



PRODUCT SALES FORECASTING TIME-SERIES ANALYSIS WITH XGBOOST AND LSTM

[GITHUB](#)

LATAR BELAKANG



Dalam era digital dan perkembangan teknologi yang sangat pesat, industri elektronik mengalami pertumbuhan yang signifikan baik dari sisi produk, permintaan pasar, maupun persaingan antar pelaku usaha. Barang-barang elektronik seperti smartphone, laptop, televisi, perangkat rumah pintar, dan aksesoris digital menjadi kebutuhan primer bagi masyarakat modern. Seiring dengan meningkatnya kebutuhan tersebut, pelaku bisnis dituntut untuk dapat memprediksi permintaan pasar dengan lebih akurat agar mampu menyusun strategi penjualan yang efektif dan efisien.

Sample Dataset



Customer ID	Age	Gender	Loyalty Member	Product Type	SKU	Rating	Order Status	Payment Method	Total Price	Unit Price	Quantity	Purchase Date	Shipping Type	Add-ons Purchased	Add-on Total
1000	53	Male	No	Smartphone	SKU1004	2	Cancelled	Credit Card	5538.33	791.19	7	2024-03-20	Standard	Accessory,Accessory,Accessory	40.21
1000	53	Male	No	Tablet	SKU1002	3	Completed	Paypal	741.09	247.03	3	2024-04-20	Overnight	Impulse Item	26.09
1002	41	Male	No	Laptop	SKU1005	3	Completed	Credit Card	1855.84	463.96	4	2023-10-17	Express		0
1002	41	Male	Yes	Smartphone	SKU1004	2	Completed	Cash	3164.76	791.19	4	2024-08-09	Overnight	Impulse Item,Impulse Item	60.16
1003	75	Male	Yes	Smartphone	SKU1001	5	Completed	Cash	41.5	20.75	2	2024-05-21	Express	Accessory	35.56
1004	41	Female	No	Smartphone	SKU1001	5	Completed	Credit Card	83	20.75	4	2024-05-26	Standard	Impulse Item,Accessory	65.78
1005	25	Female	No	Smartwatch	SKU1003	3	Completed	Paypal	7603.47	844.83	9	2024-01-30	Overnight		0
1005	25	Female	No	Laptop	SKU1005	3	Completed	Debit Card	4175.64	463.96	9	2024-06-24	Overnight	Extended Warranty,Extended Warranty	75.33
1006	24	Male	No	Smartphone	SKU1004	2	Cancelled	Debit Card	5538.33	791.19	7	2023-10-03	Standard	Impulse Item	43.05
1006	24	Male	Yes	Laptop	SKU1005	3	Completed	Cash	4175.64	463.96	9	2024-01-01	Express		0
1006	24	Male	Yes	Tablet	SKU1002	3	Completed	Paypal	2470.3	247.03	10	2024-06-25	Overnight	Impulse Item,Accessory,Impulse Item	90.38
1007	35	Male	No	Smartphone	SKU1004	2	Cancelled	Credit Card	7120.71	791.19	9	2023-11-07	Overnight	Accessory,Extended Warranty	55.48
1008	66	Female	No	Smartwatch	SKU1003	3	Completed	Cash	3379.32	844.83	4	2023-09-25	Express	Extended Warranty,Impulse Item,Extended Warranty	65.85
1011	72	Male	No	Smartphone	SKU1004	2	Completed	Credit Card	7911.9	791.19	10	2024-07-07	Standard	Accessory,Accessory,Accessory	70.17
1013	42	Female	No	Smartphone	SKU1004	2	Cancelled	Paypal	5538.33	791.19	7	2024-03-18	Overnight	Accessory,Impulse Item,Accessory	57.81
1014	73	Female	No	Smartwatch	SKU1003	3	Cancelled	Credit Card	5068.98	844.83	6	2024-01-02	Standard	Extended Warranty,Extended Warranty	69.31

Referensi Dataset Kaggle

DATA CLEANING PREPROCESSING



Dalam tahap pembersihan dan pra-pemrosesan data, dilakukan beberapa langkah berikut untuk memastikan kualitas dan konsistensi data:

1. Penghapusan Kolom yang Tidak Relevan

Kolom CustomerID dan Shipping Type dihapus karena tidak memberikan kontribusi signifikan terhadap analisis atau pemodelan yang akan dilakukan.

2. Konversi Tipe Data Tanggal

Seluruh kolom yang berisi informasi tanggal dikonversi ke format datetime yang konsisten untuk memudahkan analisis berbasis waktu.

3. Penanganan Nilai Kosong (Missing Values)

Dilakukan pengecekan terhadap keberadaan nilai kosong (NaN). Semua baris yang mengandung nilai kosong dihapus untuk menjaga integritas data.

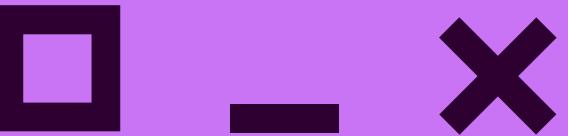
4. Identifikasi dan Penghapusan Duplikasi

Dataset diperiksa terhadap adanya baris yang duplikat. Baris duplikat yang ditemukan kemudian dihapus agar tidak memengaruhi hasil analisis.

