University of Connecticut

OPIM 5770 – Advanced Business Analytics and Project Management

Project: Atlas Air Worldwide Crew Member Schedule Optimization

Professor Jennifer Eigo

By
TEAM – 7

Akshay Kumar Pydi

Oluwademilade Ajayi

Remaliah Moses Gedala

Sagar Thorat

Suyash Singh

Yuting Hou

Contents

1. Executive Summary

Atlas Air is a cargo airline, passenger charter airline and lessor based in the United States and operating in five continents. The company has a total of 113 fleet, with 105 operating and offers "dry lease" services to ACMI (Aircraft, Crew, Maintenance, and Insurance) and CMI (Crew, Maintenance, and Insurance). This project seeks to address the high rate of turnover and the reason for the continued rise in attrition in the company and to help Atlas Air make changes where necessary. Extensive analysis was done on the rate of attrition by age group, fleet number and position, and the reasons for this attrition.

During the analysis, it was discovered that majority of the pilots had issues with the abrupt changes in the schedules, 52% of employees left Atlas Air within a year of joining, and most of them were dissatisfied with the management. With the combination of the analysis and information from external sources, possible recommendations and strategies were provided.

Data modeling was done, and factors such as age, fleet, position, change in trip count, length of trip, etc. were considered for the modeling. Seven models were compared and the best one was chosen on the bases of its accuracy and recall numbers as well as its interpretability. From this modeling, some of the key variables affecting attrition were identified as high international and domestic flights, more off-duty assignments in the actual schedules and high average delay hours. After reviewing the analysis, some of the recommendations were for Atlas Air to focus on pilot schedule and ensuring pilots are not blindsided when the actual schedules are released, to ensure that pilots are given adequate flying time, and to create an environment in which the pilots are assured that they have a voice in the workplace. Strategies were also given for the implementation of big data as a useful tool for tracking pilot performance and progress.

2. Problem Statement

Atlas Air Worldwide (AAWW) is one of the largest airlines in the United States. With four divisions - Atlas & Southern, Polar & Titan - it has a strong and authentic fleet. AAWW invests in a variety of projects aiming to promote business growth. The company oversees supply chain management and facilitates customer collaboration. Advances in the retail industry and detrimental effects on e-commerce and military forces have been perpetrated by AAWW. Global air freight yields are dependent upon it, as it accounts for 35% of all world trade. Among supply chain management methods, air freight-based supply chain management is considered a good value. In addition to investing in market giants like Amazon, FedEx, and HP, AAWW is committed to continuous expansion and growth.

Atlas Air has recorded an increased attrition rate in the last 18 months and the exit interviews show that the major reasons involve the scheduling of flights for the crew members. The percentage of crew members who left Atlas Air increased from 5% in Q4 of 2020 to 11% in Q1 of 2021. This percentage has also significantly increased to 24% in Q2 of 2021, a huge difference of 13% from what it was in the previous quarter. Increased attrition has been a major challenge to the company as additional cost is incurred because of employee replacement and the pilot training process. Atlas Air is unsure of what adjustments and improvements can be made to ensure crew member retention.

The business case that the Team 7 has come up with is –

- The crew member satisfaction will improve with an optimized schedule
- An attrition reduction of approximately $100k annually per crew member will result in cost savings

3. Methodology

Since the current contention at Atlas is the high attrition, the approach taken to tackle the problem was to explore the data and look for interesting insights regarding attrition and then to build an Employee Attrition Model using machine learning techniques.

3.1. Data Explorations Around Attrition

Several questions arose regarding the perspectives around attrition and a variety of data was also available from pilot's schedule, their payment, their feedback to the aircraft types, delays experienced on flights, etc. so the explorations were divided into 4 categories – **Pilot, Fleet, Schedule, Payment, and Feedback** based.

3.1.1. Pilot Based

Pilots moving to other companies

From the attrition data provided by Atlas, crew members that resigned from Atlas have either undergoing training or have moved to companies like United Airlines, UPS, FedEx, etc., (see Appendix – Pilot Attrition Other Companies)

Age factor

Major Attrition was observed for pilots between 30-55 age groups. (see Appendix – Pilot Attrition Age Group)

Experience - How long did the attritted pilots stay with Atlas?

Pilots experience in the company was calculated based on Active and Attrition files shared by Atlas. It was observed for pilots from 25-55 age groups that left, the average stay with Atlas was only 1.5 to 2.5 years. Also, if fleets were considered, it was observed that 777 fleet pilots of the same age group on average stayed with the company for less than a year and fleet 737 and 747 pilots stay with the company for 1-2 years while 767 fleet pilots stay on average for more than 2.5 years.

Base Attrition - From Which Base, Atlas experienced the maximum attrition?

The pilot information from different base locations was aggregated to check at which Base the attrition is maximum from. From the data, it was observed that Anchorage, Cincinnati were base locations from which maximum First Officers left. Also, maximum Captains left from Ontario and Cincinnati bases.

3.1.2. Fleet Based

Using the attrition data, for captains in Atlas, major attrition is happening for fleet 767 whereas for first officers it is fleet 747 (see Appendix – Attrition Pilots by Fleet)

3.1.3. Schedule Based

To check for insights on Schedule impact, two approaches were taken. One where the data was aggregated based on month, id, position, base and another where data was aggregated based on id, position, base, Trip No, and Month

Data Preparation

The schedule data, both published and actual, was aggregated based on pilot id, base, position, month as Schedule bidding happens monthly and pilots bid within their base. The aggregated data had total Off-duty days, total reserve days, total training days, total training trips obtained by joining Trip Nos in the schedule with the trips provided in the data dictionary and using the Category column to identify training, off duty and reserve categories.

The data also represented total R2 hours based on ‘R2’ flight number to consider the reserved hours in a month, total Deadhead hours which pilot spent in the back of a flight in a month, total flying block hours which are the hours the pilot was actually flying or in control of the plane, total deadhead trips they had in a month, total flying trips they had in a month and of those flying trips how many trips were converted to actual trips.

Then another dataset was prepared, and both the published and actual schedules were compared to look out for insights between published and actuals schedules and see if schedule had an impact. The following metrics were calculated for each pilot per month:

$$\% \, increase \, over \, Published \, Trips = \left(\frac{Actual \, Flying \, Trips \, - \, Published \, Flying \, Trips}{Published \, Flying \, Trips}\right) * \, 100$$

$$\% \, Transfer \, from \, Published \, to \, Actual = \left(\frac{\text{From Published Trips how many converted to Actual Trips}}{Published \, Flying \, Trips}\right) * \, 100$$

Observations

The aggregated data was compared for Active and Attritted employees against several measures like Average Flying Block hours, Off duty Days, etc.

It was observed that the pilots who left had lesser average flying block hours compared to active pilots each month. Also, there was a lot of fluctuations in their schedules. Pilots want to fly more and get paid for flying. Using the aggregated data, it will be easier to check where the pilot spent most of his time and whether the pilot compensated for this lack in other flying trips like more deadhead trips or more training duties. (See Appendix for Flying Block hours comparison)

It was also observed that the Actual Off Duty days is much higher for the pilots who left. Using the dashboards and drill down views the kind of Off Duty DAYS these pilots had can be investigated. Whether they were voluntarily taken by the pilots like VAC, MED or mandatorily imposed on them like Duty Rigs because of the scheduling changes needs to be explored (See Appendix for Off duty days comparison). Plotting the actual number of flying trips, the pilot went on and compares the same with the published flying trips, it was evident that the active pilots have relatively higher number of actual trips i.e. almost average 33% increase over published trips compared to inactive pilots who only had 19% increase and also the conversion of published to actual trips was more than 50% for the active pilots. So, there could be some dissatisfaction around not flying more and not flying the trips the pilots signed up for. (See Appendix for Published to Actual percent increase and Published to Actual Trip Conversion Efficiency)

Data Challenge

Since the data was monthly aggregated and monthly compared at a Base, there was a challenge as some crew didn't fly in those months and there were gaps in analysis. Moreover, crew exit is spread over the years. So, the comparisons between Active and Attrited Pilots must consider this discrepancy. Hence, another dataset was prepared where schedule data was aggregated based on Trip No, crew id but because Trip Numbers could be reused, Month-Year was also added to group the data to distinguish one trip from the other as there are instances a pilot has been assigned a trip more than once in same month or assigned the same trip no in different months. Other factors around trips were also calculated like whether the trip is international or domestic, total deadhead hours, total r2 hours , total flying block hours, total deadhead legs on the trip, whether a deadhead was used to travel to the destination where the trip actually starts, whether a deadhead was used to return to the base from the arrival where the trip ends, total flying legs, total delay hours encountered on the trip, the length of the trip and total layovers on the trip. (see Appendix Published and Actual Aggregated Data by Trip No)

Published vs Actual Schedule Analysis

For most fleets, it is observed that pilots are reserved with R2 hours in the published schedule and the reserved hours are utilized in the actual schedule during schedule conflicts like delays, or customer route change or time change requests. Fleets 747 and 777 pilots see maximum use of R2 hours while fleet 767 pilots R2 hours are still more than their flying block hours in the actual schedule.

After plotting a boxplot of Actual Flying block hours of pilots, it was observed that the average and median do not overlap and more than 50% of the pilots are flying less than average. Furthermore, there are outliers in the plot which indicate that there are trips where pilots are flying 1.5 times more than 50% of the trips above the third quartile (75% of the trip hours). So, it was deduced that there's an imbalance in flight assignment in the actual schedule.

Trip Length Analysis

From the data it was observed that Active pilots were involved in trips of different lengths whereas the attritted pilots were majorly used for trips of 5 days or less.

From a pilot survey study conducted in 2019 (iap2750.org), it was observed that most pilots prefer 6 to 10 day or 11 to 14 days trips, however, it was observed that most trips were of 5 days or less for all pilots and so for the pilots what they are getting from their schedule does not match with their expectations

Trip Layover Analysis

From the data it was observed that pilots who went on 5 days or Less trips on average had less than a day of layover. (See Appendix for Layover count)

International or Domestic Trip Analysis

From the data it was observed that all fleets are used for international and domestic flights. Fleets 747, 767 ad 777 are not used frequently for domestic trips longer than 5 days. Fleet 767 is used more frequently than fleet 777 for international trips and 747 is majorly used for international trips. It was also observed that Attrited pilots were utilized more for 5 days or less International trips relative to the Active pilots. (See Appendix for Domestic or International Trip Analysis)

3.1.4. Payment Based

The team noticed that pilots received pay credits based on how much time they spent on workday hours. Especially, pay efficiency is a vital indicator to indicate pilot working efficiency. Higher efficiency is preferred. Because the compensation data (3 files) was given, the team can by adding all the working days hours and days off-hours at each pilot level divided by the actual block hours, the pay efficiency came out. The calculation shows below:

*Pay efficiency (%) =Actual Block hours/ (workday hours +day off hours) *100*

After getting the pay efficiency of each pilot, the team used pay efficiency results to match the active pilots' data specifically. Therefore, the fleet type of operation efficiency was observed by calculating the average pay efficiency per type. 737 and 767 are lowest pay efficiency, which was 32.89 % and 45.87%. The highest pay efficiency is from 747, which is 54.46%. (Pay efficiency)

Compensation Compared to Rivals

Based on pilot wages data across multiple airlines (*Airline Pilot Central, 2019*), it was observed that the compensation is not at par with rivals. Year 1 - 3 growth, during which most pilots leave Atlas, is not as steep as rivals. Also, pilots get capped wages after 12 years of service as opposed to 15 years for rivals like UPS and FedEx. The management perspective is they cannot compete with the rivals as their business models are not same. So, the pilot perspective does not match with management. However, recently the pay raise for the pilots was negotiated. A 5- year contract was recently arbitrated, and it shows a marked increase in wages for pilots across all the fleet. (see Appendix for Renewed Contract Pay Scale).

3.1.5. Pilot Feedback Based

Internal

The Engagement survey spreadsheet with 74 records in text data format was provided by the sponsor. It is evident that many of the pilots disagree or strongly disagree with the way work schedules are allocated by roughly looking through the spreadsheet. The purpose of this is to extract the most common words from sentences. Then, implement text mining, such as removing funky characters and stop words. The team got the most frequent words and then made them into a distribution plot and word cloud image. As a result, schedule, seniority, changes, bidding process, and work stood out. The payment was also mentioned here, but it wasn't in the top five words. Schedule changes bring challenges to pilots from this stage of text analysis. Pilots are saying it would make their life easy if the scheduling department of Atlas published flights schedule efficiently. Additionally, pilots are unsatisfied with the bidding process and seniority. (Internal word image)

External

After running the text analysis of the engagement survey, the team was interested in the comments towards their workplace. Therefore, some external sources, such as Glassdoor (Glassdoor) and Twitter (Twitter) brought more valuable insights to help the team raise an insightful recommendation. Additionally, the team saw the difference under different sources.

