

FlyFast Airlines:

Reaching New Hieghts In Cutomer Satisfaction

**FlyFast Airlines Analysis Report**

IST 707 Spring 2020

A Team | Ralph Parlin and Kobi Wiseman

**Table of Contents**

[**SECTION 1: ANALYSIS** 3](#_Toc39128830)

[**Summary, Key Business Questions and Methodology** 3](#_Toc39128831)

[**Project Summary** 3](#_Toc39128832)

[**Business Questions Addressed** 3](#_Toc39128833)

[**Methodology** 4](#_Toc39128834)

[**Data Pre-processing and Preparation** 4](#_Toc39128835)

[**Dataset Description** 4](#_Toc39128836)

[**Treatment for Missing Data** 4](#_Toc39128837)

[**Feature Engineering** 5](#_Toc39128838)

[**Data Transformation and Analysis Architecture** 5](#_Toc39128839)

[**Descriptive Statistics & Visualizations** 6](#_Toc39128840)

[**Analysis and Modeling** 14](#_Toc39128841)

[**Market Segmentation** 14](#_Toc39128842)

[**Traveler Personas** 19](#_Toc39128843)

[**Sentiment Analysis** 20](#_Toc39128844)

[**Modeling** 21](#_Toc39128845)

[**SECTION 2: MARKETING PLAN** 23](#_Toc39128846)

[**Business Intelligence: Key Findings from the Analysis** 23](#_Toc39128847)

[**Current Market Positions** 23](#_Toc39128848)

[**SWOT Analysis** 24](#_Toc39128849)

[**Immediate and Actionable Recommendations to Improve Satisfaction** 25](#_Toc39128850)

[**SECTION 3: ANNEXES** 26](#_Toc39128851)

[**Annex 1: Work Flow** 26](#_Toc39128852)

[**Annex 2: Description of Features in the Survey Dataset** 28](#_Toc39128853)

[**Annex 3: Persona Classification Tree Diagrams** 30](#_Toc39128854)

**SECTION 1: ANALYSIS**

**Summary, Key Business Questions and Methodology**

**Project Summary**

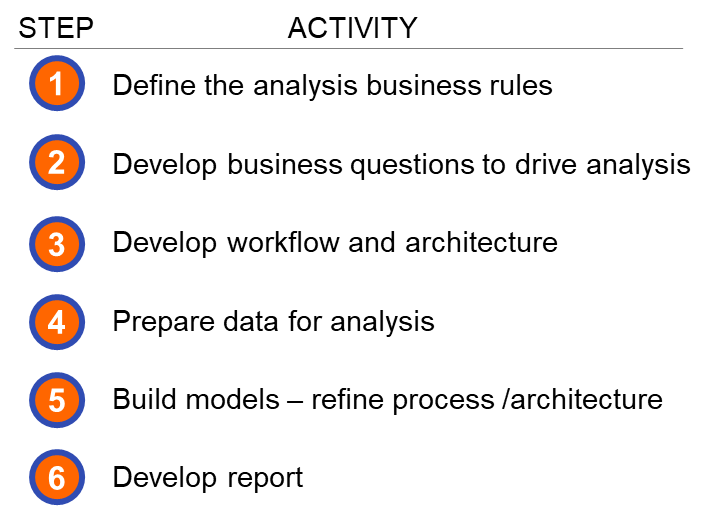
The purpose of this analysis was to identify the primary drivers of an airline passenger’s likelihood to recommend, gain a better understanding of FlyFast Airway’s market position, and offer a broad plan that leverages these findings to improve FlyFast Airway’s customer satisfaction. The data used for this analysis was collected via a survey of airline passengers traveling with FlyFast, Cheapseats and Sigma airlines from January 2014 to March 2014. The survey recorded 4,985 observations across 23 variables pertaining to traveler demographics, purchase behavior, flight information and most importantly, their likelihood to recommend the airline. Each traveler’s likelihood to recommend an airline is provided on a scale from 1 (unlikely) to 10 (very likely). For this analysis, each traveler’s likelihood to recommend measure was discretized into “Detractor”, “Passive” and “Promoter” statuses. This transformation provided us a discrete measure of traveler satisfaction and served as the cornerstone of this analysis allowing us to align our findings with the industry standard of Net Promoter Score (NPS) ratings. The analysis was conducted using the Orange suite of data analysis and machine learning tools and resulted in this report which consists of in three sections. Section one includes executive level details on the analysis process, section two provides high-level marketing recommendations based on our findings, and section three includes detailed annexes of information we reference throughout the analysis.

**Business Questions Addressed**

The following questions served as the primary guide for the direction and framing of our analysis:

1. How can we better understand the air travel market? What are the general patterns we see in the data?
2. How does segmenting the market in various ways change our understanding?
3. What is unique about highly satisfied travelers (Age, gender, loyalty status, price sensitivity)? What is unique about unhappy travelers? What is the driving factor of low satisfaction? What is required to move our unhappy customers to becoming promoters?
4. What is the marketing landscape of traveler satisfaction between FlyFast, Cheapseats and Sigma airlines look like? Who is the best, who is the worst (as a percentage of their flights)? How can we leverage these findings through SWOT analysis to improve customer experience on FlyFast and move them to Promoter status?

**Methodology**

Our methodology included breaking the analysis into a series of steps that allowed us to explore the data and identify key findings within the time available for the study. The six- step procedure below leads us to our results. We iterated on step 5 several times making adjustments as needed to derive our final analysis. Each part of the analysis was conducted jointly by two analysts to eliminate the chance of error and to explore the data in multiple ways. Furthermore, in order to focus our analysis on likelihood to recommend, we organized our analysis along two lines of effort that relate directly to the main players of air travel: the travelers, and airline carriers. To that end, we used this bisectional breakdown as a lens from which to view the dataset and provide context to the analysis. Additionally, when appropriate, we grouped all observations by their net promoter status (promoter, passive detractor) and studied each subset individually to gain a better appreciation for the relationships that exist in each. With respect to airline travelers, this approach allowed us to generate customer personas for each NPS status. For airline carriers, this approach allowed us to compare the strengths, weakness, opportunities and threats for FlyFast regarding its competitors.

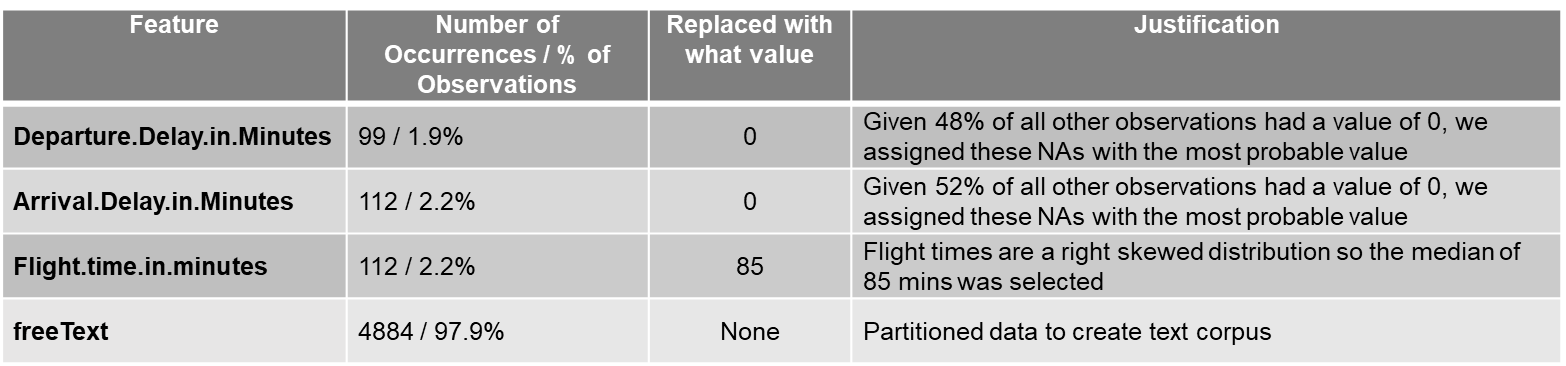
**Data Pre-processing and Preparation**

**Dataset Description**

The data for this analysis was collected via survey of 4,985 passengers flying between 01 January 2014 until 09 March 2014. The survey collected data across 23 variables pertaining to the traveler demographics, behavior, flight information and most importantly, their likelihood to recommend the airline. This information was collected for FlyFast Airways, as well as their two nearest competitors, Cheapseats Airlines and Sigma Airlines.

**Treatment for Missing Data**

Our first step was to identify missing data in the dataset and develop a mitigation strategy. The following table contains the variables in the dataset that had no recorded values. The table also shows what measures were used to replace missing values.

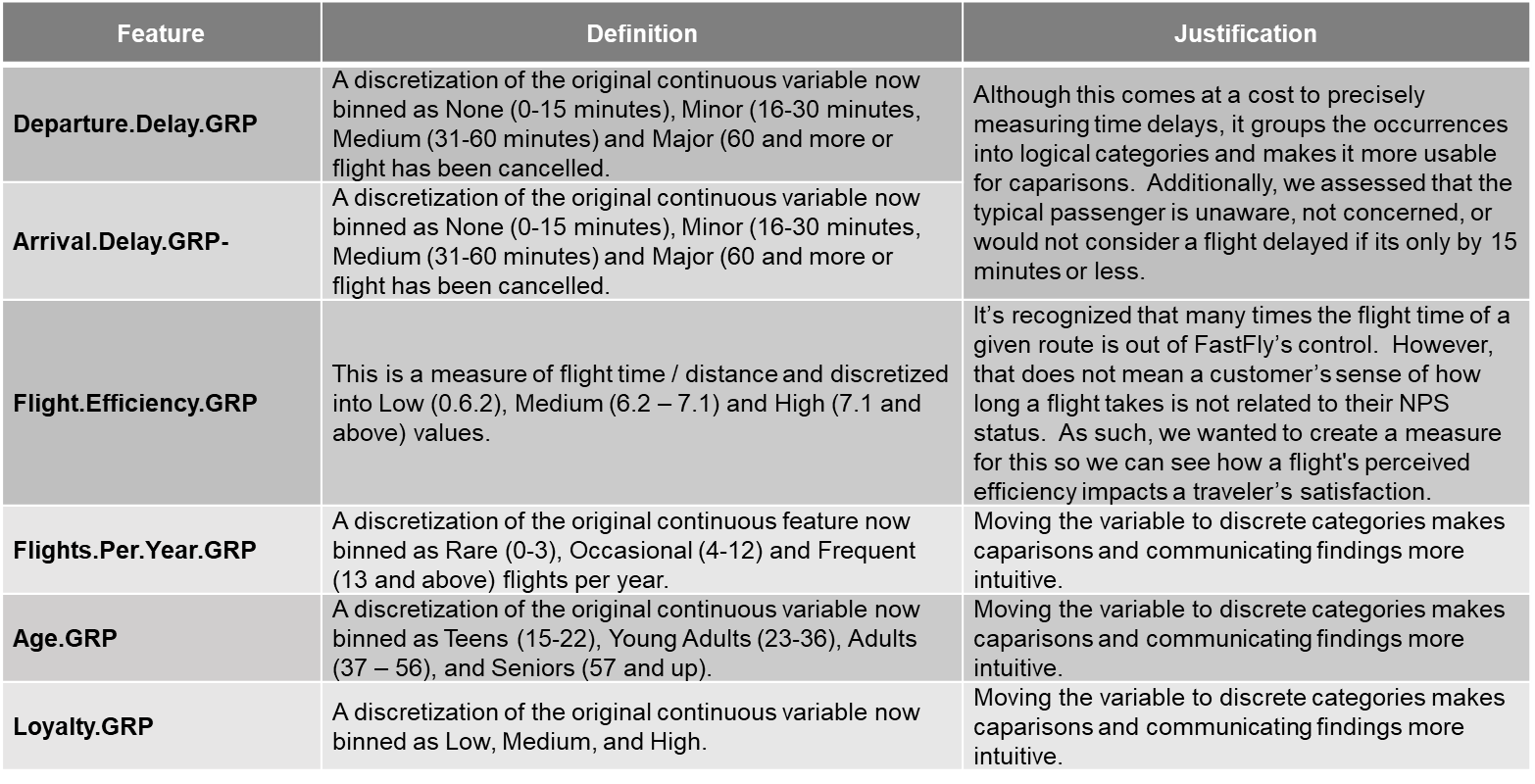


Risk with the NA replacement techniques above is mitigated by the fact that each replacement method impacted very little data (1.9% to 2.2%). No approximation measure was taken for the free text data, the blank entries were simply ignored and the observations with data were partitioned to a separate data frame for text analysis.

**Feature Engineering**

The original dataset had 23 features. The most important transformation was performed prior to this analysis where each customer’s likelihood to recommend was transformed into a discrete promotion status category where likelihood to recommend values of 6 and below were assigned as “Detractors”, 7 and 8 as “Passives”, and 9 and 10 as “Promoters.”

To improve our analysis and simplify the communication of the results, six other features were also engineered. These features, their definitions, and justifications are listed in the table below:



We also decided to exclude several variables from our analysis as well. A complete list of all features and their definitions can be found in ANNEX 2.

**Data Transformation and Analysis Architecture**

After pre-processing the data in Microsoft Excel, we moved the data into Orange and segmented the data along four broad lines to bring focus to the analysis and streamline the creation of data visualizations. This workflow can be seen in Annex 1.

The first segmentation of the data created a data table with traveler information. This data table was airline carrier agnostic allowing us to conduct analysis on industry wide traveler demographics, purchase behavior, flight information and most importantly, their likelihood to recommend the airline demographics. This data channel allowed us to analyses the differences between travelers with respect to their NPS status and build traveler personas for Promoters, Passives, and Detractors.

A second segmentation of the data created a data table for FlyFast travel only. This data channel allowed us to conduct in-depth analysis and modeling of FlyFast operations. This channel was subsequently split further when we partitioned the data into Promoters, Passives, and Detractors. This partition was performed to gain a richer understanding on the unique details of our customers with respect to NPS status.

A third segmentation of the data was to create a data table for FlyFast competitors (Cheapseats and Sigma airlines). This data channel allowed us to conduct in-depth analysis of competing airlines.

The fourth main segmentation of the data was to create a data table that contained the free text data. This data channel allowed us to conduct in-depth text analysis and gain understanding of traveler sentiment.

