Machine Learning Assignment(Airline-satisfaction)

<S5414643>

22 November 2019

Contents

List of Figures

List of Tables

1 Introduction

The objective of this report is to predict customer satisfaction using a survey taken on airline satisfaction. This information has been converted into a data-set which we will use to aid our investigation. Throughout this investigation will be using python for both data analysis and machine learning, in a attempt to predict satisfaction. In 2009 the service industry was contributing to 60% of the annual GDP in the USA. Not to mention the airline industry being relativity competitive in more economically developed countries.(An and Noh 2009) Studies have also shown that companies that are able to increase the percentile of loyal customers per year, resulted in substantial increases in overall profit.(Bowen and Chen 2001)As mentioned above, the airline industry is quite competitive, as a result companies invest a lot of resources into researching ways to give them the edge. In fact, an article by(McColl-Kennedy and Schneider 2000)states that companies gain an advantage by the level of innovation that they have implement. As a result of this constant innovation, the customer involvement in the performance equation grows. Which is why being able to predict customer satisfaction would aid their research.

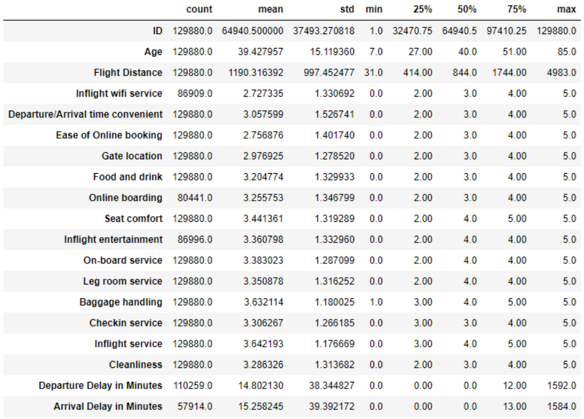
2 Exploratory Data Analysis

In this section we are going to explore the data in its current state, and to describe the data-sets properties. Which in return would answer weather the data-set can represent the problem, in addition to bringing any potential issues to light that could affect later steps in the investigation.

2.1 Data Description

Firstly Definitions of each feature can be located in the appendix which is **figure 16**. Moving on, the airline data set is a relatively large sample of data containing 24 features, a feature being an individual measurable characteristics. As mentioned prior they are all listed in the appendix excluding ID. Using the **".info()"** function we imported from the panda package, we were able to gather further information. For example this data set contains 129,880 instances. **Figure 17** highlights the way features and instances are displayed in a table, for further reference.

This data set is mostly comprised of integer values, 14 to be specific. This is because a lot of the features where being used to rate the customers overall experience with the airline. From what we can gather the ratings are from 0 to 5. We can gather this information from **figure** 1, as it is a table that displays useful statistics about the numerical features in the data set. If we focus our attention on features such as, "Online Boarding", "Gate Location" and "Seat Location" the min(minimum value) for these features are 0 and the max(maximum value) is 5. We see a similar trend with all the features that we suspect are being rated from 0 to 5. However, baggage handling is an anomaly as the min value is 1. This could be down to coincidence, and none of the passengers rated it below 1 which is a miracle, but we will find out later down the line.

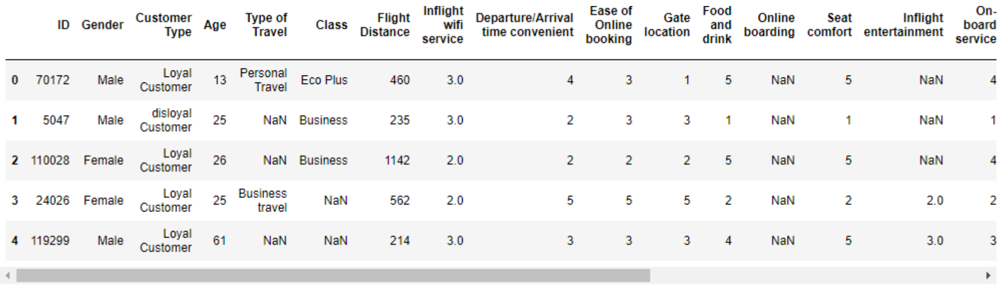


	count	mean	std	min	25%	50%	75%	max
ID	129880.0	64940.500000	37493.270818	1.0	32470.75	64940.5	97410.25	129880.0
Age	129880.0	39.427957	15.119360	7.0	27.00	40.0	51.00	85.0
Flight Distance	129880.0	1190.316392	997.452477	31.0	414.00	844.0	1744.00	4983.0
Inflight wifi service	86909.0	2.727335	1.330692	0.0	2.00	3.0	4.00	5.0
Departure/Arrival time convenient	129880.0	3.057599	1.526741	0.0	2.00	3.0	4.00	5.0
Ease of Online booking	129880.0	2.756876	1.401740	0.0	2.00	3.0	4.00	5.0
Gate location	129880.0	2.976925	1.278520	0.0	2.00	3.0	4.00	5.0
Food and drink	129880.0	3.204774	1.329933	0.0	2.00	3.0	4.00	5.0
Online boarding	80441.0	3.255753	1.346799	0.0	2.00	3.0	4.00	5.0
Seat comfort	129880.0	3.441361	1.319289	0.0	2.00	4.0	5.00	5.0
Inflight entertainment	86996.0	3.360798	1.332960	0.0	2.00	4.0	4.00	5.0
On-board service	129880.0	3.383023	1.287099	0.0	2.00	4.0	4.00	5.0
Leg room service	129880.0	3.350878	1.316252	0.0	2.00	4.0	4.00	5.0
Baggage handling	129880.0	3.632114	1.180025	1.0	3.00	4.0	5.00	5.0
Checkin service	129880.0	3.306267	1.266185	0.0	3.00	3.0	4.00	5.0
Inflight service	129880.0	3.642193	1.176669	0.0	3.00	4.0	5.00	5.0
Cleanliness	129880.0	3.286326	1.313682	0.0	2.00	3.0	4.00	5.0
Departure Delay in Minutes	110259.0	14.802130	38.344827	0.0	0.00	0.0	12.00	1592.0
Arrival Delay in Minutes	57914.0	15.258245	39.392172	0.0	0.00	0.0	13.00	1584.0

Figure 1: descriptive-stats-table

The remaining 10 features are evenly divided between objects and floats. A common trend among the objects is that they all seem to hold set options such as "Male" or "Female" and "Loyal Customer" or

"Disloyal Customer", as shown in **Figure 2**, which is an image of the first 5 rows in the data set. By focusing on features such as "Class", "Gender", and "Customer Type", we can see that this trend holds true. Lastly, the floats are used to store time values, which makes sense as these measurements can be tracked down to milliseconds. Using another data type could potentially result in data loss.