5. Ekstraksi Informasi Waktu

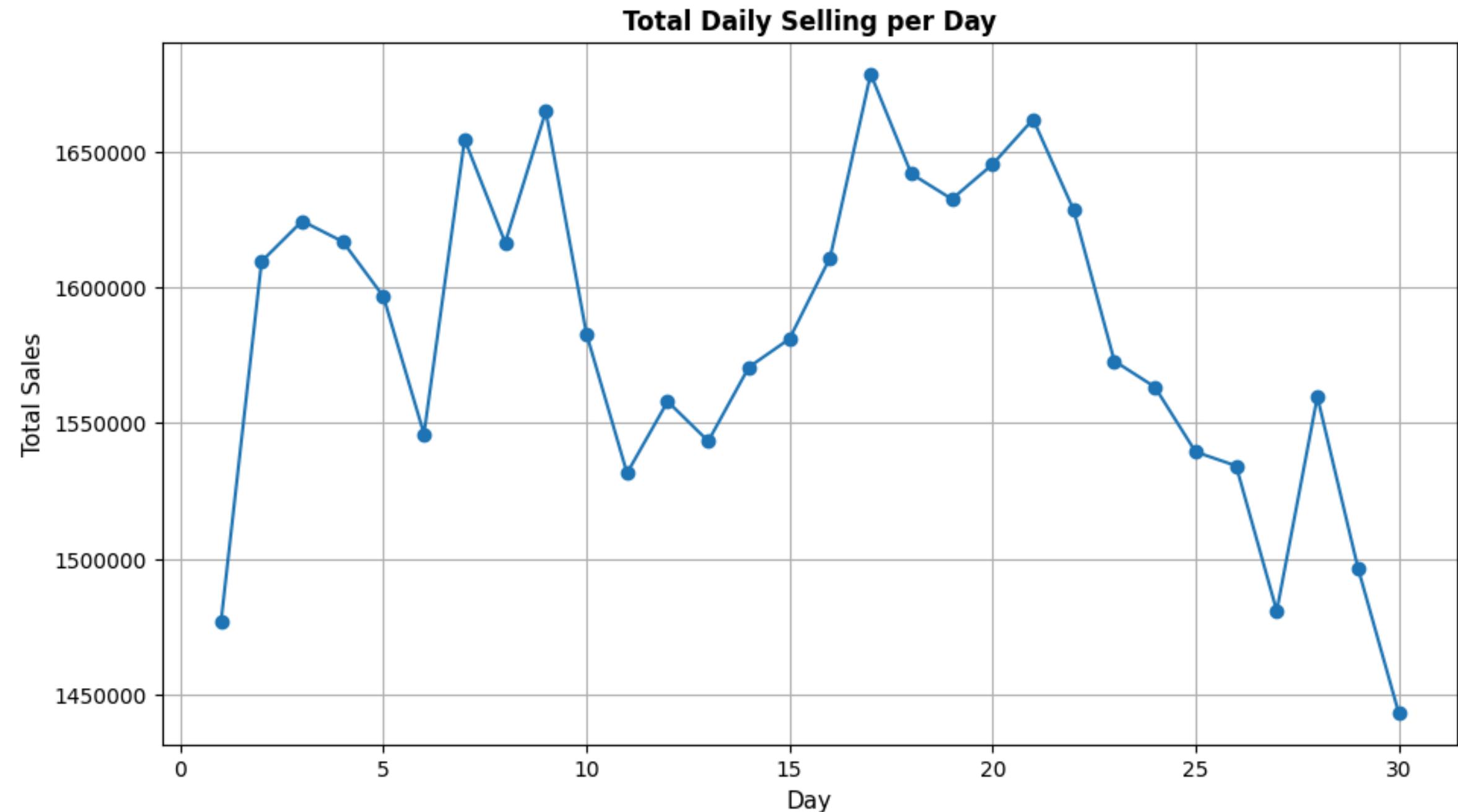
Dari kolom tanggal pembelian (Purchase Date), diturunkan informasi tambahan seperti month, day of week, week , dan day. Hal ini dilakukan untuk memperkaya fitur dan meningkatkan kualitas analisis atau prediksi berbasis waktu

DATA CLEANING PREPROCESSING RESULT



	Age	Gender	Loyalty Member	Product Type	SKU	Rating	Order Status	Payment Method	Total Price	Unit Price	Quantity	Purchase Date	Add-ons Purchased	Add-on Total	Months	Day of Week	Week	Day
0	53	Male	No	Smartphone	SKU1004	2	Cancelled	Credit Card	5538.33	791.19	7	2024-03-20	Accessory,Accessory,Accessory	40.21	March	Wednesday	12	20
1	53	Male	No	Tablet	SKU1002	3	Completed	Paypal	741.09	247.03	3	2024-04-20	Impulse Item	26.09	April	Saturday	16	20
3	41	Male	Yes	Smartphone	SKU1004	2	Completed	Cash	3164.76	791.19	4	2024-08-09	Impulse Item,Impulse Item	60.16	August	Friday	32	9
4	75	Male	Yes	Smartphone	SKU1001	5	Completed	Cash	41.50	20.75	2	2024-05-21	Accessory	35.56	May	Tuesday	21	21
5	41	Female	No	Smartphone	SKU1001	5	Completed	Credit Card	83.00	20.75	4	2024-05-26	Impulse Item,Accessory	65.78	May	Sunday	21	26

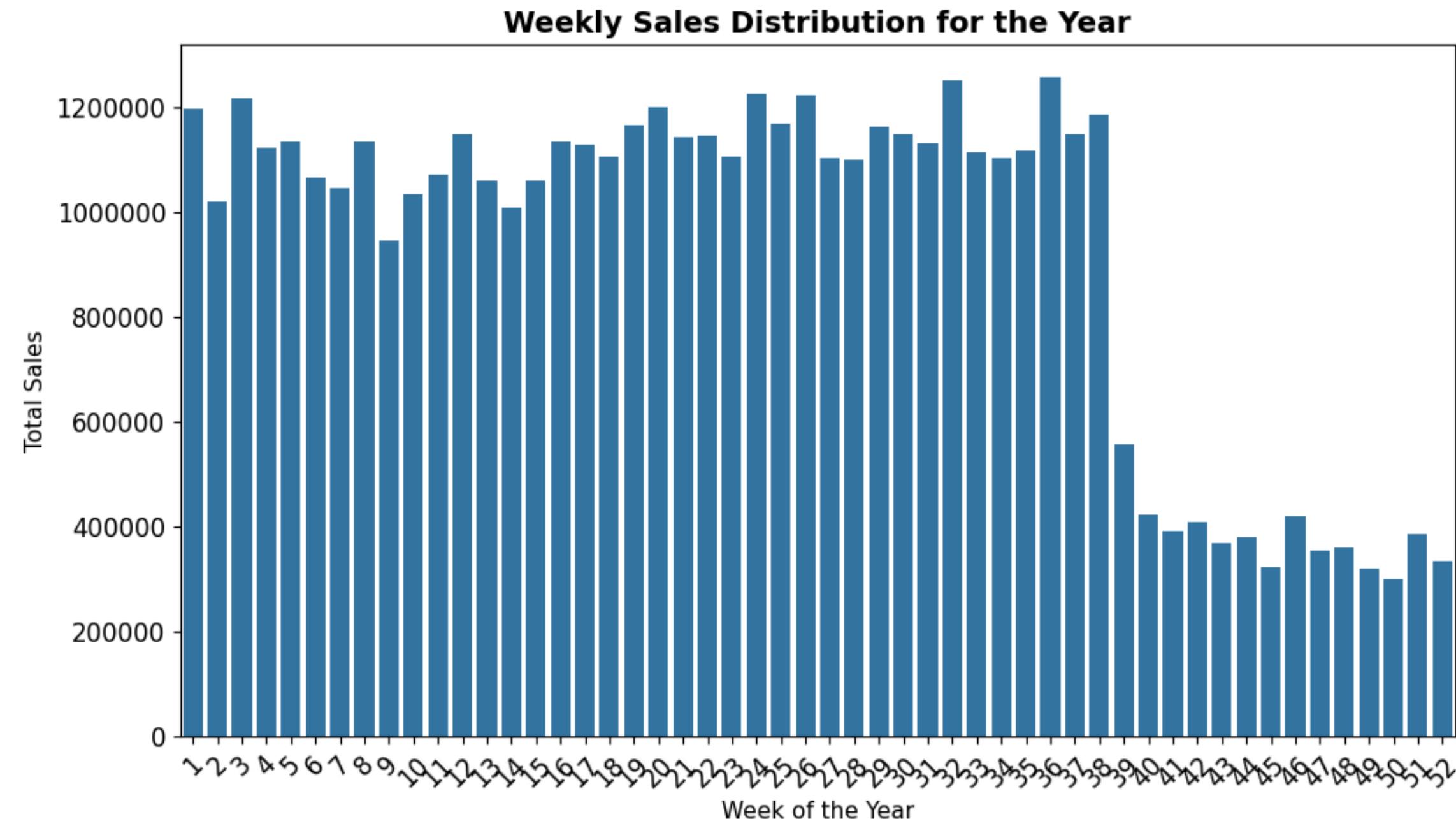
Sales Analysis (Daily Selling per Day)



Sumbu horizontal pada grafik ini diberi label "Day" yang merepresentasikan rentang waktu dalam satu bulan, yaitu dari hari ke-1 hingga hari ke-30. Setiap titik pada sumbu ini menunjukkan satu hari tertentu dalam periode waktu tersebut.

