Analysis of Airline Passenger Satisfaction

Shun Zhang
UCSD
shz094@ucsd.edu

***Abstract*—Analyzing Airline Passenger Satisfaction dataset with business suggestions and build a predictive models to predict passengers' satisfaction level based on their basic informations.**

***Index Terms*—Data EDA,Marginal Gaussian Density, Chi-square test, Feature Selection, KNN, Logistic Regression**

I. Introduction

Air travel is more popular than ever, with numerous airlines offering competitive prices. However, as the airline industry grows, it faces a crucial question: How can airlines set themselves apart and ensure passenger satisfaction in a world with so many options?

Recent reports have highlighted a concerning trend: increasing passenger frustration within the U.S. airline industry. This emphasizes the need for airlines to dig into the factors that drive passenger satisfaction or dissatisfaction. Understanding these factors can provide airlines with insights to enhance various aspects of the travel experience, cultivate customer loyalty, and expand their customer base.

This paper delves into the complex world of airline passenger satisfaction, aiming to uncover the key factors that influence travelers' experiences and choices. Using a comprehensive approach that combines both quantitative and qualitative methods, we examine customer feedback data with their traveling record. Our research seeks to identify what truly makes passengers satisfied while also considering how consumer preferences are shaping the airline industry's future.

In the upcoming sections, we will outline our research methods, present our findings, and suggest practical steps airlines can take to enhance passenger satisfaction. Our goal is to contribute meaningfully to the ongoing conversation about passenger satisfaction in the airline industry, offering valuable insights to benefit airlines, policymakers, and passengers alike.

II. Dataset

A. dataset schemes

We choose a dataset from kaggle which includes customer satisfaction scores from over 120,000 airline passengers. [1] This dataset has a binary class of either satisfied or Neutral/Unsatisfied for each passenger and several features for each record. These features not only include basic information like gender, age, travel type, travel class, etc, but also include specific areas of the flight like ease of online booking, on board Wi-Fi service, Online boarding, etc with a score from 0 to 5. Detailed features are explained as we explore deeper into the dataset later.

B. Dataset EDA

1) Customer Info proportion: We firstly take a look at the proportion of the customers that attend this survey to gain an overview of the customer types. We cluster gender, customer type, type of travel, class and satisfaction which are non-quantitative features as basic information. Detailed proportion of each feature is shown in Figure 1. We can see here the customer type and type of travel is not balanced and the overall satisfaction level is below average.

Gender		Customer Type		Type of Travel		Class		Satisfaction	
gender		type of customer		type of travel		class		Satisfaction	
Female	51%	Returning	82%	Business	69%	Business	48%	Neutral or Dissatis...	57%
Male	49%	First-time	18%	Personal	31%	Economy	45%	Satisfied	43%
						Other (9411)	7%		

Fig. 1. Basic proportion of customer.

2) Satisfaction proportion of binary features: After having a glimpse of the basic features, we want to know how satisfactions are distributed in these features. We then separate each class feature apart and plot the satisfied and neutral/unsatisfied customers as shown in figure 2. We can see overall travelers who are first-time in flight, flying with personal reasons and staying in economy classes tend to be unsatisfied, while travelers who are traveling for business reasons or in business class tend to be satisfied. Gender doesn't affect the tendency of satisfaction overall. We will take extra analysis of those type of customers who generally have low satisfaction level in the later sections.

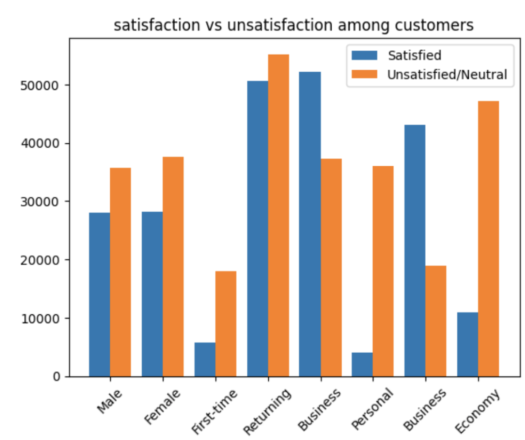

Fig. 2. Satisfaction proportion among customer types.