The first data source comes from Twitter account @AtlasAirPilot. The team had successfully downloaded all 2153 tweets by using Python to scrape all tweets. After removing all the funky characters and most common words, all the tweets were made lowercase. In addition, the team implemented stop words from nltk modules to reduce some unnecessary words. As a result, the team extracted the most frequent word from the tweets.

The next step was making it a distribution plot and word cloud image. It was noticed that most tweets are negative which means that pilots complain about what they have experienced or what they are experiencing. The most frequent words were payment, amazon, contract, and management. Interestingly, Amazon and DHL were displayed in the word cloud image. The team is guessing there are existing conflicts between clients and pilots. (Top 30 words)

The second data source is from the Glassdoor website. Text was manually collected from the official account @Atlas Air Worldwide focusing on the pilot position. There are 73 records for pros and cons which were collected from the website into an excel spreadsheet. After implementing text mining, the team got the distribution plot and word cloud image. By doing text analysis for pros, it was discovered that Atlas has some advantages they could be used for hiring new pilots. The most frequent words that come out are travel, international, flying, and gateway. Many pilots are saying that Atlas is a great place to travel the world and to engage with different fleet types (Pros of words).

However, after running the text analysis for cons, the most frequent words are payment, company management issues, training, and contract. Besides training, the other three words were also mentioned on Twitter. One valuable insight obtained from this is that pilots mentioned training sessions could be done better (Top 30 words).

3.2. Employee Attrition Model

3.2.1. Data Sampling

Having many files for analysis, the initial step taken by team was to sample the data from all the files like Delay Data, Published Schedules (4 files), Actual Schedules (4 files), Engagement Survey and Flight Data. It is important to consider various terms that are part of the Actual and Published schedules along with employee level data. But at the same time taking too many columns of same meaning from actual and published schedule will lead to many columns and hence leads for a need to dimension reduction. For this reason, the team has taken all the columns data as a difference of schedule between Actual and Published (Actual – Published) which indicates the change of the schedule from published to actual bid. This variation is the key measure for the data points that are part for the Actual and Published files.

While considering the delay, flight and attrition data, the data points are aggregated at the crew level. Once the complete aggregation of the data was present, then average was performed based on the number of months an employee worked in the organization there by the final columns showing the data at a monthly level and crew ID level. Below are the details of final sampled data:

1. Current_Status = Attrition (1) / Existing (0) - Target Variable
2. CHANGEINTRIPCOUNT = Difference in Number of Trips for an Employee
3. CHANGE_IN_HOURS = Change in the Number of Block hours worked by an Employee Excluding Dead Head Hours
4. CURRENT_AGE = Age of an Employee
5. FLEET = Fleet to which an Employee belongs
6. SEAT = CA/FO
7. CHANGEINOFFDUTY = Change in the off-duty hours

8. CHANGEINRESERVE = Change in the Reserve Duty Hours
9. CHANGEINTRAINING = Change in the Training hours
10. CHANGEINDHTIME = Change in the Dead Head Time
11. CHANGEINDHCOUNT = Change in the Dead Head Count
12. AVG_DELAY = Average delay of all the flights at the Employee level
13. AVG_NO_OF_DELAYS = Average number of delays faced by an employee during his total trips
14. AVG_FLIGHT_BLOCKHOURS = Average flight block hours
15. CONTRACTS_WORKED = Number of Distinct Contracts Worked by an Employee
16. LATENIGHT = Number of Late-night trips taken by an employee - Flights that started after 9 PM
17. Trip_Length = Average number of Trips per employees
18. Total_Layover = Average Layovers per employee
19. Int_Count = International Trips count for employees at an average per month
20. Dom_Count = Domestic Trip Count for employees at an average per month.

3.2.2. Data Exploration

This section deals with the exploration of all the sampled data in terms of their distribution for outlier detection, any patterns of the categorical or continuous variables with the attrition which is the target variable. As can be seen in Pie chart the ratio of existing to attrited employees is 88:12 which clearly shows that the target variable is biased.

While checking the correlation for all the continuous variables, it can be observed that there is a strong correlation between TRIPLENGTH, TOTAL_LAYOVERS and INT_COUNT. Also, there exists a strong correlation between CHANGEINDHTIME and CHANGEINDHCOUNT.

From the basic Exploration of the Fleet variable, it was observed that more employees attrited from Fleet 747 and 767. Also, more people from FO range are attrited compared to CA. All the Continuous variables are having normal distribution with some outliers therefore they can be used for modeling after transformation.

3.2.3. Modify

In this section any changes to the columns data and splitting of the dataset was investigated before Modeling.

Data Transformation

As per the Box plot analysis, transformation of variables like Trip count, Reserve, Off duty, Block Hours, Training, Dead head, Average delays, Average number of flight delays, late Night, Domestic Travel was needed as the data is having more outliers and is skewed. The after transformed distribution does not have any outliers.

Making Indicator Columns

Columns like Fleet and SEAT are having 4 and 2 distinct values respectively and hence they are transformed to indicator columns.

Target Variable Imbalance

The Ratio of 1 to 0 for Attrition Column is very biased. The ratio is 12:88 which ideally will give us biased results. A problem with imbalanced classification is that there are too few examples of the minority class for a model to effectively learn the decision boundary. The imbalance in the distribution of target categories can lead to a high accuracy just by predicting the majority class, but fail to capture the minority class, which is the class of interest and the point of creating the model in the first place. Thus, there is a need to balance the data before modeling. There are 2 methods followed - Under sampling or Over Sampling. In under-sampling, there will be removal of data from Majority class to match the minority class records which ideally means that not all the records are being considered for the analysis and model prediction. Hence, this method is of less scope all the records are required and then create attrition model.

Thus, the preferred technique is Oversampling. SMOTE (Synthetic Minority Oversampling Technique) randomly picks a point from the minority class and computes the k-nearest neighbors for this point. These points are added between the chosen point and its neighbors thereby creating synthetic data for the minority class.

Therefore, the dataset is first split onto 70:30 of Training:Test. Validation is avoided as the dataset is very small. Now, that the Training sample is available, SMOTE technique was used to do oversampling at 100%. 50% oversampling was also validated but that yielded very less accuracy results compared to 100% accuracy.

3.2.4. Modeling

Logit Regression was done to find out significant columns based on the p-values and then have got to remove columns like CHANGEINTRIPCOUNT_TRAN, CHANGEINRESERVE_TRAN, TRIPLENGTH which had p value greater than 0.05 and hence are insignificant as can be seen in results. These final columns are applied for various Machine Learning algorithms:

1) Logistic Regression: With this model an accuracy of 82.49 % and recall of 0.71 is achieved.
2) Decision Tree: Various iterations are done for number of splits, and it’s observed that Decision Tree with 5 splits has the maximum accuracy ad recall of 80% and 0.65 respectively.
3) K nearest neighbor: Error rate analysis was done for various values of K to find out the optimal value of K. It is observed that at K=2 the error is less, and the accuracy and recall are maximum. Please refer to Appendix.

4) Ensemble Models: Various Ensemble models like Ada Boost Classifier, Gradient Boosting Classifier, Random Forest Classifier and Extra Trees classifiers were used in the analysis. These have got a high accuracy, but the recall values are very less.

3.2.5. Assessment

In this final process of SEMMA, various parameters are assessed to decide on the Best Model that gives good prediction.

The three key factors that will be used to finalize the best model are:

1) **Accuracy**: It helps us to determine how well the True Positives and True Negatives are categorized in our model. As can be seen in the bar plot analysis, Ensemble models have high accuracies followed by KNN, Logistic and Decision tree models.
2) **Recall**: It is a very key factor for modeling part because as per our problem statement it is required to concentrate on the False Negatives. It's a loss to the company if a model predicts a crew id as not attrited where in actual scenario the employee is attrited. Hence, it is very much important to achieve a good recall value for an Attrition Model. It can be observed from bar chart that Logistic regression and Decision tree have high Recall values.
3) **Interpretability**: For an attrition model, it is very much important to know on which factors have led to attrition and any specific conditions around the significant columns. Logistic regression and Decision Tree Models are 2 models which gives high detailed information on the key variables leading to the Attrition of Employees compared to all other models as they are ensemble models and hard to Interpret because of the Ensemble in nature.

4. Results and Findings

Hence, based on the Accuracy, Recall and Interpretability – it can be observed that Logistic regression and Decision Tree model are the two which would be the possible best model for Attrition model. As Logistic regression as Higher accuracy and high recall value, hence the best model is Logistic regression model. Based on the Odds ratio in the Logistic Regression, below columns are contributing positively towards the Attrition of Employees. Odds ratios define the probability of attrition to probability of being active employee. So, an odds ratio of greater than 1 shows that with a unit increase in the variable value then there are more chances of the change leading to attrition.

- International Trips Counts
- Change in Off Duty
- Domestic Trips Count
- Average Flight Block Hours
- Average Delay
- Change in Dead Head Time

5. Conclusion and Recommendations

1. To reduce attrition and training cost, it is recommended that Atlas Air focuses on pilot schedule since a lot of the complaints were about how frequent and sudden the changes in schedule are.
2. While it is understandable that majority of these changes cannot be controlled, it might be helpful for Atlas Air to consider showing the probability of change per schedule when pilots bid their flights, so they have an idea of what to expect when the actual schedule is released.

3. In the event of schedule conflicts, flights can then be assigned to pilots who have below average flying block hours. One of the reasons for attrition was the fact that pilots were not flying enough so by assigning conflicting flights to pilots with lower flying block hours, Atlas Air should be able to achieve an even distribution of flights. From the modelling, the deviation for deadhead hours and off-duty count for a particular trip is high and so a decrease in this will in turn lead to a reduction in attrition rate.
4. Another concern was issues with management and the fact that pilots feel like they do not have a voice in the workplace. Working in an environment that seems to dismiss or invalidate their concerns can be very discouraging for the crew members so it is imperative for Atlas Air to create an environment that encourages pilots to voice their issues or dissatisfaction with the assurance that they would be heard and taken seriously. This will also increase their participation in engagement surveys.
5. The pilots and management seem to still have conflicting opinions on the new contract. A 747 Captain with 12 years of experience is expected to get 295$ which the pilots feel is still lesser than the rivals and not close to the Union's demand. Atlas Air also needs to make sure that growth of Year 1 to 11 experienced pilots is also aligning with growth of the rivals, especially in the early years, as maximum attrition is experienced in the early years. It seems that the pilots are not exactly happy with the proposed 5-year contract so their feedback can be obtained, and Atlas Air can consider making changes if feasible. (Eric Kulisch) These changes could be around differentiating with the rivals in terms of retirement packages, benefits and aligning with pilots' career expectations (Damian Brett).
6. Additionally, it was noticed that 52% of employees left Atlas Air within a year of joining and that rate of turnover is extremely high. This percentage can be reduced by offering additional

incentives to new employees. The compensation package can be reviewed to match those of the competitors and new employees can be given the opportunity to choose their schedule in every alternative bidding. Other useful strategies can be implemented to ensure that employees see Atlas Air as a great place to work and the attrition rate, especially for new employees, reduce. The training schedule can also be reviewed to ensure that training is still efficient. Surveys can be conducted with operation leaders on training efficacy to see if it tallies with training expenses and to see if adjustments need to be made.

7. Another recommendation would be for Atlas Air to build a hiring pipeline to ensure a steady workforce and manage abrupt attrition. (Oliver Wyman) Atlas Air can consider establishing good relationships with existing training schools so that they can directly hire some of the best graduating students from these schools. Long-term, the company can also look into the possibility of investing in a flight academy to build its own pilot pool.
8. The positive effects of using big data cannot be over-emphasized. By tracking the pilots in the Atlas network, the company can increase responsiveness during bad weather conditions and commercial delays as well as monitor pilot behavior in actual cockpit and simulator training modules. Another strategy would be to implement Scheduling Metrics Dashboards to monitor pilot schedules, total flying hours, total delays on a trip and conversion from published to actual trip efficiency. Lastly, a Fatigue Analytics Dashboard can be implemented to keep track of the performance of pilots, the number of trips they take in a month, the amount of rest they get and the type of trip durations they get.