Sumbu Vertical pada grafik ini diberi label "Total Sales", yang menunjukkan besarnya nilai penjualan yang dicapai setiap harinya selama periode 30 hari.

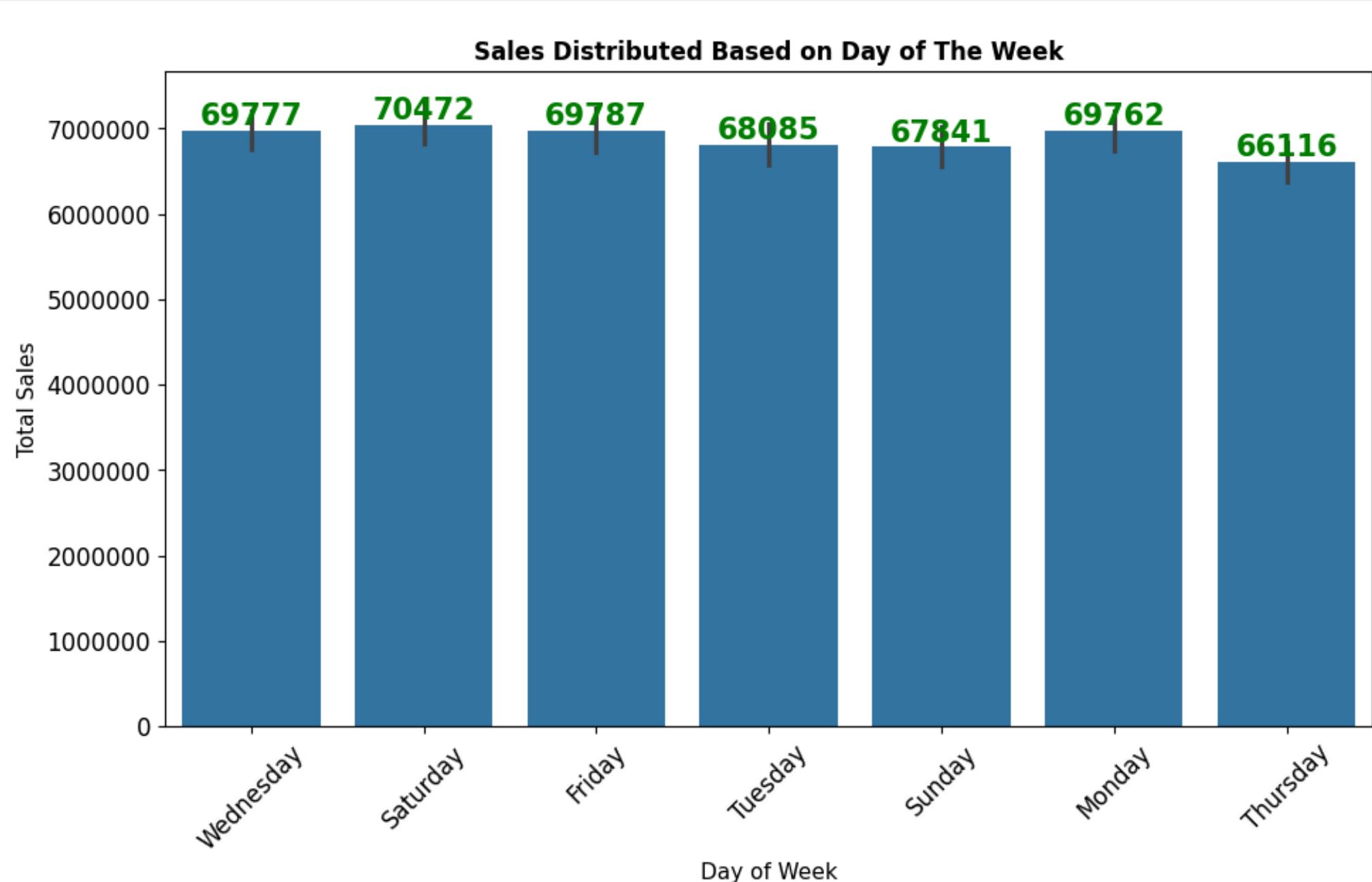
Sales Analysis (Weekly)



Sumbu vertikal menunjukkan total penjualan dalam satuan nilai. Setiap bar pada diagram akan menunjukkan jumlah penjualan yang dicapai dalam minggu tertentu.

Sumbu horizontal mewakili minggu-minggu dalam setahun. Diagram ini bertujuan untuk menunjukkan pola atau tren penjualan dari minggu ke minggu

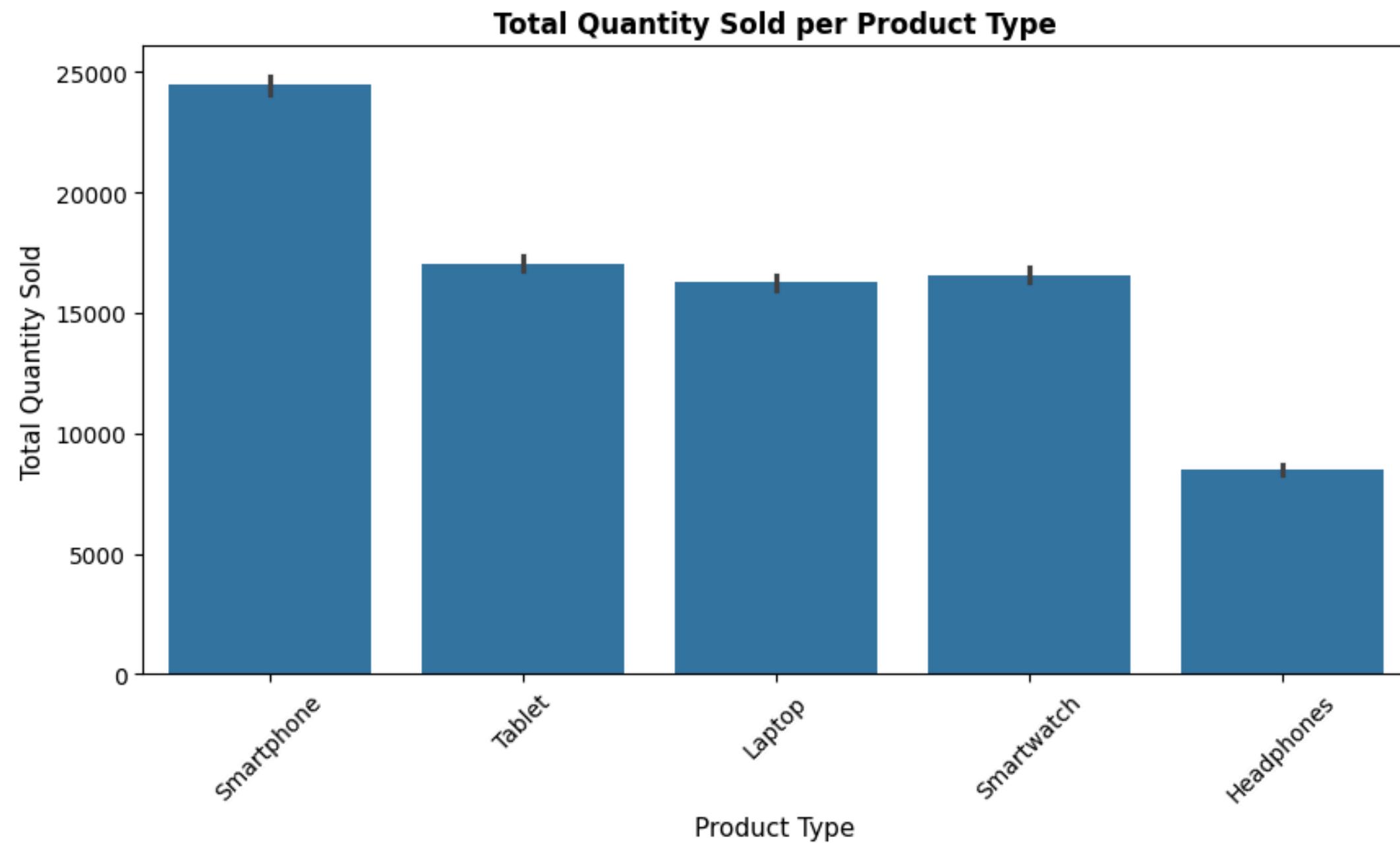
Sales Analysis (Day of the Week)



