



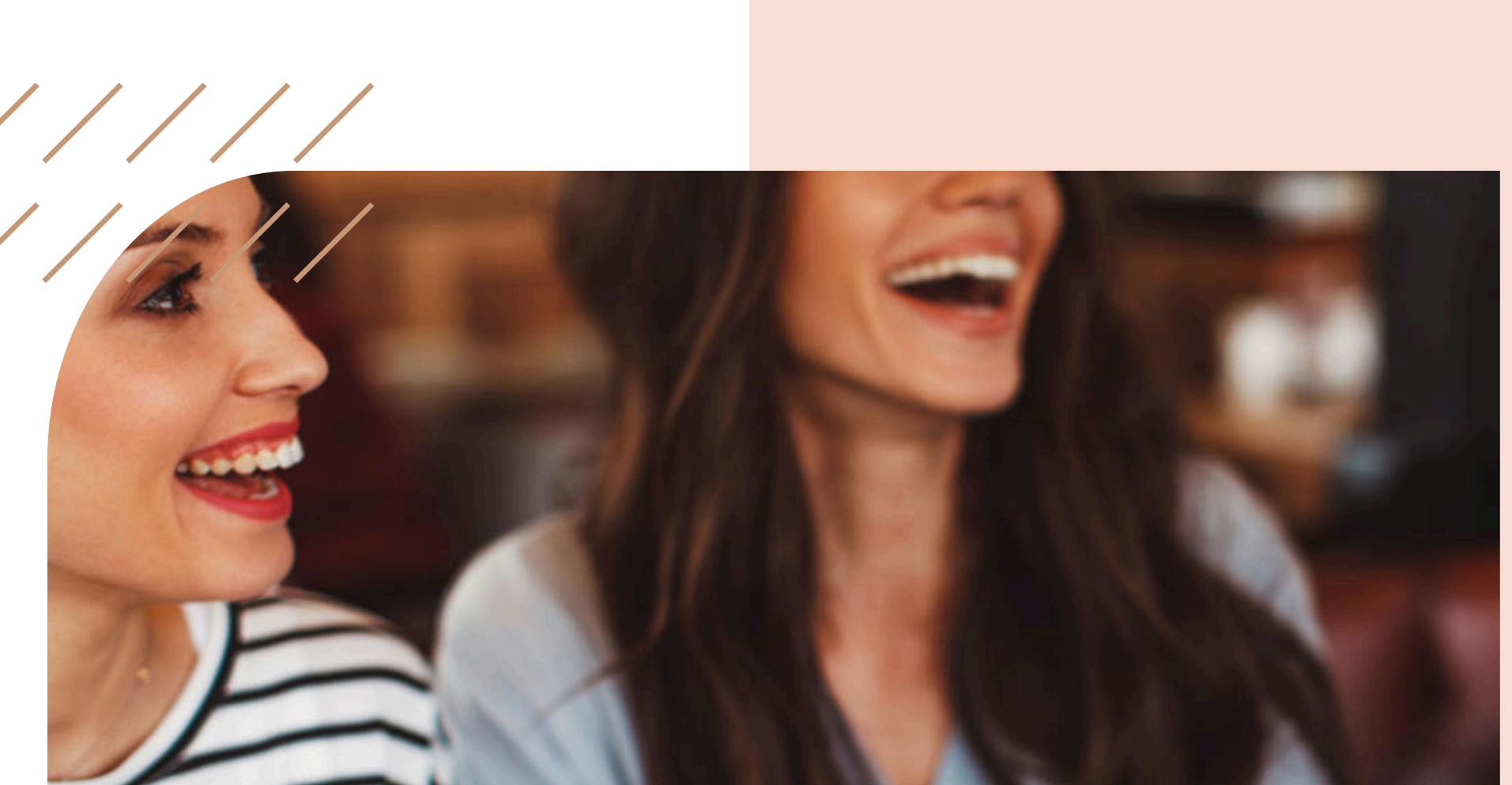
Predicting Target Sales for a Coffee Supplier in Malaysia using Machine Learning Techniques

Samantha Kok Jia Yin
TP066133

   FYP

Introduction

This project focuses on predicting target sales for a coffee supplier in Malaysia using machine learning techniques. With increasing competition in the beverage industry, businesses face challenges in understanding customer behavior, optimizing stock, and achieving sales targets. By analyzing historical invoice data, this study develops predictive models that forecast sales performance and provide insights into customer purchasing patterns. The outcome is a data-driven solution that supports better decision-making, improves forecasting accuracy, and enhances overall business growth.



Problem Statement



- **Limited Insights into Customer Purchasing Behavior**
 - Sales data collected but not analyzed effectively.
 - Lack of awareness on purchasing frequency, seasonality & demand volatility.
 - Leads to overstocking/stockouts → revenue loss & unhappy customers.
- **Lack of a Robust Predictive Sales Model**
 - Traditional models (linear regression, moving averages) too simplistic.
 - Cannot handle promotions, seasonality, or rapid market changes.
 - Results in unreliable forecasts & weak strategic planning.
- **Underutilization of Historical Invoice Data**
 - Historical invoices not fully leveraged for insights.
 - Missed opportunities in product performance tracking & customer segmentation.
 - Causes stock issues & poor customer retention strategies.
- **Inadequate Evaluation of Predictive Models**
 - Predictive models not properly validated (MAE, RMSE, R² often ignored).
 - Models become outdated over time → misleading results.
 - Reduces trust & affects sales planning, budgeting, and inventory control.

Project Aim

To leverage sales performance by developing a machine learning model to predict target sales and analyze customer behavior.

Objective

1. To analyse customer purchasing patterns and trends to enhance the accuracy of target sales predictions.
2. To develop a predictive model that forecasts sales performance and customer behavior for a coffee supplier in Malaysia.
3. To analyze historical invoice data to identify product performance for stock optimization and key customers for retention strategies.
4. To evaluate model accuracy using performance metrics for reliability and improvement.





In scope

- Compile & interpret historical data (sales, transactions, product performance).
- Analyze past sales trends (seasonality, events, customer behavior).
- Build ML predictive model to forecast sales & optimize stock management.
- Evaluate model performance using Accuracy, MAE, RMSE.

Out of scope

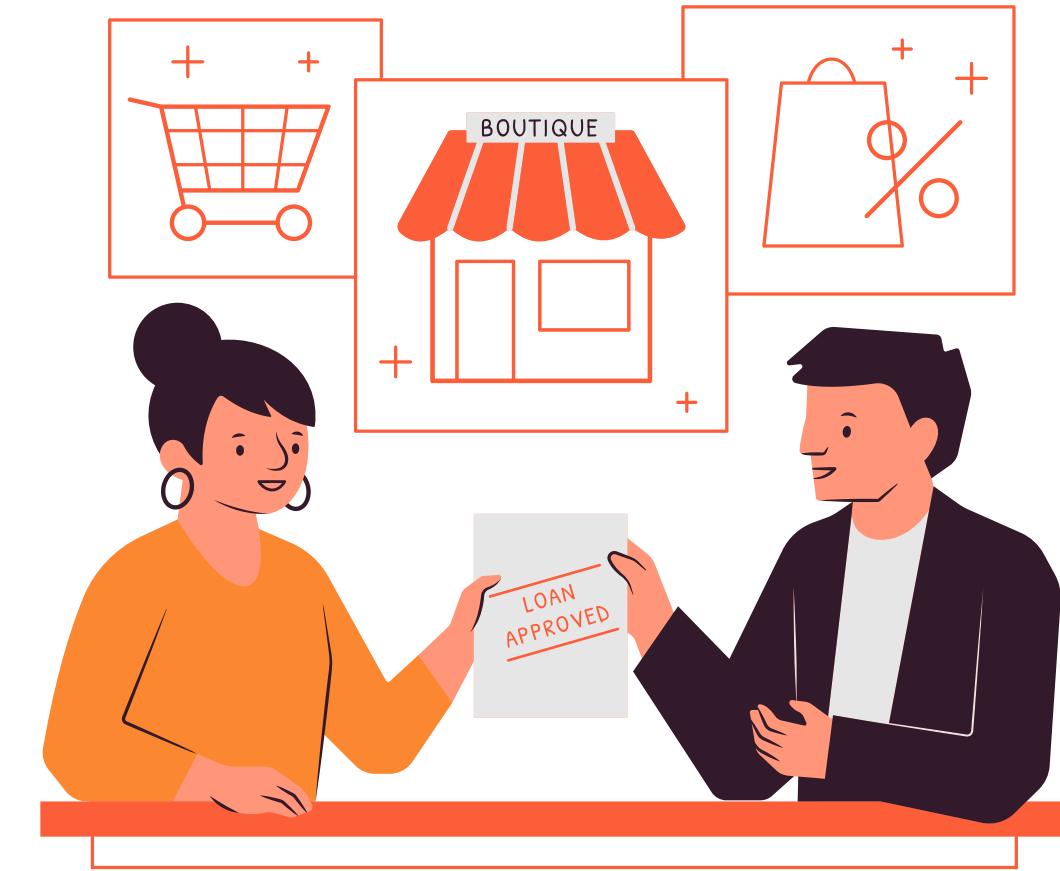
- Real-time sales forecasting & live market analysis.
- Direct implementation of sales/marketing strategies.
- External economic factors (inflation, competitor pricing, supply chain).

Target Users



Business Owners & Decision-Makers

- set realistic sales targets, improve profitability.



Sales & Marketing Teams

- gain insights on customer behavior & identify high-value customers.



Software

Python

- Main programming language for data preprocessing, model building, and evaluation.
- Libraries: Pandas, NumPy, Scikit-learn, XGBoost, Matplotlib.

