SUBJECT CODE: DATA 6000

SUBJECT NAME: CAPSTONE: INDUSTRY CASE STUDIES

ASSESSMENT TITLE: PROJECT REPORT

STUDENT NAME: RAHULKUMAR RAJMAL SUKHWAL

STUDENT ID: 1846476

Table of Contents

[Executive Summary 3](#_Toc200122142)

[Industry Problem 3](#_Toc200122143)

[Data Processing and Management 4](#_Toc200122144)

[Data Analytics Methodology 5](#_Toc200122145)

[Visualisation and Evaluation of Results 6](#_Toc200122146)

[Recommendations 8](#_Toc200122147)

[Data Ethics and Security 9](#_Toc200122148)

[Data Source 9](#_Toc200122149)

[References 10](#_Toc200122150)

[Appendix 11](#_Toc200122151)

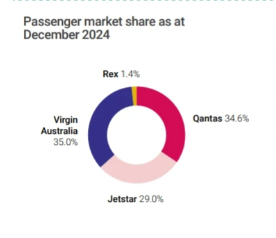
# Executive Summary

The main reason for this projects is to address the rise in customer dissatisfaction in Australian aviation caused by regular delays, cancellations and low service quality. A dataset with 25,976 records and 25 columns, focused on service and demographic parts, was used to apply descriptive and predictive analytics and find the main factors influencing dissatisfaction. Using charts, the study noticed differences in performance between the companies, while Random Forest and Logistic Regression models successfully predicted the happiness of customers. According to the research, guests are most concerned about cleanliness, the case of check-in and being on time. It is recommended that companies implement new service improvements and monitor their performance with predictive results to maintain happy and loyal customers.

# Industry Problem

The Australian aviation industry, which stands at $17 billion for 2024-25, is growing steadily each year at a rate of 4.7% (Ledovskikh, 2025). Growth in domestic travel and the addition of new routes are encouraging this development. Due to Rex Airlines collapsing, Virgin Australia now happens to be the largest domestic airline company with around a 35% share, overtaking Qantas. While the market is growing, the industry struggles because many customers are not satisfied with how it operates.

**Figure 1: Australian Domestic Airline Market share**



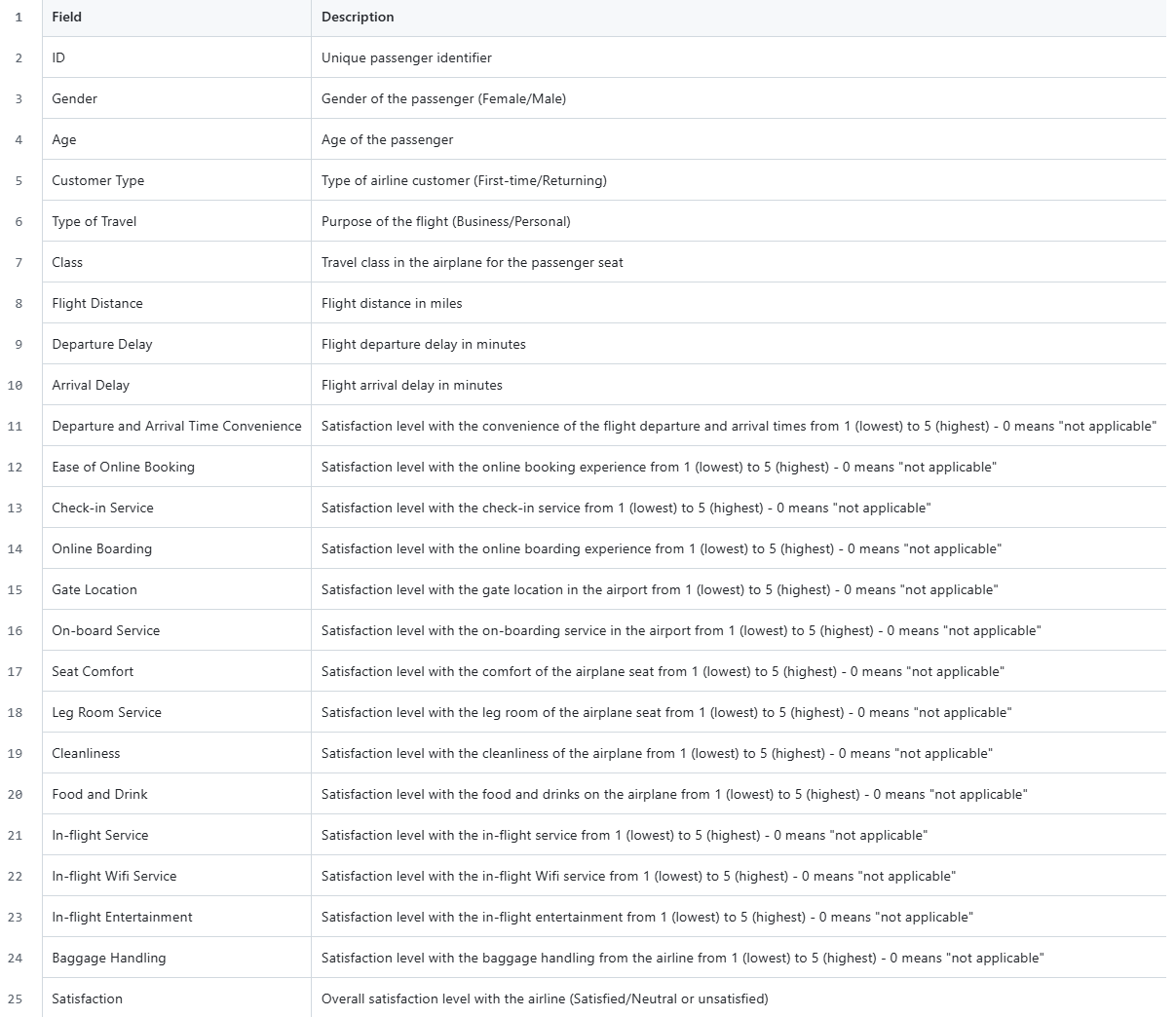
(Source: Hughes, 2025)