6. References

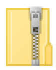
1. Iap2750.org: https://iap2750.org/wp-content/uploads/2021/09/Scheduling-Survey-2750-Q3-2021-Crew-.pdf
2. *Airline Pilot Central, 2019 - https://www.airlinepilotcentral.com/airlines/cargo/atlas_air*
3. Glassdoor:https://www.glassdoor.com/Overview/Working-at-Atlas-Air-Worldwide-EI_IE4235.11,30.htm
4. Twitter: https://twitter.com/AtlasAirPilots
5. Target Variable Imbalance: https://www.analyticsvidhya.com/blog/2020/07/10-techniques-to-deal-with-class-imbalance-in-machine-learning/
6. Oliver Wyman| Marsh and McLennan Companies | THE PILOT OF THE FUTURE - HOW THE AIRLINE INDUSTRY MUST ADAPT https://www.oliverwyman.com/content/dam/oliver-wyman/v2/publications/2018/november/2018_Flight_Ops_Survey_The_Pilot_of_the_Future_web.pdf
7. Damian Brett | Atlas Air pilots hit out at new five-year contract https://www.aircargonews.net/airlines/atlas-air-pilots-hit-out-at-new-five-year-contract/
8. Eric Kulisch, Air Cargo Editor |Disgruntled pilots blast 5-year contract with Atlas Air - FreightWaves

7. Appendix

QUERY	DATA GENERATED
Data Aggregation in published The query used in PostgreSQL to aggregate the data in published schedule – PublishedAggregated ReportExtract.sql	Published Schedule Aggregated Data.zip Aggregated Published Schedule Data
Data Aggregation in Actual Schedule The query used in PostgreSQL to aggregate the data in actual schedule – ActualAggregatedRe portExtract.sql	Actual Schedule Aggregated Data.zip Aggregated Actual Schedule Data
Published vs Actual Schedule The query used in PostgreSQL to compare actual and published schedule metrics PublishedvsActualMet rics.sql	PUBLISHEDVSACTUAL .zip Published vs Actual Data
Data Aggregation in Published Schedule by Trip No PublishedScheduleAg gregatedbyTripNo.sq	AGGREGATED_PUBLIS HED_PILOTSDATA_TRI Aggregated Published Schedule Data based on Trip No
Data Aggregation in Actual Schedule by Trip No ActualScheduleAggre gatedbyTripNo.sql	AGGREGATED_ACTUA L_PILOTSDATA__TRIPN Aggregated Actual Schedule Data based on Trip No

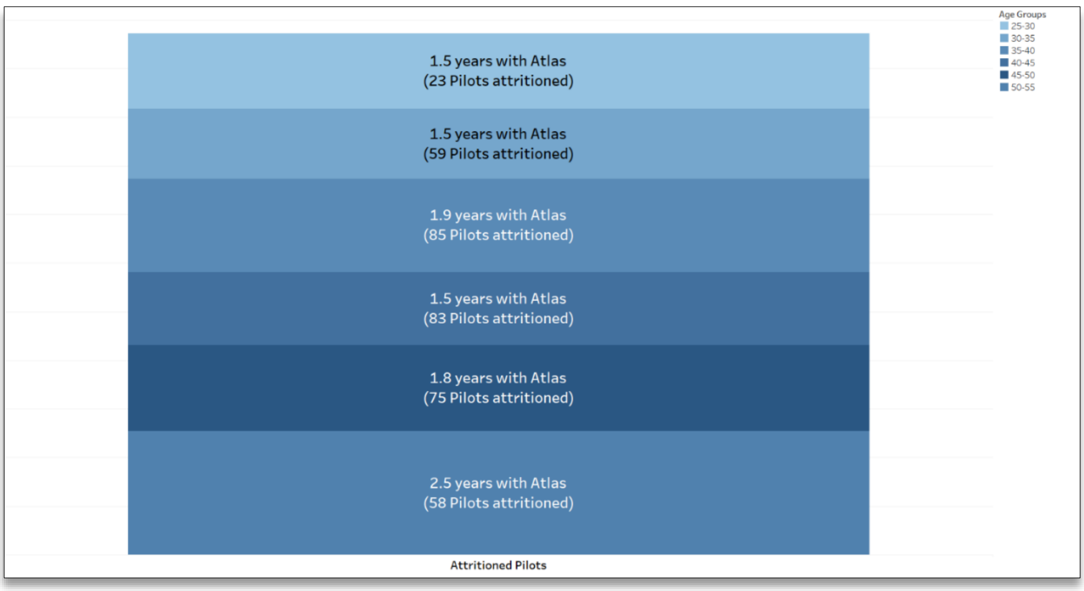
Pilot Experience at Atlas

Data –

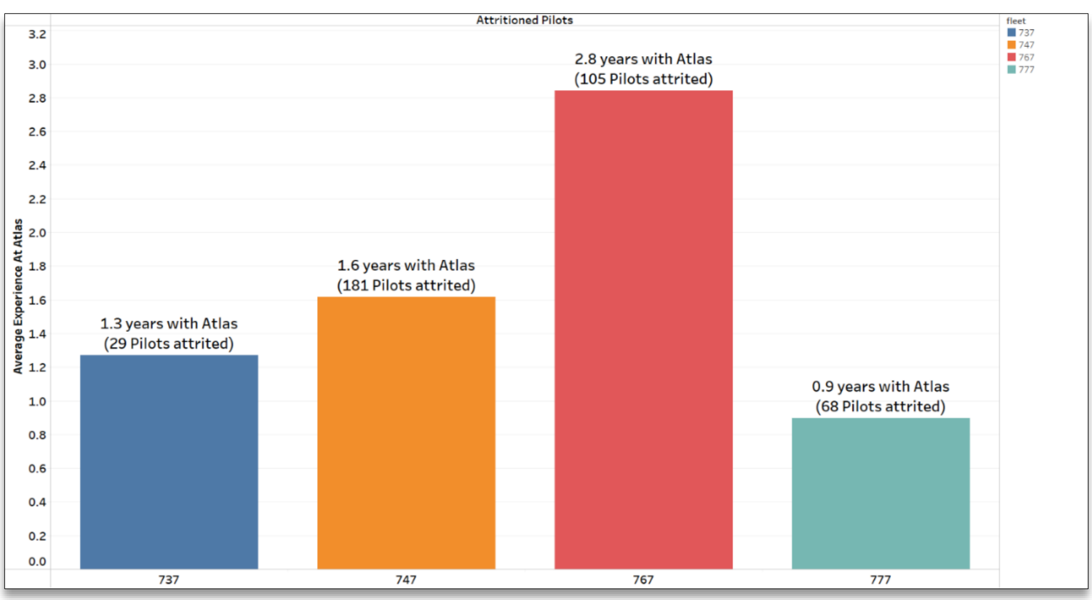


Pilots Age and Experience Data.zip

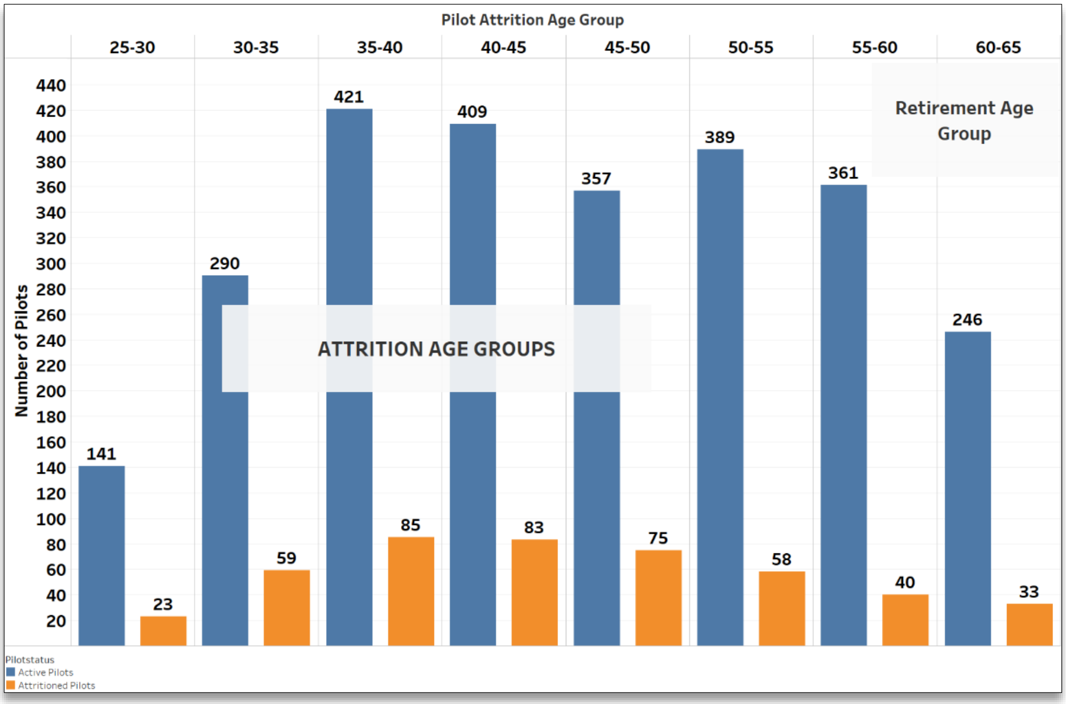
Visualization -



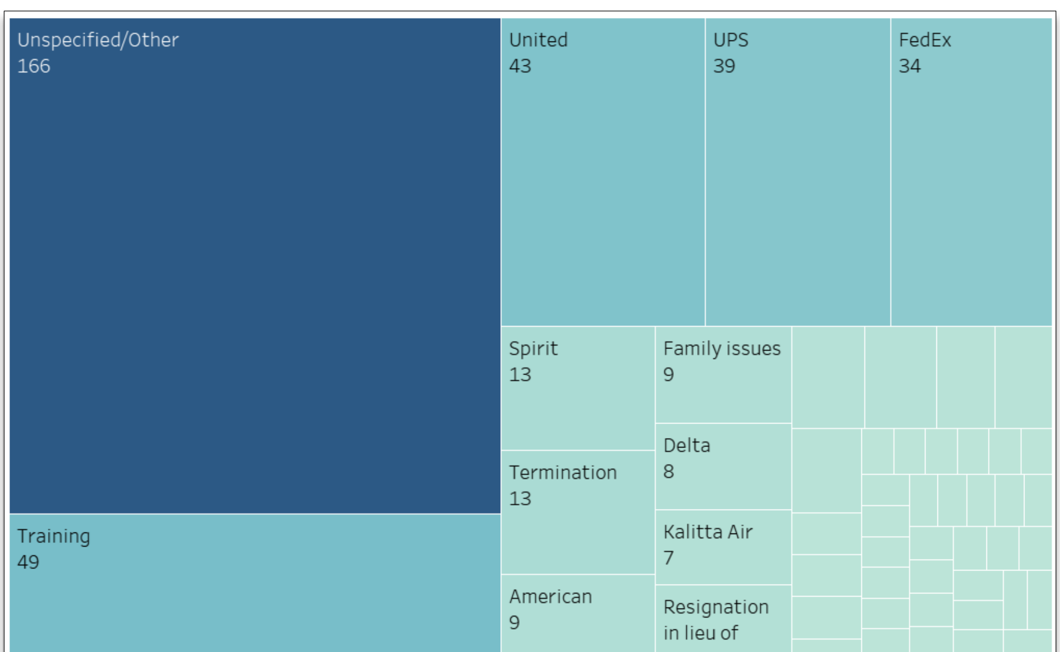
Attrition Pilots Experience at Atlas by Fleet