PyCharm

- Integrated Development Environment (IDE) for efficient coding, debugging, and version control.

Jupyter Notebook

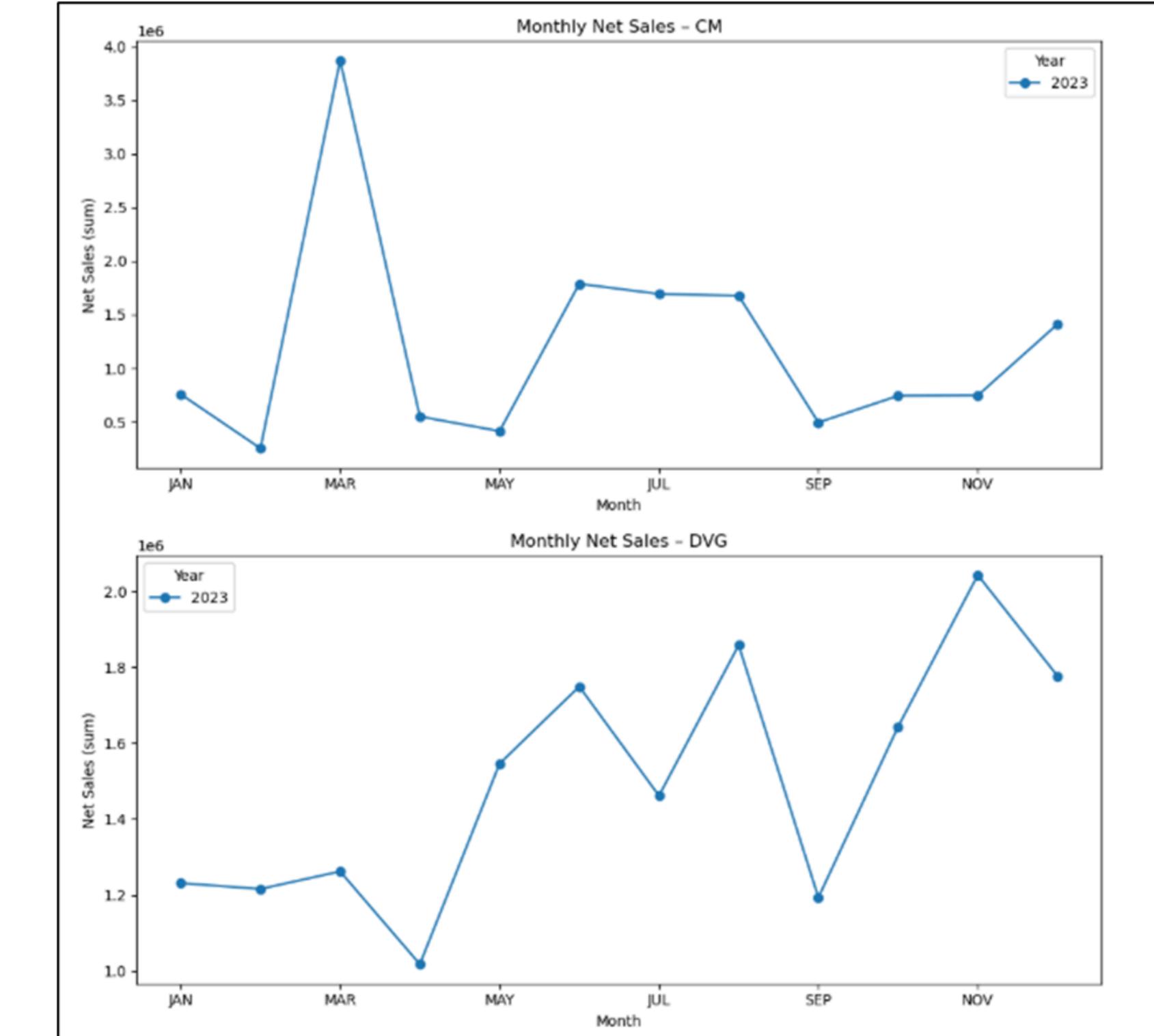
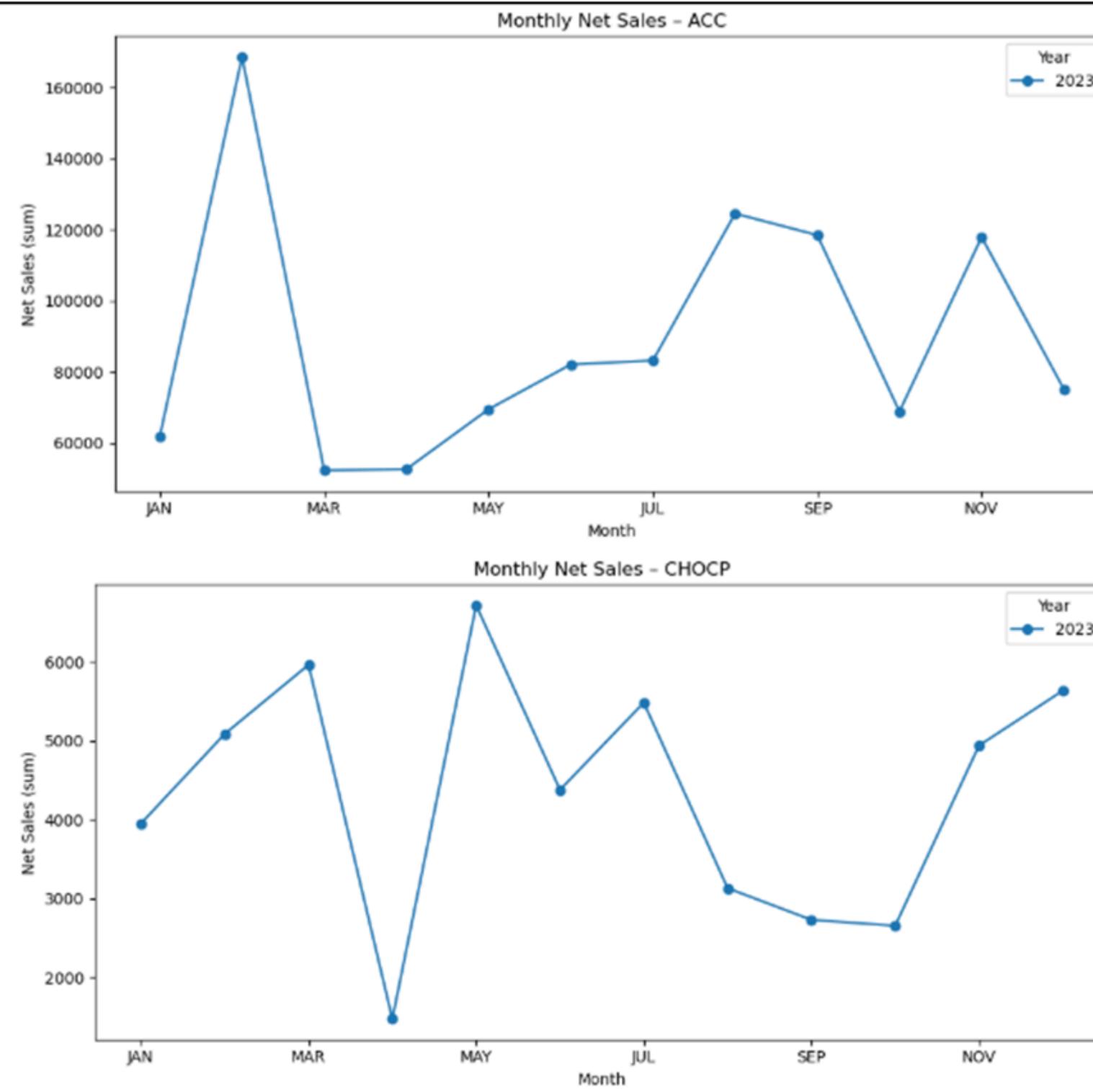
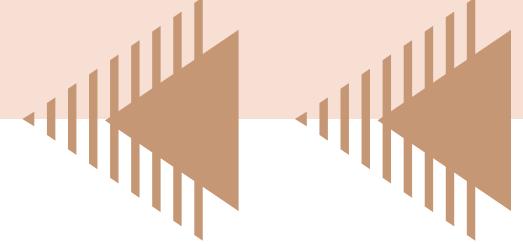
- Used for data exploration, visualization, and step-by-step model experimentation.

