DESCRIPTIVE, DIAGNOSTIC & FORECAST ANALYSIS OF PASSENGER FLIGHT IN JUANDA SURABAYA AIRPORT BASED ON GOOGLE TREND FROM 2011 TO AUGUST 2023

RIZAL IMADUDDIN YUSAN (6032222109)

SURYO UTOMO (6032222062)

MOCHAMMAD REYHAN MAULUDDI (6032222003)

MUHAMMAD FACHRY NOVA (6032222061)

DEFRIKO CHRISTIAN D (6032222098)

Contents

Background

Juanda International Airport (SUB) is the third busiest airport in Indonesia

Research Question

How to describe the analysis of big data on the trend of domestic and international air traffic managed by Juanda Surabaya Airport based on the number of passengers and air operators data in the year 2011 to August 2023 related to mobility trends during COVID-19, trends in gold price and oil price ?

Objective

Understand the characteristics of passengers and air operators based on domestic and international flight traffic data of Juanda Surabaya Airport from 2011 to August 2023

Data Source & Variable

- 1.511.660 of rows
- 28 features (variables)
- range data from 2011-August 2023

Result, Insight & Discussion

The analysis made of various aspects :
1. Annual Passenger Traffic by Airline
2. Monthly Passenger Traffic Heatmap
3. Traffic share by Destination
4.The Peak points traffic airlines (Monthly, Days and holidays)
5.Diagnostic Analysis
6.Predictive Analysis (ARIMA, LSTM, Prophet)

Conclusion

1. Explain Descriptive Analysis for Airlines Traffic
2. Explain Diagnostic Analysis effect Airlines Traffic by Covid Data Mobility, Gold Price & Oil Price
3. Explain Predictive Analysis by time series forecasting

Future Work

Air Passenger Forecasting with **macroeconomic indicators**, commodity prices, and mobility data as additional predictor variables

Company Profile, Background & Objective

Company Profile

Juanda International Airport (JIA) (Indonesian: *Bandar Udara Internasional Juanda*) (IATA: **SUB**, ICAO: **WARR**), is an international airport located in Sedati, Sidoarjo. It is now the third busiest airport in Indonesia (after Soekarno-Hatta and Ngurah Rai airport). This airport is located approximately 12 kilometers (7.5 mi) from Downtown Surabaya and serves the Surabaya metropolitan area, the metropolitan area of Surabaya plus extended urban area. Juanda International Airport is operated by PT Angkasa Pura I[1]. The aviation sector is severely affected by Covid pandemic where the aviation industry suffered a loss of $126.4 billion in 2020 and 2021 $47.7 billion. Air passenger traffic decreased by -60%.

Background

The challenges global aviation industry is experiencing financial recovery, with 2022 marking a turning point in traffic recovery

Juanda International Airport (SUB) development focuses on creating an integrated network to enhance industrial activities, employing the superhub, hub, and spoke concept

The study employs historical passenger data from Juanda Airport (2011-2020) for forecasting models, utilizing Google Trends as an exogenous variable to gauge public interest in air travel-related topics

The forecasting models include ARIMA for trends and seasonal patterns, LSTM for nonlinear characteristics, and Prophet for integrating seasonal components, trends, and external variables

Objective

The research aims to provide more accurate predictions of Juanda Airport's passenger numbers, incorporating insights from Google Trends. Model evaluation will use statistical metrics like MAE, RMSE, and MAPE.

The research potential impact extends to offering valuable guidance for authorities and stakeholders involved in planning, capacity management, infrastructure development, flight planning, operational management, and marketing strategies in (SUB)

Furthermore, the findings of this research can assist in improving flight planning, operational management, and marketing strategies at the airport

Literature Study, Data Source & Variable

Literature Study

"The Success of Lufthansa in Adding Value to Its Business through the Shift from a Selling Dominant to a Service Dominant Mindset Using Big Data."

“Big data is being mined to discover innovative service provision and operational optimizations that were not possible before.”
(Chen, 2016)

“We observe a negative and statistically significant stock price reaction to the announcement of COVID-19 as a global pandemic to the airline industry in US”
(Martins, 2022)

“The optimal weights for gold were substantially higher than those of other assets, reaching a peak during the pandemic, implying that investors consider gold a flight-to-safety asset”
(Nassar, 2023)

Data Source

Data were sourced from PT Angkasa Pura I Airport Collaborative Decision Making System (SIOPSKOM X-CDM), Yahoo Finance, and Google Mobility Data

List of Important Variables

Variable	Description	Data Type
Activity Period	Date of Flight	DateTime
OPERATOR	Airline	Text
DOIN	Domestic/International	Category
Passenger Count	Paying, Infant, Transfer	Integer
GC=F	Gold Price	Continuous
CL=F	Oil Price	Continuous
Mobility Score	Covid - Indonesia Google Mobility Score	Continuous

Key figures

1,511,660	28	16	7
# of rows in the dataset	# of variables in the dataset	# of text variables in the dataset	# of numeric variables in the dataset

Related Works

Time Series Forecasting Models Used:

- ARIMA and Exponential Smoothing which are the traditional methods.
- LSTM neural networks effectively capture long-range dependencies.
- Prophet, developed by Facebook, is suitable for strong seasonality and holiday effects.

Hybrid Forecasting Approaches:

- Researchers combined strengths of ARIMA, LSTM, and Prophet models.
- Enhances accuracy and robustness of predictions.

Application of Models:

- ARIMA and Exponential Smoothing for traditional forecasting.
- LSTM for capturing long-range dependencies.
- Prophet for datasets with seasonality and holiday effects.

Exogenous Variables in Forecasting:

- Incorporating external factors like economic indicators and online search trends improves accuracy.
- Google Trends data provides insights into public interest and influences time series data.

Research Context:

- Specific emphasis on the aviation industry and Juanda Airport.
- Goal is to develop a comprehensive forecasting framework.

Practical Significance:

- Forecasting assists in capacity planning and operational decision-making.
- Hybrid models provide a more accurate understanding of temporal patterns.

Future Works:

- Building upon existing studies to develop a robust forecasting framework.
- Further exploration of hybrid models and the incorporation of additional exogenous variables.

Theory and Methodology

Data Collection

Historical data on the number of passengers at Juanda Airport from 2011 to 2020 is collected from official sources. This dataset serves as the foundation for model development and evaluation.

Preprocessing

Data preprocessing involves cleaning, handling missing values, and aggregating the data into relevant time intervals (e.g., monthly data). Additionally, Google Trends data related to air travel keywords is acquired and preprocessed for use as an exogenous variable.

Data Splitting

The dataset is split into two subsets: a training set and a testing set, with a split ratio of 80% for training and 20% for testing. The training set contains the majority of historical data, while the testing set covers a more recent portion.

ARIMA Modeling

Autoregressive Integrated Moving Average (ARIMA) modeling is applied to capture the temporal patterns and autocorrelations in the passenger data. The model is chosen based on the stationarity of the series and the results of autocorrelation and partial autocorrelation functions.

LSTM Modeling

Long Short-Term Memory (LSTM) neural networks are used for time series forecasting. The model architecture is designed with suitable hyperparameters. LSTM can capture complex temporal dependencies and nonlinear relationships in the data.

Prophet Modeling

The Prophet forecasting model is employed, considering its capabilities in handling seasonality and exogenous variables. The model is configured with information about yearly and weekly seasonality, holiday effects, and the inclusion of Google Trends data as an exogenous input.

Theory and Methodology

Evaluation Matrix

The accuracy of the forecasting models is assessed using various evaluation metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE).

Model Comparison

The ARIMA, LSTM, and Prophet models are compared based on their performance in predicting passenger numbers. The models are evaluated using the chosen metrics to determine which model provides the most accurate forecasts.