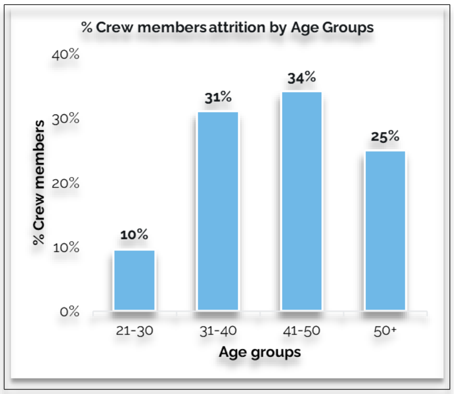
Pilot Attrition Age Group

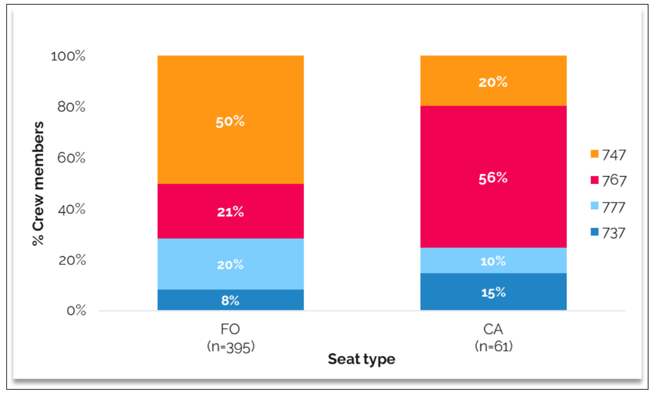


Pilot Attrition – Companies that Pilots are preferring to move



Pilot Attrition Age Group and Fleet/Seat Type



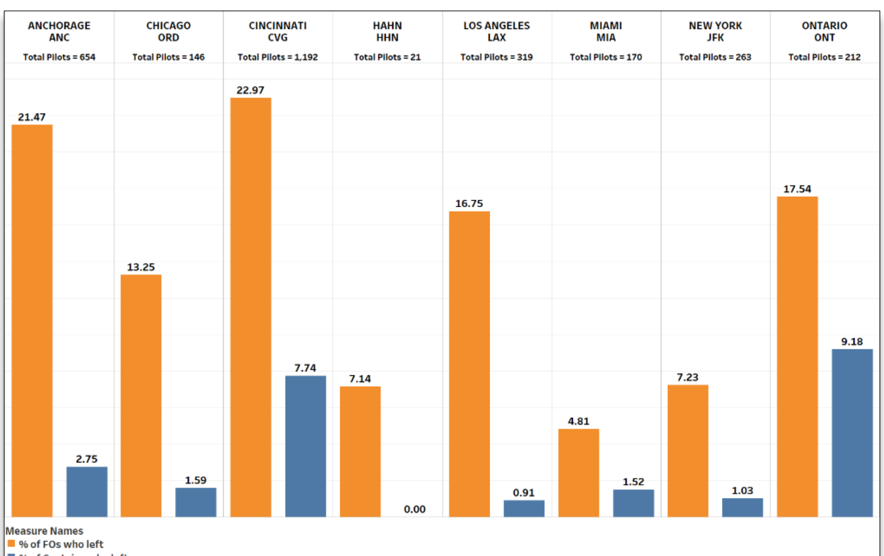


Pilot Attrition from Base

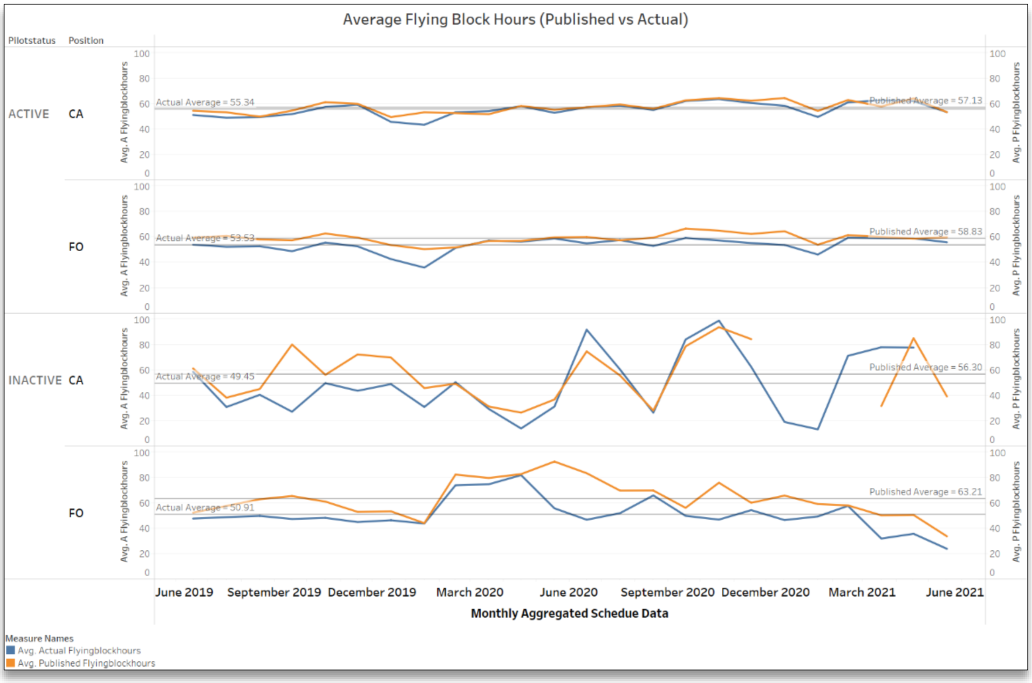
Data

Pilot Statistics from Base.xlsx

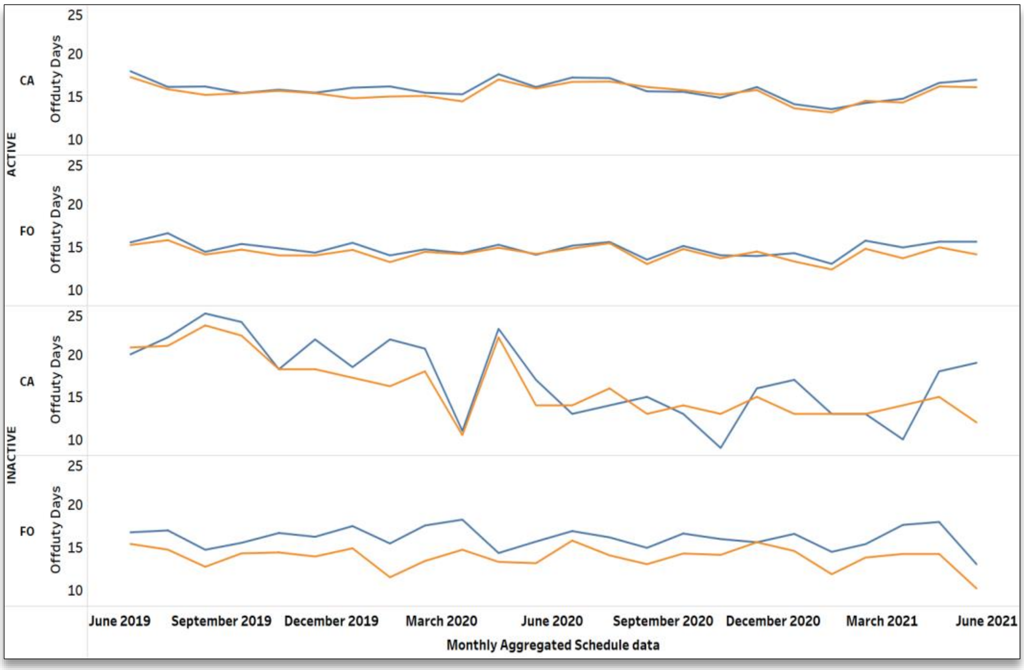
Visualization



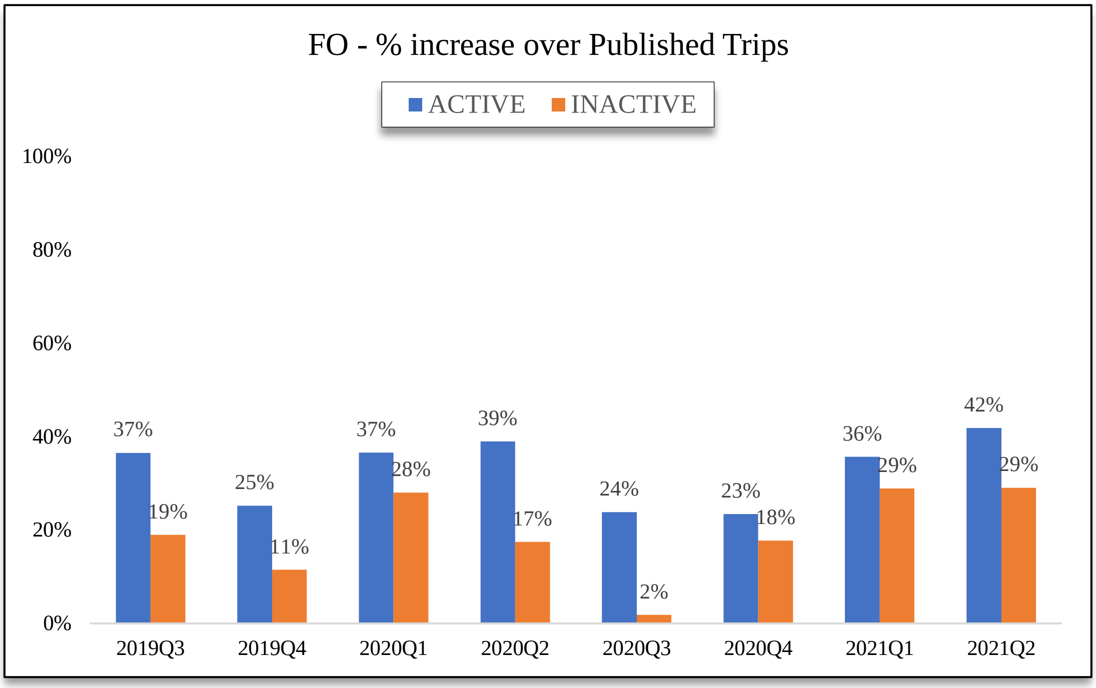
Published vs Actual Monthly Flying Block Hours



