

A photograph of a professional meeting. In the foreground, a man in a dark suit and striped tie is gesturing with his hands. Behind him, a woman with her hair pulled back is looking down at something. To the right, another person's face is partially visible, smiling. Overlaid on the top right is a white grid with a blue bar chart showing increasing values from left to right, and a red line graph showing an upward trend.

Project 1 - SALES FORECASTING

by 360

Meet the Team



Yanto



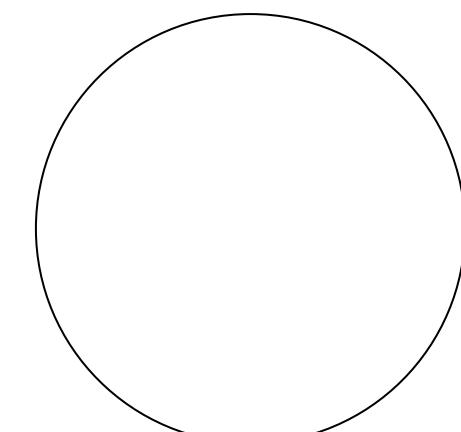
Muhammad Sutisna



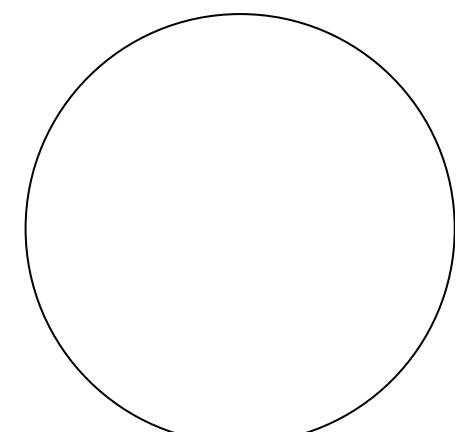
Vika Oktarina



Ryandito Mahendradani



A. Mambaus Sholihin



Mukhammad Fatkhur Rozi



Background & Problem Statement

Dalam usaha toko groceries persediaan stock barang menjadi masalah tersendiri bagi pelaku usaha. Sering kali toko akan mengalami kerugian apabila toko tidak dapat menyiapkan jumlah barang sesuai dengan permintaan. Alhasil banyak barang tidak terjual dan menjadi kadaluarsa sehingga menimbulkan kerugian, atau sebaliknya barang yang laku dijual tapi toko malah kehabisan stock sehingga tidak dapat meraup keuntungan.

Oleh karena itu, penyediaan stock yang pas dalam waktu yang tepat sangat diperlukan untuk memaksimalkan potensi laba usaha.

Objectives & Scope

Objectives

Sesuai dengan latar belakang permasalahan yang ada. Maka untuk itu Tim 360 mengembangkan sistem AI untuk meramalkan penjualan toko, sehingga pemilik toko dapat menyiapkan stock sesuai dengan permintaan dan memaksimalkan laba

Scope

Peramalan dibatasi dalam scope sebagai berikut:

1. Data yang dianalisa kategori Groceries I pada toko groceries kode 5
2. Peramalan penjualan dengan sistem AI model time series ARIMA & LSTM



Data Collection & Preparation

Dalam melakukan eksplorasi data, ditemukan hal-hal berikut:



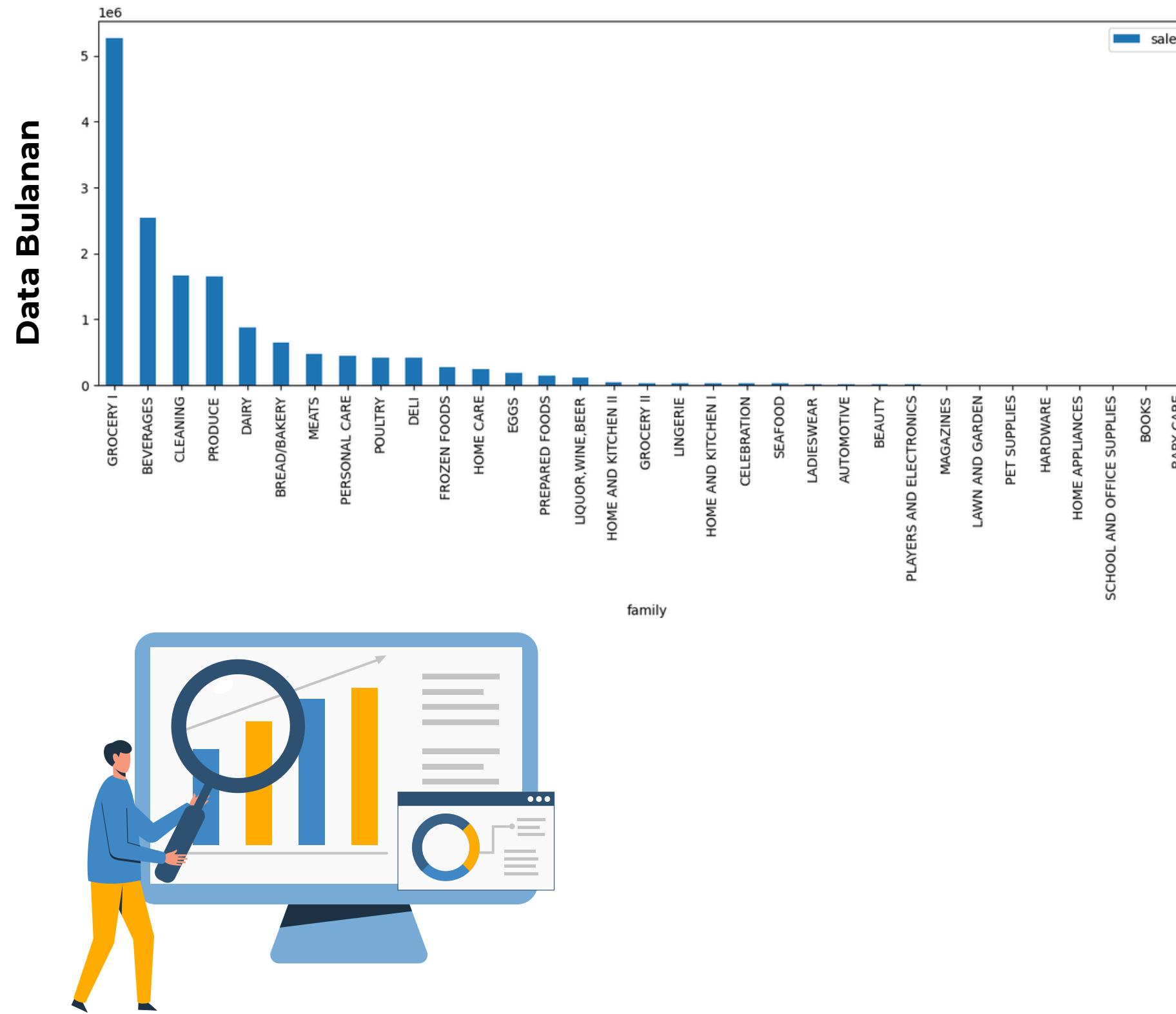
Terdapat 5 fitur dalam data yaitu:

1. Family: Jenis barang
2. Store_nbr: Nomor cabang
3. Sales: Jumlah penjualan
4. Onpromotion: Data apakah barang dalam promosi atau tidak
5. Dcoilwtico: Harga minyak mentah

Persiapan data dengan sebagai berikut:

1. Menghapus kolom store_nbr karena hanya kode toko dan memang dataset sudah terfilter ke kolom 5 sesuai objectives dan scope
2. Menghapus kolom Dcoilwtico karena banyak data kosong
3. Memilih kategori produk / Family yang paling layak diteliti
4. Melakukan uji korelasi antara OnPromotion dengan Sales

Data Collection & Preparation

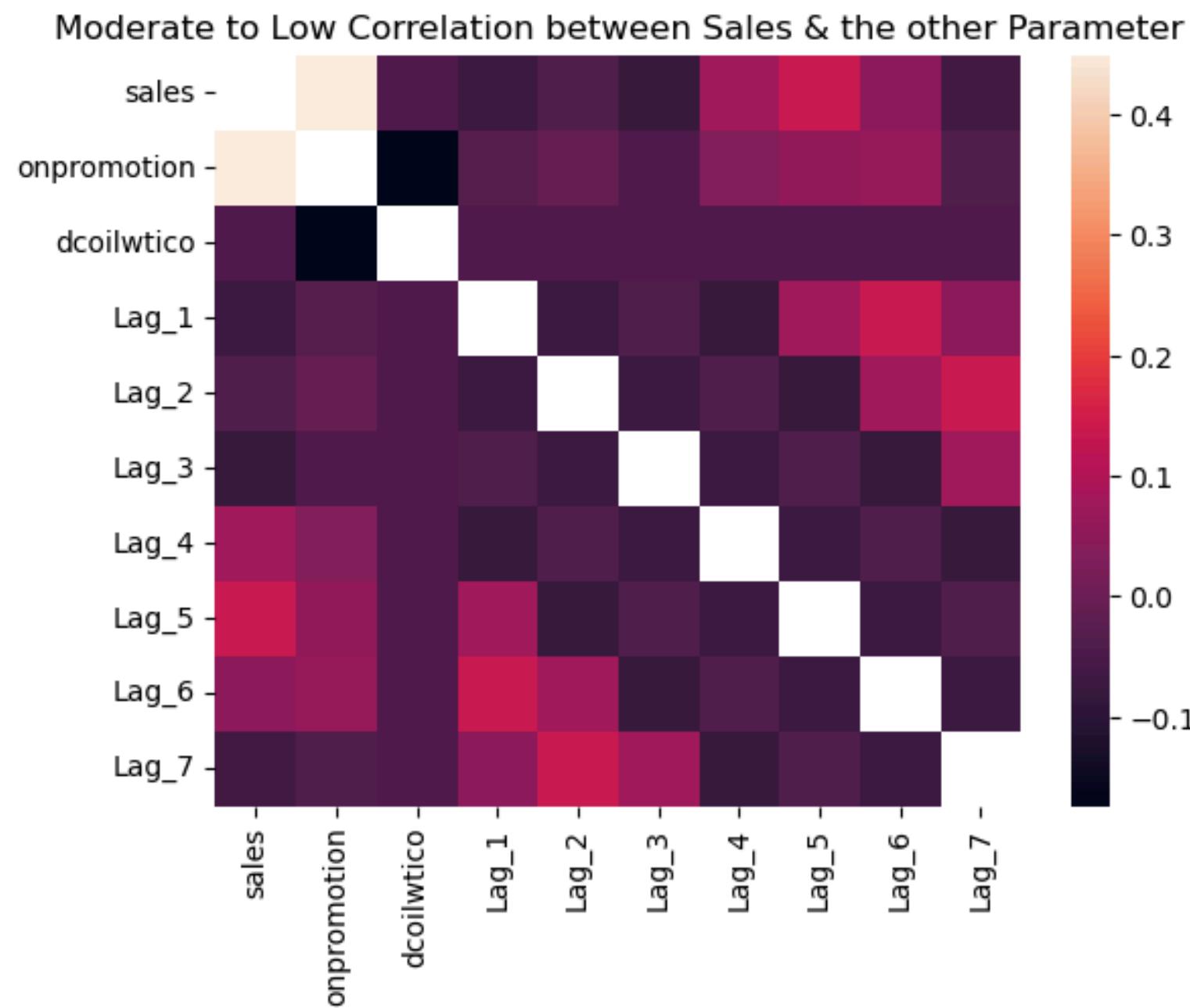


Data yang dipakai difokuskan pada kategori Groceries I dikarenakan:

1. Groceries (bahan pokok) merupakan kategori produk dengan permintaan yang relatif stabil dan frekuensi penjualan tinggi
2. Karena frekuensi penjualan yang tinggi, kategori ini memberikan kontribusi pendapatan yang tinggi juga
3. Dengan volume penjualan yang tinggi, groceries berisiko tinggi terhadap overstock atau stockout
4. Kategori groceries sering kali memiliki data historis yang lebih lengkap karena konsisten dijual setiap hari. Ini sangat berguna untuk time series forecasting menggunakan model seperti LSTM, ARIMA, atau SARIMA.
5. Pola konsumsi groceries dapat merefleksikan tren ekonomi makro atau perubahan perilaku konsumen.

Data Collection & Preparation

Dalam hal uji korelasi Penjualan Groceries I dengan beberapa variabel seperti promotion, harga minyak dunia, dan lag, tidak ditemukan korelasi yang kuat ($\text{cor} > |\text{abs } 0.5|$)



Hipotesa yang menyebabkan hal ini adalah kategori Groceries merupakan kebutuhan pokok akan selalu dibutuhkan oleh pelanggan dan tidak tergantung pada ketersediaan promosi dan perubahan harga minyak dunia.

Model Development

Model Yang Digunakan



1

ARIMA (AutoRegressive Integrated Moving Average):

Cocok untuk pola linier dengan asumsi stasioneritas.



2

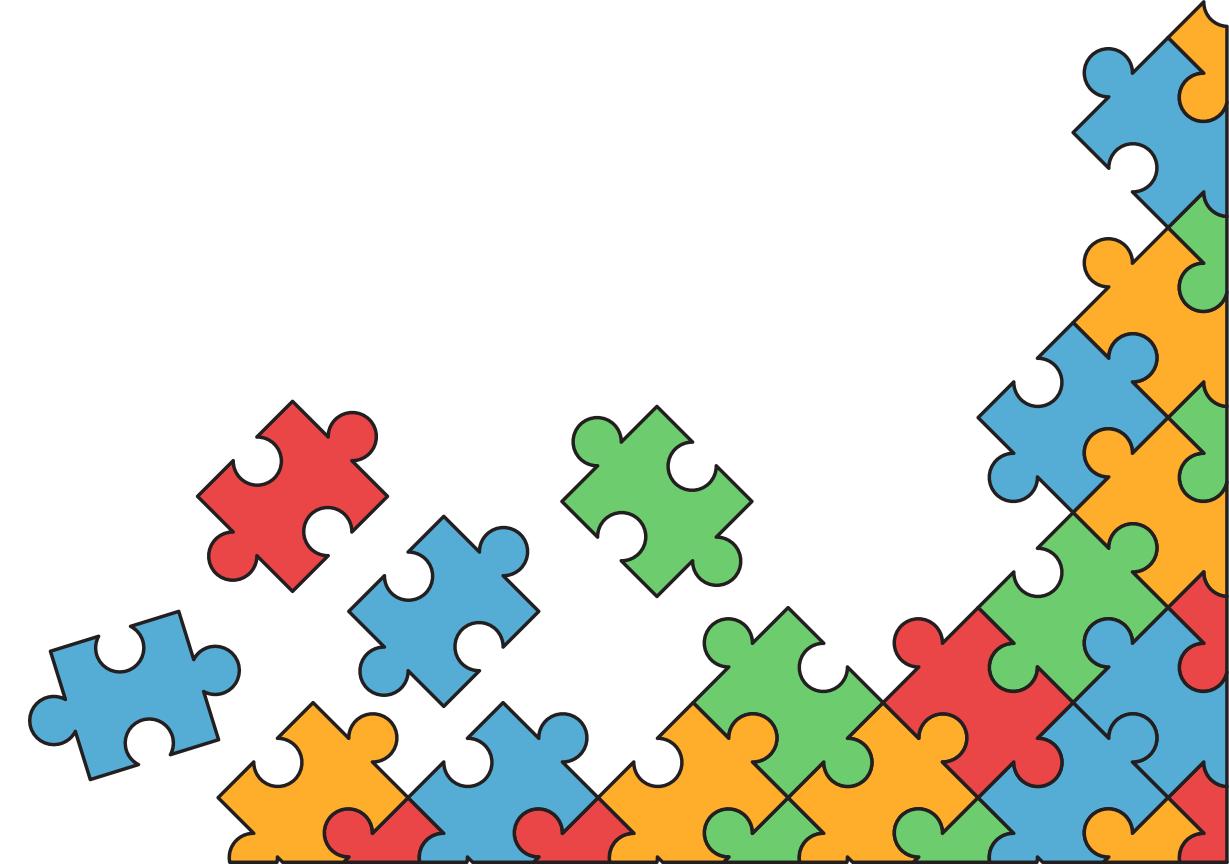
LSTM (Long Short-Term Memory):

Recurrent Neural Network yang unggul untuk pola time series kompleks dengan memori jangka panjang.

