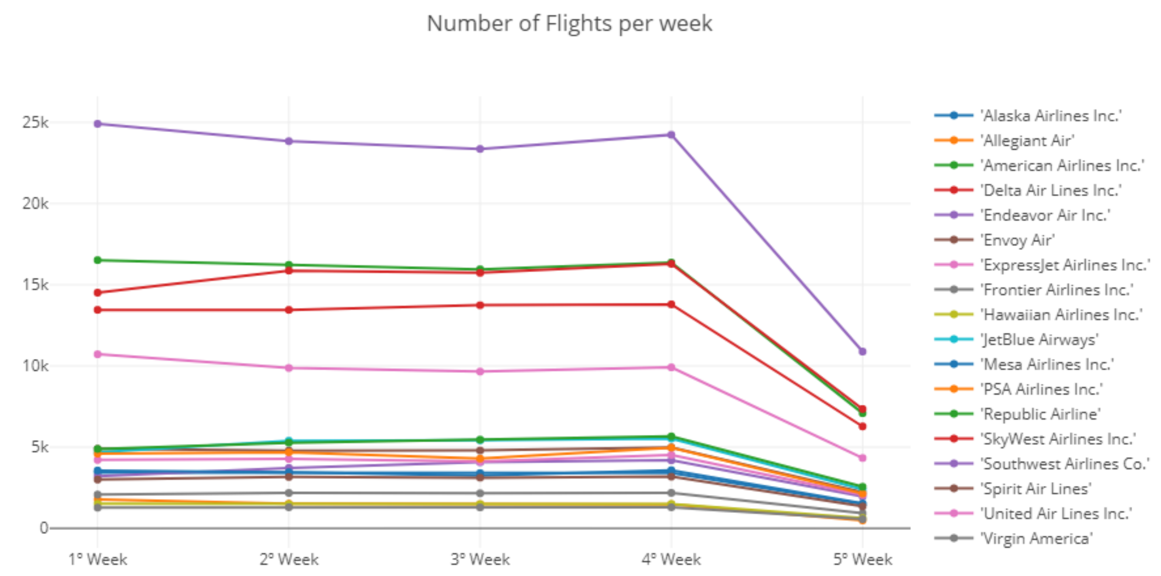
Annex. 2 : Number of Flights per day in January 2018



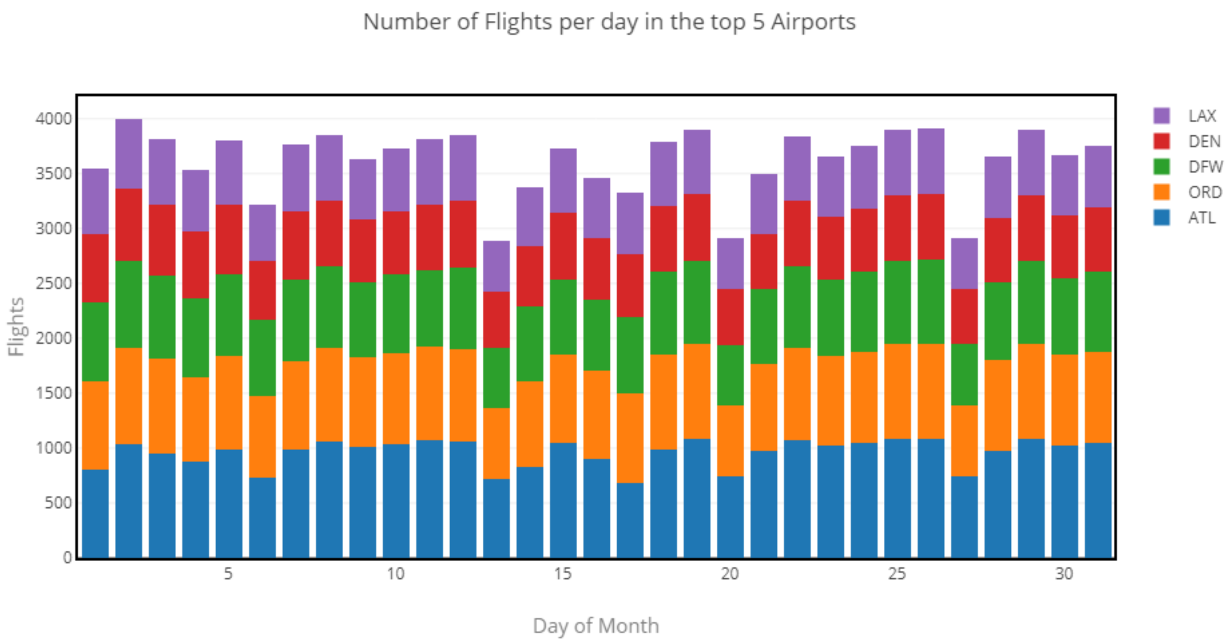
Annex. 3 : Number of Flights per Airline.