	ID	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	Food and drink	Online boarding	Seat comfort	Inflight entertainment	On-board service
0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3.0	4	3	1	5	NaN	5	NaN	4
1	5047	Male	disloyal Customer	25	NaN	Business	235	3.0	2	3	3	1	NaN	1	NaN	1
2	110028	Female	Loyal Customer	26	NaN	Business	1142	2.0	2	2	2	5	NaN	5	NaN	4
3	24026	Female	Loyal Customer	25	Business travel	NaN	562	2.0	5	5	5	2	NaN	2	2.0	2
4	119299	Male	Loyal Customer	61	NaN	NaN	214	3.0	3	3	3	4	NaN	5	3.0	3

Figure 2: Data-set.Head()

With regards to the features, satisfaction is a stat that is being measured and recorded in this data set. This is the most important feature for this investigation as it is what we are trying to predict, this feature contains three results which are, neutral, satisfied and dissatisfied. The existence of this features allows us to compare it to our rating features that where mentioned above. Which in return allows us to compare other features to the satisfaction outcome. **Figure 3** demonstrates the distribution of the satisfaction level in a pie chart.

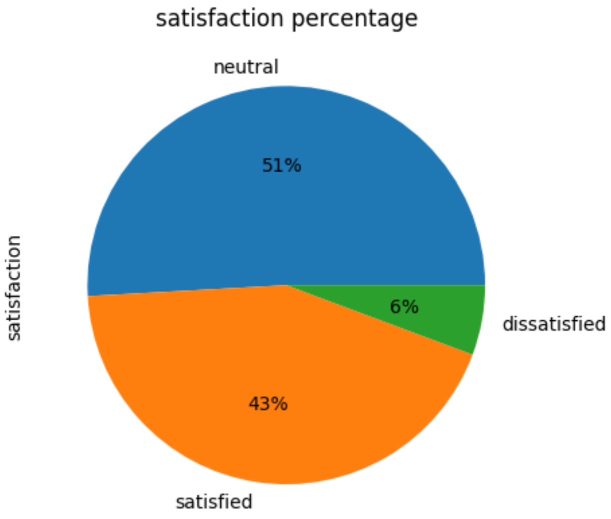


Figure 3: Satisfaction Pie Chart

We've examined the distribution of the ratings features, as shown in **Figure 4**. These charts display the frequency of each rating. It's important to note that some categories have received no ratings, but they are still recorded. They are such a small fraction of data that they aren't visible. This data is useful as it identifies areas where the company could allocate extra resources to raise the average rating. It also shows where they are succeeding and could apply those practices to areas that are falling behind. For example, "In-flight WiFi service" has the lowest average rating while "In-flight service" has the highest. This is demonstrated in **Figure 1** and **Figure 4**.