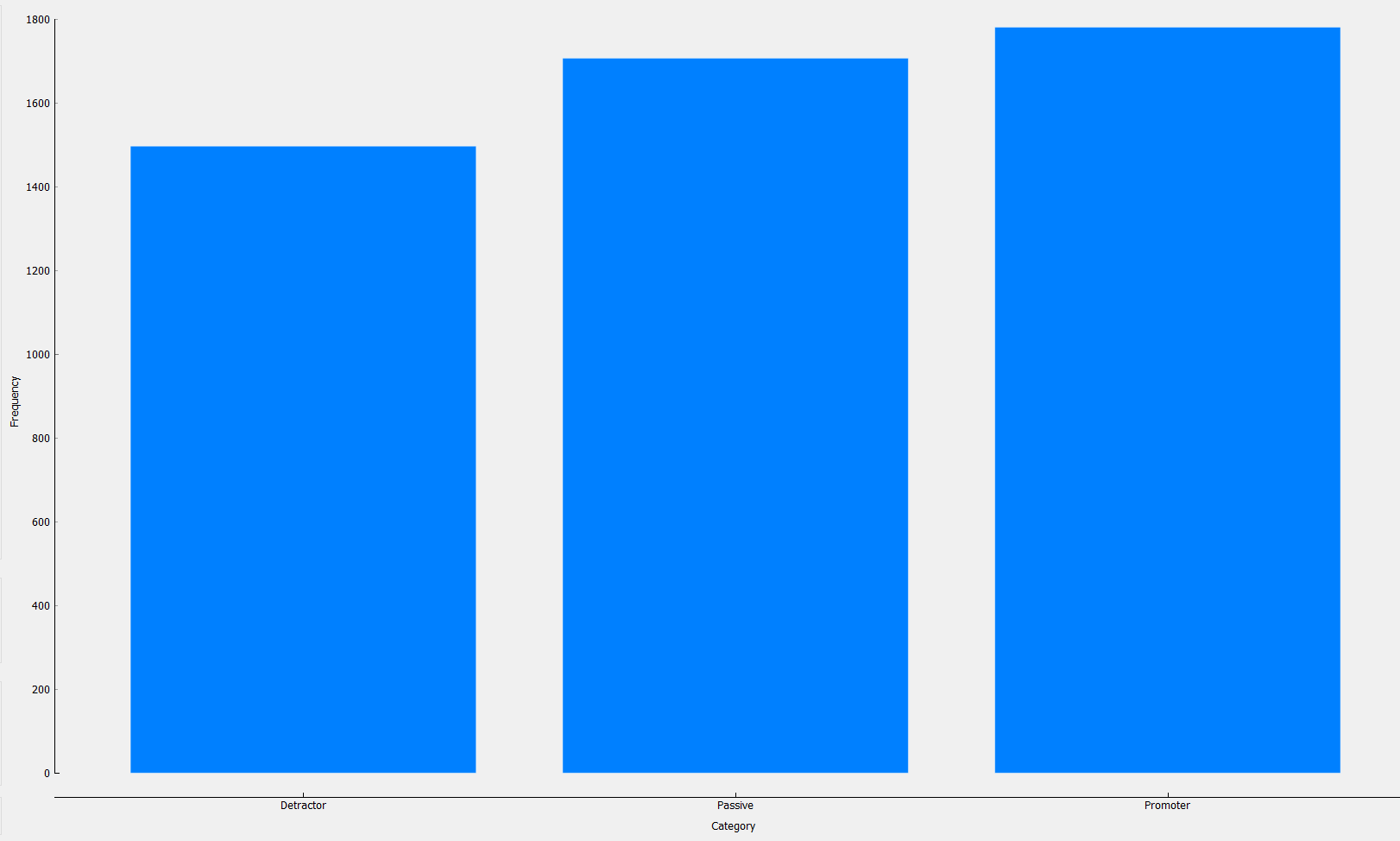
**Descriptive Statistics & Visualizations**

The following section provides a brief overview of the descriptive statistics and visualizations for several of the most salient features used in the analysis. However, a list of the most interesting findings is below:

Key findings:

* Generally, the air travel market is split evenly across NPS status with Promoter taking the majority stake at 35.73%, Passives at 34.24%, and Detractors at 30.03%.
* Number of flights per year is lead predominately by “Frequent” meaning most travelers conduct 13 or more flights in a given year.
* “Blue” is the overwhelming majority for airline status.
* Female passengers constituted 57.39% of all air travelers.
* 59% of travelers are categorized as Low loyalty.
* Most travelers (61%) are categorized as business travelers
* Most travelers (82.7%) travel Economy Class
* 66% of Travelers have between 2 to 3 frequent flyer accounts

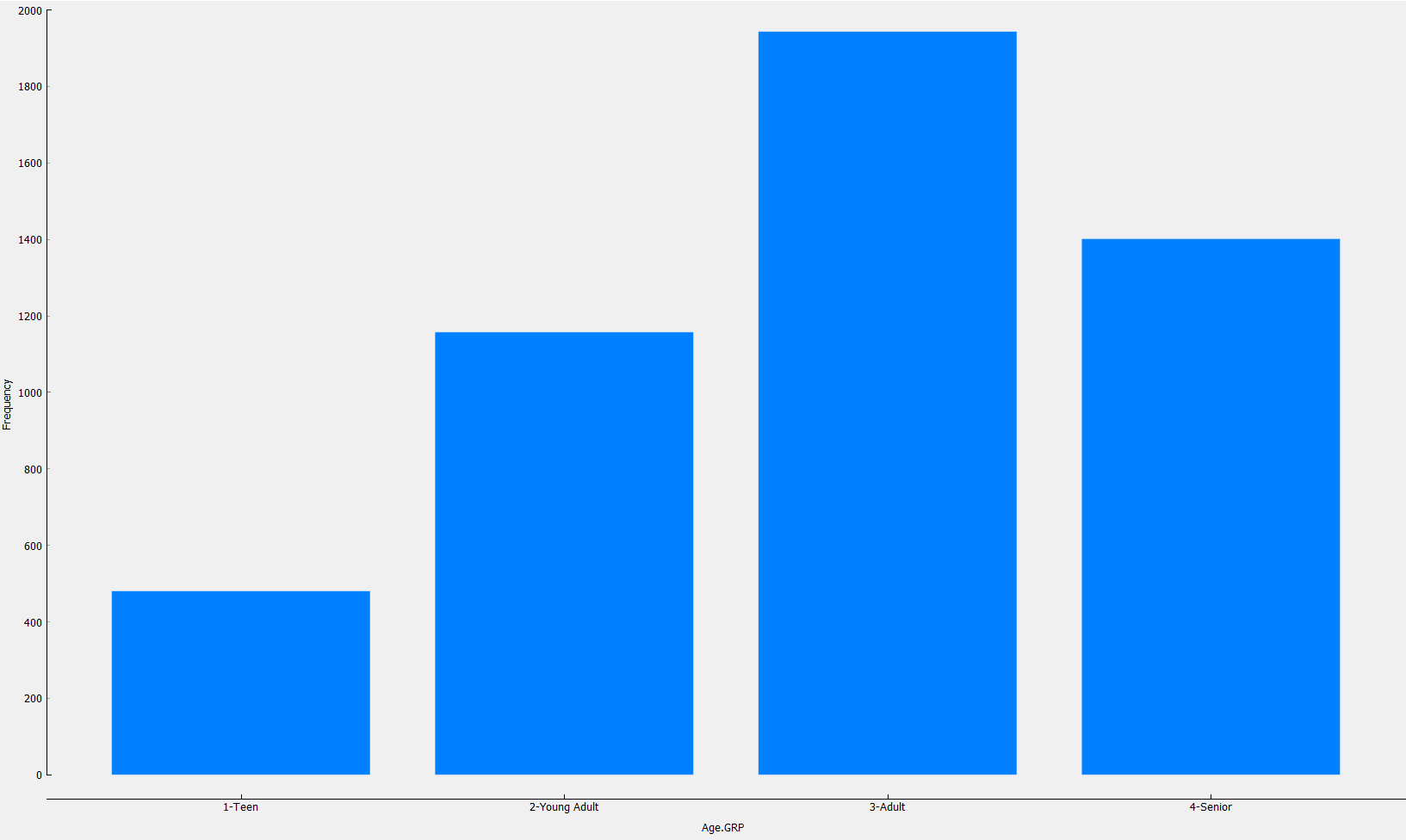
Category (Net Promoter Status)



A critical first step in determining how to increase the Promoters for FlyFast airways is to gain a better appreciation of the total market breakdown by NPS status. As seen in the graph above, generally, the market is split evenly across all NPS status with Promoter taking the majority stake at 35.73%, Passives at 34.24%, and Detractors at 30.03%.

---------------------------------------------------------------------------------------------------------------------

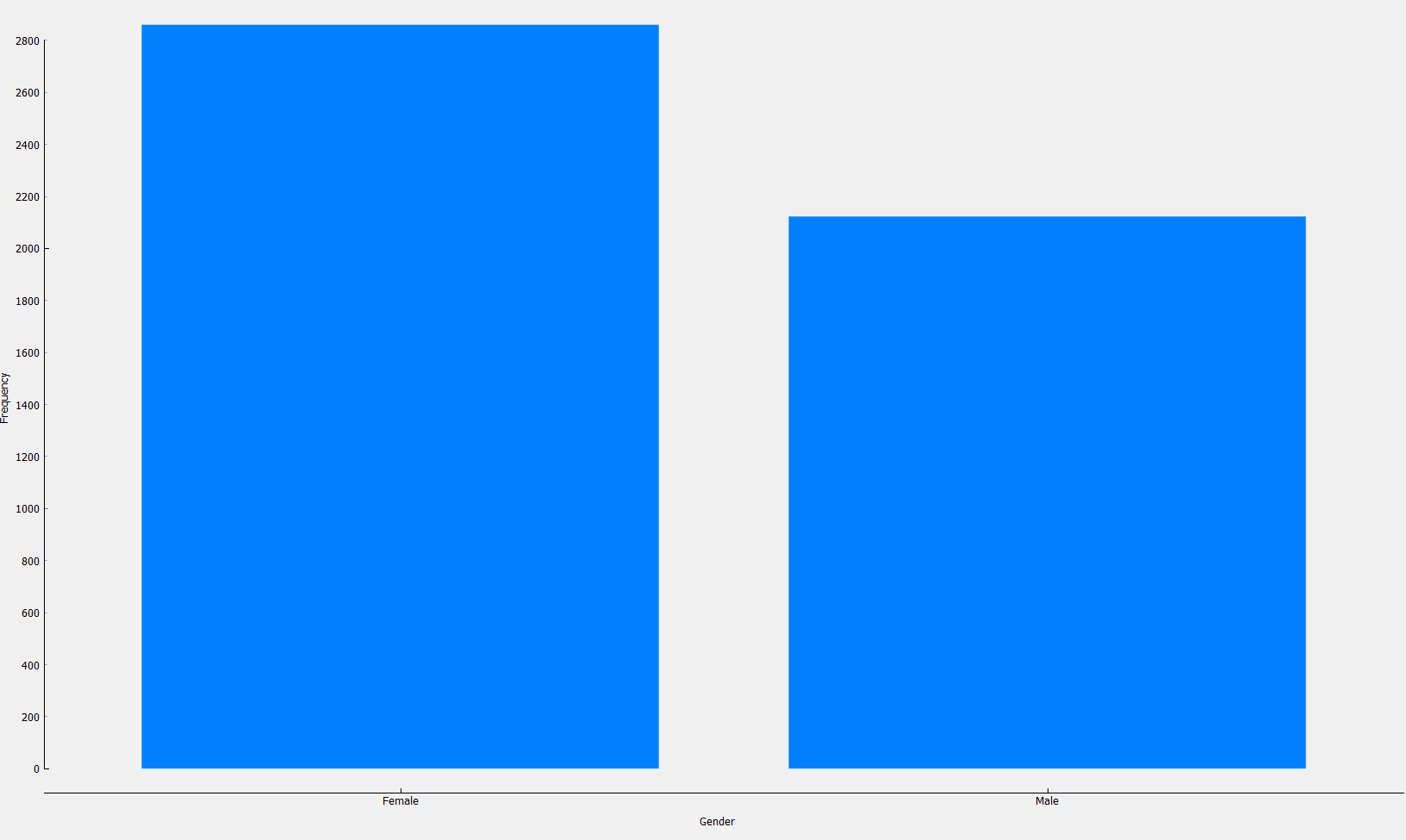
Age



Passenger age is also a fundamental aspect in understanding the data. Adults and Seniors (Ages 37 and above) make up the majority (67.12%) of airline travelers.

---------------------------------------------------------------------------------------------------------------------

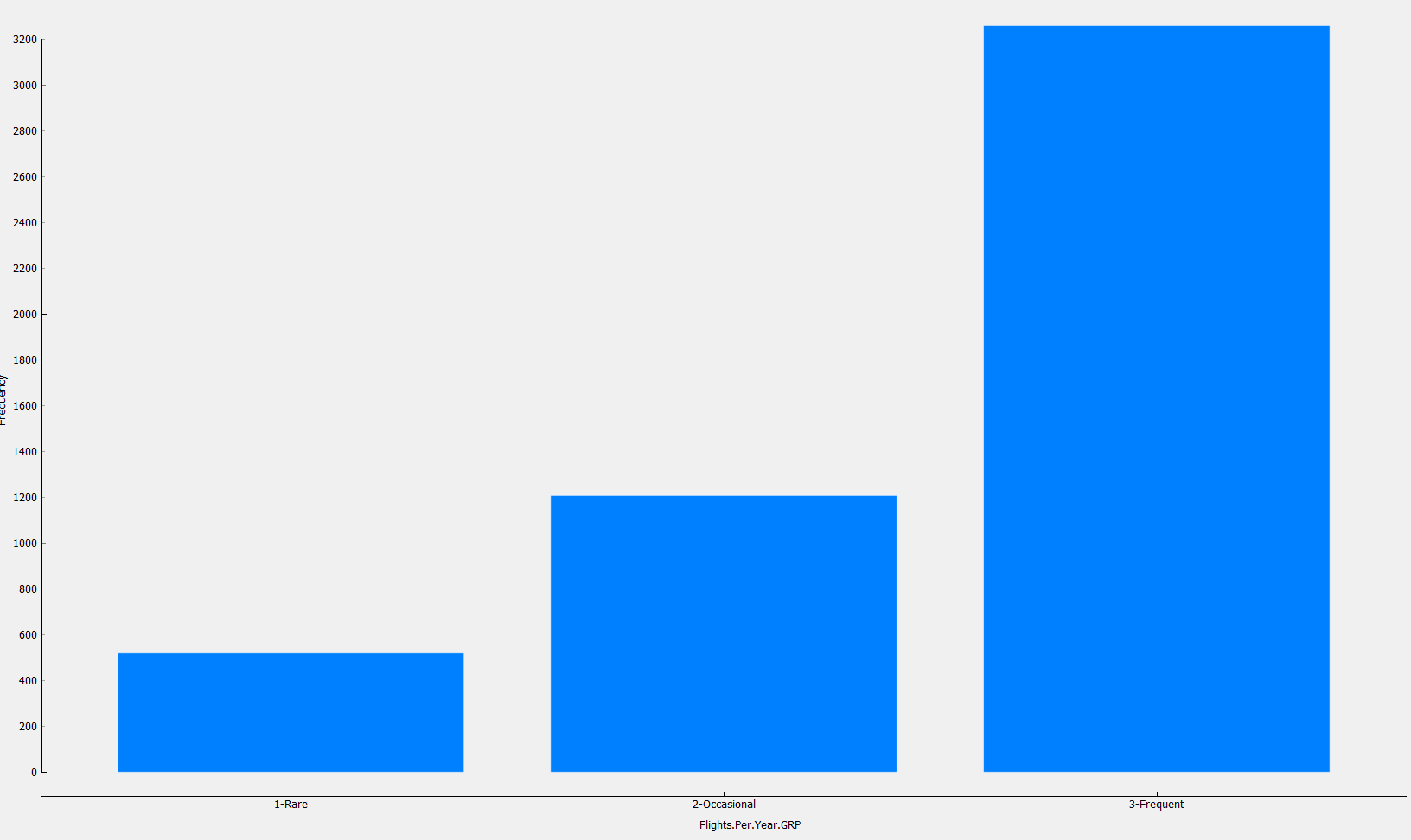
Gender



With some correlation to their representation on general population, female travelers make up the majority (57.39%) of airline travelers.