Model Development

ARIMA MODEL (dengan Data Bulanan & Harian)

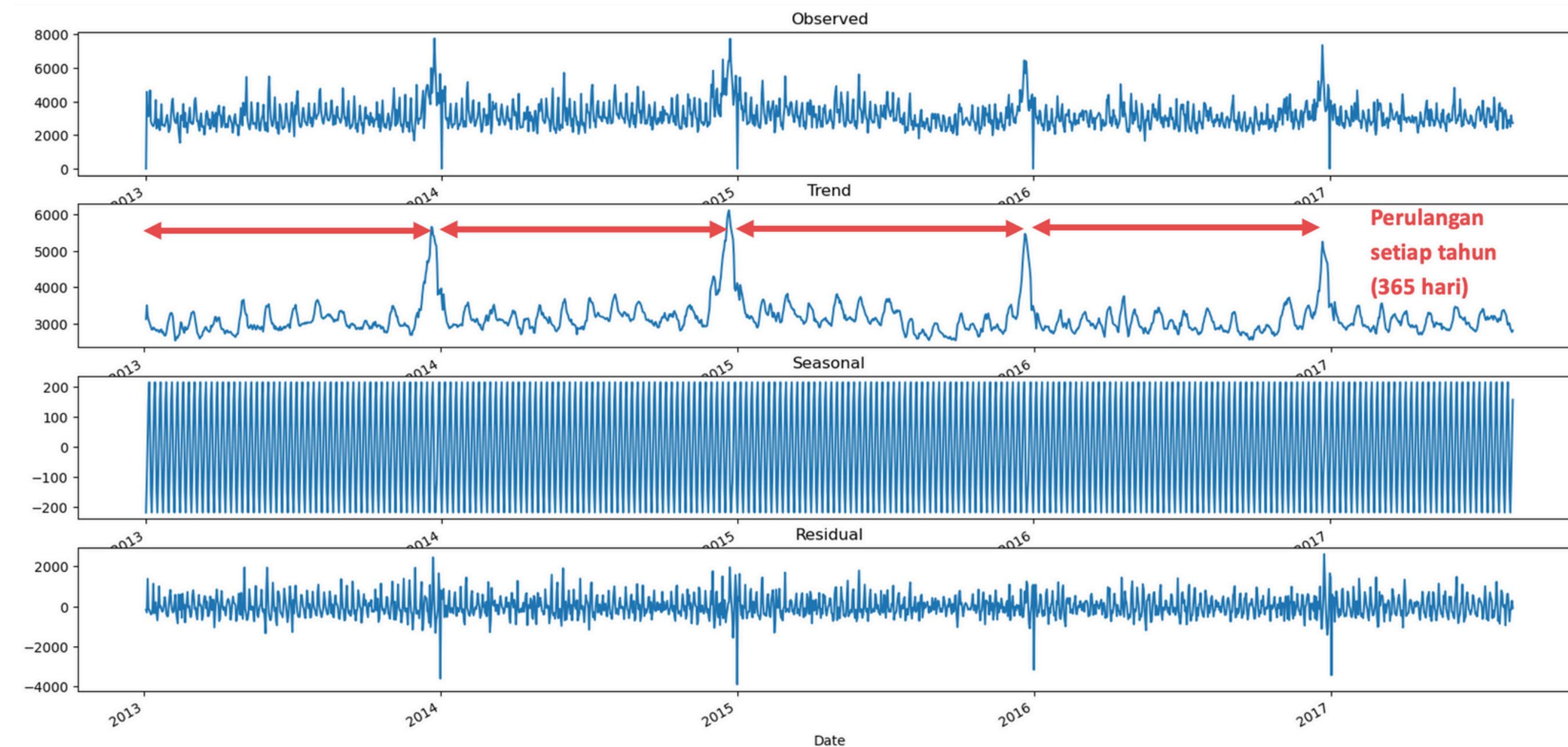
- Plot data time series
- Uji stasioneritas
- Diferensiasi Data
- Analisis ACF & PACF
- Bangun model ARIMA
- Training Model ARIMA (p,d,q)
- Forecasting
- Evaluasi Model



Model Development

ARIMA MODEL (dengan Data Harian)

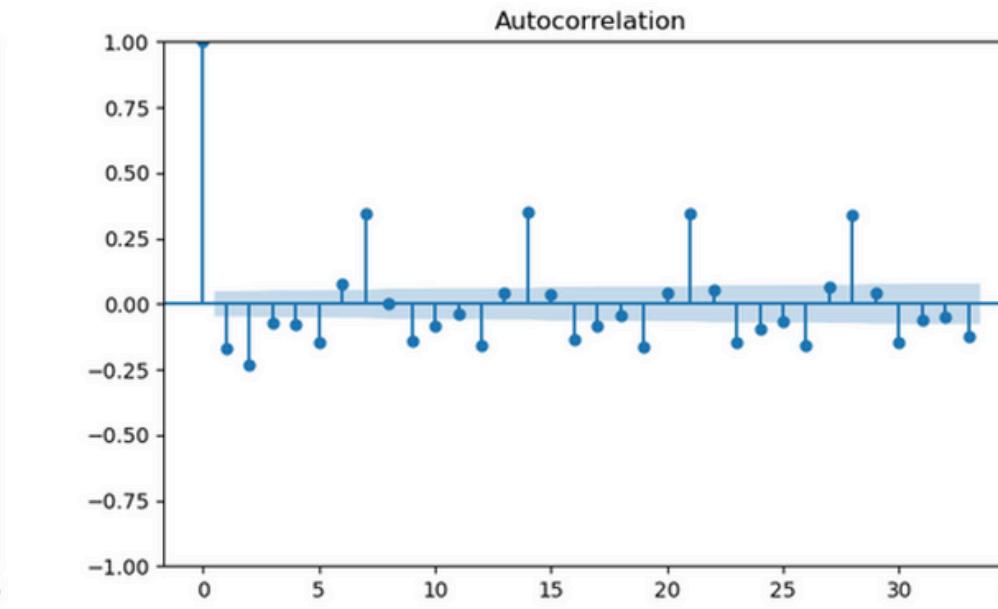
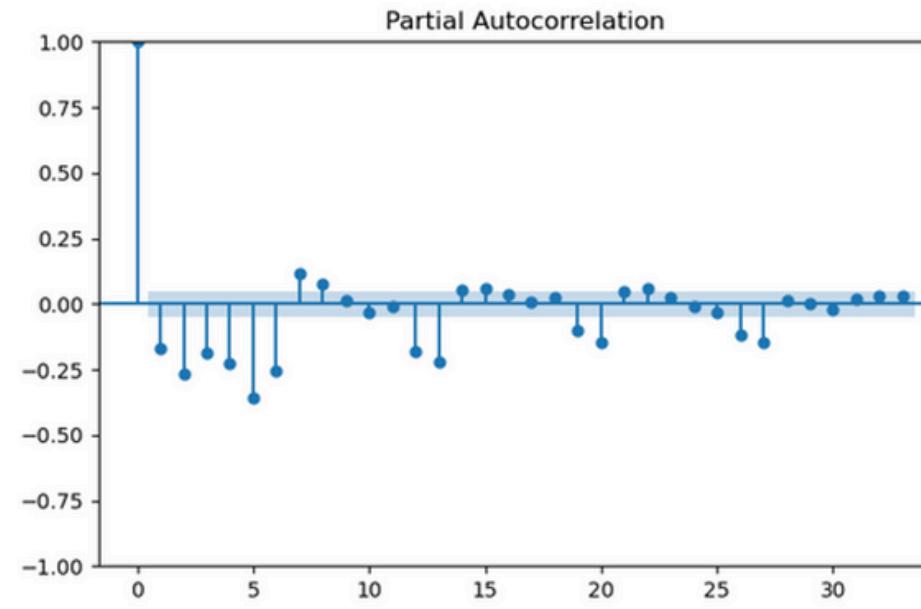
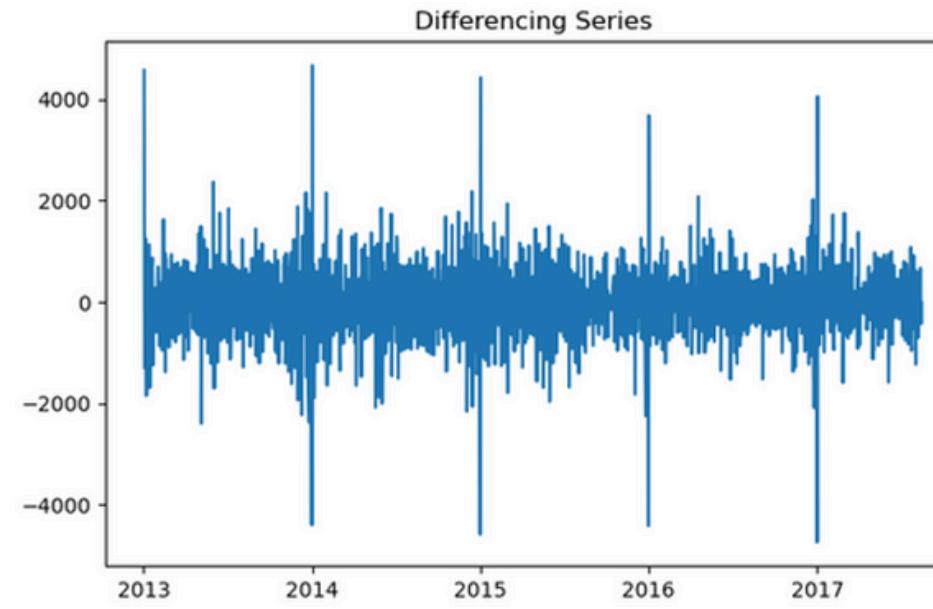
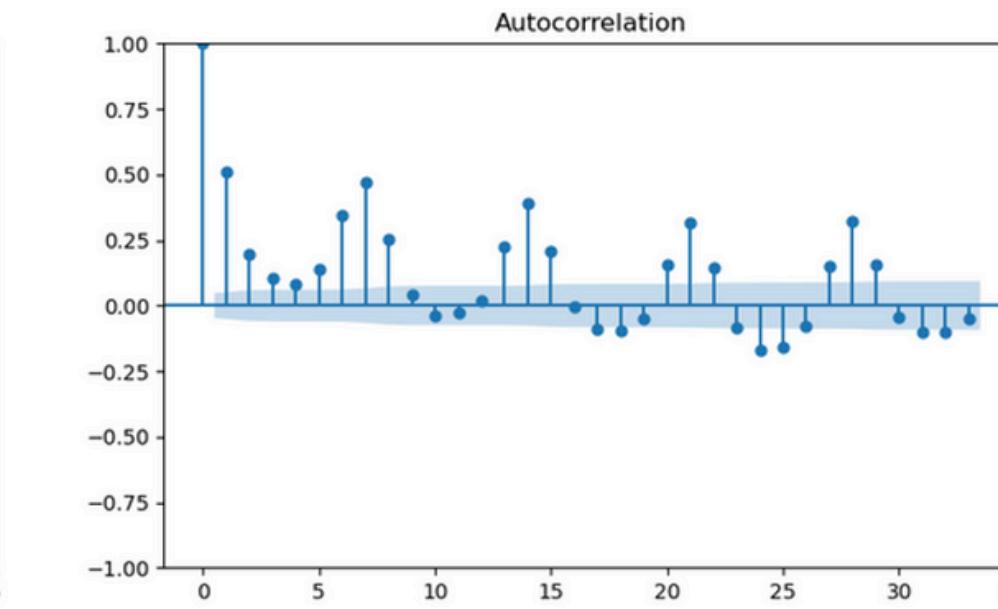
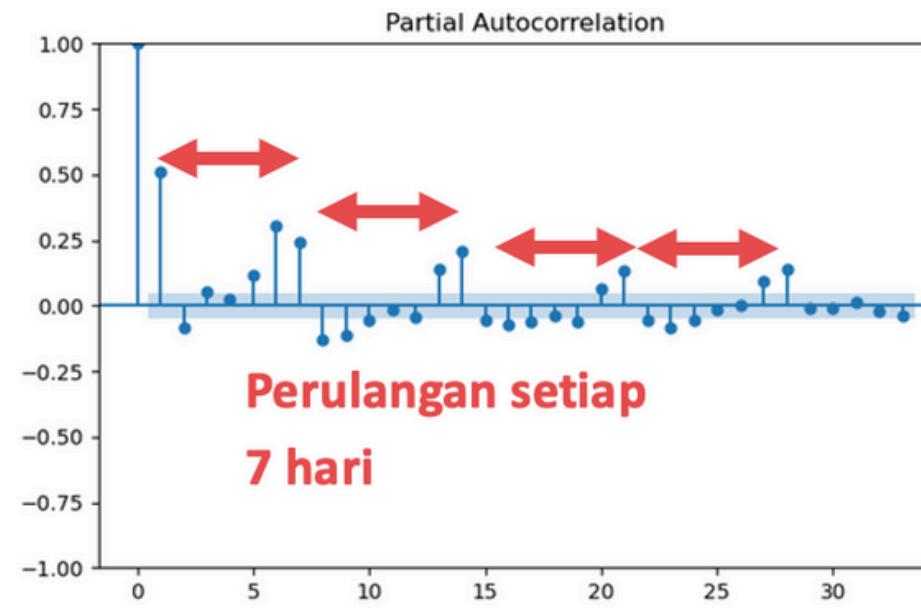
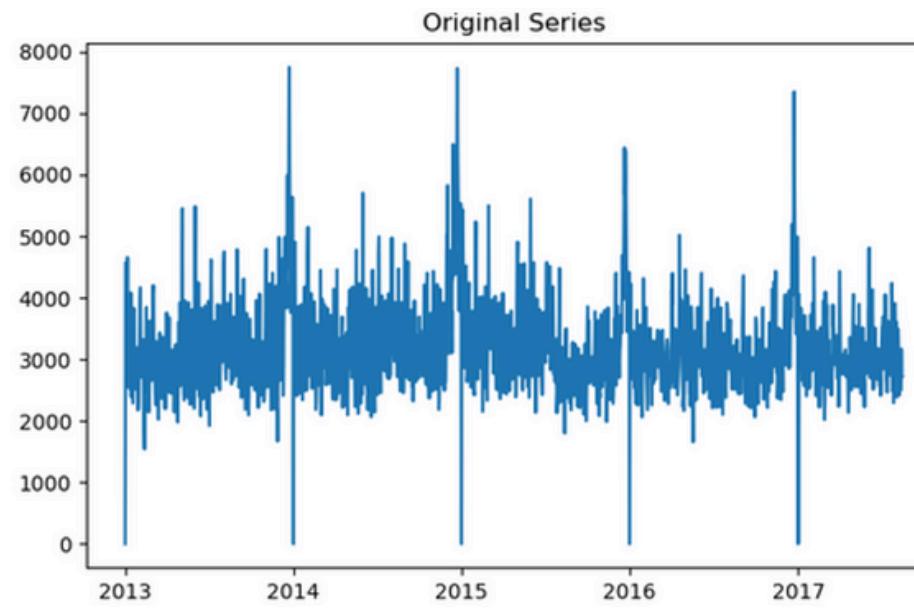
Eksplorasi Data (Data Harian)



Model Development

ARIMA MODEL (dengan Data Harian)

MODEL ARIMA (p,d,q) (P,D,Q,s)



Model Development

ARIMA MODEL (dengan Data Harian)

Dari PACF, kita bisa mengambilkan nilai AR(p) = 1,

Dari ACF, kita bisa mengambilkan nilai MA(q) = 1,

Dari hasil uji stationarity, kita bisa mengambilkan nilai d = 1,

Dari PACF, kita bisa mengambil kesimpulan seasonal = 7

Dari Data Decompose 'Trend', kita bisa mengambil kesimpulan seasonal = 365

Berdasarkan data tersebut, diputuskan bahwa akan dilakukan 3 metode ARIMA dalam pemodelan:

1.ARIMA Orde (1,0,1), Orde Seasonality (1,0,1,365)

2.ARIMA Orde (1,0,1), Orde Seasonality (1,0,1,7)

3.Auto ARIMA

Results

ARIMA MODEL (dengan Data Harian)