One of the biggest current problems in the industry is that passengers are becoming less satisfied, mainly because flights are often delayed, and a lot of journeys are cancelled. On-time performance in March 2024 was only 77.2%, compared to the regular average of 81%, and cancellations reached 5%. Problems during operations are also made worse by people rating airline check-in services, cabins and onboard services poorly. For example, Singapore Airlines reaches a satisfaction rating of 94.3%, while American Airlines only ranks 89.3% (Jones, 2025).

The main question of this project is “What techniques can be used by airlines to study and minimize customer dissatisfaction?”

Improving this problem will boost brand loyalty, increase competitiveness and sustain the growth of the domestic aviation industry. This study used a dataset with 25,976 records and 25 features covering demographics, type of service and how people travel. By using this organised data, airlines were able to identify patterns and forecast how customers could be unhappy, so they were able to address these issues and make improvements in advance.

**Figure 2: Dataset**

****

(Source: Self-Created)

# Data Processing and Management

Information on airline passenger satisfaction from 25,976 passengers is contained in the 25-columndatset used for this project. Two types of factors are covered, such as demographics including gender and age, along with travel aspects, such as flights distance, seat class and travel type. It also ranks different parts of the service, including check-in, the food offered, using online boarding, dealing with baggage and seat enjoyment. The column called “Satisfactions” is set up in two categories, including “Satisfied” and “Neutral or Dissatisfied”. As the structured data focuses on customer dissatisfaction, it is closely connected to the business issue.

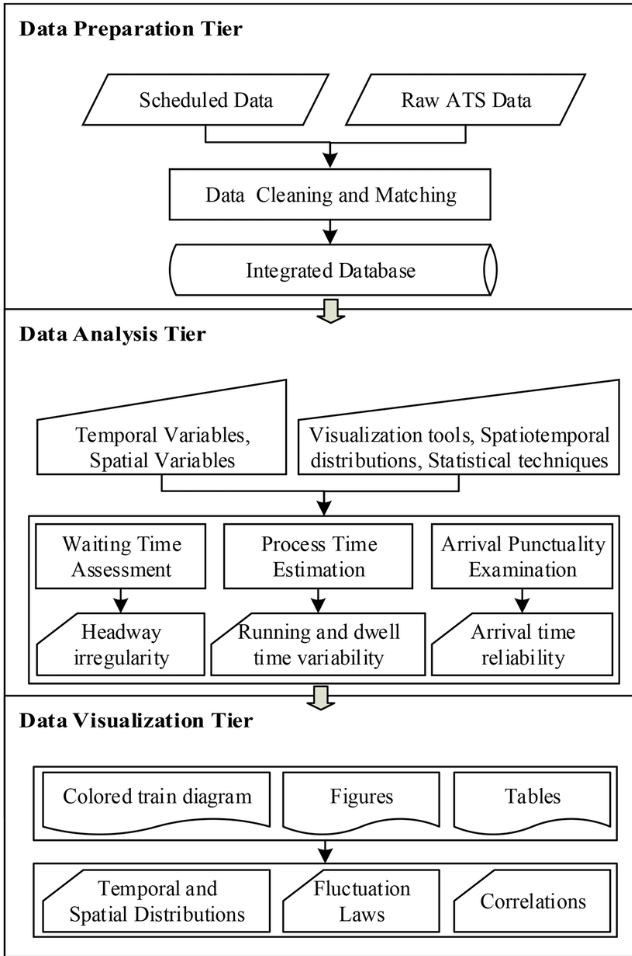
Descriptive and predictive analytics tools were used to meet the business questions. Descriptive analytics made it possible to look back on how the ratings and behaviour of customers have changed throughout the years. Comparing data using bar charts, pie charts and area charts in Power BI has contributed to learning that customers who are not satisfied rated the levels for check-in, cleaning standards and baggage services more negatively. The visuals helped to spot the main service issues that influence customer satisfaction.