Sumbu horizontal pada grafik ini diberi label "Day of Week", yang menampilkan nama hari dalam 1 minggu. Masing-masing bar merepresentasikan total penjualan yang terjadi pada masing-masing hari tersebut selama periode pengamatan.

Sumbu Vertical pada grafik ini diberi label "Total Sales", yang menunjukkan volume penjualan kumulatif yang berhasil dicapai pada masing-masing hari

Product Analysis



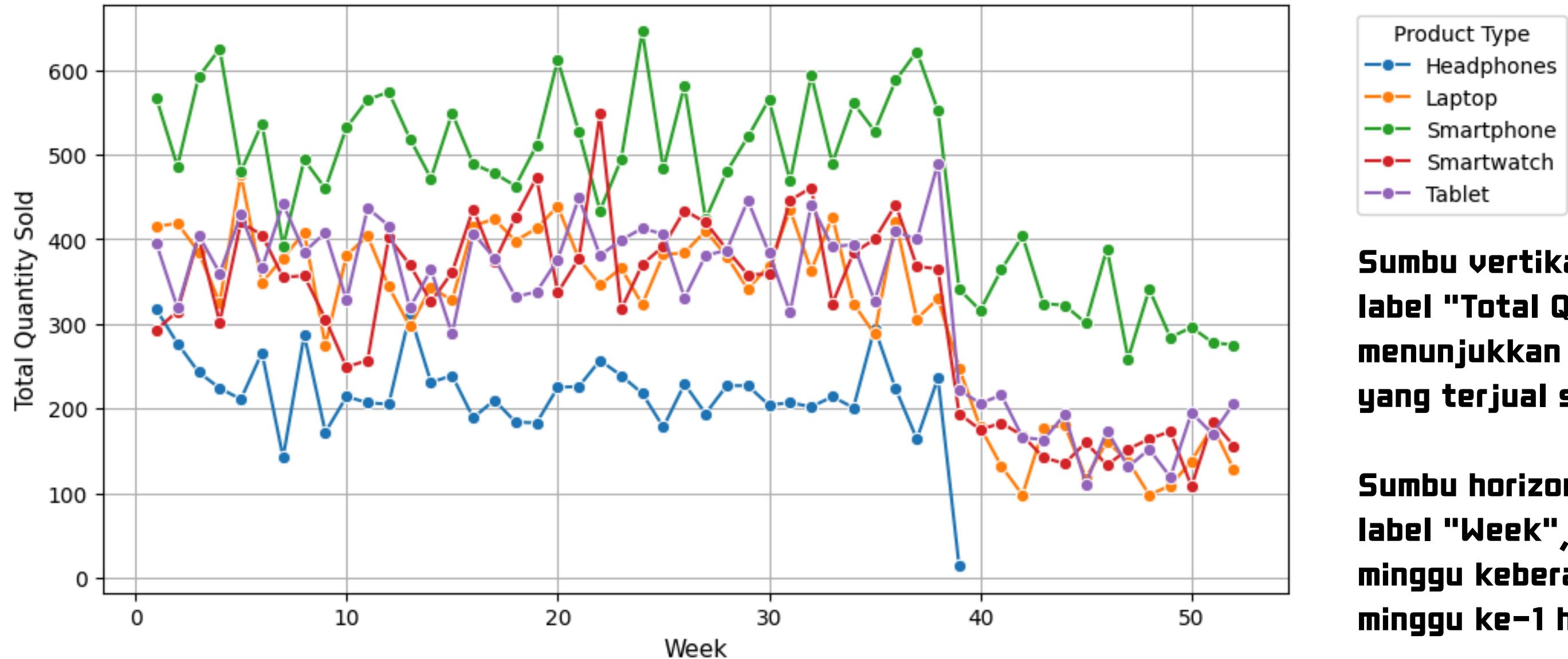
Sumbu vertical pada grafik ini diberi label "Total Quantity Sold", yang menampilkan total quantity yang terjual dari tiap product.

Sumbu horizontal pada grafik ini diberi label "Product Type", yang menampilkan nama dari product yang terjual.