Exogonus Variable Analysis

The influence of Google Trends data on the forecasting accuracy is analyzed. This involves assessing the impact of search trends related to air travel and identifying the keywords that have the most significant correlation with passenger numbers

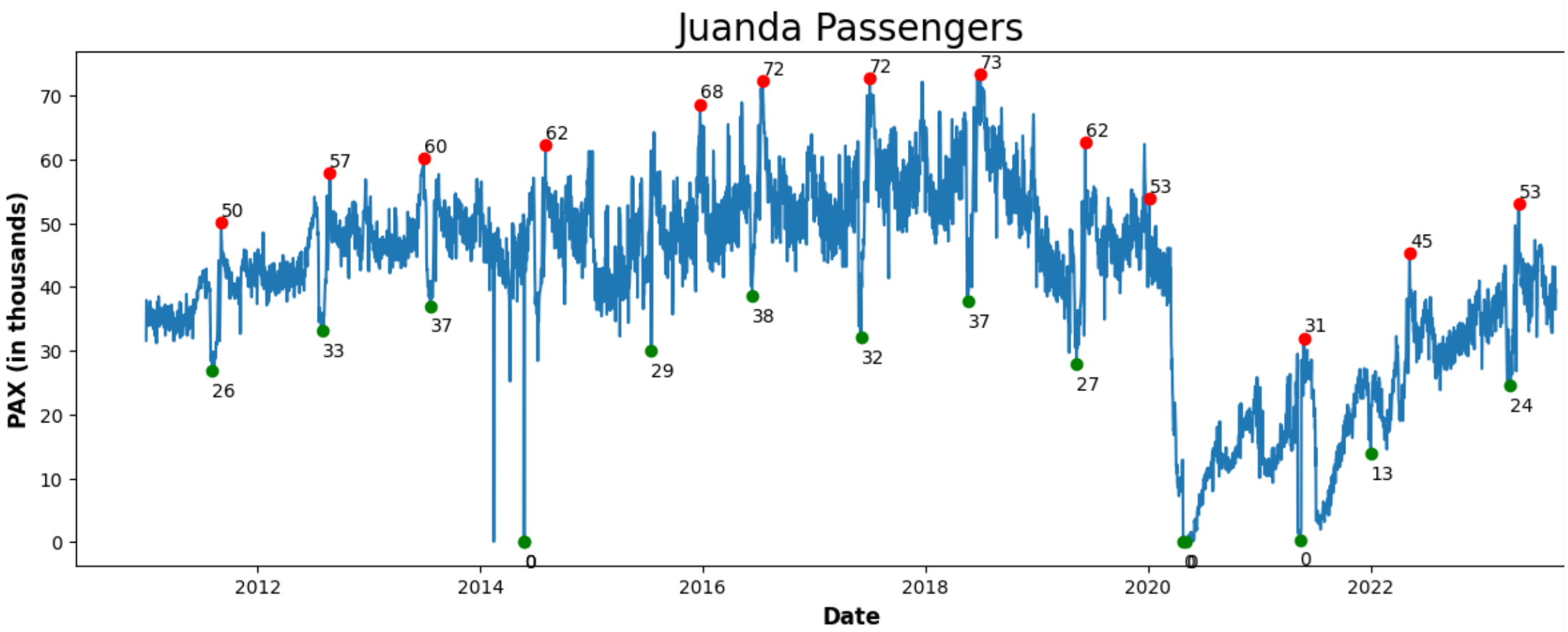
Forecast for 2021 - 2023

After models have been selected and validated, the final models are utilized to forecast the number of passengers at Juanda Airport for the years 2021 to 2023, incorporating the learned patterns, trends, and exogenous factors.

Interpretation and Conclusions

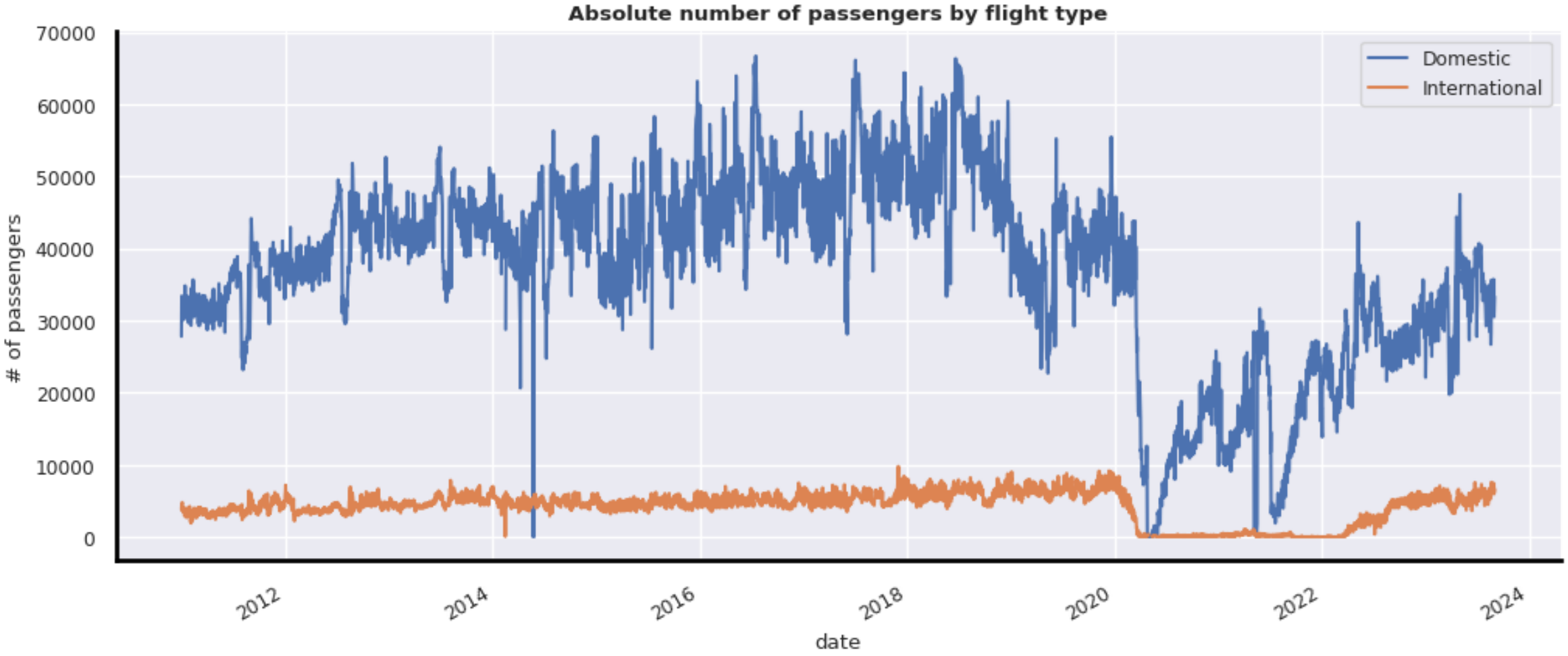
The results are interpreted to provide insights into the accuracy of each model and the role of Google Trends data in improving predictions. Conclusions are drawn regarding the most suitable forecasting approach for Juanda Airport passenger numbers and the practical implications for airport management and planning.