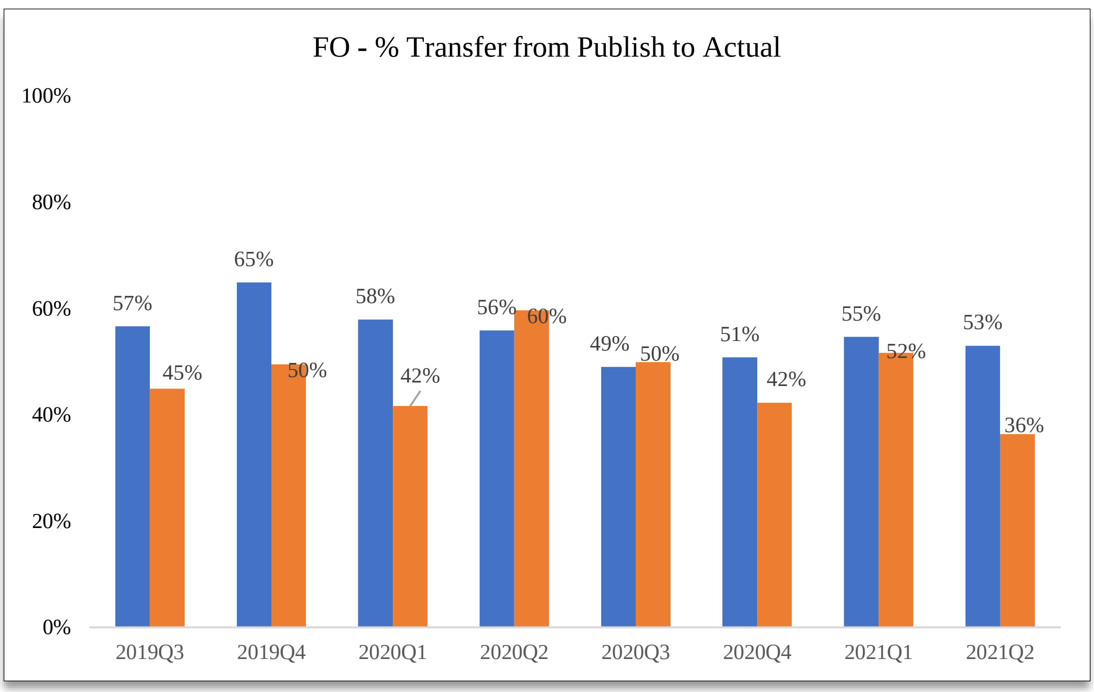
Published vs Actual Off Duty Days



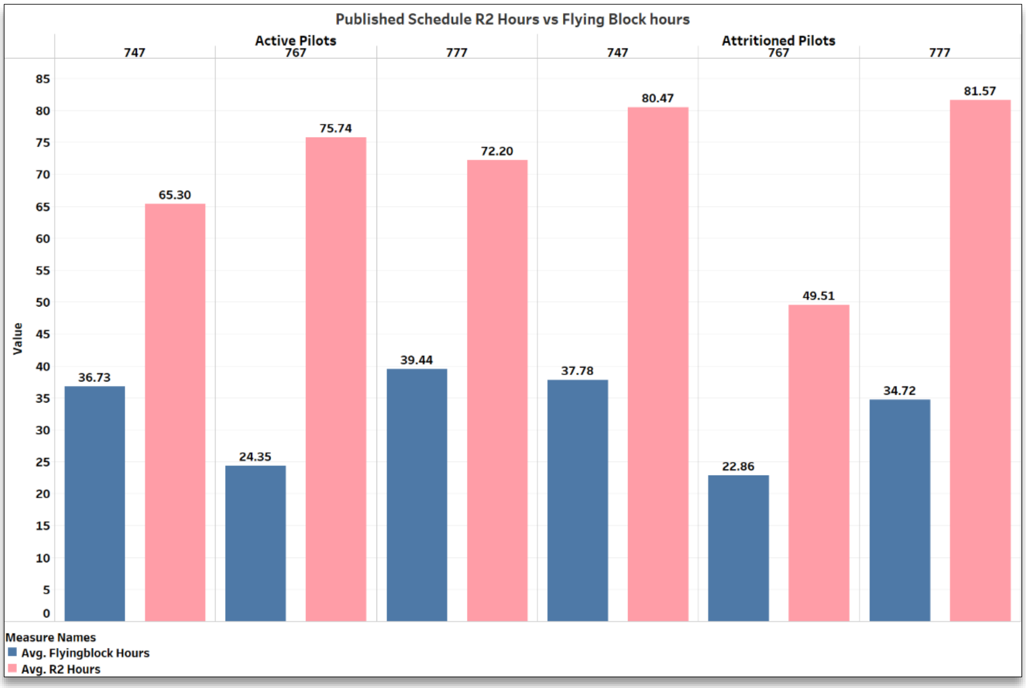
% Increase over Published Trips



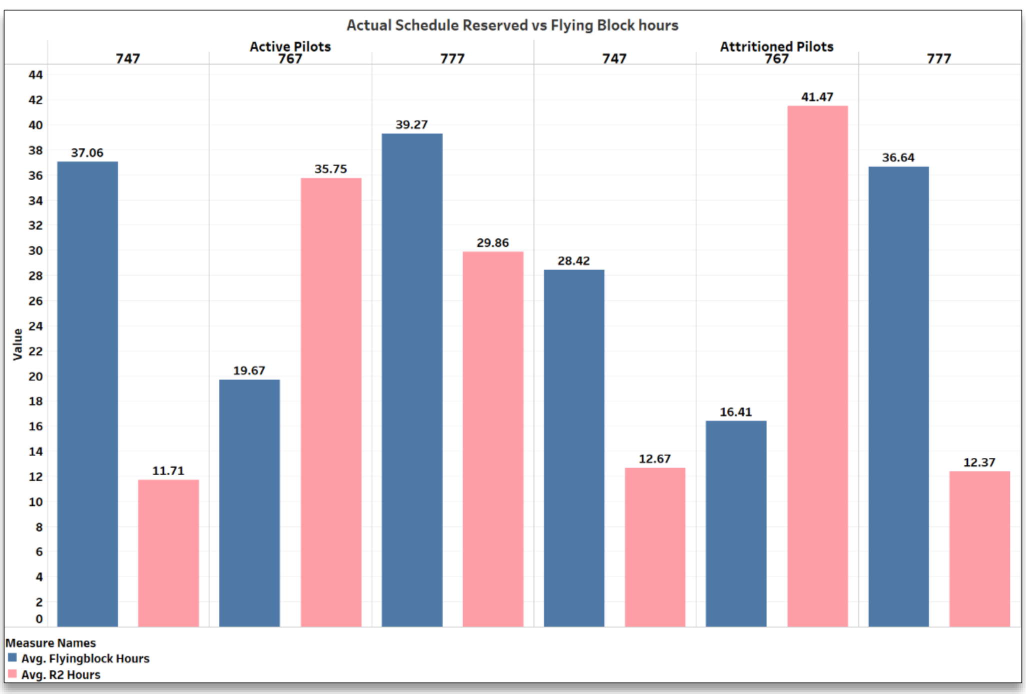
% Transfer from Publish to Actual



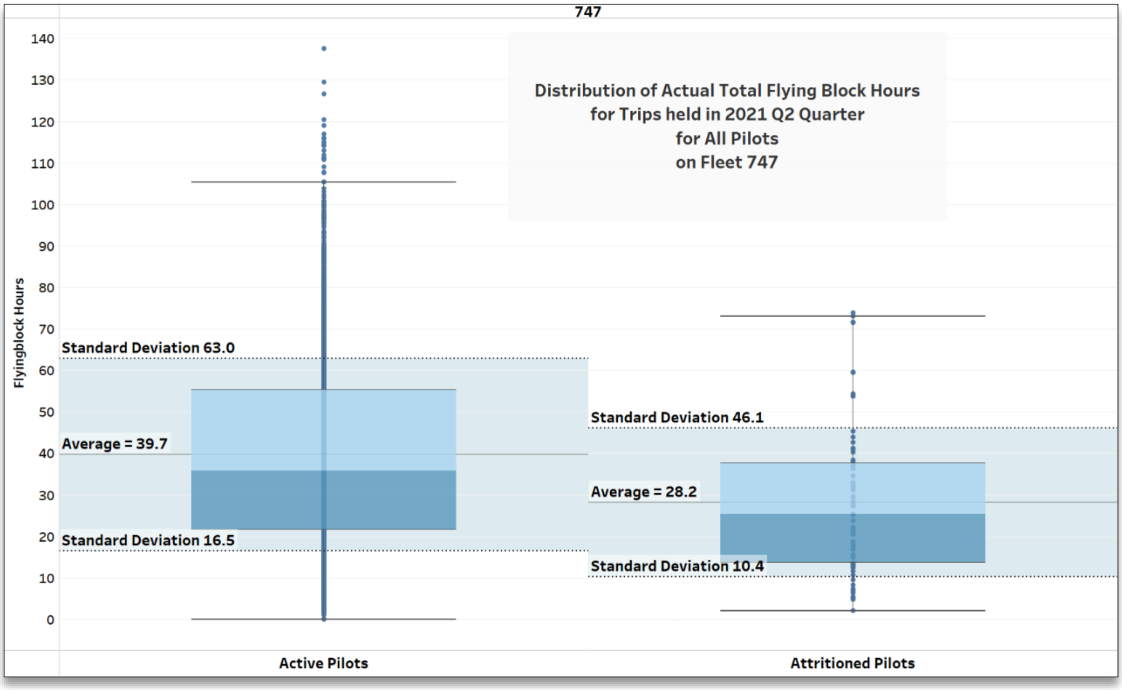
Published schedule - R2 hours vs Flying Block Hours



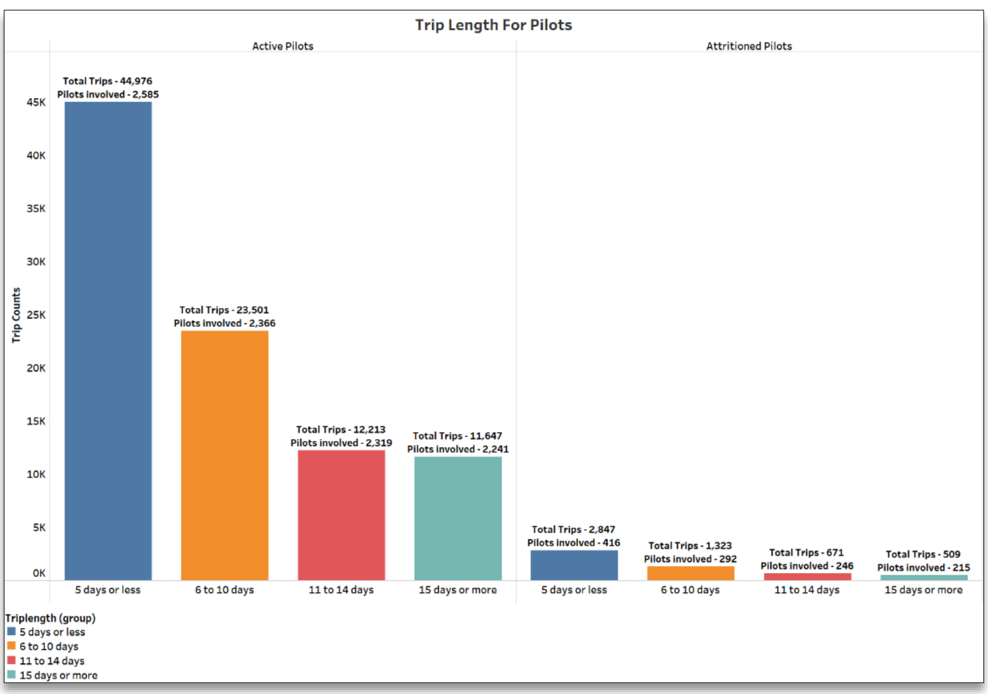
Actual schedule - R2 hours vs Flying Block Hours



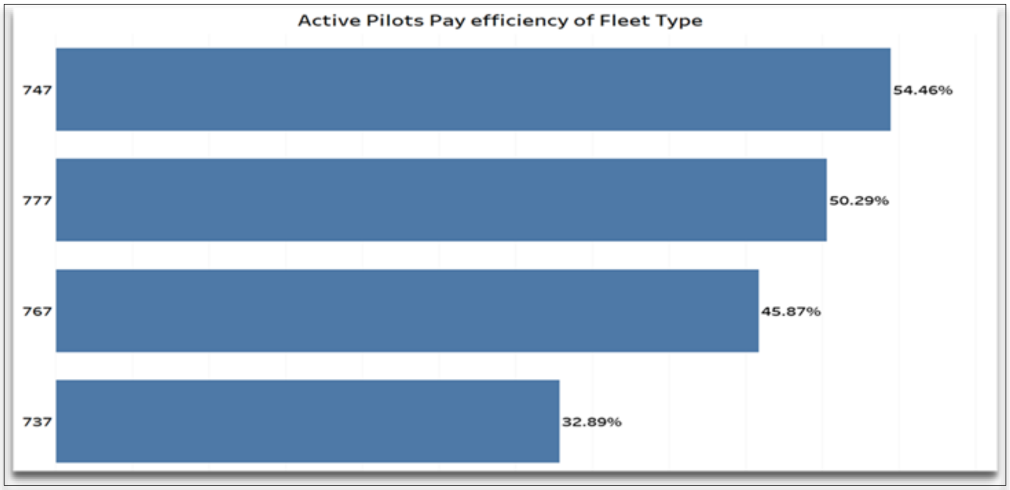
Box Plot of Actual Flying Block Hours for Q2 2021 trips



