

COFFEE SHOP SALES

Statistical Data Analysis



Content

01

Our Team

02

Data Overview

03

Data Structure

04

Results

05

Summary



Our Team



Putri Ayu
Desita

5003211051



Nabilah
Azharudin

5003211093

Overview

The dataset contains transactions, stores, products, and the price for each products.

01



Terdapat 214k total transaksi yang telah terjadi.

Terdapat **tiga** *store location* yang sudah ada, yaitu **Lower Manhattan**, **Hell's Kitchen**, dan **Astoria**.



02

03



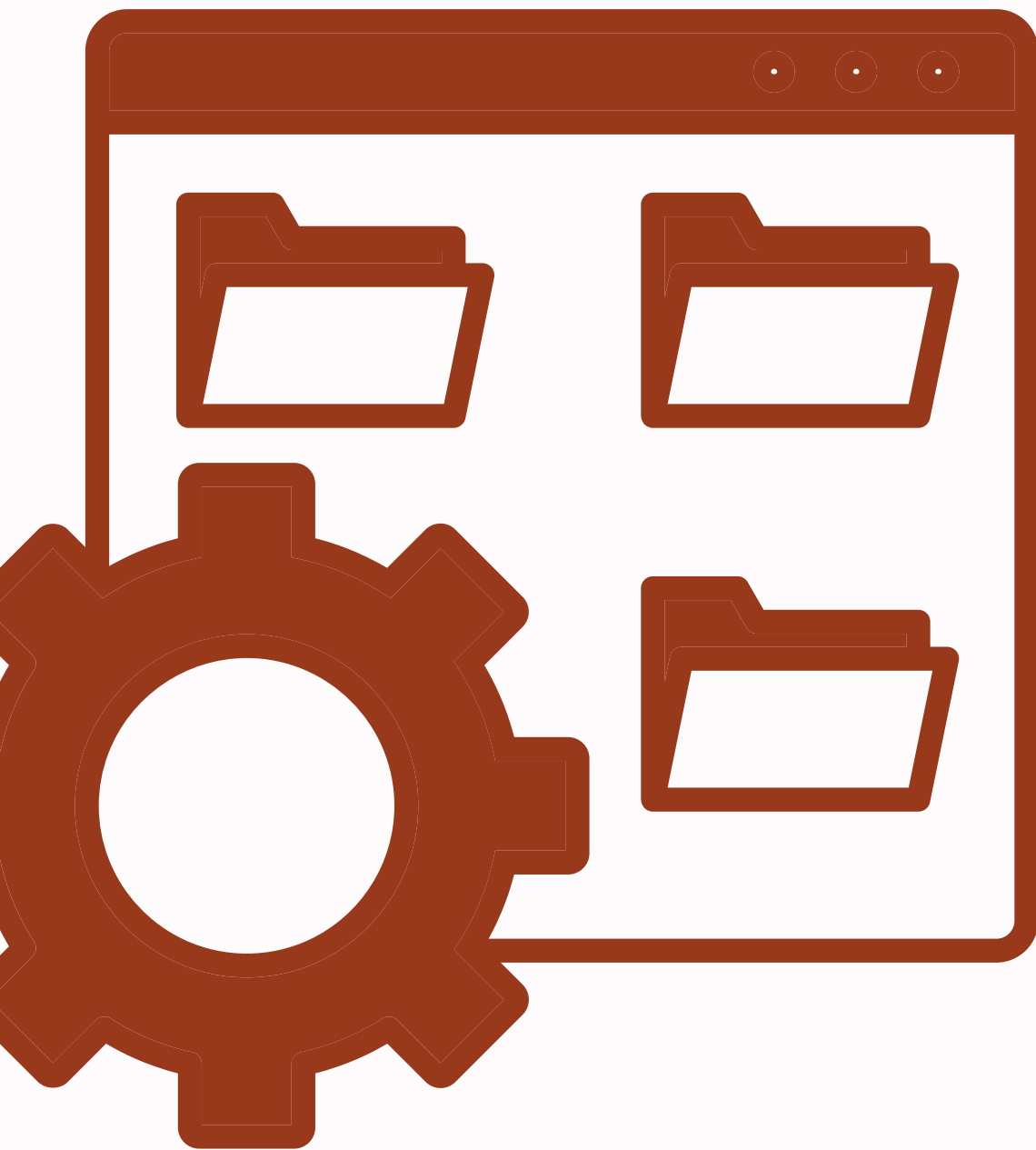
Setiap *product details* memiliki harga yang berbeda-beda.

Terdapat **80** *product details* yang tersedia di seluruh toko.



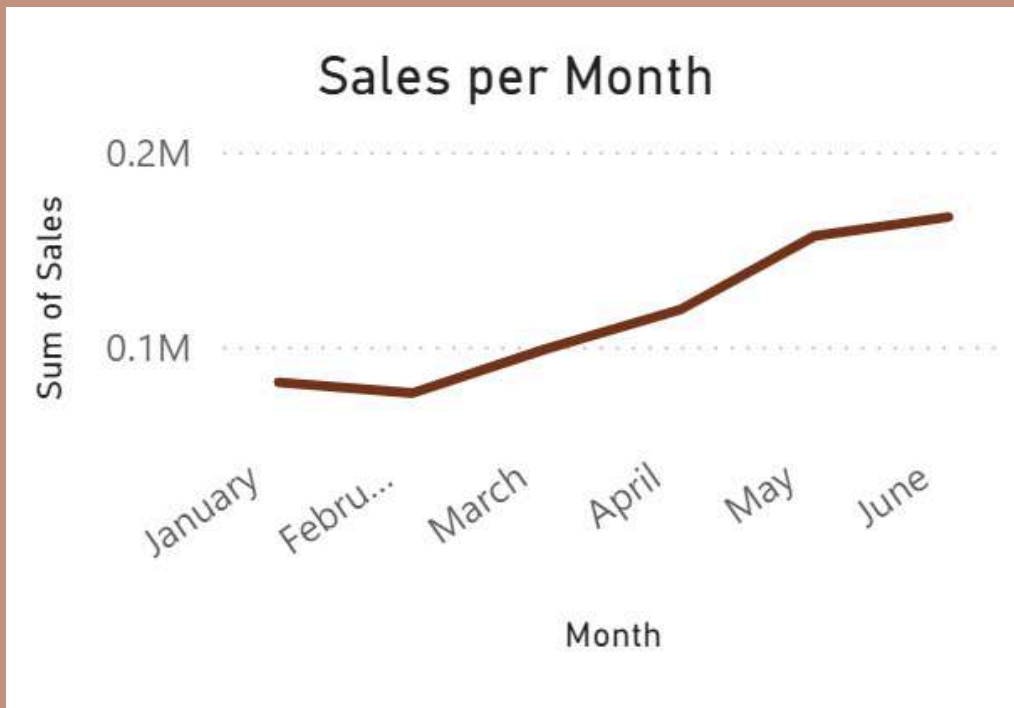
04

Data Structure



Variabel	Struktur
Transaction Date	Ordinal
Transaction Date	Ordinal
Transaction Quantity	Integer
Strore ID	Integer
Store Location	String
Product ID	Integer
Unit Price	Float
Product Category	String
Product Type	String
Product Detail	String

Overall Sales Report



01

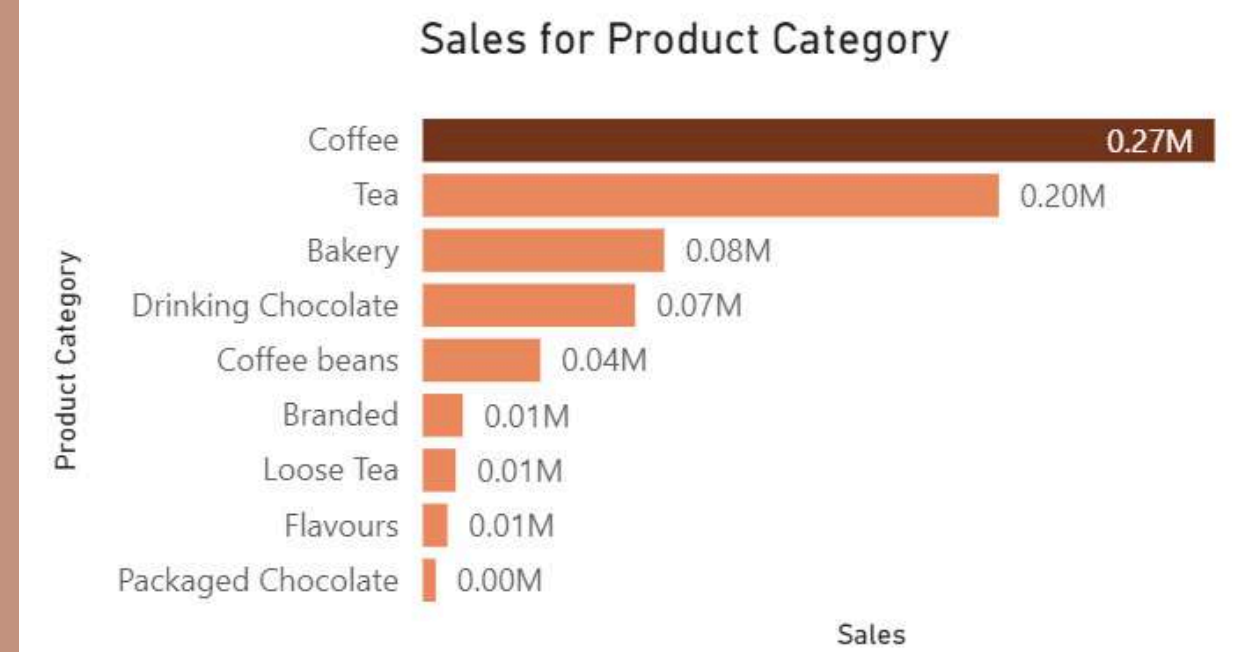
Secara keseluruhan, *sales* mengalami kenaikan di setiap bulannya yang menandakan adanya *potential growth* yang menjanjikan.