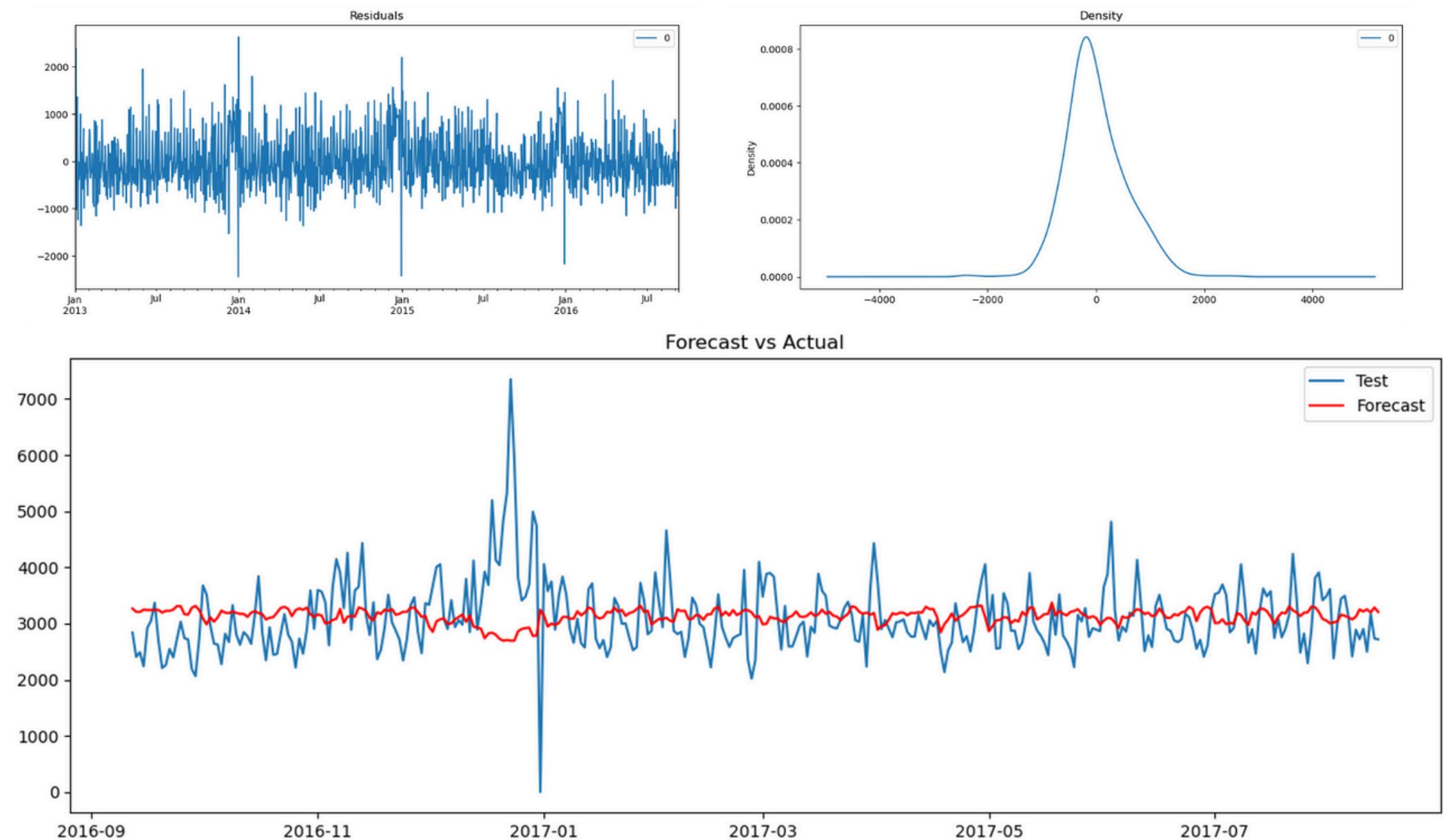
MODEL ARIMA - 365

```
SARIMAX Results
=====
Dep. Variable: Sales No. Observations: 1347
Model: ARIMA(1, 0, 1)x(1, 0, 1, 365) Log Likelihood: 10456.580
Date: Fri, 30 May 2025 AIC: 20872.611
Time: 18:24:01 BIC: 20903.845
Sample: 01-01-2013 HQIC: 20884.399
Covariance Type: opg
=====

coef std err z P>|z| [0.025 0.975]
-----
const 3133.8059 28.804 108.799 0.000 3077.352 3190.260
ar.L1 0.3509 0.041 8.466 0.000 0.270 0.432
ma.L1 0.2792 0.042 6.634 0.000 0.197 0.362
ar.S.L365 0.3355 0.281 1.195 0.232 -0.215 0.886
ma.S.L365 -0.4925 0.276 -1.785 0.074 -1.033 0.048
sigma2 3.106e+05 1.12e+04 27.720 0.000 2.89e+05 3.33e+05
=====

Ljung-Box (L1) (Q): 0.03 Jarque-Bera (JB):
Prob(Q): 0.87 Prob(JB): 110.57
Heteroskedasticity (H): 0.71 Skew: 0.00
Prob(H) (two-sided): 0.00 Kurtosis: 4.14
=====

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```



Results

ARIMA MODEL (dengan Data Harian)

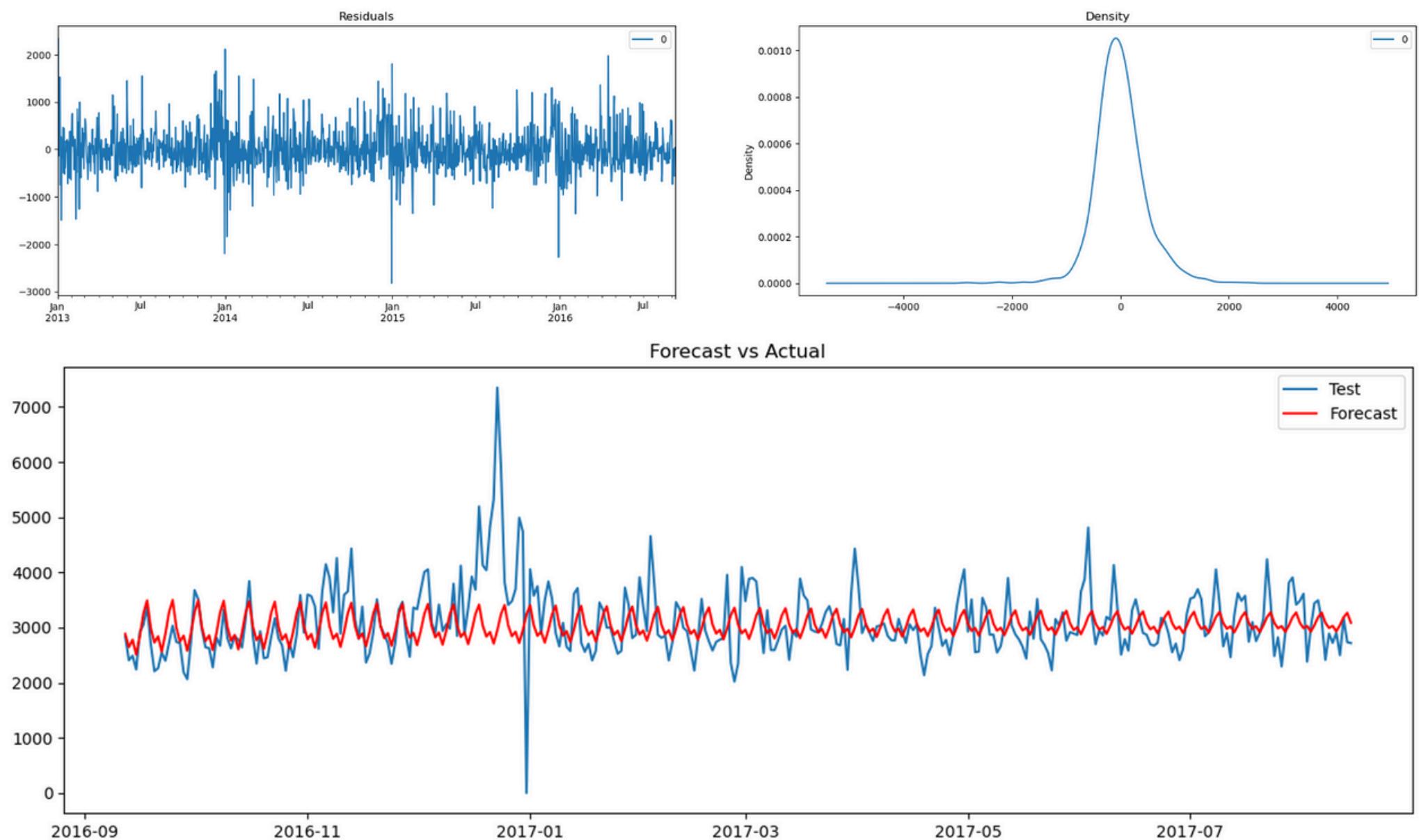
MODEL ARIMA - 7

```
SARIMAX Results
=====
Dep. Variable: Sales No. Observations: 1347
Model: ARIMA(1, 0, 1)x(1, 0, 1, 7) Log Likelihood: -10155.832
Date: Fri, 30 May 2025 AIC: 20323.664
Time: 18:24:06 BIC: 20354.897
Sample: 01-01-2013 HQIC: 20335.362
Covariance Type: opg
=====

coef std err z P>|z| [0.025 0.975]
-----
const 3133.7039 169.536 18.484 0.000 2801.419 3465.989
ar.L1 0.7240 0.031 23.250 0.000 0.663 0.785
ma.L1 -0.3281 0.040 -8.305 0.000 -0.406 -0.251
ar.S.L7 0.9781 0.005 192.401 0.000 0.968 0.988
ma.S.L7 -0.8028 0.018 -44.962 0.000 -0.838 -0.768
sigma2 2.059e+05 5264.906 39.116 0.000 1.96e+05 2.16e+05
=====

Ljung-Box (L1) (Q): 0.74 Jarque-Bera (JB): 650.53
Prob(Q): 0.39 Prob(JB): 0.00
Heteroskedasticity (H): 0.84 Skew: 0.21
Prob(H) (two-sided): 0.06 Kurtosis: 6.38
=====

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```



Results

ARIMA MODEL (dengan Data Harian)