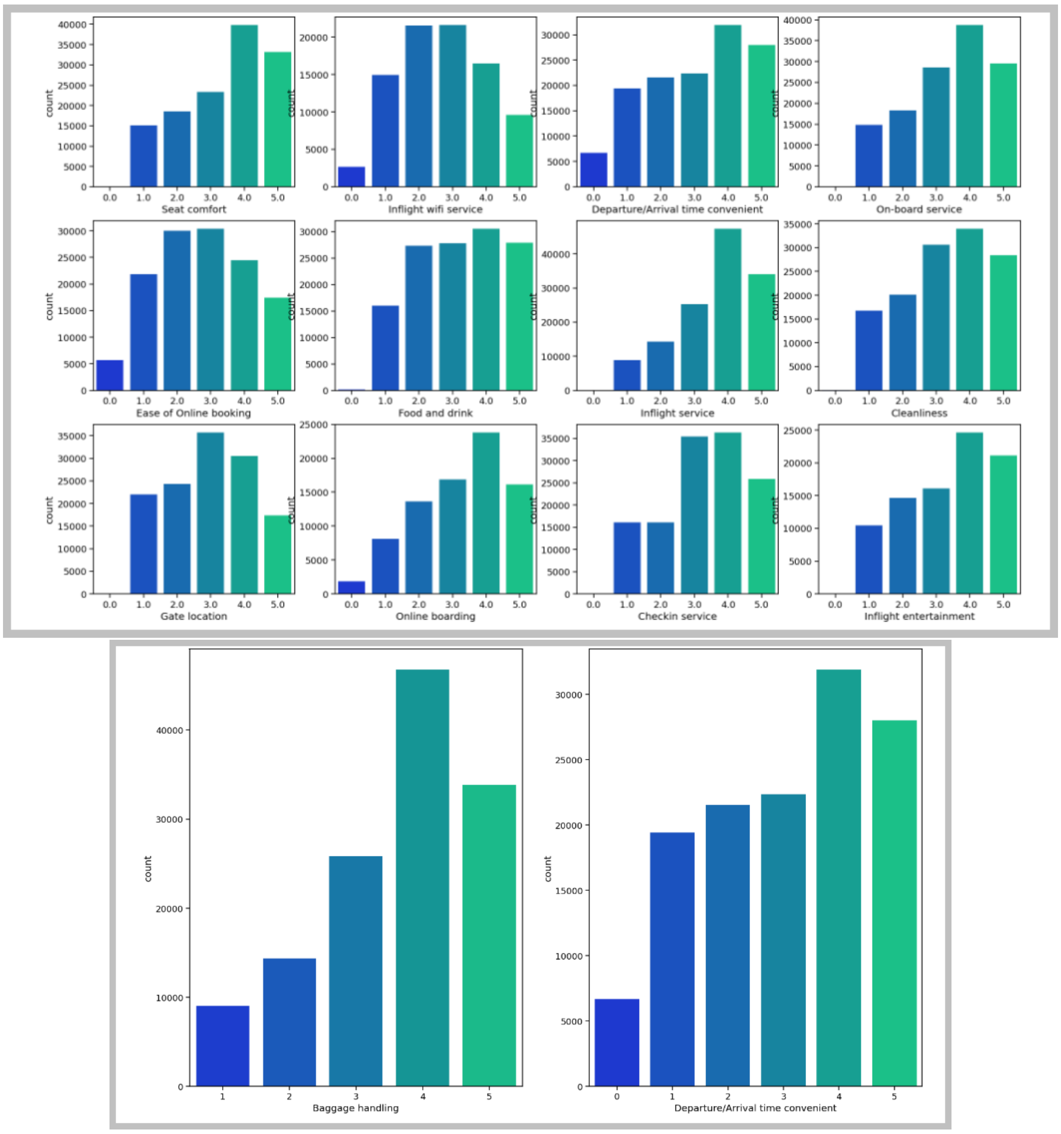


Figure 4: Frequency of ratings features

We also took a look at how these rating features compared to satisfaction to see if we could spot any relationships between them, which is to be expected however we want to look at how strongly correlated they are. This is what **figure 5** demonstrates

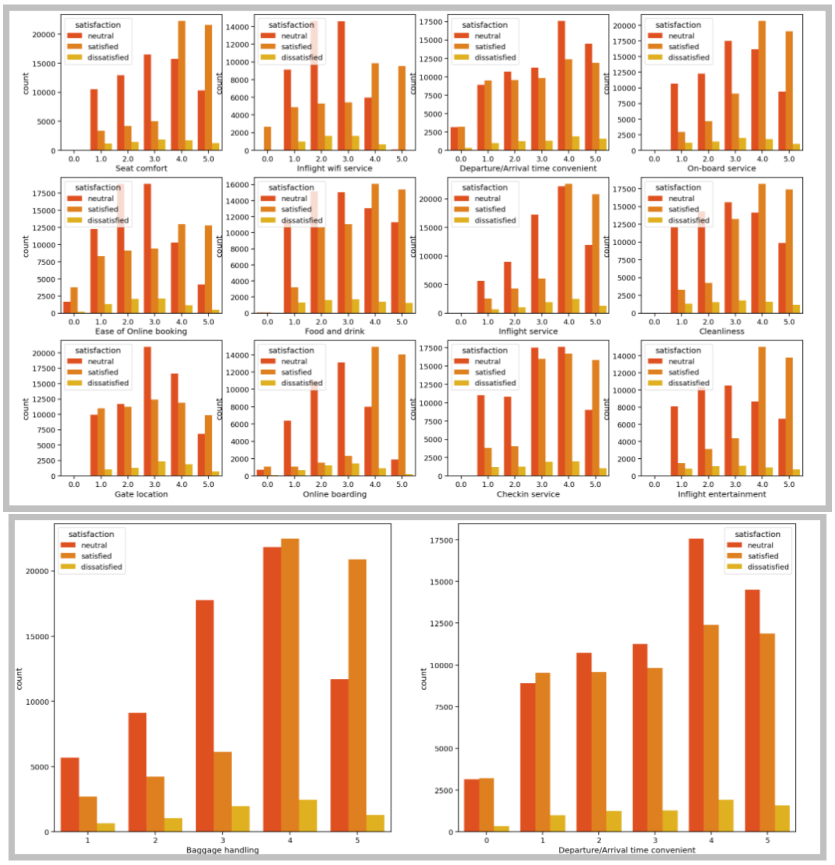


Figure 5: Satisfaction and rating bar chart

The analysis of customer ratings revealed a neutral trend across all categories, with no significant instances of dissatisfaction. However, a clear pattern emerged in the categories of "on-board service" and cleanliness, where a high rating of 4 or 5 correlated with high levels of satisfaction, while lower ratings of 1 to 3 corresponded with low levels of satisfaction. This suggests that these factors have a significant impact on a passenger's overall level of satisfaction.

Furthermore, a data visualization of an age and satisfaction histogram revealed that the average customer which is 39 years or older, also displayed high levels of satisfaction, as shown in **Figure 6**.

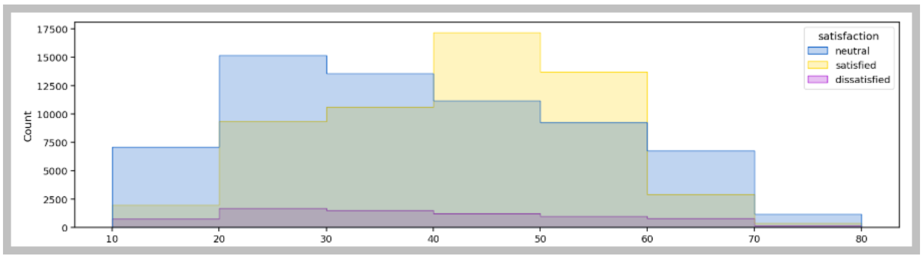


Figure 6: Age histogram

The gender distribution is relatively balanced and so are the satisfaction levels between the genders, which means this feature not having any drastic impacts on satisfaction as seen in these figure 7 and 8.