02

Setiap toko memiliki total penjualan yang hampir sama untuk penjualan *overall products*.

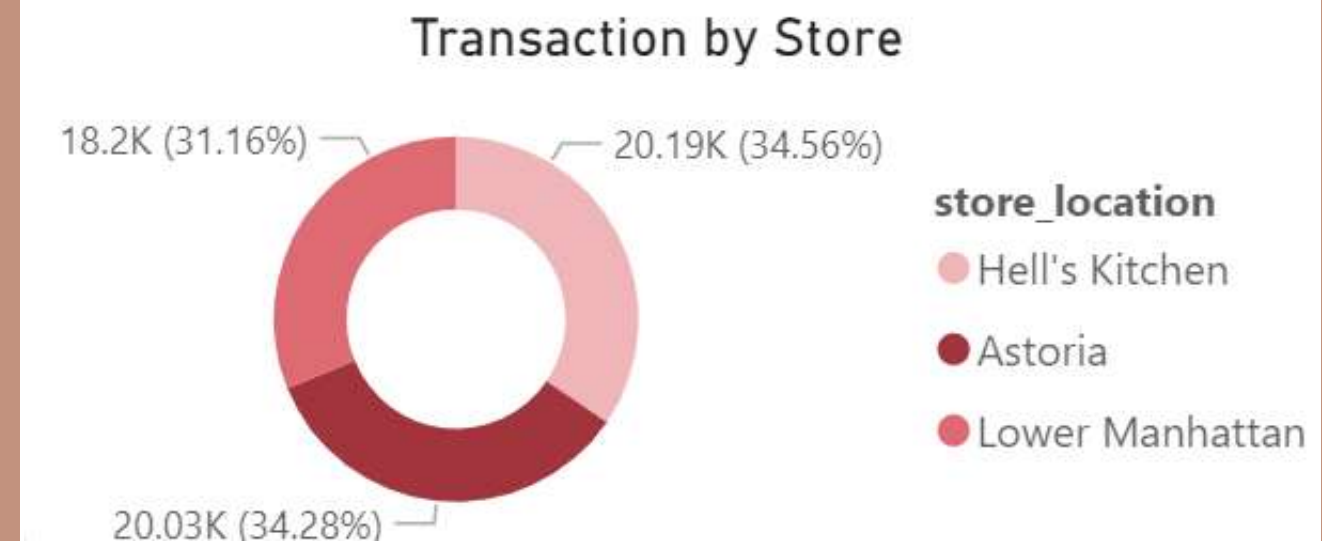
Overall Sales Report



03

Dari 9 *product categories*, **coffee** memiliki penjualan yang jauh lebih besar dibandingkan dengan kategori produk lainnya.

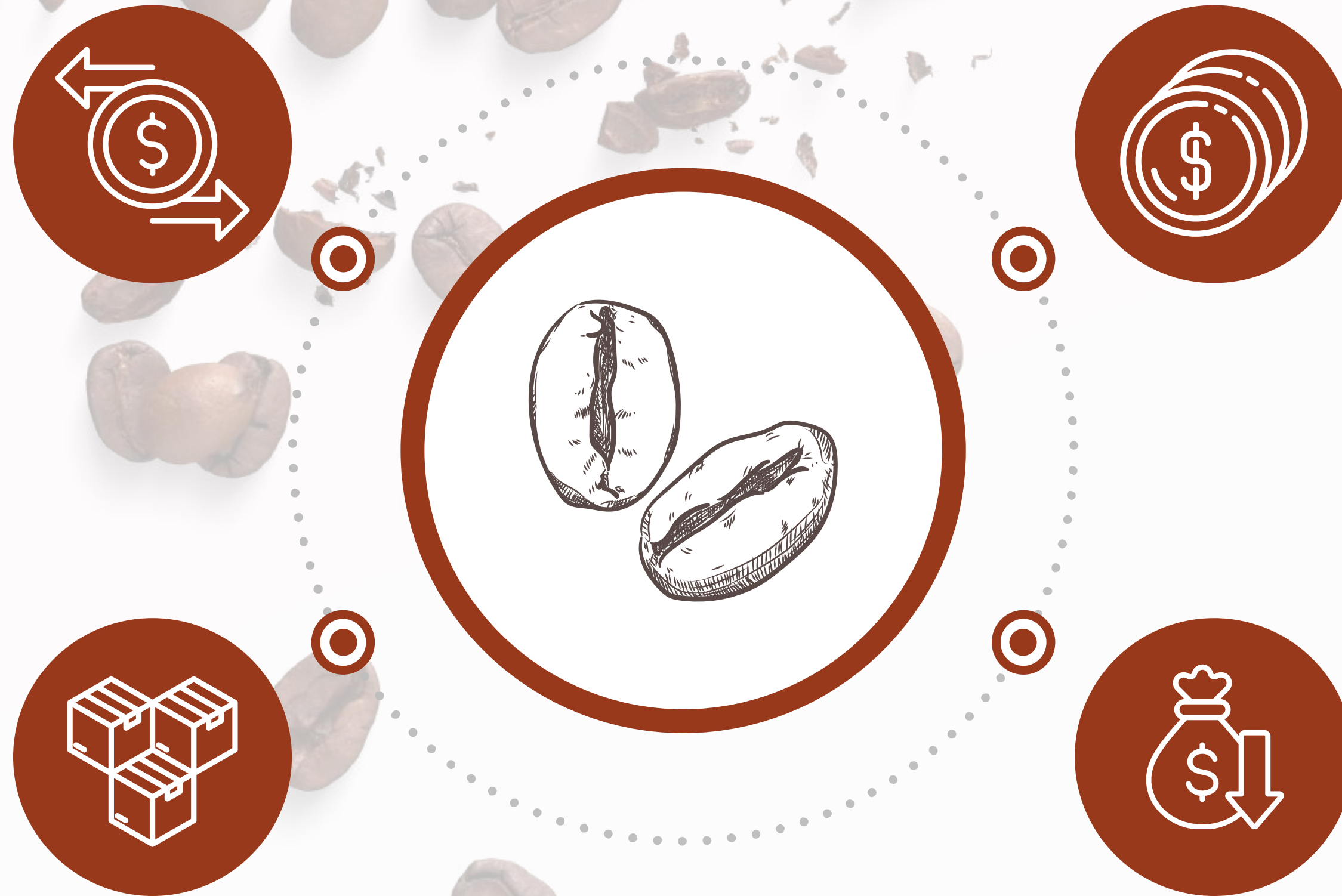
Dikarenakan **coffee** memiliki *sales* tertinggi, maka analisis akan dibataskan dalam data **coffee** saja.



04

Tidak ada perbedaan jumlah transaksi antar-toko yang besar yang menandakan bahwa setiap toko memiliki *growth potential* yang sama.

Coffee Sales Report



Total Transaction

Total transaksi untuk **coffee** berada di angka **89.000**

Highest Price

Product termahal adalah **Cappucino Large** dengan harga \$4.25

Lowest Price

Produk dengan harga terendah adalah **Our Old Time Diner Blend Small** dengan harga \$2

Total Product Details

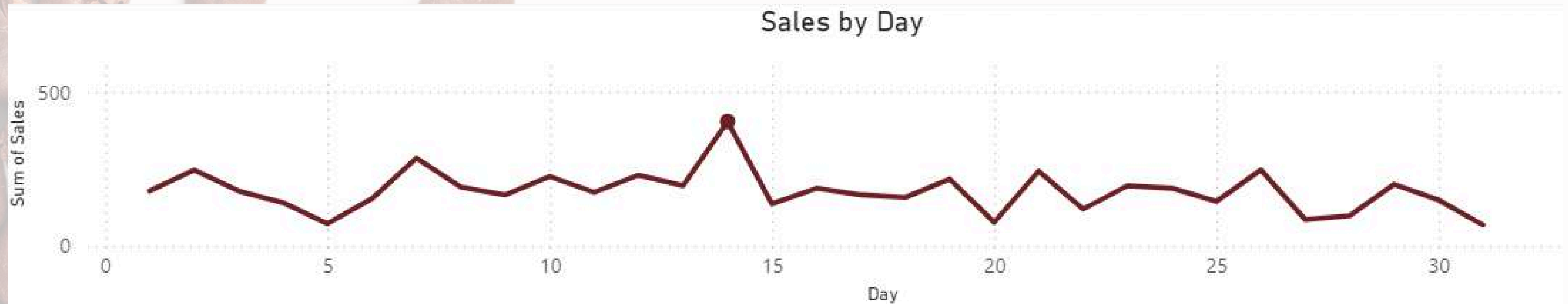
Kategori **coffee** memiliki 21 *product details* yang sudah dijual

Coffee Sales per Day



Di **Astoria**, penjualan untuk **Cappucino** cenderung lebih stabil jika dibandingkan dengan produk lainnya. Penjualan ini memiliki rekor tertinggi di hari kedua penjualan dengan jumlah sales sebesar \$318.75

Coffee Sales per Day



Di **Hell's Kitchen**, penjualan untuk **Cappuccino Large** cenderung lebih stabil tinggi jika dibandingkan dengan produk lainnya. Penjualan ini memiliki rekor tertinggi di hari ke-14 penjualan dengan jumlah sales sebesar \$403.75