III. Key feature detection

As part of the analysis of this project, we want to know what are some key features that passengers are most concerned with. In other words, among all the survey areas that get a feedback from 0 to 5, what are the most determining features that affect passengers's satisfaction. We will use two different ways to detect feature relevance with satisfaction: Marginal Gaussian Density and Chi-Square test.

A. Marginal Gaussian Density

As all the services provided by the airline company is scaled from 0 to 5, we are assuming they are all Gaussian distributed which follows a pattern as below:

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

There are two advantages to assuming the distribution as Gaussian. First, it is simple and widely accepted, as Gaussian distribution is only characterized by two parameters, mean and standard deviation, and many statistical methods like hypothesis tests and confidence intervals are based on normal distribution assumption. Another advantage is Gaussian distribution can be quite robust to deviations from perfect normality, especially for large sample sizes. This means that even if the data isn't exactly Gaussian, statistical methods assuming normality can still provide reasonably accurate results.

1) Plotting: Marginal density, also known as a marginal probability density function (PDF) or marginal distribution, refers to the probability distribution of a subset of random variables from a joint probability distribution. Here each subset is one feature of customer feedback about airline company service. We plot each marginal density for each feature, while assuming they are gaussian distributed, and separate the satisfied and neutral/unsatisfied class as below shown in figure 3.

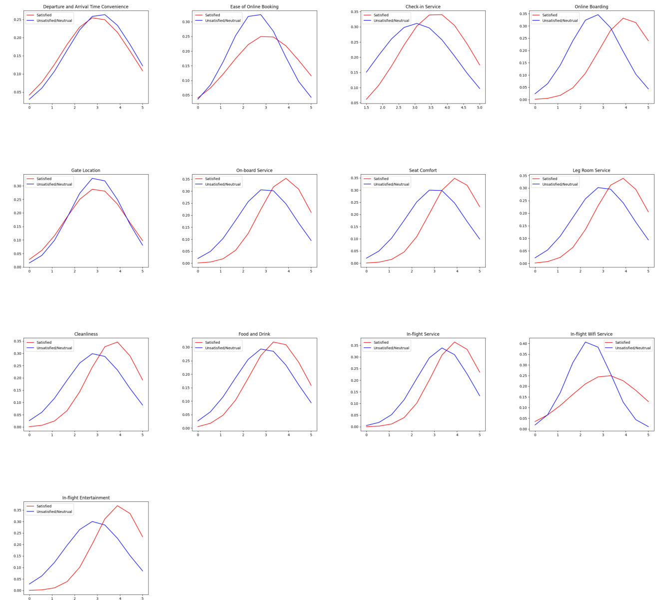

Fig. 3. Marginal Gaussian distribution for each service feature.

2) Analysis: We can then quickly distinguish which features are more deterministic and which are less deterministic from figure 3 by visually checking the difference between two distribution lines. Larger differences represent higher importance of this feature affecting the customer satisfaction and otherwise lower importance. For instance, the 'Departure and Arrival Convenience' and 'Gate Location' are obviously not affecting customer satisfaction level too much as the distribution from satisfied customer and neutral/unsatisfied customer are roughly the same. On the other hand, 'Online Boarding', 'In-flight Wifi Service' and 'In-flight Entertainment' have visually more separated distribution lines and we can conclude that these three are comparatively more important features affecting customer satisfaction.

3) Limitations: Although this analysis helps us to visualize the correlation between individual features and the result satisfaction, we can only have a rough idea of whether a feature is important or not. It is very hard for us to rank them as the idea of 'how different the two distribution lines are' is too abstract. For instance, Visually we would consider 'In-flight Wifi Service' has a larger difference of the marginal distribution than 'Online Boarding' as the former one has a larger difference in the pattern determined by the standard deviation. However, 'Online Boarding' has a larger difference in the mean of the distribution and it might be more correlated with the satisfaction. We can't tell as we are not aware of the weight importance between variance and mean. Thus we need a metric that can give us a quantitative value to determine the correlation and we turn to the Chi-Square test.

