

Group 3: Britney Odria, Brett McKeehan, Ngoc Pham, Satyam Palkar and Arnab Tiwari

The sky's the limit: How Australia's airline industry can identify inefficiencies and improve passenger satisfaction

Business problem:

With more than 57 million passenger trips recorded in 2023 alone (BITRE, 2024), the aviation industry is a pivotal part of Australia's transport network. Competition is fierce, the rewards immense – Qantas, the nation's flagship carrier, generated \$19.8 billion in revenue last year (Hughes, 2023). Yet profitability is far from guaranteed. Extreme weather, holiday bottlenecks, fuel prices, customer satisfaction and the knock-on effects of operational inefficiencies all present significant challenges to airlines, and optimisation is critical for long-term success. Leveraging official data, this project will analyse internal and external factors influencing airline deliverables and propose a series of actionable improvements for stakeholders to minimise disruptions, maximise profits and help offset operational instability.

Data overview:

Dataset 1: Domestic airline performance, 2004-2024

Source and dimensions: [Australian government](#); 14 features, 101,248 records

Description: This dataset offers a clear picture of the Australian aviation industry's performance over decades, with cancellations and delays key features for analysis. Such comprehensive time-series statistics will enable the research team to identify patterns of delayed and cancelled flights, benchmark performance across different airlines and airports, and explore strategies to minimise disruptions and improve the overall travel experience.

Dataset 2: Passenger transit by airport, 2004-2022

Source and dimensions: [Australian government](#); 17 features, 4467 records

Description: This dataset focuses on air passenger movements across Australia over 18 years, offering researchers a clear picture of the impacts of consistent bottlenecks, holidays and potentially significant global or economic events on airport operations. These insights are crucial for airports to optimise schedules and perhaps plan for infrastructure upgrades.

Dataset 3: Short-term visitors arrivals, 1991-2024

Source and dimensions: [Australian Bureau of Statistics](#); 11 features, 577 records

Description: This dataset contains numerous features representing different aspects of short-term visitor stays, ranging from less than a week to less than a year, as well as the reasons for journey. This categorical data is valuable to assess global passenger trends into Australia and the impact of seasonality on tourism and the broader aviation industry.

Research questions

1. Which routes have the highest frequency of delays?
2. What are the peak times for passenger traffic, and how do they align with the highest instances of flight delays and cancellations?
3. Can we predict flight delays based on airport operational data and passenger traffic trends?

Methodology

Data preparation

- Cleaning: Delete duplicates, check for missing values, identify anomalies or outliers.
- Transformation: Normalise numerical values, encode categorical, feature scaling.
- Feature engineering: Focus on characteristics with a strong correlation to the target variable; develop additional features as required.

Exploratory data analysis (EDA)

- Descriptive statistics: Generate feature assessment metrics using Python.
- Visualisations: Create plots, histograms and heatmaps to explore the distribution, relationships and trends within the data and detect outliers using Python.
- Geographical analysis: Explore geospatial patterns using Tableau.

Statistical analysis and modelling

- Regression: Explore the impact of dependent features to the target using linear, logistic (L1, L2) and ElasticNet models.
- Time-series: Track seasonality and patterns over time, and anticipate future trends
- Machine learning: Random Forest, Decision Trees and K-Nearest Neighbours (KNN) to predict delays and cancellations.

Validation and testing

- K-fold cross-validation: Tests prediction models on unknown data.
- Performance metrics: RMSE, MAE, R-square etc for regression tasks; accuracy, precision, recall, F1-score for classification tasks.
- Model tuning: Use gridsearchCV or random search for hyperparameter optimisation.

Interpretation of results

- Plot geospatial visualisations to identify regular flight bottlenecks across Australia and define what's working and what's not for airlines and airports.
- Suggest operational improvements so airports can handle more passengers.
- Mitigate the impact of weekends and holidays to help airlines minimise delays.

Conclusion

Vigorous investigation, analysis and interpretation of public aviation datasets offer significant opportunities for operational improvements – potentially impacting millions of travellers across Australia every month. Our research explores and visualises a wide variety of factors to develop a cross-functional understanding of industry dynamics. By leveraging data analysis, we aim to establish a more efficient and customer-centric aviation ecosystem.

References

Hughes, Christopher (2023), 'Total revenue of Qantas in Australia from the financial year 2015 to 2023' in Statista.com. [Accessed 16/03/2024](#)

The Bureau of Infrastructure and Transport Research Economics (2024) 'Domestic aviation activity'. [Accessed 16/03/2024](#)