Descriptive Analysis Highlight



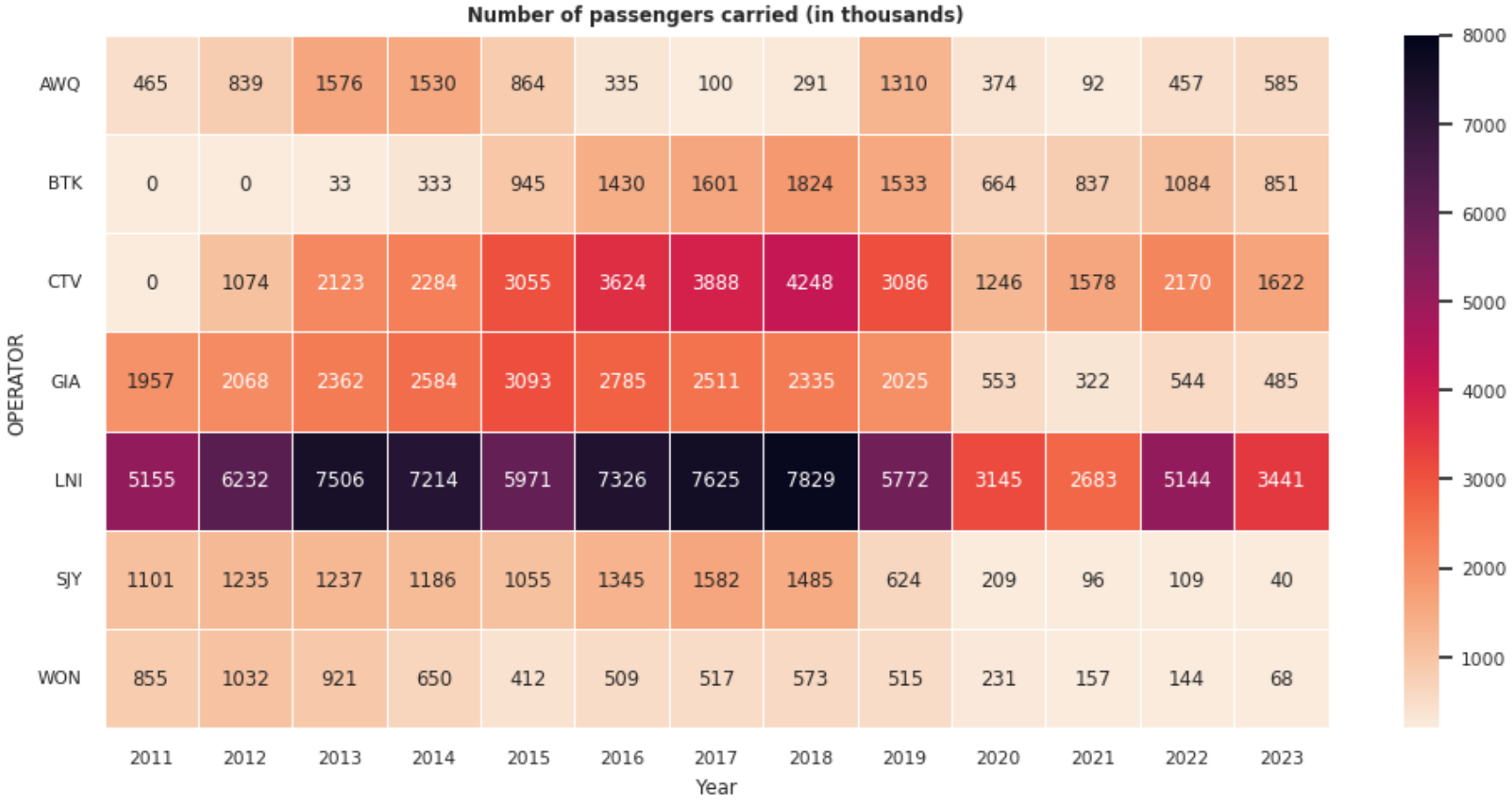
- Passengers traffic **steadily grows** from 2011 through 2019 and 2022 through 2023, **drastically decreases** in 2020 due to Covid 19 effects.
- **The highest number** of passengers is reported so far from 2017 through 2019 between 72 - 73 thousands passengers
- **There are few zero numbers. The first** zero number is reported on 2014 because of kelud volcanic eruption and **the second and third** zero number are in 2020 because of Covid 19 lockdowns.

Traffic share by Destination



- Based on international and domestic categories, Juanda Airport is predominantly dominated by domestic flights overall.
- For domestic traffic, in 2023 the number of passengers has not experienced an increasing trend as it did before 2019.
- In 2020, there was an extreme decrease to the point of reaching zero due to the impact of COVID on the airlines industry

Annual Passenger Traffic by Airline

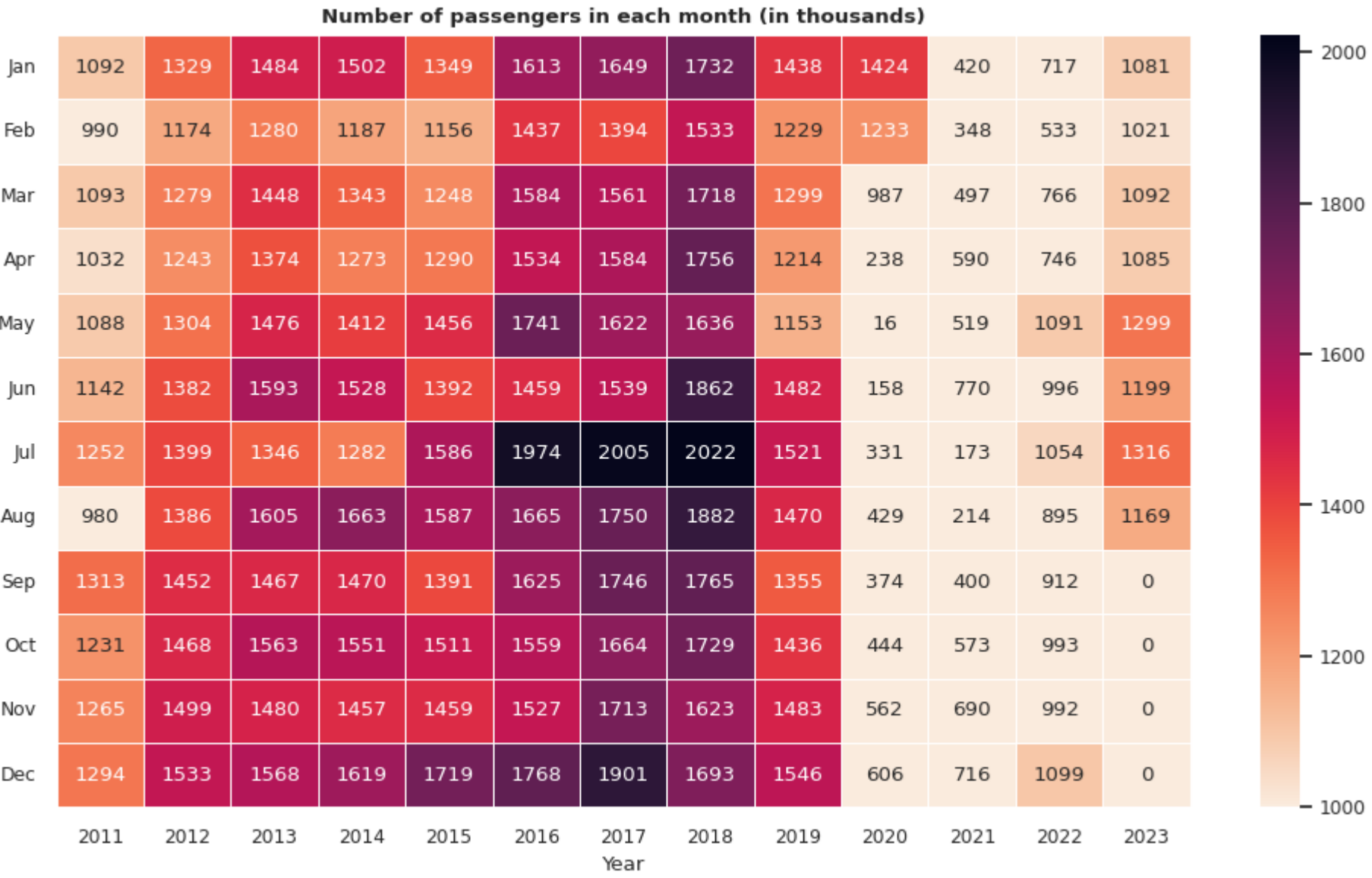


Airlines Operators :

1. AWQ = Air Asia Air
2. BTK = Batik Air
3. CTV = Citilink
4. GIA = Garuda Indonesia
5. LNI = Lion Air
6. SJY = Sriwijaya
7. WON = Wings

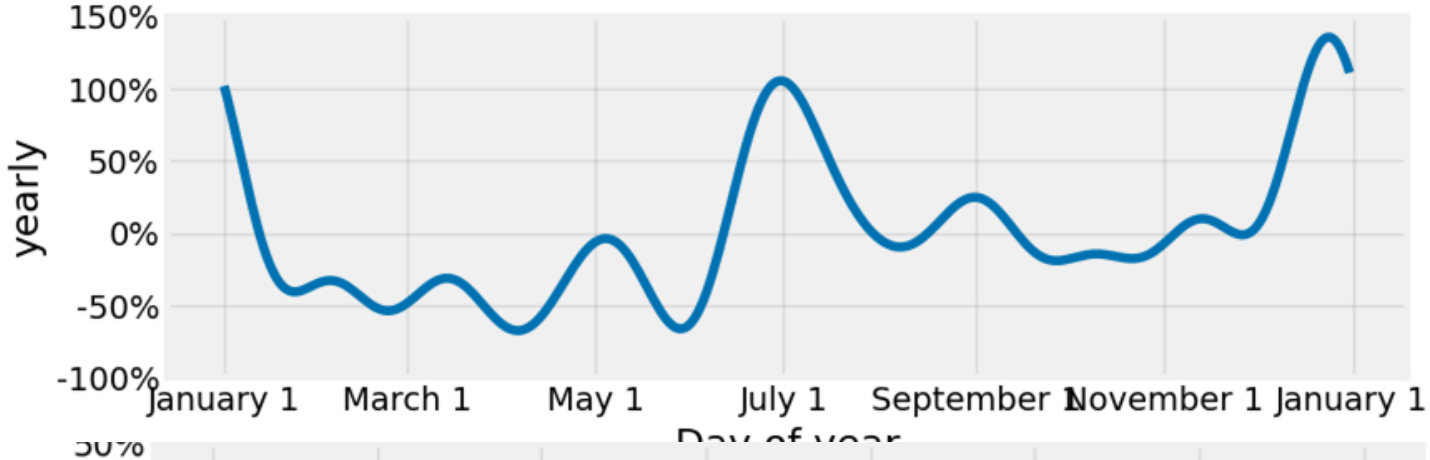
- The heatmap shows seven most dominating airline operators for annual passenger traffic in Juanda Airport. Lion Air is the most dominant among them.
- Passengers traffic **steadily grows** from 2011 through 2018 but from 2019 through 2023 **drastically decreases** because of covid 19's effects.
- In 2023 the number of passengers has not experienced an increasing trend as it did before 2019.
- **The highest number** of passengers is reported so far is in 2016-2018. Lion Air has the most dominating number of passengers of all operators.

