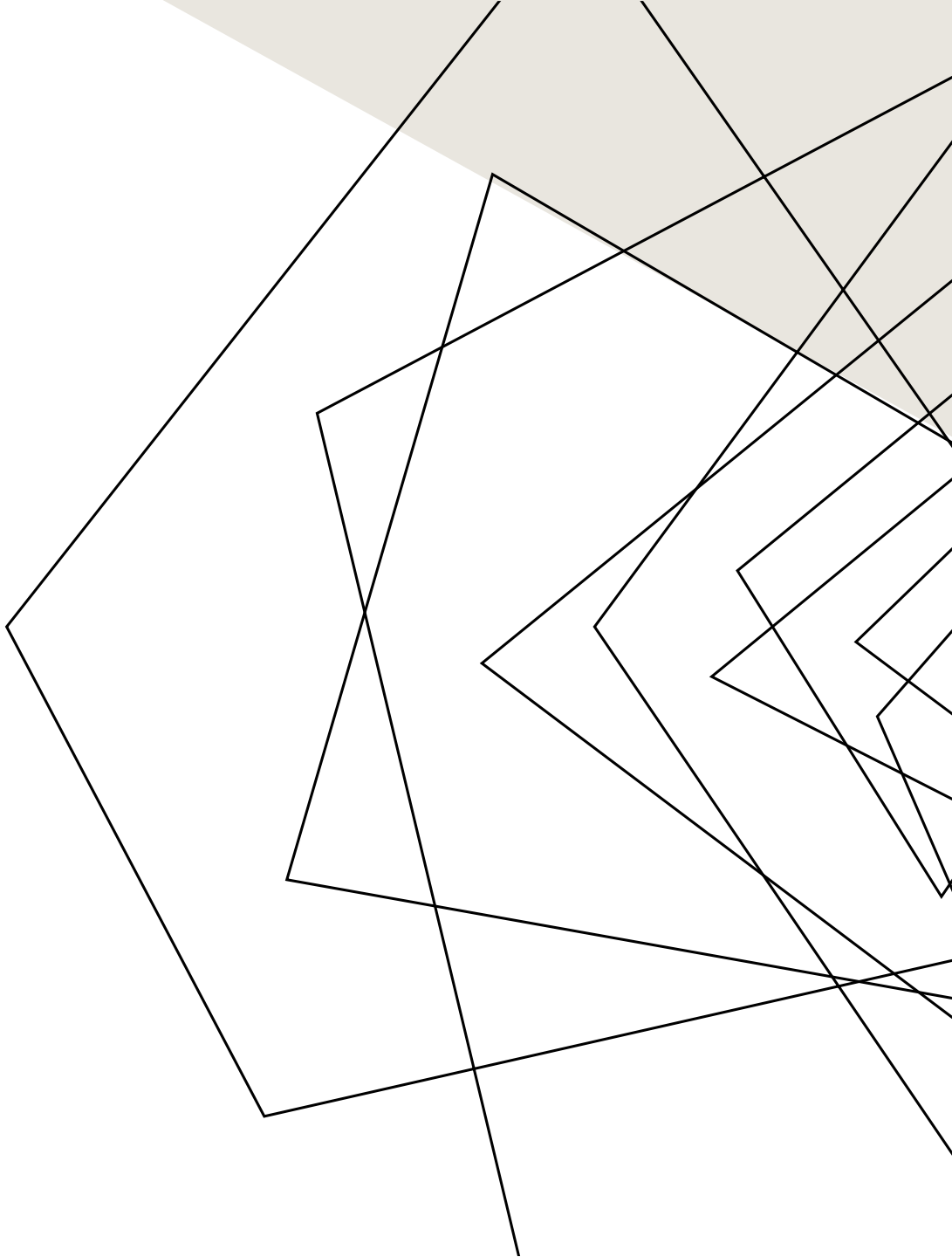
E-COMMERCE SALES ANALYSIS & INSIGHTS

USING PYTHON FOR DATA EXPLORATION & MODELING

KATHERINE PONCE | FEB 2025

INTRODUCTION

- **Objective:** Analyze sales trends, customer behavior, and product performance using Python.
- **Key Steps:** Data cleaning, exploratory analysis, correlation testing, and predictive modeling.
- **Outcome:** Identify key sales drivers and provide actionable business recommendations.



DATASET OVERVIEW

- **Total Records:** 1000
- **Columns Included:**
  - Customer ID, Order ID, Product, Sales Amount, Order Date
  - Region, Payment Method, Discount, Customer Segment
- **Observations:**
  - The dataset contains **structured transaction data**.
  - Missing values and duplicates were handled appropriately.