Description	Number of Flights
'Southwest Airlines Co.'	107240
'American Airlines Inc.'	72157
'Delta Air Lines Inc.'	69754
'SkyWest Airlines Inc.'	60714
'United Air Lines Inc.'	44494
'Republic Airline'	23845
'JetBlue Airways'	23420
'Envoy Air'	21670
'PSA Airlines Inc.'	20687
'ExpressJet Airlines Inc.'	19175
'Endeavor Air Inc.'	17136
'Mesa Airlines Inc.'	15381
'Alaska Airlines Inc.'	15179
'Spirit Air Lines'	13816
'Frontier Airlines Inc.'	9527
'Allegiant Air'	6761
'Hawaiian Airlines Inc.'	6617
'Virgin America'	5722

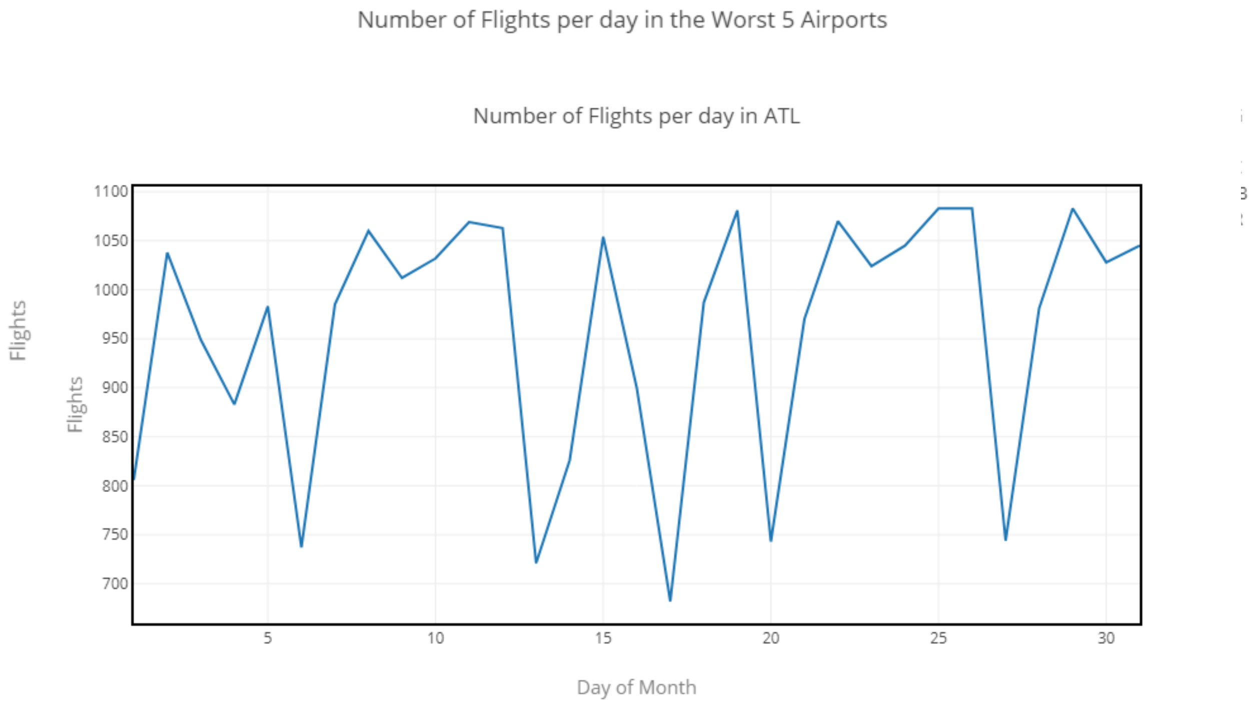
Annex. 4 : Number of Flights per week for all Airlines.



