

FORECASTING ON-TIME PERFORMANCE



OF AUSTRALIAN DOMESTIC AIRLINES
IN JUNE 2024



QUYNH THO CHU

TABLE OF CONTENTS

I. EXECUTIVE SUMMARY	3
II. INDUSTRY PROBLEM	4
III. DATA PROCESSING AND MANAGEMENT	8
IV. DATA ANALYTICS METHODOLOGY	10
V. VISUALISATION AND EVALUATION OF RESULTS	11
VI. RECOMMENDATIONS	14
1. Data insights	14
2. The role of data analytics	16
3. The limitations	16
4. Other analytics techniques	16
VII. DATA ETHICS AND SECURITY	17
REFERENCE LIST	18

I. EXECUTIVE SUMMARY

On-time performance is one of the most critical factors in determining airlines' customer service reliability and competitiveness in the Australian domestic aviation industry. A such, this project used datasets on Australian Bureau of Infrastructure and Transport Research Economics (BITRE) website to **forecast the on-time performance of Australian domestic airlines in June 2024**. In this study, ARIMA and VAR models are applied to predict three significant variables: On Time Departures rate (%), On Time Arrivals rate (%) and Cancellations rate (%).

The results demonstrate that the Australian domestic airlines will not have any significant improvement in their on-time performance until June 2024. If airline companies fail to take this issue into account, they may lose their passengers or cease operations due to poor reliability and low competitive ability.



II. INDUSTRY PROBLEM

The airline industry is currently facing major challenges due to aircraft cancellations and delays, as shown in Figures 1, 2, and 3. It significantly affects passengers, businesses, transportation, and the global travel economy.

The Airports To Avoid This Summer

Airports with most delayed flights that had been scheduled between May 26-July 19, 2022 (in %)



Figure 1: Airports where the most flights are delayed between May 26 and July 19, 2022 (Anna Fleck 2022)

Thousands of Flights Canceled & Delayed Over July 4 Break

Canceled and delayed flights landing at/departing from U.S. airports between July 2, 5am and July 4, 4am 2022

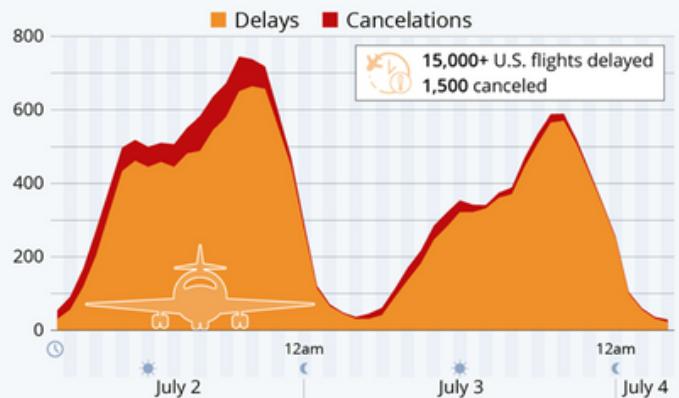


Figure 2: Flights at US airports that are delayed or cancelled between July 2, 2022, and July 4, 2022 (Anna Fleck 2022)

Grounded Flights

Europe's biggest hubs see rising cancellations

■ Cancelled Flights

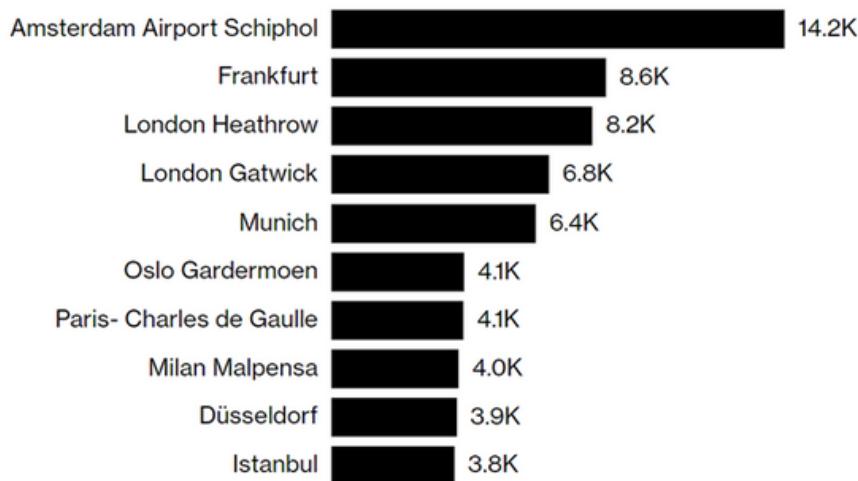


Figure 3: Number of cancellations in Europe between April 1, 2022 and June 29, 2022 (Siddharth Vikram Philip 2022)

According to the BITRE report, the on-time performance over all routes in November 2023 of Australian domestic airlines was significantly lower than the long-term average performance for all routes (BITRE 2023). The following Figures 4 and 5 also indicate that this has not been significantly improved after the Covid-19 pandemic.