DATA CLEANING & PREPARATION

Steps Taken:

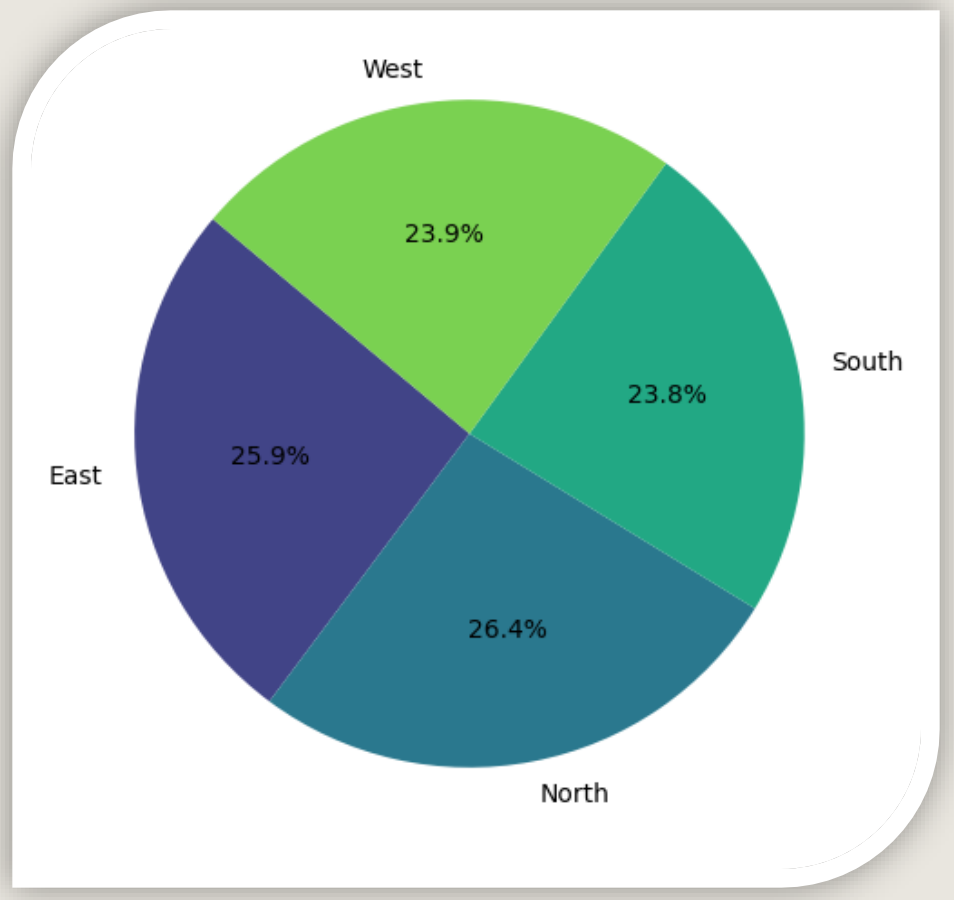
- Converted `Order_Date` to **datetime format**.
- Handled missing values by replacing them with **0 or median values**.
- Removed duplicate rows (**X duplicates found and removed**).
- One-hot encoding applied for categorical variables (Region, Payment Method).

EXPLORATORY DATA ANALYSIS (EDA)