Coffee Sales per Day

Sales by Day

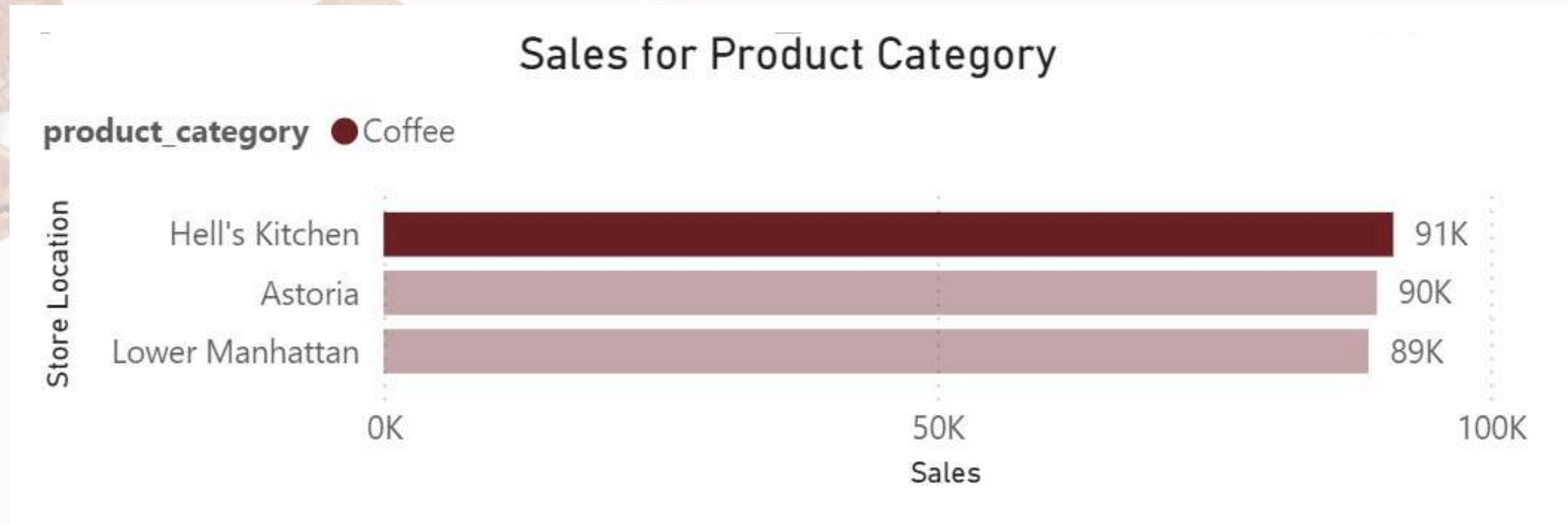


Transaction by Store

Sales for Product Category

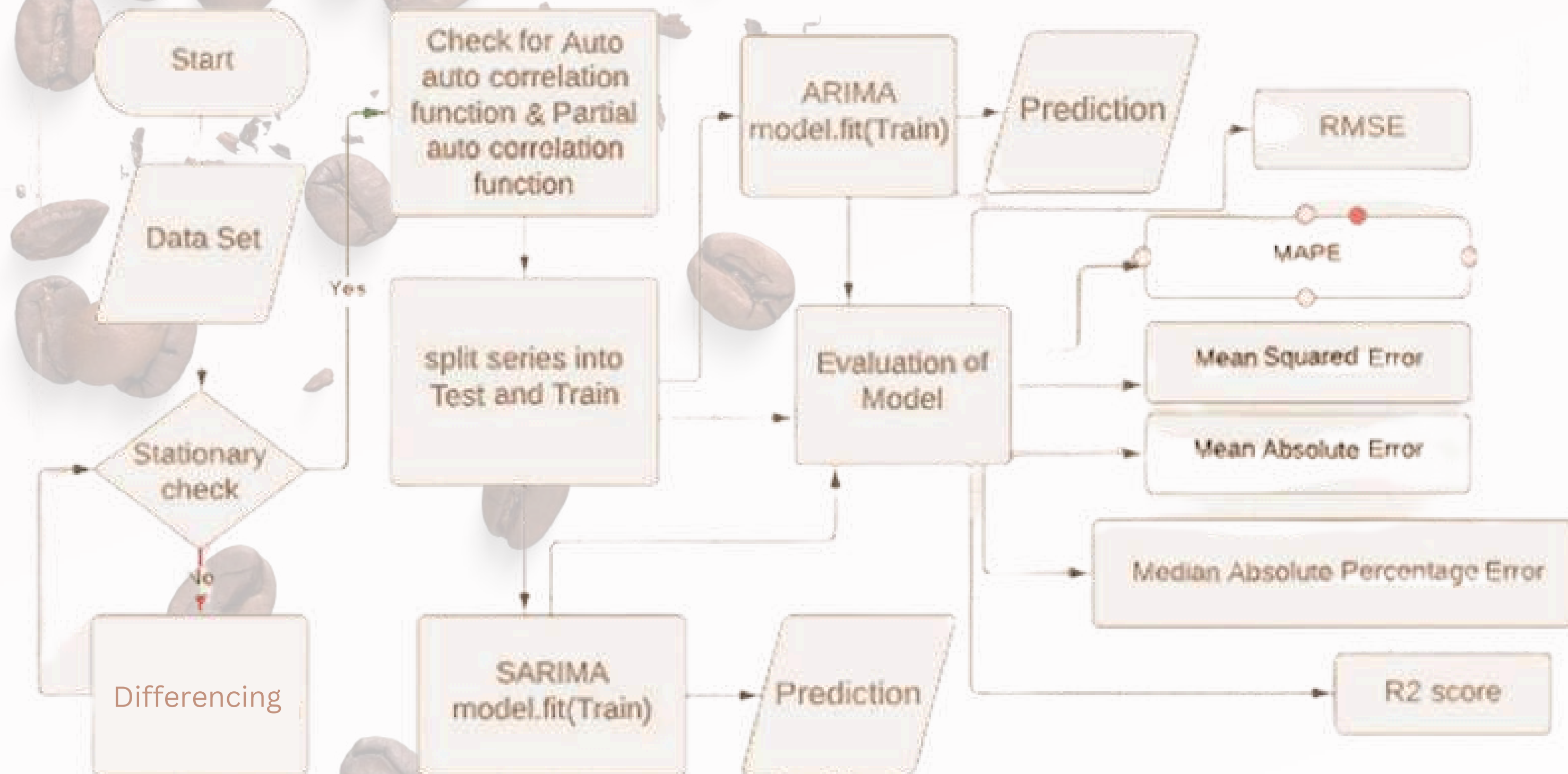
Di **Lower Manhattan**, penjualan untuk **Latte** cenderung lebih stabil tinggi jika dibandingkan dengan produk lainnya. Penjualan ini memiliki rekor tertinggi di hari ke-9 penjualan dengan jumlah sales sebesar \$465