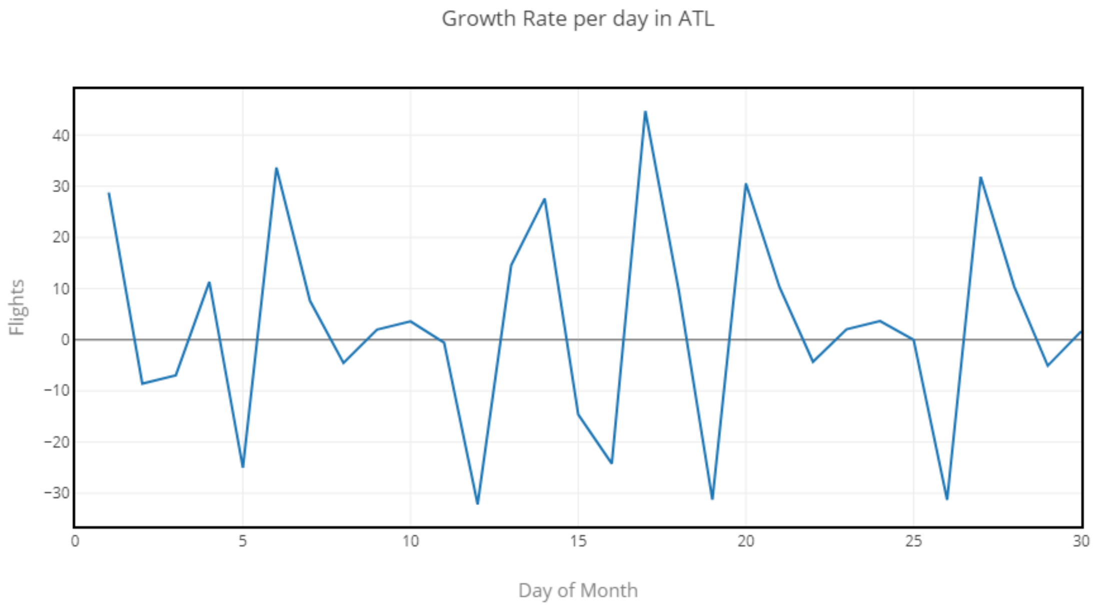
Annex. 5 : Number of flights per day in the top 5 largest Airports



Annex. 6 : Number of flights per day in the top 5 smallest Airports.



Annex. 7 : Number of flights per day per airline and it's growth rate.