Monthly Passenger Traffic Heatmap

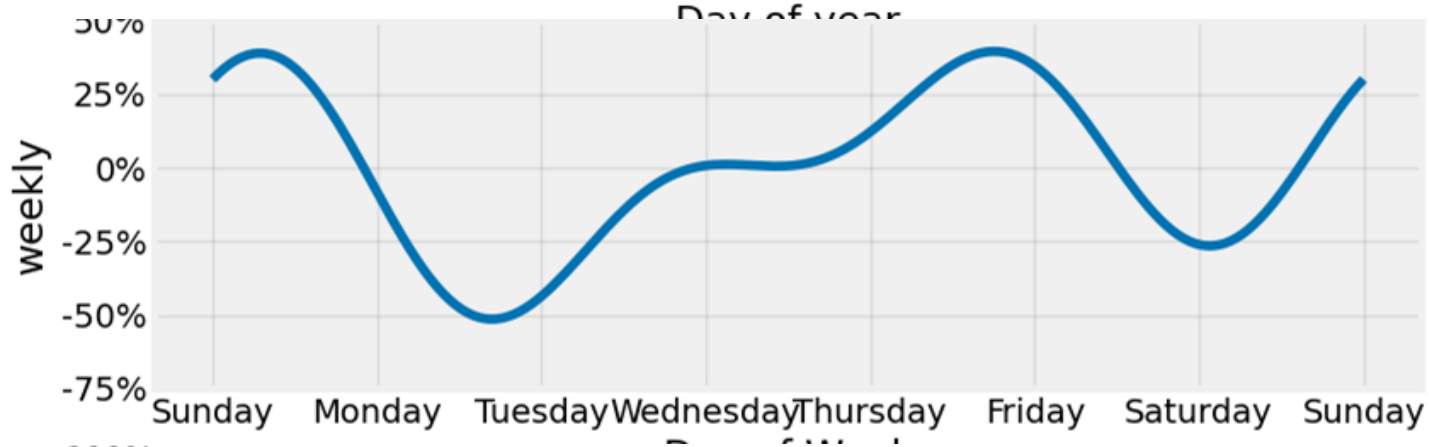


- The heatmap shows monthly passenger traffic in Juanda Airport from 2011-2023.
- Passengers traffic **steadily grows** from 2011 through 2018 but from 2019 through 2023 **drastically decreases** because of covid 19's effects.
- In 2023 the number of passengers has not experienced an increasing trend as it did before 2019.
- **The highest number** of passengers is reported so far is in 2016-2018.
- **July-Dec** has the **peak** of monthly passenger traffic every year from 2011-2019
- **Feb** has the smallest number of monthly passenger traffic every year from 2011-2019..

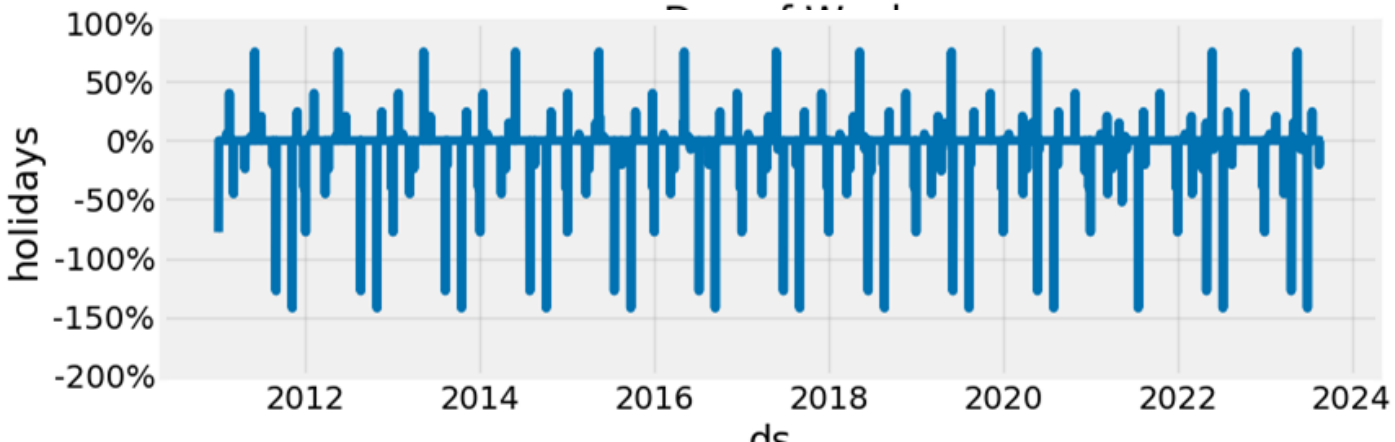
Time Series Decomposition



- There are two peak points every year from 2011-2023. The peak points are in July (**School Holiday**) and December (**Christmas and New Year Holiday**).
- There are two trough points every year from 2011-2023. The trough points are in April and June.