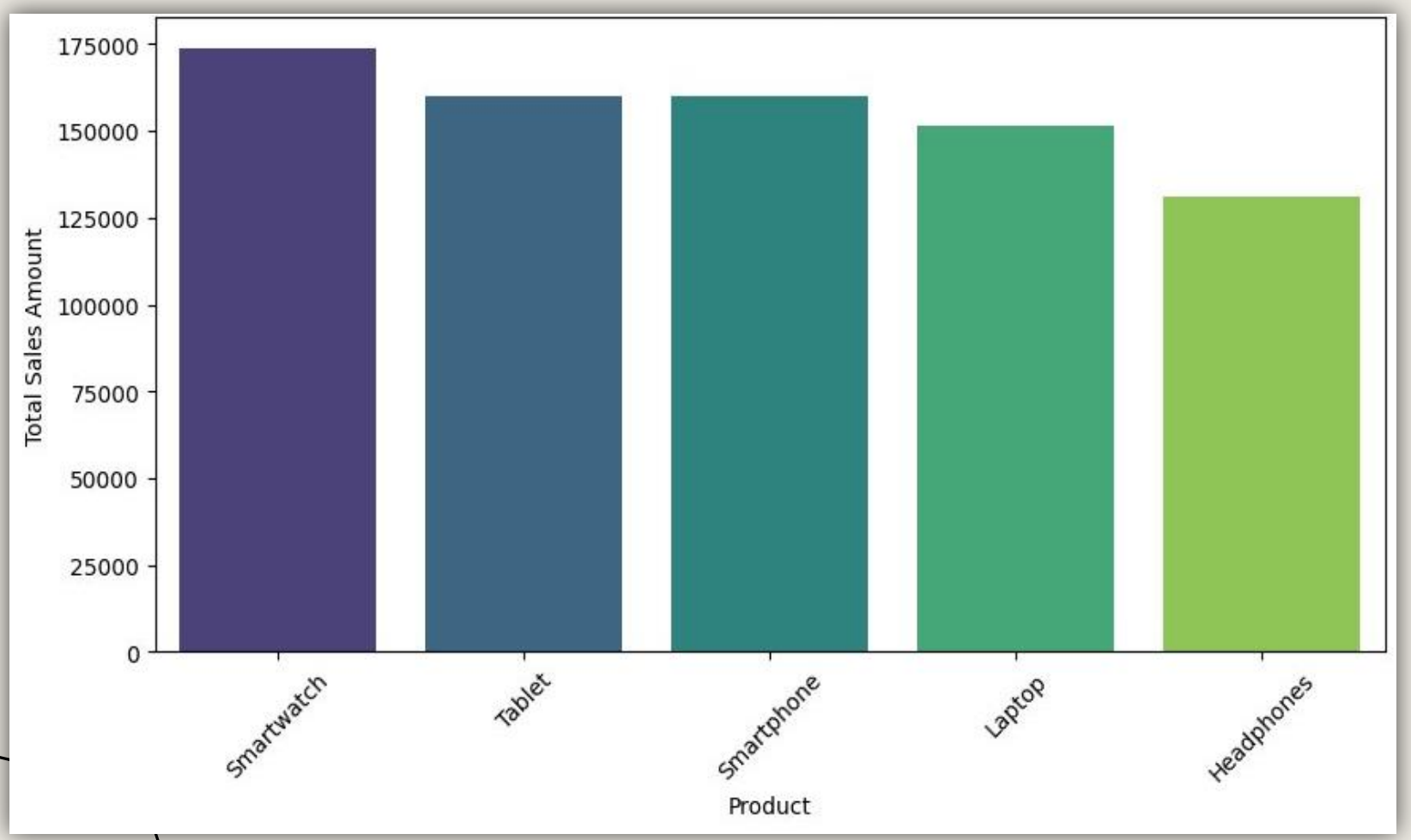
Revenue and Best- Selling Products

TOTAL REVENUE BY REGION

- **Regions contributing highest sales** were analyzed.
- **Insights:** North Region contributes the most revenue.