B. Chi-Square Test

Chi-Square Test is a statistical hypothesis test used in the analysis of contingency tables when the sample sizes are large. In simpler terms, this test is primarily used to examine whether two categorical variables (two dimensions of the contingency table) are independent in influencing the test statistic (values within the table). [2] The test is valid under the assumption that the sample data is Chi-square Distributed where Chi-Square Distribution is defined below: If Z1, ..., Zk are independent, standard normal random variables, then the sum of their squares:

$$Q = \sum_{i=1}^{k} Z_i^2 \tag{2}$$

is distributed according to the chi-squared distribution with k degrees of freedom. Here as mentioned before, we assume that each feature is a independent normal distribution and thus we can use the Chi-Square test to determine the correlation between each feature customer satisfaction with a Chi Score defined as:

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{3}$$

Where X square is the chi-squared, O is the observed value and E is the expected value.

1) Overall Chi Score: We use the Chi-Squared test to plot the Chi-Score out as shown in Figure 4. Larger Chi Score represents higher correlation between the feature and the satisfaction. Here it helps to solve the problem of Marginal Gaussian density where we can hardly rank their correlation. From this figure we can see that the top three correlations are 'Online Boarding', 'In flight Entertainment' and 'Seat Comfort' and the least three are 'Gate Location', 'Departure and Arrival Time Convenience' and 'Ease of Online Booking'. The trend is corresponding to the result we get from Marginal Gaussian Density but here it is much clearer. We can then use this metric to further check feature importance of the passenger segments.

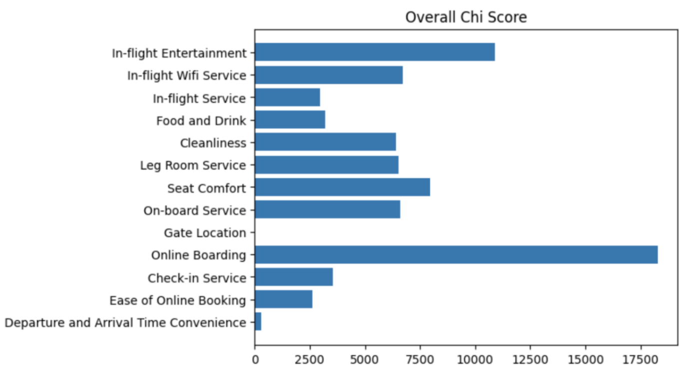

Fig. 4. Chi-Score over all passenger.

2) First Time Traveler: The plot of First Time traveler is shown in Figure 5. We can see the top three changed into 'In-flight Wifi Service', 'Ease of Online Booking' and 'Online Boarding' and the least three changed into 'Food and Drink', 'In-flight Entertainment' and 'Cleanliness'.

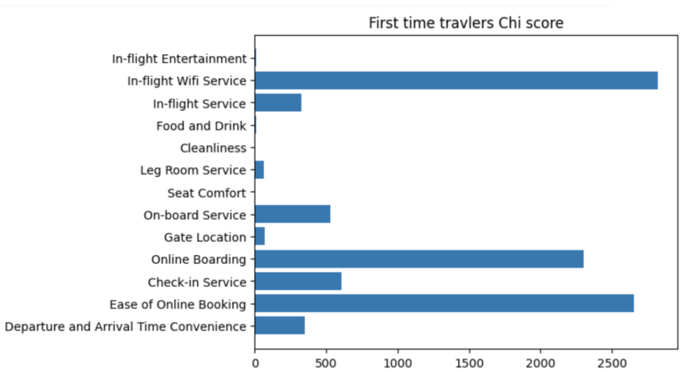

Fig. 5. Chi-Score of First time travelers.

3) Personal Travelers : The plot of Personal travelers is shown in Figure 6. We can see the top three changed into 'In-flight Wifi Service', 'Online Boarding' and 'Ease of Online Booking' while other features have significantly lower correlations.