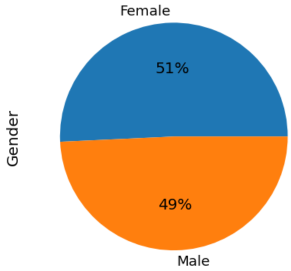


Figure 7: Gender distribution Pie chart

We looked at how type of travel and class where affecting satisfaction. It appears that majority of customers traveling for business in business class ended up being satisfied customers. None of the other groups displayed anything out of the blue or particularly interesting.

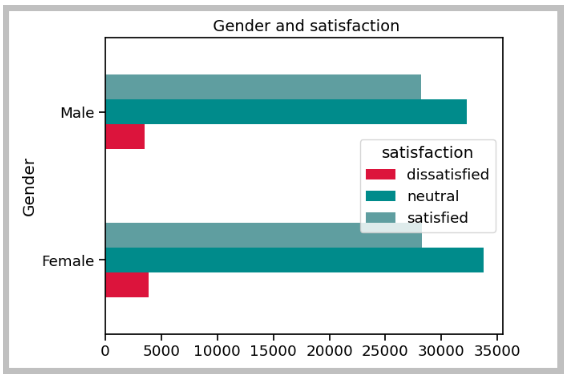


Figure 8: Gender and satisfaction levels

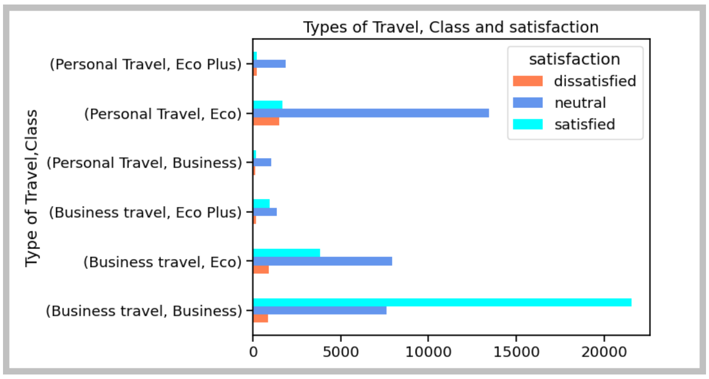
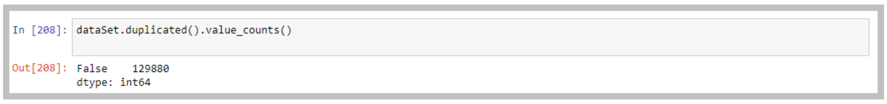


Figure 9: This figure is a bar chart that shows the satisfaction distribution based the reason your flying

2.2 Data Analysis

This should go into analyzing the data for quality and insights. So, check for, e.g., duplicates, missing values and noise. For analysis, you can do, e.g., correlation analysis and clustering analysis.

Before advancing to the testing phase of this investigation we needed to check for any potential issues that could invalidate our test. For example the data set has been checked for any duplicate values, as this could skew the results our model gives back to us, at least in this case all the passengers should be unique. **Figure 10** below displays the results of the check, if any rows came up true that would mean a duplicate was present.

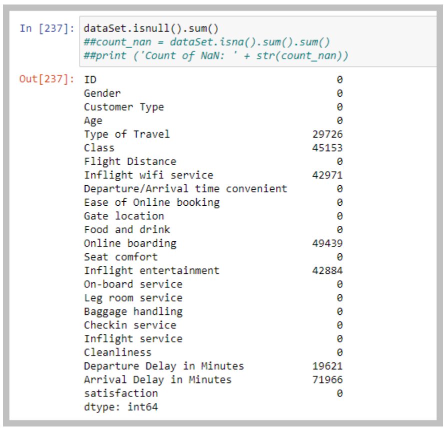


```
In [208]: dataSet.duplicated().value_counts()

Out[208]: False    129880
          dtype: int64
```

Figure 10: This figure demonstrates the check for duplicates in the data set

Unfortunately, the data set was plagued with missing data, this would greatly affect our investigation, as it impacts the credibility of the results that are models would produce(Farhangfar *et al.* 2008). **Figure 11** below illustrates the extent of missing data for each feature in the data set. As a preliminary observation, there are 7 features with missing values, accounting for a total of 301,760 missing values, or approximately 0.00968073...% of the data. While this may seem insignificant, it is important to note that in order to achieve the most accurate results from our analysis, it is ideal to address this missing data.. As shown in the **figure 11** “Arrival delay in minutes” is the feature with the most missing values.



```
In [237]: dataSet.isnull().sum()
          ##count_nan = dataSet.isna().sum().sum()
          ##print ('Count of NaN: ' + str(count_nan))

Out[237]: ID                                       0
          Gender                                   0
          Customer Type                            0
          Age                                      0
          Type of Travel                       29726
          Class                                45153
          Flight Distance                          0
          Inflight wifi service                42971
          Departure/Arrival time convenient        0
          Ease of Online booking                   0
          Gate location                            0
          Food and drink                           0
          Online boarding                      49439
          Seat comfort                             0
          Inflight entertainment               42884
          On-board service                         0
          Leg room service                         0
          Baggage handling                         0
          Checkin service                          0
          Inflight service                         0
          Cleanliness                              0
          Departure Delay in Minutes           19621
          Arrival Delay in Minutes             71966
          satisfaction                             0
          dtype: int64
```

Figure 11: This figure look at the amount of missing values that are in the entry

Flight distance, which is one of the features has a few outlier present.This is bad because outliers in a data set can have a significant impact on Machine Learning models. This is because they lie considerably outside the range of majority of the data and can skew the models performance this can result in poor predictions and like the missing data needs to be handled(Hua and Pei 2007). Everything outside the whiskers of the box-plot would be consider an outlier as show in **figure 12** below.