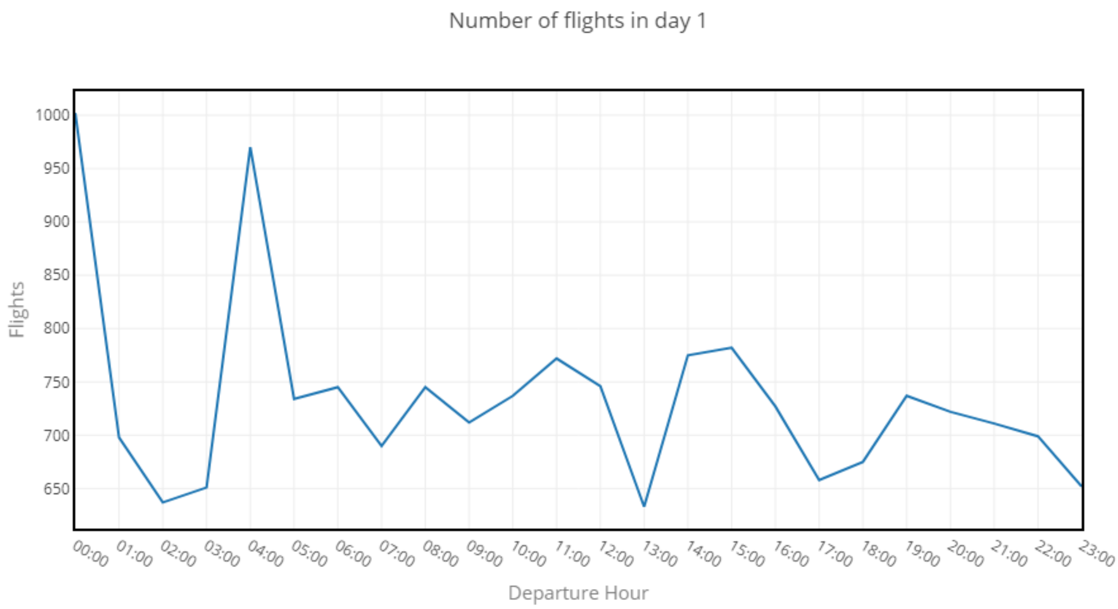


Annex. 8 : Possible destinations and Origins that go to "LAX".

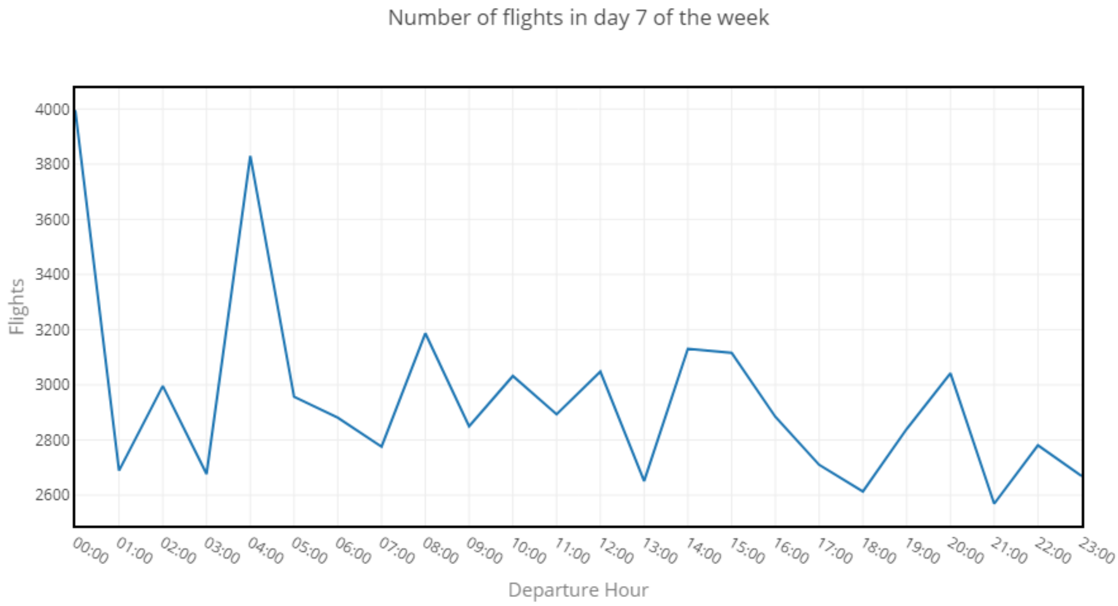
ORIGIN	DEST	Number of Flights
LAX	AUS	7
LAX	BOS	30
LAX	BWI	6
LAX	CLE	37
LAX	DEN	211
LAX	DFW	7
LAX	EWR	295
LAX	HNL	114
LAX	IAD	167
LAX	IAH	294
LAX	ITO	31
LAX	KOA	41
LAX	LAS	60
LAX	LIH	41
LAX	MCO	38
LAX	OGG	69
LAX	ORD	289
LAX	PHX	28
LAX	SAN	22
LAX	SAT	3
LAX	SFO	319