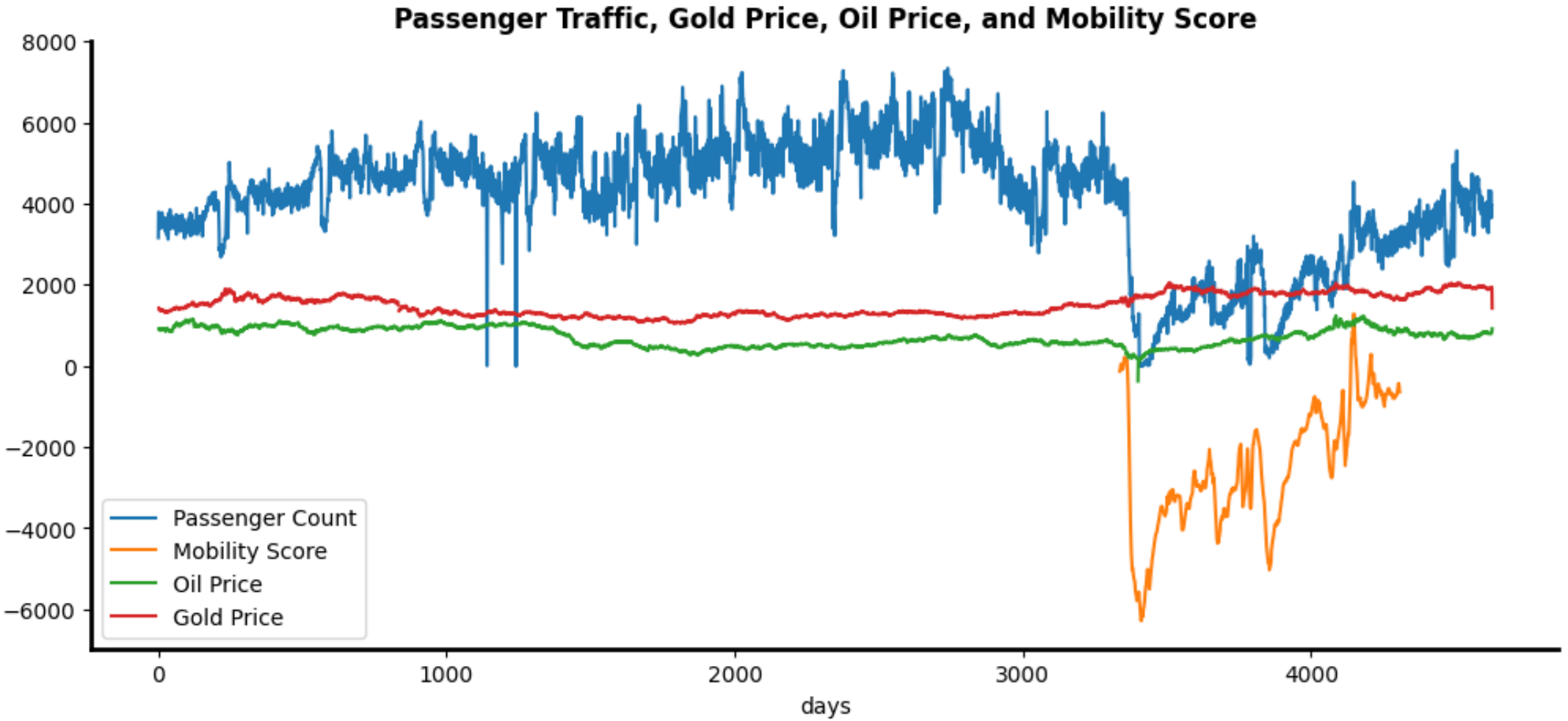


- There are two weekly peak points from 2011-2023. The peak points are in Sunday and Friday.
- There is one weekly trough points every year from 2011-2023. The trough point is in Tuesday.



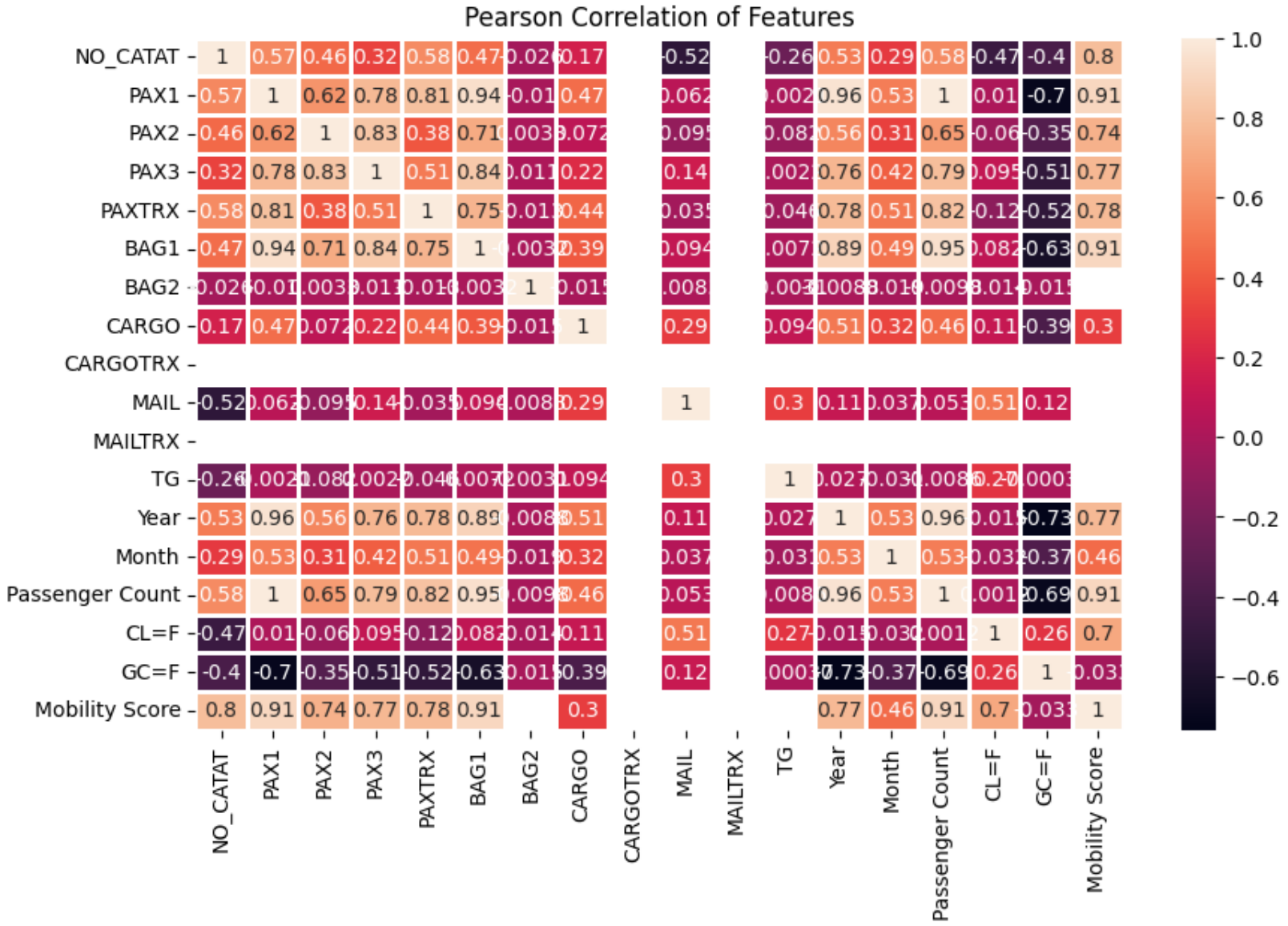
- On the holiday, there is a negative effect on passenger decreasing number. This occurs because customers travel on the day before the holiday.

Diagnostic Analysis



- The trend in passenger traffic has significantly declined during the COVID-19 pandemic. This is in line with the decreasing mobility score trend.
- The emergence of COVID-19 has caused the price of gold to rise.
- On the contrary, the emergence of COVID-19 has caused the movement of global oil prices to decrease due to the global economic downturn.