MODEL ARIMA - Auto

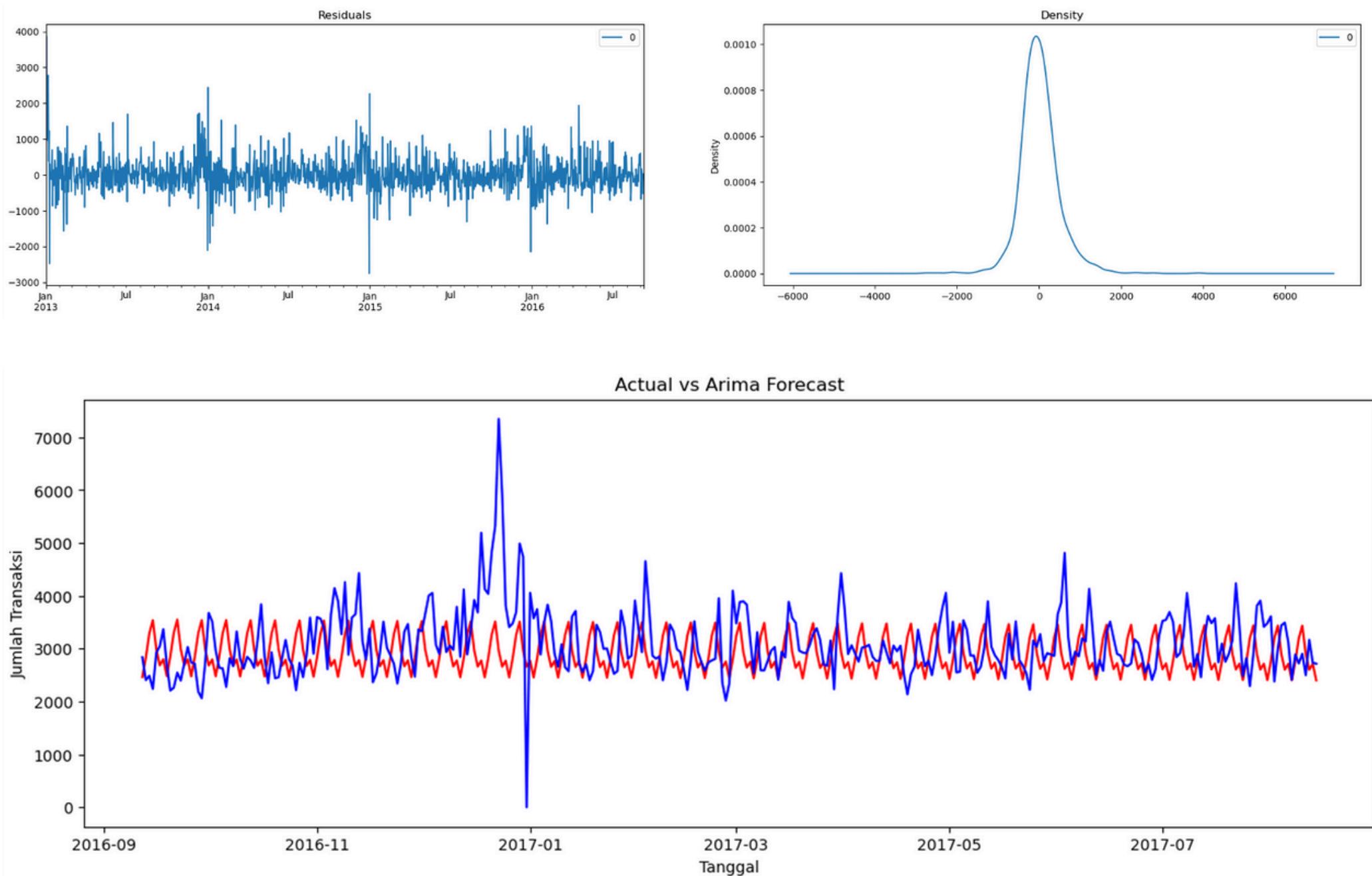
```
Best model: ARIMA(1,0,0)(2,0,2)[7]
Total fit time: 149.300 seconds

SARIMAX Results
=====
Dep. Variable: y No. Observations: 1347
Model: SARIMAX(1, 0, 0)x(2, 0, [1, 2], 7) Log Likelihood: -10183.46
Date: Fri, 30 May 2025 AIC: 20378.92
Time: 18:26:36 BIC: 20410.16
Sample: 01-01-2013 HQIC: 20380.62
Covariance Type: opg
=====

coef std err z P>|z| [0.025 0.975]
-----
ar.L1 0.4981 0.018 27.483 0.000 0.463 0.534
ar.S.L7 0.4969 1.430 0.347 0.728 -2.307 3.300
ar.S.L14 0.5021 1.429 0.351 0.725 -2.299 3.304
ma.S.L7 -0.3164 1.424 -0.222 0.824 -3.108 2.476
ma.S.L14 -0.4274 1.174 -0.364 0.716 -2.729 1.874
sigma2 2.132e+05 5372.583 39.682 0.000 2.03e+05 2.24e+05
=====

Ljung-Box (L1) (Q): 4.01 Jarque-Bera (JB): 780.77
Prob(Q): 0.05 Prob(JB): 0.00
Heteroskedasticity (H): 0.82 Skew: 0.22
Prob(H) (two-sided): 0.03 Kurtosis: 6.70
=====

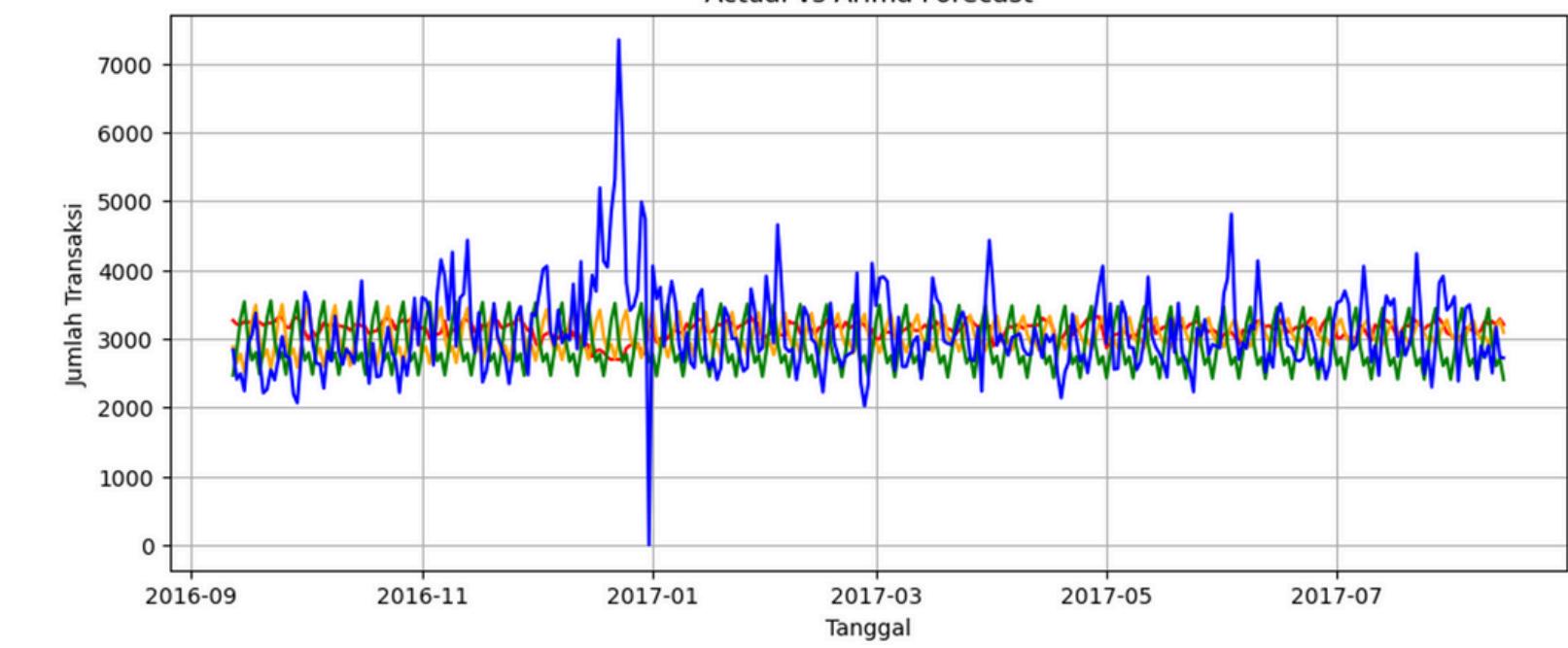
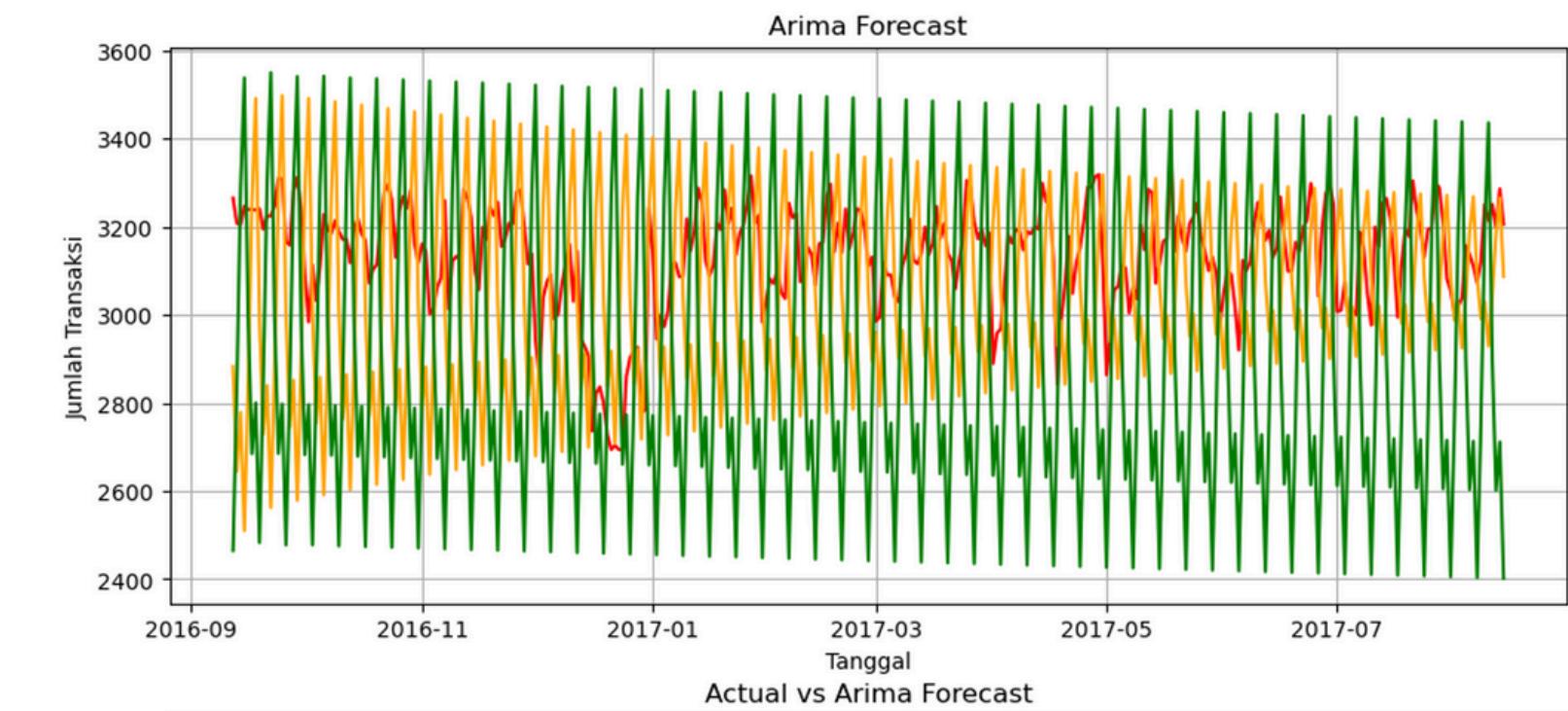
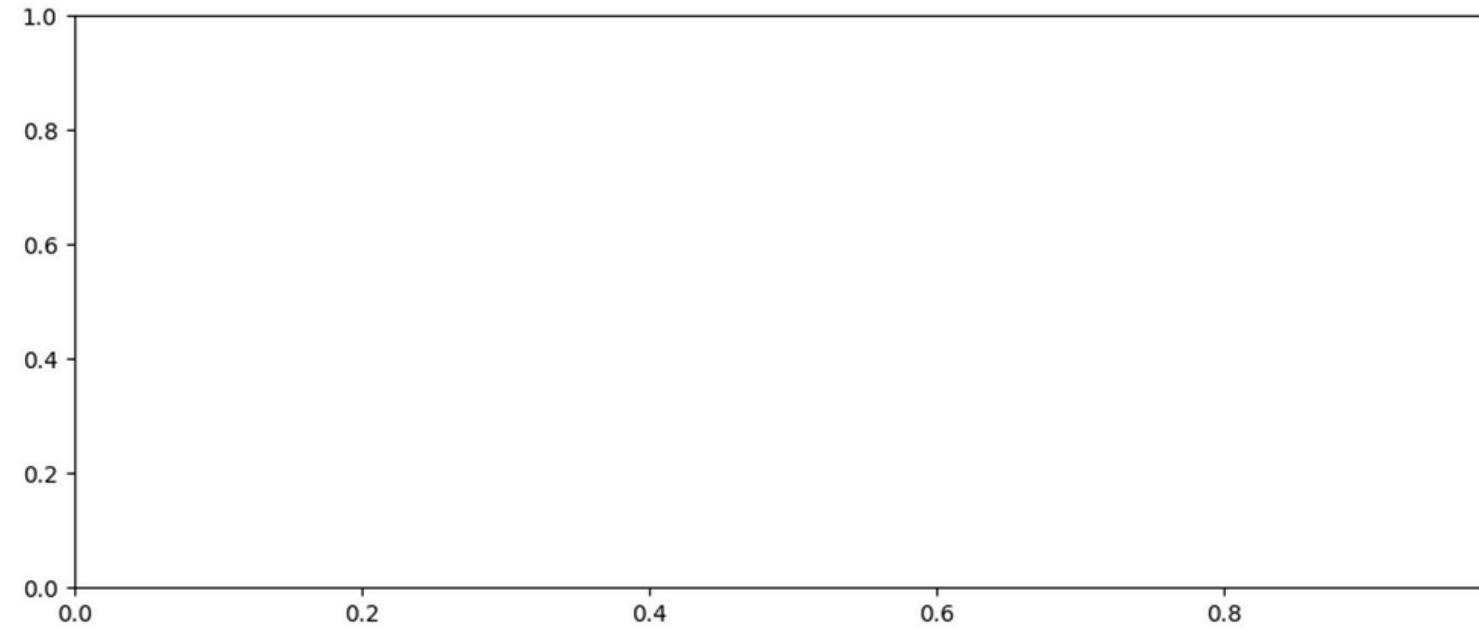
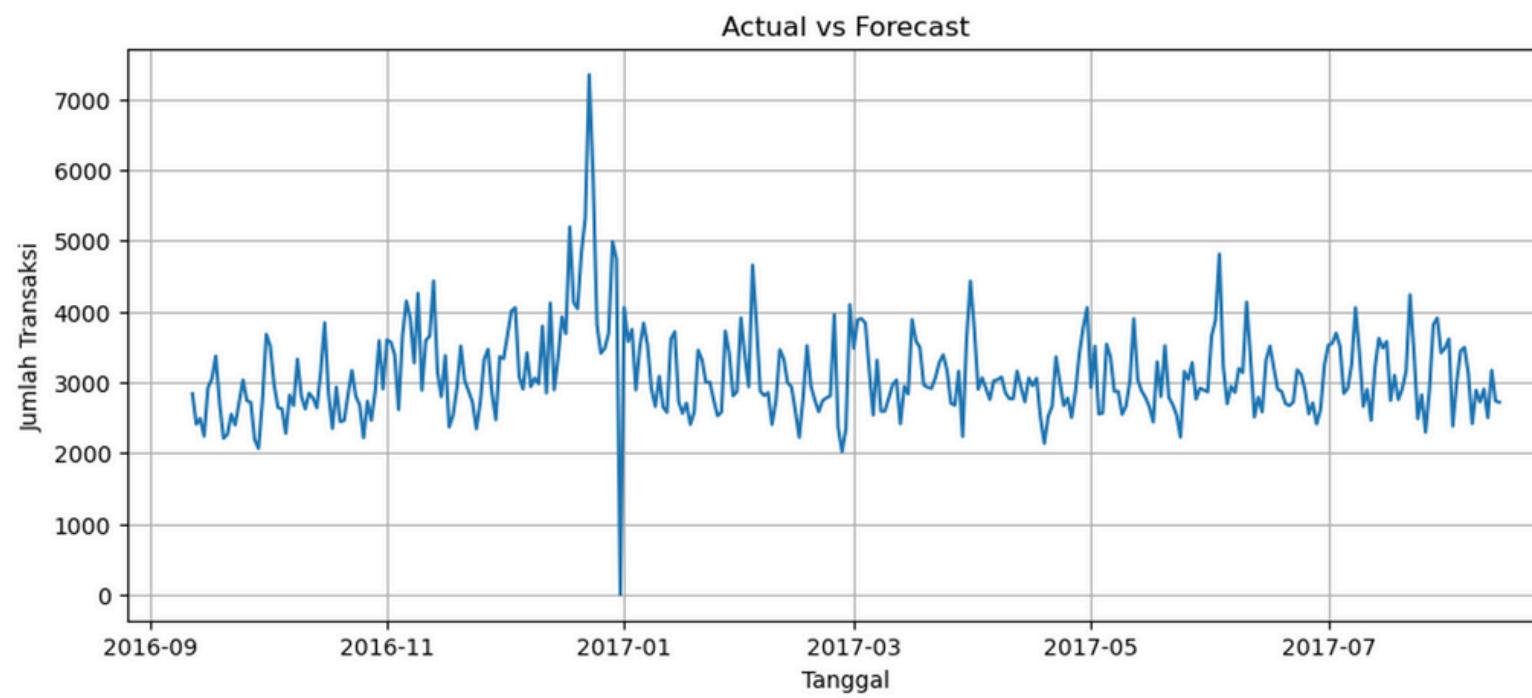
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```



Results

ARIMA MODEL (dengan Data Harian)

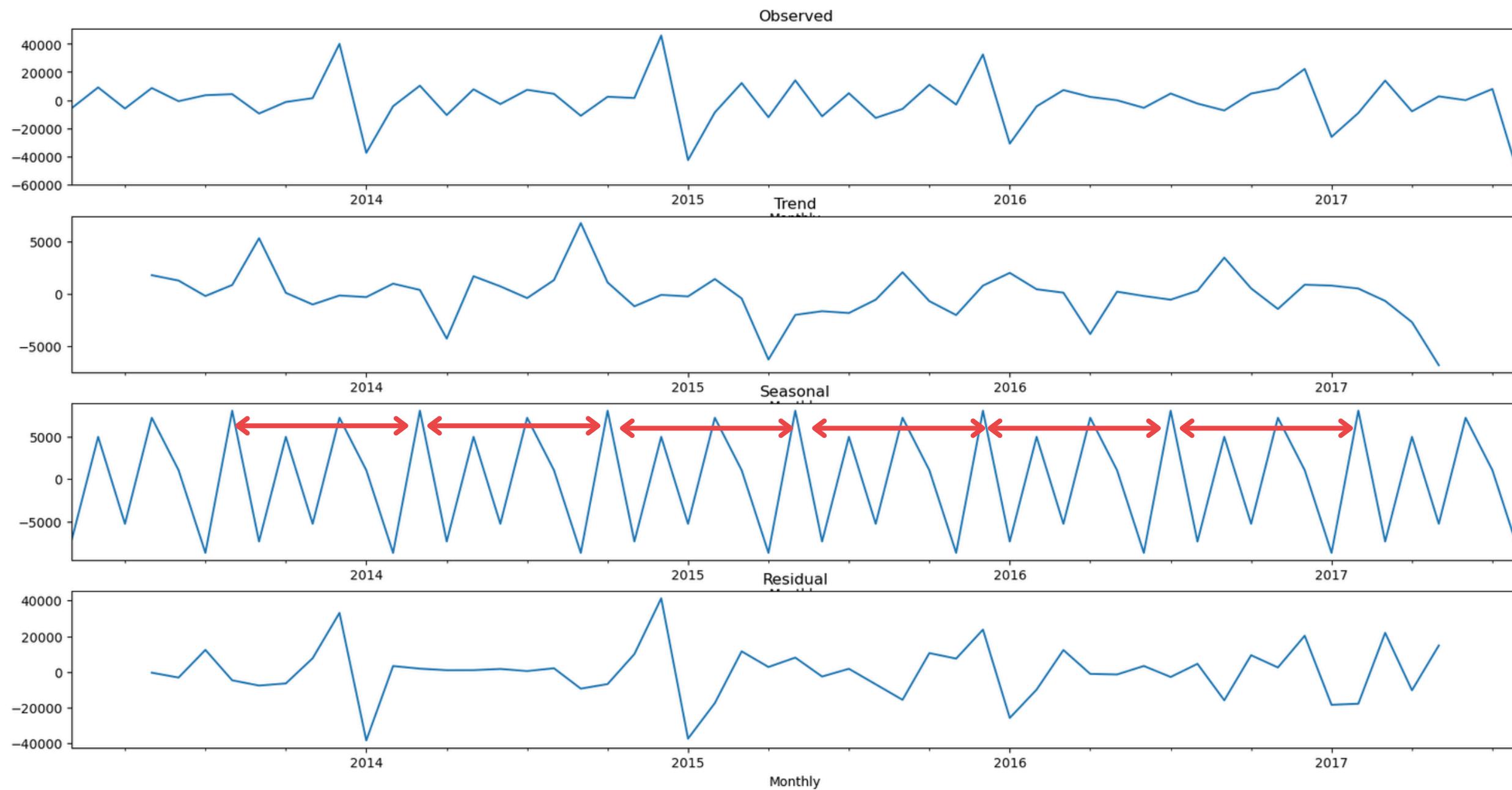