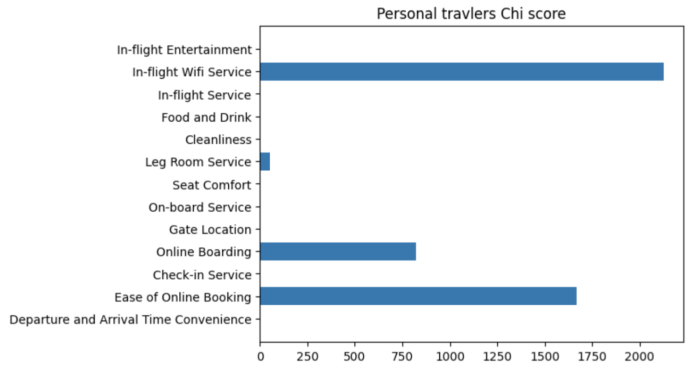

Fig. 6. Chi-Score of Personal travelers.

4) Economy Travelers : The plot of Economy travelers is shown in Figure 7. We can see the top three are the same with personal travelers and the least three changed into 'In-flight Service', 'Gate Location' and 'Departure and Arrival Time Convenience'.

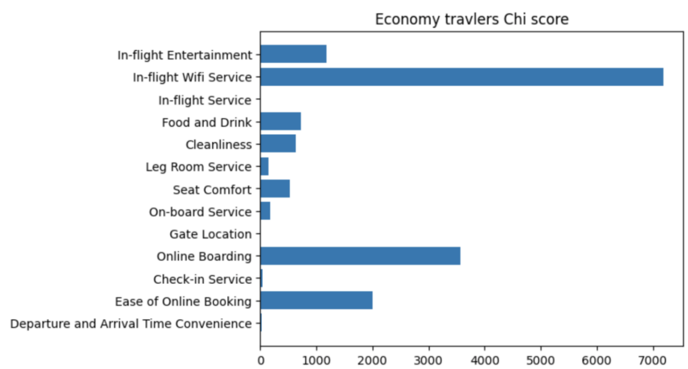

Fig. 7. Chi-Score of Economy travelers.

5) Business Travelers : We also want to see the correlation from the business travelers who have the highest satisfaction level and it is plotted in Figure 8. We can see they consider more features than other segmentations of passengers and they don't care about the 'Ease of Online Booking' much.

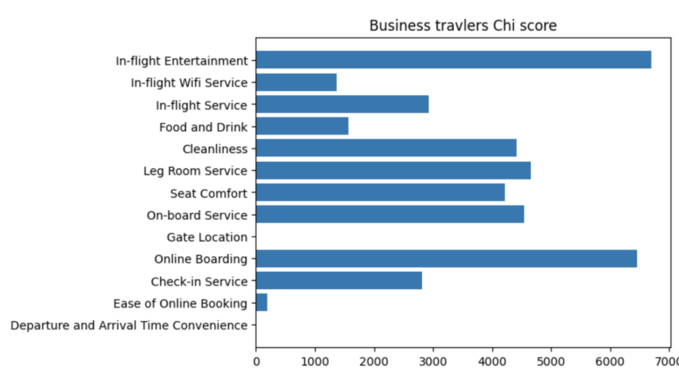

Fig. 8. Chi-Score of Business Class travelers.

6) Limitations: It is obvious that the Chi-Square test is a better way to test correlations between features and results as we can have a quantitative result other than abstract 'visual difference'. However, it is under the assumption that each feature has the same range of data and thus for those features without a standard scale, this might not be a good way to distinguish.

IV. Predictive Task

We also try to use this dataset to do a predictive analysis based on basic information we can collect from the passengers, which are the basic information including gender, age, customer type, type of travel, class, flight distance, departure delay and arrival delay. As the other features are not collectable for passengers who don't fill out the survey, we will ignore them in the predictive task. Here in this section we will show and explain some methods included in our predictive task.