Coffee Sales per Day

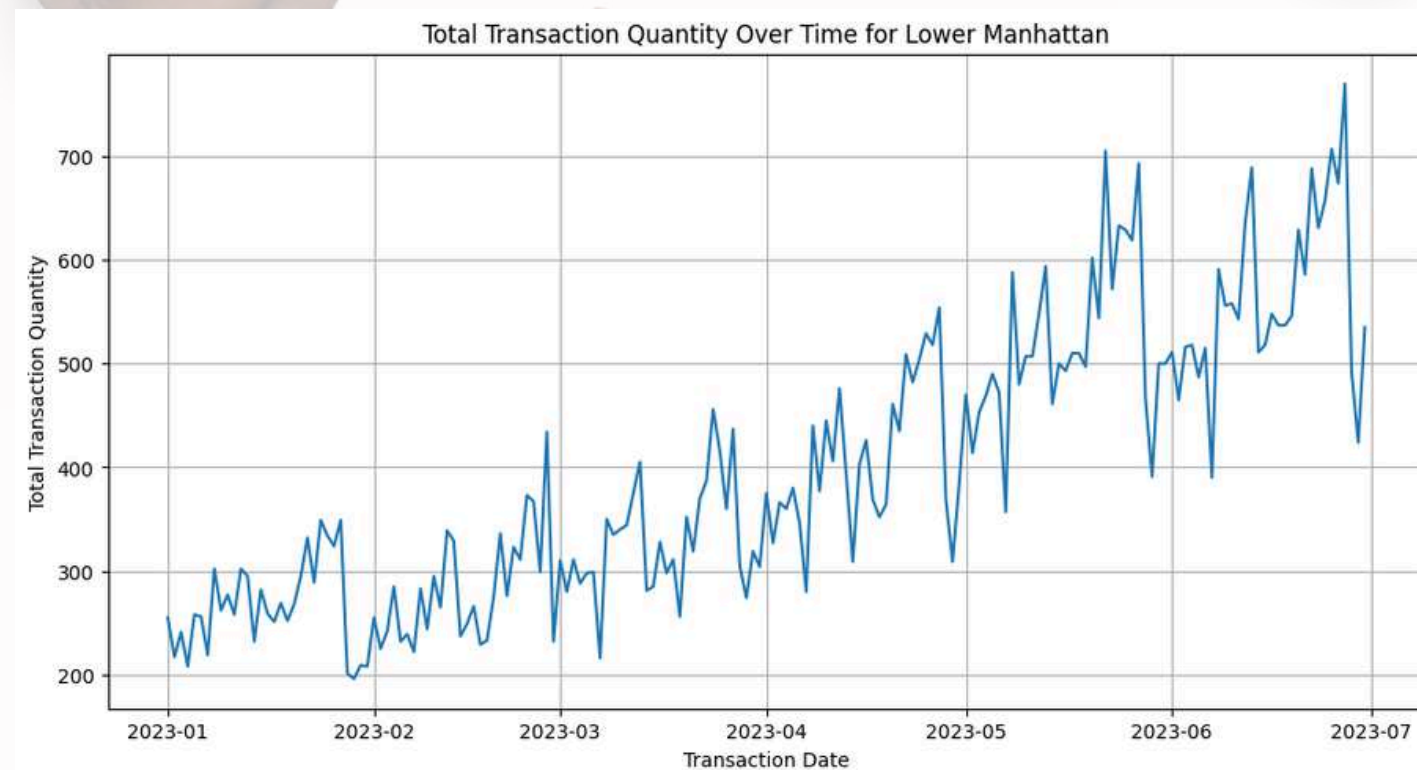
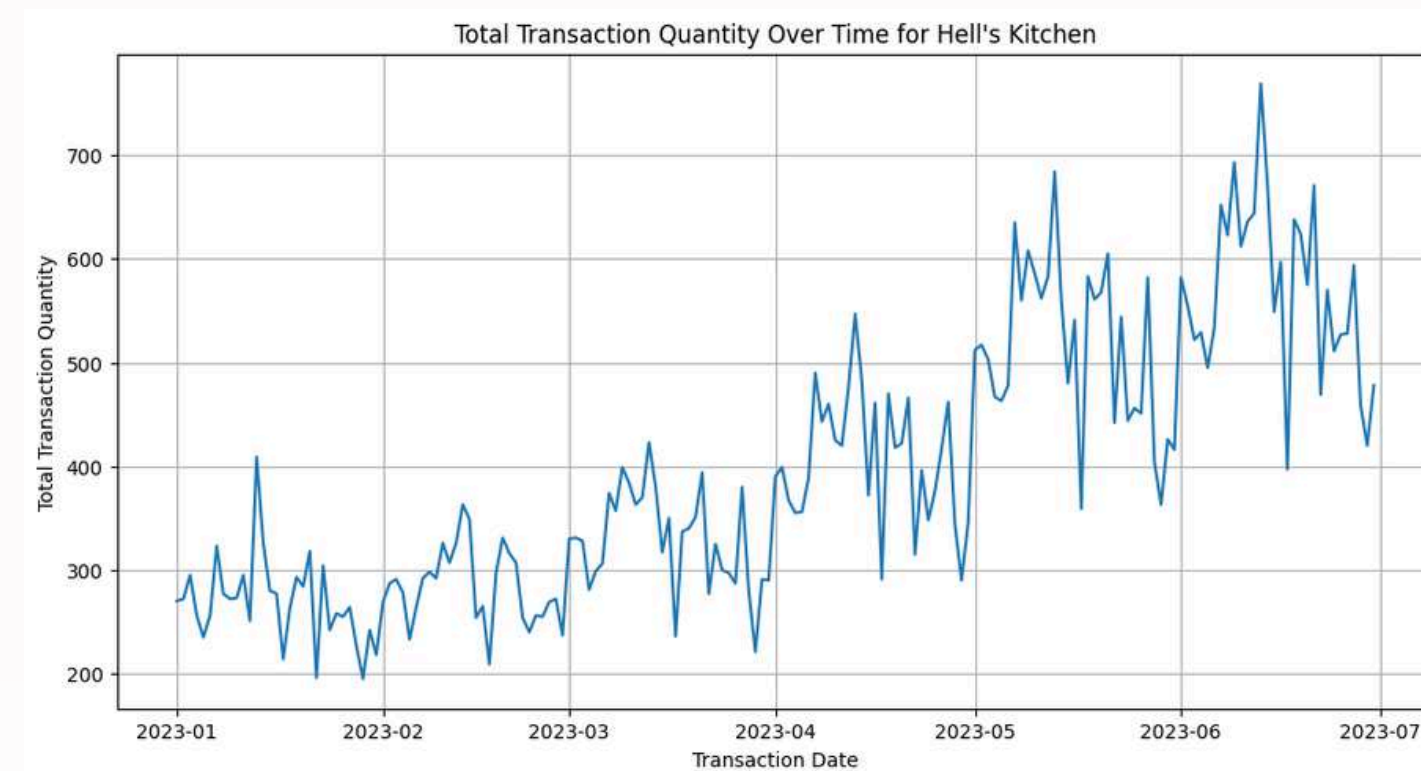
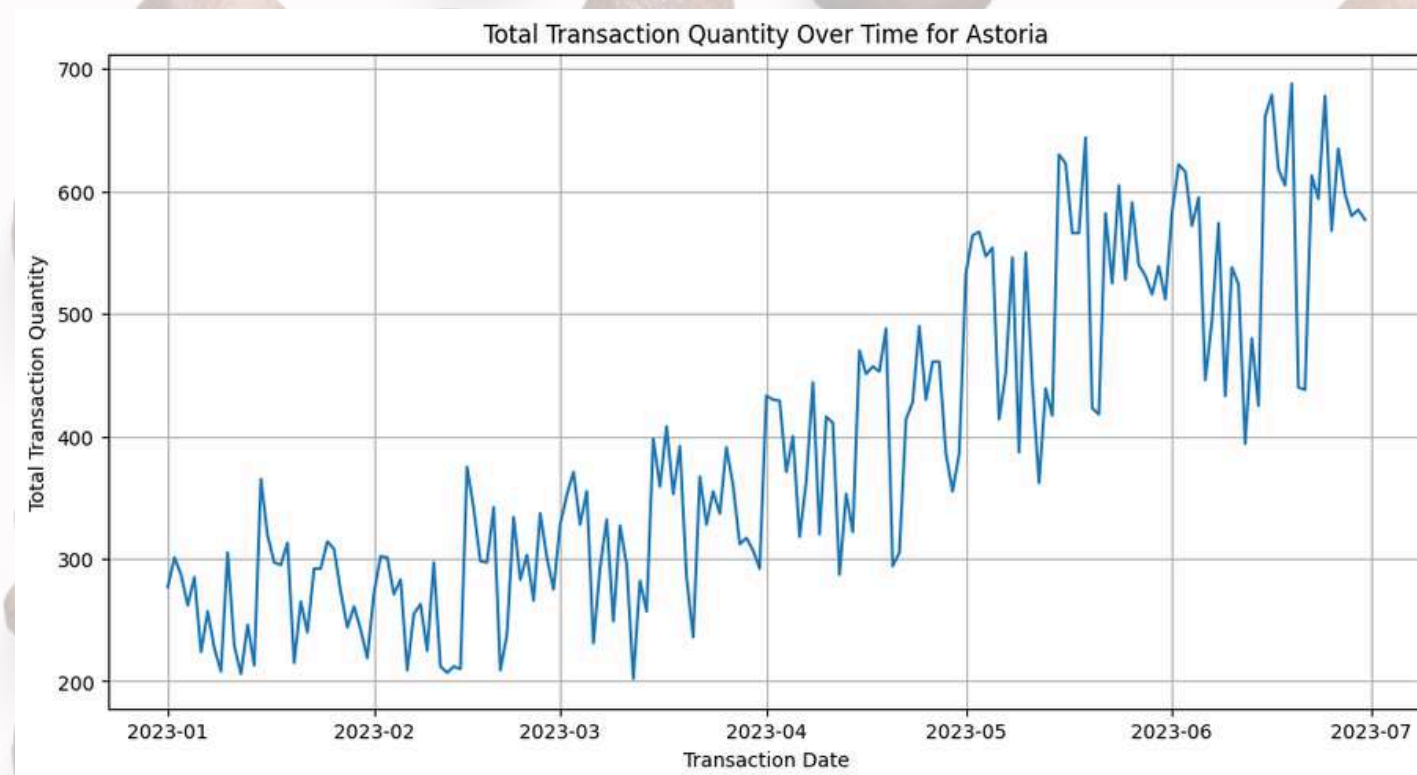


Lokasi dengan penjualan **coffee** tertinggi adalah **Hells's Kitchen** dengan total penjualan sebesar \$91k sehingga lokasi ini dapat menjadi prioritas untuk strategi pemasaran atau peluncuran produk baru yang berkategori kopi karena adanya daya tarik pelanggan yang lebih besar.

Flowchart Forecasting Menggunakan SARIMA



Original Plot



- Di Astoria, Hells Kitchen, dan Manhattan, transaction qty selalu terjadi trend naik dari Januari sampai Juni.
- Selain itu dapat dilihat juga bahwasanya transaction qty membentuk pola seasonal sehingga akan memakai SARIMA sebagai pembuatan model untuk forecasting qty dari produk di hari-hari berikutnya

ADF Test

Sebelum Differencing

Hipotesis Uji

H_0 : data tidak stasioner dalam rata-rata

H_1 : data stationer dalam rata-rata

Tingkat signifikansi: 5%

Revenue	P-Value
Hell's Kitchen	0,98
Astoria	0,59
Lower Manhattan	0,96

- Dari p-value data transaction qty untuk ketiga store belum stasioner secara mean. Sehingga dibutuhkan penanganan dengan melakukan differencing pada data

Setelah Differencing

Hipotesis Uji

H_0 : data tidak stasioner dalam rata-rata

H_1 : data stationer dalam rata-rata

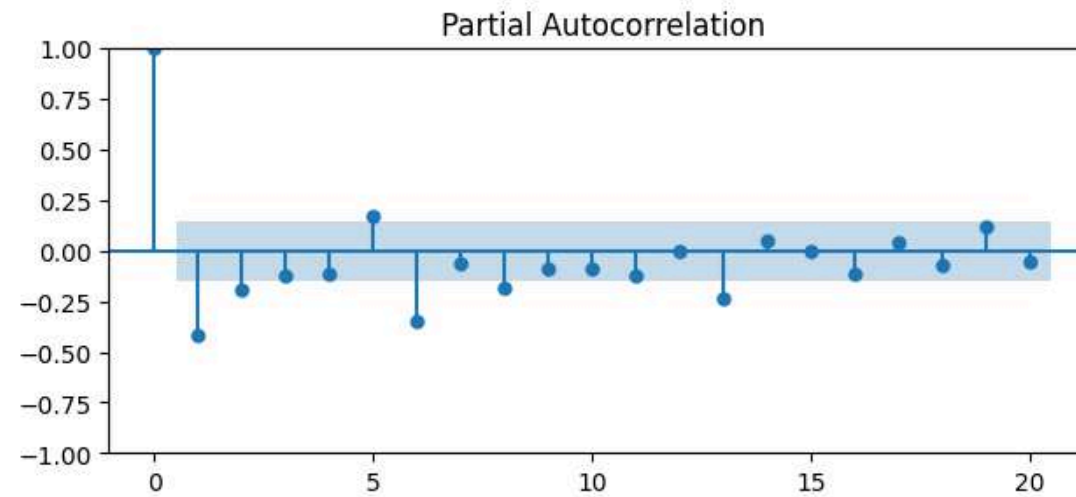
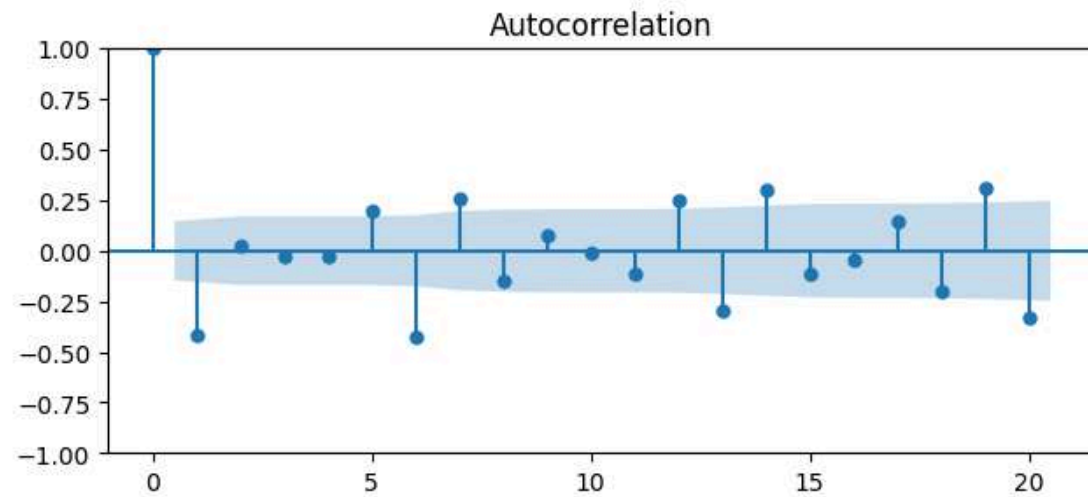
Tingkat signifikansi: 5%