Product Analysis



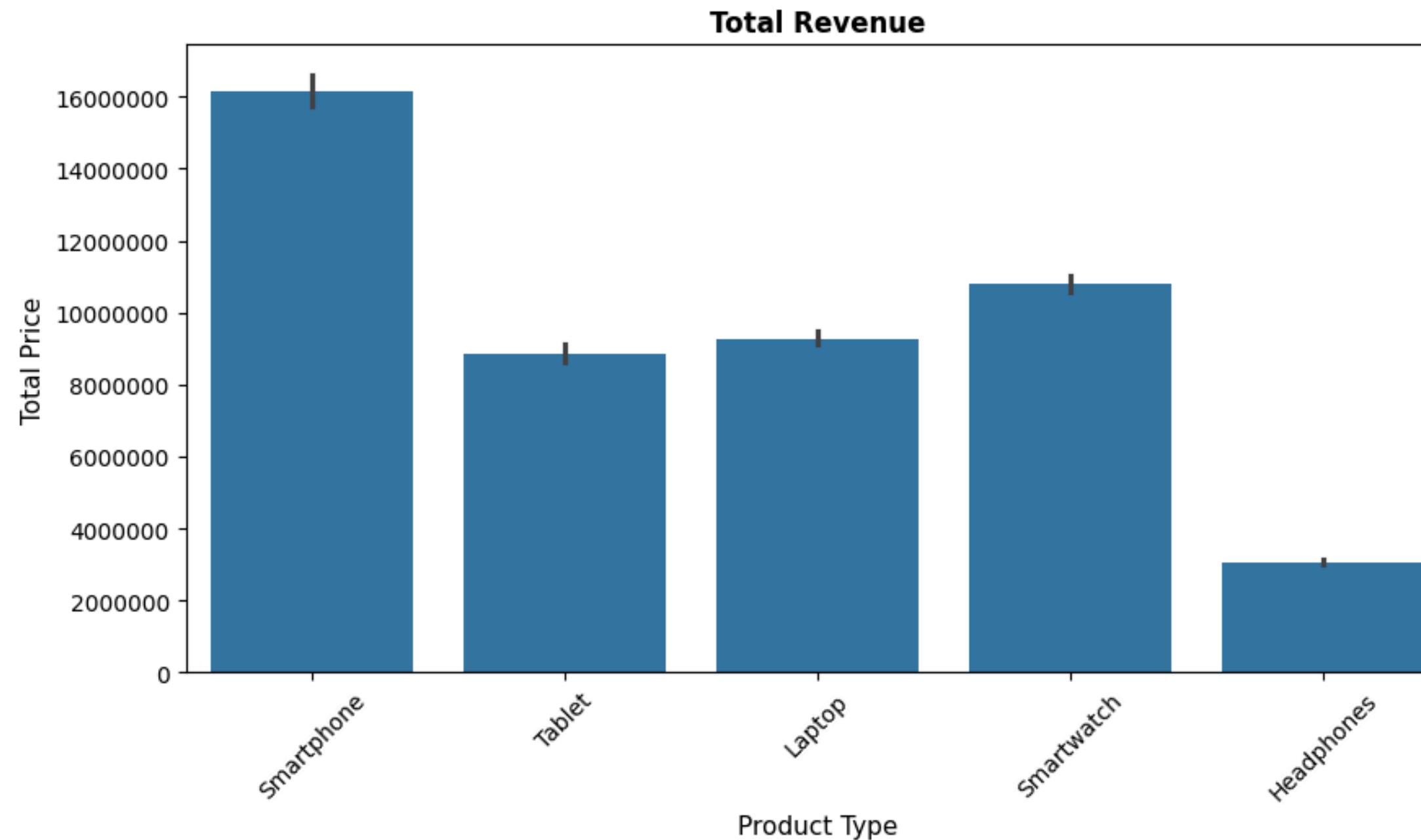
Weekly Best Selling Products



Sumbu vertikal pada grafik ini diberi label "Total Quantity Sold", yang menunjukkan total kuantitas produk yang terjual setiap minggu.

Sumbu horizontal pada grafik ini diberi label "Week", yang menunjukkan minggu keberapa dalam setahun (dari minggu ke-1 hingga minggu ke-53).

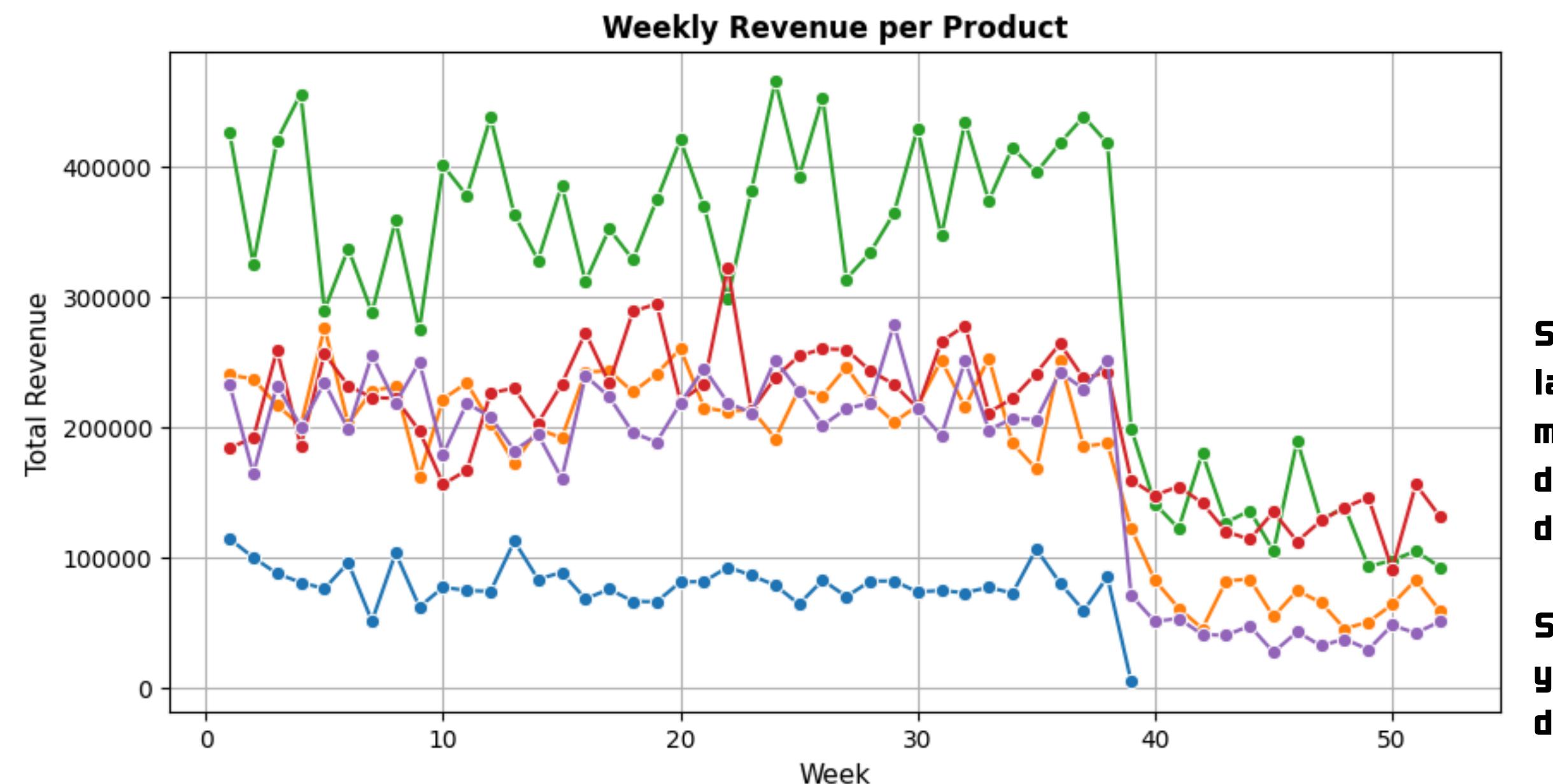
Product Analysis



Sumbu vertikal diberi label "Total Price", yang menunjukkan total pendapatan kumulatif yang dihasilkan oleh masing-masing jenis produk sepanjang tahun.

Sumbu horizontal diberi label "Product Type", yang menampilkan kategori produk: Smartphone, Tablet, Laptop, Smartwatch, dan Headphones.

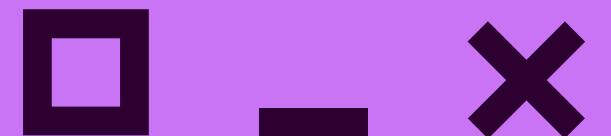
Product Analysis



Sumbu vertikal pada grafik ini diberi label "Total Revenue", yang menunjukkan total pendapatan yang dihasilkan dari penjualan setiap produk dalam satu minggu.