Transport Minister Catherine King also criticized for the worst on-time performance and alarming amounts of cancellations in over a year of Australia's domestic airlines. These incredibly disappointing results reflect why so many passengers are dissatisfied with Australian Airlines (Amelia McGuire 2023).

In order to monitor the monthly punctuality and reliability of major domestic airlines, the BITRE has collected reports from them and published them on its website. However, the predictions and solutions for increasing the on-time performance of domestic airlines have not been reported (BITRE 2023). As such, this project looks up information related to Australian domestic aviation and discovers datasets on the websites of BITRE and Kaggle.com which include flight cancellation and delay figures.

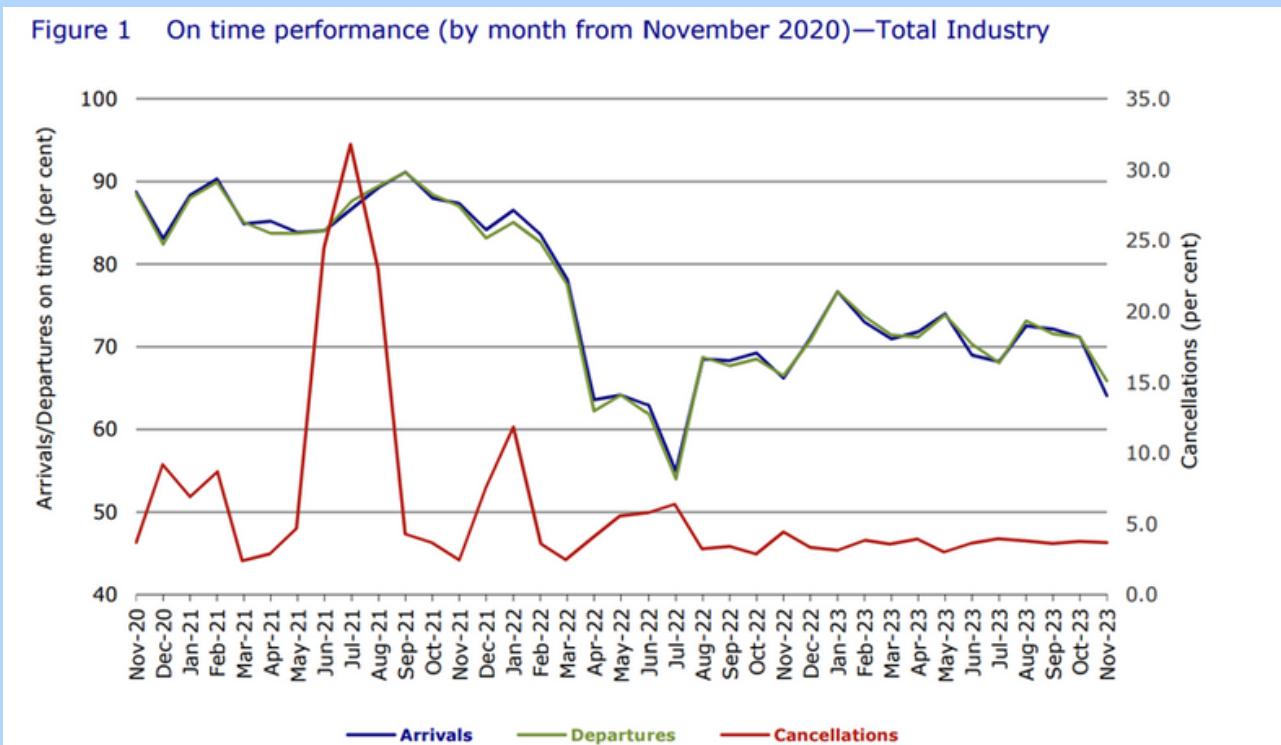


Figure 4: On-time performance of total industry in Australia from Nov 2020 to Nov 2023 (BITRE 2023, p.4)

Figure 4 Cancellations (by month from November 2022)