Revenue	P-Value
Hell's Kitchen	0,00
Astoria	$5,7 * 10^{-5}$
Lower Manhattan	0,00

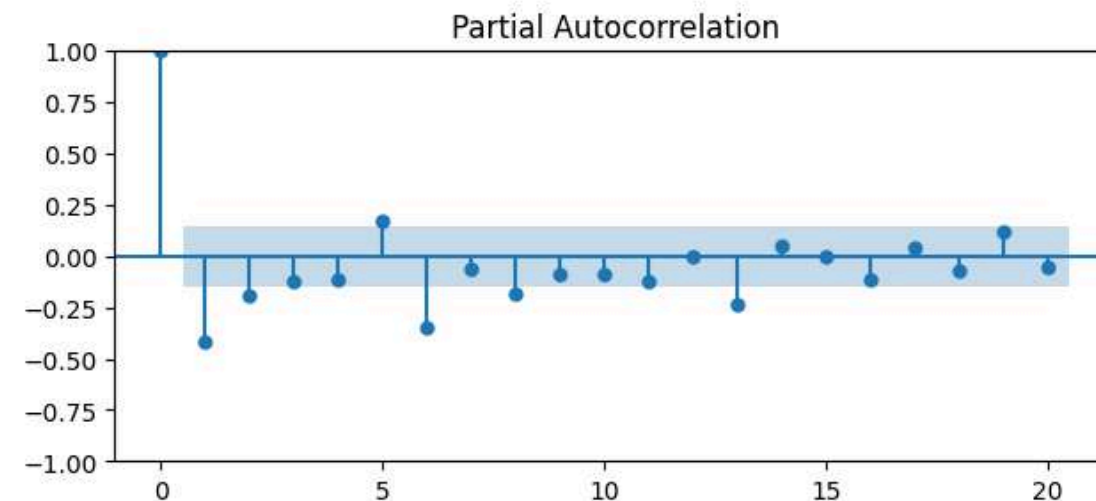
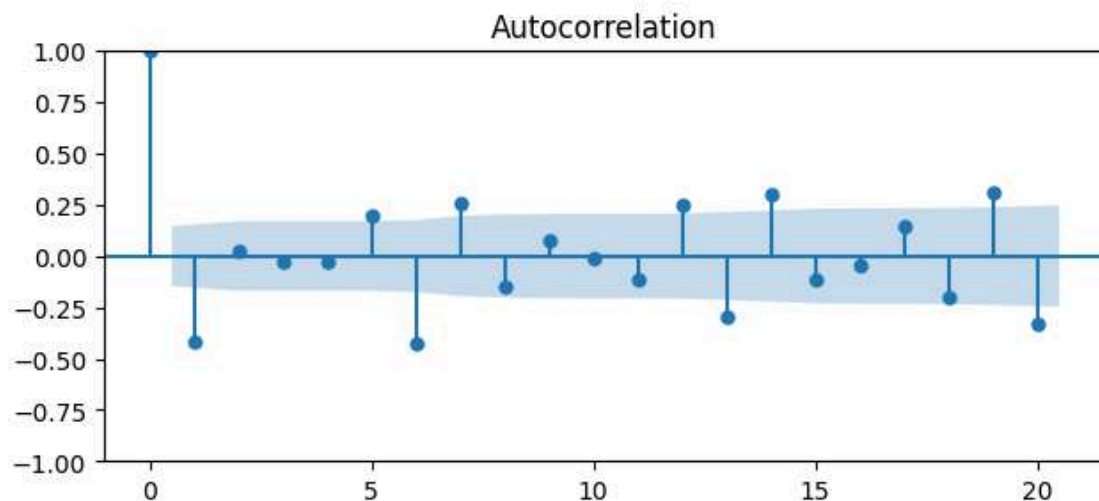
- Dari p-value data transaction qty untuk ketiga store setelah dilakukan differencing, sudah stasioner dalam mean sehingga bisa dilanjutkan ke pembuatan model SARIMA

ACF dan PACF

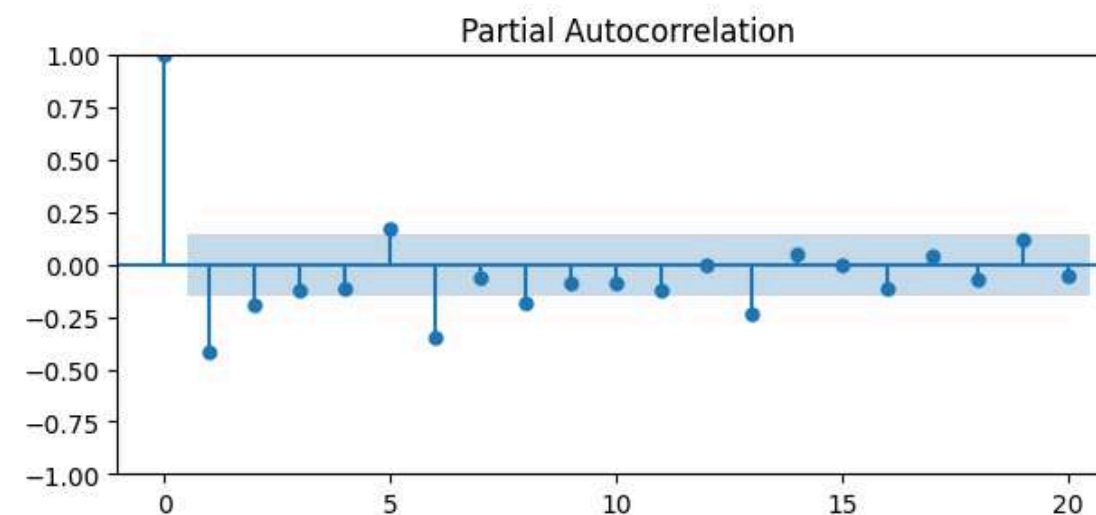
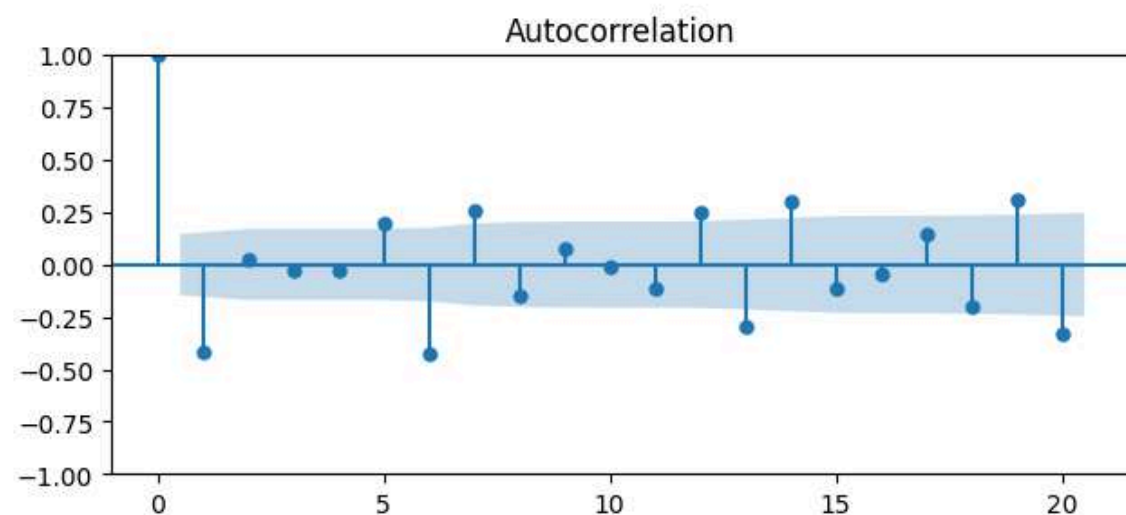
ACF and PACF for Astoria



ACF and PACF for Hell's Kitchen



ACF and PACF for Lower Manhattan



- Dapat dilihat untuk ACF dan PACF plot untuk ketiga store
- Astoria: $q(1)$, $p(1,2,3,4)$
- Hells Kitchen: $q(1)$, $p(1,2,3,4)$
- Lower Manhattan: $q(1)$, $p(1,2,3,4)$
- Sehingga dibuatlah kombinasi untuk parameter AR dan MA dari ACF dan PACF tersebut kemudian dilakukan pemilihan kombinasi terbaik dengan nilai AIC terkecil

Uji Diagnostic SARIMA Model Untuk Astoria

SARIMA Model Summary for Astoria:

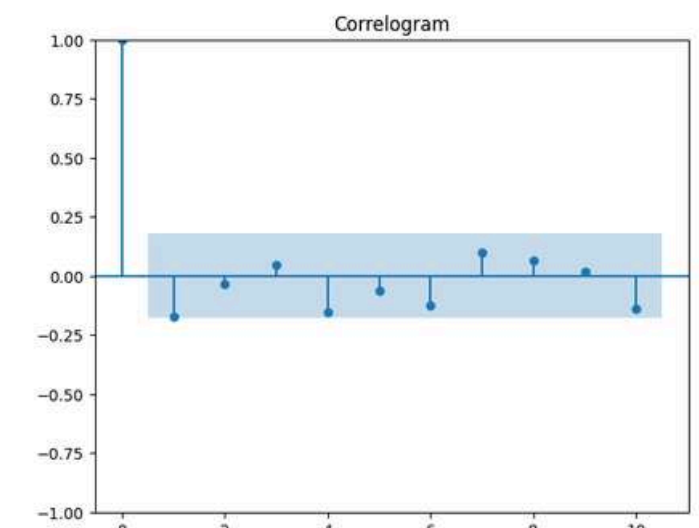
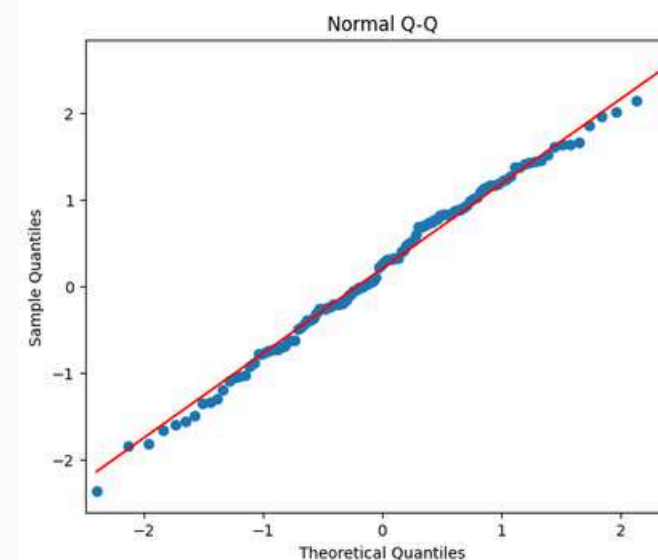
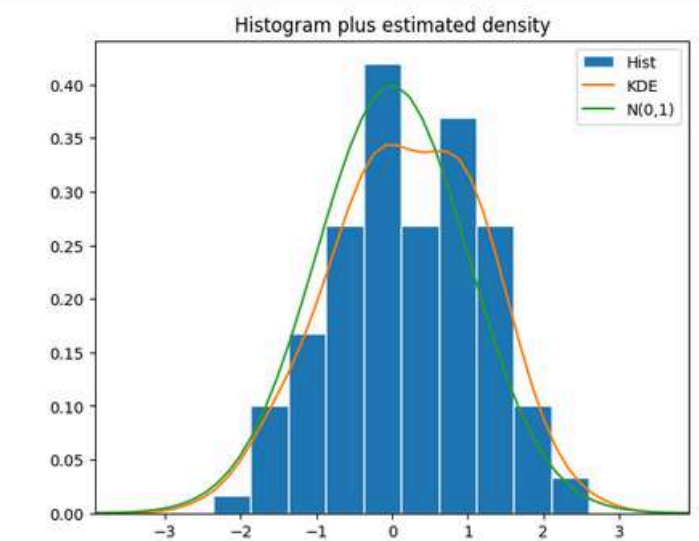
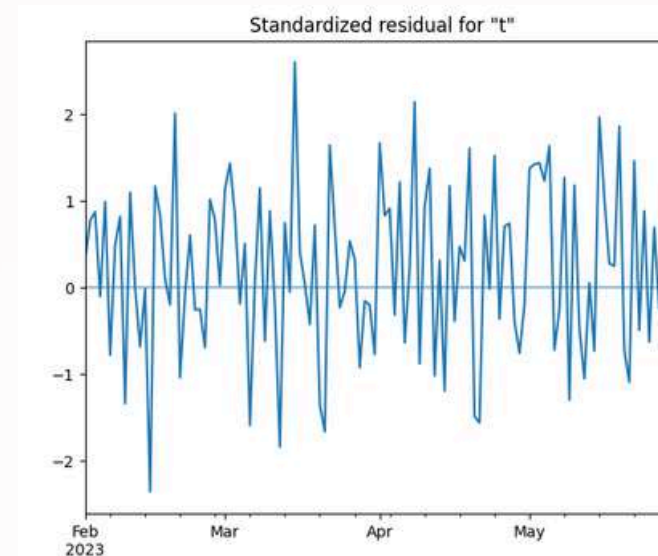
SARIMAX Results

```
=====
Dep. Variable:      transaction_qty    No. Observations:      151
Model:              SARIMAX(0, 1, 1)x(0, 1, 1, 30)  Log Likelihood        -658.845
Date:               Fri, 22 Nov 2024    AIC                   1323.689
Time:               05:57:19           BIC                   1332.052
Sample:             01-01-2023         HQIC                  1327.085
                  - 05-31-2023
=====
```

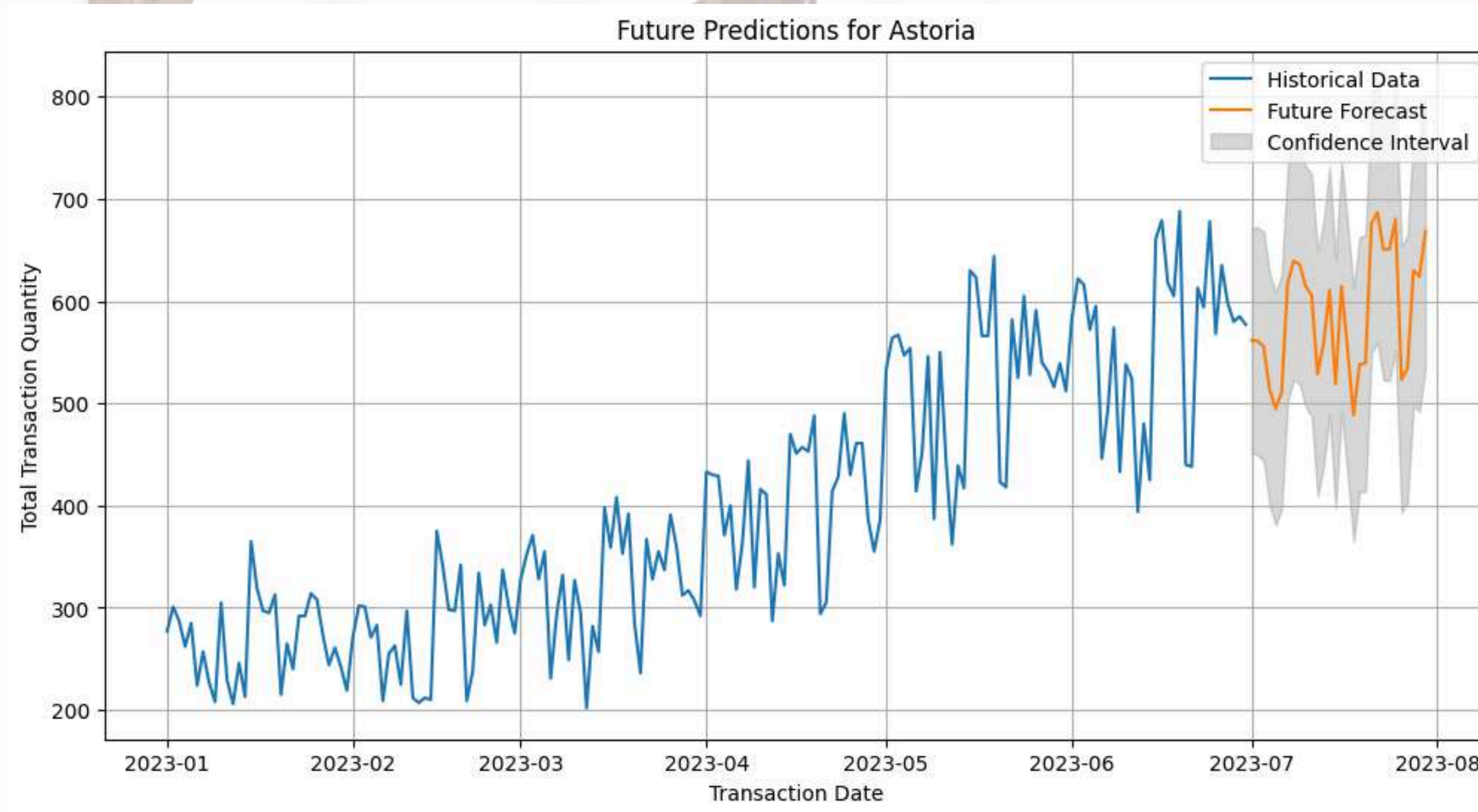
Covariance Type: opg

```
=====
              coef    std err          z      P>|z|      [0.025    0.975]
-----
ma.L1         -0.8750     0.047    -18.487     0.000    -0.968    -0.782
ma.S.L30       -0.6680     0.157     -4.245     0.000    -0.976    -0.360
sigma2        2949.3354   535.821     5.504     0.000   1899.145   3999.526
=====
```

```
=====
Ljung-Box (L1) (Q):      3.70    Jarque-Bera (JB):      1.41
Prob(Q):                 0.05    Prob(JB):           0.49
Heteroskedasticity (H):   1.13    Skew:               -0.15
Prob(H) (two-sided):      0.70    Kurtosis:           2.56
=====
```



Forecasting Transaction Qty Untuk Lokasi Astoria



Secara keseluruhan, data historis menunjukkan adanya tren meningkat dalam jumlah transaksi. Meski terdapat fluktuasi harian yang signifikan, tampak bahwa transaksi secara umum naik dari sekitar 300 di awal tahun menuju ke 600 di bulan Juni-Juli 2023, sehingga Astoria dapat merencanakan pengadaan stock.