Streamlit

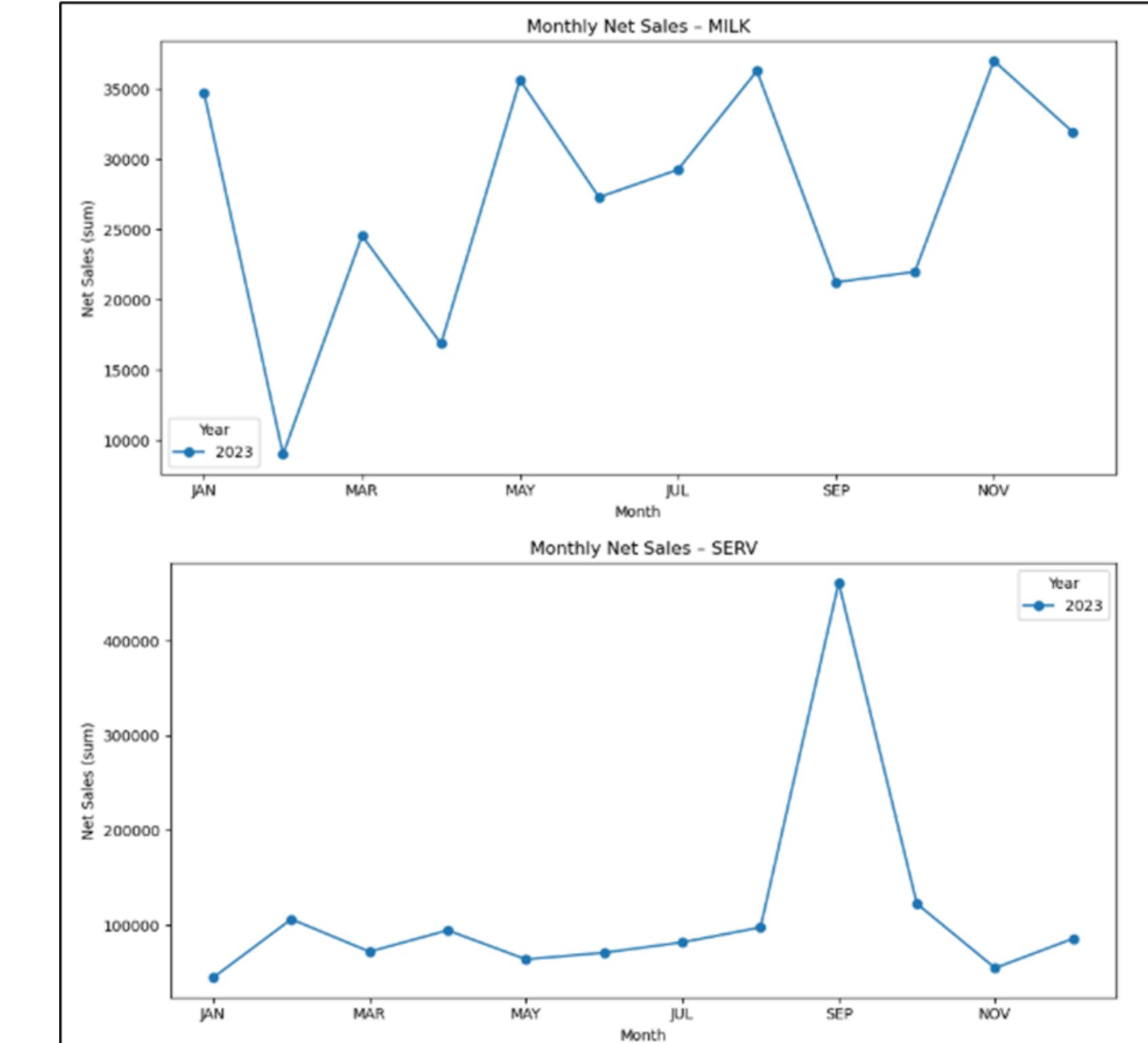
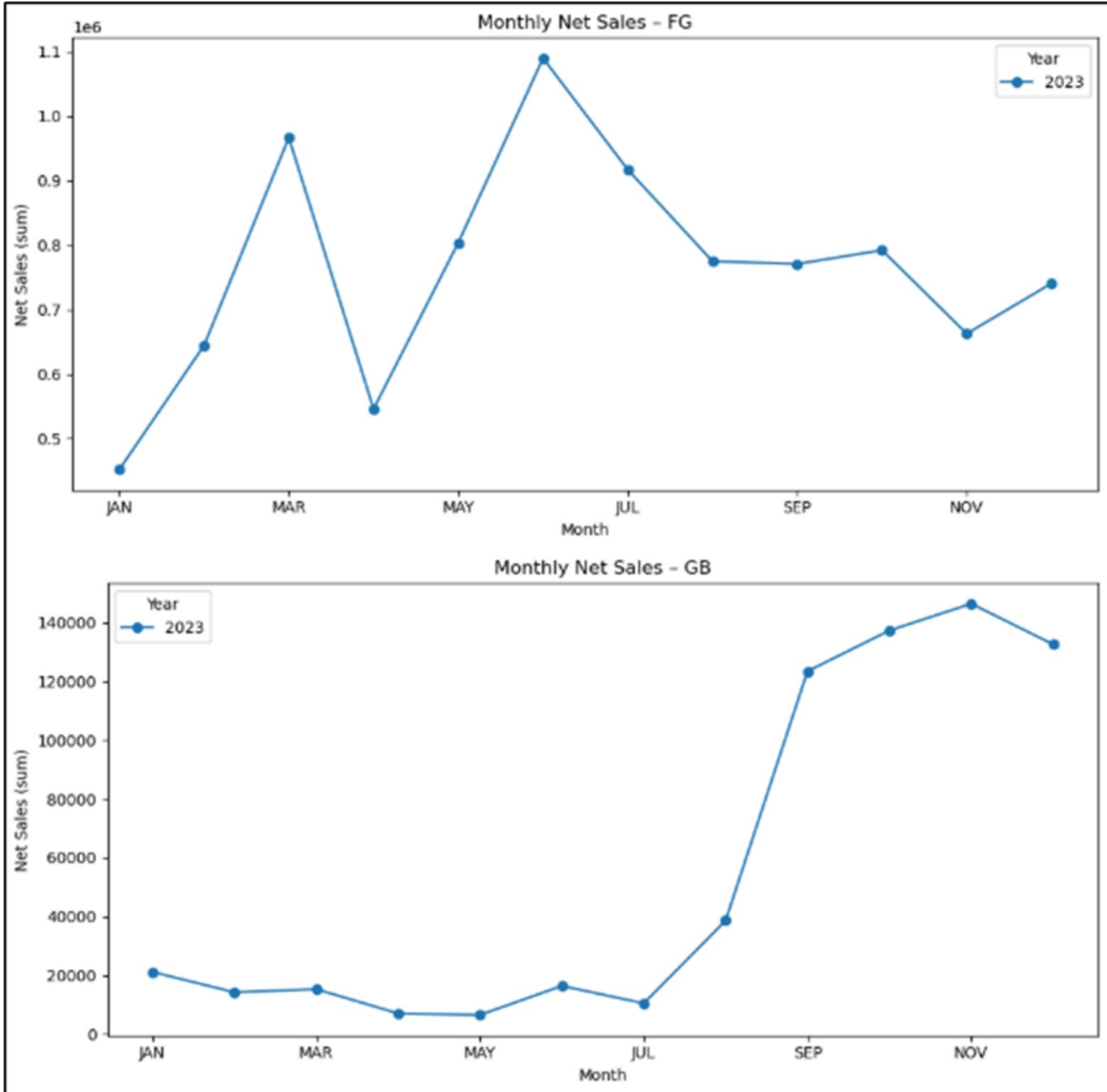
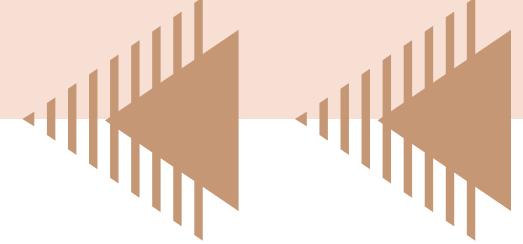
- Framework for building the interactive web dashboard to visualize and predict target sales.