ORIGIN	DEST	Number of Flights
AUS	LAX	5
BOS	LAX	30
BWI	LAX	7
CLE	LAX	37
DEN	LAX	210
DFW	LAX	8
EWR	LAX	296
HNL	LAX	114
IAD	LAX	167
IAH	LAX	295
ITO	LAX	31
KOA	LAX	41
LAS	LAX	62
LIH	LAX	41
MCO	LAX	38
OGG	LAX	69
ORD	LAX	290
PHX	LAX	28
SAN	LAX	22
SAT	LAX	3
SFO	LAX	318

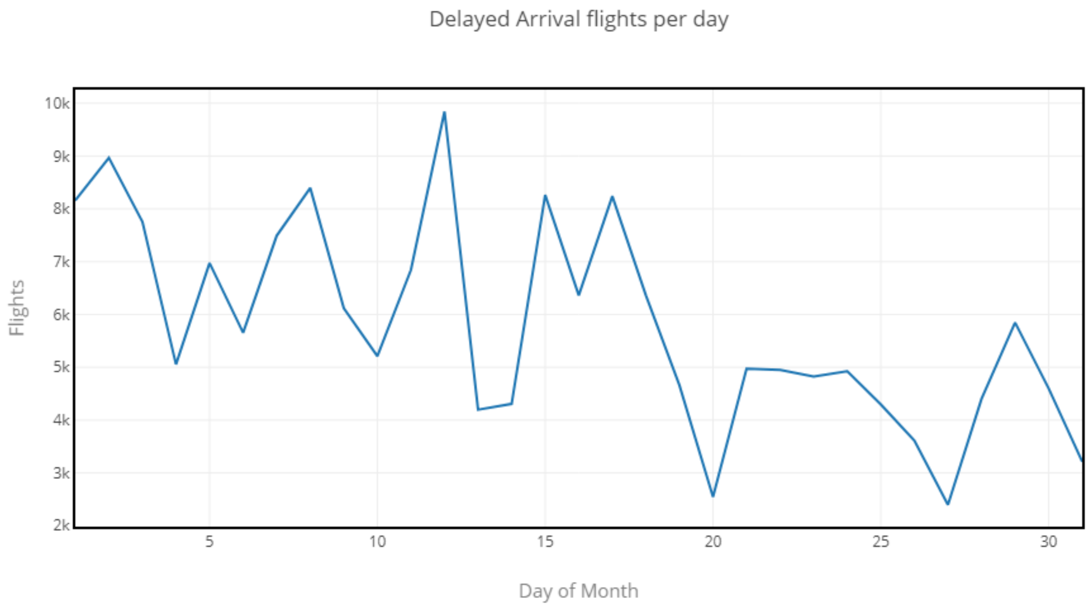
Annex. 9 : Number of flights in day 1.