MODEL ARIMA - Comparison



Model Development

ARIMA MODEL (dengan Data Bulanan)

Eksplorasi Data (Data Bulanan)

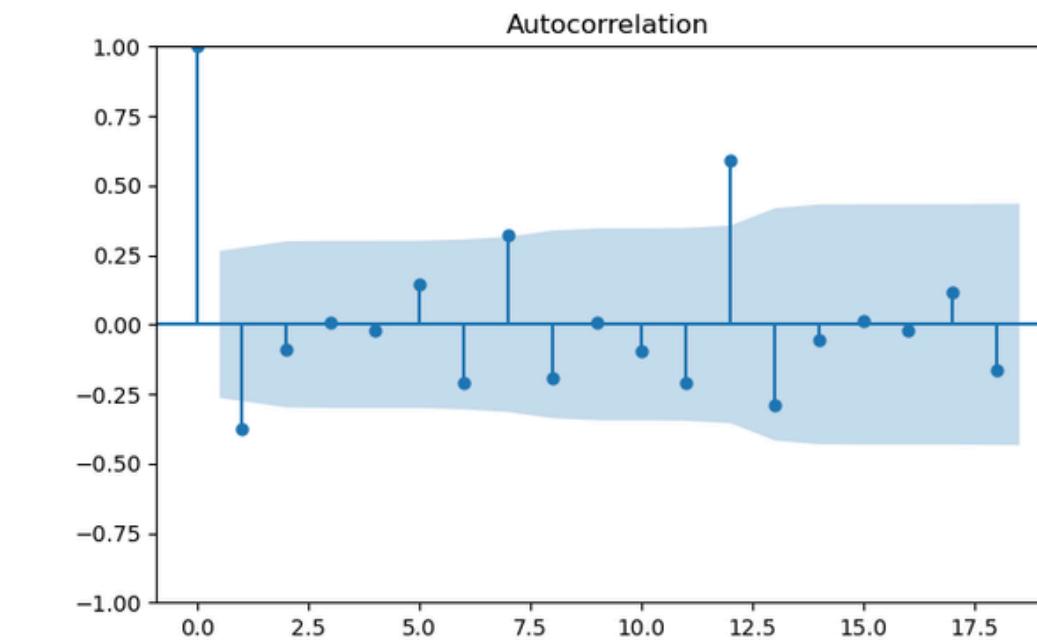
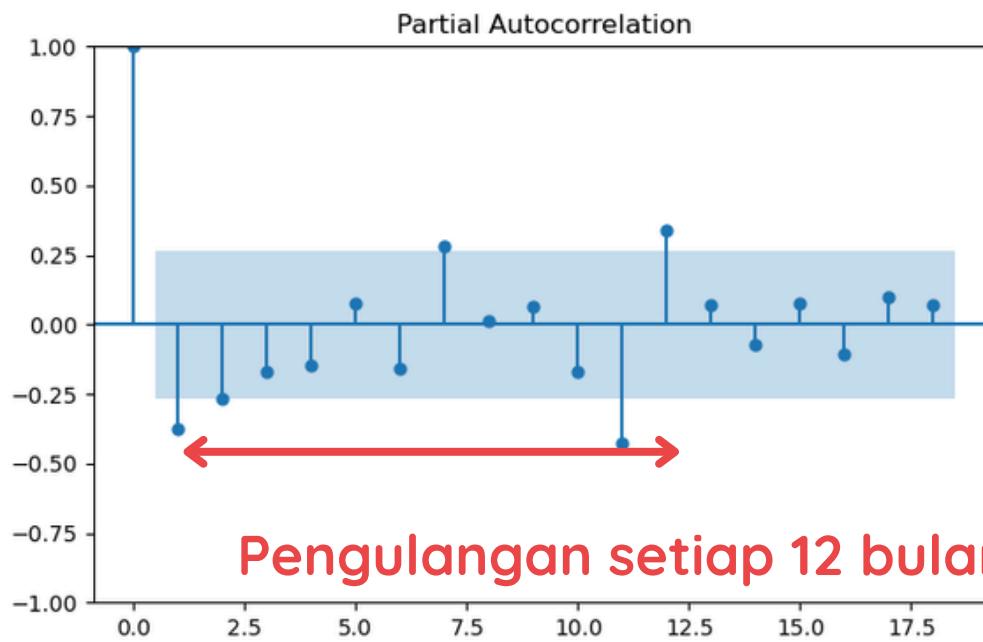
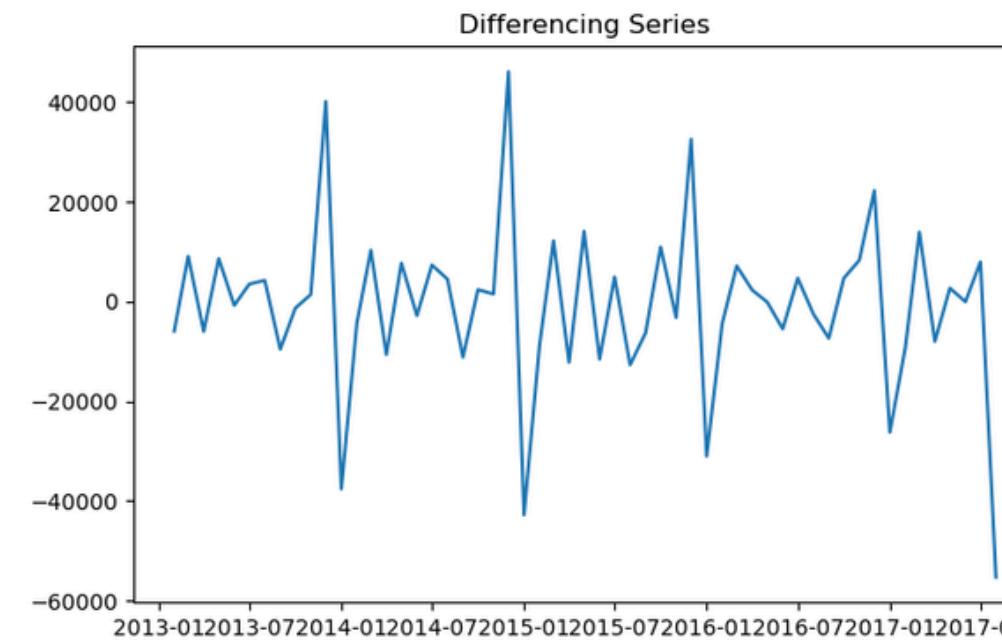
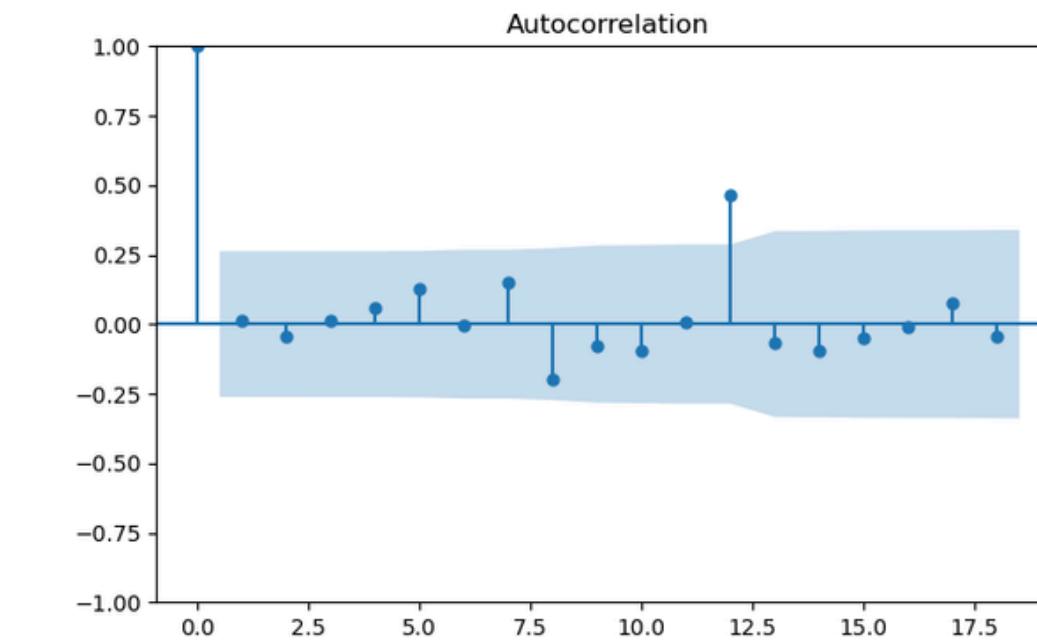
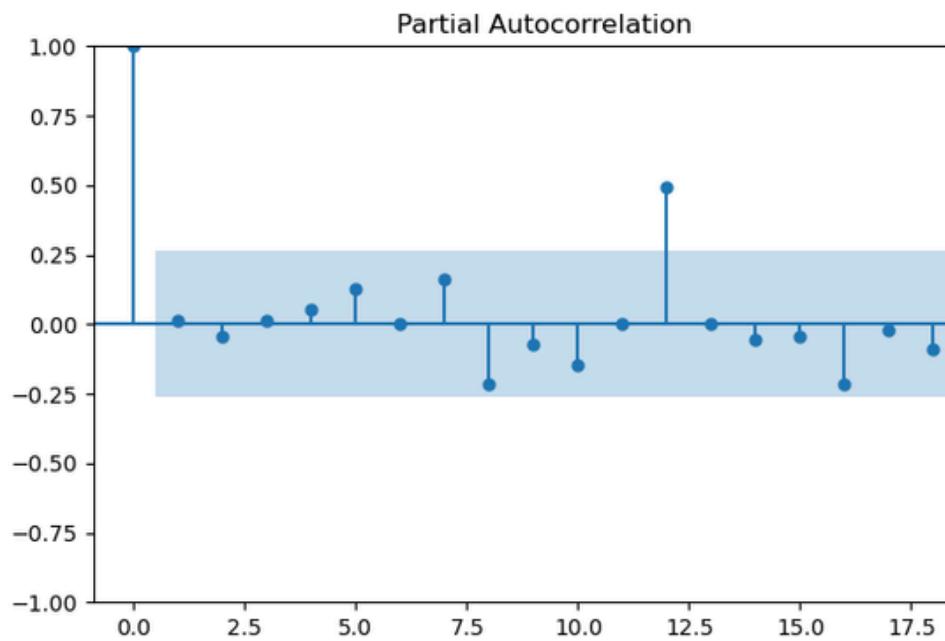
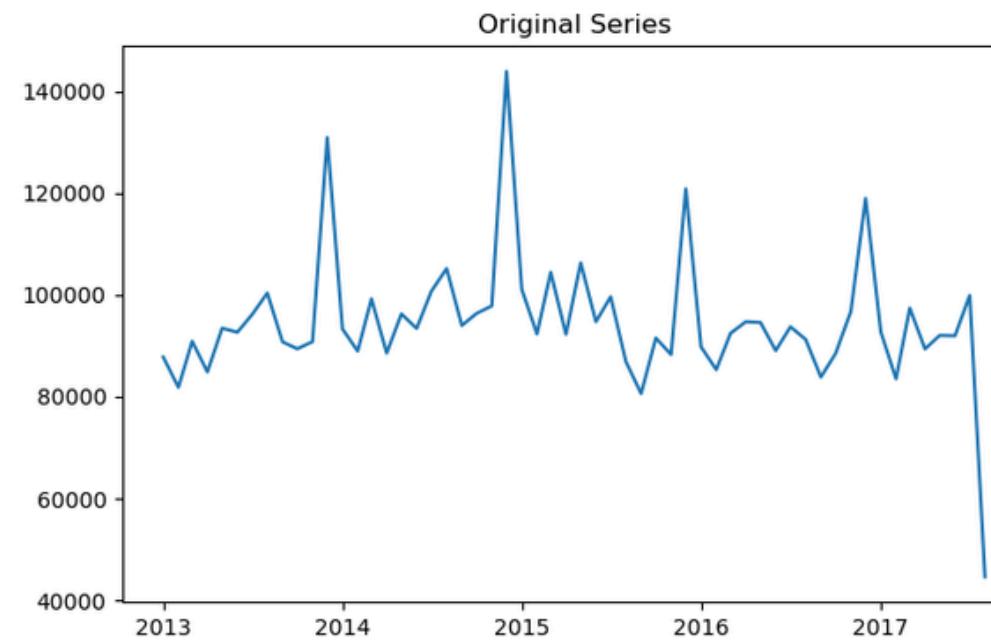


Pengulangan setiap 7 bulan

Model Development

ARIMA MODEL (dengan Data Bulanan)

MODEL ARIMA (p,d,q) (P,D,Q,s)



Pengulangan setiap 12 bulan

Model Development

ARIMA MODEL (dengan Data Bulanan)

Dari PACF, kita bisa mengambilkan nilai AR(p) = 1,

Dari ACF, kita bisa mengambilkan nilai MA(q) = 1,

Dari hasil uji stationarity, kita bisa mengambilkan nilai d = 0,

Dari PACF, kita bisa mengambil kesimpulan seasonal = 12

Dari Data Decompose 'Trend', kita bisa mengambil kesimpulan seasonal = 7

Berdasarkan data tersebut, diputuskan bahwa akan dilakukan 3 metode ARIMA dalam pemodelan:

1.ARIMA Orde (1,0,1), Orde Seasonality (1,0,1,12)

2.ARIMA Orde (1,0,1), Orde Seasonality (1,0,1,7)

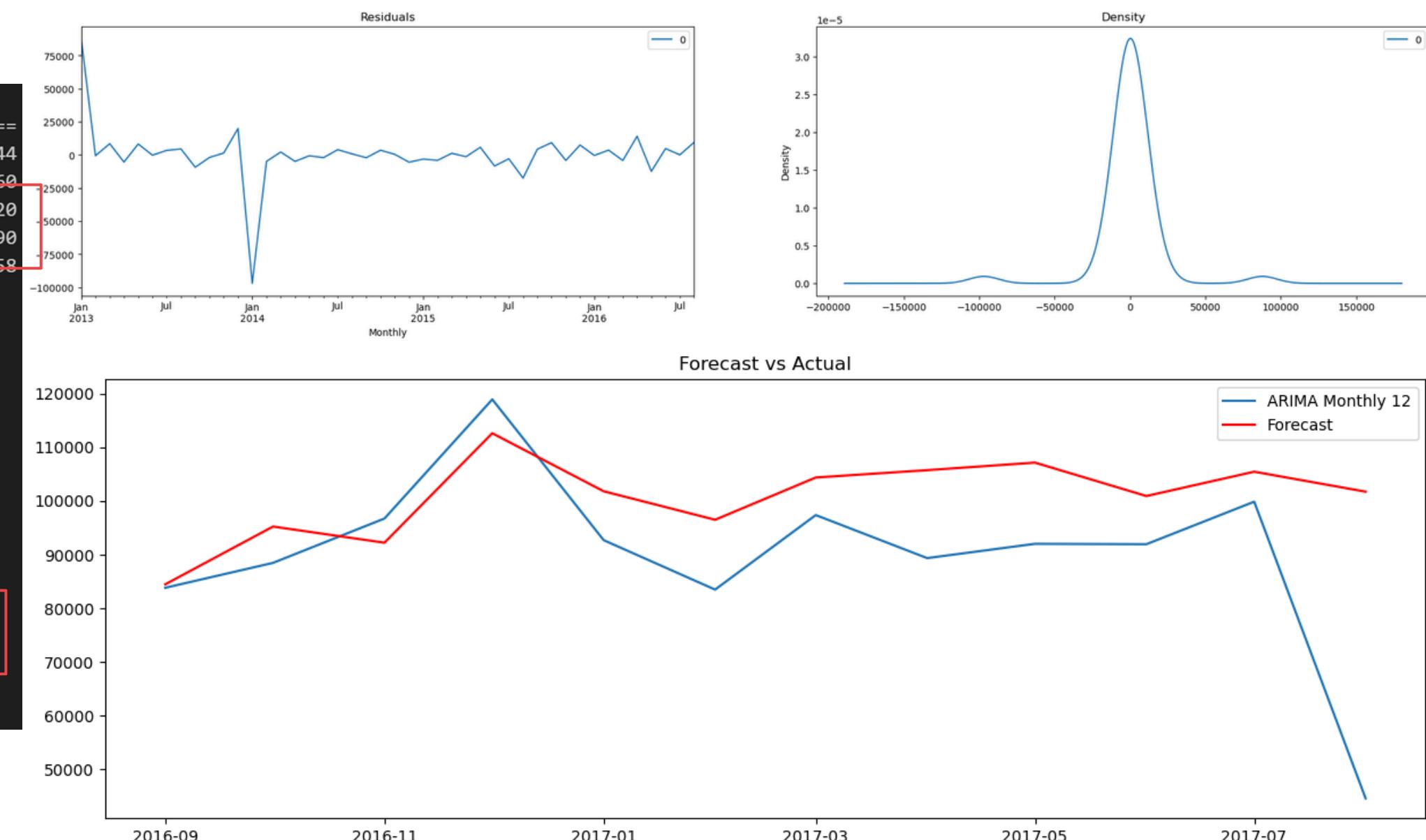
3.Auto ARIMA

Results

ARIMA MODEL (dengan Data Bulanan)

MODEL ARIMA - 12