Monthly net sales by Product class

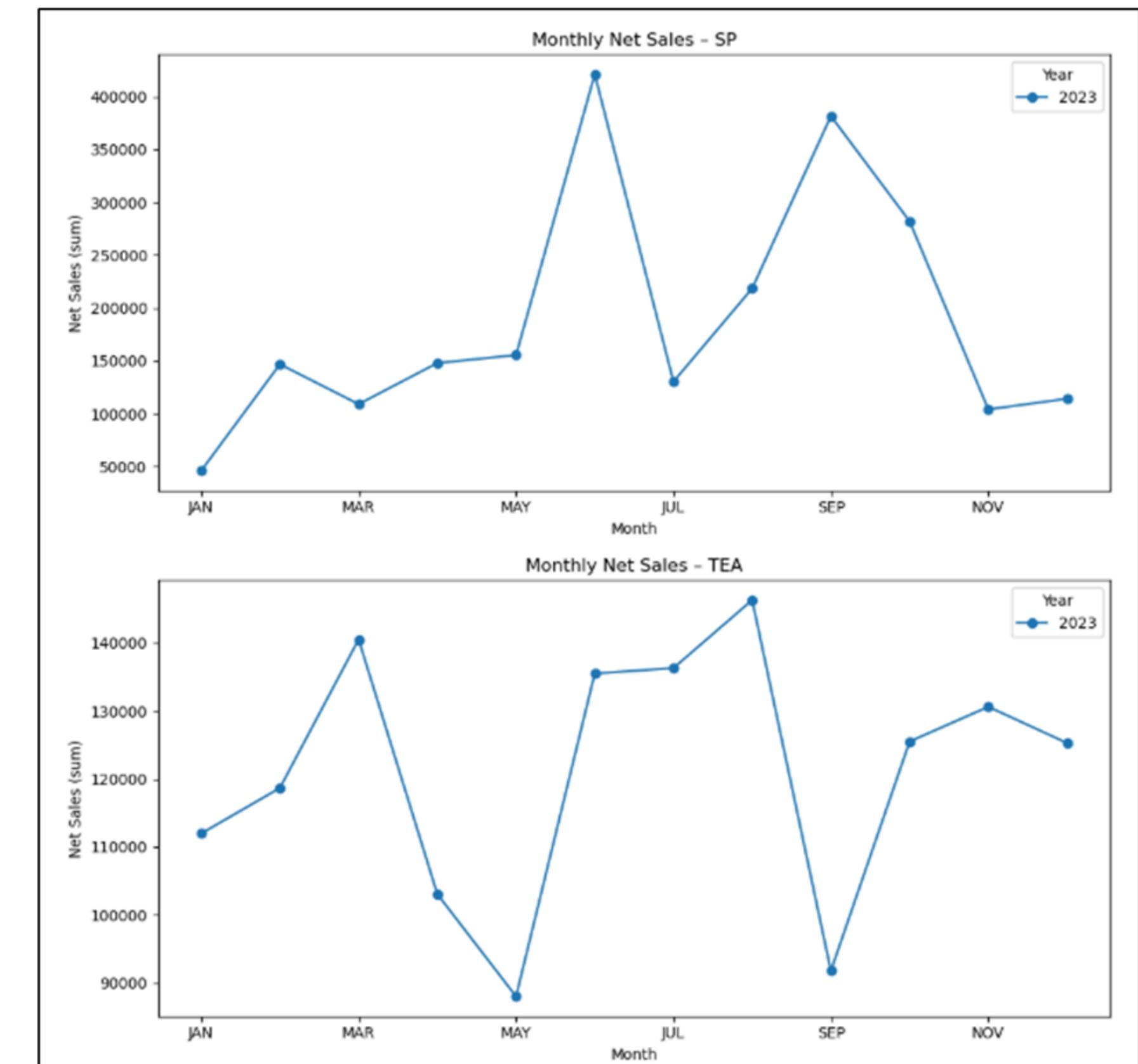


Monthly net sales by Product class

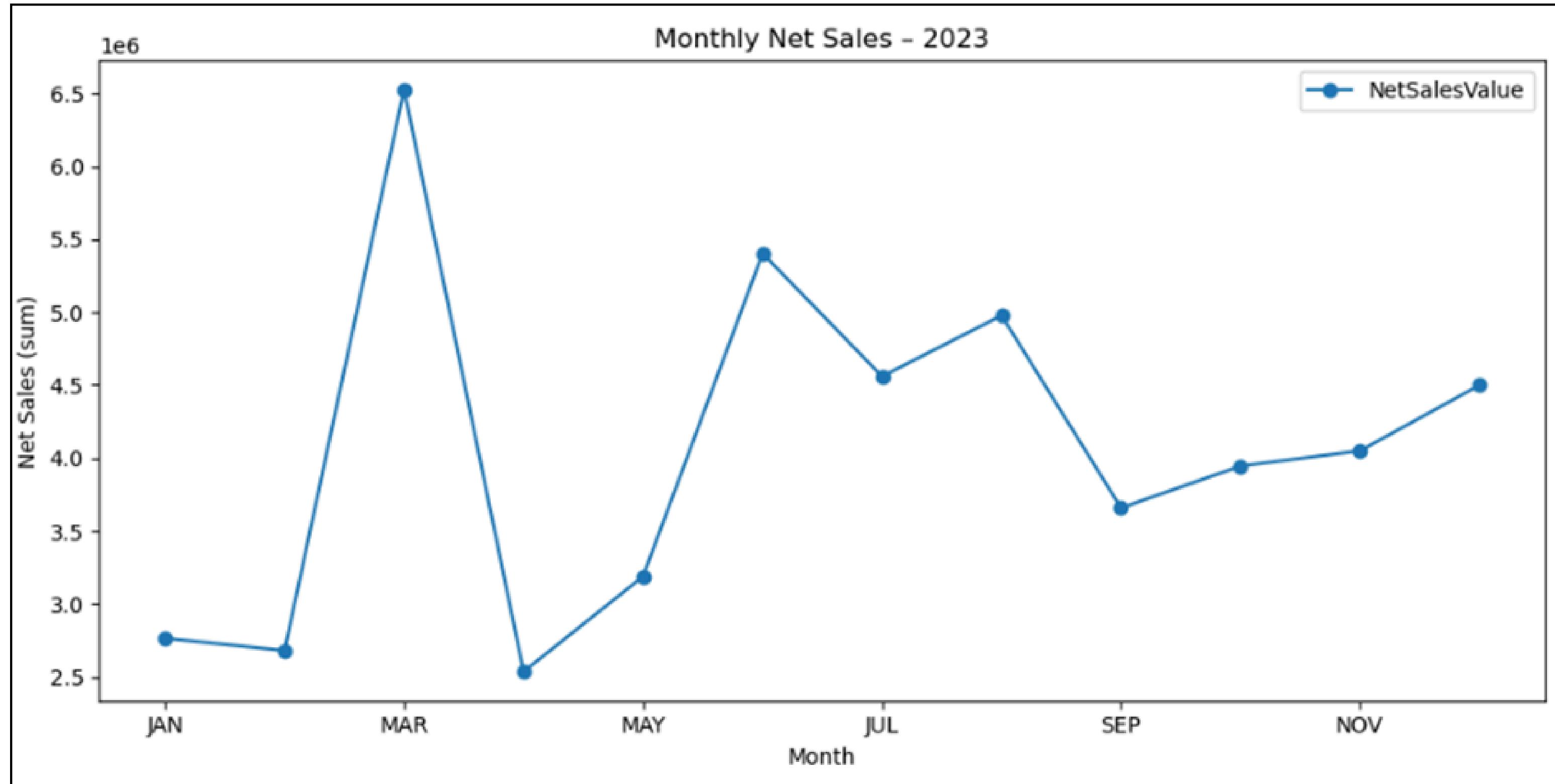
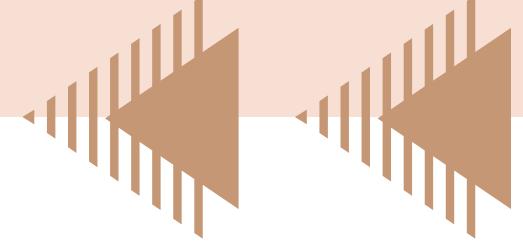