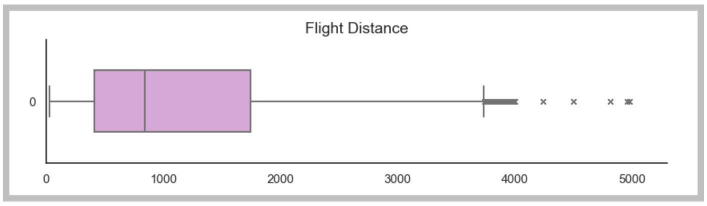


Figure 12: Flight Distance box plot

(Chai 2020)

To determine whether any relationships could prove useful, we also examined the correlation between the numerical features. Heat-maps make information easier to digest. Thus, we found that most features have poor relationships, whether positive or negative. However "Departure delay in minutes" and "arrival delay in minutes" had a close to perfect positive correlation. It makes complete sense when one considers what they measure. That being said I don't think having both of them is particularly useful for our end goal. This could pose problem because features that are too highly correlated can result in over fitting, which is when the machine learning model performs well on the training data but not on test.

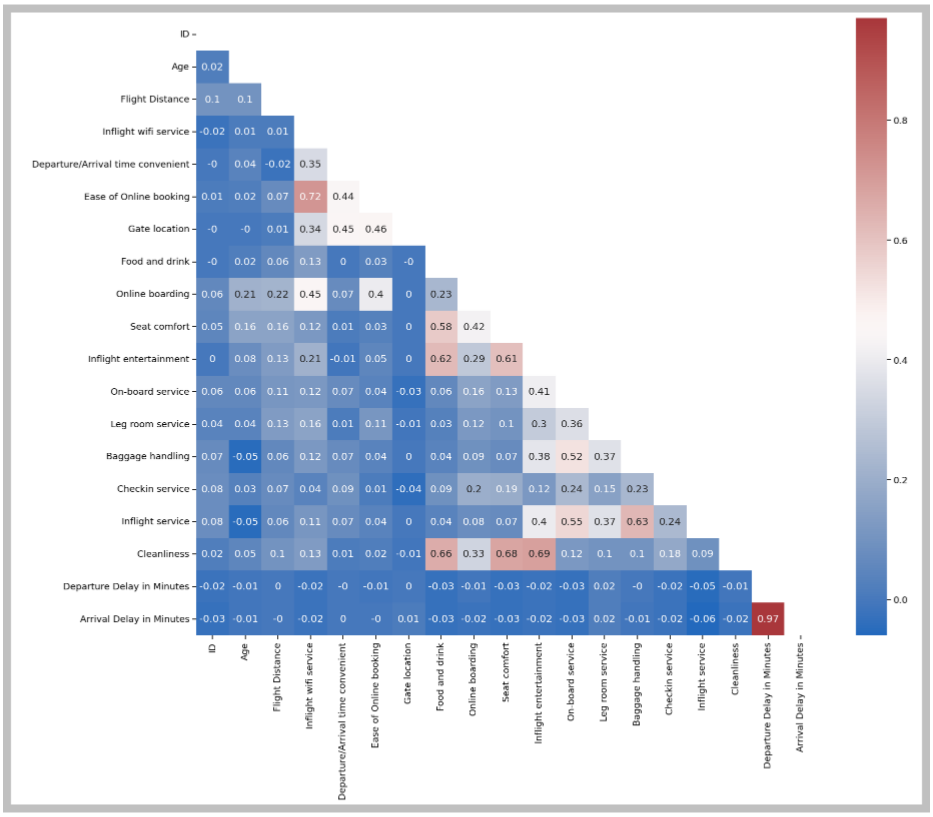


Figure 13: Heat-map which demonstrates the correlation between features

When "Departure delay in minutes" and "arrival delay in minutes" was converted to a scatter plot, as expected it shows a really strong correlation, however it also reveals some outliers, they are still following the data pattern but they would be considered extreme examples, as shown in **Figure 14**.

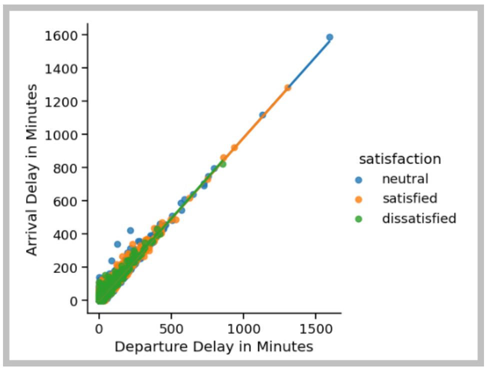


Figure 14: Departure delay and Arrival Delay scatter plot

3 Method

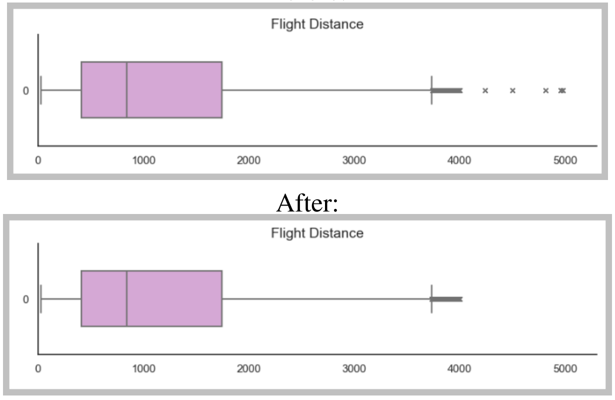
In this section we are going to talk about our aims and objectives, how we tackled data processing. Our chosen classifiers the validation method we are going to be using and how we are going to evaluate the results

3.1 Aims and Objectives

The primary objective for handling missing data in the data set is to find a solution that minimizes loss of the original data. Additionally, a secondary objective is to achieve an overall accuracy of at least 90% on at least one model, while also striving for an F1 score of 92% or higher. This objective is closely tied to evaluating the performance of various machine learning models on this data set, in order to determine the most suitable model for this data set.