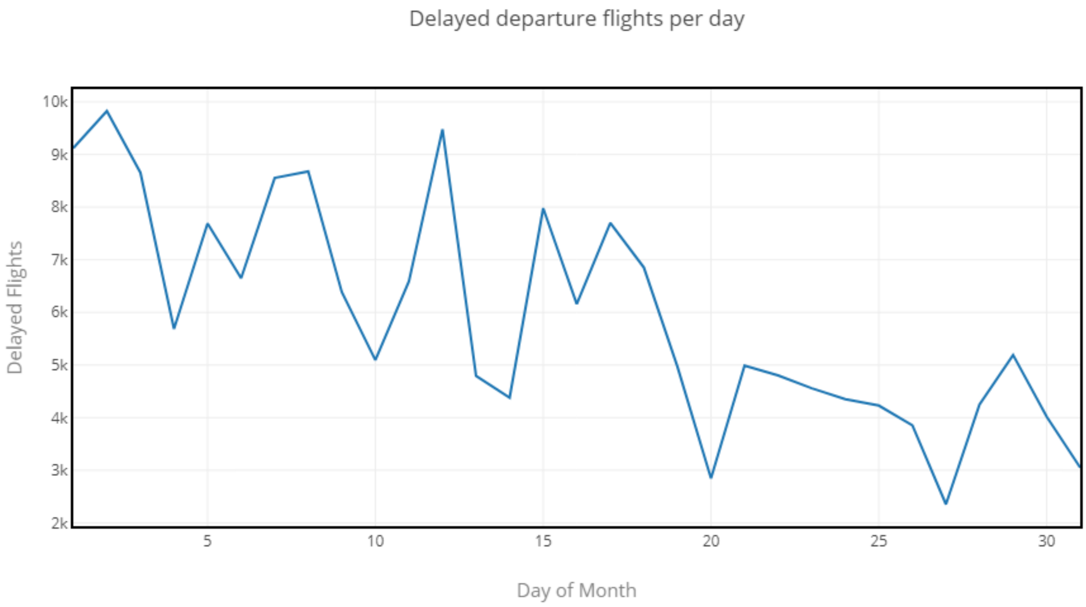
Annex. 10: Number of flights in day 7 of the week.



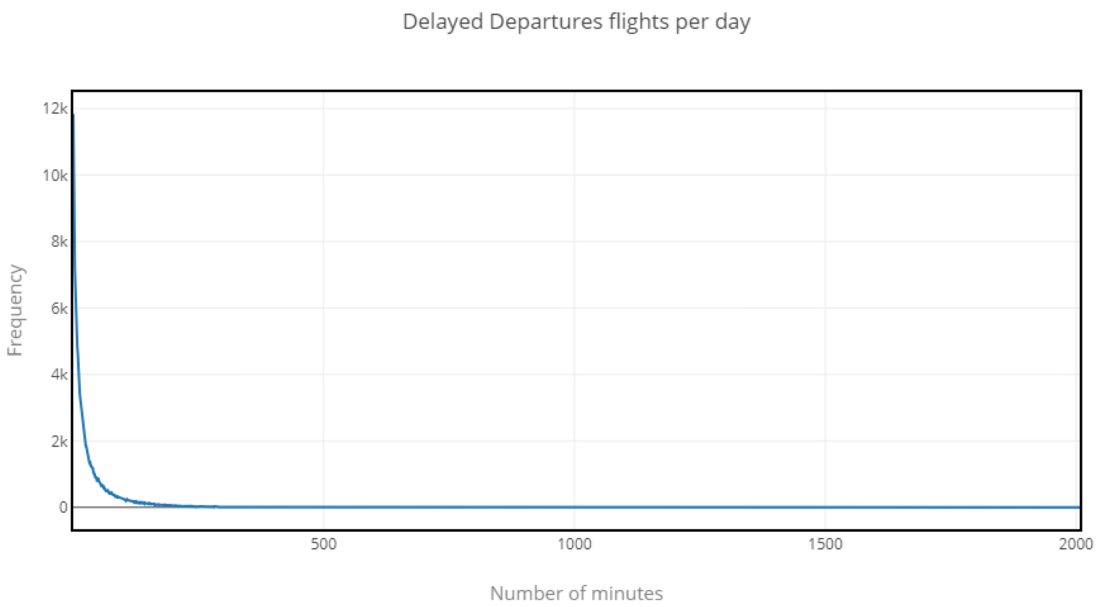
Annex. 11: Number of delayed arrivals per day.