---------------------------------------------------------------------------------------------------------------------

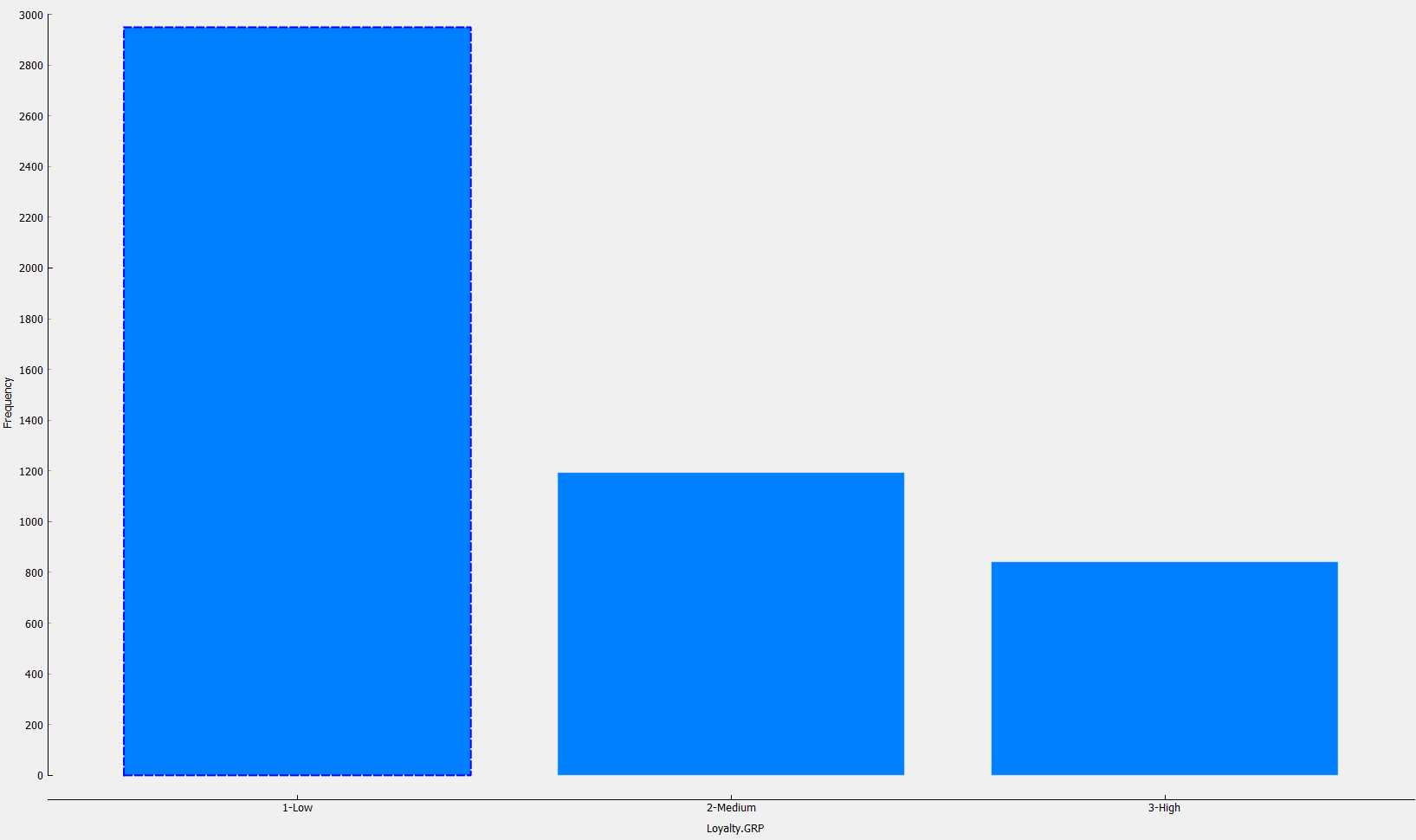
Flights Per Year



A large majority of travelers (65.38%) fly at least once a month on average. (Frequent flyer = 13 or more flights per year).

---------------------------------------------------------------------------------------------------------------------

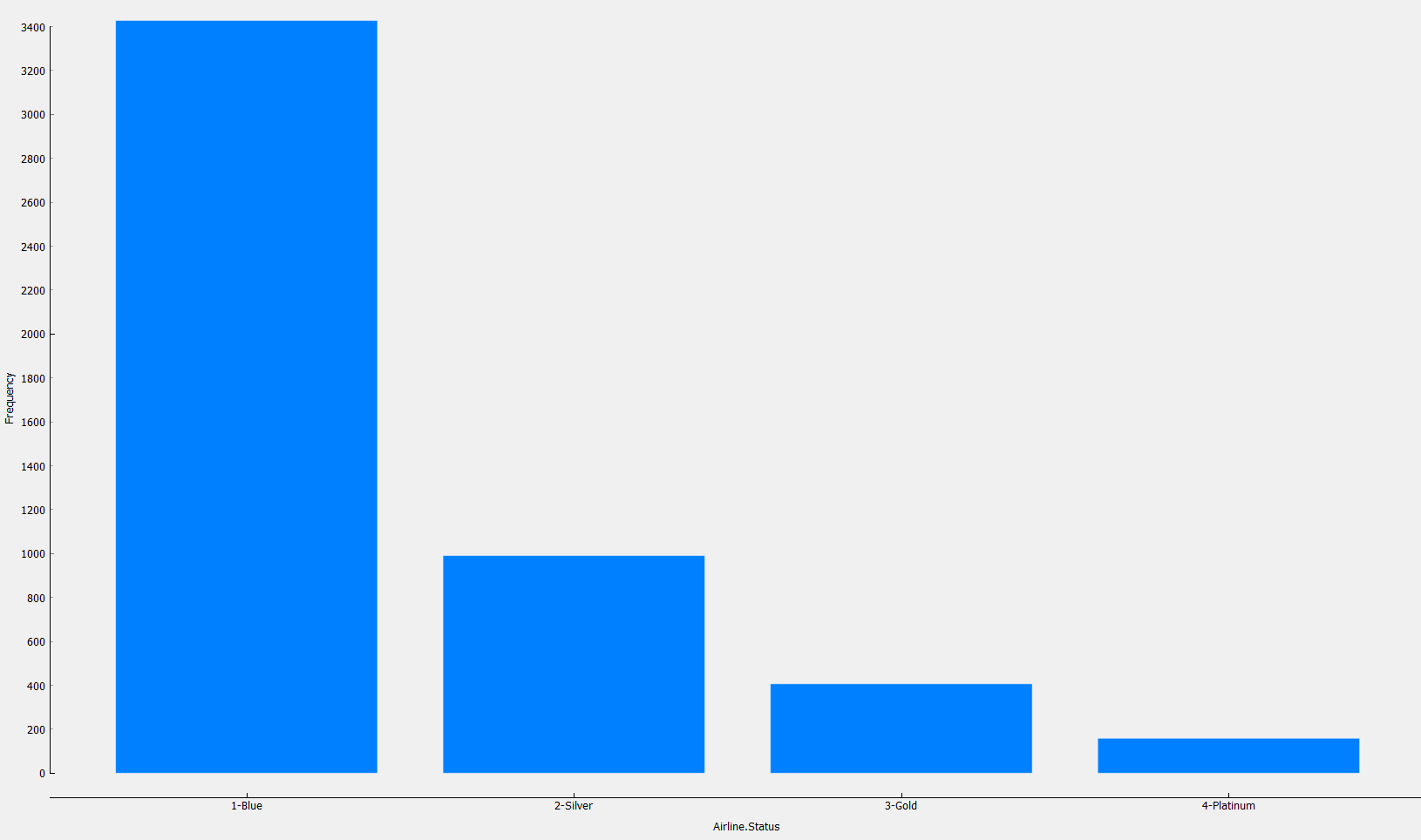
Loyalty



A cross section of air travel loyalty reflects that most (59.16%) are categorized as Low loyalty.

---------------------------------------------------------------------------------------------------------------------

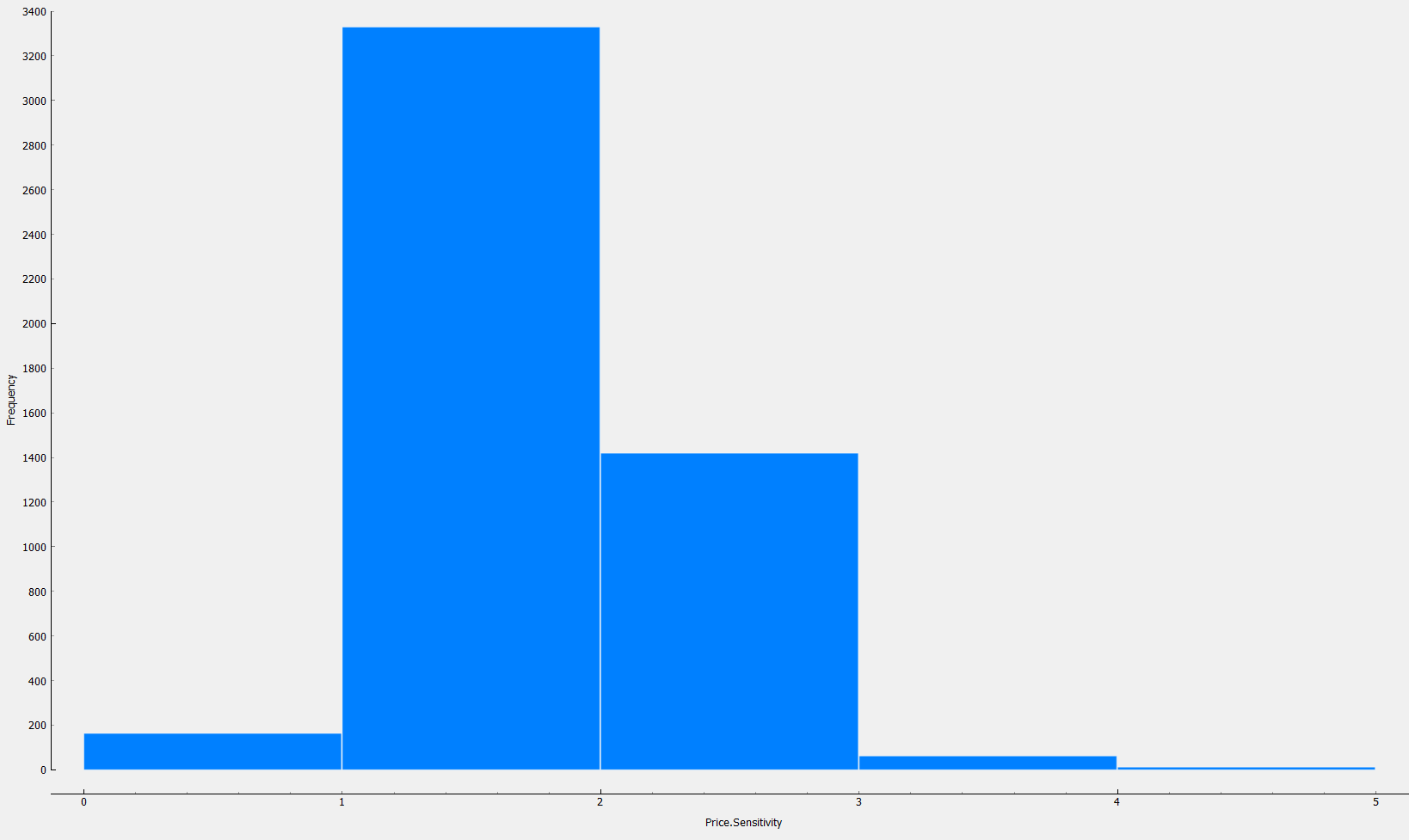
Airline Status



Most travelers (68.77%) appear to be reluctant to become a member of an exclusive upgrading program as they choose to fly with the basic ‘Blue’ status.

---------------------------------------------------------------------------------------------------------------------

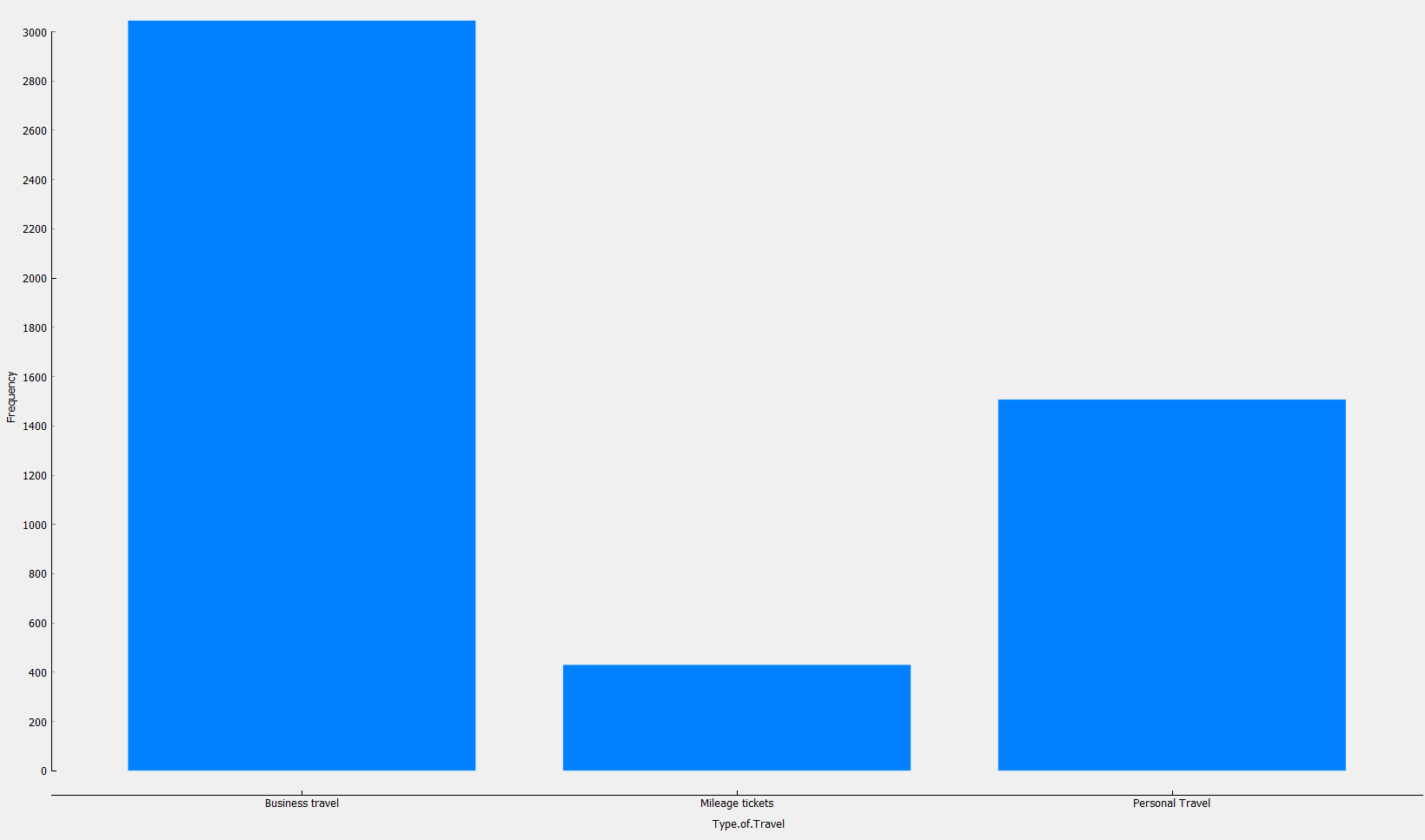
Price Sensitivity



Price sensitivity reflects the grade to which the price impacts customer purchasing behavior. The price sensitivity has a range from 0 to 5, with 5 being the most price sensitive. We observed that the count of price sensitivity is the highest at 1, with a moderate count at 2 and very little at 0, 3, 4 and 5.

---------------------------------------------------------------------------------------------------------------------

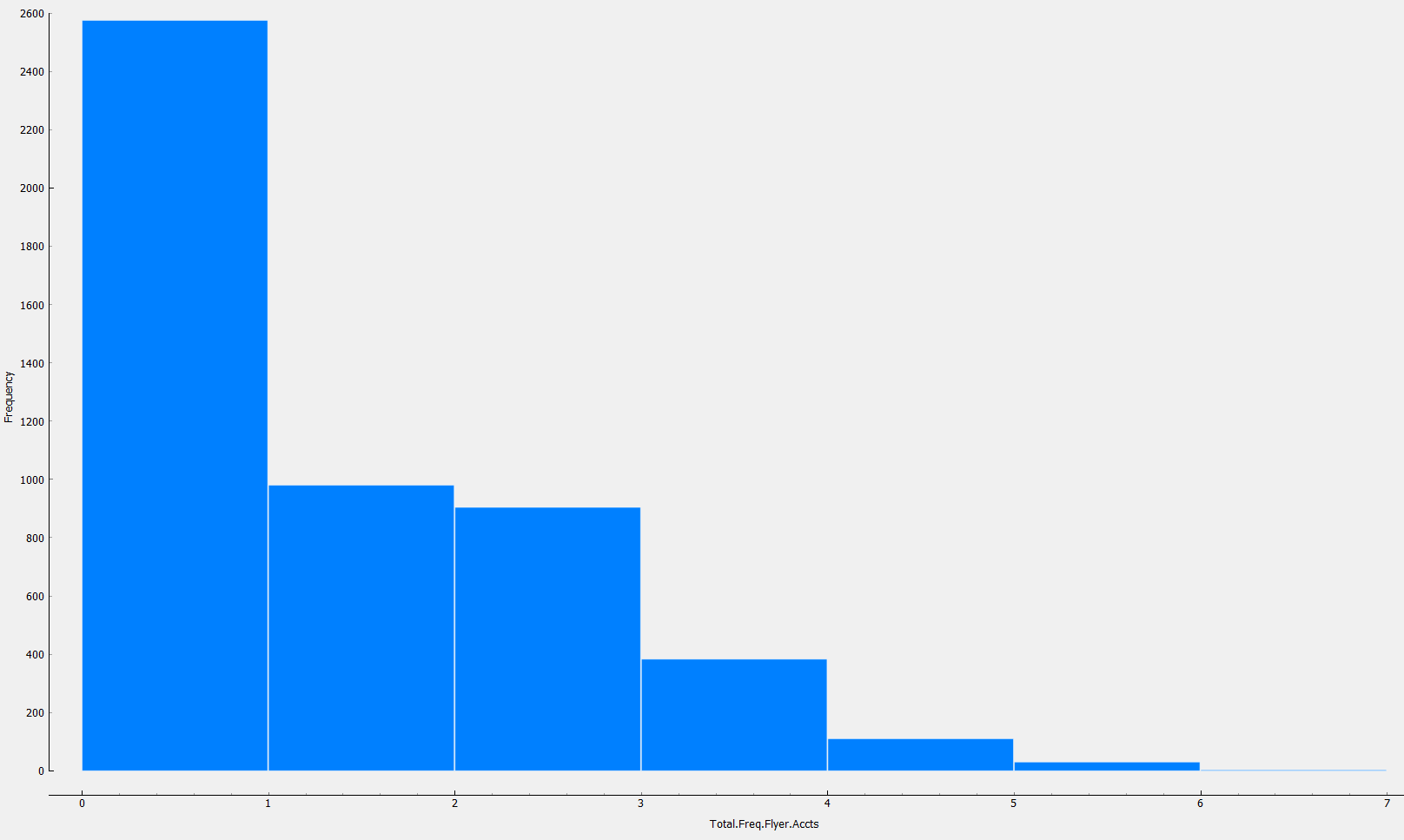
Type of Travel



Most of the travelers are classified as Business, followed next by Personal.