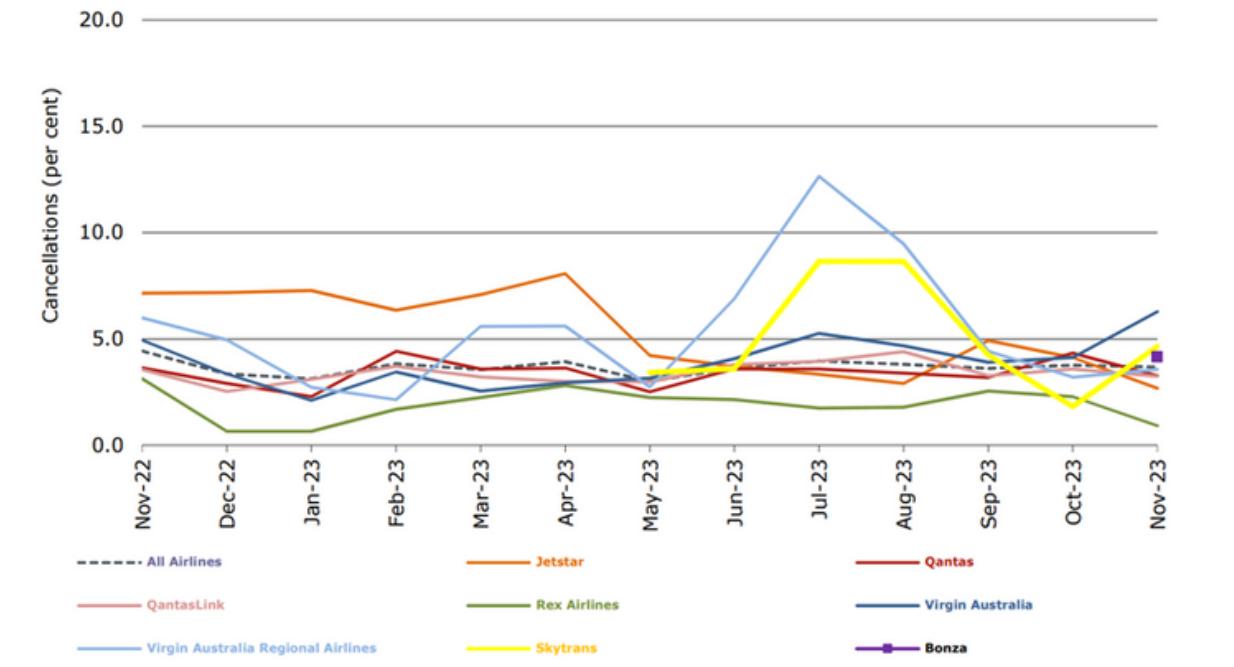


Figure 5: Cancellations of major Australian airlines from Nov 2022 to Nov 2023
(BITRE 2023, p.5)

The datasets from the BITRE website are used in this project because they are more recent and contain important variables that can be used to forecast the on-time performance of airlines in the future. Nevertheless, these datasets include unnecessary variables and are divided up into separate sheets and files which make it difficult to use models in forecasting.

In response to my colleague's feedback, I eliminated extraneous data and combined all of the sheets and files into a single file with necessary data that is better for forecasting. Additionally, my lecturer advised me to combine the ARIMA and VAR models predictions into a single table or graph to facilitate better comparison.

Based on the forecasting results, the industry may assess and put into practice methods to enhance the punctuality of Australian airlines and assist the government in developing policies to protect passengers. Furthermore, travelers can select their best airlines based on the on-time performance predictions for each major airline.

III. DATA PROCESSING AND MANAGEMENT

The datasets are collected from Australian domestic airlines by BITRE, whose analysis and statistics on infrastructure and transport-related issues have assisted the Australian Government in policy development and fostering greater community understanding (BITRE 2023). They include two files with separate data sheets from 2003 to 2023, which contain the total number of flight cancellations, delays, and on-time departures rate, on-time arrivals rate, cancellations rate by month, year of all ports, and each route of all airlines (BITRE 2023).

Route	Departing Port	Arriving Port	Airline	Month	Sectors Scheduled	Sectors Flown	Cancellations	Departures On Time	Arrivals On Time	Departures Delayed	Arrivals Delayed	OnTime Departures (%)	OnTime Arrivals (%)	Cancellations (%)
All Ports-All Ports	All Ports	All Ports	All Airlines	Nov-03	32 560	32 341	219	28 439	28 494	3 902	3 847	87.9	88.1	0.7
All Ports-All Ports	All Ports	All Ports	Qantas	Nov-03	9 884	9 823	61	8 484	8 687	1 339	1 136	86.4	88.4	0.6
All Ports-All Ports	All Ports	All Ports	QantasLink	Nov-03	10 591	10 469	122	9 196	9 099	1 273	1 370	87.8	86.9	1.2
All Ports-All Ports	All Ports	All Ports	Regional Express	Nov-03	4 245	4 229	16	3 666	3 714	563	515	86.7	87.8	0.4
All Ports-All Ports	All Ports	All Ports	Skywest	Nov-03	712	712	0	687	687	25	25	96.5	96.5	0.0
All Ports-All Ports	All Ports	All Ports	Virgin Australia	Nov-03	7 128	7 108	20	6 406	6 307	702	801	90.1	88.7	0.3
All Ports-All Ports	All Ports	All Ports	All Airlines	Dec-03	33 412	33 170	242	28 175	28 229	4 995	4 941	84.9	85.1	0.7
All Ports-All Ports	All Ports	All Ports	Qantas	Dec-03	10 352	10 273	79	8 620	8 897	1 653	1 376	83.9	86.6	0.8
All Ports-All Ports	All Ports	All Ports	QantasLink	Dec-03	10 401	10 262	139	8 702	8 535	1 560	1 727	84.8	83.2	1.3
All Ports-All Ports	All Ports	All Ports	Regional Express	Dec-03	4 647	4 634	13	3 865	3 892	769	742	83.4	84.0	0.3
All Ports-All Ports	All Ports	All Ports	Skywest	Dec-03	690	690	0	669	669	21	21	97.0	97.0	0.0
All Ports-All Ports	All Ports	All Ports	Virgin Australia	Dec-03	7 322	7 311	11	6 319	6 236	992	1 075	86.4	85.3	0.2
All Ports-All Ports	All Ports	All Ports	All Airlines	Jan-04	31 913	31 685	228	27 599	27 475	4 086	4 210	87.1	86.7	0.7
All Ports-All Ports	All Ports	All Ports	Qantas	Jan-04	10 221	10 147	74	8 861	9 082	1 286	1 065	87.3	89.5	0.7
All Ports-All Ports	All Ports	All Ports	QantasLink	Jan-04	10 008	9 884	124	8 762	8 525	1 122	1 359	88.6	86.3	1.2
All Ports-All Ports	All Ports	All Ports	Regional Express	Jan-04	3 798	3 789	9	3 265	3 205	524	584	86.2	84.6	0.2