```
SARIMAX Results
=====
Dep. Variable: Sales No. Observations: 44
Model: ARIMA(1, 1, 1)x(1, 1, 1, 12) Log Likelihood 315.360
Date: Fri, 06 Jun 2025 AIC 640.720
Time: 09:55:51 BIC 647.890
Sample: 01-01-2013 HQIC 643.058
Covariance Type: opg
=====
            coef    std err        z   P>|z|      [0.025]     [0.975]
---+-----+-----+-----+-----+-----+-----+
ar.L1    -0.1091    2.185   -0.050    0.960    -4.391     4.173
ma.L1     0.0458    2.212    0.021    0.983    -4.290     4.381
ar.S.L12  -0.9901    0.266   -3.726    0.000    -1.511    -0.469
ma.S.L12  0.9629    0.536    1.796    0.072    -0.088     2.013
sigma2   3.947e+07  3.12e-08  1.26e+15  0.000    3.95e+07  3.95e+07
=====
Ljung-Box (L1) (Q): 3.73 Jarque-Bera (JB): 1.02
Prob(Q): 0.05 Prob(JB): 0.60
Heteroskedasticity (H): 6.29 Skew: -0.28
Prob(H) (two-sided): 0.01 Kurtosis: 3.69
=====
```



Results

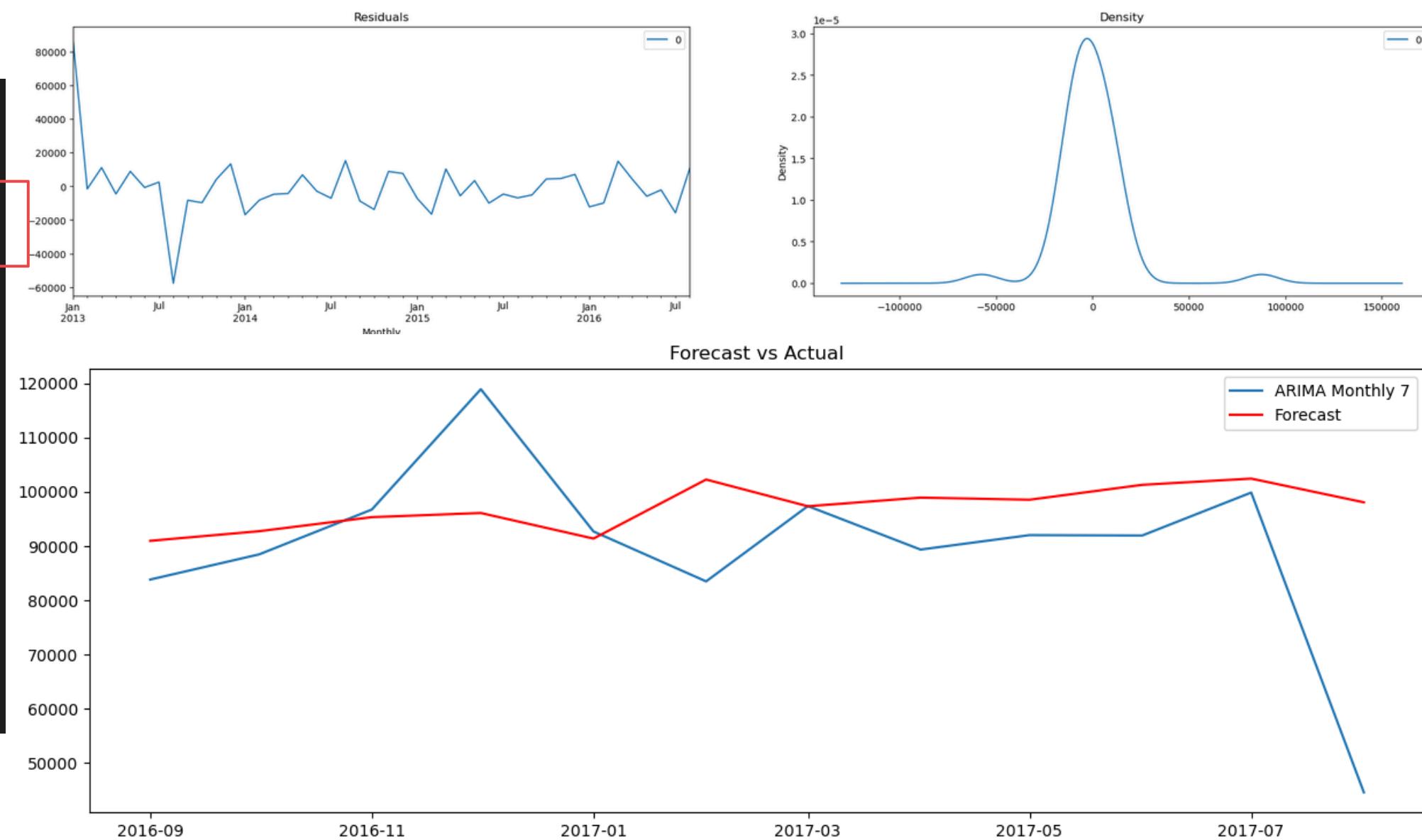
ARIMA MODEL (dengan Data Bulanan)

MODEL ARIMA - 7

```
SARIMAX Results
=====
Dep. Variable: Sales No. Observations: 44
Model: ARIMA(1, 1, 1)x(1, 1, 1, 7) Log Likelihood: -380.363
Date: Fri, 06 Jun 2025 AIC: 770.725
Time: 09:55:54 BIC: 778.643
Sample: 01-01-2013 HQIC: 773.489
Covariance Type: opg
=====

            coef    std err        z   P>|z|      [0.025      0.975]
-----+-----+-----+-----+-----+-----+-----+
ar.L1    0.4955    1.414    0.350    0.726    -2.276     3.267
ma.L1   -0.5493    1.356   -0.405    0.685    -3.207     2.109
ar.S.L7  0.4132    0.262    1.575    0.115    -0.101     0.927
ma.S.L7 -0.9167    0.551   -1.663    0.096    -1.997     0.164
sigma2  9.138e+07  7.66e-09  1.19e+16  0.000    9.14e+07  9.14e+07
=====

Ljung-Box (L1) (Q): 1.37 Jarque-Bera (JB): 2.01
Prob(Q): 0.24 Prob(JB): 0.37
Heteroskedasticity (H): 1.02 Skew: 0.27
Prob(H) (two-sided): 0.97 Kurtosis: 1.98
=====
```



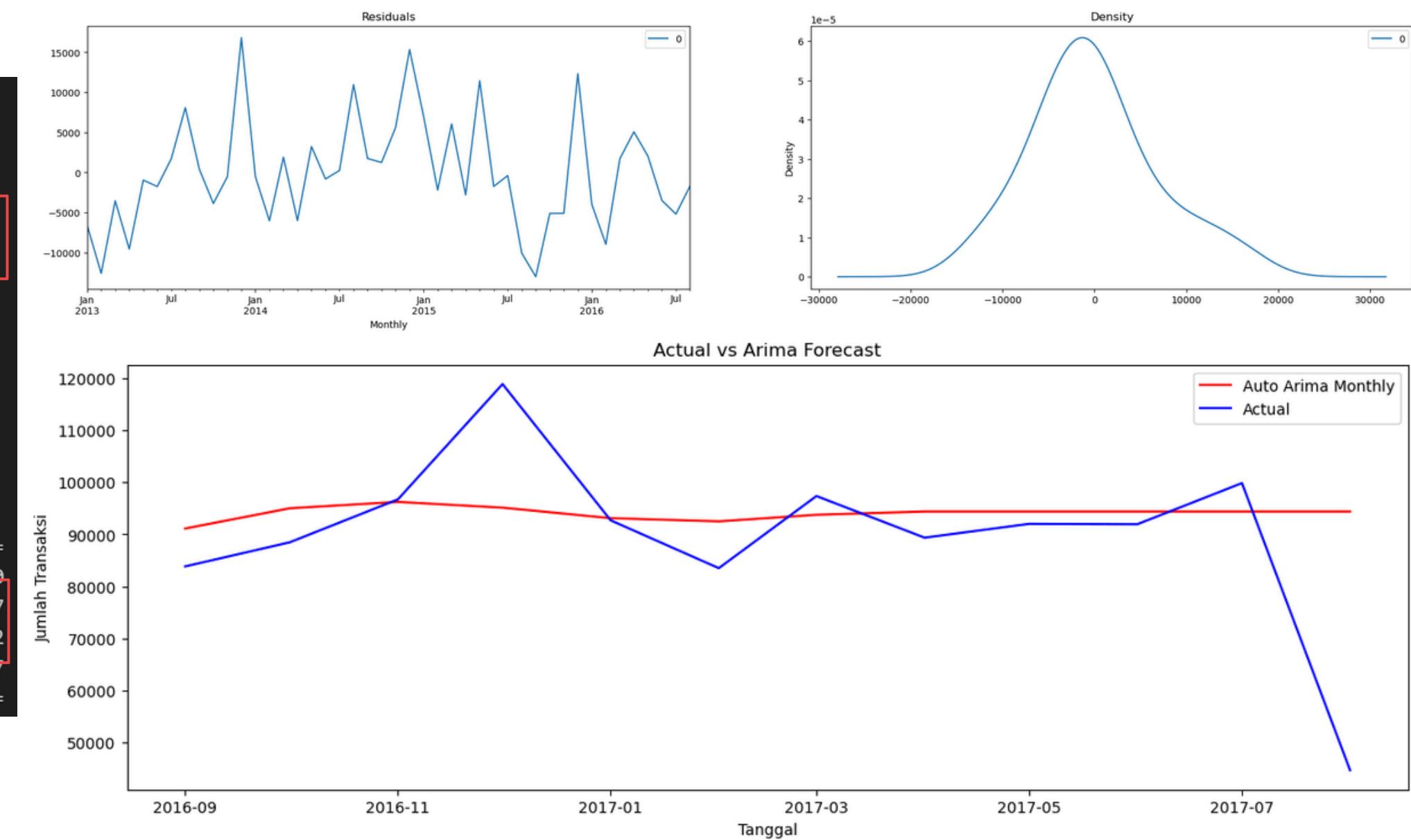
Results

ARIMA MODEL (dengan Data Bulanan)

MODEL ARIMA - Auto