A. Feature Selection

Like discussed in Key Feature Detection, we use the Gaussian marginal distribution to determine the feature that can be ignored and the result is plotted in Figure 9. For plots of 'Departure delay' and 'Arrival delay' we get rid of the outliers when plotting for better visualization and standardize the x axis. We eliminate the outliers with 1.5 IQR rules which eliminate Any observations that are more than 1.5 IQR below Q1 or more than 1.5 IQR above Q3 are considered outliers. We can see here from below that 'Gender', 'Departure Delay' and 'Arrival Delay' have less correlation with the final satisfaction level. As for Chi Square test, as mentioned above it is not appropriate when the range of each feature is significantly different so we skip this metric here.

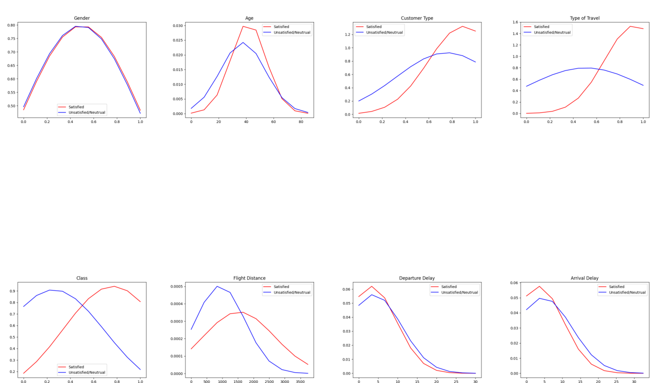

Fig. 9. Marginal Density for basic information.

B. Evaluation metrics

We are using Accuracy as our evaluation metric, which is a simple one for us to quickly recognize the performance of a model and it is simply derived by:

$$Accuracy = \frac{NumberofCorrectPredictions}{NumberofTotalPredictions} \quad (4)$$

C. Cross validation

Using only one portion of the data is not accurate enough as we aren't sure if the part we select is biased or not, we include cross validation in our predictive task. Cross-validation is a resampling method that uses different portions of the data to test and train a model on different iterations. [3]. We use 5 sample cross validation in this project.

D. Principle Component Analysis (PCA) dimension reduction

As we concluded before, Departure and Arrival delay is not very correlated to passenger satisfaction level. However, we still want to somehow use it in our model. Thus we include Principal Component Analysis for analyzing large datasets containing a high number of dimensions/features per observation, increasing the interpretability of data while preserving the maximum amount of information. [4] We will use it to combine delay information with distance information to create a new feature and see if it helps to improve the model performance.

V. Model Used

A. K Nearest Neighbours (KNN) model

We use K Nearest Neighbours (KNN) as our predictive model metric. This model can be used in both classification and regression tasks and here in a classification task, An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors. [5] In this dataset as we have proven Gender and Departure and Arrival Delay doesn't have a strong correlation with the result, which leaves us mostly binary classes as data. We will, however, still have models including those features as comparison of the performance. Moreover, as we can't tell which K number will give the best prediction under a certain metric, we also use cross validation for different K values, which are odd numbers from 1 to 19, to find the best K value. The use of odd numbers is to avoid equal votes for a datapoint from two classes. An example is shown in Figure 10 below. In most cases like the example, the accuracy start to stabilize after K passes 15 or 17.

B. Logistic Regression

We also use Logistic Regression as an alternative model where it is estimating the parameters of a logistic model [6]. Formally, in binary logistic regression there is a single binary dependent variable, coded by an indicator variable, where the two values are labeled "0" and "1", while the independent variables can each be a binary variable or a continuous variable. We will also perform metrics we discussed above on logistic regression and compare the result.

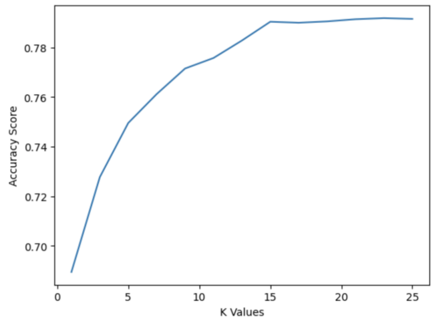

Fig. 10. Marginal Density for basic information.