Figure 6: The screenshot of the dataset

To identify the trend and seasonality of the on-time performance, this project uses Power BI to efficiently represent and analyze the changes in historical periods. Then time series forecasting methods like ARIMA and VAR models are applied to solve the problem.

In the datasets, there are different values like the value for each specific route, the value for All Ports, the value for All Airlines, and the value for each airline. The datasets are combined and processed to create a data file with 4 columns and 253 rows that includes the On Time departures (%), On Time arrivals (%), and Cancellation (%) of the All Ports and the All Airlines from November 2003 to November 2023. This file is imported into Orange software to forecast the on-time performance of domestic airlines.

Month	Nov-03	Dec-03	Jan-04	Feb-04	Mar-04
On-time Departures (%)	87.9	84.9	87.1	90.1	91.6
On-time Arrivals (%)	88.1	85.1	86.7	89.3	91.1
Cancellations (%)	0.7	0.7	0.7	0.6	0.7

Figure 7: The sample of data imported into Orange

IV. DATA ANALYTICS METHODOLOGY

Power BI is a platform that provides tools to link unrelated datasets, clean and modify the data (Jesse Scardina 2024). Moreover, it gives insights through visualizations and support organizations in forecasts to solve problems and achieve their goals. With the monthly data in many years, which were stored in separated files and sheets, Power BI is useful to link the datasets and visualize data continuously by month or year. From the visualizations, we can identify the seasonality of the data, predict the trend and tell stories.

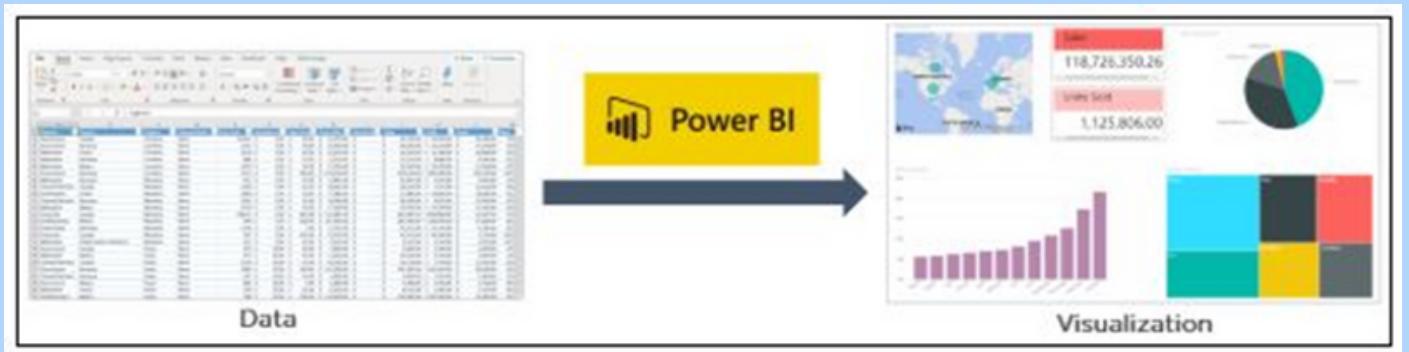


Figure 8: Power BI create graphs to visualize the data (Avijeet Biswal 2023)

ARIMA model is a technique that uses historical time series to forecast future values. Especially, it can be applied to non-stationary data by using differencing (Selva Prabhakaran 2024). The datasets in this project are the monthly data and are not stationary, that are applicable to use the ARIMA model for predicting the on-time performance of airlines in the next months.



VAR model is a statistical method, which also be applied for predictions. This method has an advantage of not imposing any restrictive assumptions on the distribution or structure of the data. So the non-stationary data can be forecasted by VAR model after applying differencing (LinkedIn 2023). In addition, VAR model is used in this project to compare with ARIMA model for outcome forecasting in order to assess the models' accuracy and select the more effective method for resolving the business problem.