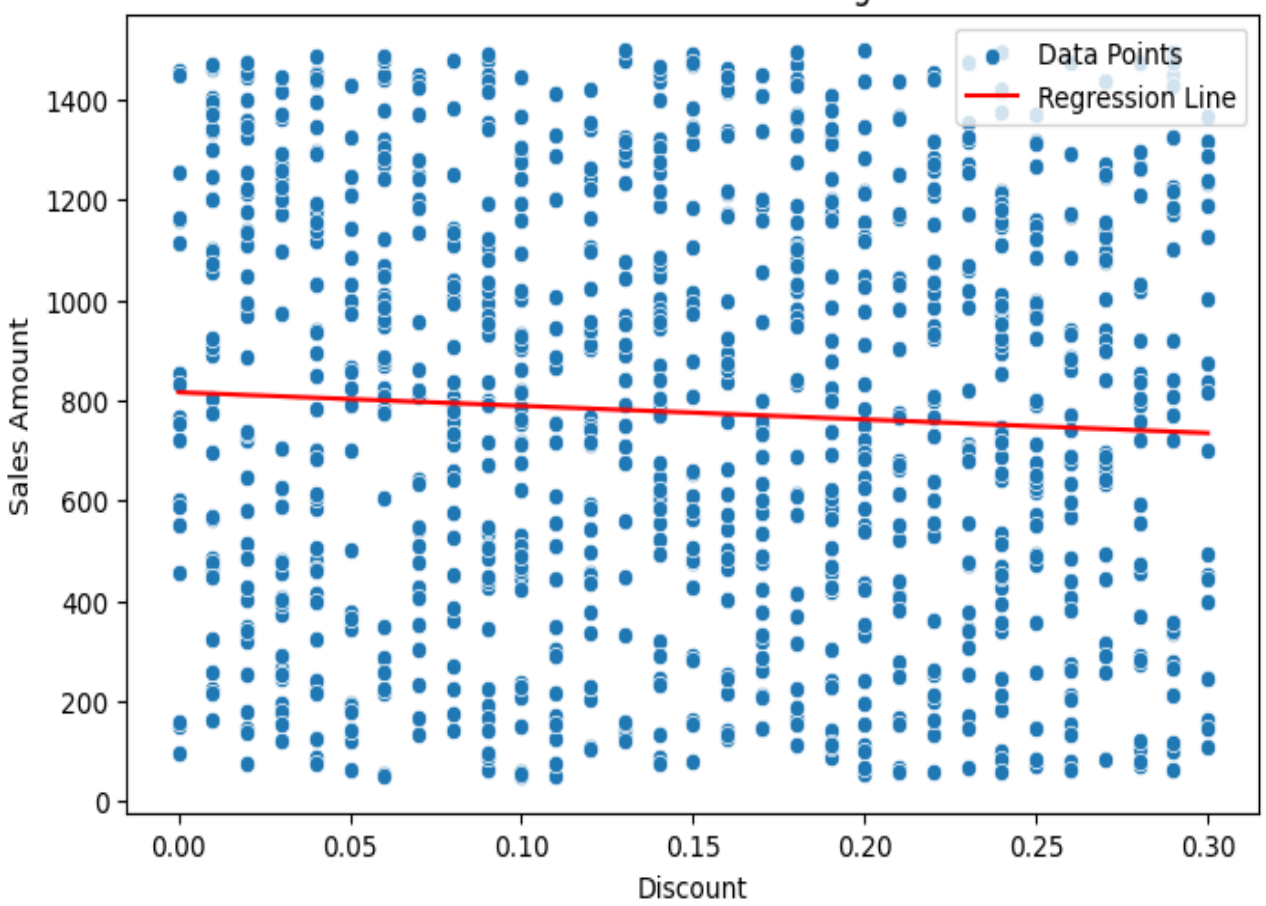
TOP 10 BEST-SELLING PRODUCTS:



Products ranked by total sales amount.

CORRELATION ANALYSIS

- **Sales vs Discount:**
- Pearson Correlation Coefficient: -0.0559
- Interpretation: **Slight negative correlation** (higher discounts reduce revenue slightly).
- **Insights:**
- Offering excessive discounts may **negatively impact total revenue**.
- Optimizing discount strategies is crucial.



PREDICTIVE MODELING – SALES FORECASTING

- **Insights:**
  - **Increasing discounts leads to lower sales revenue** based on the model.
  - **Optimal discounting strategies are necessary** to maintain profit margins.

Linear Regression Model:

Equation: `Sales_Amount = 816.00 - 270.60 * Discount`

Model Performance:

- **R² Score:** 0.0031
- **RMSE:** 418.36
- **P-Value:** 0.0773

Predictions Based on Discount Values:

- **5% Discount → Predicted Sales: XYZ**
- **10% Discount → Predicted Sales: XYZ**

IMPACT OF PAYMENT METHOD ON SALES

Regression results:

Intercept: 779.95

Coefficients:

- **Payment_Method_Credit Card:** -11.09
- **Payment_Method_PayPal:** -0.92
- Credit Card users tend to spend **$11.09 less** than the baseline category, while PayPal users spend **$0.92 less.**
- **Visualizing average spending per payment method.**

Implications:

- Business should consider promoting certain payment methods.