3.2 Data Processing

First of all, we will remove the "Arrival Delay in Minutes" table because it is so closely related to "Departure Delay in Minutes" that it is redundant, excluding this could also simplify the data and make and allow the models to understand easier. Additionally, it has the most missing values. To remove outliers from flight distance we used interquartile range. It didn't get rid of everything beyond the upper whisker however it did eliminate the extreme outliers, as this solely dependent on where people are flying to the distance can vary, which is the outliers left are reasonable, as shown in **Figure 15**


Before:

After:

Figure 15: Block-plot before and after

Based on the data analysis, it can be concluded that the missing data in the data set is MCAR (Missing Completely At Random), indicating that the absence of this data is unrelated to any other feature in the data set and was not influenced by any other data.(Acock 2005) There are so many rows with missing data, as shown in figure X, that dropping all the instances with 'N/A' would result in a large loss of data. It is imperative to avoid underfitting the data to the chosen classifiers(Koehrsen 2018), as this will have a significant impact on the results. Furthermore, there is still a great deal of authentic data available. As a result, imputation appears to be the most appropriate method.(Royston 2004) The reason why we selected multivariate imputation is that it utilizes multiple features that result in more accurate estimates. This is because it considers the relationships between variables rather than just one column to make its determination. For this we chose to use KNN-Imputer. It is a machine learning model that allows you to impute values based on a chosen amount of surrounding variables.

The first step was to convert object features such as "class" to numeric data since the imputer only accept numerical values. A majority of objects had only two options, for example gender, was converted assigned '1' or '0' by using ".map"

Other objects such as "Class" and "Satisfaction" were challenging because they held three options, in addition to having missing values which made it difficult to convert. This was solved by using ordinal encoding. This was suitable because these features have a hierarchy of values from best to worst which is why this style of encoding was suitable. It also allows me to keep the N/A fields and allow the imputer to correct them.

3.3 Classifiers and Configuration

we decided to select gradient boosting as one of the classifiers, it is considered one of the more computationally demanding classifiers. However, it is said to be good on data sets that contain many features and it is said to perform particularly well with image and text classification. It is interesting because the idea behind it is to take weak learning algorithms in this case regression tress. These regression trees then output real values, during this process new learners are created, while the old ones remain unchanged, then they are concatenated to correct errors in predictions.(Bentéjac *et al.* 2021)

KNN(K-Nearest Neighbor) Even though it is said not to perform very well on large data set, I want to investigate to what extent this is true.(Dreiseitl and Ohno-Machado 2002)

We are also going to be doing testing on random forest, another decision tree based algorithm like gradiant boosting. However the way they use decision trees to make prediction are different. Random forest builds multiple decision trees independently which is a bagging method whereas gradiant boosting builds them sequentially which is a boosting method.(Probst *et al.* 2019)

3.4 Validation Method

Based off the analysis of this data set, hold-validation is the most suitable for this data set. Considering that it is a larger data set, hold-validation seems suitable because of the way it splits up the data into subsection which in result splits the work load. In return speeding i[the process and making less computationally demanding.(Raschka 2018)

3.5 Evaluation Metrics

Which metrics are you going to use and why?

To effectively evaluate the performance of the model, we will utilize a classification report that includes not only accuracy, but also metrics such as precision, recall, and F1-score. This approach is valuable as it provides a more comprehensive understanding of the model's performance, even if the overall accuracy is not high. Additionally, utilizing a confusion matrix in conjunction with the classification report will allow us to visualize the specific decisions made by the model and gain further insight into the results.

4 Results and Discussion

The results of the investigation show, that the highest overall accuracy we were able to achieve was 88%, this was achieved using Random Forest. This was expected as is considered to perform exceptionally well on larger data-sets, not to mention to be extremely robust against over-fitting**. Table 1** displays the results pulled from the classification report on random forest. Taking a closer look, we can see that it did an exceptional job at predicting the more prominent results which were satisfied and neutral. However really struggled with dissatisfaction, which your going to see is a common theme with all the classifiers. I believe this is because the "support" size which is the number of samples each metric had access to was considerably smaller than its counterparts. Treating it almost as if it were an outlier, as a result the models where bias towards the other 2 outcomes. In fact, the model that performed the worst which was KNN was able to make the most right predictions on dissatisfaction as show in the **table 3**.

Table 1: Gradiant Boosting

	Gradiant boosting			
Satisfaction	Precision	Recall	F-score	support
Satisfied	92%	91%	92%	10973
Neutral	84%	94%	89%	12984
Dissatisfied	8%	0%	0%	1447

Overall Accuracy : 87%

Table 2: Random Forest

	Random Forest			
Satisfaction	Precision	Recall	F-score	support
Satisfied	92%	91%	92%	10973
Neutral	84%	94%	89%	12984
Dissatisfied	8%	0%	0%	1447

Overall Accuracy : 88%

Table 3: K-Nearest Neighbors

	K-Nearest Neighbors			
Satisfaction	Precision	Recall	F-score	support
Satisfied	69%	71%	70%	10973
Neutral	70%	75%	72%	12984
Dissatisfied	9%	0%	0%	1447

Overall Accuracy : 69%

“

The confusion matrix in this study indicates a bias towards recognizing dissatisfaction as neutral. The corresponding confusion matrices can be found in the appendix, with 0 representing "satisfied", 1 representing "neutral", and 2 representing "dissatisfied". Hyper-parameter tuning was done using varied learning rates for gradient boosting and the best results were found within a range of 0.75 to 1.2. Random forest was also attempted but did not yield significant results. The overall performance on the test data was good, but it is believed that better results could have been achieved with a more balanced distribution of dissatisfaction. Despite addressing missing values, it would have been preferable to have collected new data, as a portion of the data is theoretical and could have affected the results.

5 Conclusions and Further Work

If more time were elected we could run test on smaller samples of the data, where the distribution of satisfaction wasn’t so large, maybe that would help the models ability to recognize the less prominent options. More in-depth hyper parameter tuning was needed to get maximum results, however with the current computation power available it would take to much time to run. More models would also be tested to see how they performed stacked up against our current ones. Additionally a better sample of the data set could be acquired, to see if the data was truly lost.