V.VISUALISATION AND EVALUATION OF RESULTS

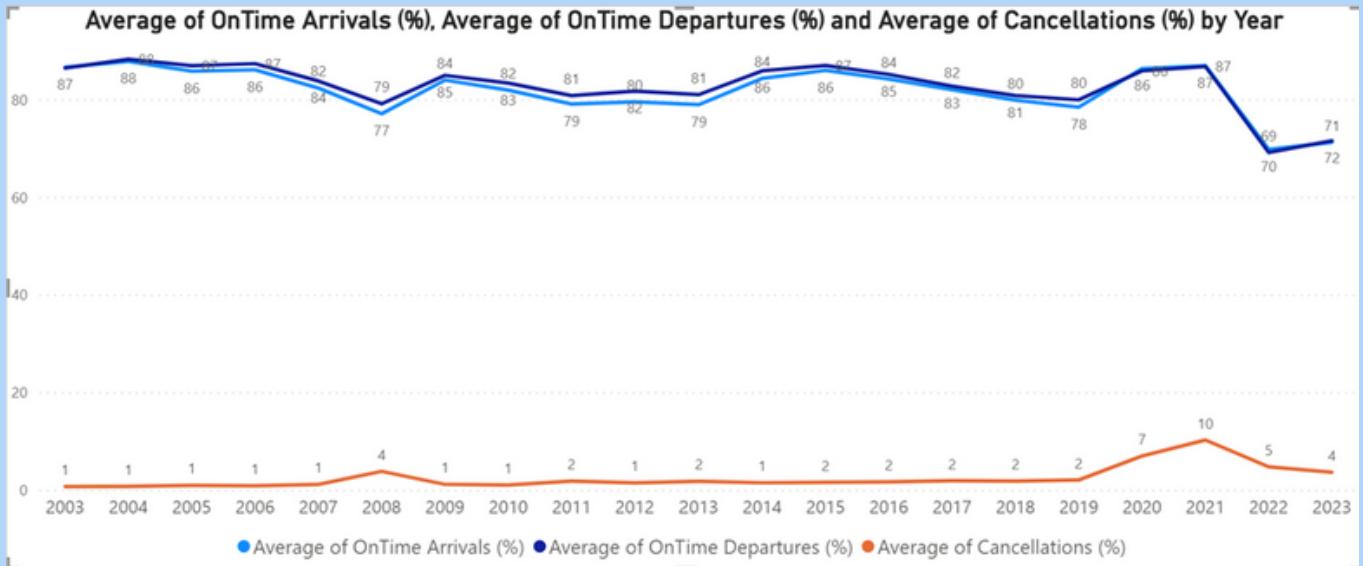


Figure 9: The Average of Cancellations (%), On time Departures (%) and On time Arrivals (%) by Year graph created by Power BI

The line chart shows the changes in average of cancellations rate, On time departures rate and On time arrivals rate by year of all airlines. It reflects that the on time performance of all airlines was high from 2003 to 2007 but then decreased in 2008 may be due to the Great Financial Crisis of 2007–2008. Especially, it was low in 2011 and in 2019 because of the Chilean volcanic eruption and the Covid-19 pandemic, respectively.

The on time departures rate and on time arrivals rate seemed to be better during 2020–2021 while the cancellations rate was high in this pandemic period because fewer flights were taken overall during those years, and fewer flights experienced delays.

After the pandemic, airlines' on-time performance increased in 2022–2023, with a decrease in cancellation rates and an increase in on-time arrivals and departures. Based on the above analysis, we found that this visual might help us to identify that Australian domestic airlines' on-time performance will be on the rise in 2024.

ON-TIME DEPARTURES RATE		
	VAR Model	ARIMA Model
1	67.3088	66.9765
2	67.1121	66.7434
3	67.2074	66.7925
4	67.1827	66.7821
5	67.1483	66.7843
6	67.0909	66.7839
7	67.0214	66.784

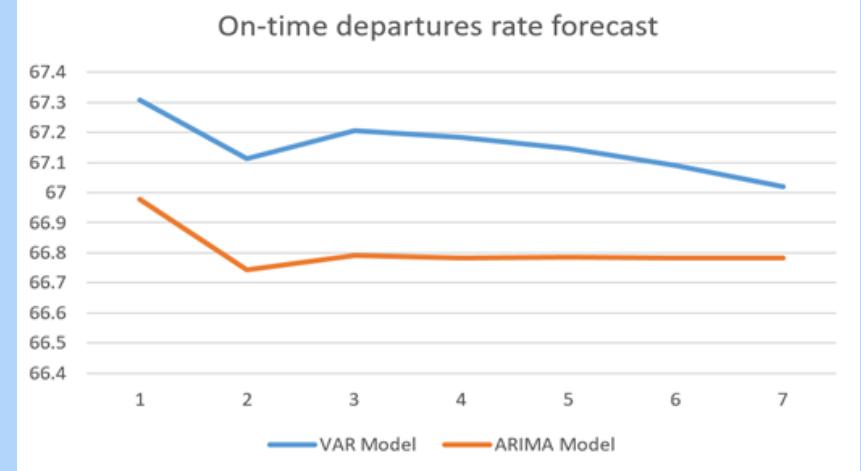


Figure 10: Predictions of VAR model and ARIMA model for on-time departures rate in the next 7 months from Orange

ON-TIME ARRIVALS RATE		
	VAR Model	ARIMA Model
1	66.1065	65.6204
2	65.8788	65.2955
3	66.0608	65.3653
4	66.0762	65.3503
5	66.0788	65.3535
6	66.0477	65.3529
7	65.9983	65.353