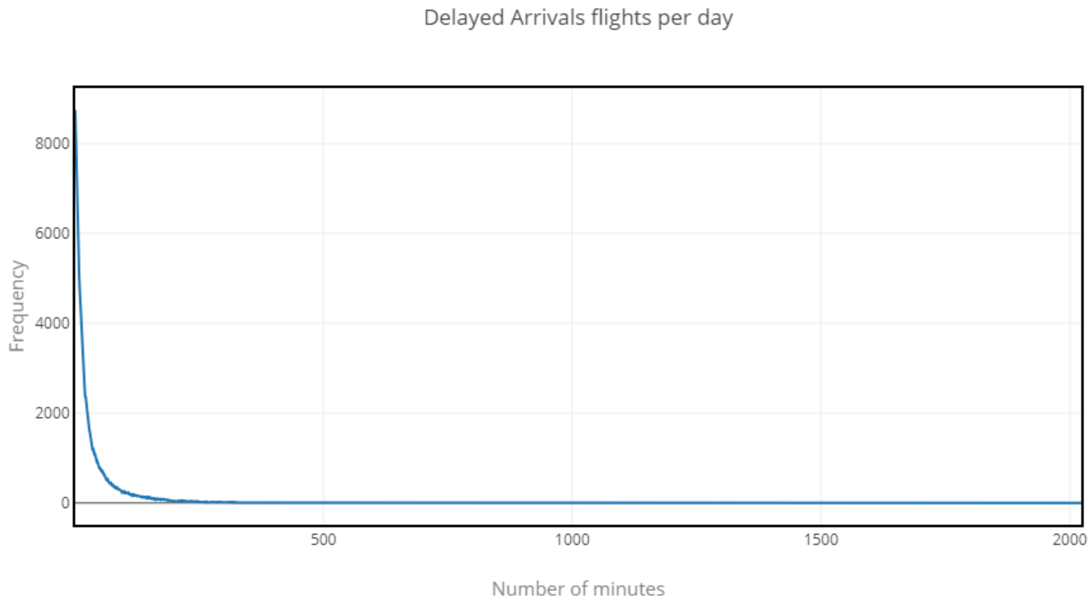
Annex. 12: Number of delayed departures per day.