```
SARIMAX Results
=====
Dep. Variable:                      y     No. Observations:                  44
Model:             SARIMAX(0, 0, [1], 7)   Log Likelihood:                -450.859
Date:                Fri, 06 Jun 2025   AIC:                         907.717
Time:                      09:55:56     BIC:                         913.070
Sample:                 01-01-2013   HQIC:                        909.702
                           - 08-01-2016
Covariance Type:            opg

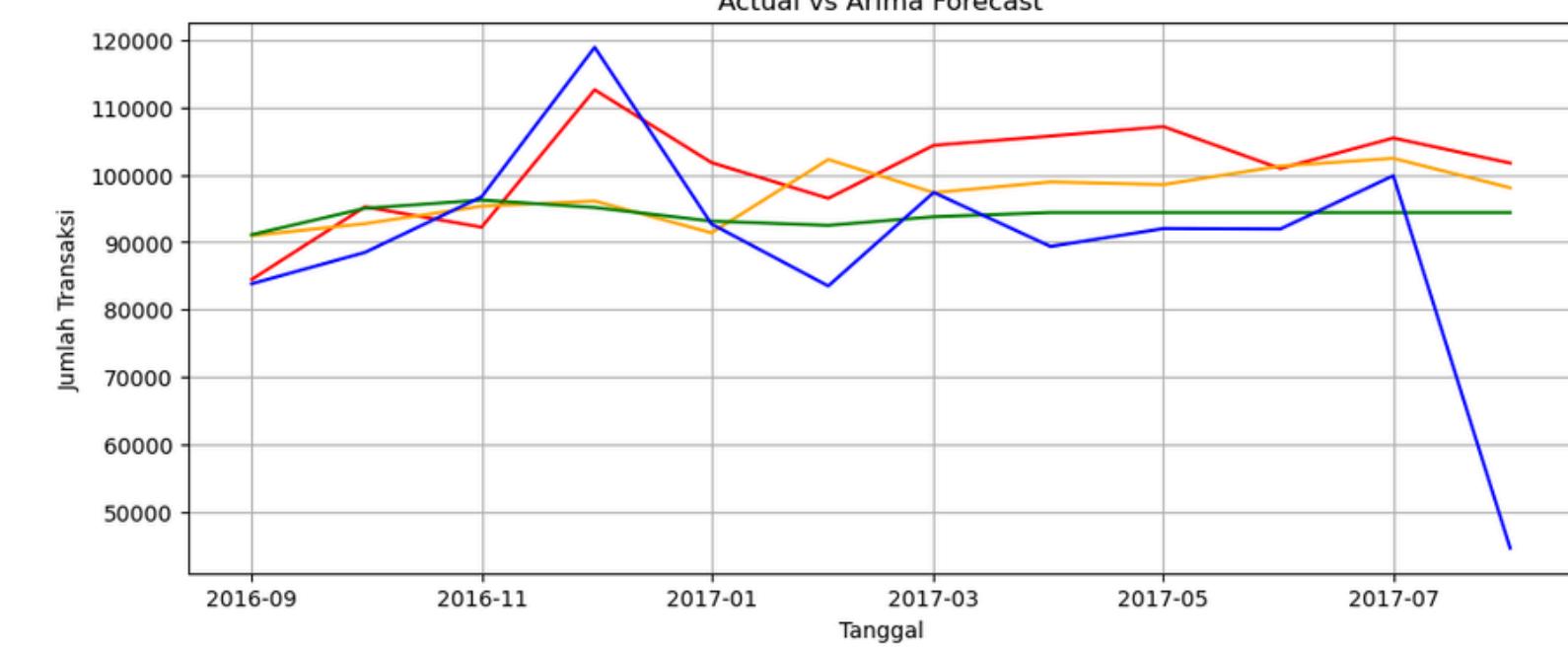
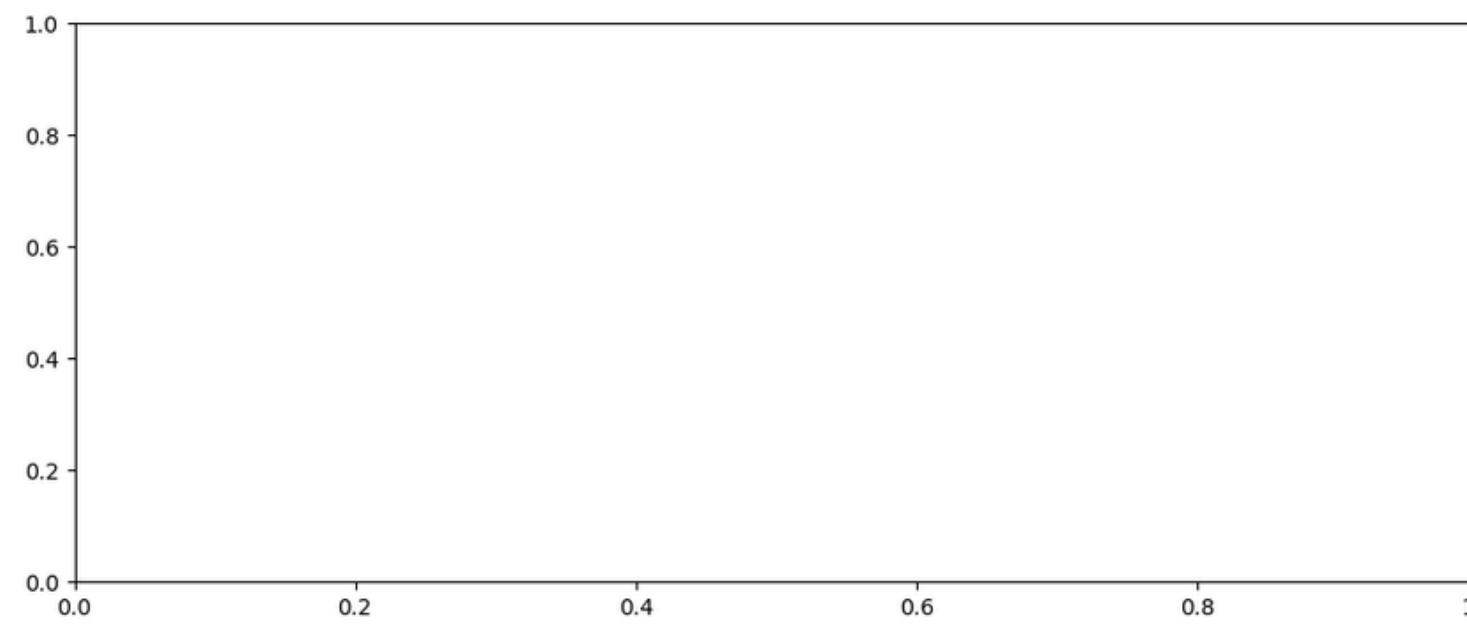
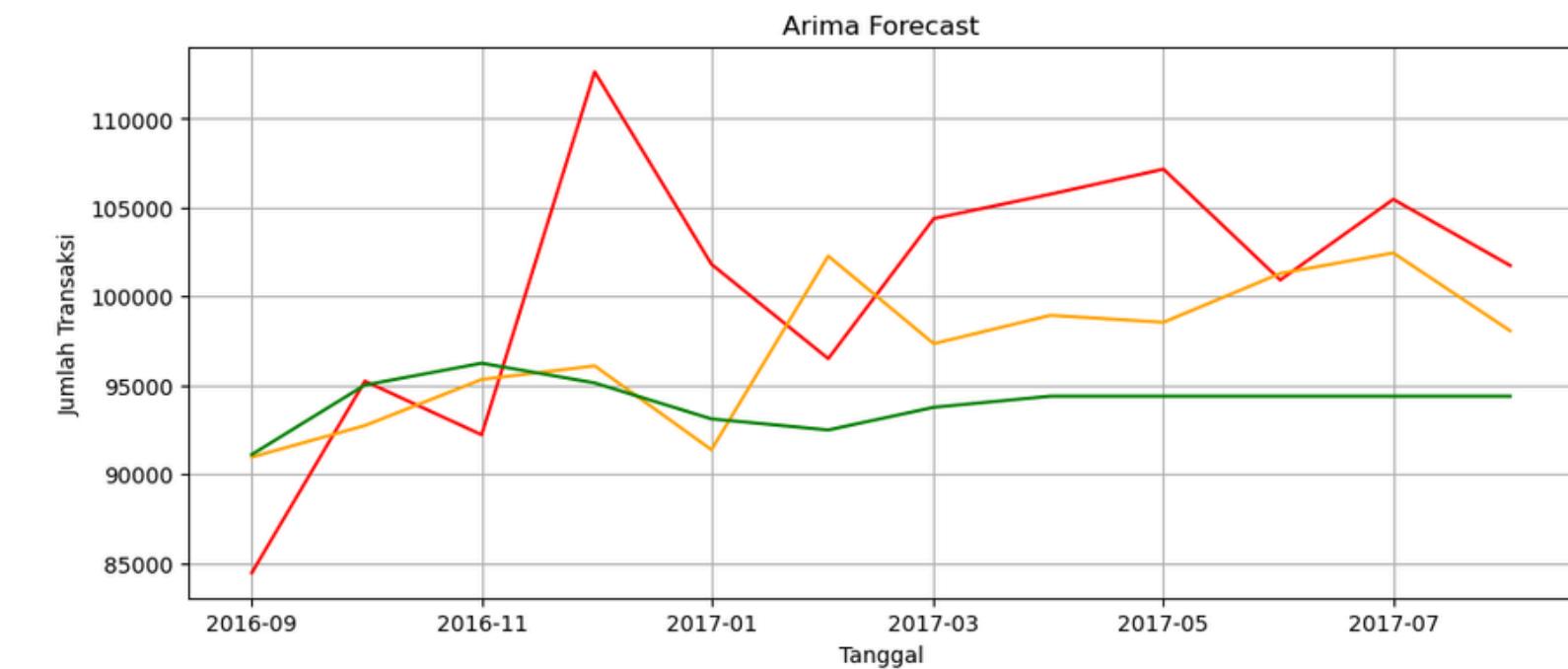
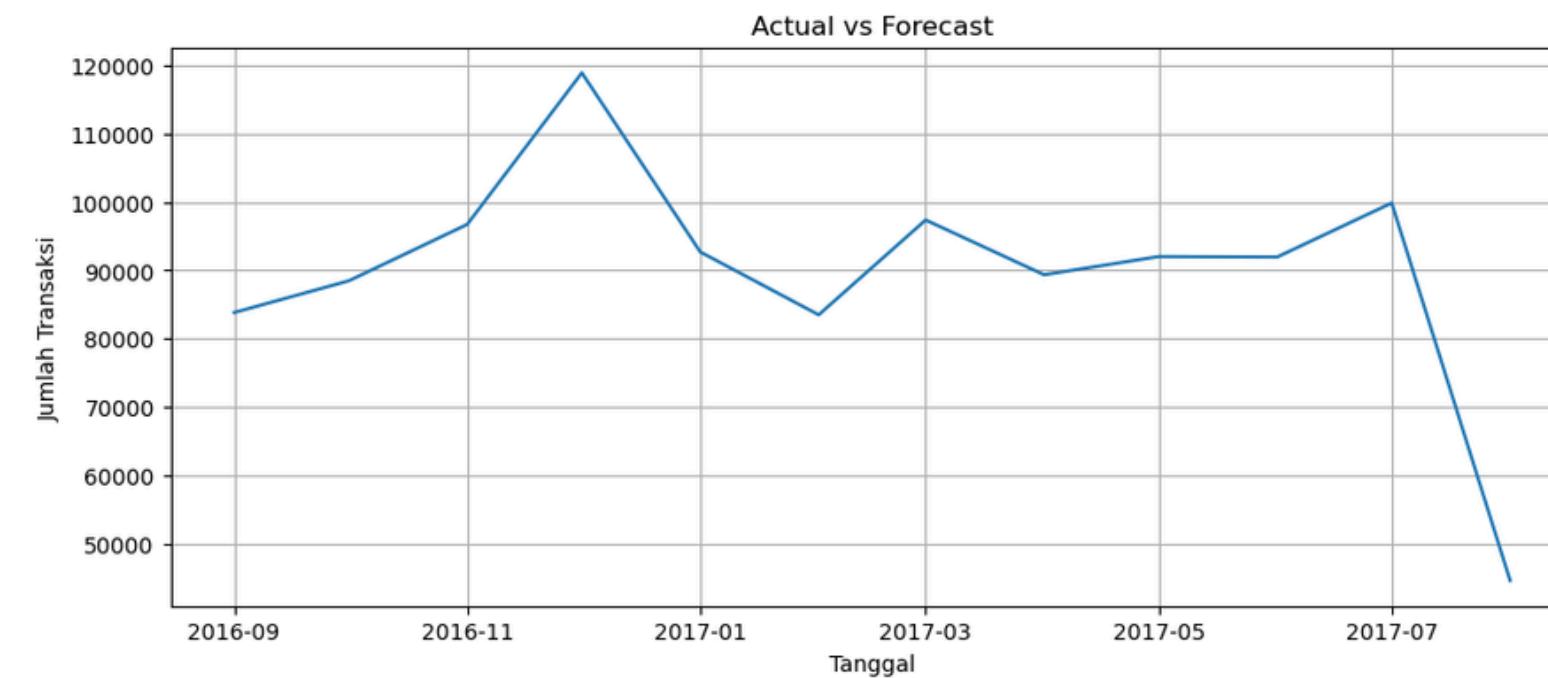
coef      std err      z      P>|z|      [ 0.025      0.975]
-----
intercept    9.44e+04    1564.096    60.355      0.000    9.13e+04    9.75e+04
ma.S.L7        0.3660      0.238     1.538      0.124     -0.101      0.832
sigma2       5.099e+07     0.045  1.12e+09      0.000     5.1e+07     5.1e+07
Ljung-Box (L1) (Q):      1.99
Prob(Q):          0.16
Heteroskedasticity (H):  0.96
Prob(H) (two-sided):  0.94
Jarque-Bera (JB):      2.00
Prob(JB):          0.37
Skew:              0.52
Kurtosis:          3.07
=====
```



Results

ARIMA MODEL (dengan Data Bulanan)

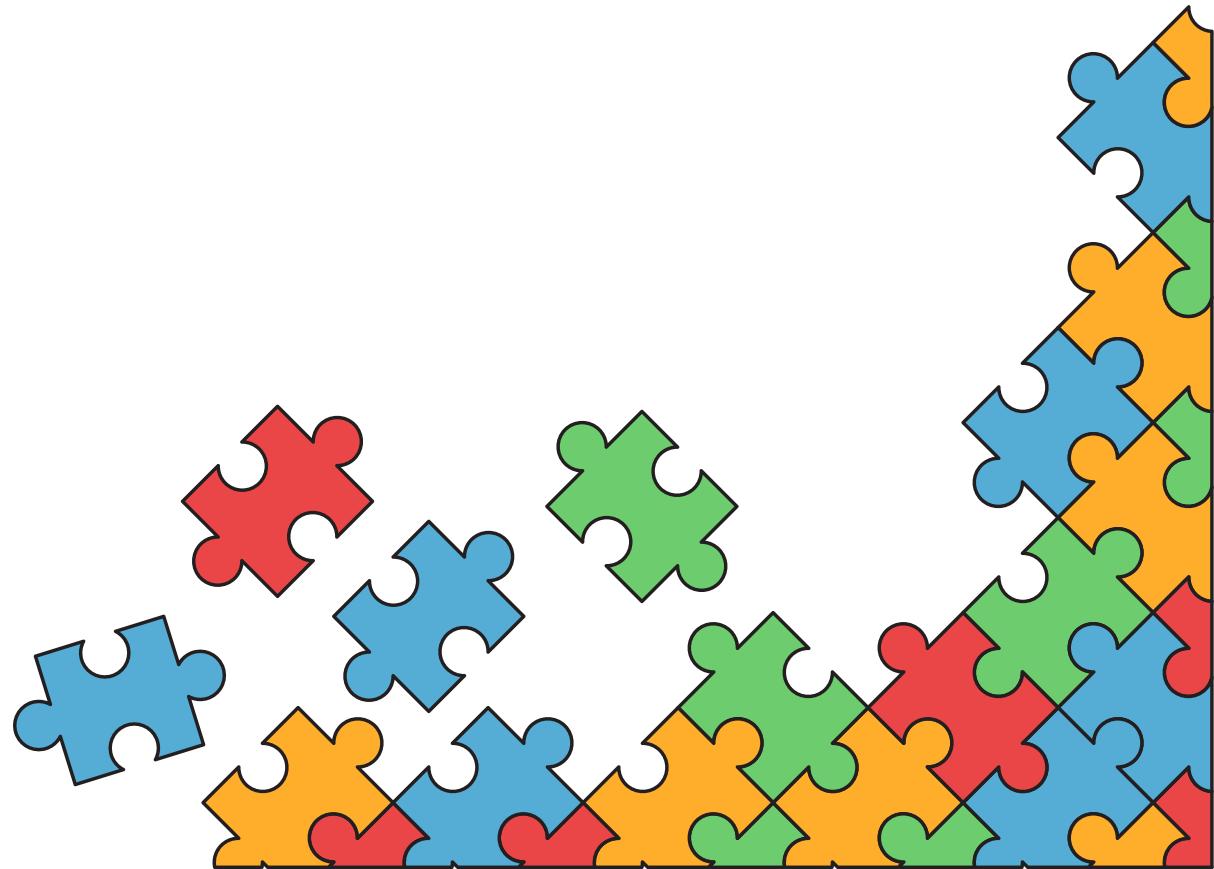
MODEL ARIMA - Comparison



Model Development

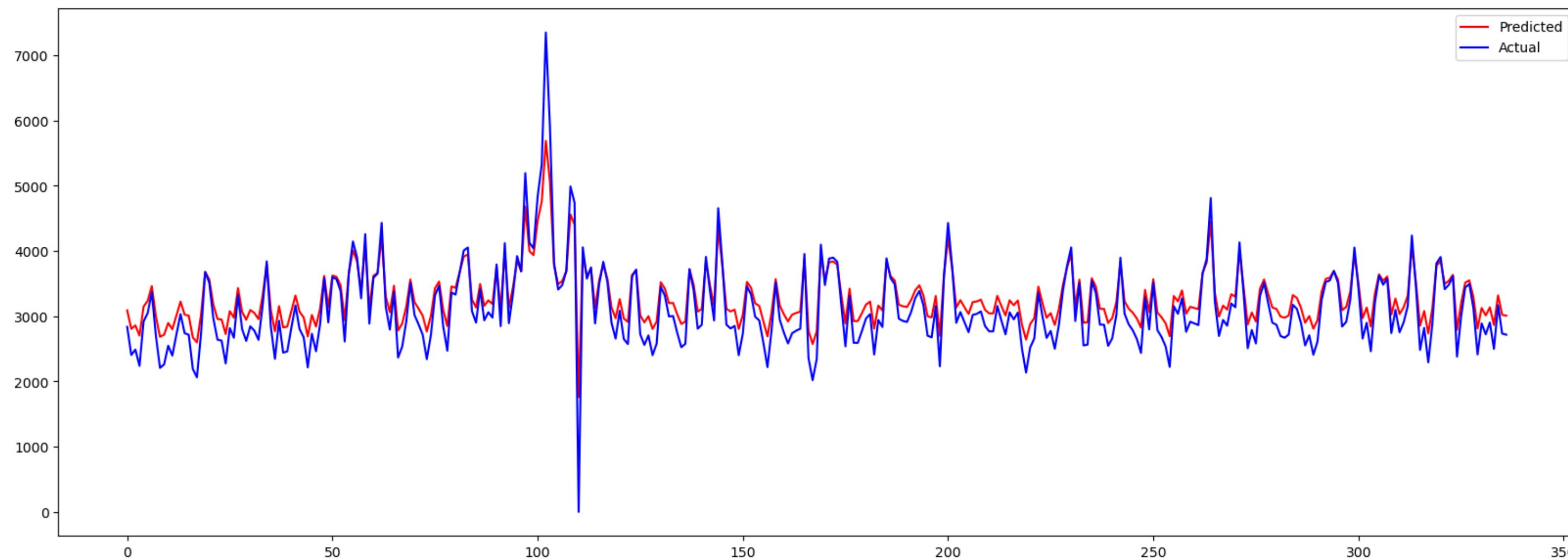
LSTM Model (dengan Data Bulanan & Harian)

- Preprocessing Data
- Reshape Input untuk LSTM
- Bangun Arsitektur Model LSTM
- Training Model
- Forecasting
- Evaluasi Model



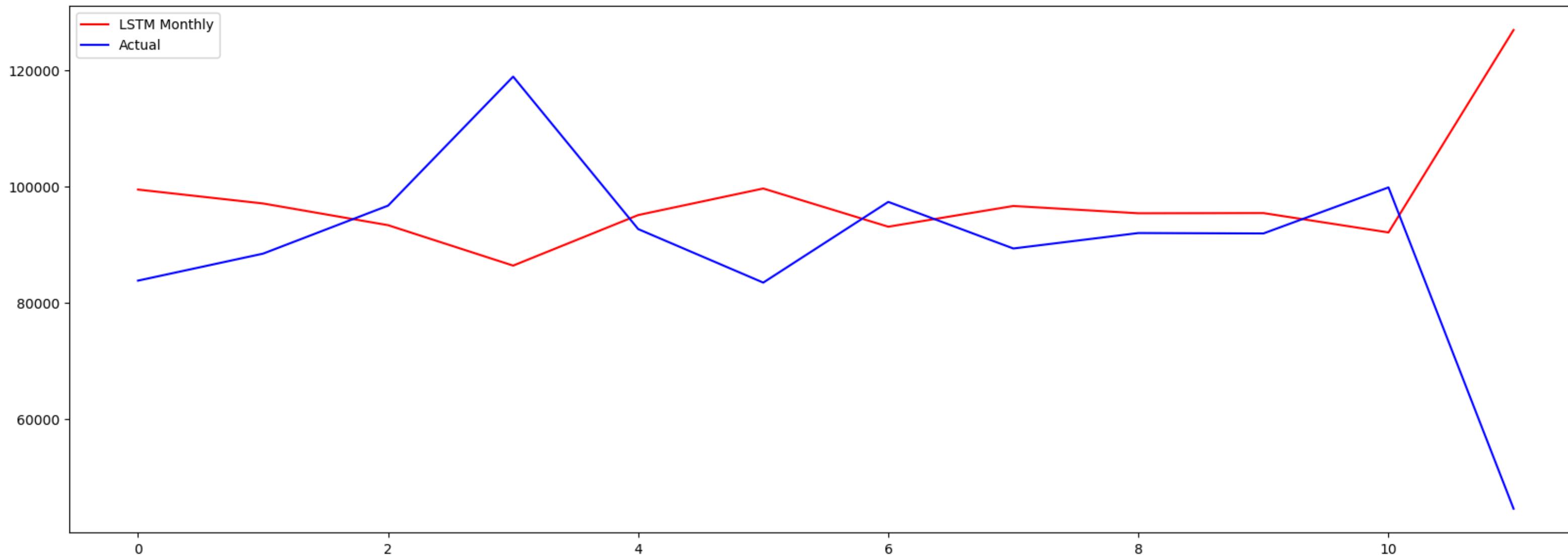