References

Alan C Acock. 2005. Working with missing values. *Journal of Marriage and family*, **67**, 1012–1028.

Myungsook An and Yonghwi Noh. Sep 2009. Airline customer satisfaction and loyalty: impact of in-flight service quality. *Service Business*, **3**, 293–307. ISSN 1862-8508. URL `https://doi.org/10.1007/s11628-009-0068-4`. (doi:10.1007/s11628-009-0068-4)

Candice Bentéjac, Anna Csörgő and Gonzalo Martínez-Muñoz. 2021. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, **54**, 1937–1967.

John T. Bowen and ShiangâLih Chen. Jan 2001. The relationship between customer loyalty and customer satisfaction. *International Journal of Contemporary Hospitality Management*, **13**, 213–217. ISSN 0959-6119. URL `https://doi.org/10.1108/09596110110395893`. (doi:10.1108/09596110110395893)

Christine P Chai. 2020. The importance of data cleaning: Three visualization examples. *Chance*, **33**, 4–9.

Stephan Dreiseitl and Lucila Ohno-Machado. 2002. Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics*, **35**, 352–359.

Alireza Farhangfar, Lukasz Kurgan and Jennifer Dy. 2008. Impact of imputation of missing values on classification error for discrete data. *Pattern Recognition*, **41**, 3692–3705.

Ming Hua and Jian Pei. Cleaning disguised missing data: a heuristic approach. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 950–958, 2007.

Will Koehrsen. 2018. Overfitting vs. underfitting: A complete example. *Towards Data Science*.

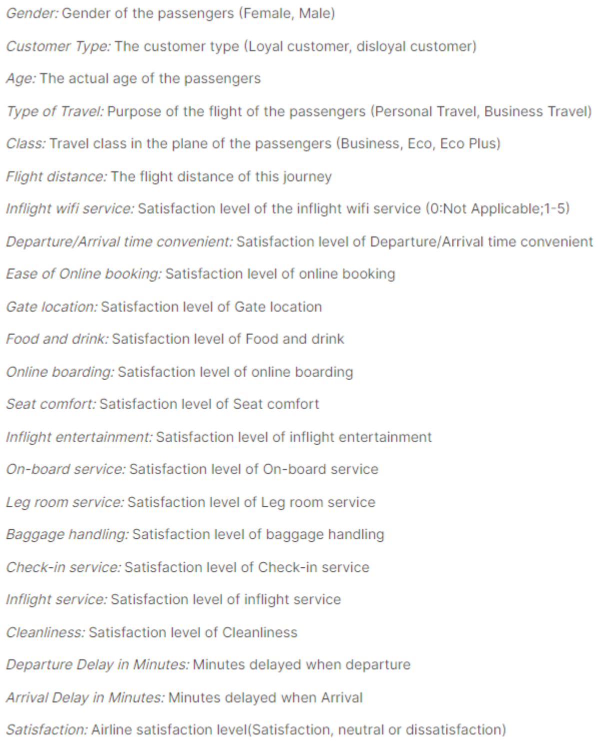
Janet McColl-Kennedy and Ursula Schneider. 2000. Measuring customer satisfaction: Why, what and how. *Total Quality Management*, **11**, 883–896. URL `https://doi.org/10.1080/09544120050135434`. (doi:10.1080/09544120050135434)

Philipp Probst, Marvin N Wright and Anne-Laure Boulesteix. 2019. Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, **9**, e1301.

Sebastian Raschka. 2018. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808*.

Patrick Royston. 2004. Multiple imputation of missing values. *The Stata Journal*, **4**, 227–241.

A Data Analysis



Gender: Gender of the passengers (Female, Male)

Customer Type: The customer type (Loyal customer, disloyal customer)

Age: The actual age of the passengers

Type of Travel: Purpose of the flight of the passengers (Personal Travel, Business Travel)

Class: Travel class in the plane of the passengers (Business, Eco, Eco Plus)

Flight distance: The flight distance of this journey

Inflight wifi service: Satisfaction level of the inflight wifi service (0:Not Applicable;1-5)

Departure/Arrival time convenient: Satisfaction level of Departure/Arrival time convenient

Ease of Online booking: Satisfaction level of online booking

Gate location: Satisfaction level of Gate location

Food and drink: Satisfaction level of Food and drink

Online boarding: Satisfaction level of online boarding

Seat comfort: Satisfaction level of Seat comfort

Inflight entertainment: Satisfaction level of inflight entertainment

On-board service: Satisfaction level of On-board service

Leg room service: Satisfaction level of Leg room service

Baggage handling: Satisfaction level of baggage handling

Check-in service: Satisfaction level of Check-in service

Inflight service: Satisfaction level of inflight service

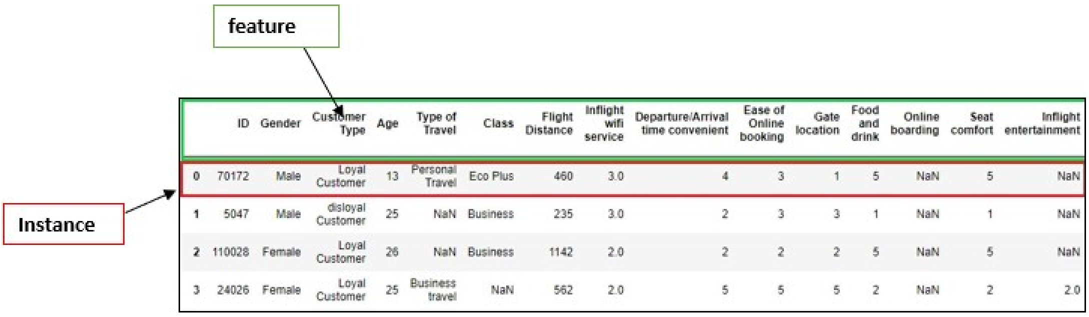
Cleanliness: Satisfaction level of Cleanliness