MAPE : 12.3%

Uji Diagnostic SARIMA Model Untuk Hell's Kitchen

SARIMA Model Summary for Hell's Kitchen:

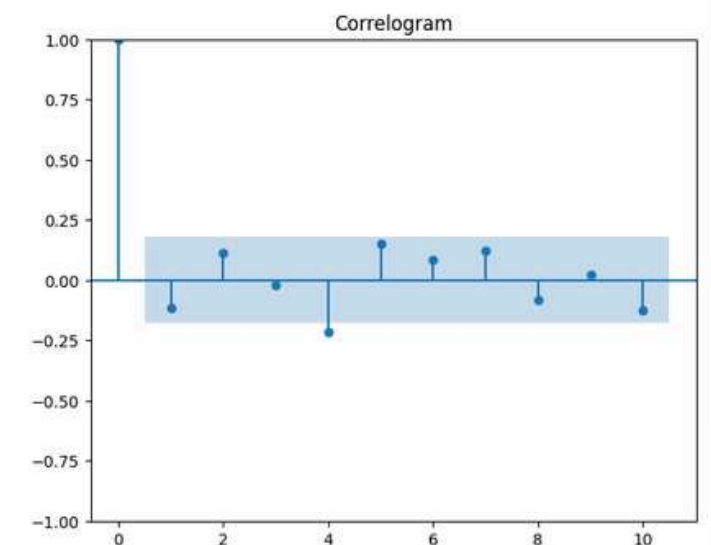
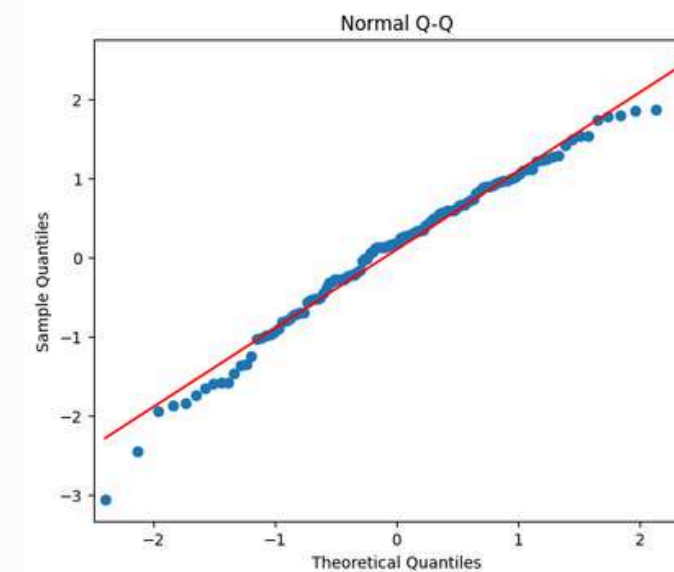
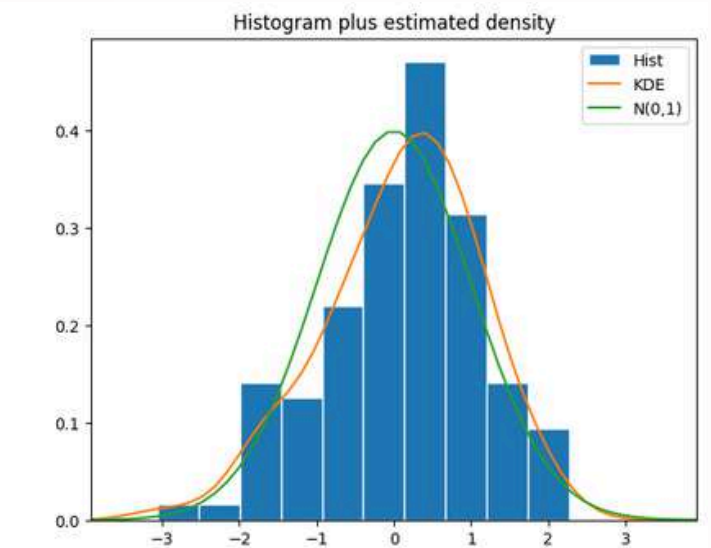
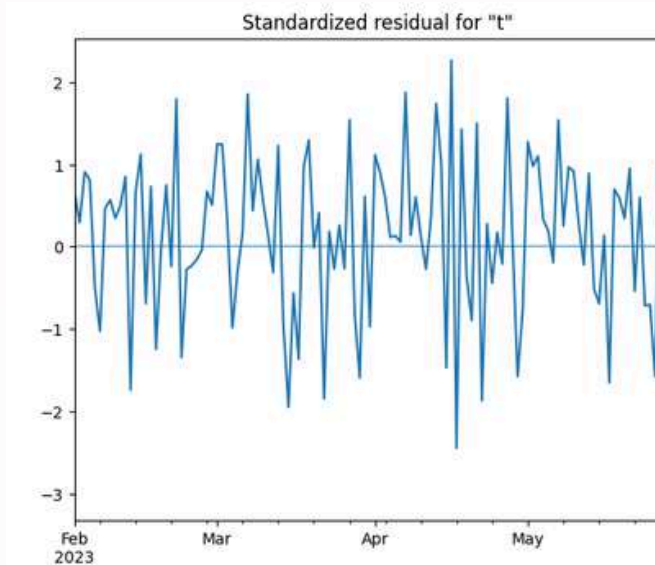
SARIMAX Results

```
=====
Dep. Variable:      transaction_qty    No. Observations:      151
Model:              SARIMAX(0, 1, 1)x(0, 1, 1, 30)    Log Likelihood        -649.958
Date:               Fri, 22 Nov 2024    AIC                   1305.916
Time:               05:57:41           BIC                   1314.278
Sample:             01-01-2023         HQIC                  1309.312
                  - 05-31-2023
Covariance Type:    opg
=====
```

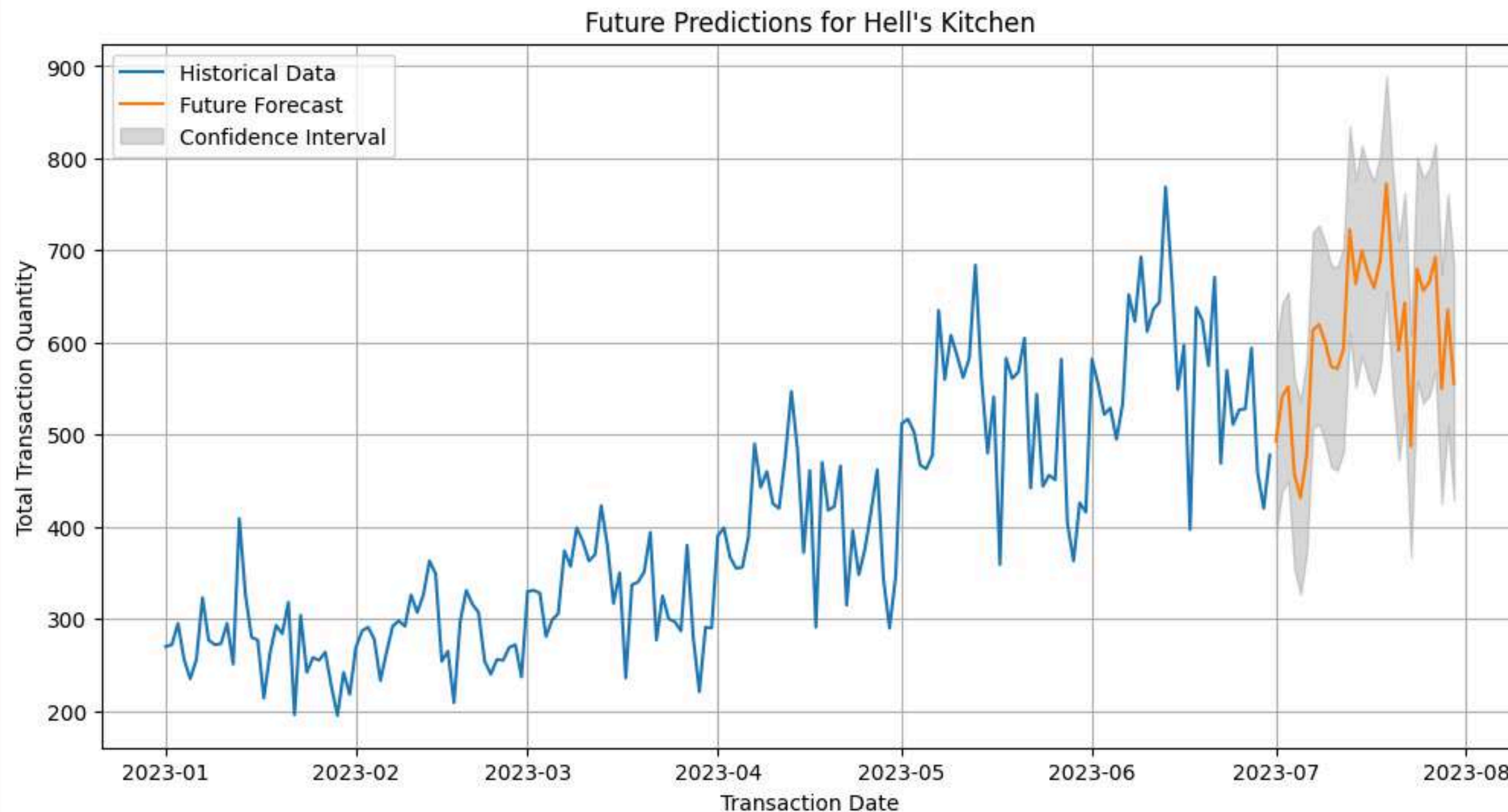
	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.8355	0.055	-15.122	0.000	-0.944	-0.727
ma.S.L30	-0.3553	0.114	-3.115	0.002	-0.579	-0.132
sigma2	2841.2499	458.853	6.192	0.000	1941.915	3740.585

Ljung-Box (L1) (Q):	1.68	Jarque-Bera (JB):	4.67
Prob(Q):	0.19	Prob(JB):	0.10
Heteroskedasticity (H):	1.46	Skew:	-0.48
Prob(H) (two-sided):	0.24	Kurtosis:	3.11

```
=====
```