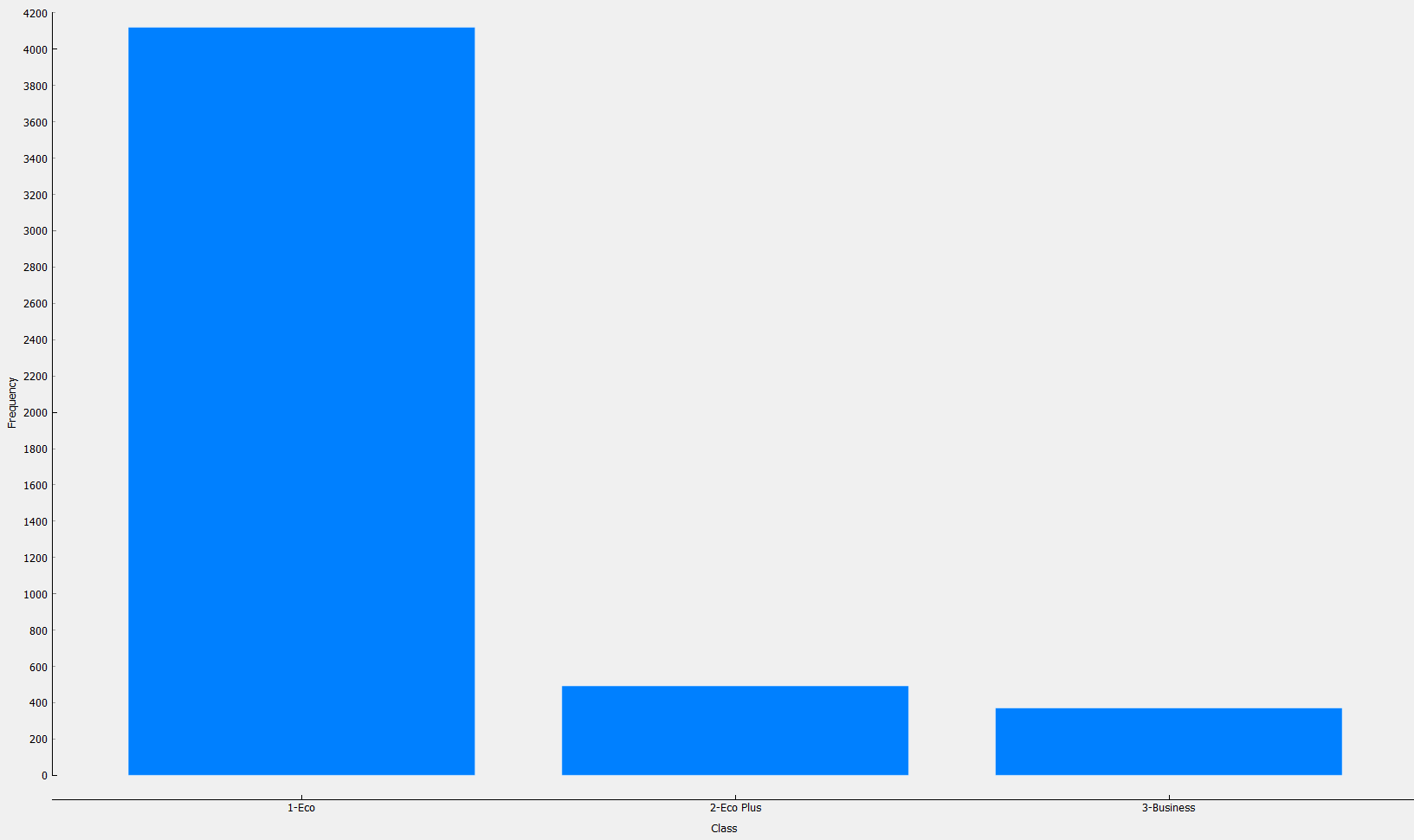
---------------------------------------------------------------------------------------------------------------------

Total Frequent Flyer Accounts



Most travelers do not have any Frequent Flyer Accounts.

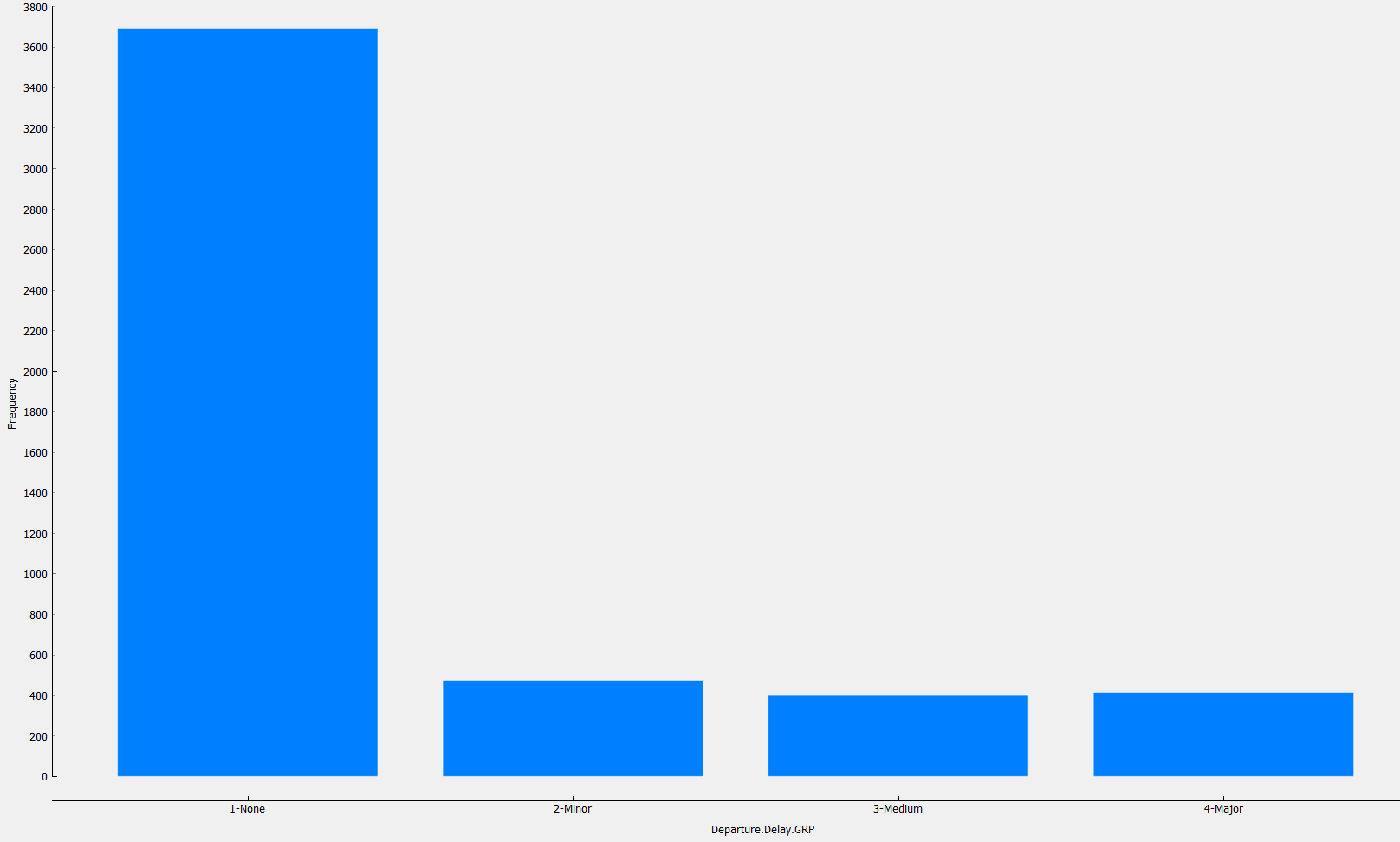
Class



Almost all travelers fly Eco Class.

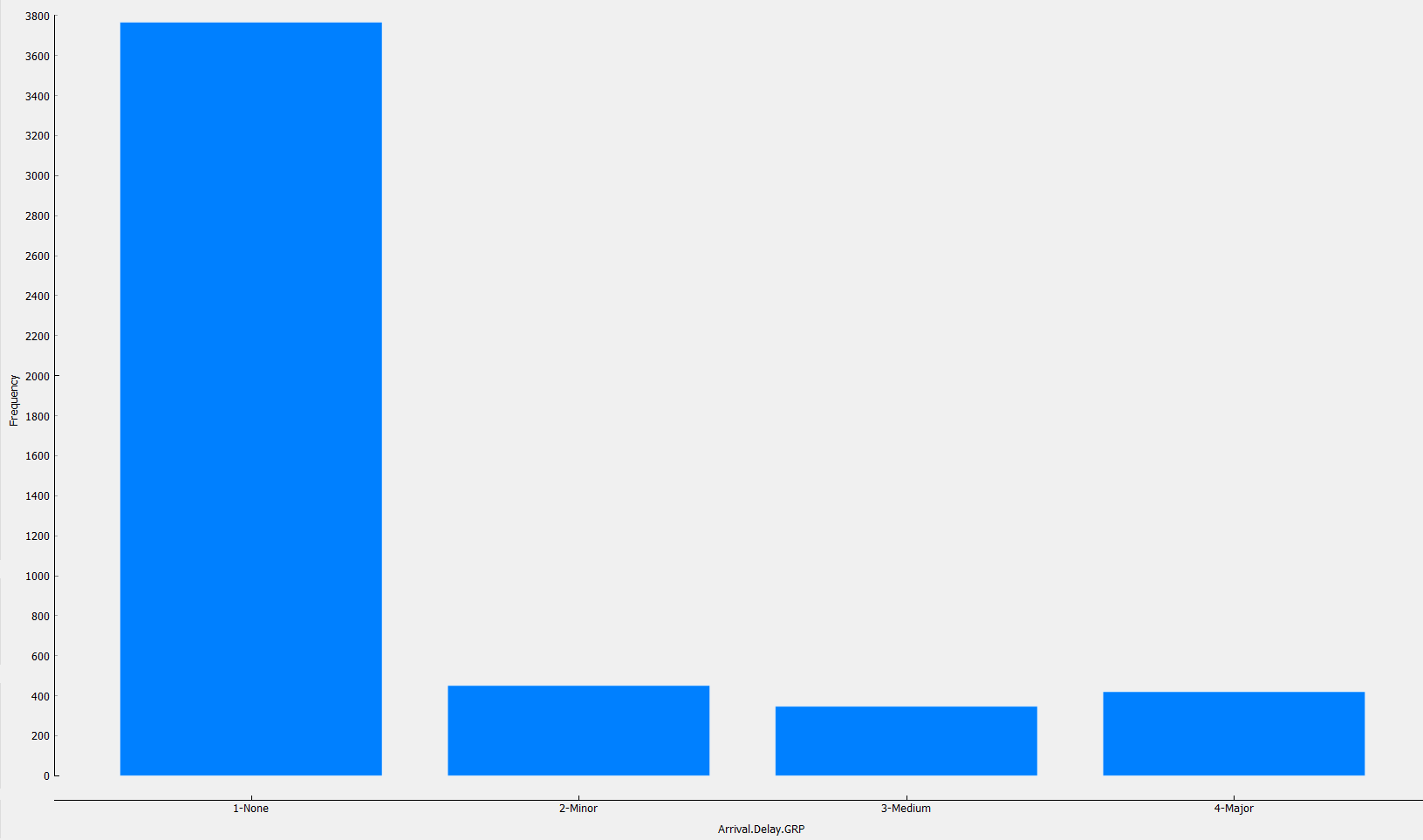
---------------------------------------------------------------------------------------------------------------------

Departure Delay



Most travelers had no delay in their flight’s departure.

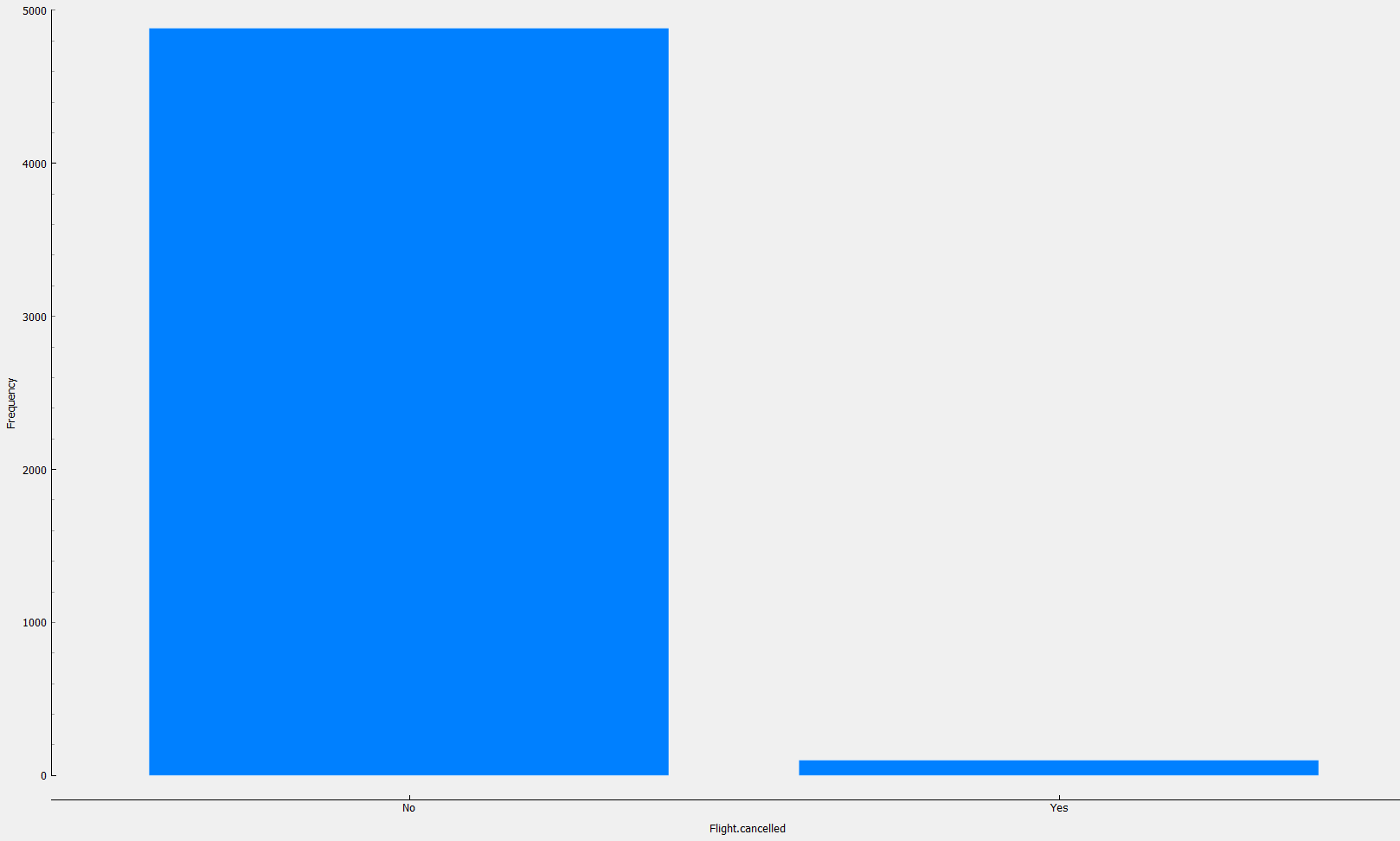
Arrival Delay



Most travelers had no delay in their flight’s arrival.