Trip Length for Pilots

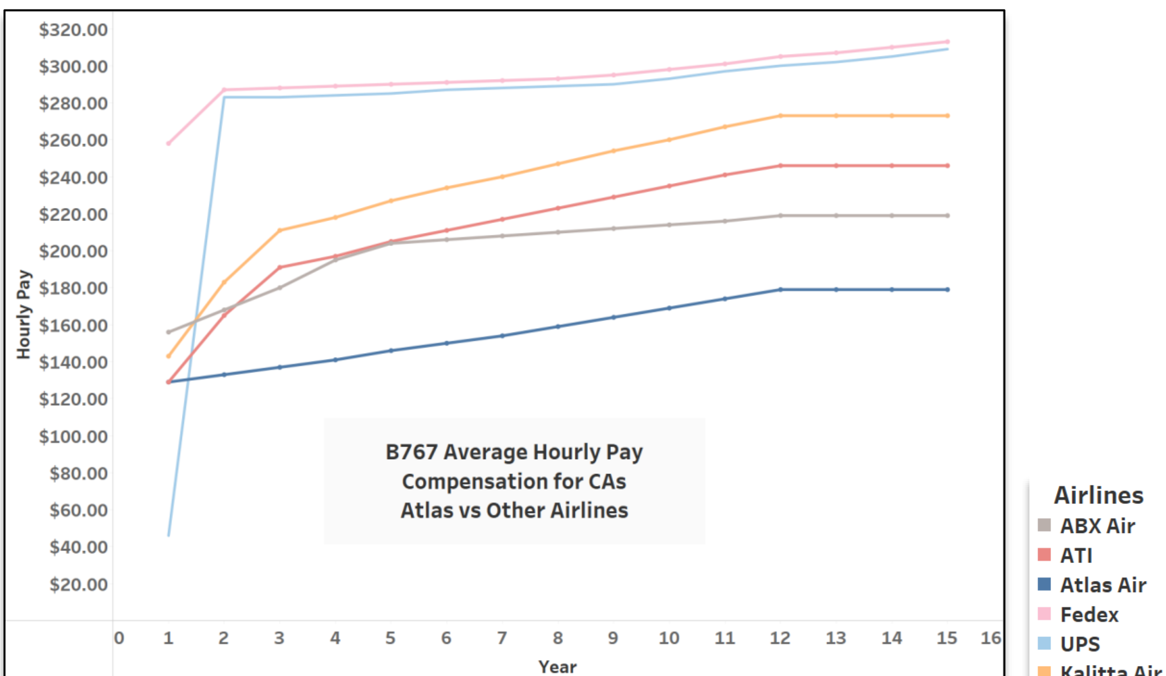


Pay efficiency:



Hourly Pay Compensation compared with Airlines

AtlasHourlyPayRatesVsOthers.xlsx

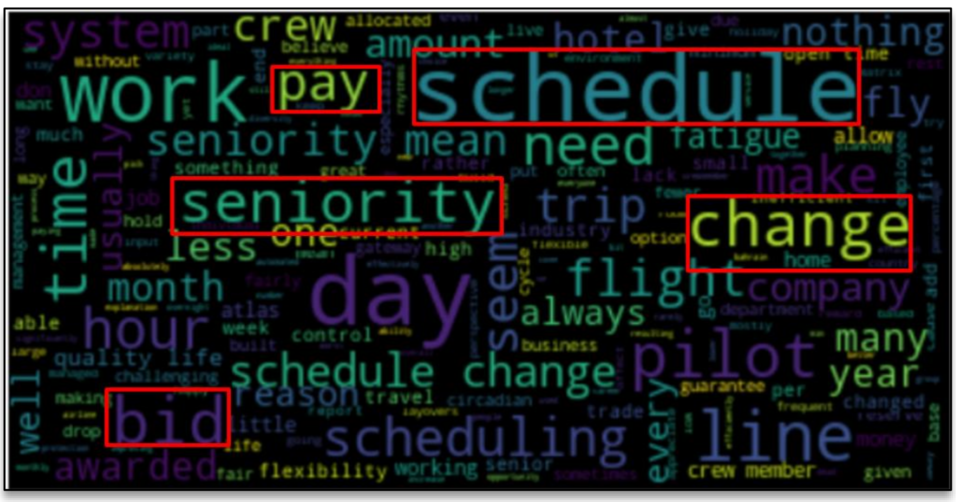


Renewed Contract Pay Scale

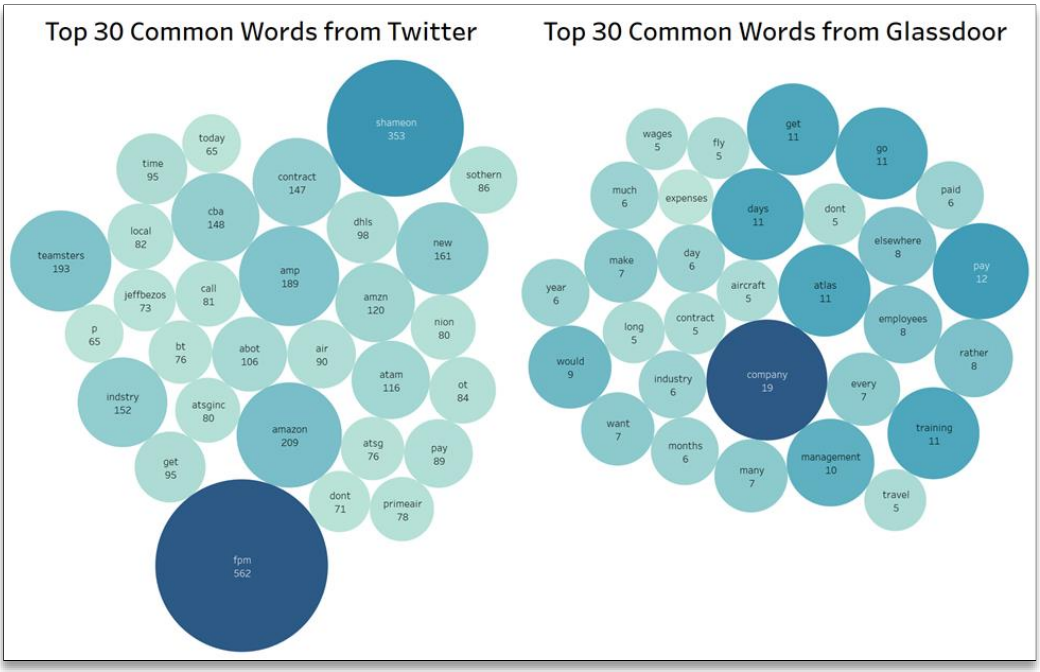
Hourly Wage Rate Analysis

	Current	Union	Atlas	Award
B747 %	100.00%	100.00%	100.00%	100.00%
B747 TOP Rate	$234.65	$325.00	$293.44	$295.58
B777 %	91.98%	100.00%	97.46%	97.46%
B777 TOP Rate	$215.84	$325.00	$285.98	$288.07
B767 %	83.94%	92.52%	94.92%	94.92%
B767 TOP Rate	$196.97	$300.70	$278.52	$280.56
B737 %	75.91%	76.92%	75.91%	75.91%
B737 TOP Rate	$178.12	$250.00	$222.75	$224.37
COLA	N/A	3.0%	2.5%	3.0%
Longevity AVG YOS 1-12	3.92%	5.69%	6.01%	4.33%
FO% of CA (Median/Average)	68%/67.51%	68%/64.47%	68%/67.97%	68%/65.44%

Internal word image:

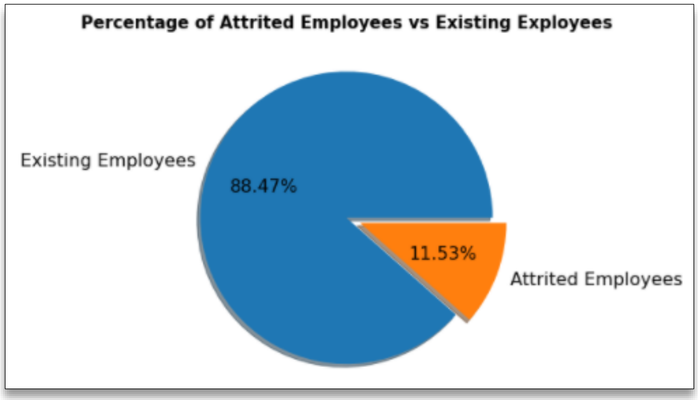


Top 30 words:

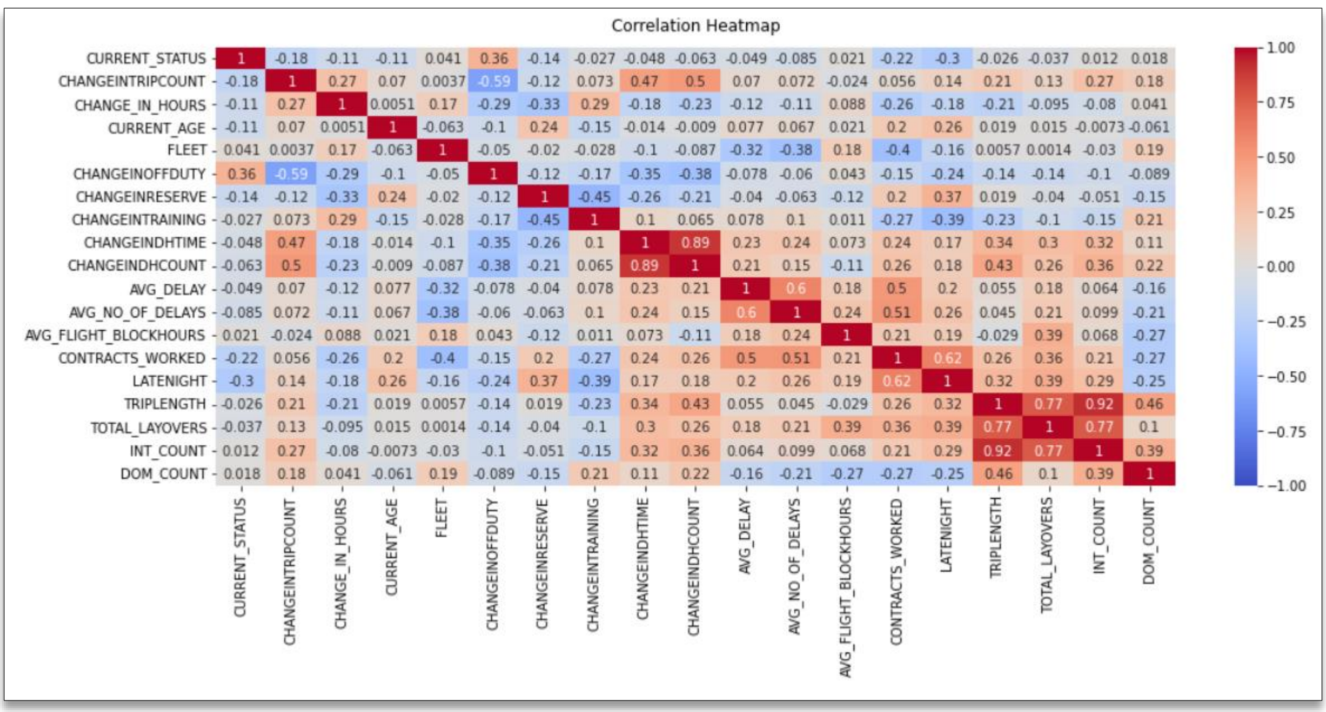


Pros of Text

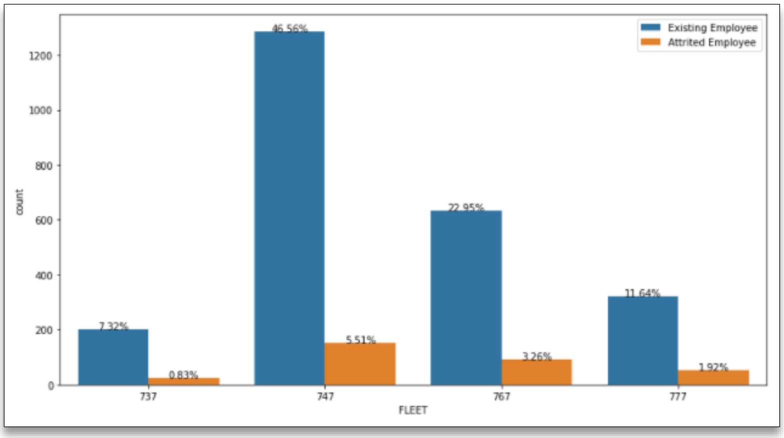
AttritionVsActive:



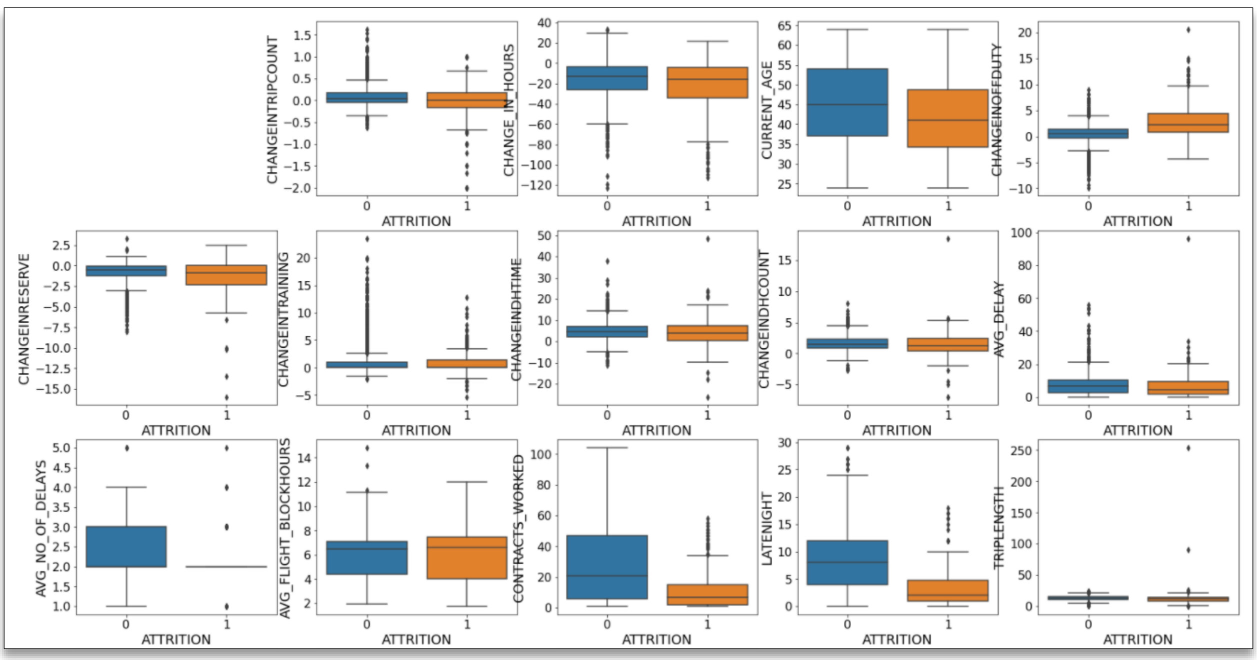
CorrelationMatrix:



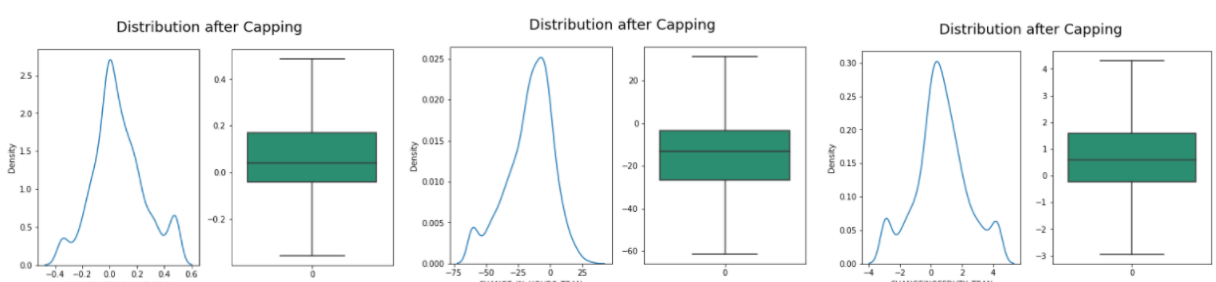
FleetExploration:

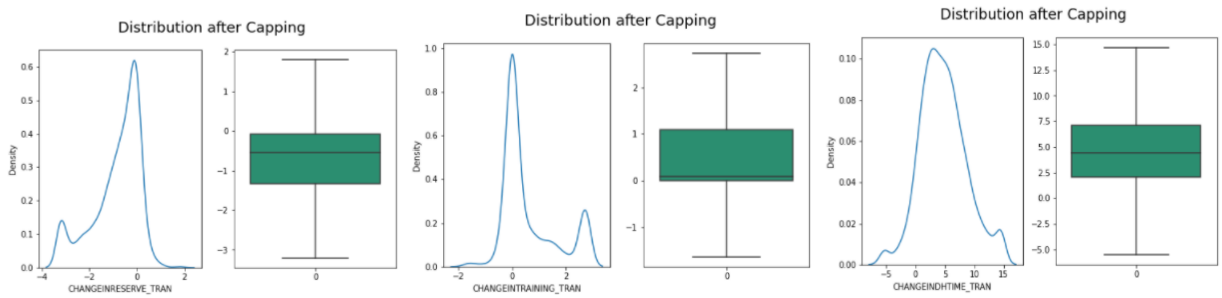


ContinuousVariablesExploration:



DataTransformation:





IndicatorColumnsCreation:

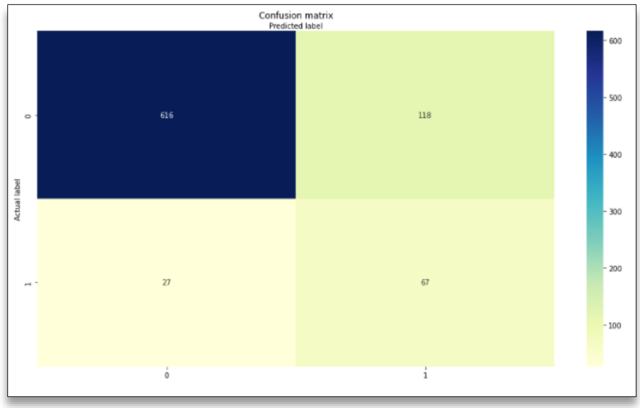
FLEET_747	FLEET_767	FLEET_777	SEAT_FO
0	1	0	1
1	0	0	0
1	0	0	1
1	0	0	1
1	0	0	1

LogitRegressionAnalysis:

```
Optimization terminated successfully.
         Current function value: 0.362850
         Iterations 8
                           Logit Regression Results
==============================================================================
Dep. Variable:         CURRENT_STATUS   No. Observations:                 3412
Model:                          Logit   Df Residuals:                     3394
Method:                           MLE   Df Model:                           17
Date:                Thu, 09 Dec 2021   Pseudo R-squ.:                  0.4765
Time:                        21:27:56   Log-Likelihood:                -1238.0
converged:                       True   LL-Null:                       -2365.0
Covariance Type:            nonrobust   LLR p-value:                     0.000
===========================================================================================
                              coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------------
CURRENT_AGE                -0.0306      0.005     -6.283      0.000      -0.040      -0.021
AVG_FLIGHT_BLOCKHOURS       0.5056      0.049     10.416      0.000       0.410       0.601
CONTRACTS_WORKED           -0.0619      0.006    -10.631      0.000      -0.073      -0.051
TOTAL_LAYOVERS             -0.1685      0.064     -2.619      0.009      -0.295      -0.042
INT_COUNT                   1.6751      0.200      8.372      0.000       1.283       2.067
DOM_COUNT                   6.2854      1.722      3.650      0.000       2.910       9.661
CHANGE_IN_HOURS_TRAN       -0.0112      0.004     -3.051      0.002      -0.018      -0.004
CHANGEINOFFDUTY_TRAN        0.6433      0.045     14.194      0.000       0.554       0.732
CHANGEINTRAINING_TRAN      -0.3096      0.065     -4.797      0.000      -0.436      -0.183
CHANGEINDHTIME_TRAN         0.0875      0.018      4.982      0.000       0.053       0.122
AVG_DELAY_TRAN              0.0993      0.016      6.053      0.000       0.067       0.132
AVG_NO_OF_DELAYS_TRAN      -0.3067      0.125     -2.454      0.014      -0.552      -0.062
LATENIGHT_TRAN             -0.2601      0.020    -13.043      0.000      -0.299      -0.221
DOM_COUNT_TRAN             -7.9031      1.843     -4.287      0.000     -11.516      -4.290
FLEET_747                  -1.6993      0.226     -7.516      0.000      -2.142      -1.256
FLEET_767                  -1.3229      0.154     -8.601      0.000      -1.624      -1.021
FLEET_777                  -3.4587      0.297    -11.659      0.000      -4.040      -2.877
SEAT_FO                    -0.2658      0.129     -2.062      0.039      -0.518      -0.013
===========================================================================================
```

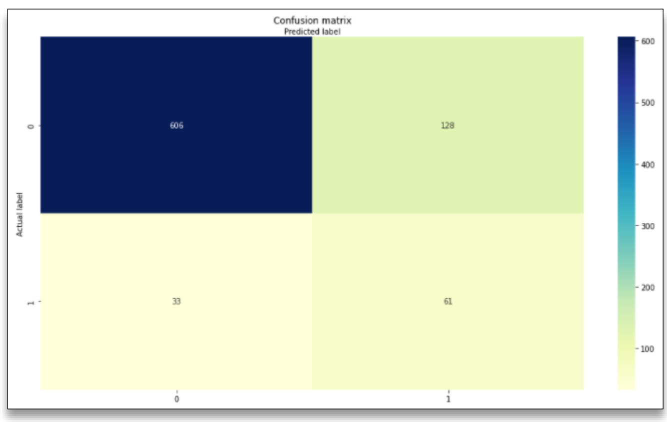
LogisticRegressionResults:

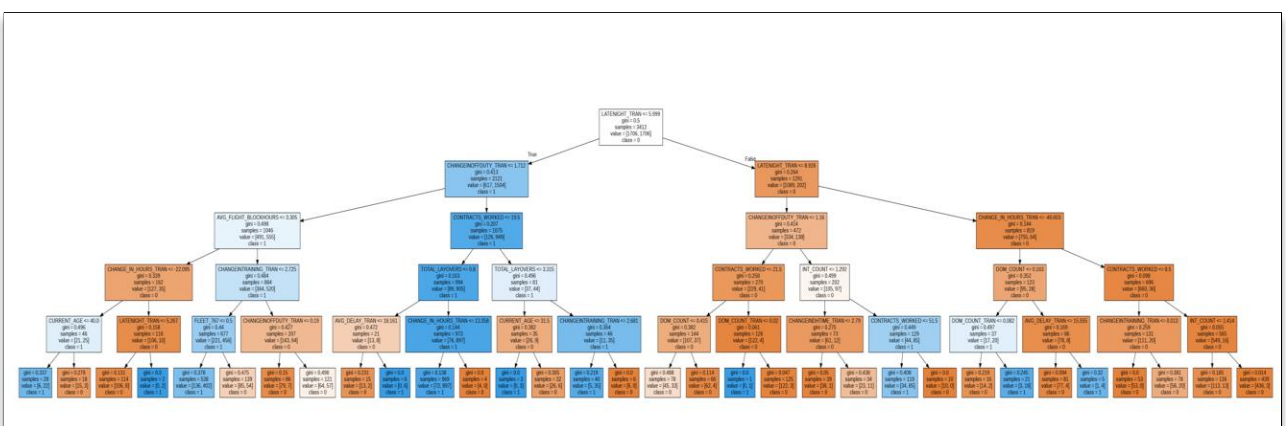
```
Accuracy: 82.487922705314

Precision: 0.3621621621621622

Recall: 0.7127659574468085

ConfusionMatrix:
 [[616 118]
 [ 27  67]]

Classification_report:
              precision    recall  f1-score   support

           0       0.96      0.84      0.89       734
           1       0.36      0.71      0.48        94

    accuracy                           0.82       828
   macro avg       0.66      0.78      0.69       828
weighted avg       0.89      0.82      0.85       828
```



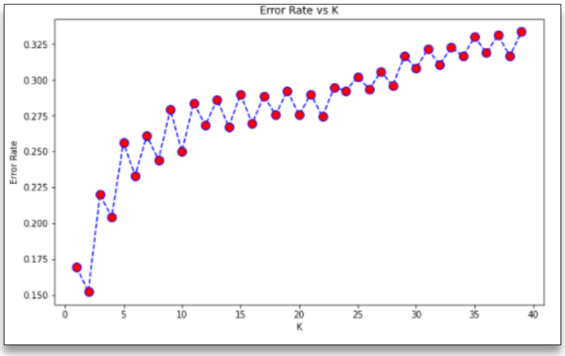
DecisionTreeResults:

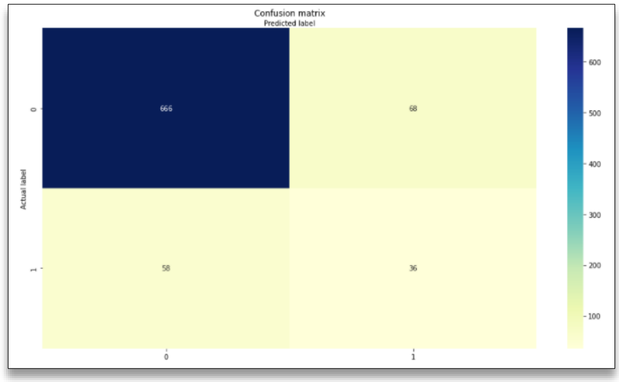
```
Accuracy: 80.55555555555556

Precision: 0.32275132275132273

Recall: 0.648936170212766

ConfusionMatrix:
 [[606 128]
 [ 33  61]]

Classification_report:
              precision    recall  f1-score   support

           0       0.95      0.83      0.88       734
           1       0.32      0.65      0.43        94

    accuracy                           0.81       828
   macro avg       0.64      0.74      0.66       828
weighted avg       0.88      0.81      0.83       828
```