Sumbu horizontal diberi label "Week", yang menunjukkan minggu keberapa dalam satu tahun (1–53).

Performance Metrics



LSTM

Overall Model Accuracy:

- MAE: 14.39
- RMSE: 17.16
- R² Score: 0.16

Dalam proyek ini model yang digunakan adalah XGBoost dimana ini merupakan sebuah model yang dipakai untuk pemodelan supervised learning berbasis decision tree.

XGBoost

Overall Model Accuracy:

- MAE: 10.93
- RMSE: 14.17
- R² Score: 0.65

XGboost dipakai, karena dalam kasus ini tugas utamanya adalah prediksi numerik yang memprediksi jumlah penjualan harian dari produk-produk berdasarkan data historis.

Model Output



Forecast for the next 7 days:		
Purchase Date	Product Type	Predicted Quantity
2024-09-24	Laptop	54.036285
2024-09-25	Laptop	78.200409
2024-09-26	Laptop	124.764938
2024-09-27	Laptop	117.807045
2024-09-28	Laptop	86.838318
2024-09-29	Laptop	91.154480
2024-09-30	Laptop	90.378334
2024-09-24	Smartphone	64.832855
2024-09-25	Smartphone	75.718918
2024-09-26	Smartphone	82.263252
2024-09-27	Smartphone	82.023056
2024-09-28	Smartphone	86.529182
2024-09-29	Smartphone	90.980530
2024-09-30	Smartphone	91.488327
2024-09-24	Smartwatch	43.851822
2024-09-25	Smartwatch	30.508842
2024-09-26	Smartwatch	9.257240
2024-09-27	Smartwatch	1.680744
2024-09-28	Smartwatch	17.220343
2024-09-29	Smartwatch	13.724504
2024-09-30	Smartwatch	12.439455
2024-09-24	Tablet	64.612808
2024-09-25	Tablet	61.520088
2024-09-26	Tablet	60.810642
2024-09-27	Tablet	65.683113
2024-09-28	Tablet	85.172424
2024-09-29	Tablet	105.026382
2024-09-30	Tablet	100.459190
2024-09-24	Headphones	27.494343
2024-09-25	Headphones	21.163782
2024-09-26	Headphones	14.101980
2024-09-27	Headphones	13.565019
2024-09-28	Headphones	9.632977
2024-09-29	Headphones	10.547470
2024-09-30	Headphones	19.495998

Top-Selling Product Types:		
	Product Type	Total Sales
1	Smartphone	24207
2	Tablet	16880
3	Smartwatch	16389
4	Laptop	16125
5	Headphones	8176

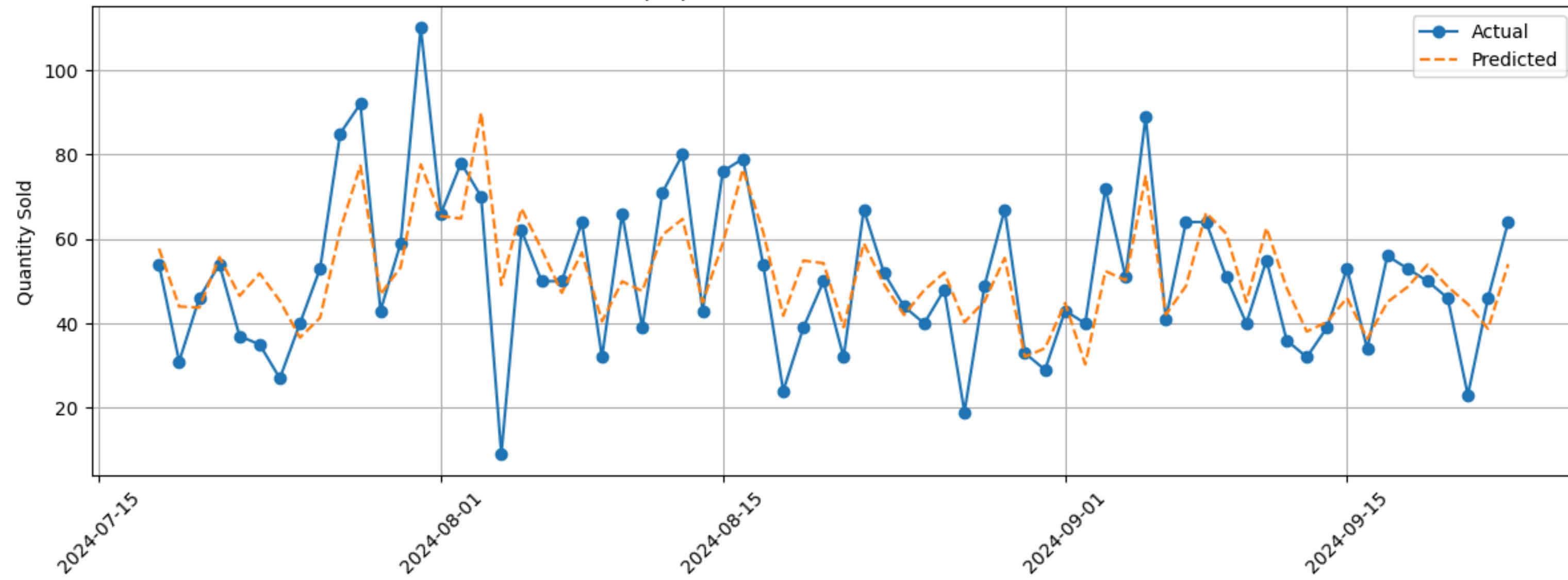
Tujuan dari model ini adalah untuk memperkirakan jumlah penjualan harian dari masing-masing jenis produk selama tujuh hari ke depan, berdasarkan data historis dan pola yang telah dianalisis dari dataset.

Di dalam output terdapat prediksi kuantitas penjualan harian untuk produk-produk elektronik yang dijual. Dan juga terdapat daftar produk yang terlaris berdasarkan total penjualan historis dimana outputnya menunjukkan bahwa smartphone merupakan produk dengan penjualan terlaris.