Forecasting Transaction Qty Untuk Lokasi Hell's Kitchen



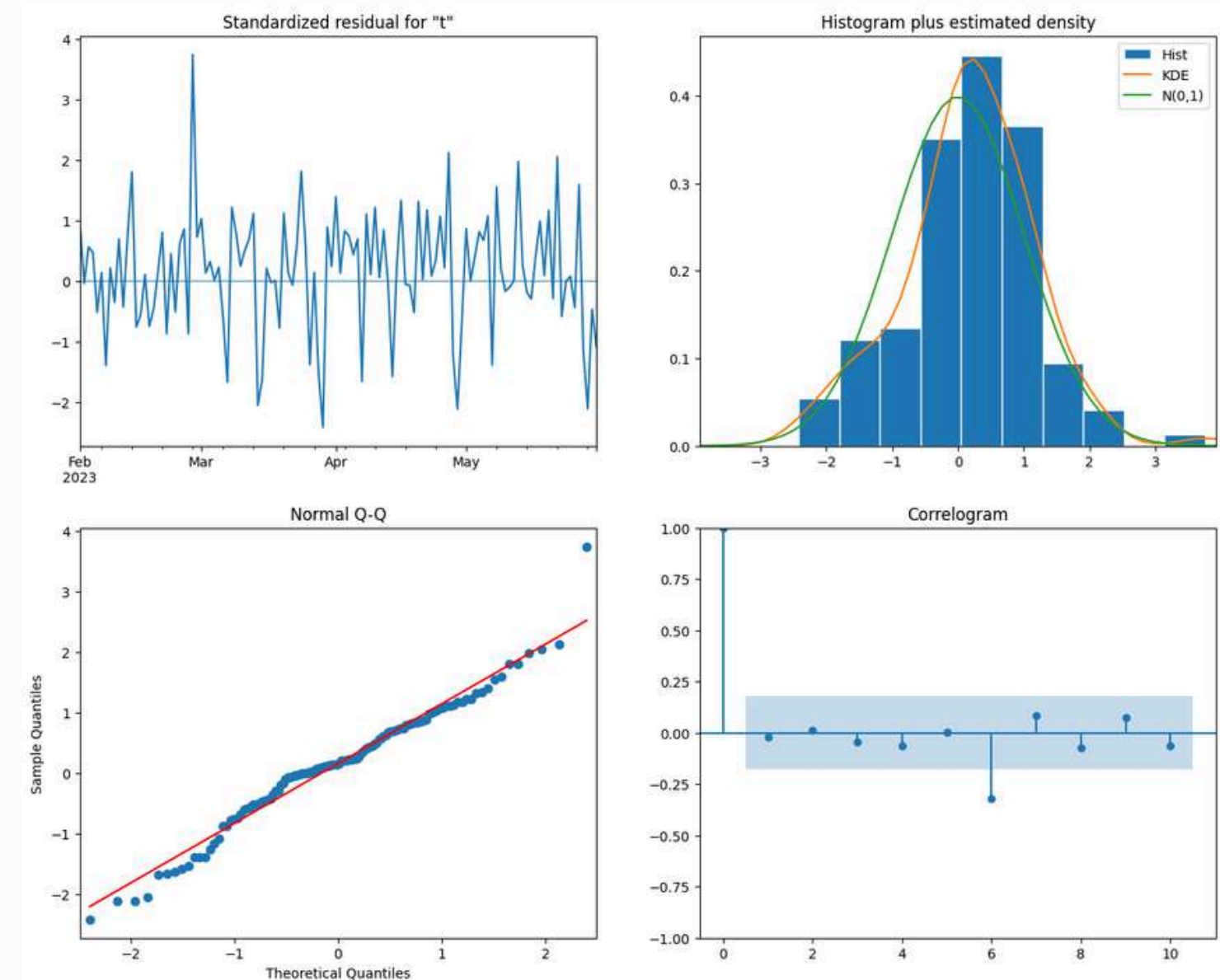
MAPE : 12%

Hell's Kitchen mengalami peningkatan konsisten dalam jumlah transaksi selama paruh pertama tahun 2023. Hal ini dapat mencerminkan pertumbuhan bisnis, peningkatan popularitas lokasi, atau faktor ekonomi lokal.

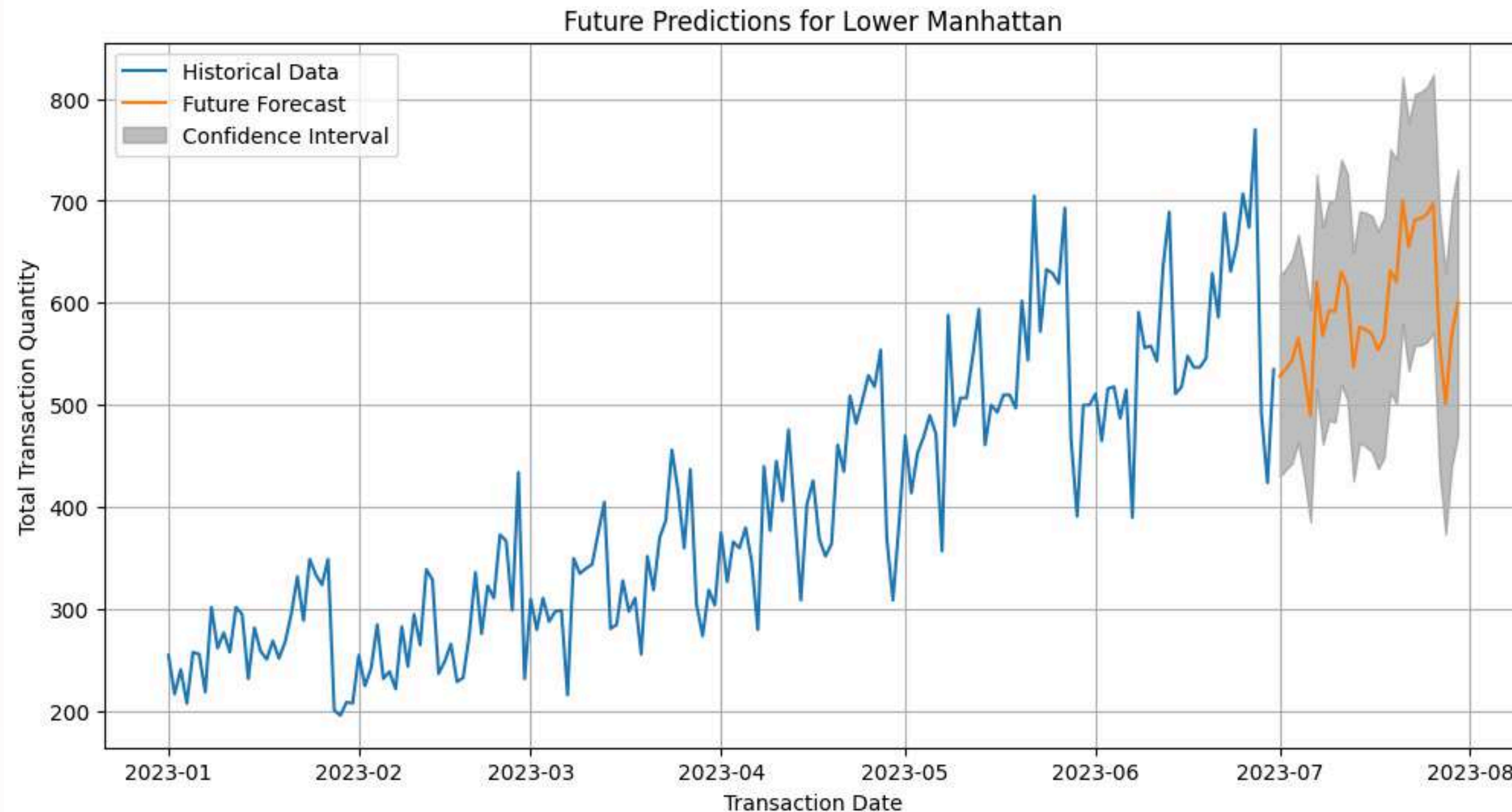
Dengan hasil forecasting transaction qty yang stabil, Hell's Kitchen dapat merencanakan operasional yang lebih efisien untuk pengadaan barang.

Uji Diagnostic SARIMA Model Untuk Lower Manhattan

```
SARIMA Model Summary for Lower Manhattan:
=====
                        SARIMAX Results
=====
Dep. Variable:          transaction_qty    No. Observations:      151
Model:                 SARIMAX(1, 1, 1)x(0, 1, 1, 30)    Log Likelihood        -647.561
Date:                 Fri, 22 Nov 2024    AIC                  1303.123
Time:                 05:57:49    BIC                  1314.273
Sample:               01-01-2023    HQIC                 1307.651
                   - 05-31-2023
Covariance Type:      opg
=====
              coef    std err          z      P>|z|      [0.025    0.975]
-----
ar.L1         -0.2280     0.092     -2.485     0.013     -0.408     -0.048
ma.L1         -0.8008     0.066    -12.206     0.000     -0.929     -0.672
ma.S.L30      -0.6241     0.129     -4.848     0.000     -0.876     -0.372
sigma2        2498.6100    357.159      6.996     0.000    1798.592    3198.628
=====
Ljung-Box (L1) (Q):           0.05    Jarque-Bera (JB):           4.75
Prob(Q):                     0.82    Prob(JB):                 0.09
Heteroskedasticity (H):       1.19    Skew:                     -0.04
Prob(H) (two-sided):          0.58    Kurtosis:                 3.97
=====
```



Forecasting Transaction Qty Untuk Lokasi Lower Manhattan



Data historis menunjukkan pertumbuhan stabil di Lower Manhattan, yang dapat mencerminkan peningkatan aktivitas ekonomi atau promosi yang efektif di daerah tersebut. Model forecasting menunjukkan stabilitas pada jumlah transaksi. Ini berarti model memperkirakan bahwa pola historis akan terus berlanjut.

MAPE : 11.17%

ATTACHMENTS

Power BI File

Python



**THANK
YOU**