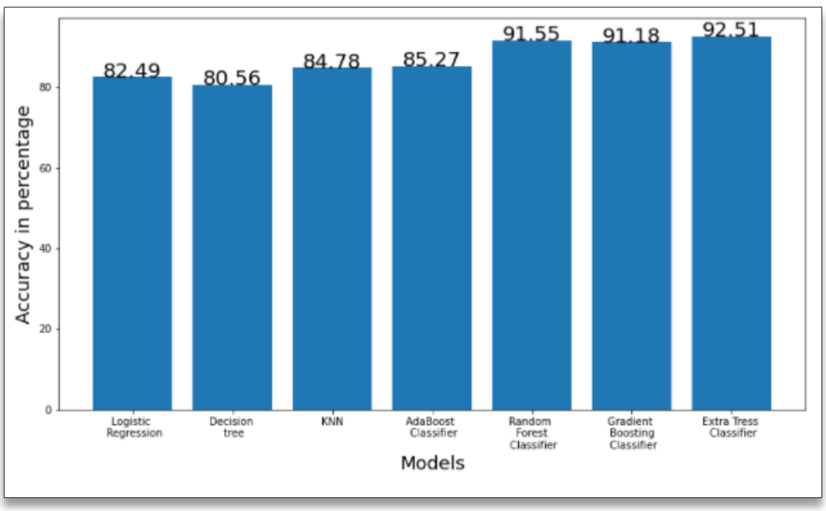


KNNResults:

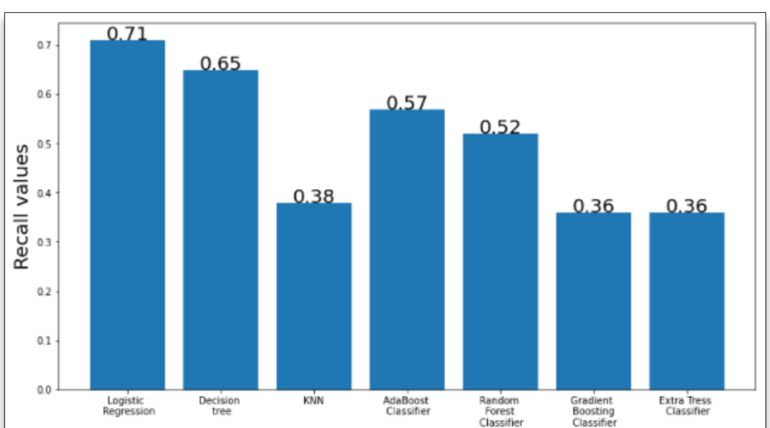




AccuracyComparison:



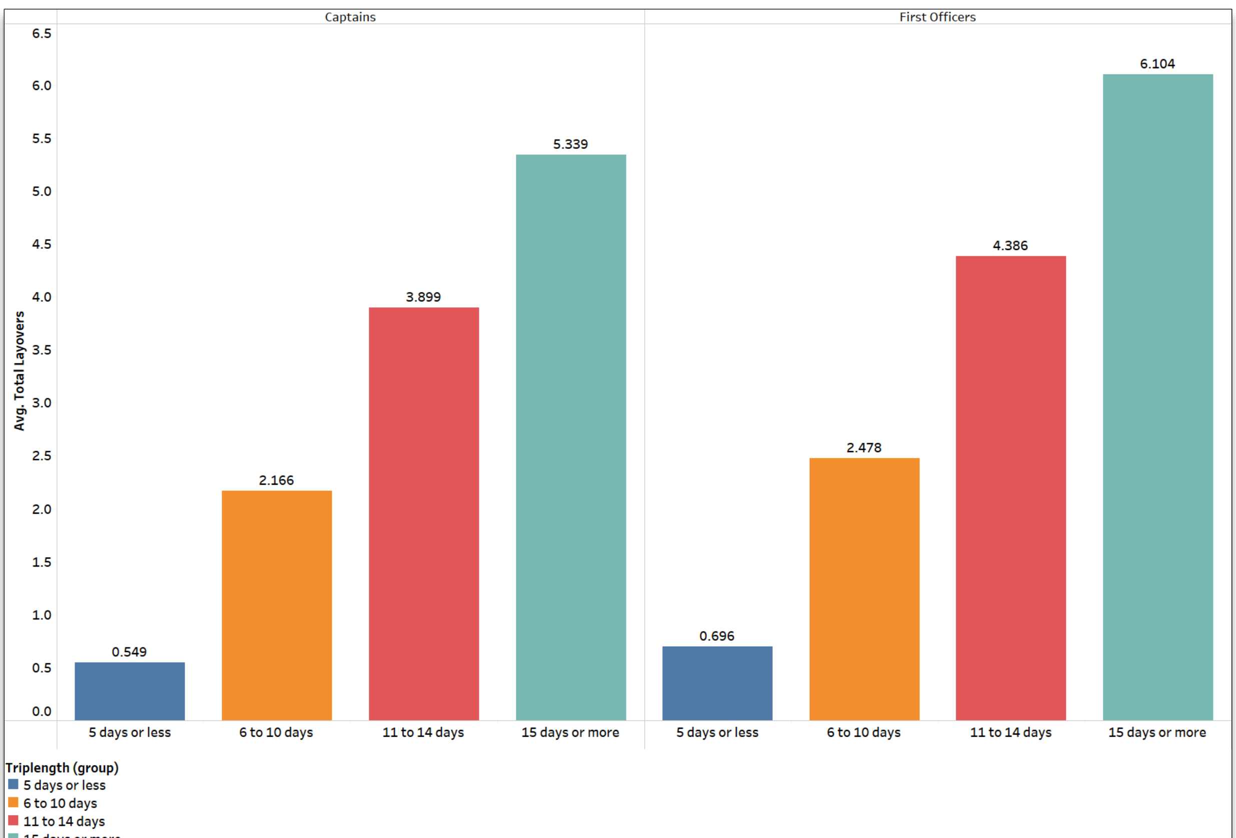
RecallComparison:



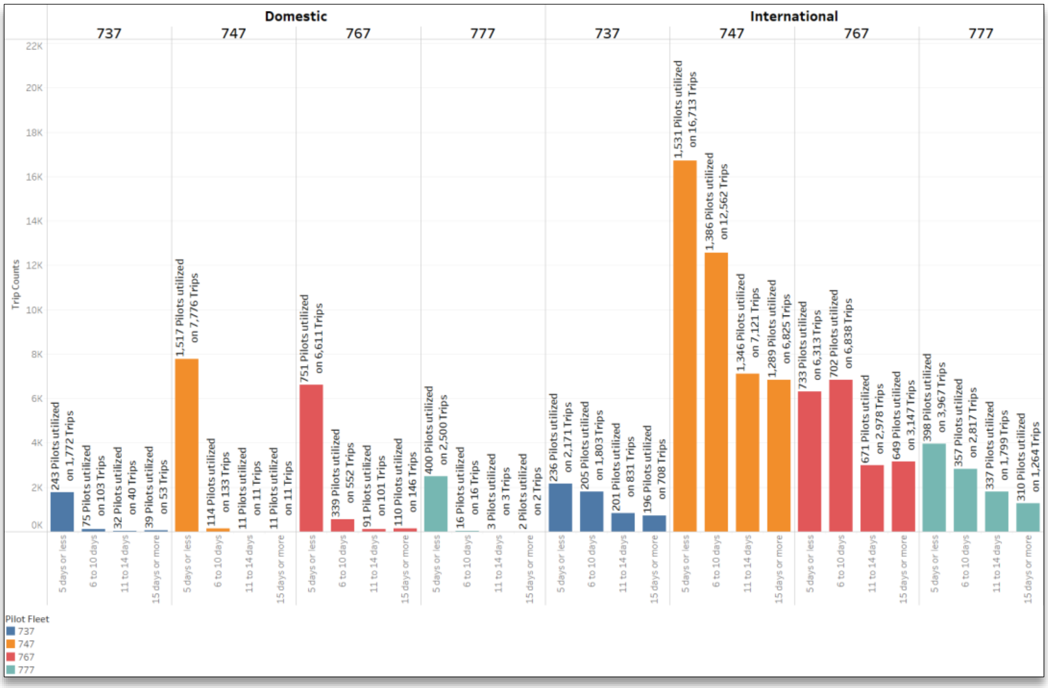
ODDsRatio:

	odds_ratio	variable
4	5.834932	INT_COUNT
7	1.893836	CHANGEINOFFDUTY_TRAN
5	1.738159	DOM_COUNT
1	1.695209	AVG_FLIGHT_BLOCKHOURS
10	1.093586	AVG_DELAY_TRAN
9	1.085503	CHANGEINDHTIME_TRAN
6	0.987420	CHANGE_IN_HOURS_TRAN
0	0.977285	CURRENT_AGE
2	0.938965	CONTRACTS_WORKED
3	0.845626	TOTAL_LAYOVERS
11	0.838339	AVG_NO_OF_DELAYS_TRAN
17	0.794104	SEAT_FO
12	0.772256	LATENIGHT_TRAN
8	0.739696	CHANGEINTRAINING_TRAN
15	0.314488	FLEET_767
13	0.233279	DOM_COUNT_TRAN
14	0.225925	FLEET_747
16	0.036376	FLEET_777

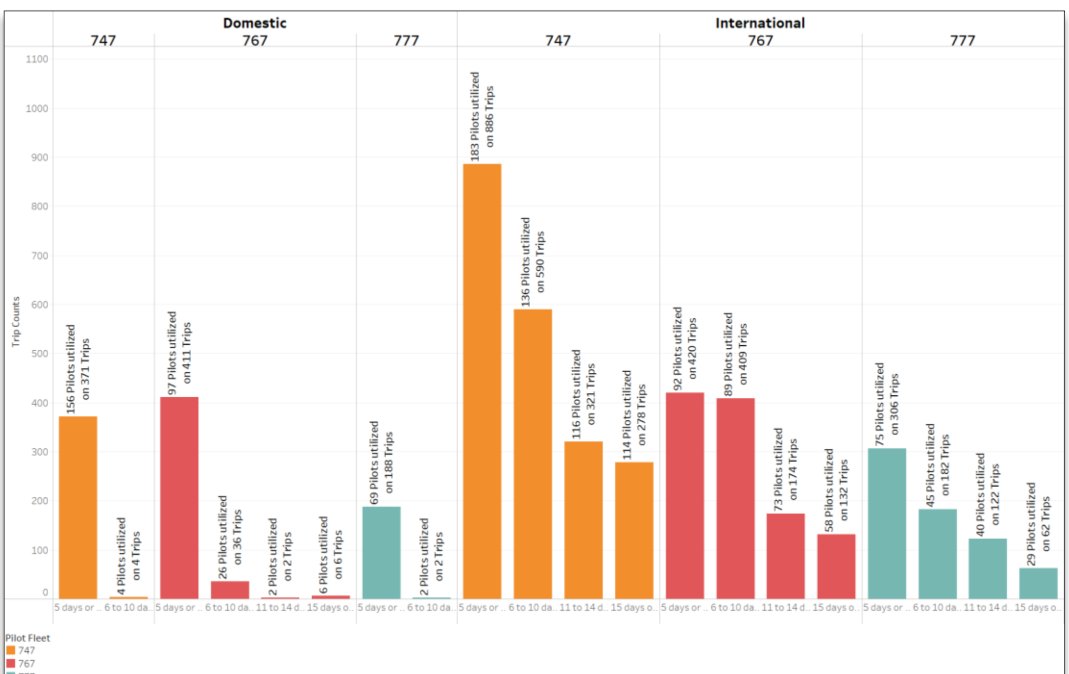
Layover Count



Domestic or International Trip Analysis



Domestic or International Trip Analysis for Attrition Pilots



Domestic or International Trip Analysis for Active Pilots