Monthly net sales by Product class

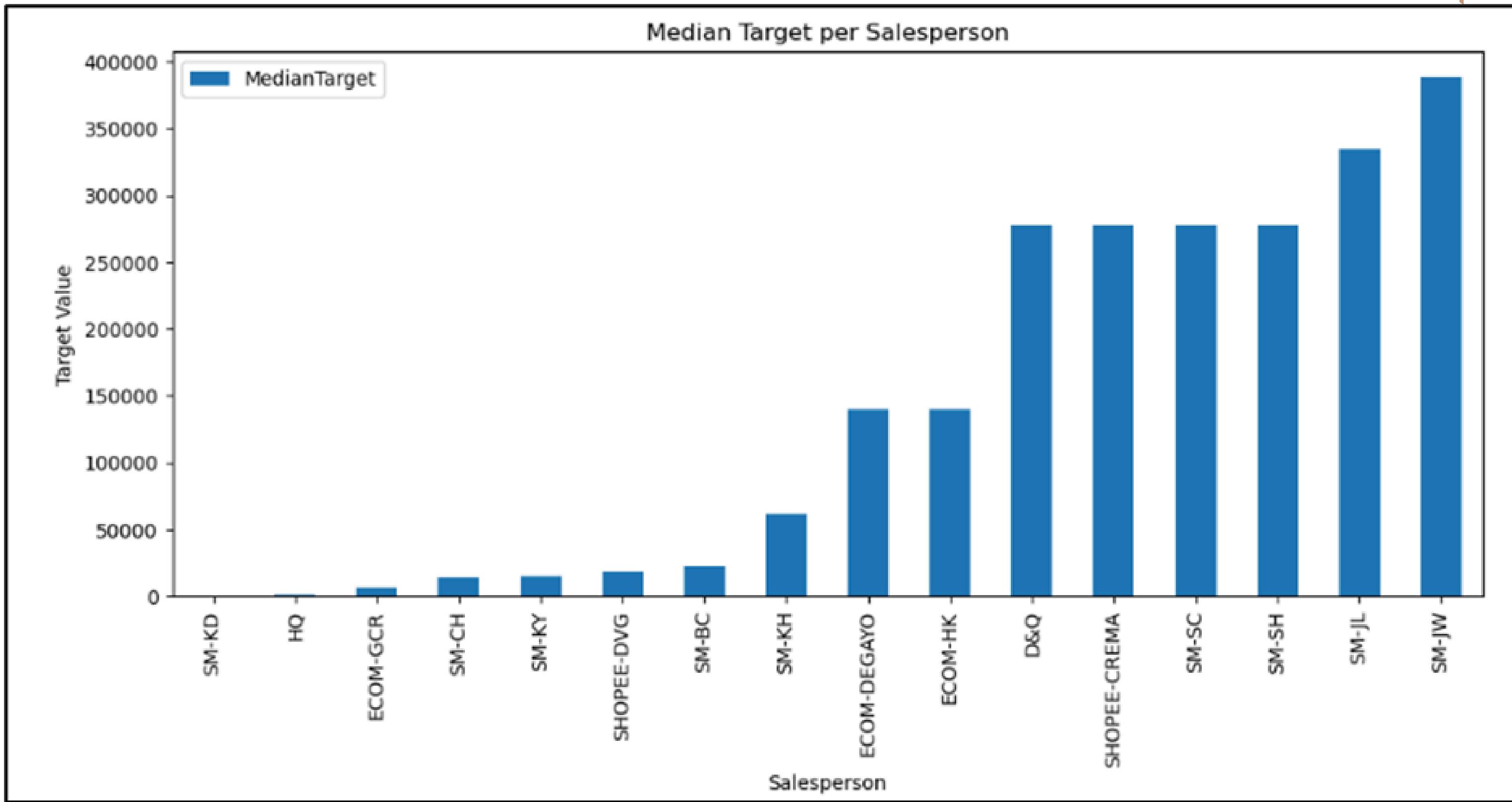
10 highest - revenue
product class



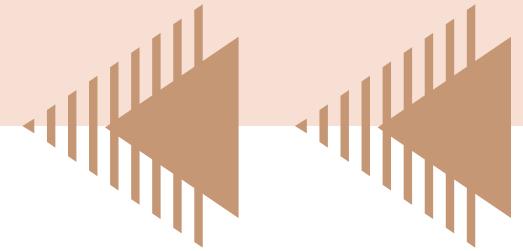
Monthly net sales 2023- Line plot



Median Target per salesperson



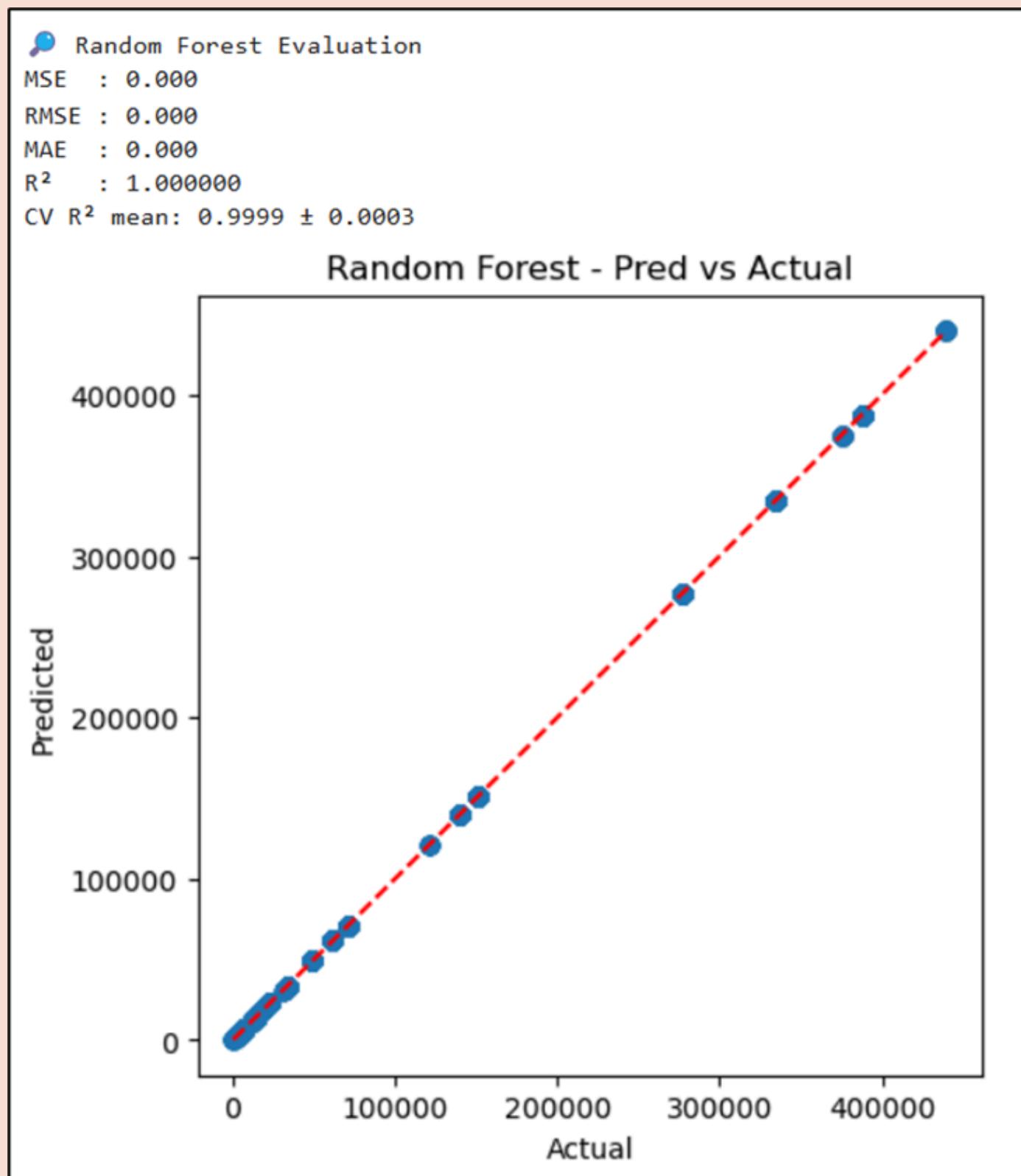
Model Evaluation



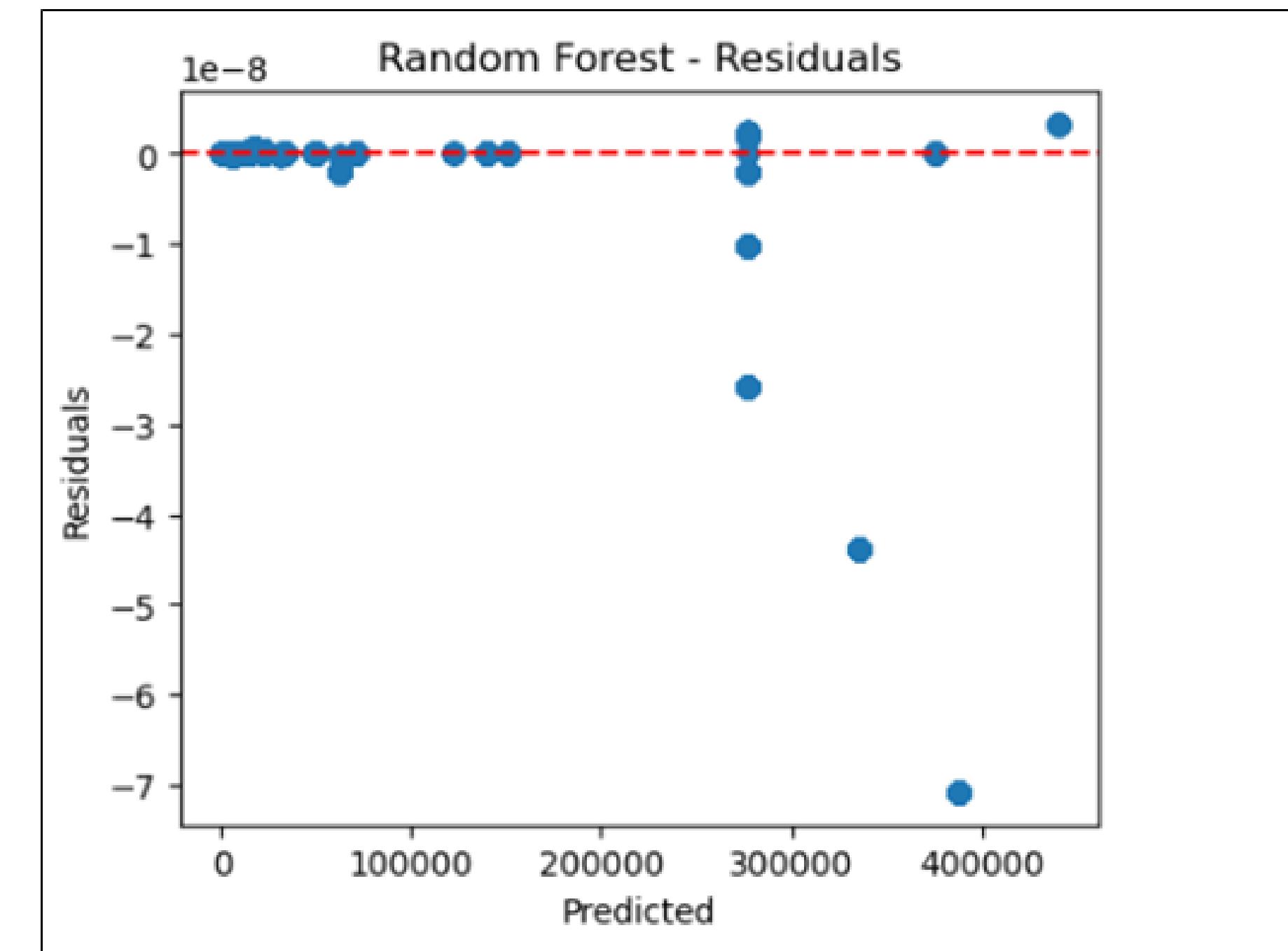
Model	RMSE	MAE	MAPE (%)	R ²
Gradient Boosting	88.61	10.02	0.08408	1.0000
KNN	105,366.10	70,316.22	387.99	0.518598
Linear Regression	41,499.87	26,474.52	248.80	0.925321
Random Forest	3.7e-08	2.47e-08	7.49e-12	1.0000
XG Boost	54.23	1.23	0.08166	1.0000

- **Gradient Boosting** → Good accuracy, RMSE = 88.61.
- **KNN** → Very poor performance, high RMSE & MAPE, R² = 0.52.
- **Linear Regression** → Moderate accuracy, but high error values.
- **Random Forest** → Extremely accurate (near perfect, R² = 1.0).
- **XGBoost** → Best balance of accuracy & robustness, low RMSE (54.23), MAE (1.23), R² = 1.0.