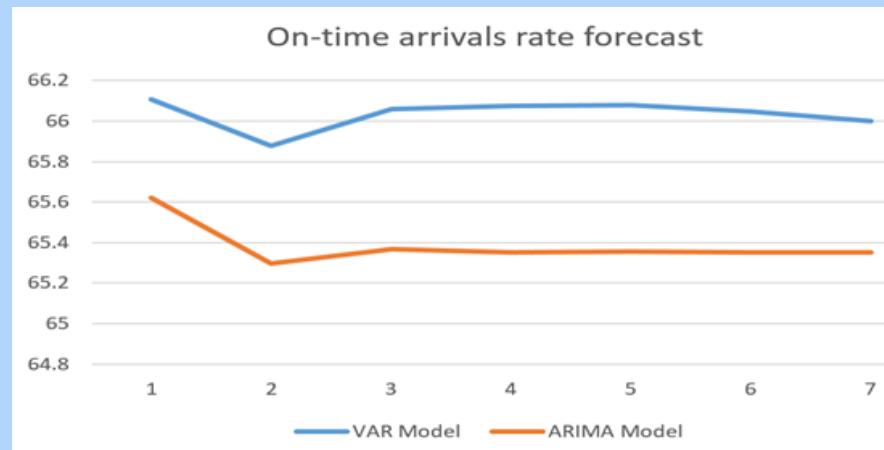


Figure 11: Predictions of VAR model and ARIMA model for on-time arrivals rate in the next 7 months from Orange

CANCELLATIONS RATE		
	VAR Model	ARIMA Model
1	2.97543	3.69302
2	2.49066	3.69277
3	2.257	3.69279
4	2.16726	3.69278
5	2.14199	3.69278
6	2.14028	3.69278
7	2.14475	3.69278

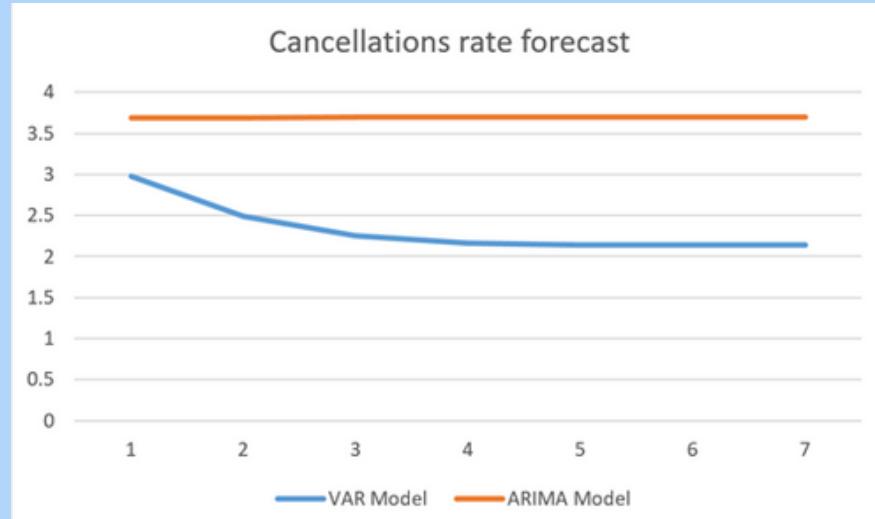


Figure 12: Predictions of VAR model and ARIMA model for cancellations rate in the next 7 months from Orange

According to the above numbers and visualizations in figures 10, 11 and 12, we can see that the outcomes of ARIMA model are mostly the same after three months. So it is unable to foresee the changes in the long period. To find better results for the business problem, the VAR model is also applied to compare with the ARIMA model. These figures indicate effectively how the VAR model is more efficient in predicting results over an extended period of time, as evidenced by the fluctuating line graph and changing numbers.