Model Prediction

□ - X

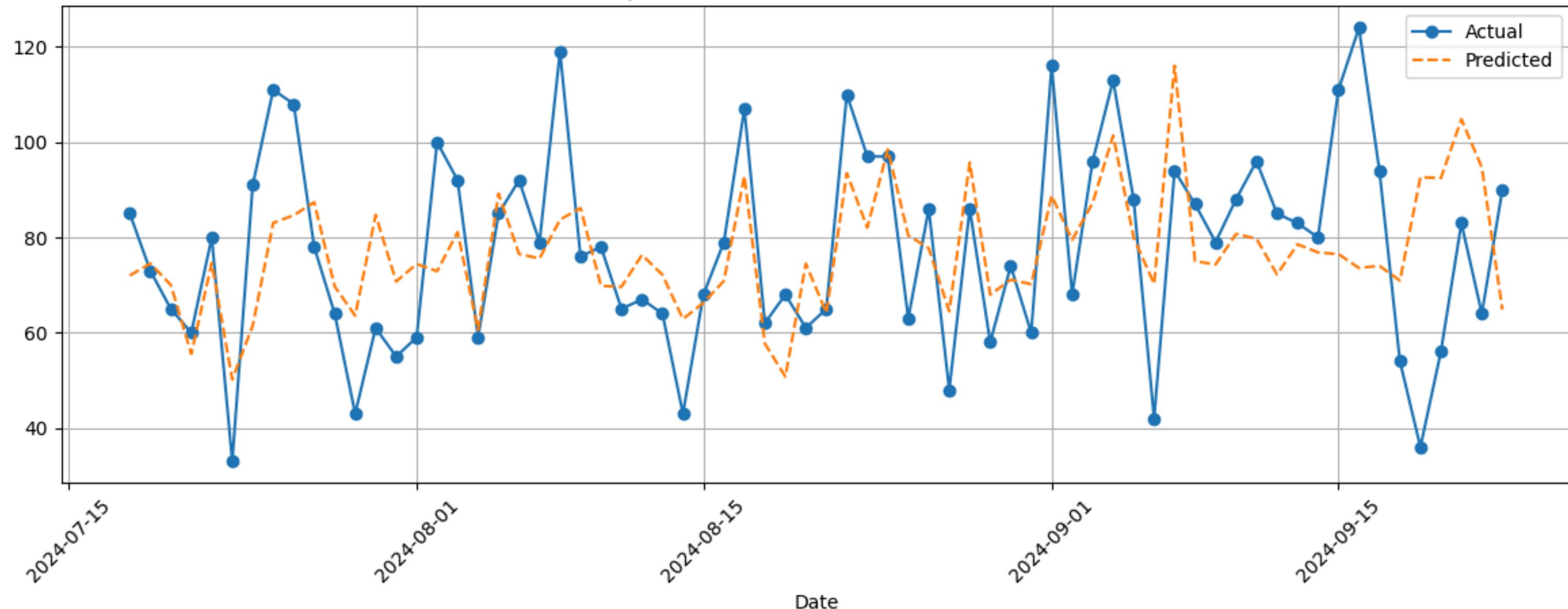
Laptop - Actual vs Predicted (MAE: 9.3)



Model Prediction

□ - X

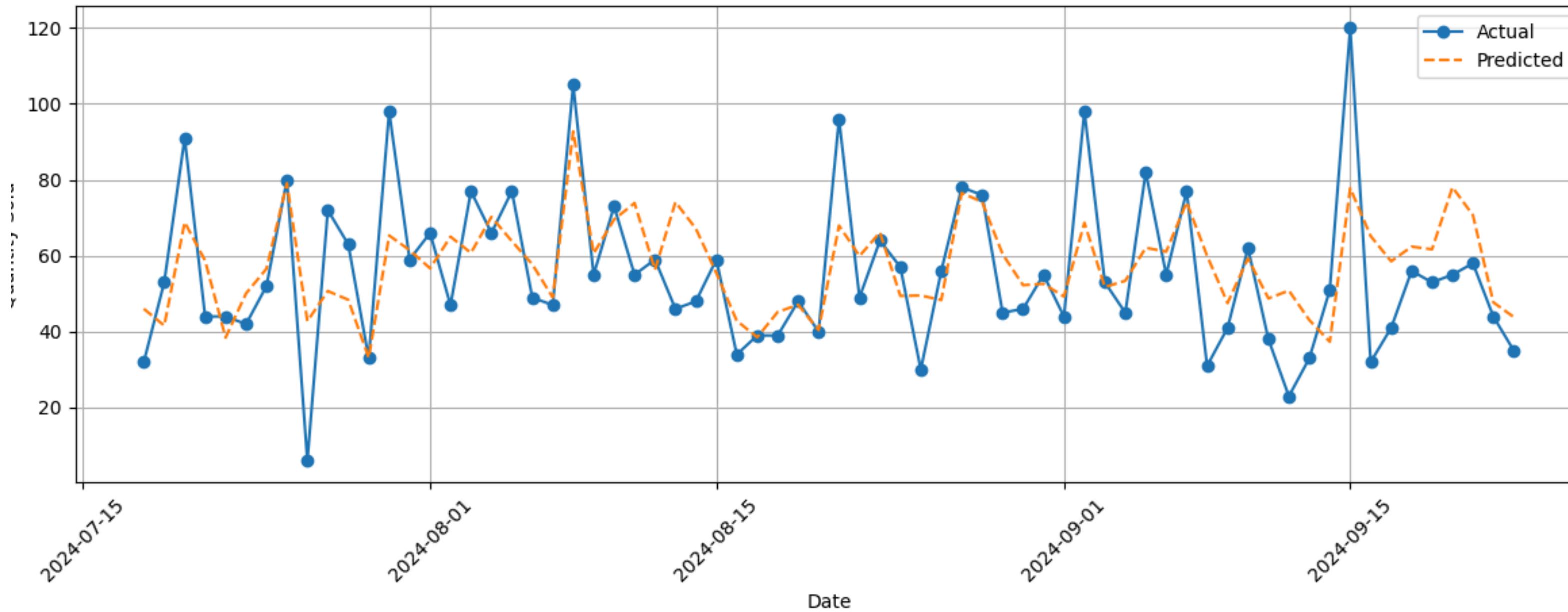
Smartphone - Actual vs Predicted (MAE: 14.9)