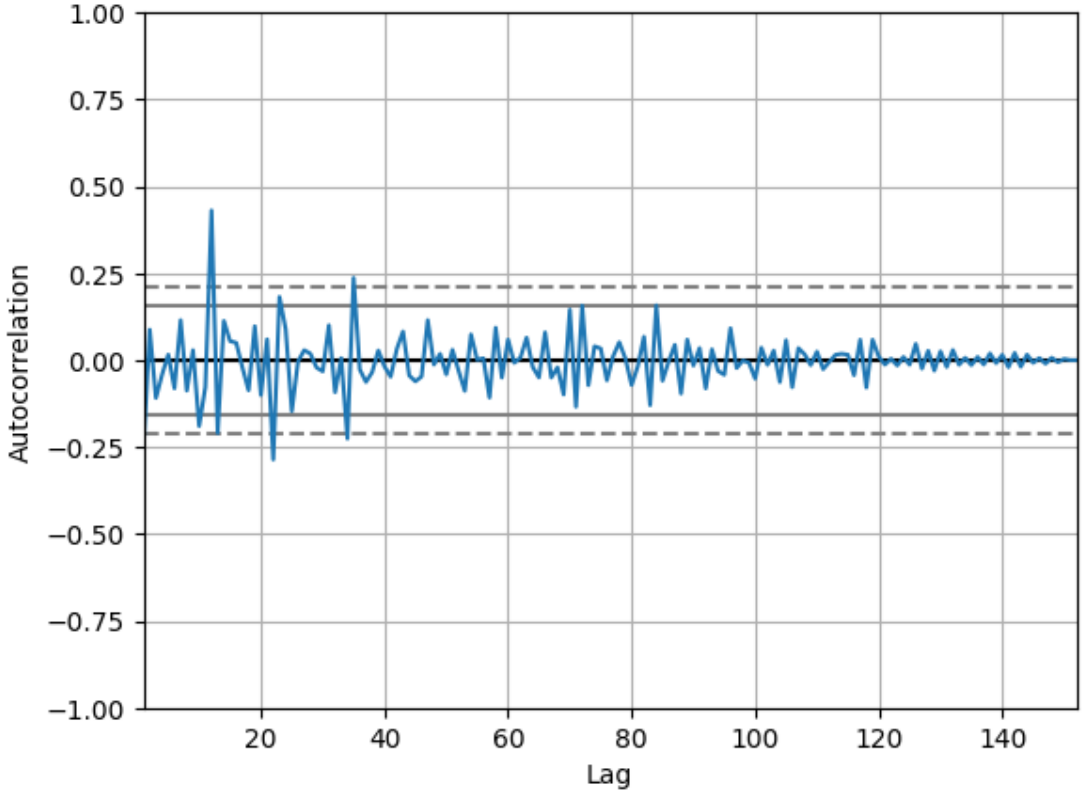
Diagnostic Analysis: Correlation

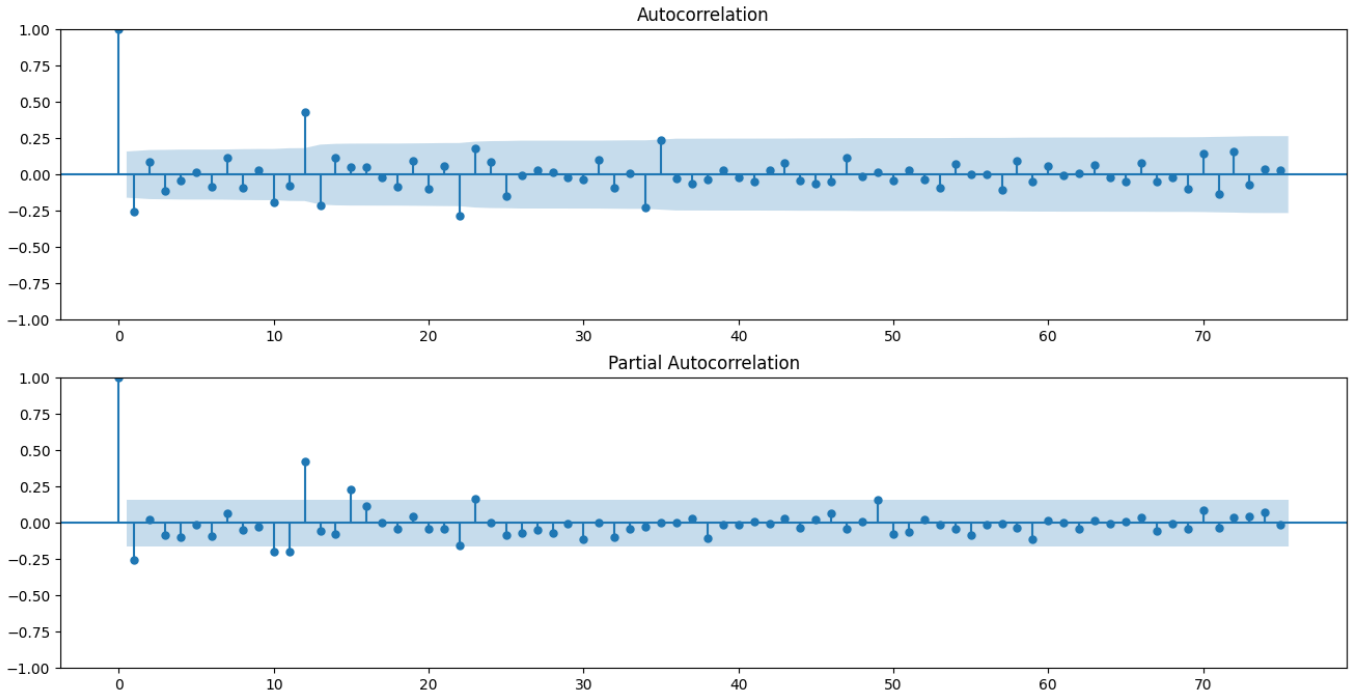


NO_CATAT	0.580343
PAX1	0.998722
PAX2	0.648513
PAX3	0.793570
PAXTRX	0.816014
BAG1	0.946740
BAG2	-0.009841
CARGO	0.460257
CARGOTRX	NaN
MAIL	0.052539
MAILTRX	NaN
TG	-0.008595
Year	0.961057
Month	0.532147
Passenger Count	1.000000
CL=F	0.001181
GC=F	-0.694954
Mobility Score	0.908650

- The number of passengers has a negative correlation with the price of gold (GC=F) with a coefficient of **-0.69**
- The number of passengers has a very small positive correlation with the global oil price (CL=F) with a coefficient of 0.001.
- The number of passengers has a great positive correlation with the mobility score with a coefficient of **0.9**