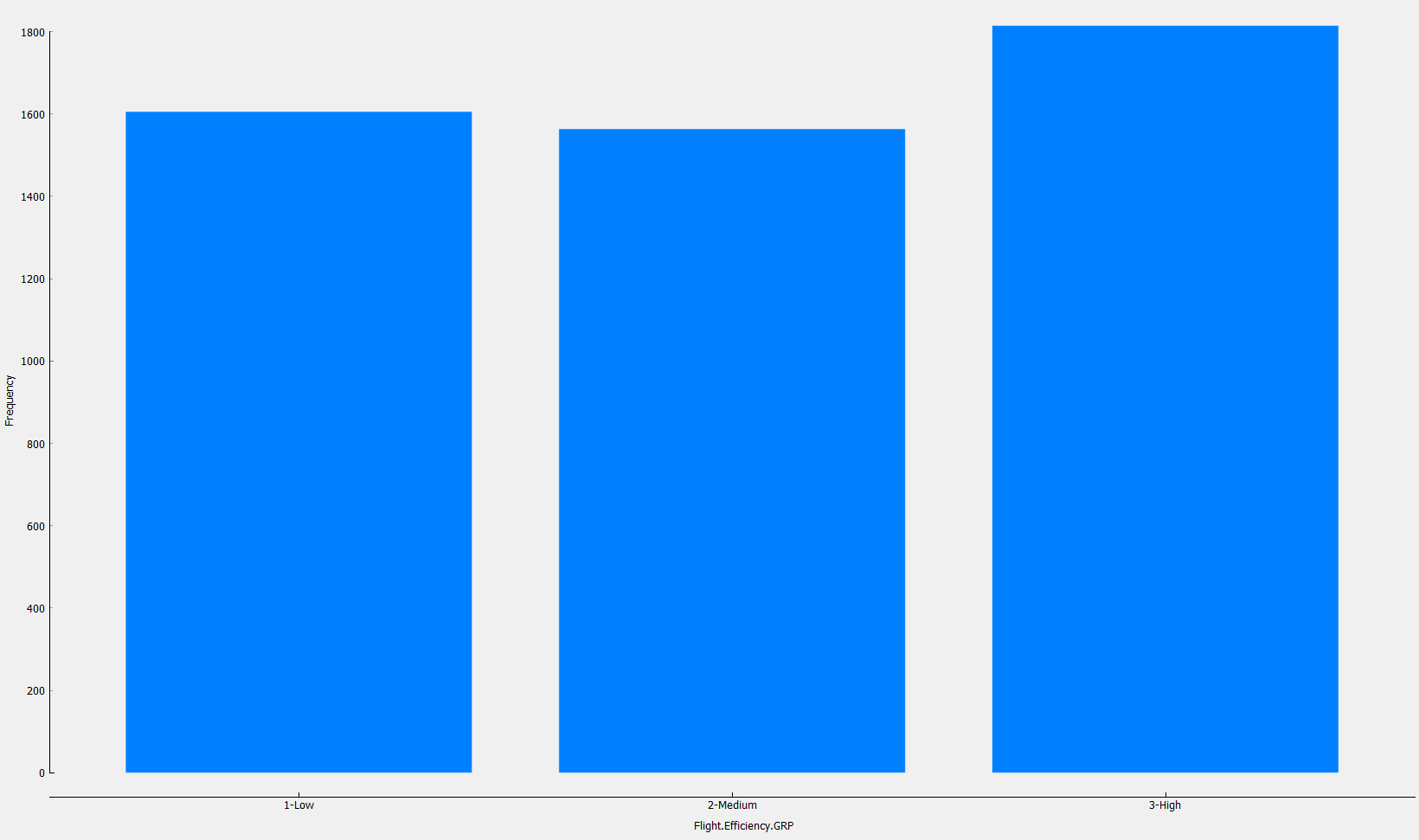
---------------------------------------------------------------------------------------------------------------------

Flight Cancelled



Most travelers had no cancellation of their flight.

Flight Efficiency (Flight distance / Flight time)



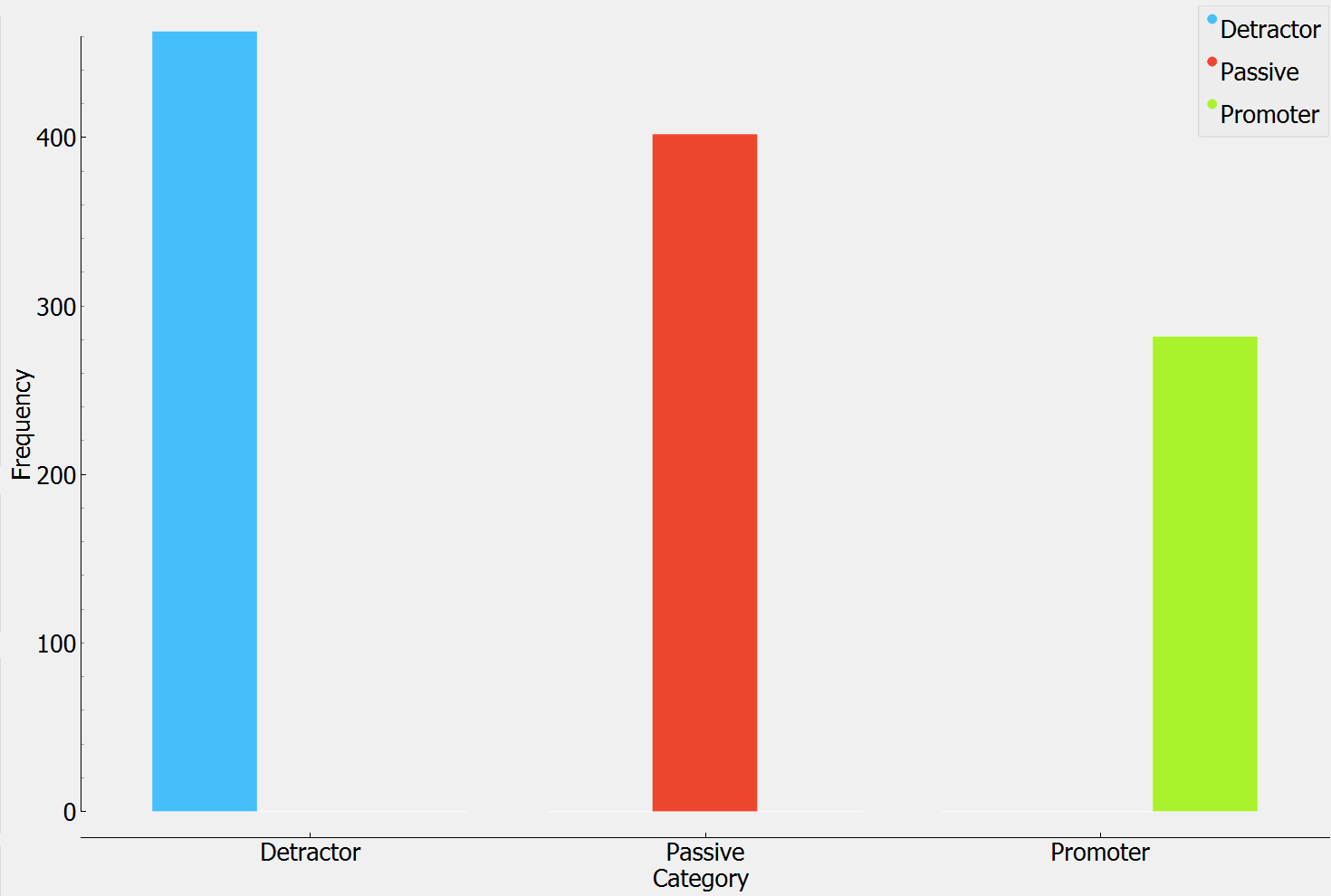
The average flight distance covered over the flight time splits quite equally.

**Analysis and Modeling**

**Market Segmentation**

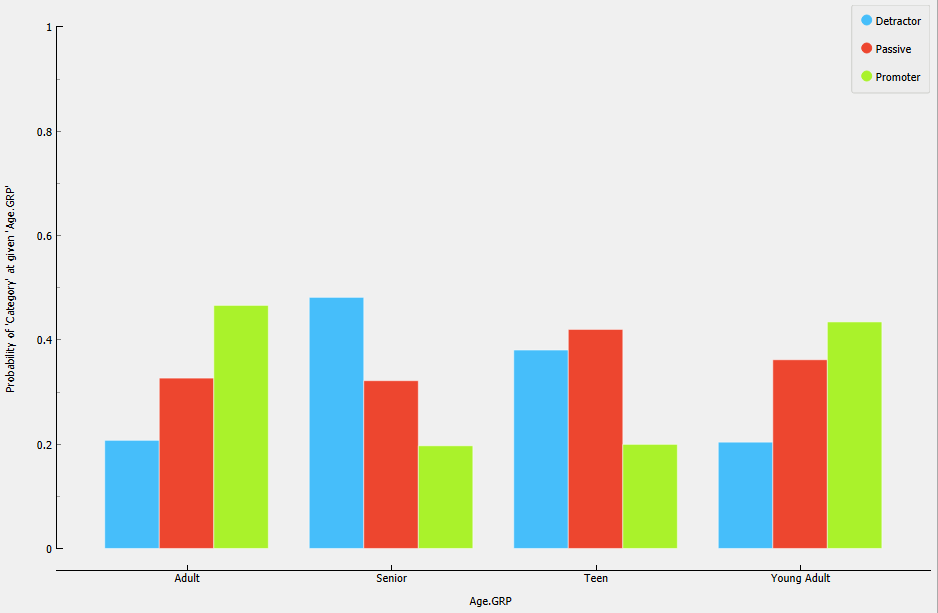
The following section takes the analysis beyond descriptive statistics and provides a more comprehensive understanding of how our most salient features relate to NPS. This analysis improves clarity on market segments by examining underlying patters and relationships in the data with respect to NPS status.

As noted early in the descriptive statistics section, generally, the airline traveler market is split evenly across all NPS status with Promoter taking the majority stake at 35.73%, Passives at 34.24%, and Detractors at 30.03%. All things being equal, this suggests that the expected proportion of FlyFast customers should reflect similar proportions. However, as seen in the graph below, the NPS breakdown for FlyFast is inconsistent with general market proportions. Promoters constitute only 24.59% of FlyFast customers, while passives constitute 35.05%, and Detractors make up the majority at 40.37%. This finding indicates that FlyFast is underperforming when compared to the broader market.



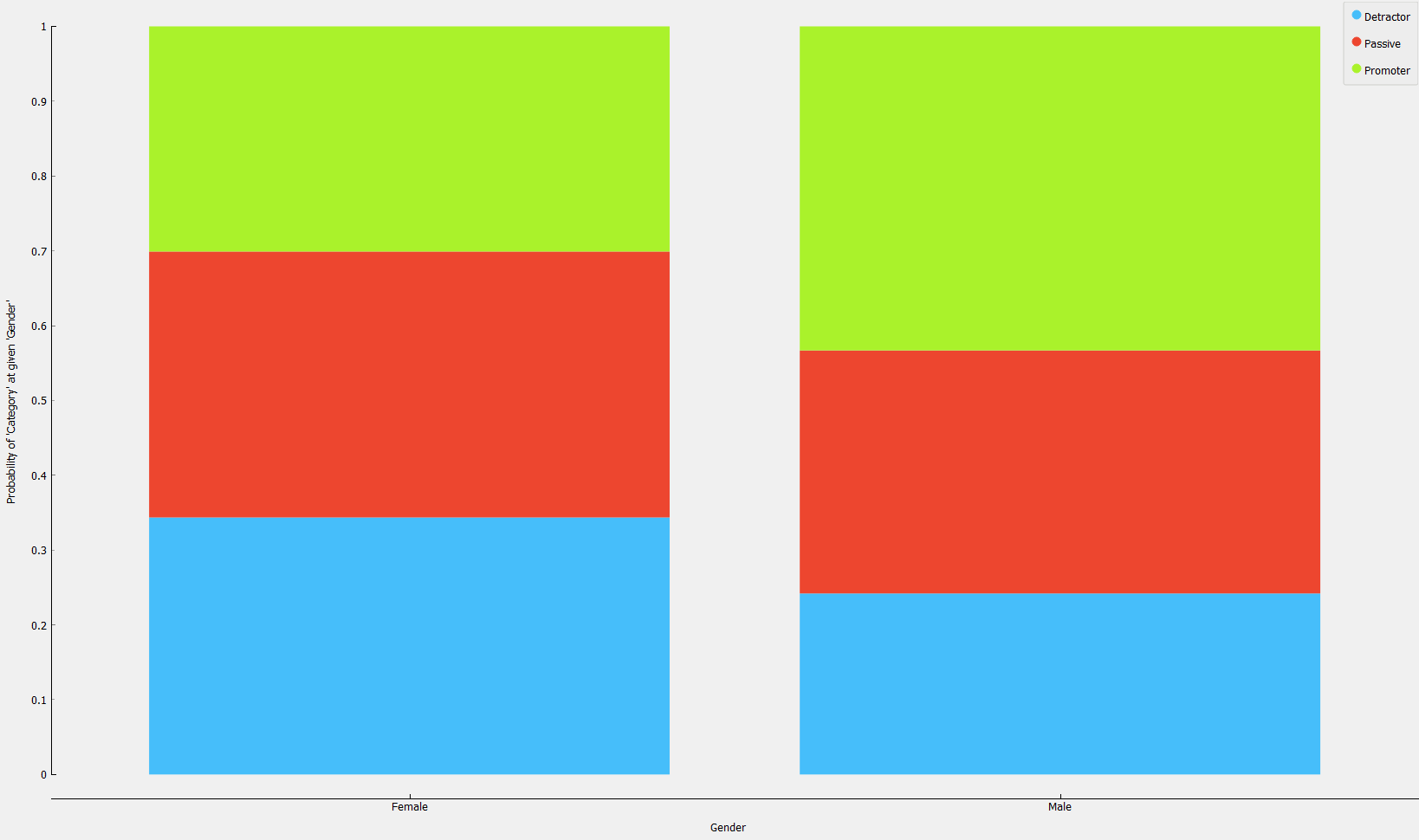
---------------------------------------------------------------------------------------------------------------------

When reviewing NPS across age, we see that Promoters carry the lead with Adults and young Adults; Passives are found leading primarily in Teen travelers; and Detractors carry the lead among Senior travelers which account for 45% total Detractors. Within seniors, Detractors make up 48.15% of this market segment. Given Seniors make up 28.12% of the total traveler market, additional research should be conducted to determine why these travelers are unsatisfied.