Model Development

LSTM Model (Daily Data)



Model Development

LSTM Model (Monthly Data)



Results

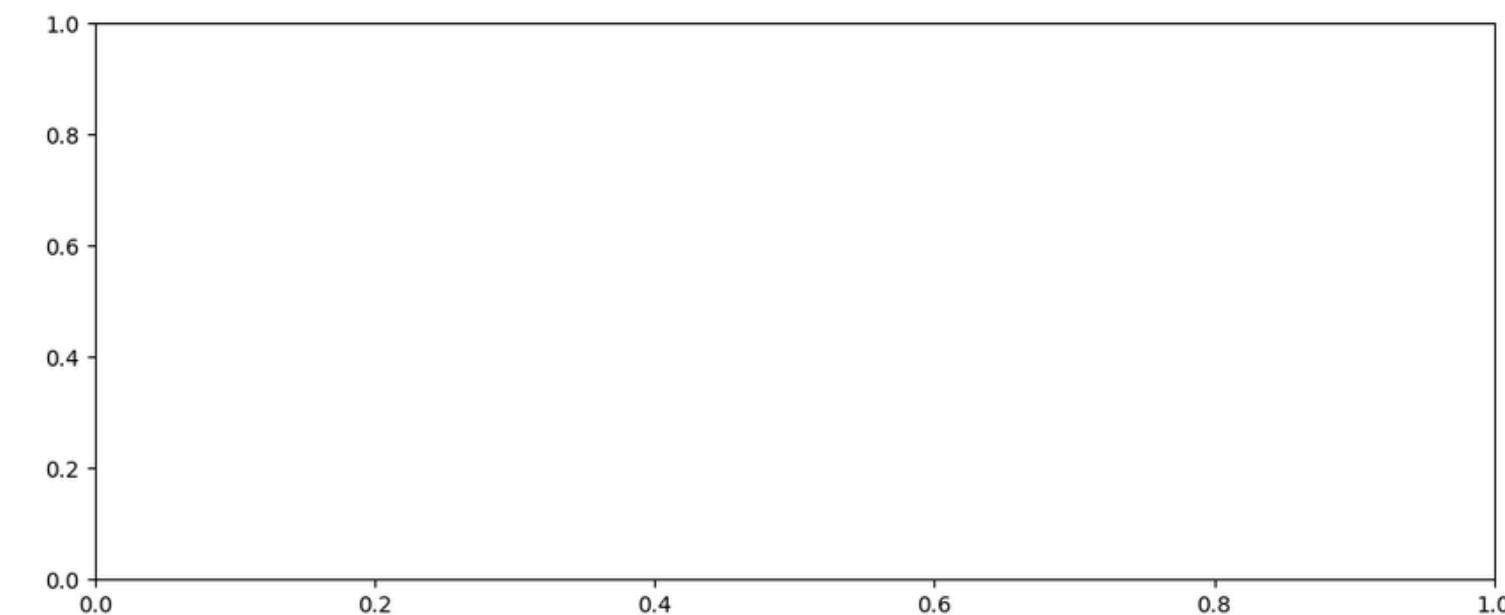
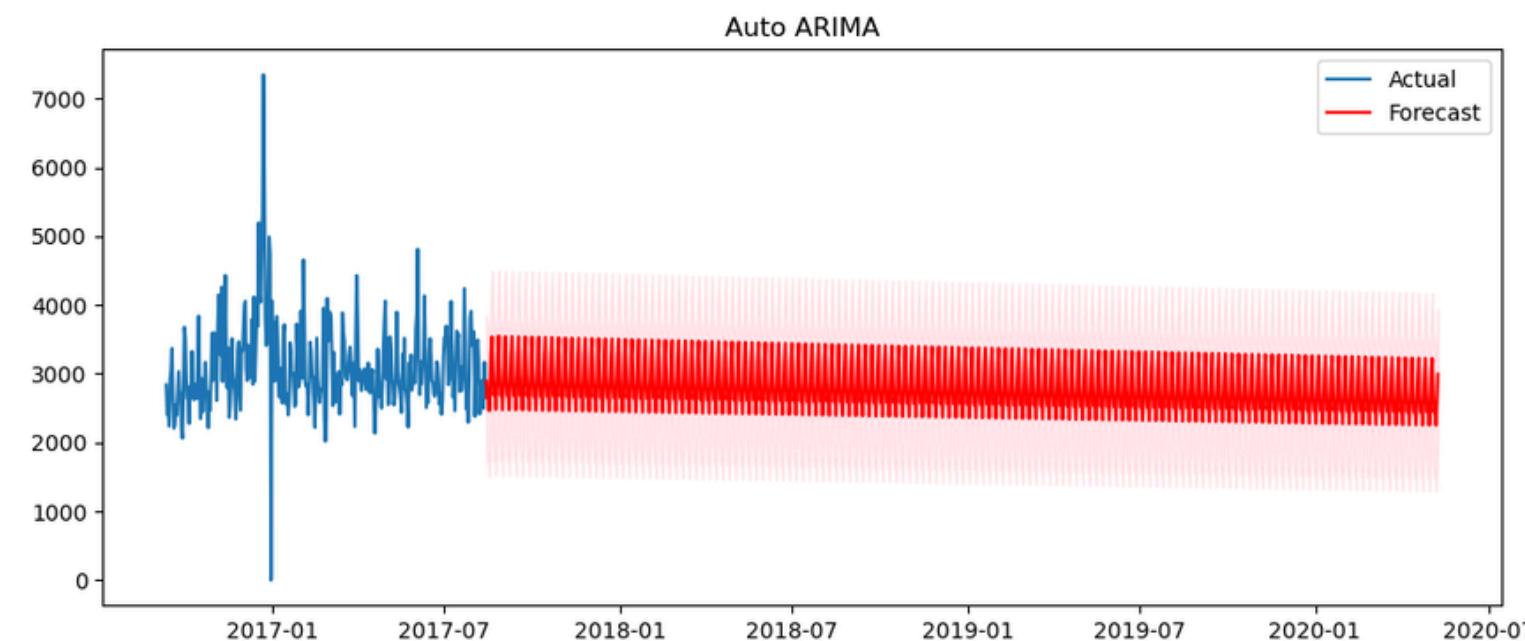
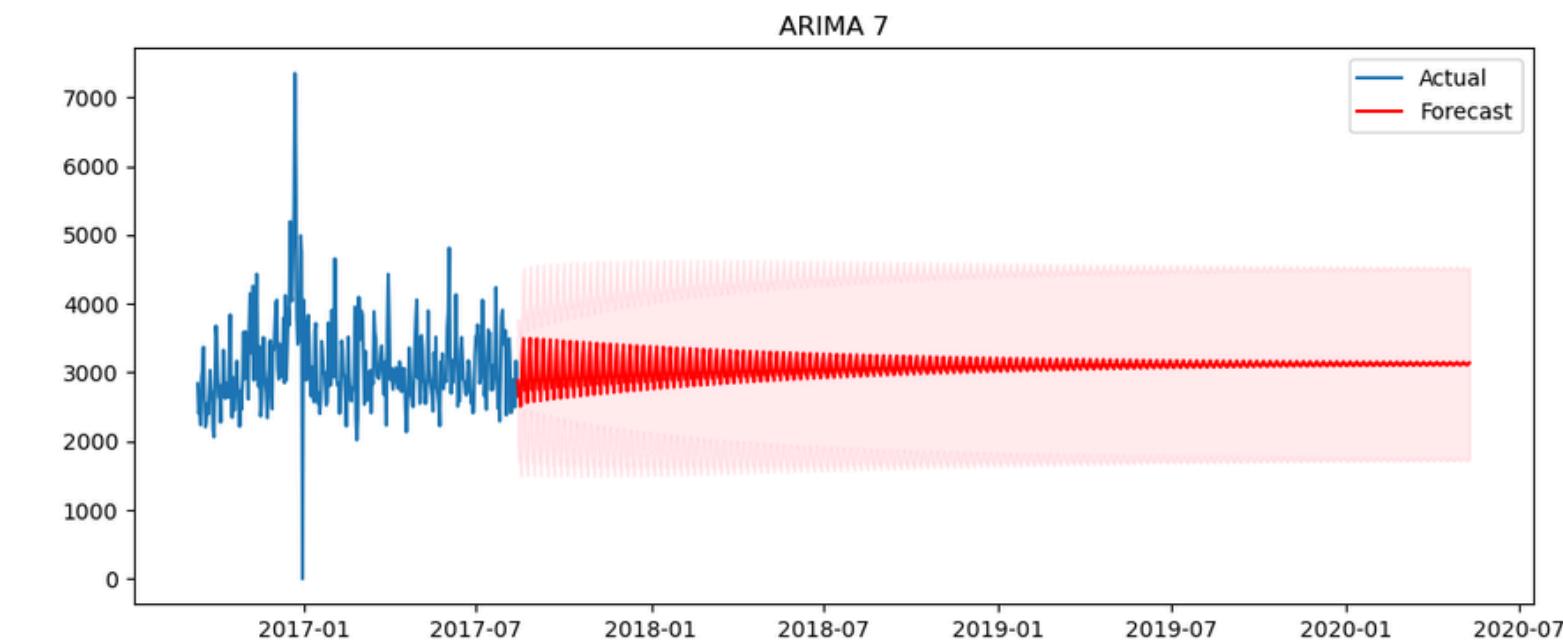
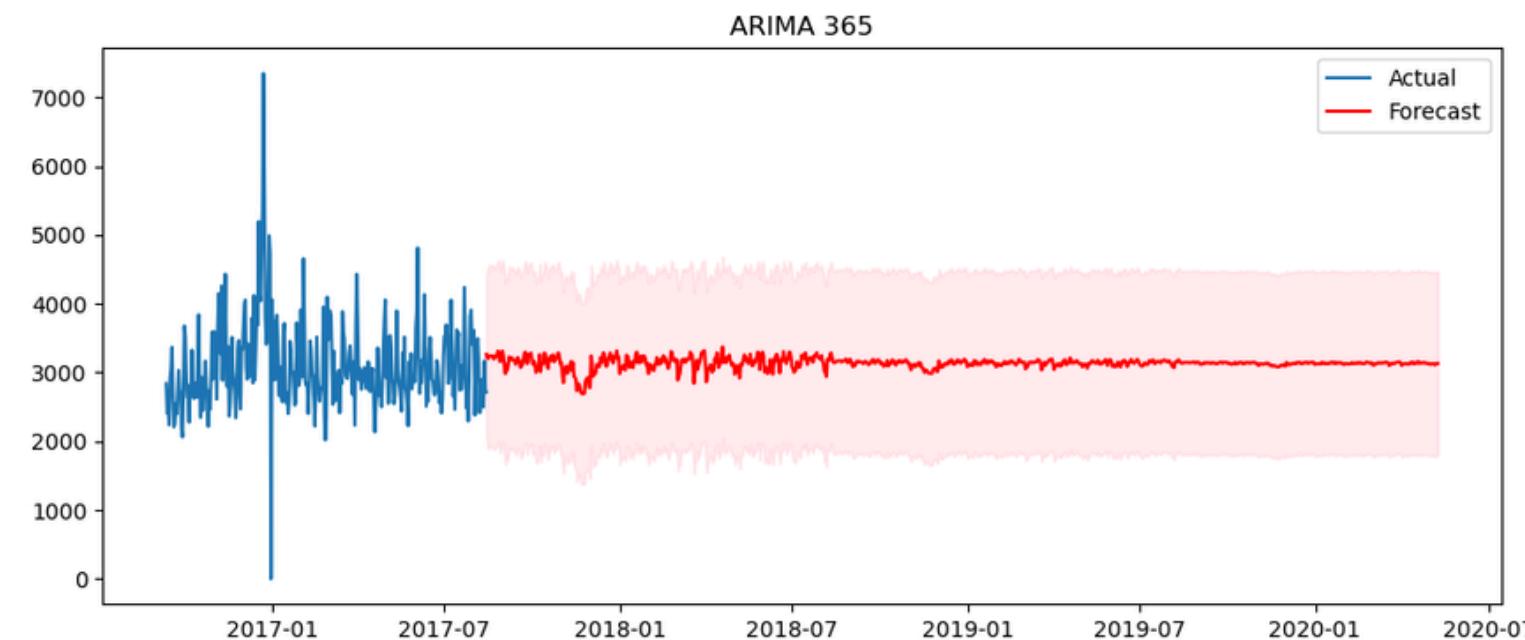
ARIMA & LSTM Model Comparison

Comparison MAE , MAPE & RSME

Category	MAPE	MAE	RMSE
ARIMA Daily 365	15.990586	509.514655	701.396566
Arima Daily 7	13.018163	434.049301	635.692635
Auto Arima Daily	18.094961	589.298621	774.275163
LSTM Daily	7.800415	226.391062	285.350773
ARIMA Monthly 12	19.119256	12554.115774	18900.419547
Arima Monthly 7	17.378982	11444.167712	18345.308551
Auto Arima Monthly	14.976582	9694.803512	16600.808415
LSTM Monthly	24.499753	15621.962891	26776.410614

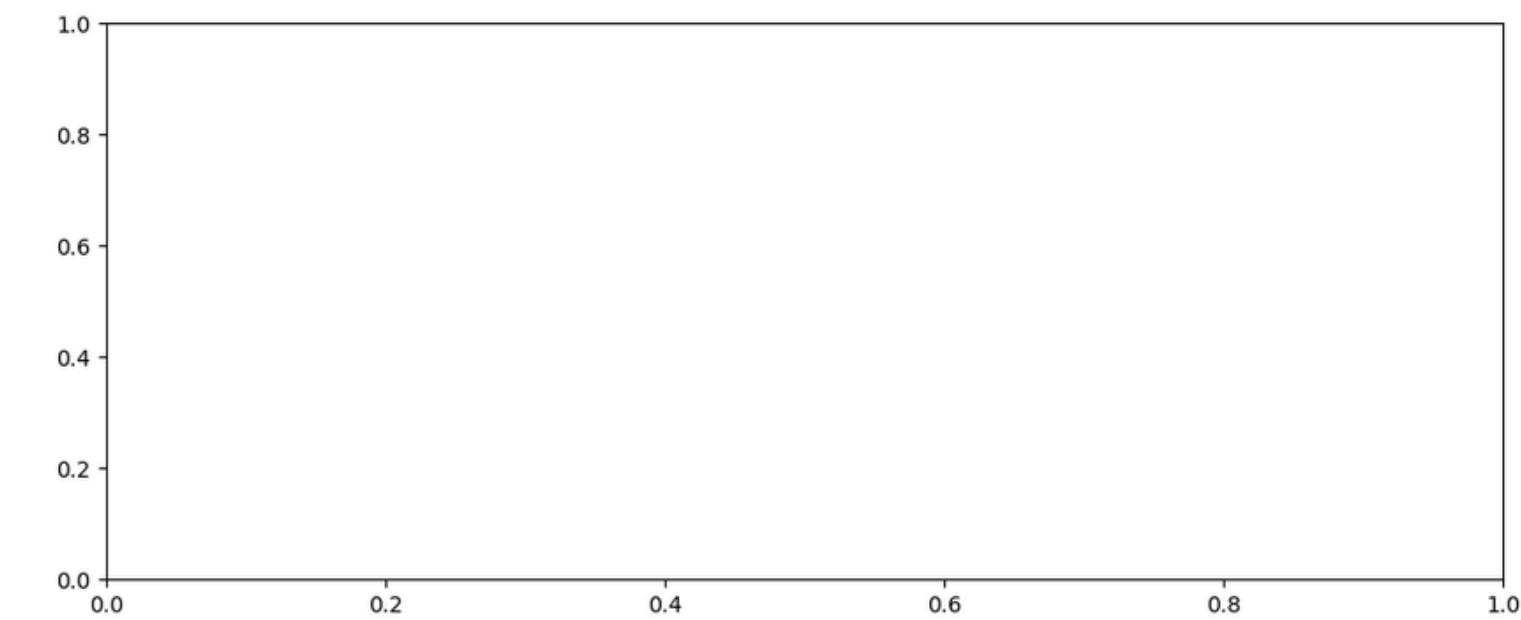
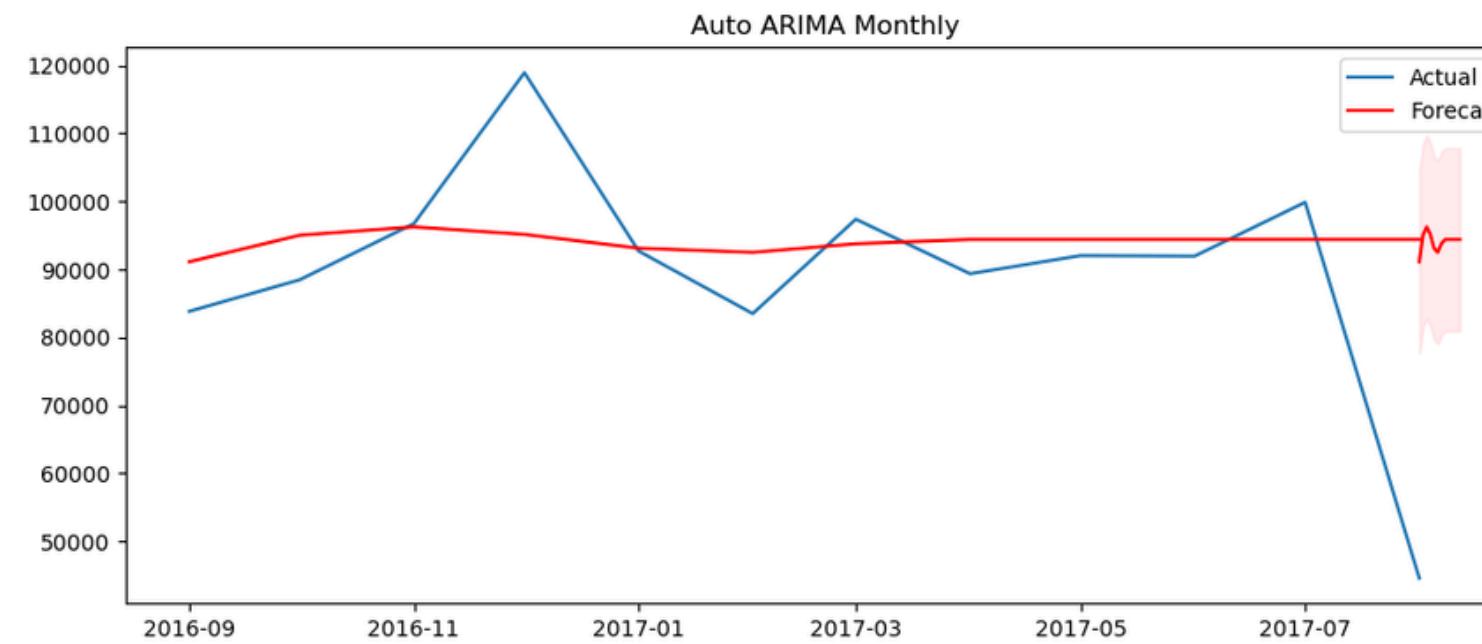
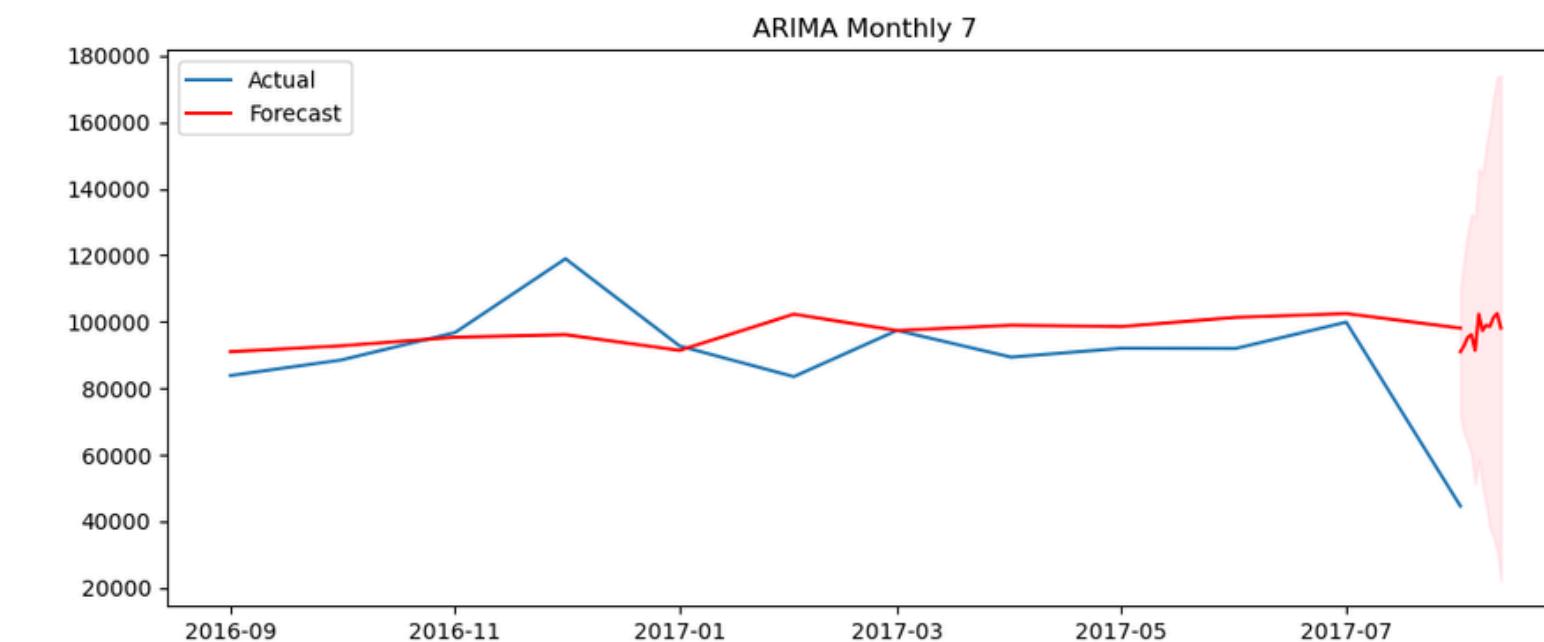
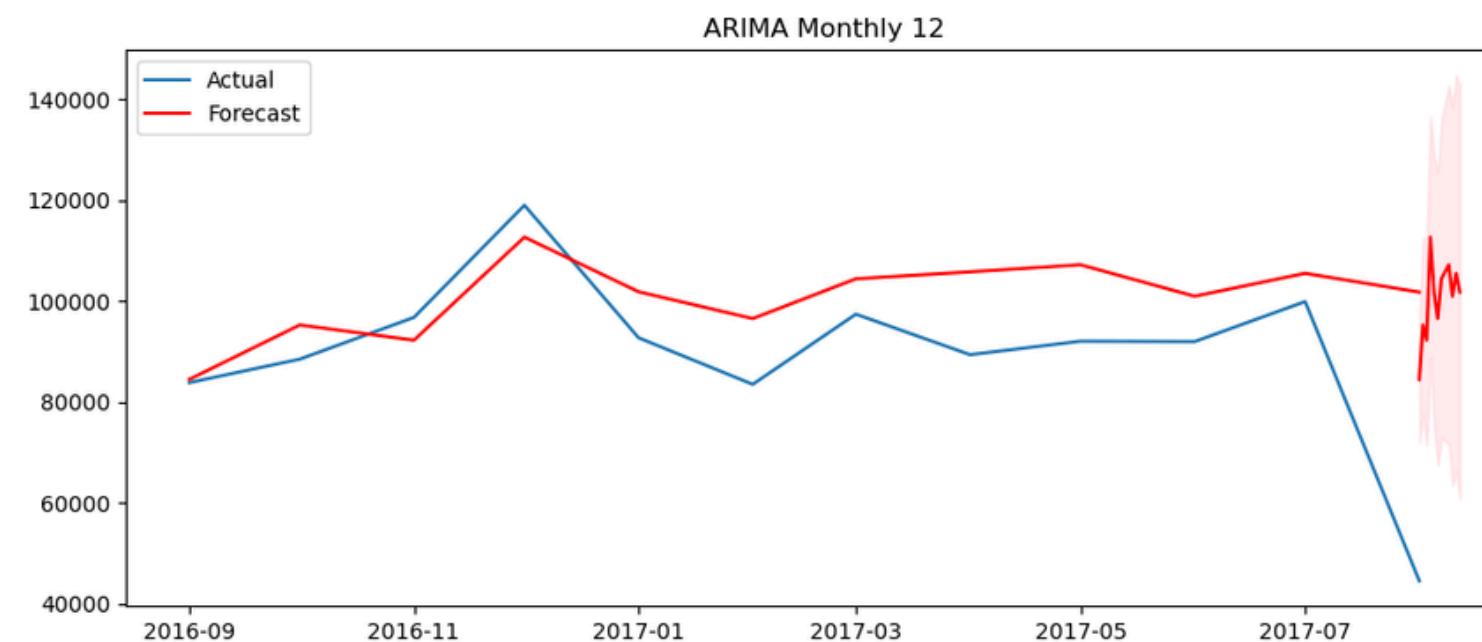
Model Development

ARIMA Model (Daily Data Forecasting Opsional)



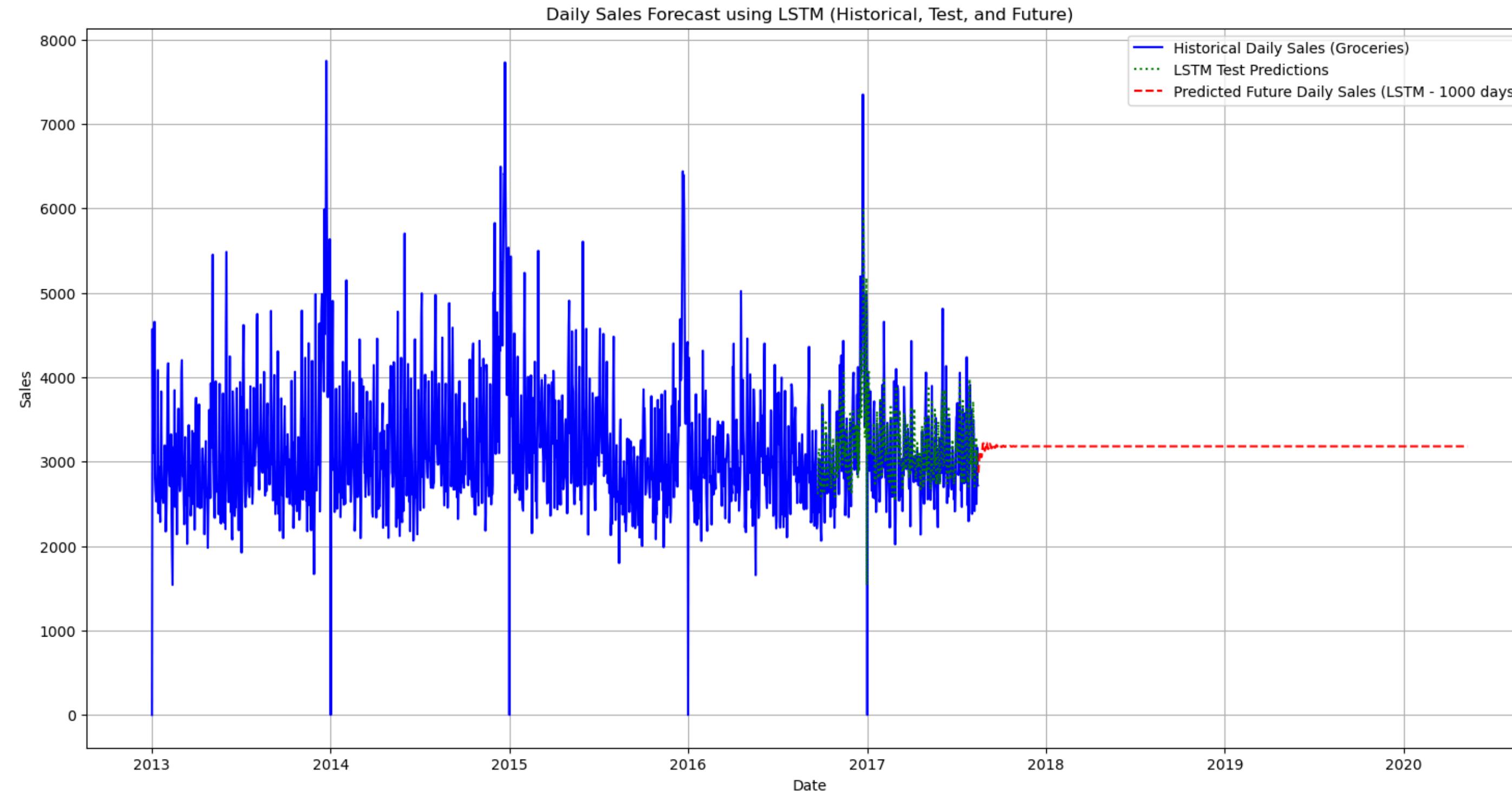
Model Development

ARIMA Model (Monthly Data Forecasting Opsional)



Model Development

LSTM Daily (Forecasting The Best Model)



Future Improvement

1. Model dapat dikembangkan untuk menganalisa kategori lainnya
2. Hasil model dapat dikembangkan untuk insight bisnis seperti analisa stock mingguan, prediksi lonjakan penjualan, prediksi penurunan
3. Model dapat diintegrasikan ke sistem untuk notifikasi.



Conclusion

Model ARIMA lebih cocok untuk data time series jangka pendek

sedangkan untuk Model LSTM lebih cocok untuk data time series yang kompleks.