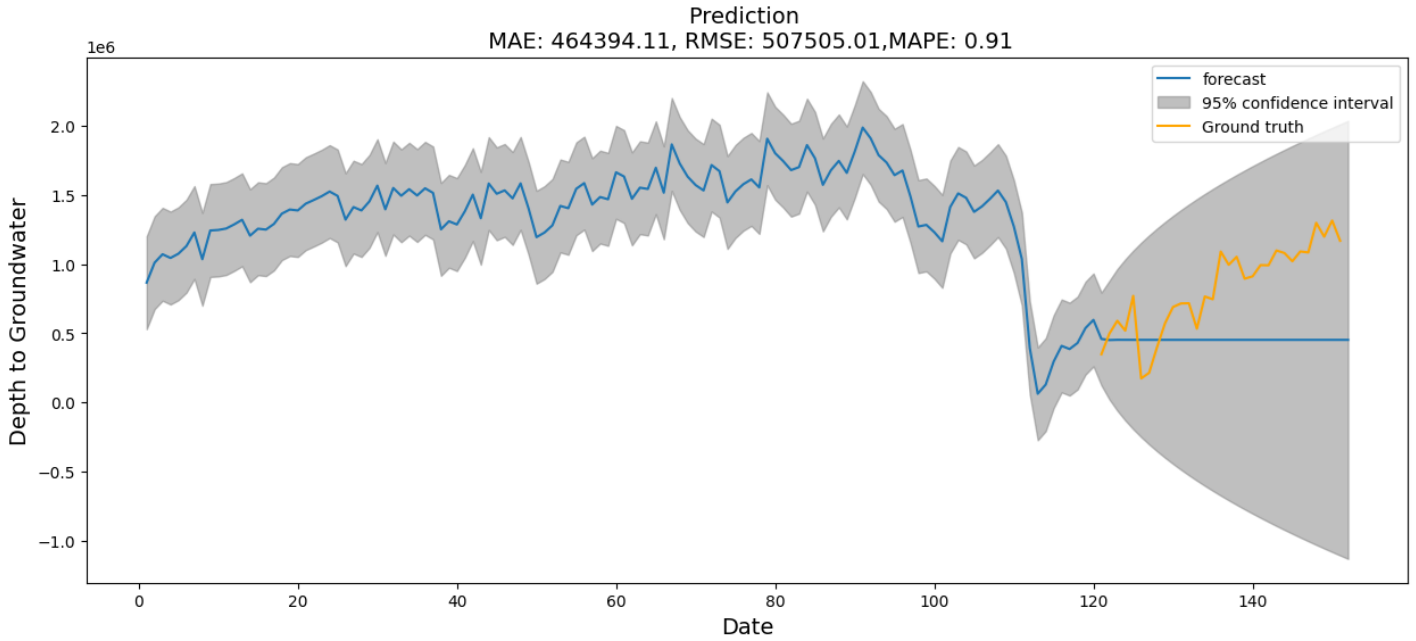
Time Series Forecasting: ARIMA

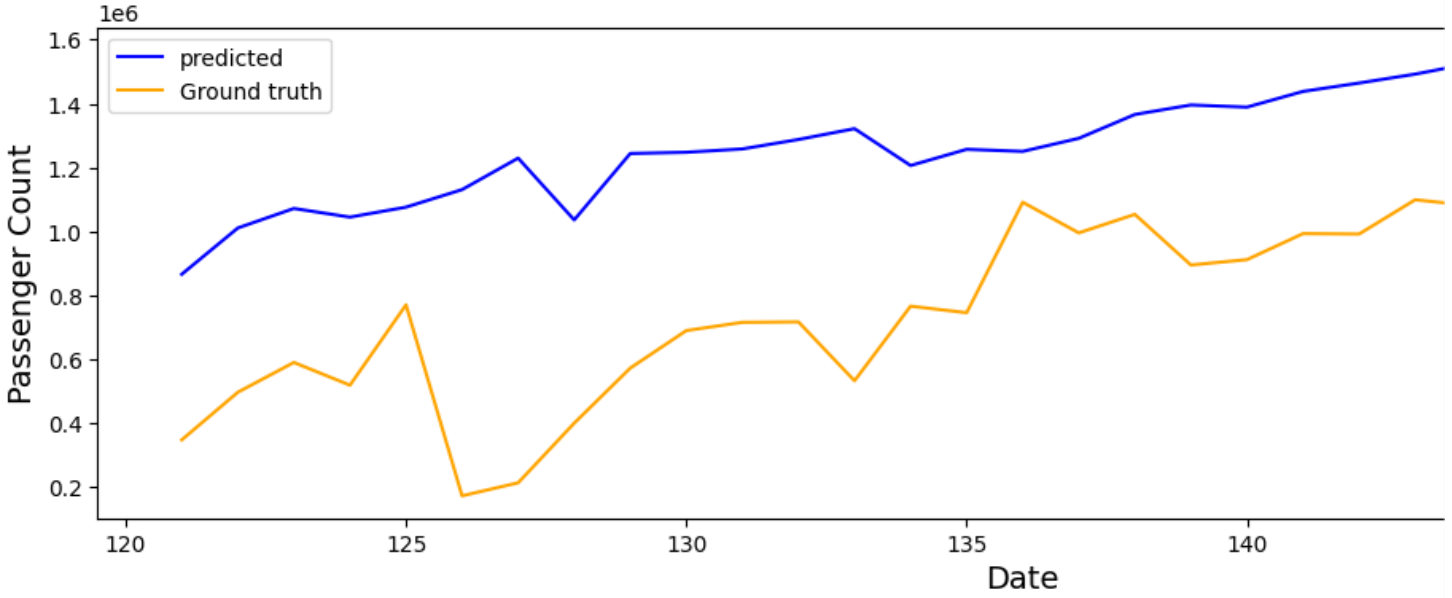




- Based on the ACF and PACF plots, we can select the appropriate orders (p, d, q) for ARIMA model. The orders represent the number of autoregressive terms (p), differencing (d), and moving average terms (q) in the model.

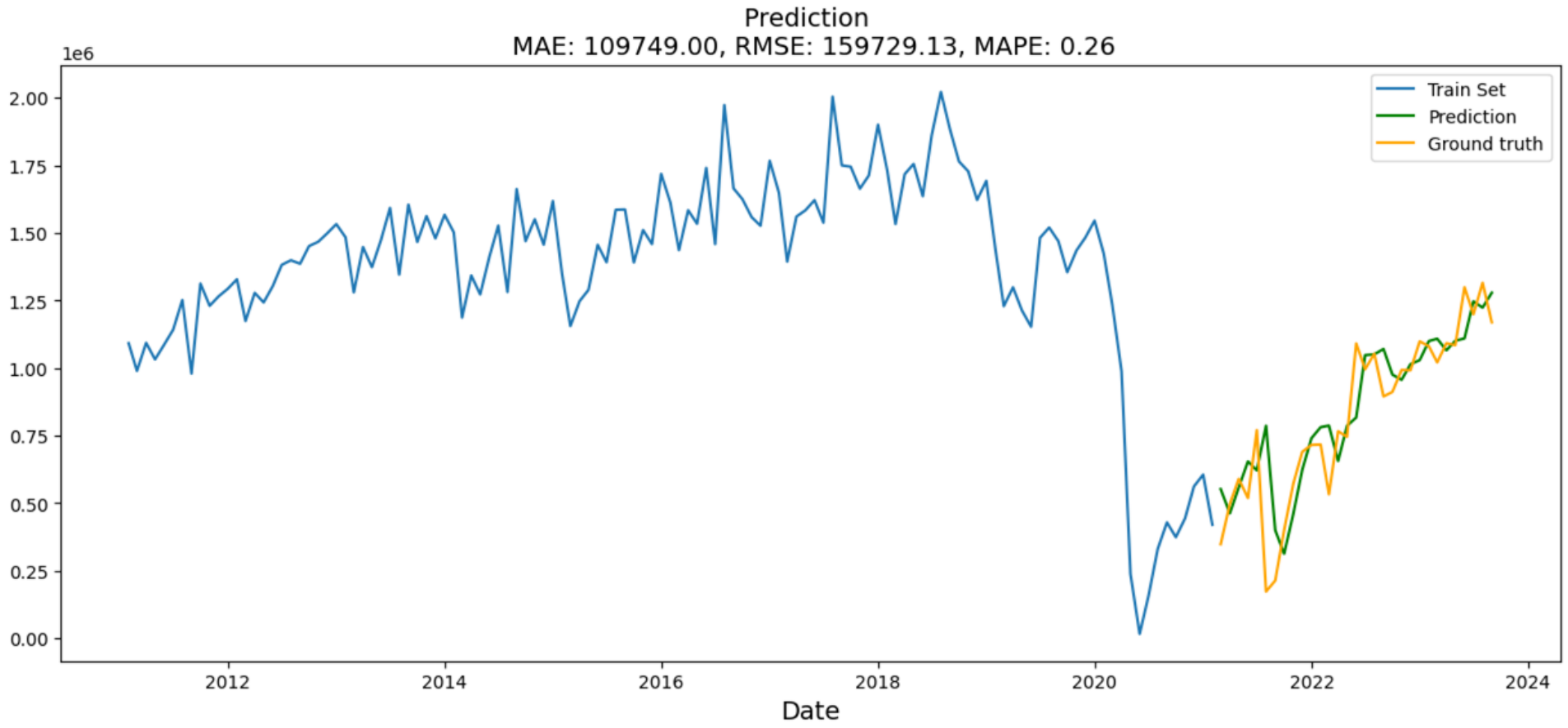
Time Series Forecasting: ARIMA





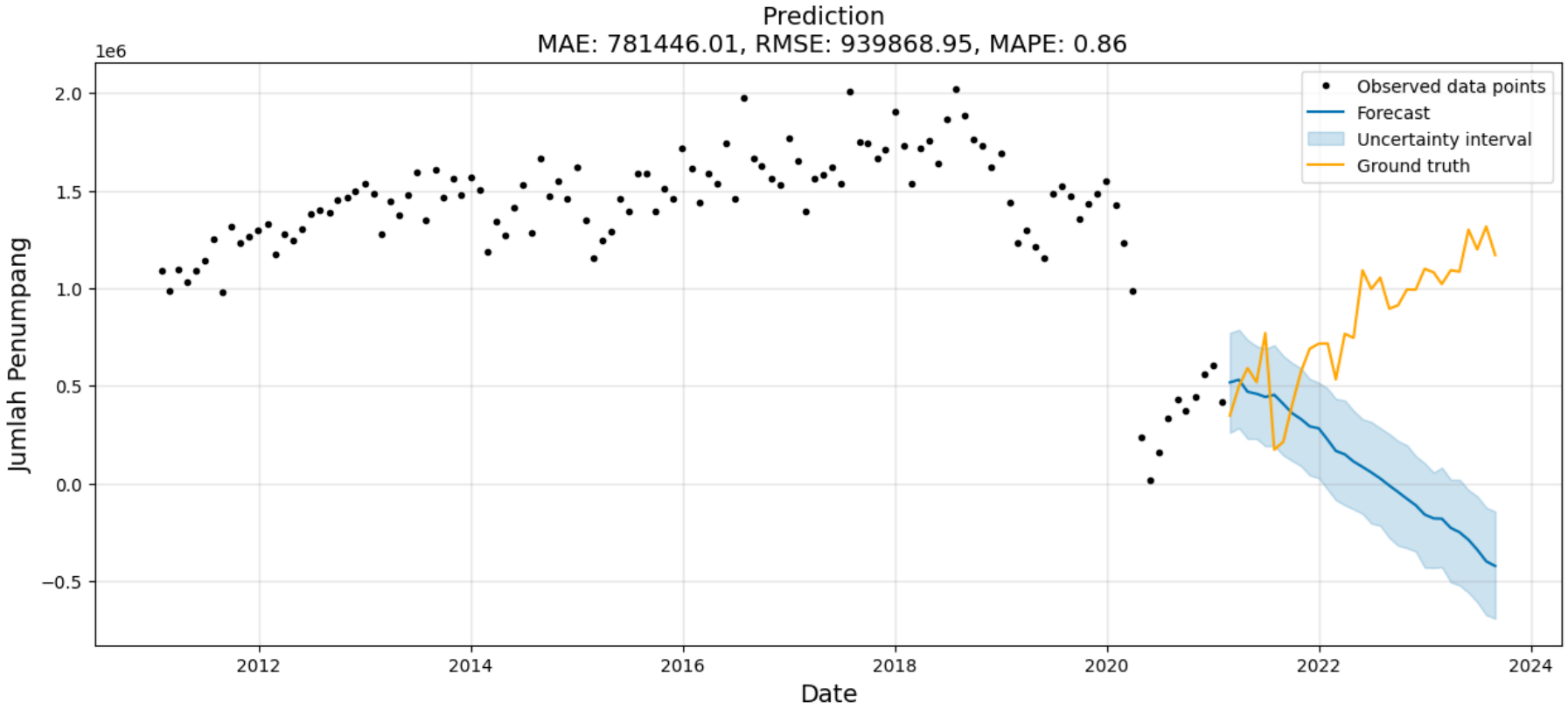
- The RMSE for the ARIMA model is approximately 507,505
- The MAPE for the ARIMA model is approximately 0.906, which is equivalent to 90.6%. A lower MAPE signifies a more accurate forecast.