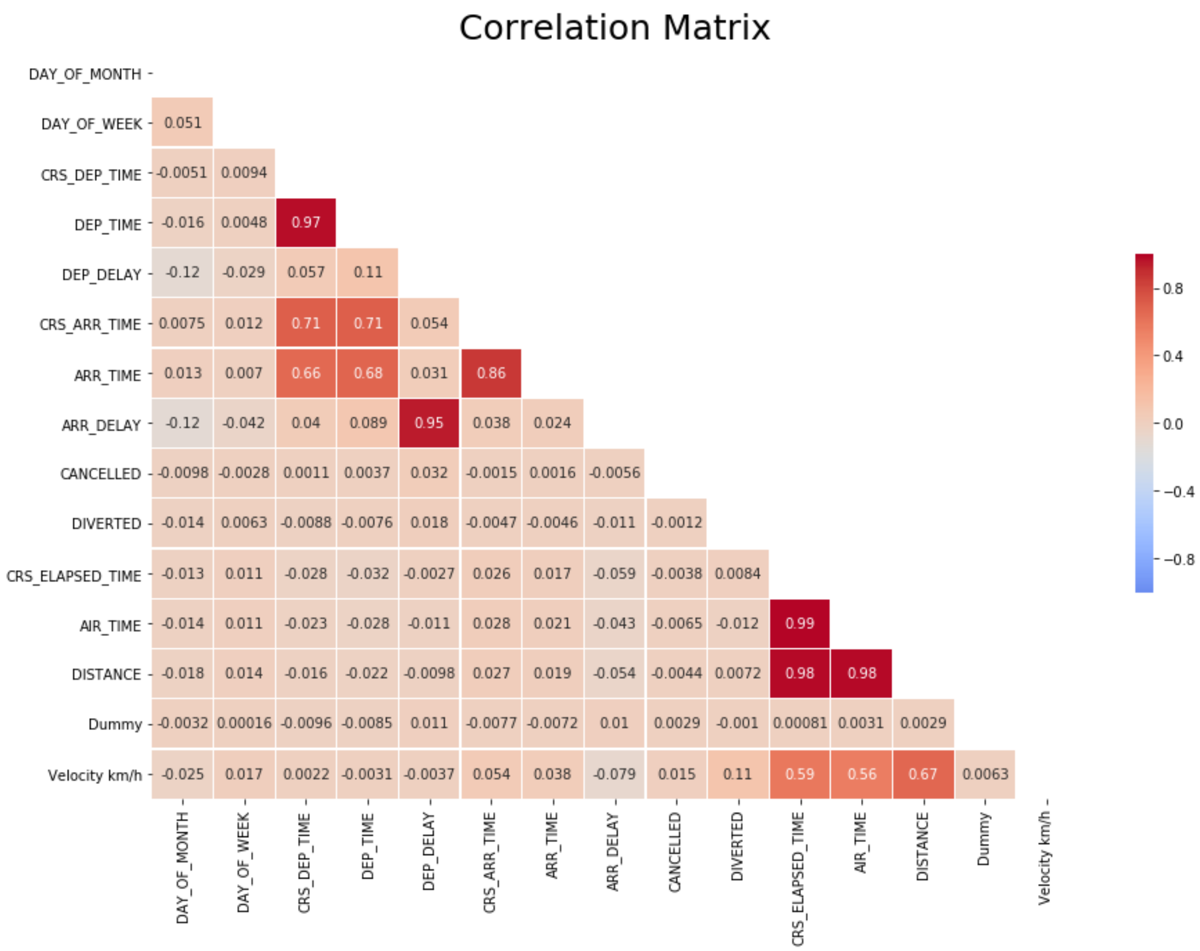
Annex. 13: Number of departures per time delayed.



Annex. 14: Number of arrivals per time delayed.


Delayed Arrivals flights per day
8000
6000
4000
2000
0
Frequency
500
1000
1500
2000
Number of minutes

Annex. 15: Correlation Matrix.



Annex. 16: OLS Regression Results.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              ARR_DELAY   R-squared:                       0.907
Model:                            OLS   Adj. R-squared:                  0.907
Method:                 Least Squares   F-statistic:                 1.109e+06
Date:                Tue, 20 Nov 2018   Prob (F-statistic):               0.00
Time:                        15:55:51   Log-Likelihood:            -2.3372e+06
No. Observations:              570118   AIC:                         4.674e+06
Df Residuals:                  570113   BIC:                         4.674e+06
Df Model:                           5
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Velocity km/h   -0.0146      0.000   -109.978      0.000      -0.015      -0.014
Dummy           -0.2129      0.907     -0.235      0.814      -1.990       1.564
DISTANCE        -0.0016   3.88e-05    -42.247      0.000      -0.002      -0.002
DAY_OF_MONTH     0.0140      0.002      6.723      0.000       0.010       0.018
DEP_DELAY        0.9953      0.000   2340.825      0.000       0.994       0.996
==============================================================================
Omnibus:                   801793.921   Durbin-Watson:                   1.758
Prob(Omnibus):                  0.000   Jarque-Bera (JB):      11243380187.355
Skew:                          -6.955   Prob(JB):                         0.00
Kurtosis:                     690.833   Cond. No.                     4.94e+04
==============================================================================
```