Random Forest Graphs

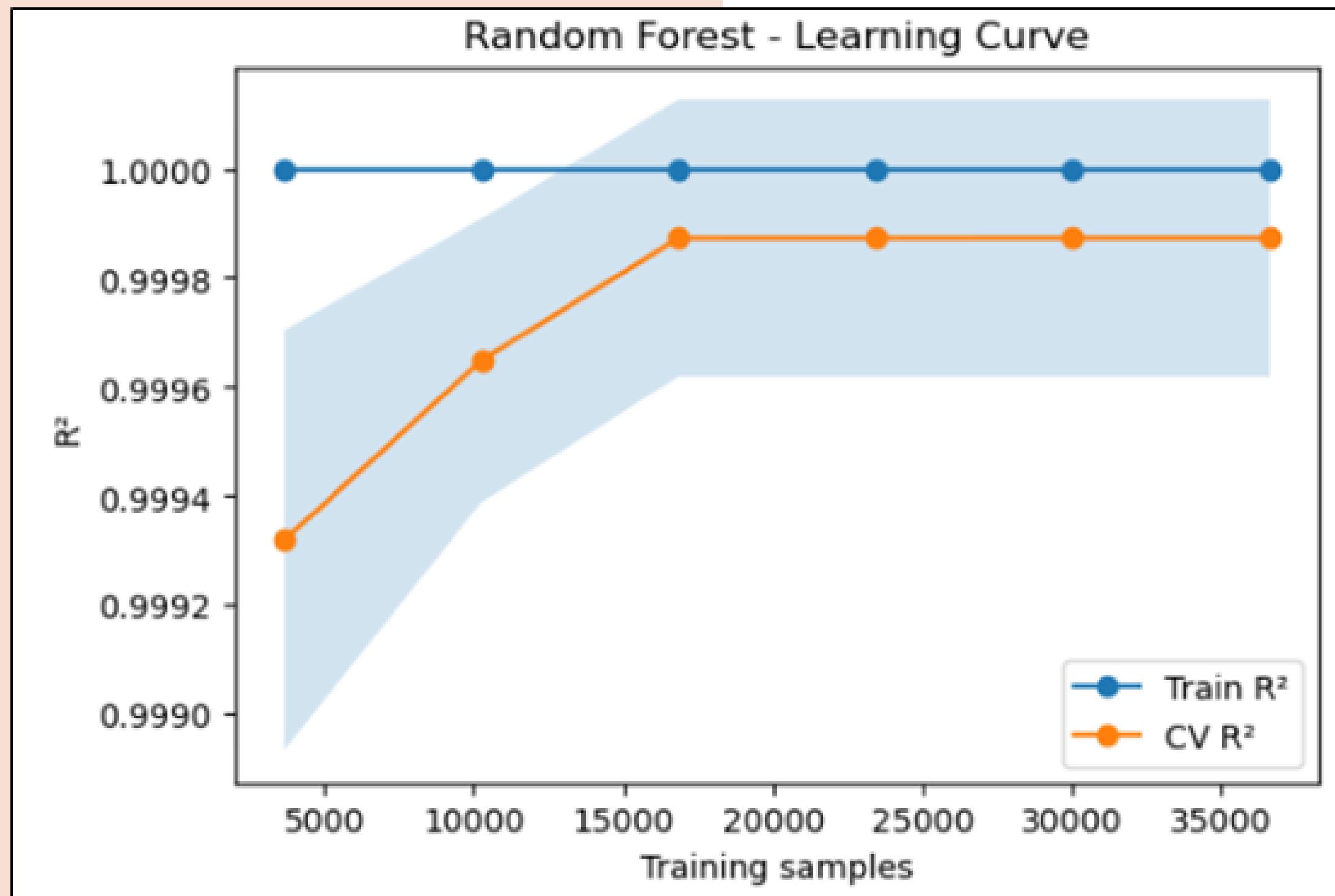


- Points perfectly align on the red line → very strong fit ($R^2 = 1.0$).
- Indicates excellent model learning.



- Most residuals near zero, but a few deviations at high values.
- Suggests possible overfitting — model may memorize training data.

Random Forest Learning Curve



- Training R² stays at 1.0, validation R² slightly lower.
- The gap implies overfitting, meaning model performance may drop on new data.