Time Series Forecasting: LSTM



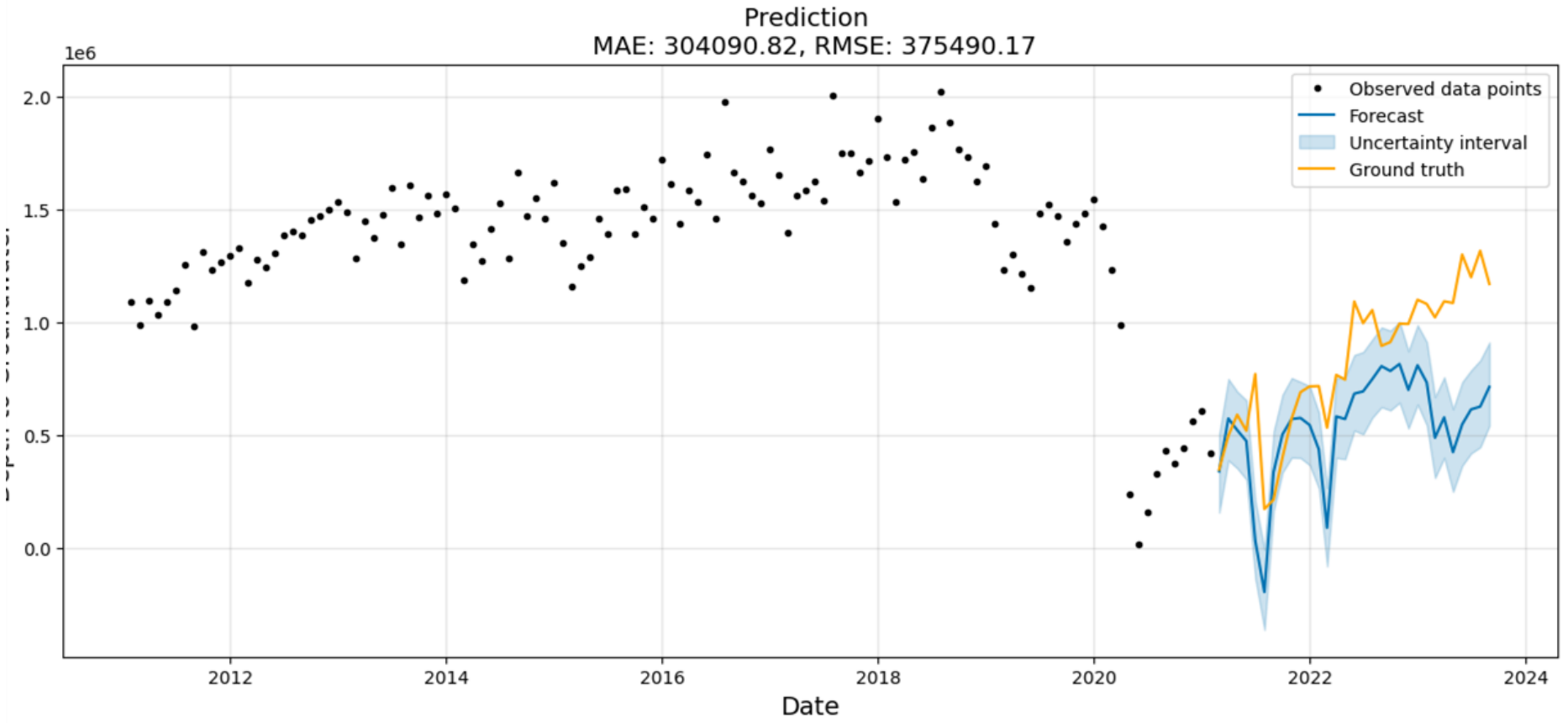
- The RMSE for the LSTM model is approximately 159,729
- The MAPE for the LSTM model is approximately 0.261, which is equivalent to 26.1%

Time Series Forecasting: Prophet



- The RMSE for the Prophet model is approximately 939,869
- The MAPE for the Prophet model is approximately 0.863

Time Series Forecasting: Prophet with Exogenous Variable



- The RMSE for the Prophet model with exogenous variables is approximately 363,306
- The MAPE for the Prophet model with exogenous variables is approximately 36.87%

Time Series Forecasting: Model Performance Summary

Model	MAPE	RMSE
ARIMA	0.906	507,505.00
LSTM	0.261	159,729.13
Prophet	0.863	939,869.95
Prophet with Exogenous	0.36	363,306.00

- The LSTM model stands out as the most accurate model among the evaluated approaches, as it achieved the lowest RMSE and MAPE.
- The ARIMA model also provides reasonably accurate predictions, outperforming both the Prophet models.
- The inclusion of Google Trends data as exogenous variables in the Prophet model improved its forecasting accuracy, reducing the MAPE from 0.863 to 0.36. However, this model still fell short of the LSTM and ARIMA models in terms of accuracy

Descriptive & Diagnostic Conclusion

- Passengers traffic steadily grows from 2011 through 2018 but from 2019 through 2023 drastically decreases because of covid 19's effects.
- In 2023 the number of passengers has not experienced an increasing trend as it did before 2019.
- Based on international and domestic categories, Juanda Airport is predominantly dominated by domestic flights overall.
- Lion Air is the most dominant among the seven dominant airlines operators.
- There are two peak points every year from 2011-2023. The peak points are in July (School Holiday) and December (Christmas and New Year Holiday).
- The number of passengers has a negative correlation with the price of gold (GC=F) with a coefficient of -0.69.
- The number of passengers has a great positive correlation with the mobility score with a coefficient of 0.9

Forecasting Model Conclusion

- The study aimed to forecast the number of passengers at Juanda Airport from 2011 to 2023.
- Four different forecasting models were applied and evaluated for their accuracy. Results showed significant differences in forecasting accuracy among the models.
- The Long Short-Term Memory (LSTM) model emerged as the most accurate, achieving the lowest Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE).
- The Autoregressive Integrated Moving Average (ARIMA) model also provided reasonably accurate predictions, outperforming the Prophet models.
- The Prophet model, when supplemented with Google Trends data as exogenous variables, improved forecasting accuracy by reducing MAPE from 0.863 to 0.36.
- Despite the enhancement with exogenous variables, the Prophet model still fell short of the accuracy achieved by the LSTM and ARIMA models.

References

Annual Report PT Angkasa Pura I Tahun 2022

Chen, H.M. Schutz, R. Kazman, R. Matthes, F. *“Amazon in the Air: Innovating with Big Data at Lufthansa.”* Hawaii International Conference on System Sciences 2016 49th.

Khafidli, MK., Choiruddin, A.: Forecast of Aviation Traffic in Indonesia Based on Google Trend and Macroeconomic Data using Long Short-Term Memory. 2022. International Conference on Data Science and Its Applications (ICoDSA)

Nassar, S. Boubaker, S. Chaibi, A. Makram, B. *“In search of hedges and safe havens during the COVID—19 pandemic: Gold versus Bitcoin, oil, and oil uncertainty.”* Quarterly Review of Economics and Finance 90 (2023): 318-332.

Martins, A.M. Cro, S. *“Airline stock markets reaction to the COVID-19 outbreak and vaccines: An event study.”* Journal of Air Transport Management 105 (2022) 1-10.

Zachariah, RA., Sharma, S., Kumar, V. Systematic Review of Passenger Demand Forecasting in Aviation Industry. 2023. Multimedia Tools and Applications

https://www.google.com/covid19/mobility/

BPS Provinsi Jawa Timur. Statistik Transportasi Provinsi Jawa Timur Tahun 2022.

Lampiran, Data Source Dll

Colab Notebook:
https://drive.google.com/file/d/1ZMrzCP0ZYQ0lfOIMbL4CNgyt5L1pGTXU/view?usp=drive_link

THANK YOU