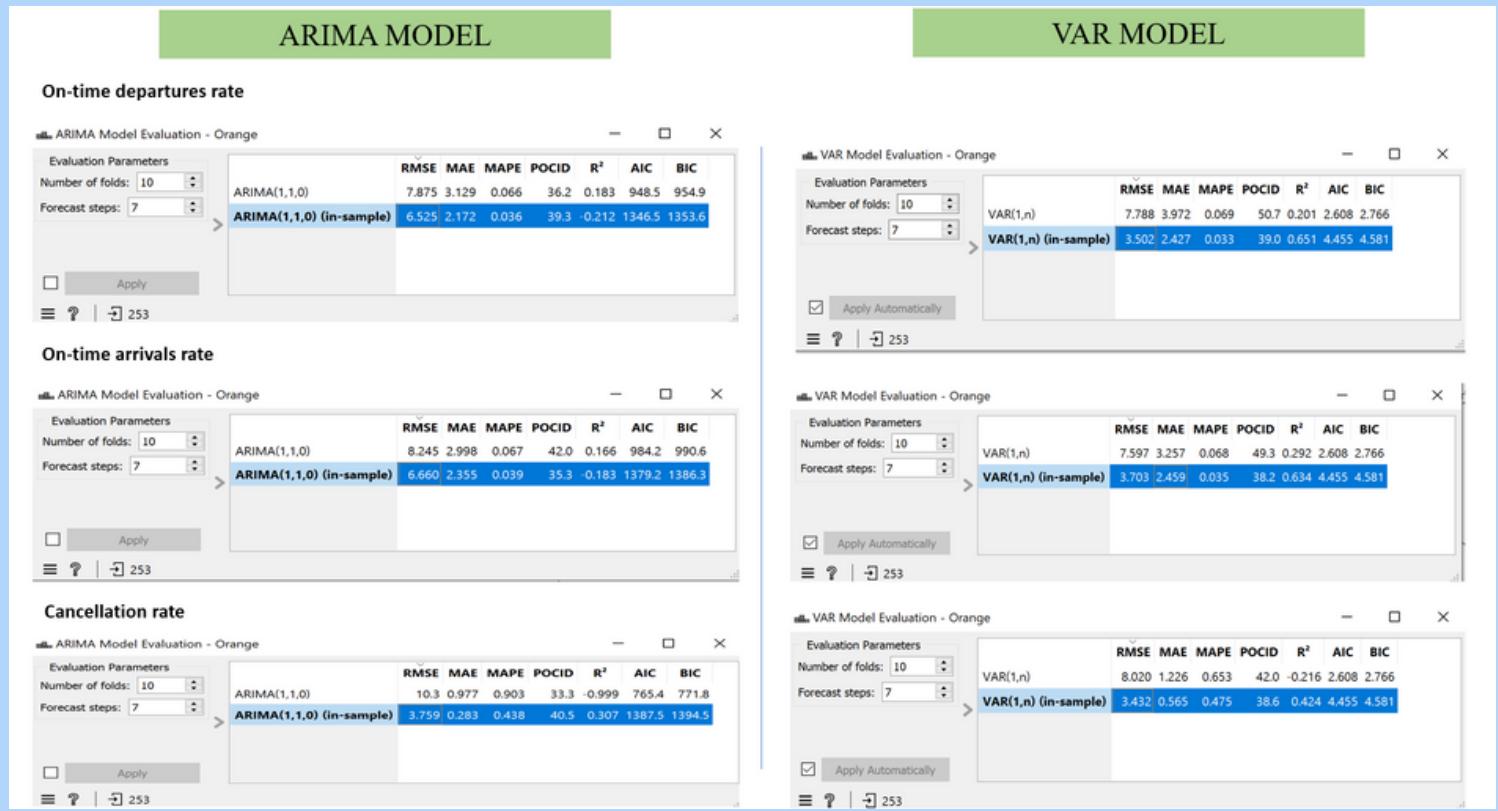


Figure 13: The evaluation metrics of ARIMA and VAR Model's from Orange

Based on the evaluations metrics in Figure 13, the VAR model with the smaller MAPEs is more accurate in forecasting the on time departures rate and on time arrivals rate. However, it is not as accurate as ARIMA model in forecasting the cancellations rate of domestic airlines because of the higher MAPE metric.

VI.RECOMMENDATIONS



1. Data insights

ON-TIME PERFORMANCE OF AUSTRALIAN DOMESTIC AIRLINES IN JUNE 2024

No .	On time performance	November 2023	June 2024- ARIMA Model	June 2024- VAR Model	Long term average of industry
1	On Time Departures rate	65.9%	66.78%	67.02%	81.2%
2	On Time Arrivals rate	64.1%	65.35%	65.99%	82.3%
3	Cancellations rate	3.7%	3.69%	2.14%	2.2%

The predictions of ARIMA and VAR model for the on-time performance of Australian domestic airlines in the table of Figure 13 indicate that:

- The Australian Airlines' on-time performance in June 2024 will be slightly better than its performance in November 2023.
- The on-time departures rate and the on-time arrivals rate will still be significantly lower than to the long-term average performance, which are 81.2% and 82.3%, respectively. The cancellations rate forecasted by the ARIMA model will be much higher than the long-time average of 2.2% while this rate forecasted by VAR model is only 2.14%.
- In conclusion, it seems doubtful that the Australian domestic airlines will see a major improvement in their on-time performance in the next seven months.



Based on the above results and insights, the following are the recommendations for the business problem:

- The Government may amend laws and policies to enhance competition in the domestic airline sector which could better encourage airlines to invest in their customer service and protect passengers' benefits such as compensation rights for cancellations and delays.
- The domestic airlines should report the reasons or factors of their cancellations and delays and suggest solutions to improve their on-time performance in the future.
- The future on-time performance rate, cancellation rate, and delay rate can be forecasted periodically by BITRE or each airline to support the airlines in controlling their punctuality and adjusting their operation timely to minimize the impacts of cancellations and delays on passengers.
- The evaluation report for punctuality and reliability of airlines and the predictions for on-time performance should be public and easily approached for customers to assist them in choosing the best airlines for their travel.



2.The role of data analytics:

01

Power BI is the descriptive method that represents effectively the changes over time in cancellations, delays, and on-time performance rate of domestic airlines. Additionally, it may support us to identify the trend in the next period.

02

Both ARIMA and VAR model are appropriate for times series data and they can account for non-stationary data.

03

The VAR model forecasts outcomes over a longer time horizon with greater accuracy and effectiveness.