Departure Delay in Minutes: Minutes delayed when departure

Arrival Delay in Minutes: Minutes delayed when Arrival

Satisfaction: Airline satisfaction level(Satisfaction, neutral or dissatisfaction)

Figure 16: Feature Descriptions



	ID	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	Food and drink	Online boarding	Seat comfort	Inflight entertainment
0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3.0	4	3	1	5	NaN	5	NaN
1	5047	Male	disloyal Customer	25	NaN	Business	235	3.0	2	3	3	1	NaN	1	NaN
2	110028	Female	Loyal Customer	26	NaN	Business	1142	2.0	2	2	2	5	NaN	5	NaN
3	24026	Female	Loyal Customer	25	Business travel	NaN	562	2.0	5	5	5	2	NaN	2	2.0

Figure 17: The split between the features and the instances

B Results

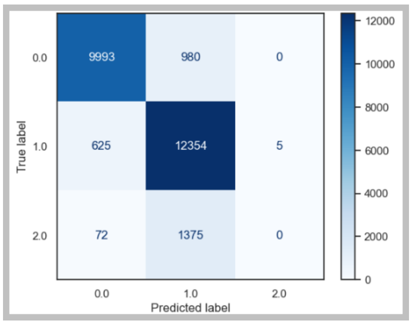


Figure 18: Random Forest Heat map

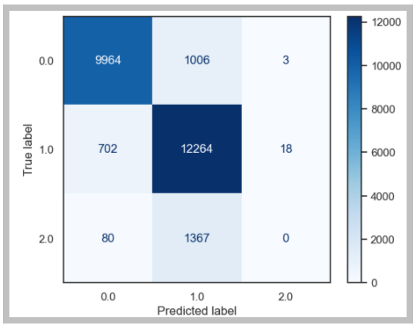


Figure 19: Radiant Boost Heat-map

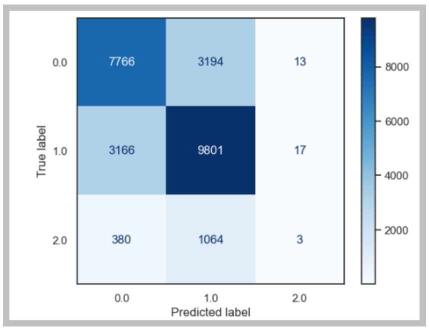
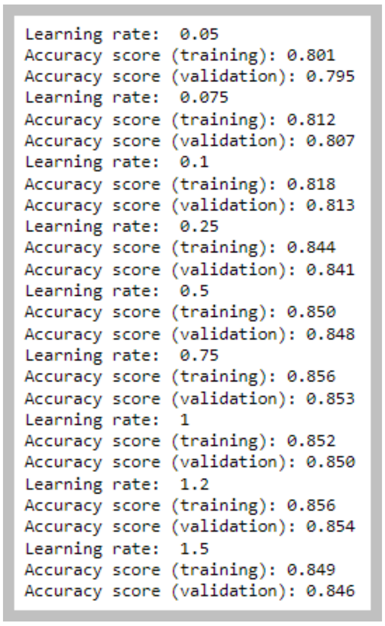
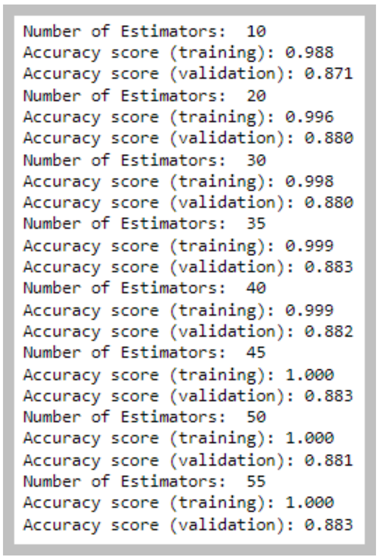


Figure 20: KNN_Heat-map



```
Learning rate:  0.05
Accuracy score (training): 0.801
Accuracy score (validation): 0.795
Learning rate:  0.075
Accuracy score (training): 0.812
Accuracy score (validation): 0.807
Learning rate:  0.1
Accuracy score (training): 0.818
Accuracy score (validation): 0.813
Learning rate:  0.25
Accuracy score (training): 0.844
Accuracy score (validation): 0.841
Learning rate:  0.5
Accuracy score (training): 0.850
Accuracy score (validation): 0.848
Learning rate:  0.75
Accuracy score (training): 0.856
Accuracy score (validation): 0.853
Learning rate:  1
Accuracy score (training): 0.852
Accuracy score (validation): 0.850
Learning rate:  1.2
Accuracy score (training): 0.856
Accuracy score (validation): 0.854
Learning rate:  1.5
Accuracy score (training): 0.849
Accuracy score (validation): 0.846
```

Figure 21: Gradient boosting Learning rate results



```
Number of Estimators:  10
Accuracy score (training): 0.988
Accuracy score (validation): 0.871
Number of Estimators:  20
Accuracy score (training): 0.996
Accuracy score (validation): 0.880
Number of Estimators:  30
Accuracy score (training): 0.998
Accuracy score (validation): 0.880
Number of Estimators:  35
Accuracy score (training): 0.999
Accuracy score (validation): 0.883
Number of Estimators:  40
Accuracy score (training): 0.999
Accuracy score (validation): 0.882
Number of Estimators:  45
Accuracy score (training): 1.000
Accuracy score (validation): 0.883
Number of Estimators:  50
Accuracy score (training): 1.000
Accuracy score (validation): 0.881
Number of Estimators:  55
Accuracy score (training): 1.000
Accuracy score (validation): 0.883
```

Figure 22: Number of estimators Hyper parameter results