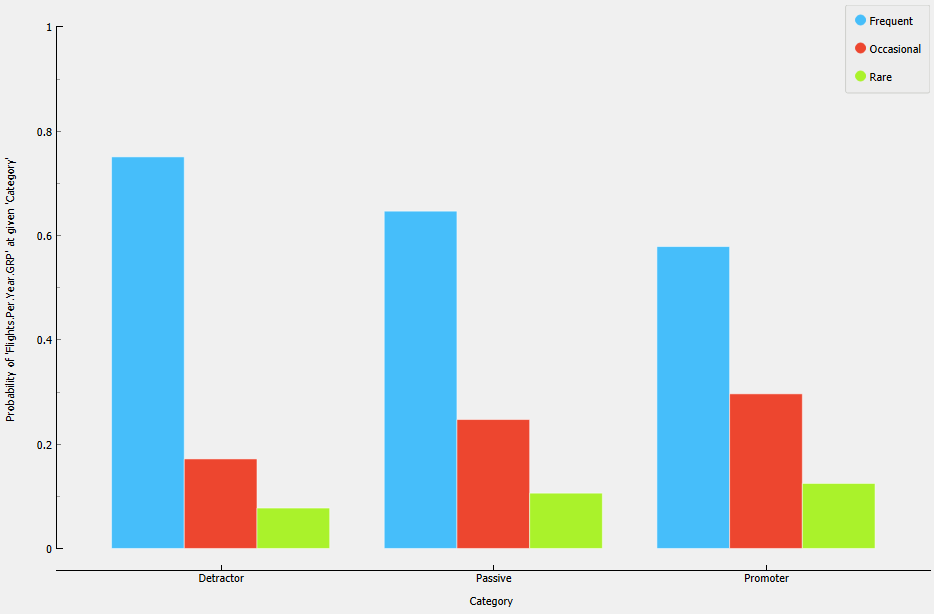
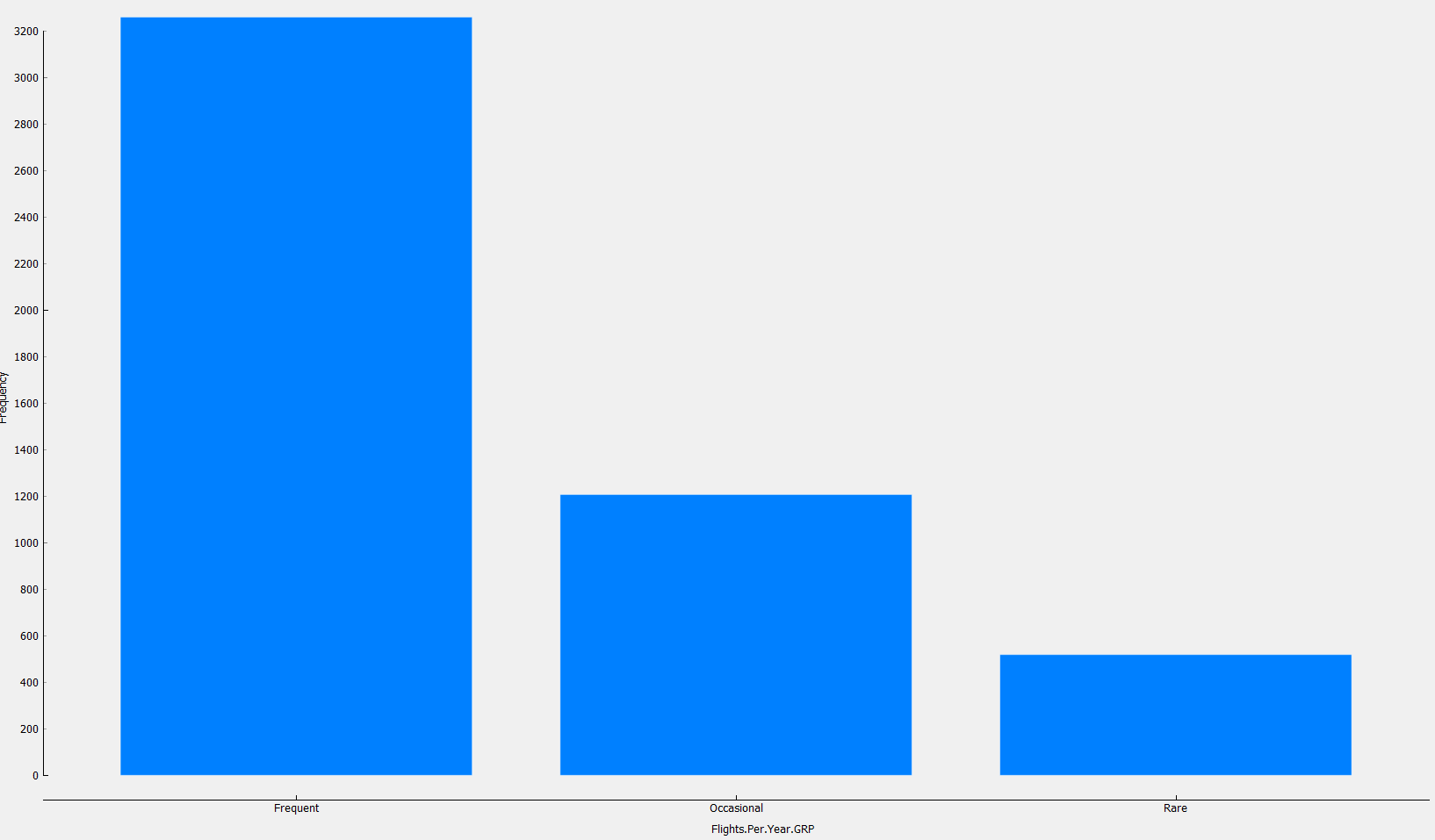
---------------------------------------------------------------------------------------------------------------------

When NPS status is viewed by Gender we see the biggest difference being that Males constitute a larger percentage of Promoters than Females, and conversely, Female travelers have a higher proportion of Detractors. This finding, coupled with the fact that females constitute 56.3% of travelers, demonstrates that additional marketing effort should focus on Female travelers, specifically those Adult and Senior women as they make up 67.8% of the Female market. The chart below highlights the breakdown of NPS status across Gender.



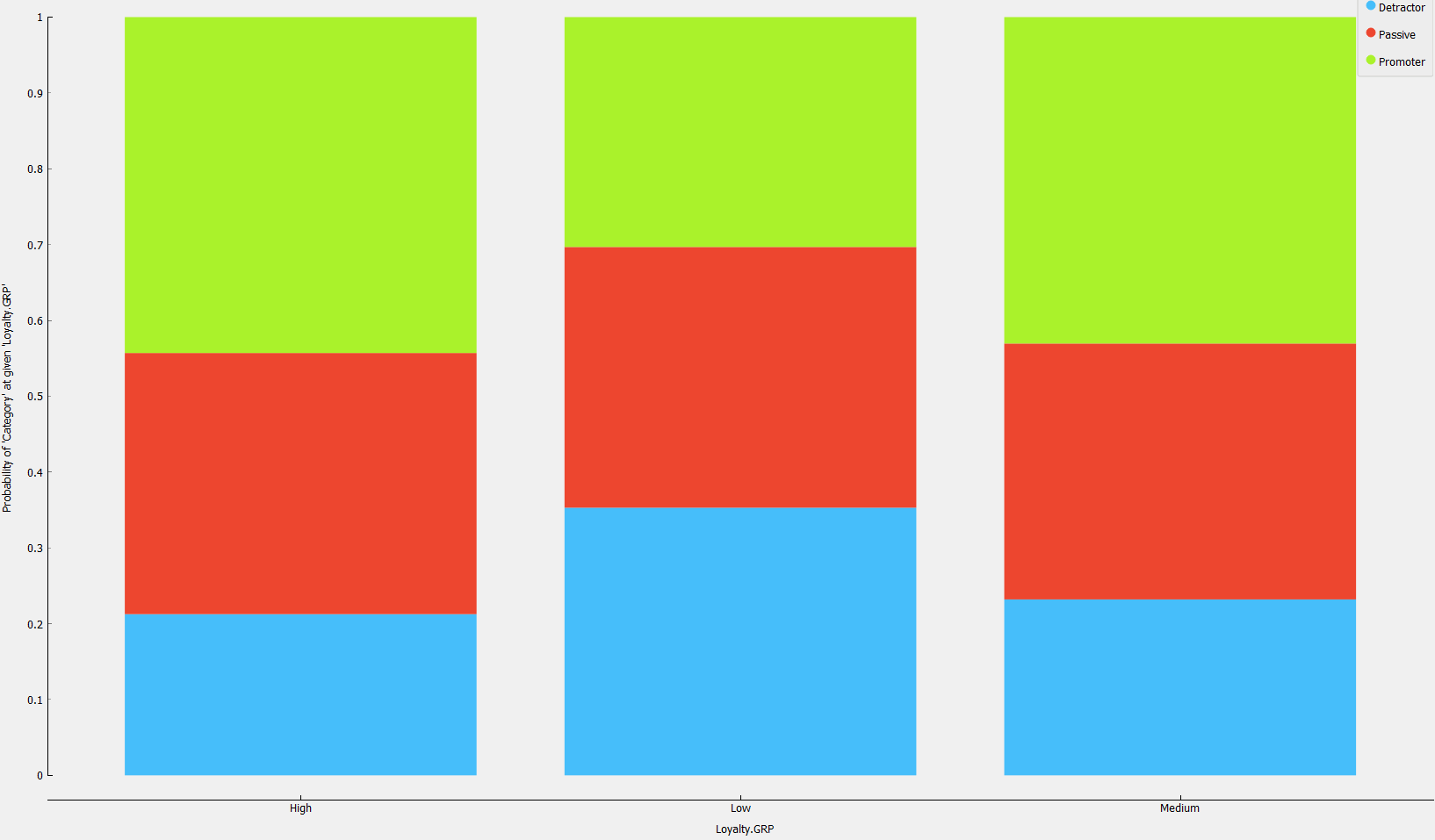
---------------------------------------------------------------------------------------------------------------------

When viewed by flights per year, we see that there is no significant relationship between the frequency a traveler flies and their (its?) NPS status. This becomes even more apparent when compared with the distribution of flights per year more broadly as seen in the image to the lower right. NPS statuses of Detractor, Passive and Promoter are somewhat equally proportional to the distribution of flight per year suggesting there is no significant relationship between how often a traveler flies and their (its?) NPS status.

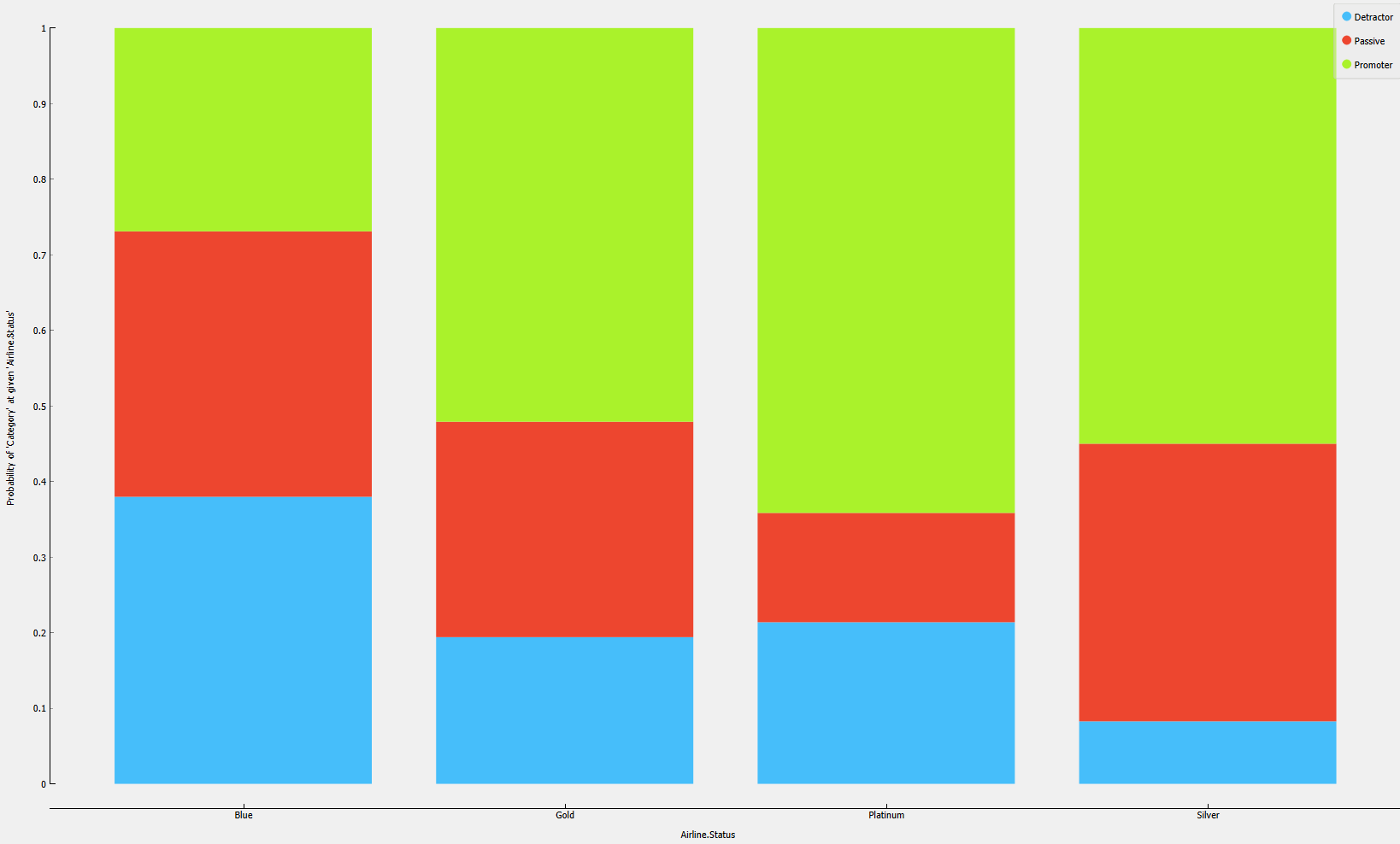
---------------------------------------------------------------------------------------------------------------------

When NPS is viewed in the context of loyalty group we can see that Detractors have a lower proportional percentage of Highly loyal travelers, and conversely, a higher proportional percentage of Low loyalty travelers. Proportionality we see that loyalty increases as you move from Detractor to Promoter, and conversely, low loyalty decreases as you move from Detractor to Promoter. This finding highlights that there is a positive correlation (~0.32) relationship between customer satisfaction and loyalty.



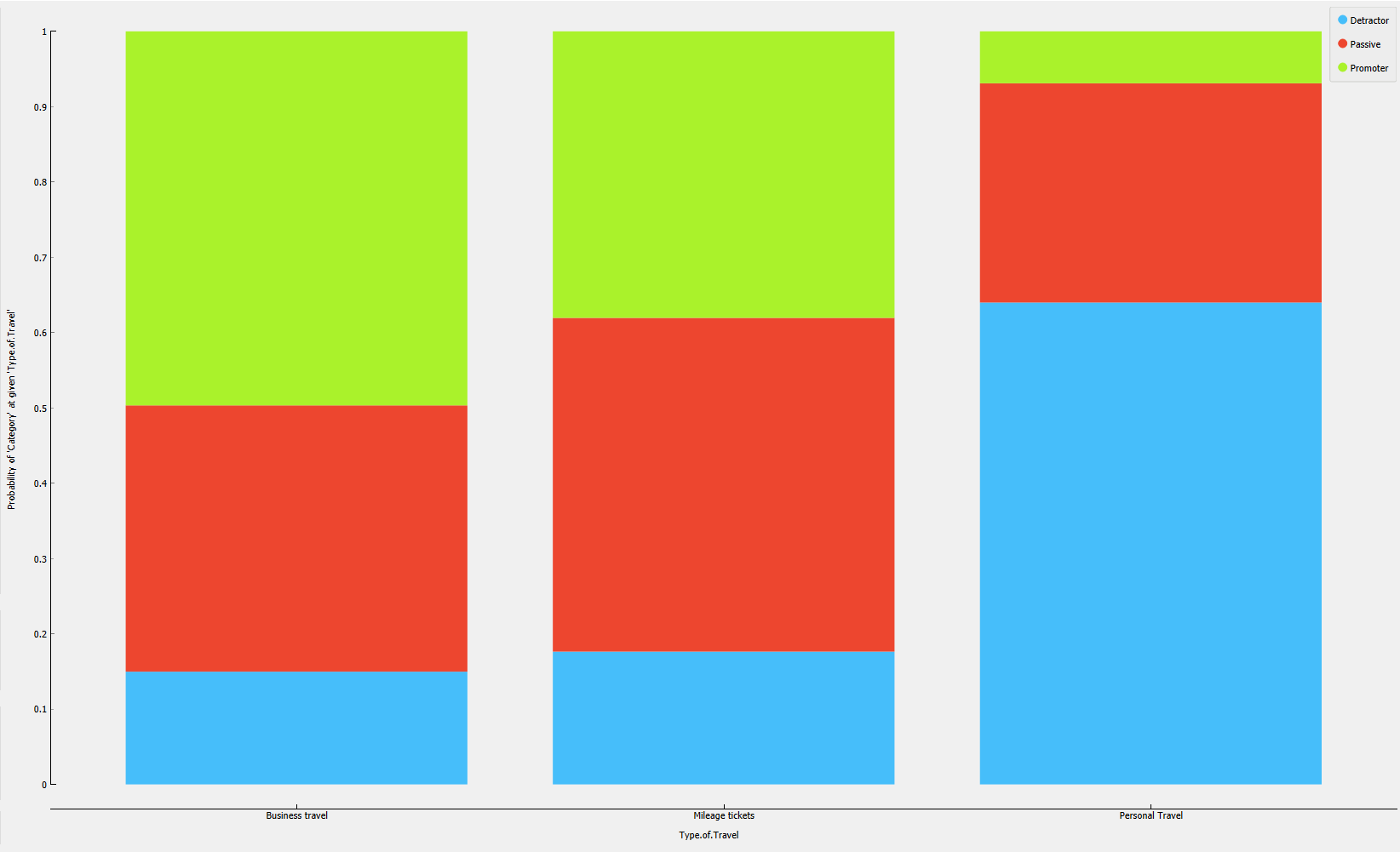
---------------------------------------------------------------------------------------------------------------------

With regard to Airline Status we see a trend that as airline status increase, generally so does the Promoter NPS status, and vice versa. (Note: the table below is not in the correct order, status level goes from Blue, Silver, Gold, Platinum). For example, most of the Blue status is made up of Detractors. This suggest entry-level airline status flyers are not overly impressed with their level of service. On the other end of the spectrum, it shows that customers with a Platinum status, are overwhelmingly Promoters. And perhaps most interesting, is that Silver—a middle of the pack status level--has very few Detractors. This analysis suggests an audit of benefits associated with each level of status to identify potential opportunities to adapt features to maximize NPS outcomes.