Classification models were built using predictive analytics to predict how customers would rate their satisfaction. The Orange data mining platform was used to build machine learning models, such as Random Forest and Logistic Regression, predicting satisfaction based on the given features. Their use made it easier to find passengers who could not be satisfied, so action could be taken early.

Before creating models, the dataset was pre-processed in various ways. The columns of “Unnamed: 0” and “ID” were not needed and were removed from the table. In the “Arrival Delay in Minutes” column, 83 missing values were removed from the table. The travel type and class data were categorised, and those numbers were normalised as required to ensure the model performed correctly. As the data was clean and complete, it could be used for effective plotting and reliable forecasting model design. The effort put into data preparation played a big role in making the insights reliable.

# Data Analytics Methodology

**Figure 3: Flowchart**



(Source: Self-Created)

To analyse passenger satisfaction in the aviation industry, this study applies both descriptive analytics and predictive analytics with classification models. Applying this mixed method gives valuable knowledge about history and provides insights for the future, which is important for a situation such as dissatisfied customers.

The study used descriptive analytics to look at and summarise the data. It pays attention to locating the main trends, frequencies, and patterns in the data (Ticing, 2024). Visual tools were used by the study to make the unprocessed data simpler to understand and clearer for audiences to see how things such as check-in service, seat comfort and cleanliness are distributed between various passenger groups. Descriptive analytics helped in understanding what kinds of service quality issues the data involved.

A significant part of the study used predictive analytics with machine learning classification models. In the case of a categorical target, classification is the proper approach. People can report satisfaction on the survey, with choices being “Satisfied” and “Neutral/Dissatisfied”. Logistic Regression and random Forest are two algorithms from Orange that the study included. Each method was considered suitable due to its benefits, including that Logistic Regression is easy to understand, and Random Forest is reliable with complex data. Being easy to use, reliable and with a convenient visual interface, Orange is a perfect tool for trying out and comparing numerous models by simply dragging them into place (Dobesova, 2024).

As both descriptive and predictive approaches were included, the problem could be assessed in an organised manner. Descriptive analysis helped find what causes satisfaction, and predictive modelling was employed to build a tool that predicts how much customers would be satisfied based on earlier information.

Applying these steps together, the process can be improved by running the analysis again. Once new information appears, including details from new travel seasons or what customers suggest, pressing the same process can show how satisfied they remain. It is also possible to easily update models using Orange and Power BI helps by regularly monitoring updates dashboards. The approach is suitable for many business problems. Using both history-based and predictive methods, airlines are better able to improve their services and mange the experiences of their customers today more than ever.

# Visualisation and Evaluation of Results

The project used different visual tools to extract and explain insights from both types of analytics. To show the patterns in service ratings by customer groups, bar charts and pie charts were used in Power BI. Satisfaction with cleanliness, the check-in process and food onboard were rated lower consistently by passengers who said they were dissatisfied. As a case in point, satisfied customers gave cleanliness a score of 4.09 and check-in service 3.74, whereas dissatisfied passengers rated them as low as 2.87. They played a key role in finding out what causes customers to feel unhappy with the services provided.

**Figure 4: Descriptive Visuals**

A graph of blue rectangular bars

AI-generated content may be incorrect.A graph with blue lines and numbers

AI-generated content may be incorrect.