3. The limitations:

However, the following limitations of the dataset and analytics methods are discovered based on the data processing and outcomes predictions:

- The original data kept in separate files and sheets needs to be accurately processed before applying the descriptive and predictive methods.
- Due to the non-stationary nature of the data, differencing must be used in order to eliminate any seasonal patterns or trends before forecasting.
- The ARIMA model seems to be ineffective at predicting outcomes in the period over 3 months. So the accuracy of predictions may be reduced if using this model to forecast outcomes over the long period.

4. Other analytics techniques:

According to the above analysis and evaluations, this project suggests applying VAR model in forecasting the outcome to solve the business problem. Furthermore, in the future we can consider to apply Long Short-term memory (LSTM) technique to forecast the outcomes over extended periods. Time series forecasting tasks like stock price, weather, and energy consumption prediction have been tackled by LSTM. They can identify patterns in time series data and apply those patterns to forecast future occurrences (Geeksforgeeks 2023).

VII. DATA ETHICS AND SECURITY

The dataset used in this project is freely accessible on the BITRE website, so it is assumed that further use and analysis are permitted. However, it is necessary to acknowledge that the airlines own their original data. Furthermore, the data needs to be assessed based on its content, accuracy, collection period and purpose to make sure that further analysis of the data is appropriate because the original data was not gathered for secondary data analysis (Jaya Prasad Tripathy 2013).

To avoid misleading predictions and inaccurate conclusions, data integrity and accuracy are necessary for ethical data analysis. As such, the dataset collected on the public website should be ensured its high quality or its limitations must be disclosed. Moreover, the analytics methodologies and findings must be transparent and thoroughly verified (Chiara Scopigno 2023). In this project, the bias and discrimination of the data are recognized and acknowledged before processing and applying analytics methods. Additionally, all the limitations of data and methods are shown clearly with the visualizations of the ARIMA and VAR models' forecasting.



REFERENCE LIST

1. Amelia McGuire 2023, Transport minister savages airlines after worst performance in more than a year, The Sydney Morning Herald, retrieved 01 February 2024, <<https://www.smh.com.au/business/companies/transport-minister-savages-airlines-after-worst-performance-in-more-than-a-year-20231221-p5esz2.html>>
2. Anna Fleck 2022, The Airports to avoid this summer, graphic, Statista, retrieved 01 February 2024, <<https://www.statista.com/chart/27865/worst-offending-airports-for-delays/>>
3. Anna Fleck 2022, Thousands of flights canceled and delayed over July 4 Weekend, graphic, Statista, retrieved 01 February 2024, <https://www.statista.com/chart/27718/july-4-travel-chaos/>
4. Avijeet Biswal 2023, What is Power BI?: Everything You Need to Know!, Simplilearn, retrieved 28 December 2023, <<https://www.simplilearn.com/tutorials/power-bi-tutorial/what-is-power-bi>>
5. BITRE 2023, Domestic airline on-time performance, BITRE, retrieved 20 November 2023, <<https://www.bitre.gov.au/statistics/aviation/otphome>>
6. BITRE 2023, Domestic airline on-time performance November 2023, retrieved 01 February 2024, <https://www.bitre.gov.au/sites/default/files/documents/BITRE OTP_Report_November_2023.pdf>
7. BITRE 2023, Airline On-time performance monthly reports, BITRE, retrieved 20 December 2023, <https://www.bitre.gov.au/publications/ongoing/airline_on_time_monthly>
8. Chiara Scopigno 2023, Ethical Considerations in Data Analysis: Navigating Privacy, Bias and Responsibility, Linkedin, retrieved 08 January 2024, <https://www.linkedin.com/pulse/ethical-considerations-data-analysis-navigating-privacy-scopigno>
9. Geeksforgeeks 2023, Deep Learning Introduction to Long Short Term Memory, Geeksforgeeks, retrieved 15 January 2024, <<https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/>>.
10. Jaya Prasad Tripathy 2013, Secondary Data Analysis: Ethical Issues and Challenges, National Library of Medicine, retrieved 08 January 2024, <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4441947/>>
11. Jesse Scardina 2024, Microsoft Power BI, Tech Target Network, retrieved 07 January 2024, <<https://www.techtarget.com/searchcontentmanagement/definition/Microsoft-Power-BI>>
12. Linkedin 2023, What are the advantages and disadvantages of using vector autoregression for forecasting?, Linkedin, retrieved 06 January 2024, <<https://www.linkedin.com/advice/0/what-advantages-disadvantages-using-vector-autoregression>>
13. PI Exchange n/d, Which metrics are used to evaluate a regression model's performance?, PI.Exchange, retrieved 29 December 2023, <<https://www.pi.exchange/knowledgehub/metrics-to-consider-when-evaluating-a-regression-models-performance>>
14. Selva Prabhakaran 2024, ARIMA Model – Complete Guide to Time series forecasting in Python, Machine Learning, retrieved 06 January 2024, <<https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/>>
15. Siddharth Vikram Philip 2022, How Europe became the epicenter for this summer's travel chaos, photograph, retrieved 01 February 2024, <<https://www.bloomberg.com/news/features/2022-07-01/record-flight-delays-cancellations-make-europe-2022-s-worst-place-to-travel>>