XGBoost Graphs

XGBoost Evaluation

MSE : 2941.184

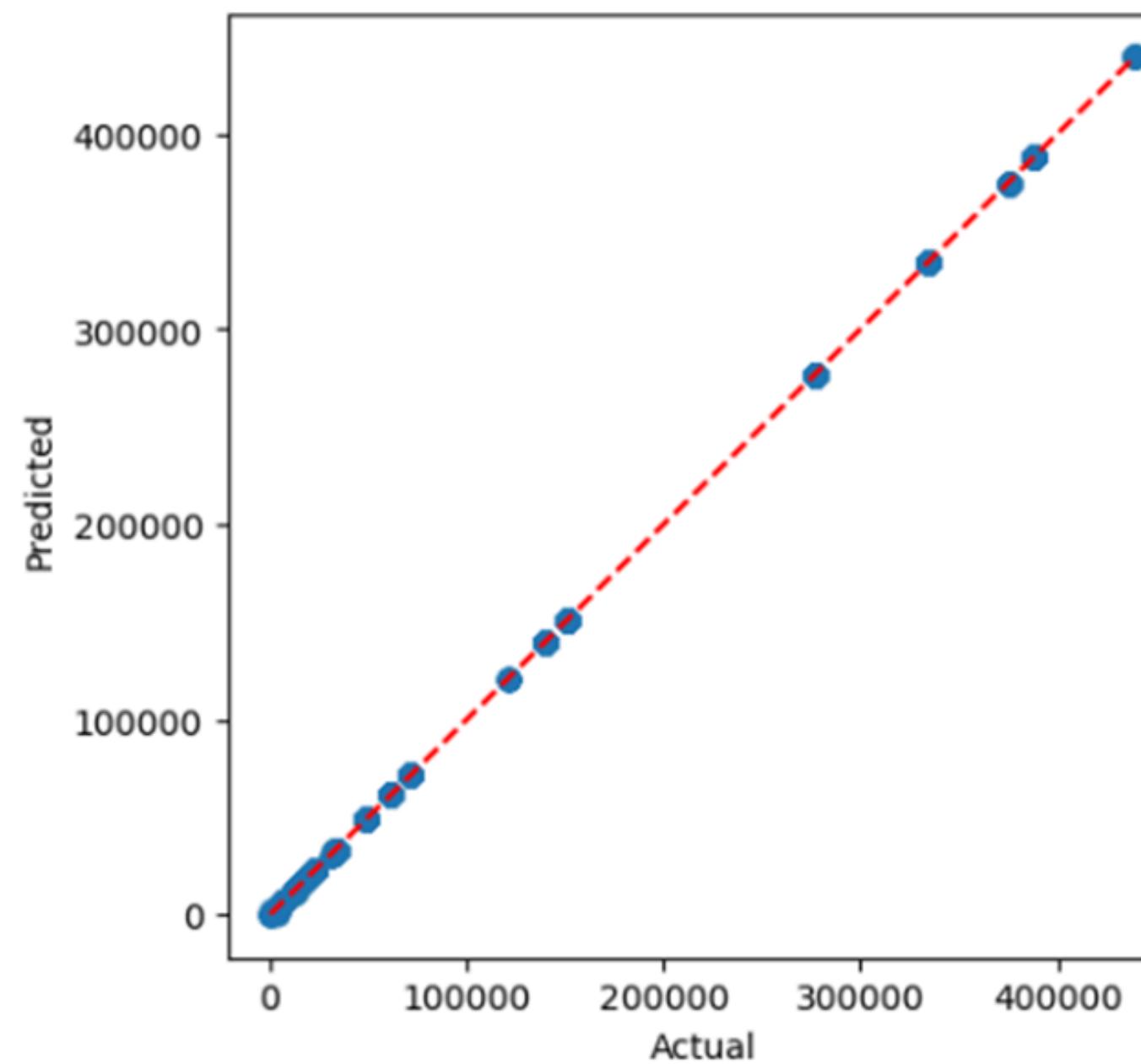
RMSE : 54.233

MAE : 1.233

R² : 1.000000

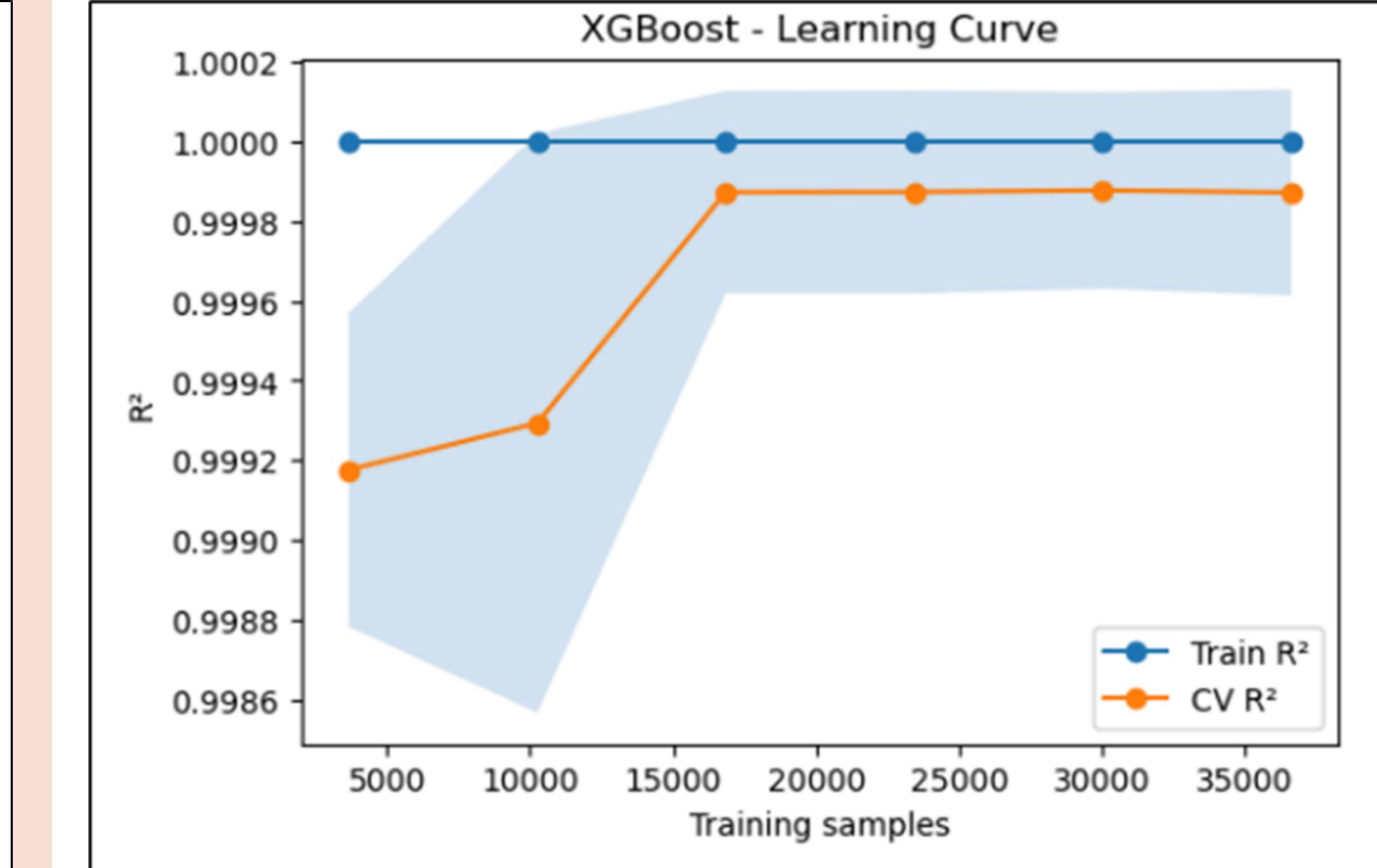
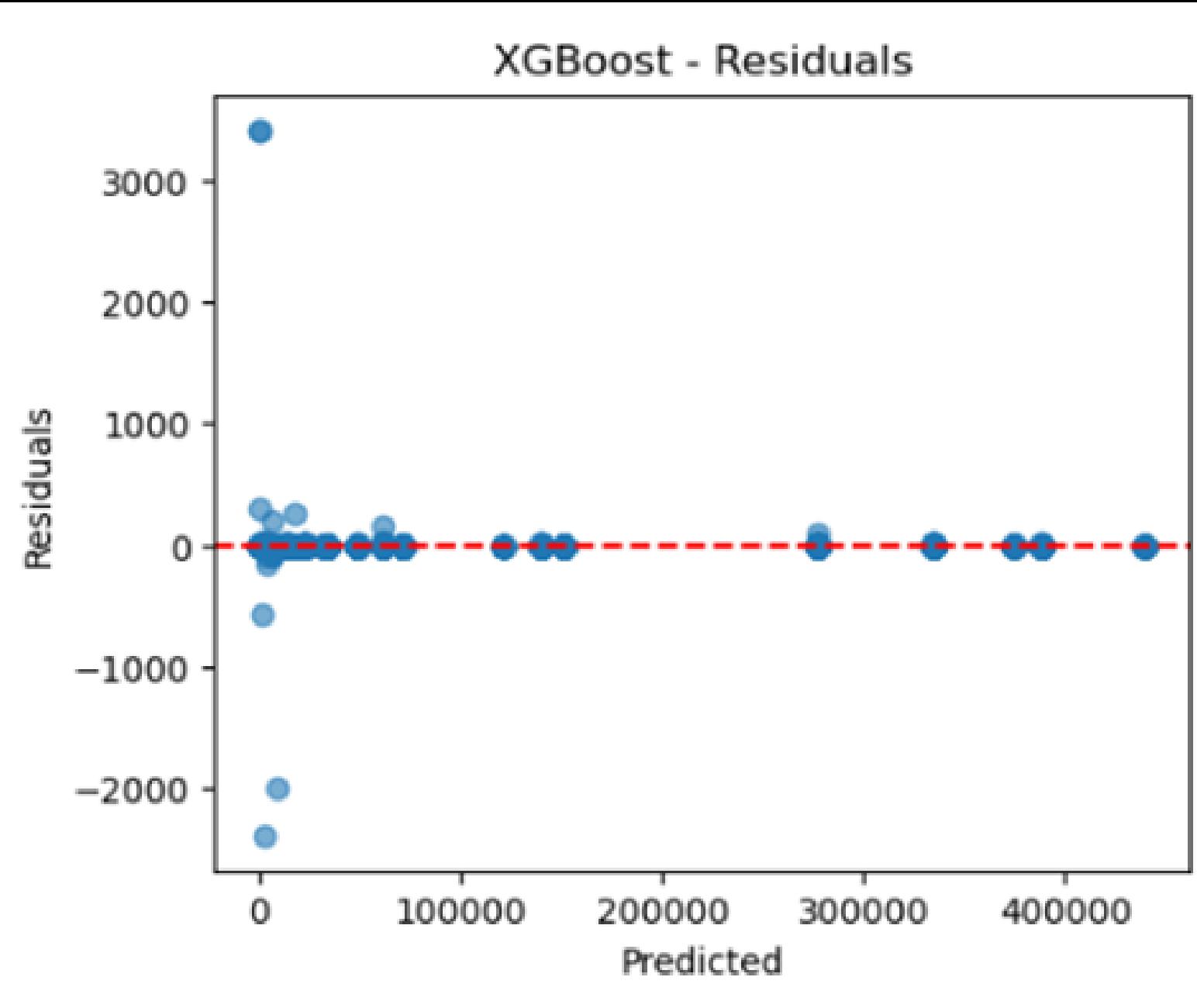
CV R² mean: 0.9999 ± 0.0003

XGBoost - Pred vs Actual



- **R² = 1.0, indicating a perfect fit between predicted and actual values.**
- **RMSE = 54.23, MAE = 1.23 → very small error margin.**
- **Points align almost perfectly on the red line → predictions closely match actual sales.**
- **Shows high accuracy and stability across all data points.**
- **Confirms that XGBoost effectively captures sales patterns for reliable forecasting**

XGBoost Learning Curve



- Residuals are small and evenly distributed around zero.
- Shows better error stability and less overfitting compared to Random Forest.

- Both training and validation R^2 lines converge near 1.0.
- Indicates good generalization and consistent performance as data increases.

Model deployment

Target Input Dashboard

Enter / Update a Salesperson Target

Pick from data Add new salesperson

Salesperson (code)

D&Q

Salesperson name (reference)

Detected last-year total: 153,968

Suggested target (info)

192460 - + cap

Your Target 2024

192460 - +

Add / Update Remove this salesperson

Target Sales Plan (Live)

	Code	Name	Total Net Sales 2023	Your Target 2024
0	D&Q		153,968	192,460
1	ECOM-HK		610	50,000

Total Net Sales 2023: 154,578

Target Sales 2024: 242,460

Achievability %: 100.0%

↑ +56.9% vs LY

↑ 242,460 reachable out of 242,460 target

Headcount (entered): 2

Download plan (CSV)

Model deployment

Target Input Dashboard

Enter / Update a Salesperson Target

Pick from data Add new salesperson

Salesperson code (required)

SM-RJ

Salesperson name (reference)

Rebecca

Total Net Sales 2023

3722 - +

Your Target 2024

5893 - +

Add / Update (new)

Remove this salesperson

Target Sales Plan (Live)

	Code	Name	Total Net Sales 2023	Your Target 2024
0	D&Q		153,968	192,460
1	ECOM-HK		610	50,000
2	SM-RJ	Rebecca	3,722	5,893