A screenshot of a graph

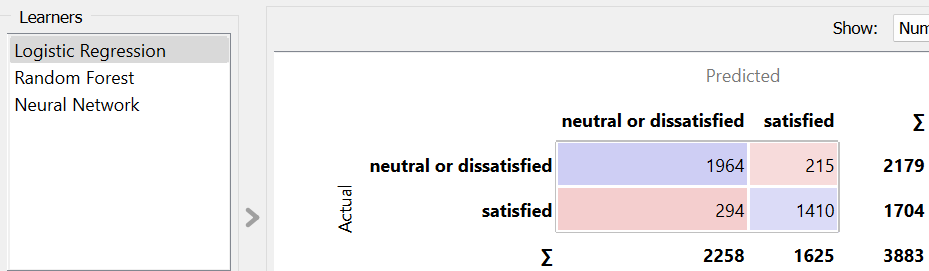
AI-generated content may be incorrect.A graph of blue bars

AI-generated content may be incorrect.

(Source: Self-Created)

Results from predictive analytics were shown using evaluation metrics generated using Orange. Some important models were tested, including Logistic Regression and Random Forest. The Random Forest model gave top results, yielding 95.6% of correct classifications and an AUC of 0.991. by looking at the visual results of each algorithm, the study was able to distinguish which algorithm was the strongest for forecasting satisfaction. Random Forest showed precision and recall that supported the effective detection of both the satisfied and dissatisfied customer emotions. This data showed that the models were able to pick out important satisfaction indicators from the given dataset.

**Figure 5: Predictive Visuals**



A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

(Source: Self-Created)

This project was carried out effectively using Power BI and Orange. Power BI made it possible to create interactive dashboards that made understanding and sharing service metrics much easier. Orange’s easy-to-arrange interface reduced time spent on building, training and testing classification models, so I could easily compare them. As a result, these tools provided a complete and logical environment where users could find answers and make predictions. Making use of the visuals and metrics made it easier to find the reasons behind problems and to trust in the performance of the various models.

Using charts and numbers contributed to simplifying the data and sharing key insight regarding the aviation industry and its diverse aspects. Using visual tools made it easier to understand the information and use it to fix customer satisfactions issues in the airline business.

# Recommendations

With the findings and visual outputs, some solutions can be proposed to solve the rising dissatisfaction of customers in the Australian aviation sector. The use of visuals and graphics revealed that dissatisfied customers found cleanliness, check-in and the food less satisfactory. Airline companies can improve their services in these areas by training staff, updating operating routines and handling facility improvements. As business travellers were more satisfied, particularly with online booking and longer flights, making personal and economic travel similar needs led to better ratings for hotels.

The accuracy of the models, especially Random Forest, was extremely high in locating dissatisfied passengers using data such as travel type, how late the plane arrived and ratings of the airline’s services. Based on these insights, airlines can create systems that recognise early warning signs of dissatisfied passengers and offer them special treatment, including quicker check-ins, extra messages or some kind of compensation. Well-thought-out targeted actions, based on predictive analytics, can lower the chance of bad experiences and increase customer loyalty.

The data and models contributed to giving a meaningful understanding. However, they had some key drawbacks. Since the dataset had a single missing column, it was solved, but its outdated nature cannot show modern changes in the market and what consumers now expect. The classification method employed numerical ratings and could not take into account qualitative information about the reasons for dissatisfaction or customers’ emotions.

Data analytics helped detect and predict different patterns of dissatisfaction, making it possible to make strong management decisions with facts at hand. It is important to add social media analysis and real-time feedback systems in future steps to discover more about customers’ feelings (Idris and Mohamad, 2023). NLP techniques could be used to make sense of unstructured reviews and feedback, so the company can understand passenger sentiment better (Samir, Abd-Elmegid and Marie, 2023).

Using data to make decisions is key to improving services in the aviation industry (Bakir et al., 2025). By using constant feedback with descriptive and predictive data, airlines will become more responsive, adaptive and in line with what customers require.

# Data Ethics and Security

Using data ethically is very important while studying customer satisfaction, mainly when handling personal and behavioural details (Okorie et al., 2024). Though the data in the dataset was anonymous, all legal and privacy standards needed to be followed. The visualisations in Power BI were precise, made with tidy information and described customer input without affecting its quality. Going ahead, when using real-time or personal data from social media, it will be important to comply closely with privacy rules and obtain permission from users. In the future, data ethics needs to shape analytical work to ensure information is clear, decisions are fair, and people’s privacy is maintained (Junsree, 2023).