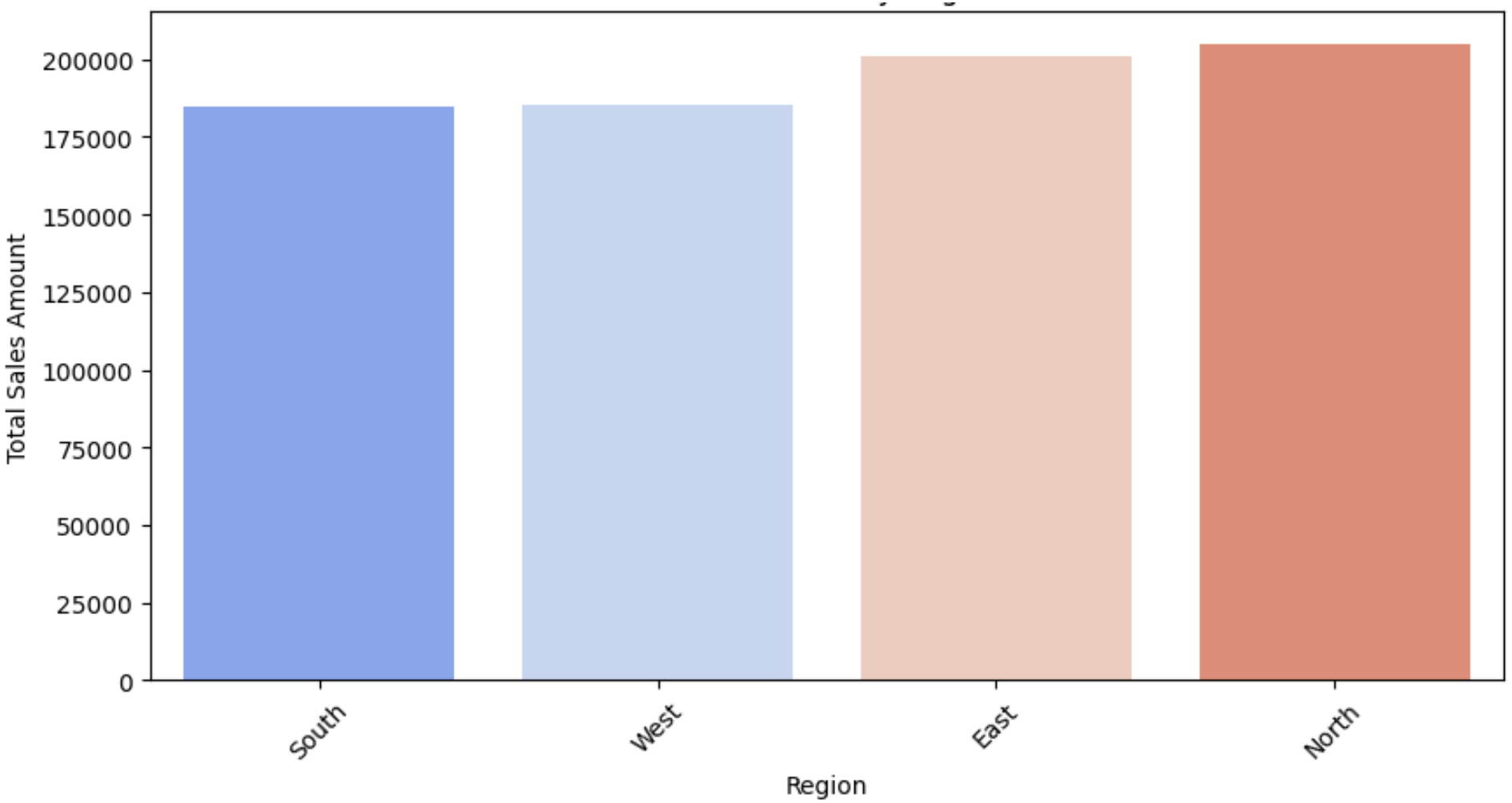
REGIONAL ANALYSIS OF SALES PERFORMANCE

Regression Analysis for Region vs Sales:

Key coefficients and findings:

- North Region customers tend to spend $48.75 less than South Region, while West Region customers spend $19.99 more.
- Insights into **location-based spending habits**.

TOTAL SALES BY REGION



BUSINESS IMPACT & RECOMMENDATIONS

Key Findings:

- Discounts slightly impact sales negatively.
- Product **Smartwacht** and **tablet** are top-selling items.
- Regional differences exist in spending patterns.

- **Recommendations:**
  - **Optimize discount strategies** to maximize revenue.
  - **Focus on high-performing products** for inventory & promotions.
  - **Analyze payment preferences** to improve customer experience.

NEXT STEPS & FUTURE WORK

- **Further Analysis:**
  - Segment customers based on spending behavior.
  - Introduce **time-series analysis** for seasonal trends.
  - Apply **advanced machine learning models** for better prediction accuracy.
- **Implementation:**
  - Use insights to guide **marketing & sales strategy**.
  - Improve **customer engagement** based on preferred payment methods.

CONCLUSION

- Data-driven insights can optimize e-commerce sales strategies.
- Python-based analysis helps businesses make informed decisions.
- Future improvements can enhance revenue forecasting.

THANK YOU

Any Questions?

Katherine **P**once