Total Net Sales 2023: 158,300

Target Sales 2024: 248,353

Achievability %: 99.9%

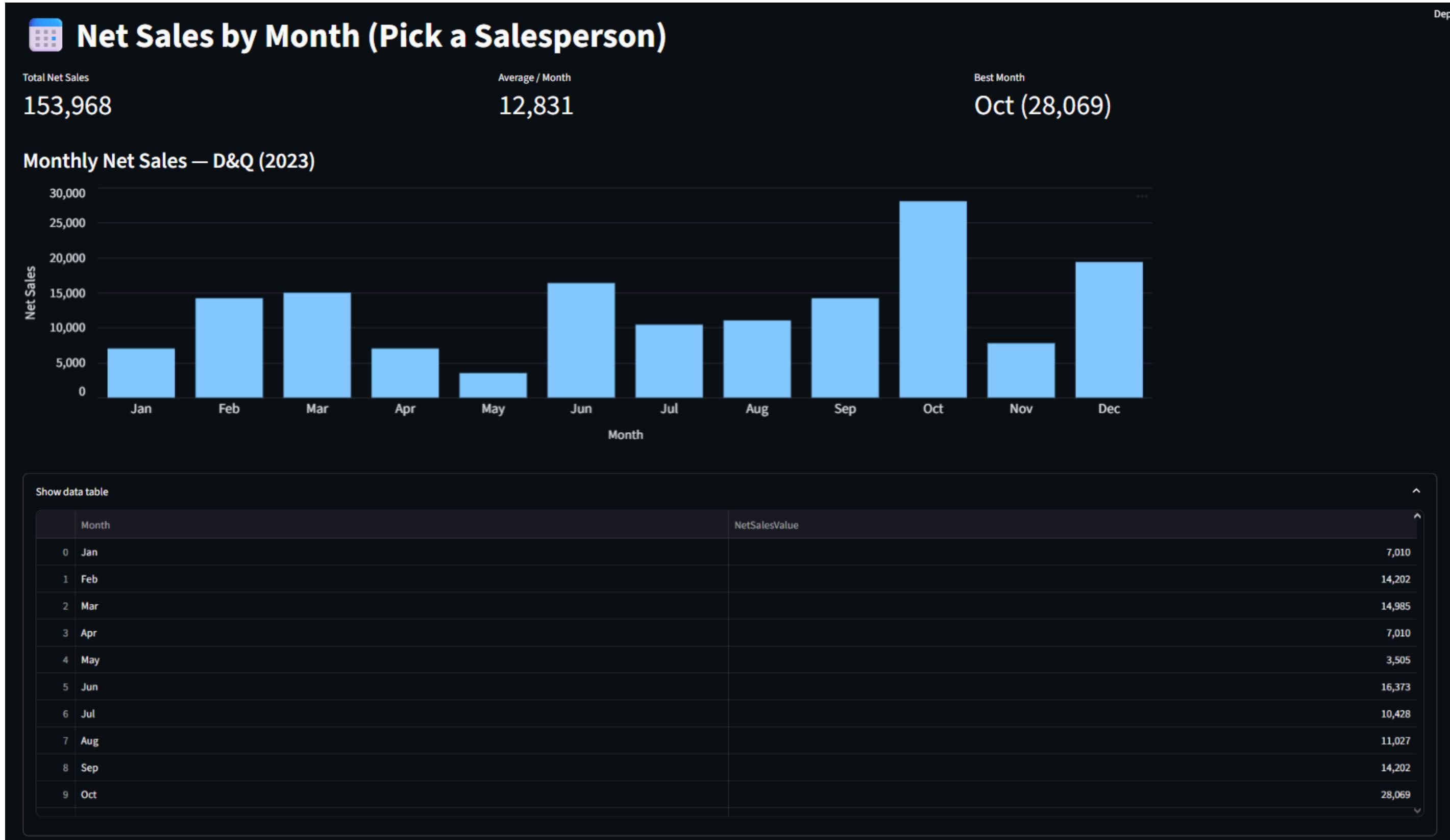
↑ +56.9% vs LY

↑ 248,043 reachable out of 248,353 target

Download plan (CSV)



Dashboard Visualization





Dashboard Visualization

Predicted Target Sales for Next Year (Annual KPI)

Guardrail applied (cap): floor=123,174, cap=192,460 (233,945 → 192,460)

Predicted Target for D&Q in 2024

192,460.12

Show short list for all salespeople

	Salesperson	Total Net Sales 2023	Predicted Target 2024	Bound	Seen in training
10	SM-JW	24,784,174	22,305,756.38	floor	<input checked="" type="checkbox"/>
9	SM-JL	7,378,523	6,640,670.57	floor	<input checked="" type="checkbox"/>
15	SM-SH	7,067,541	6,360,787.30	floor	<input checked="" type="checkbox"/>
14	SM-SC	6,274,811	5,647,330.22	floor	<input checked="" type="checkbox"/>
12	SM-KH	832,937	740,082.76	-	<input checked="" type="checkbox"/>
7	SM-BC	795,394	636,315.21	floor	<input checked="" type="checkbox"/>
8	SM-CH	457,999	366,398.89	floor	<input checked="" type="checkbox"/>
13	SM-KY	446,888	357,510.13	floor	<input checked="" type="checkbox"/>
6	SHOPEE-DVG	220,965	233,734.17	-	<input checked="" type="checkbox"/>
0	D&Q	153,968	192,460.12	cap	<input type="checkbox"/>

Download predictions (CSV)



Product Performance Visualization

📦 Product Performance & 👤 Customer Overview

A) Product priorities by sales (ABC)

	ProductClass	Total net sales	Cumulative % of sales	Priority band
0	DVG	17,999,850	36.89	A (high)
1	CM	14,404,562	66.41	A (high)
2	FG	9,159,231	85.17	B (medium)
3	SP	2,256,888	89.80	B (medium)
4	TEA	1,453,589	92.78	B (medium)
5	SERV	1,356,014	95.56	C (low)
6	ACC	1,075,717	97.76	C (low)
7	GB	669,274	99.13	C (low)
8	MILK	325,635	99.80	C (low)
9	CHOCP	52,147	99.91	C (low)





Customer Class Visualization

B) Customer overview (top classes by sales & orders)

Class references: CTRD = trade customer, CNTRD = non-trade customer, SD = sub distribution, KA = key account, HRC = independent.

Top class by total sales

CTRD-KA

↑ 34,482,622

Show top N classes

1

Top class by number of orders

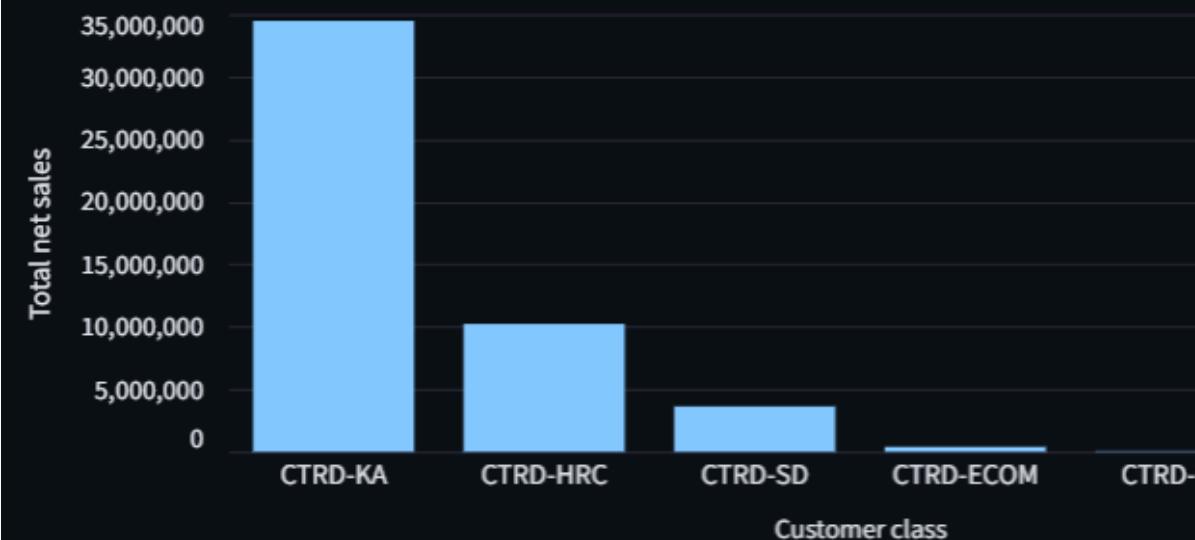
CTRD-HRC

↑ 6,857

6

6

	CustomerClass	Meaning	Total net sales	Number of orders	Top salesperson	Top salesperson sales
3	CTRD-KA	Trade customer Key account	34,482,622	6,077	SM-JW	24,022,729
2	CTRD-HRC	Trade customer Independent	10,229,324	6,857	SM-JL	4,310,659
5	CTRD-SD	Trade customer Sub distribution	3,632,667	1,195	SM-JL	3,020,951
1	CTRD-ECOM	Trade customer E-commerce	399,632	2,755	SHOPEE-DVG	220,965
4	CTRD-OS	Trade customer Other segment	53,420	18	SM-JL	46,913
0	CNTRD	Non-trade customer	450	3	HQ	450



[Download customer class summary \(CSV\)](#)



Salesperson Visualization

Sales Target Plan – Overview

How to use this page

1. Choose the base year on the left (plan is base_year + 1).
2. Review KPIs: company total, planned total (Δ vs LY), and reps at floor/cap.
3. See the Top N chart and the Full Plan table below.

Method: We predict each rep's 12 monthly values with the model, sum them, then apply guardrails

($\geq 5M$: floor/cap; 1–5M: floor/cap; 100k–1M: floor/cap; <100k: absolute min) and an optional company growth goal that distributes extra only to reps with cap headroom.

Total Net Sales (base year)

48,798,114

Planned Target 2024

51,238,020

↑ +5.0%

Reps at floor

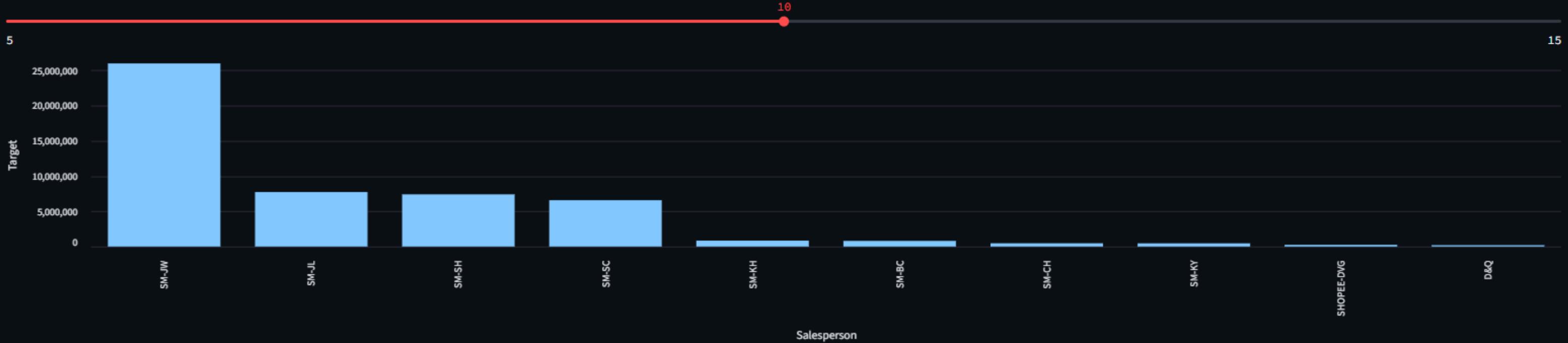
11 / 16

Reps at cap

4 / 16

Guardrails applied after model prediction: $\geq 5M$ 100–114%, 1–5M 95–120%, 100k–1M 90–125%, <100k min 50,000. Company growth goal: 5.0%

Show top N by target





Thank You

For your attention



FYP