# Data Source

1. <https://github.com/aryakghosal/airline-passenger-satisfaction-analysis/blob/master/Dataset/data_dictionary.csv>
2. <https://github.com/Rahulkumarsukhwal/Data6000-capstone-project/upload/main>

# References

Ane, Y. 2022. *The 6 Steps of Predictive Analytics*. [online] Analytics Vidhya. <<https://www.analyticsvidhya.com/blog/2022/09/the-6-steps-of-predictive-analytics/>.>

Bakir, N., Nemar, S.E., Gehchan, F.B., Samrouth, K. and Vrontis, D. 2025. Optimizing airline services: leveraging data-driven strategies for enhanced customer satisfaction and engagement. *Journal of Asia Business Studies*. <<https://doi.org/10.1108/jabs-09-2024-0559>.>

Dobesova, Z. 2024. Evaluation of Orange data mining software and examples for lecturing machine learning tasks in geoinformatics. *Computer applications in engineering education*. <<https://doi.org/10.1002/cae.22735>.>

Hughes, D. 2025. *Virgin Australia Beats Qantas by 0.4% to Become Largest Domestic Airline*. [online] Aviation A2Z. <<https://aviationa2z.com/index.php/2025/02/18/virgin-australia-becomes-largest-domestic-airline/>.>

Idris, S.L. and Mohamad, M. 2023. A Study on Sentiment Analysis on Airline Quality Services: A Conceptual Paper. *Information management and business review*, 15(4(SI)I), pp.564–576. <<https://doi.org/10.22610/imbr.v15i4(si)i.3638>.>

Jones, G. 2025. *Singapore Airlines named Roy Morgan Research 2024 International Airline of the Year*. [online] Travel Weekly. <<https://travelweekly.com.au/singapore-airlines-named-roy-morgan-research-2024-international-airline-of-the-year/>>

Junsree, K. 2023. *Ethics in Data Analysis: Transparency, Privacy, and Fairness*. [online] Medium. <<https://kritjunsree.medium.com/ethics-in-data-analysis-transparency-privacy-and-fairness-b67e120fb9e1>.>

Ledovskikh, A. 2025. *Domestic Airlines in Australia - Market Research Report (2014-2029)*. [online] Ibisworld.com. <<https://www.ibisworld.com/australia/industry/domestic-airlines/472/>.>

Okorie, G.N., Udeh, C.A., Adaga, E.M., DaraOjimba, O.D. and Oriekhoe, O.I. 2024. ETHICAL CONSIDERATIONS IN DATA COLLECTION AND ANALYSIS: a REVIEW: INVESTIGATING ETHICAL PRACTICES AND CHALLENGES IN MODERN DATA COLLECTION AND ANALYSIS. *International Journal of Applied Research in Social Sciences*, 6(1), pp.1–22. <<https://doi.org/10.51594/ijarss.v6i1.688>.>

Samir, H.A., Abd-Elmegid, L.A. and Marie, M.M. 2023. Sentiment analysis model for Airline customers’ feedback using deep learning techniques. *International journal of engineering business management*, 15. <<https://doi.org/10.1177/18479790231206019>.>

Ticong, L. 2024. *4 Types of Data Analytics To Enhance Your Decisions*. [online] Datamation. <<https://www.datamation.com/big-data/types-of-data-analytics/>.>

# Appendix

|  |  |
| --- | --- |
| **Section** | **Details** |
| **Descriptive Analytics Process** | * Summarized historical data on customer satisfaction, focusing on key service features like check-in, cleanliness, food, and baggage handling. * Used bar charts, pie charts, and area charts to visually identify patterns and trends in satisfaction ratings. * Highlighted lower ratings in check-in service, cleanliness, and food as primary drivers of dissatisfaction. |
| **Predictive Analytics Process** | * Built machine learning classification models to predict customer satisfaction (Satisfied vs. Neutral/Dissatisfied). * Models used: Logistic Regression, Random Forest, * Trained on features such as flight distance, travel class, type of travel, and service ratings. |
| **Software Used** | **Power BI:** Created descriptive visualisations (bar, pie, area charts) for clear insights.  **Orange Data Mining:** Built and evaluated classification models, compared algorithm performance using an interactive, drag-and-drop interface. |
| **Limitations** | * The dataset can not fully reflect current market dynamics (dated nature). * Only numeric and categorical features are included without any qualitative feedback, such as customer comments. * Social media sentiment and real-time data are not yet integrated. * Missing values in “Arrival Delay in Minutes” (83 nulls) required imputation. |