---------------------------------------------------------------------------------------------------------------------

With regard to type of travel, business class travelers disproportionately make-up the Promoter customers. Conversely, customers conducting personal travel are the largest population of Detractors. Further analysis should be done to review the policies and benefits surrounding these traveling categories with a focus on making personal class travel more consistent with business class.



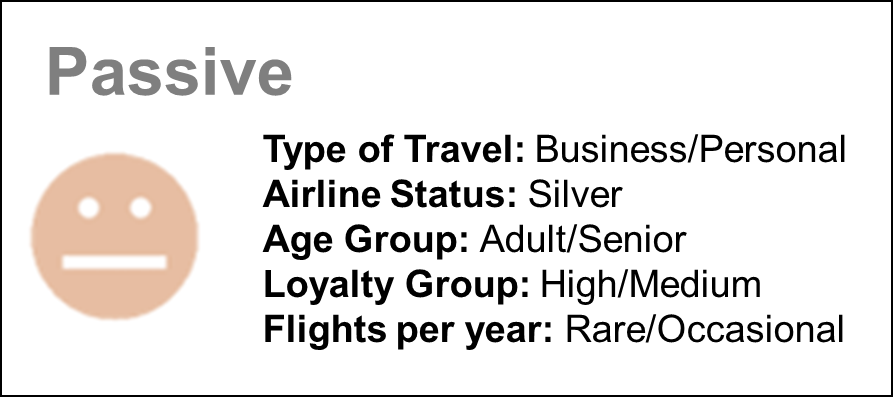
**Traveler Personas**

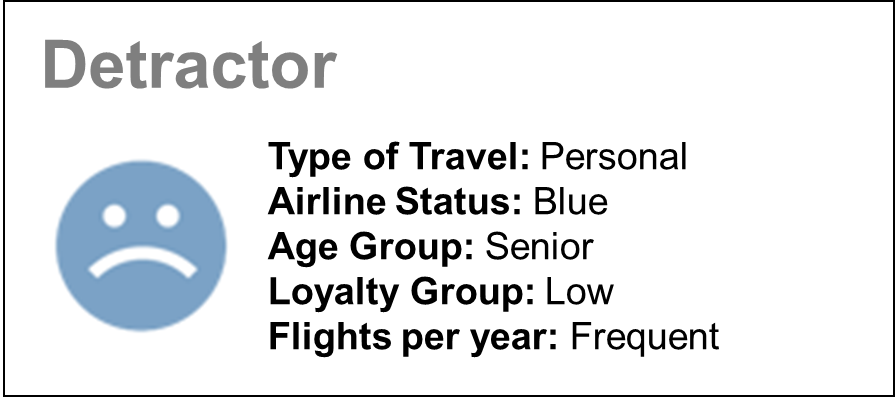
Based on our classification tree analysis (ANNEX 3), we determined the top 5 features in terms of information gain that determines where a traveler falls with respect to promoter status are determined most heavily by the following features:

* Type of Travel
* Airline Status
* Age Group
* Loyalty Group
* Flights per year

Based on the feature above, we have created the following personas that provide details for each type of travel in reference to the above features. The features in each persona are not made up exclusively by the values presented, but the values do account for an overwhelming majority in most cases. If two values are listed, for example Loyalty Group: High/Medium, that indicated there was no single majority, but rather a shared majority. The three personals for NPS status are below.







Based on the personas above, moving customers out of the Detractor Status requires focusing marketing effort on frequently traveling Seniors. 28.16% of FlyFast customers are Seniors, of which, 17.35% are detractors. FlyFast’s high percentage of Detractors among senior travelers is the largest departure from what is seen in industry trends from competing airlines. Improving this metric to be more closely aligned with industry standards could yield approximately a 5% reduction in Detractor status.

**Sentiment Analysis**

The most powerful text analysis that could be performed is sentiment analysis of the reviews contained in the freeText data field. Only 2.02% of the data contained comments but analyzing what was provided help provide additional context to the analysis as it is a great way to gain a sense of a passenger’s experience of a specific flight and the airline. Analyzing this data may capture aspects hidden in traditional statistics.

Focusing on the difference between FlyFast and Sigma in terms of Promoters and Detractors proportions, a word-cloud comparison was made between the two airlines’ passengers. Due to the low number of surveys filled in, this analysis cannot show the entire emotional spectrum. However, there is a clear difference between these two airlines in terms of volumes of positivity:

FlyFast’s word cloud seen below primarily consists of ordinary or neutral terms without that provide little indication of positive or negative sentiment.



Conversely, Sigma’s word cloud below contains some very positive responses (“Excellent”, “Great”, “Good”) by their customers.



Three key terms appear in both airlines: Service, Delayed, and Time. This suggests that Sigma’s key to success does not necessarily mean avoiding flight delays, but may result from keeping those who experience a delay well satisfied, entertained and compensated.

**Modeling**

Five well-known data mining classification models were used in order to find the best predictor of a given traveler’s NPS status (Promoter, Passive or Detractor). This model will provide FlyFast better understand air traveler’s and most importantly, allow them to predict a traveler’s NPS status for new data FlyFast receives. This analysis modeled the data using the five techniques listed below, each using a validation split of 70% train, 30% test split.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Running Time (s)** | **AUC** | **CA** | **F1** | **Precision** | **Recall** |
| **Neural Network** | **0.013** | **0.994** | **0.930** | **0.930** | **0.930** | **0.930** |
| SVM | 0.111 | 0.909 | 0.745 | 0.740 | 0.791 | 0.745 |
| Decision Tree | 0.000 | 0.764 | 0.584 | 0.588 | 0.606 | 0.584 |
| kNN | 0.209 | 0.764 | 0.572 | 0.563 | 0.579 | 0.572 |
| Naïve Bayes | 0.001 | 0.744 | 0.554 | 0.555 | 0.555 | 0.554 |

Since every algorithm has varying strengths and weaknesses, determining the best model requires understanding the methods used to evaluate model performance. This is important not just to get the most out of each model, but to know how to select the correct model for the analysis task. The best measure of a model’s performance is the one that best captures whether the model is successful at its intended purpose, specifically, how well it will extrapolate to future cases. Given this, the analysis considered the five evaluation measures in the table above. A review of each model’s quality measurement shows that the Neural Network model provides the best results. While the other models are averaging about 60-70% accuracy (percent predicted right/percent predicted wrong) in properly predicting a traveler’s NPS status, the Neural Network model is performing at about 93% accuracy. Additionally, processing a prediction on a dataset of this size with the existing features takes only a fraction of a second which makes deploying and hosting the model across the organization easier.

Another useful tool is in model evaluation is the confusion matrix, which is a table that categorizes predictions made by the classifier according to whether they matched the actual value recorded in the test data. Confusion matrices are extremely helpful in providing the analyst details in the way prediction errors were made, normally referred to as Type I and Type II errors, or False Positive and False Negative errors. The performance of the Neural Network model is detailed in the in the confusion matrix below (proportions of actual cases is displayed):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Predicted** | | |
| Detractor | Passive | Promoter |
| **Actual** | Detractor | 95.9 % | 2.2 % | 1.9 % |
| Passive | 5.2 % | 92.0 % | 2.7 % |
| Promoter | 2.8 % | 7.4 % | 89.7 % |

For each one of the NPS categories, the model predicts the correct outcome for almost 96% of actual Detractors, 92% of Passives and nearly 90% of Promoters. Based on the differences in classification performance seen above, generally we can conclude that it is easier to model travelers classified as Detractors, but more challenging to determine the difference in whether a traveler is a Passive or a Promoter.

Although not very pronounced, this helps with targeted marketing efforts. A worst-case scenario in a resource constrained marketing environment is to spend marketing dollars on a segment of customer that does not require conversion, e.g. a Promoter. An inability to clearly see who is a Promoter risks this outcome. As such, the smart choice would be to spend marketing dollars on a segment that is clearly defined and distinct from the other segments (assuming it requires conversion). This approach allows marketers to more accurately assess the results of their marketing efforts and helps them make precise adjustments to meet their objectives. Based on the contingency matrix above, we can see that it is much easier to “see” what travelers don’t like (Detractors), than it is to “see” what they sort of like (Passives) or really like (Promoters).

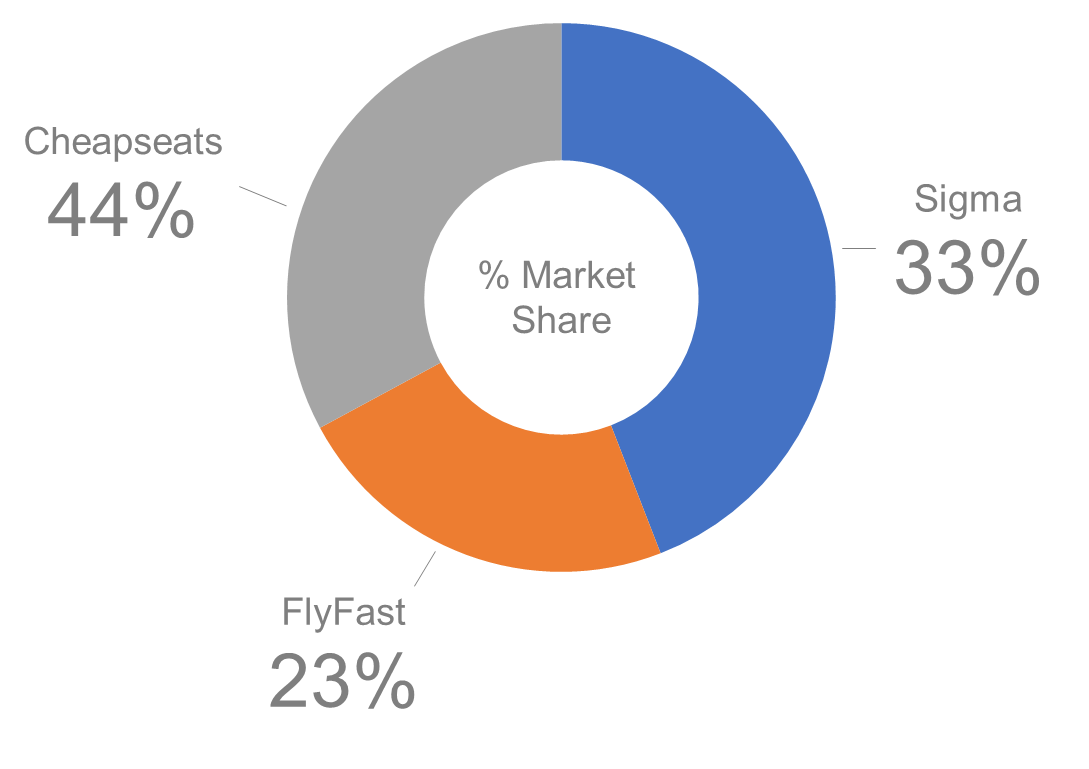
**SECTION 2: MARKETING PLAN**

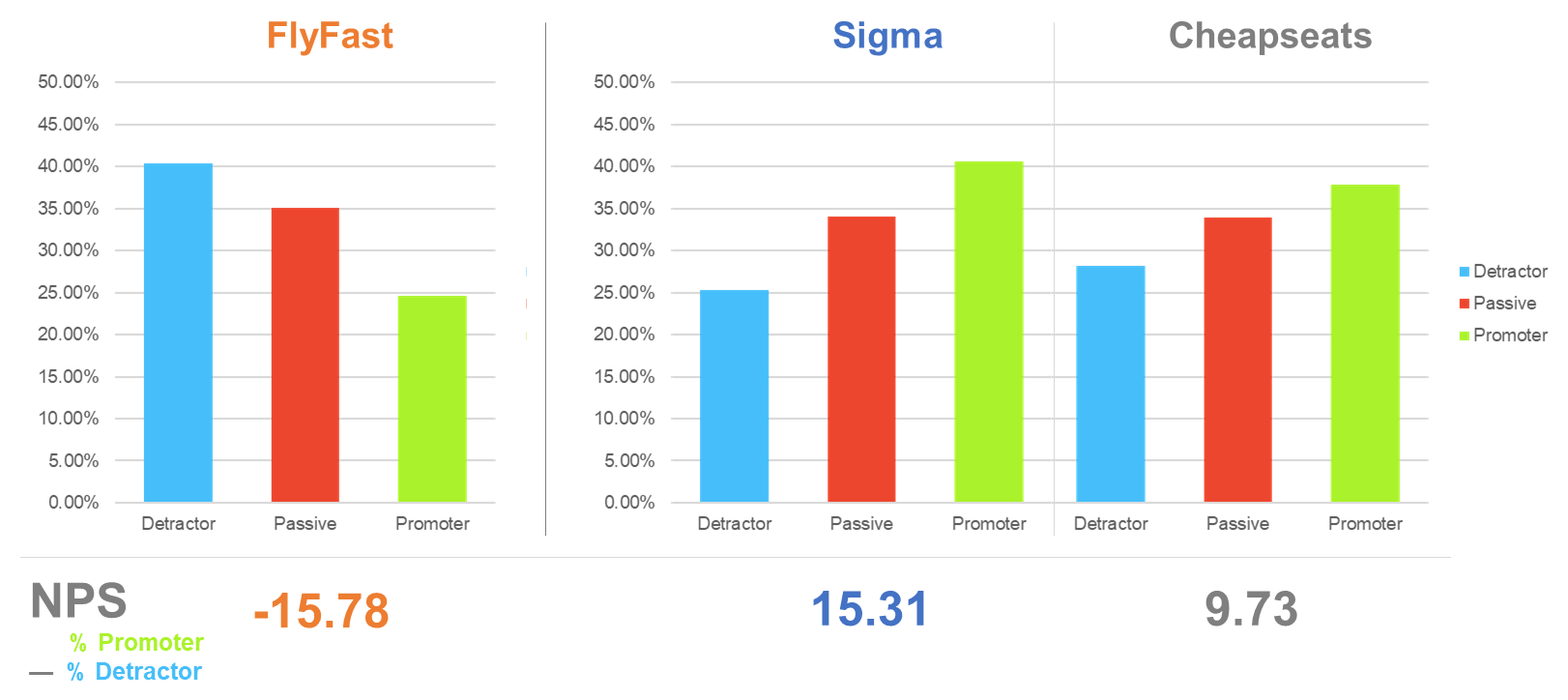
**Business Intelligence: Key Findings from the Analysis**

Building on what was learned in the analysis section, outlined below is an aggregated assessment of current market positions, a summary level Strengths, Weakness, Opportunities and Threats assessments, and five marketing recommendations that will help FlyFast Airways move more customers to a Promoter status.