Model Prediction

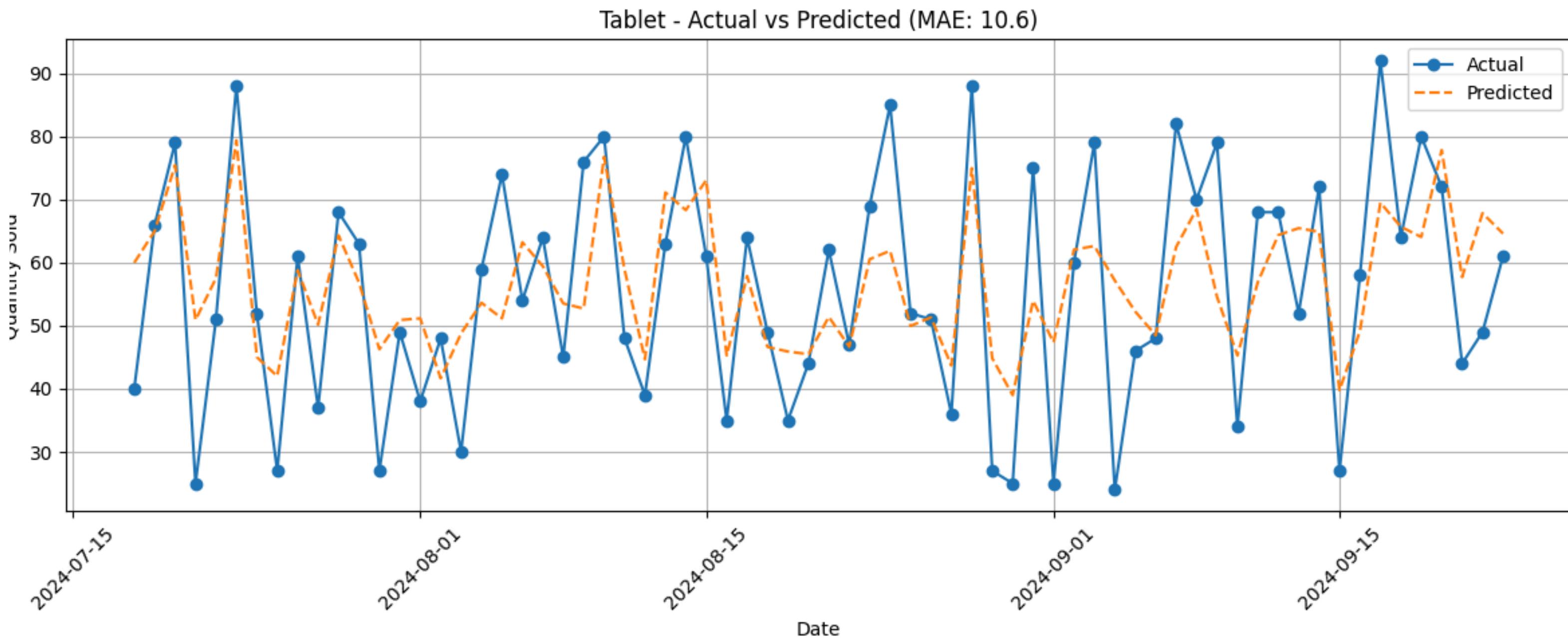
□ - X

Smartwatch - Actual vs Predicted (MAE: 11.8)



Model Prediction

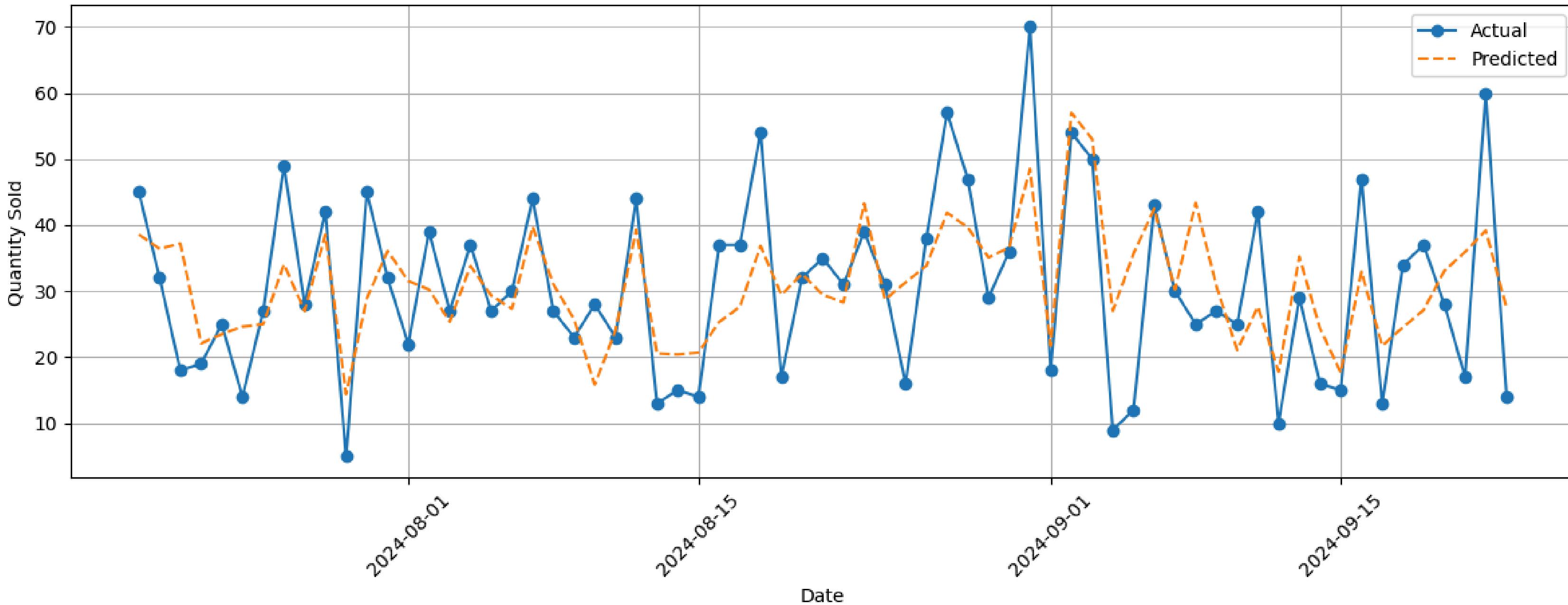
□ - X



Model Prediction

□ - X

Headphones - Actual vs Predicted (MAE: 7.9)



Dashboard Visualization by Tableau

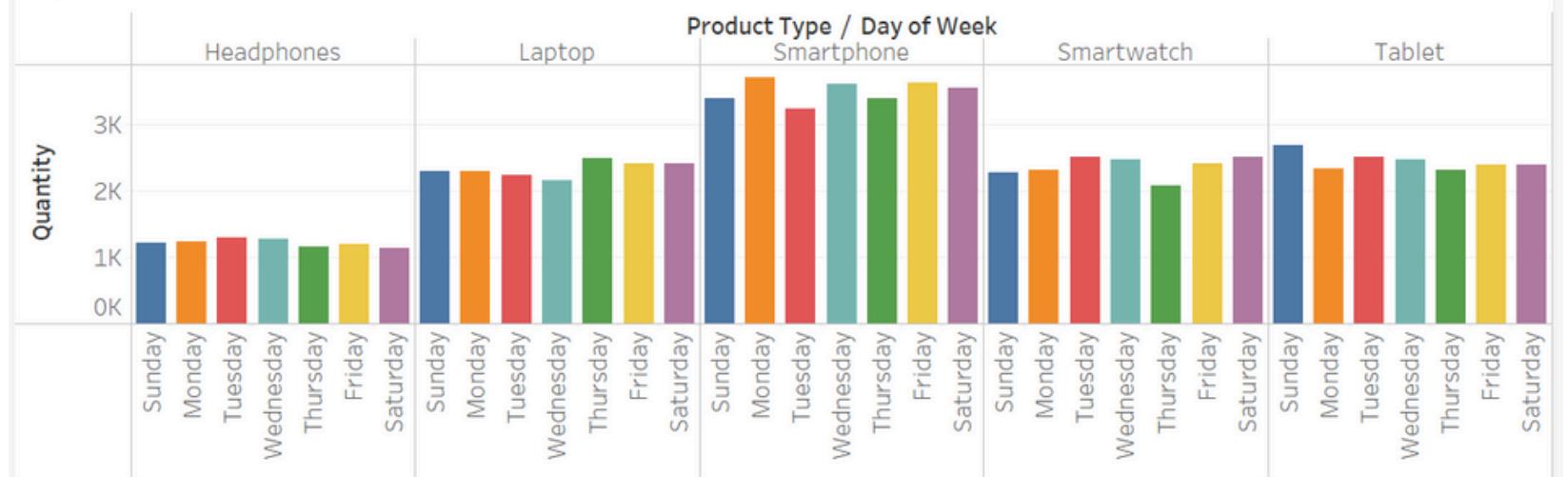


Electronic Sales Analysis

Quantity
82.821

Revenue
48.184.250

Day of Week Sales Performance Overview



Weekly Sales Performance Overview



Headphones
Laptop
Smartphone
Smartwatch
Tablet



**TERIMA
KASIH**