VI. Model Result

In this section we will show and discuss the results of the predictive task including model performance and comparisons of different metrics' results. Models and evaluations metrics are referenced with scikit-learn library. [7]

A. *Performance result*

The model performance result of KNN is shown in TABLE I and the model performance result of Logistic regression is shown in TABLE II. Below are some explanations of abbreviations in the table:

- PCA 1: combine distance, departure delay, arrival delay
- PCA 2: combine distance, departure delay
- PCA 3: combine distance, arrival delay
- Reduce 1: no gender
- Reduce 2: no gender, departure delay and arrivial delay
- Reduce 3: no gender and departure delay
- Reduce 4: no gender and arrival delay

Note that if a any column includes a feature that is been reduced by PCA, it is a N/A in the cell as we cannot reuse it.

	All Data	Reduce 1	Reduce 2	Reduce 3	Reduce 4
No PCA	0.791	0.794	0.790	0.797	0.794
PCA 1	N/A	N/A	0.787	N/A	N/A
PCA 2	N/A	N/A	0.786	0.785	N/A
PCA 3	N/A	N/A	0.786	N/A	0.786

TABLE I
KNN Performance Result

	All Data	Reduce 1	Reduce 2	Reduce 3	Reduce 4
No PCA	0.783	0.783	0.782	0.784	0.782
PCA 1	N/A	N/A	0.772	N/A	N/A
PCA 2	N/A	N/A	0.772	0.784	N/A
PCA 3	N/A	N/A	0.772	N/A	0.783

TABLE II
Logistic Regression Performance Result

B. *Result Analysis*

From the result table above, the overall performance of KNN is roughly 0.8 to 1 percent better than Logistic Regression. Reducing the un-correlated features have a slight performance increase when we delete the gender and departure delay. The PCA metric doesn't have an improvement on the performance. On the other hand, it decreases the performance.

VII. Summary and Business Suggestions

For this project, we analysis the airline passenger survey's user feedback and passenger basic information, together with a model for predicting passenger's satisfaction based on their flying records with a accuracy of 79 percent. Also we conclude several business suggestions below:

- For Passengers who are first-time of the airline, consider providing free-trial in-air Wifi Service and better pre-boarding online instructions to increase their satisfaction level
- For passengers who are flying for personal reasons and economy class, consider products or SMS messages for easier online boarding and simplify the online booking steps.
- For business travelers, consider more options of online entertainments like more video channels or movie options.

Also, as we find out during predictive tasks that travel delay is not very correlated with satisfaction, where intuitively we consider longer delay can lead to higher dissatisfaction, maybe considering ways to get user feedback earlier after their trips. We suspect most passengers took this survey after a long time since they finished their last trip. As we don't have more detailed information of how this survey is taken, this suggestion might not be persuasive. Another possibility is that travelers indeed don't consider travel delay time as an important factor of their satisfaction level with the airline.

.

References

[1] "Airline Passenger Satisfaction," www.kaggle.com. https://www.kaggle.com/datasets/mysarahmadbhat/airline-passenger-satisfaction/data?select=datadictionary.csv

[2] "Chi-Square - Sociology 3112 - Department of Sociology - The University of utah". soc.utah.edu. Retrieved 2022-11-12.

[3] Allen, David M (1974). "The Relationship between Variable Selection and Data Agumentation and a Method for Prediction". Technometrics. 16 (1): 125–127

[4] Jolliffe, Ian T.; Cadima, Jorge (2016-04-13). "Principal component analysis: a review and recent developments". Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 374 (2065)

[5] Cover, Thomas M.; Hart, Peter E. (1967). "Nearest neighbor pattern classification". IEEE Transactions on Information Theory. 13 (1): 21–27.

[6] Tolles, Juliana; Meurer, William J (2016). "Logistic Regression Relating Patient Characteristics to Outcomes". JAMA. 316 (5): 533–4.

[7] "Scikit-learn, Machine Learning in Python" https://scikit-learn.org/stable/