### **Current Market Positions**

As seen in the graphics below, FlyFast Airways currently has the smallest market share with respect to size when compared to its two nearest competitors. Regarding market share, Cheapseats leads between the three brands. However, as seen in the lower bar charts, in terms of NPS, Sigma has the largest percentage of Promoters, and smallest percentage of Detractors (both as a percentage of their total flights).





As seen above, Sigma airlines is the market leader in terms of NPS with a score of 15.31. As such, Sigma is the benchmark competitor from which we compare and model FlyFast. This analysis looked across all features to determine key differences between these carriers. In terms of Age, Price Sensitivity, Flights Per Year, Type of Travel, Total Number of Frequent Flyer accounts, and Class of Travel, there are very few differences between FlyFast and Sigma.

In terms of, Gender, Loyalty, Departure and Arrival Delays, Cancelled flights, and Flight efficiency, there are significant variations between FlyFast and Sigma that help explain the difference in NPS statuses between the two airlines.

* **Gender:** 60.28% of Sigma’s customers are Female compared to 56.41% for FlyFast
* **Loyalty:** Sigma’s Low Loyalty customers makes up 58% of their market share, while FlyFast’s Low Loyalty travelers make up 65% of their market share.
* **Departure Delays:** Sigma has the highest percent of on-time flights at 82.12%, and lowest percent of Major Delay at 6.77%; while FlyFast comes in second with 73% on-time flights, but has the highest percent of Major Delay at 10.46%. A similar pattern is observed in arrival delays as well.
* **Cancelled flights:** Sigma Airlines leads the pack with only 0.79% of their flights being cancelled, while FlyFast is the worst carrier with 4.97% of its flights being cancelled.
* **Flight efficiency:** Sigma leads the pack with 46.80% categorized as High Efficiency, while FlyFast is the most inefficient with only 22.06% percent of flight categorized as high efficiency.

These key differences, along with current market positions of FlyFast and its competitors, helps identify key strengths, weakness, opportunities, and threats that can be leverage in developing marketing recommendations.

**SWOT Analysis**

Strengths

* As the smallest carrier with respect to market share, a dynamic pivot to refocusing the brand on customer satisfaction (NPS) is less challenging
* FlyFast has the largest share of Passives (as a percent of total travelers)
* FlyFast travelers are less price sensitive

Weakness

* Smallest market share in terms of size
* Worst performing carrier in terms of NPS status
* More major delays than any other carrier (as a percent of total flights)

Opportunities

* Seize the Adult and Senior Female market
* Develop marketing campaign aimed at Detractors since this segment is well defined; which facilitates campaign assessments and tuning
* 57.03% of Cheapseats and 58.51% of Sigma passengers are “Low Loyalty” customers which suggests they can be persuaded with the correct incentives

Threats

* Implications of Corona virus
* Our main competitor Sigma airlines is small enough to be adaptive and replicate our success

**Immediate and Actionable Recommendations to Improve Satisfaction**

Based on the collective understanding provided by the analysis and market positions identified in the report, the following recommendations to increase FlyFast’s NPS Promoter status are offered below:

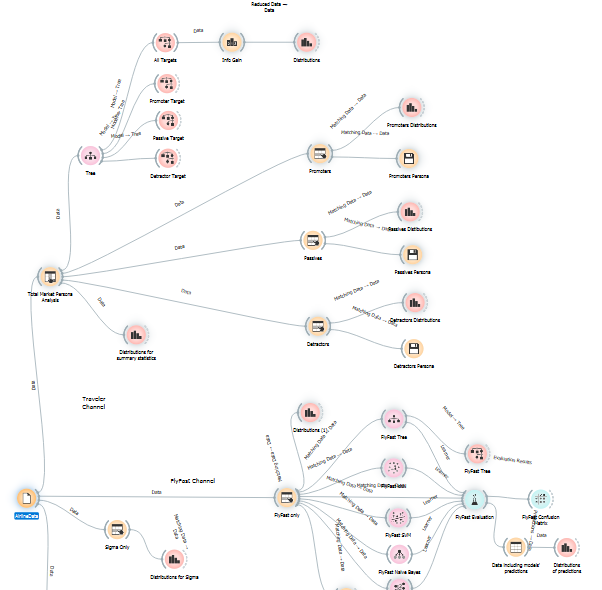
1. FlyFast must conduct a focus group with Adult and Senior Female travelers to better determine what they value most in an airline company. Our analysis shows Adult and Senior Female travelers are a high-payoff market segment and with their needs were met, they could radically improve FlyFast’s Promoter numbers.
2. FlyFast has the opportunity to become the market leader for Senior flyers by learning what sort features, policies, and alike are preferred by Seniors, and making them available. Also, after teens, Seniors are the second lowest represented age group in the Platinum Airline status—a status associated with Promoters. FlyFast could offer free Airline Status upgrades to Seniors even if this means increasing costs slightly as Senior are only moderately price sensitive. Adding a “Senior Aline Status” membership where senior travelers receive unique benefits (e-readers, books, early boarding, closer seating to the front, reserved overhead storage in front of planes, free drinks, etc.) would help improve FlyFast’s reputation amongst Senior travelers and better align their NPS status proportions to levels amongst our competitors.
3. Prioritize marketing and testing resources on moving Detractors to Promoters before trying to move Passives to Promoters. As stated earlier, a worst-case scenario is to spend marketing dollars on a segment of customer that does not require conversion. The modeling of FlyFast travelers shows that it is much easier to identify the Detractor segment than the Passive segment. As such, it is easier to identify their needs, and easier to assess if changes made to improve customer service are effective or if they need refinement.
4. Data shows that travelers with Medium to High loyalty tend to be Promoters. As FlyFast begins making progress with improving customer satisfaction, a parallel campaign designed to increase loyalty should also be executed. This could be done through club accounts, access to certain privileges, frequent flyer accounts, and alike.

**SECTION 3: ANNEXES**

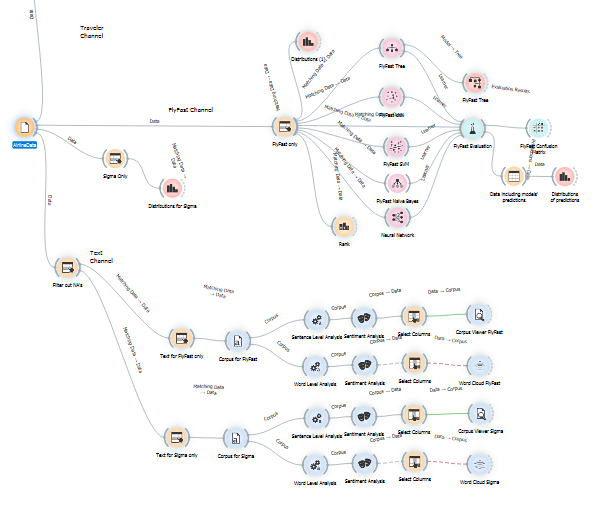
## **Annex 1: Work Flow**

Image from Orange

Part 1 of 2:



Part 2 of 2:



## **Annex 2: Description of Features in the Survey Dataset**

Engineered Features​

1. **Departure.Delay.GRP-** A discretization of the original continuous variable now binned as None (0-15 minutes), Minor (16-30 minutes, Medium (31-60 minutes) and Major (60 and more or flight has been cancelled.
2. **Arrival.Delay.GRP-** A discretization of the original continuous variable now binned as None (0-15 minutes), Minor (16-30 minutes, Medium (31-60 minutes) and Major (60 and more or flight has been cancelled.
3. **Flight.Efficiency.GRP-** This is a measure of flight time / distance and discretized into Low (0-6.2), Medium (6.2 – 7.1) and High (7.1 and above) values.
4. **Flights.Per.Year.GRP –** A discretization of the original continuous feature now binned as Rare (0-3), Occasional (4-12) and Frequent (13 and above) flights per year.
5. **Age.GRP -** A discretization of the original continuous variable now binned as Teens (15-22), Young Adults (23-36), Adults (37 – 56), and Seniors (57 and up).
6. **Loyalty.GRP** - A discretization of the original continuous variable now binned as Low, Medium, and High.

Features provided in the dataset

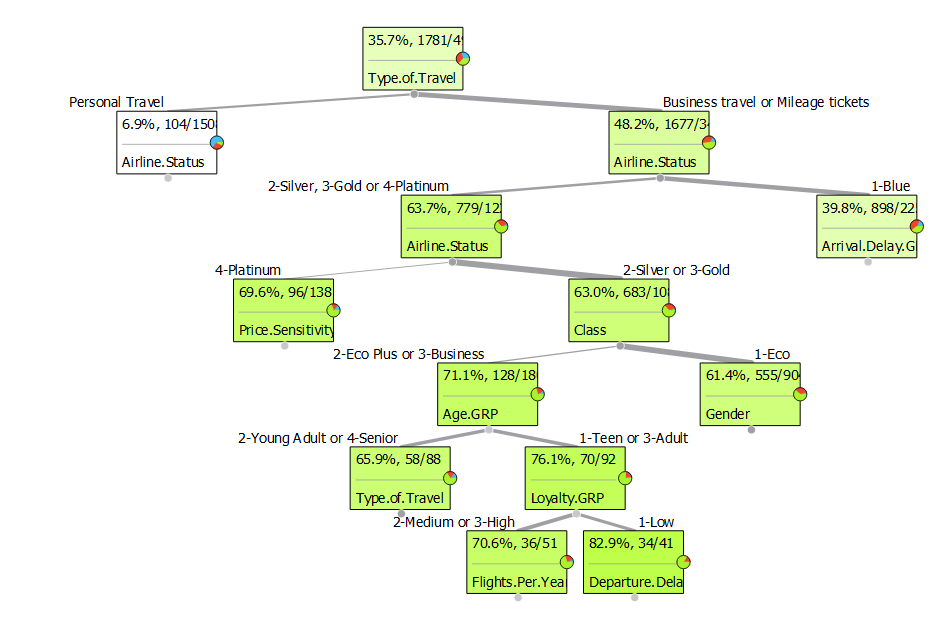
1. **Airline.Status** – each customer has a different type of airline status, which are platinum, gold, silver, and blue (based on level of travel with the airline)
2. **Age** – the specific customer’s age. Ranging from 15 to 85 years old.
3. **Gender** – male or female.
4. **Price.Sensitivity** – the grade to which the price affects to customers purchasing. The price sensitivity has a range from 0 to 5, with 5 being very price sensitive.
5. **Flights.Per.Year** – The number of flights that each customer has taken in the most recent 12 months. The range starting from 0 to 100.
6. **Loyalty** – An index of loyalty ranging from -1 to 1 that reflects the proportion of flights taken on other airlines versus flights taken on this airline. A higher index means more loyalty.
7. **Type.of.Travel** – One of business travel, mileage tickets, or personal travel (ex. vacation)
8. **Total.Frequent.Flyer.Accounts** – How many frequent flyer accounts the customer has
9. **Class** – three different kinds of service level (business, economy plus, and economy).
10. **Partner.Name** – These are the full names of the partner airline companies.
11. **Departure.Delay.in.Minutes** – How long the flight’s departure was delayed, when compared to schedule.
12. **Arrival.Delay.in.Minutes** – How long the arrival was delayed.
13. **Flight.Cancelled** – occurs when the airline does not operate the flight.
14. **Flight.time.in.minutes** –the length of time, in minutes, to reach the destination.
15. **Flight.Distance** – the distance between the departure and arrival destination.
16. **Likelihood.to.Recommend –** rated on a scale of 1 to 10, which shows how likely the customer is to recommend the airline to their friends (10 is very likely, and 1 is not very likely).
17. **Category -** This is a traveler’s net promoter stance which assigned likelihood to recommend values of 6 and below as “Detractors”, 7 and 8 as “Passives”, and 9 and 10 as Promoters.
18. **freeText**– a free form text field of the passenger comment, with respect to the flight.

Features removed from the dataset

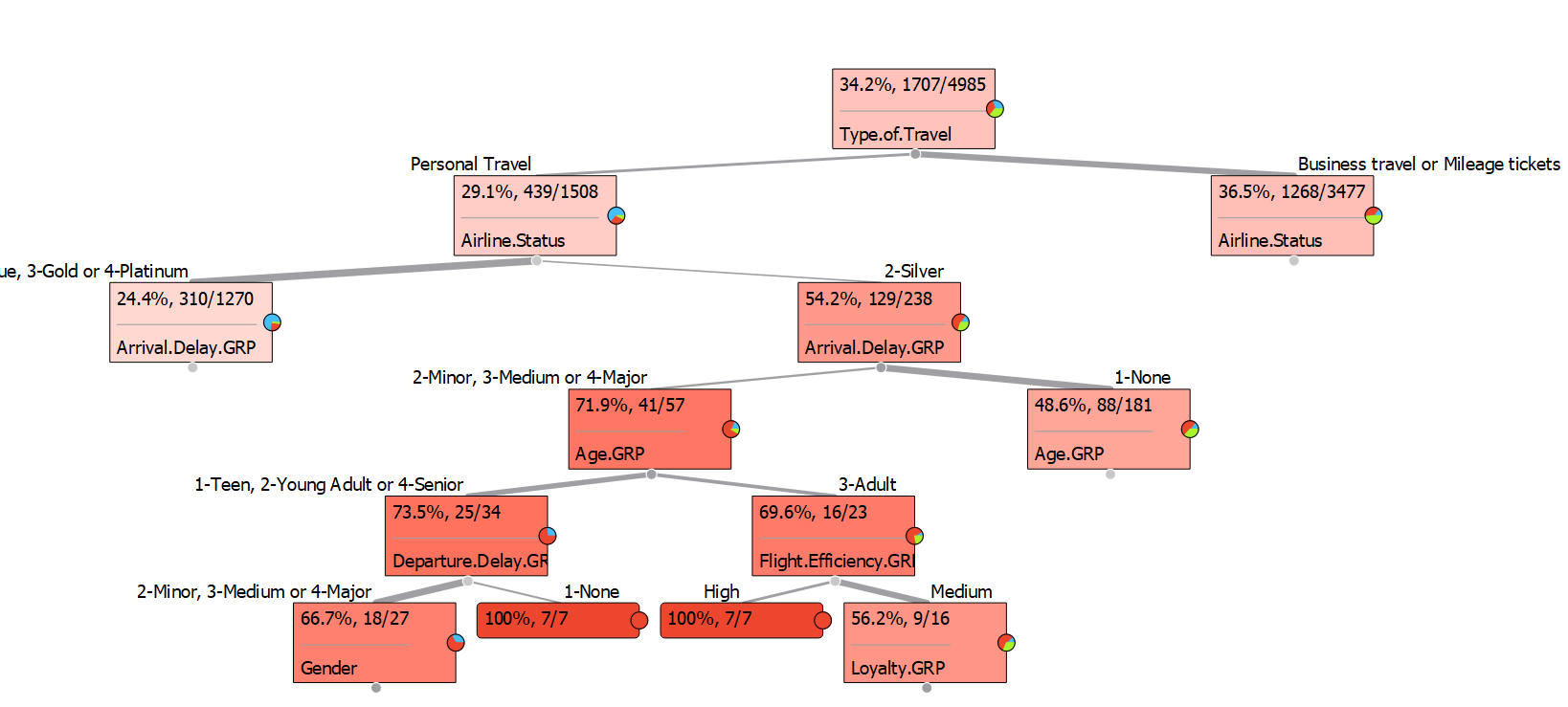
1. **REMOVED: Flight date** – the passenger’s flight date of travel.
2. **REMOVED: Origin City** – the place where passenger departed from. For example, Boston MA.
3. **REMOVED: Origin State** – the place where passenger departed from. For example, Texas.
4. **REMOVED: Destination City** – the place to which passenger travels to. For example, Boston MA.
5. **REMOVED: Destination State** – the place to which passenger travels to. For example, Texas.

## **Annex 3: Persona Classification Tree Diagrams**

**Promoters**



**Passives:**



**Detractors:**

