CAPSTONE PROJECT:

SALES DATA ANALYSIS & INSIGHTS

Data Analysis Essentials using Google Colab

PROJECT REPORT PRESENTATION

Presented By : **Sumaiya Tasnim**

Email: **sumaiyaa.tasnim.18@gmail.com**

Excelerate

WHAT ARE OBJECTIVES?

Assigned Dataset: "Sales_Transactions_3500.csv"

Project Deliverables Overview

Day 1: Python basics & data handling

Day 2: Data Cleaning & Exploratory Data Analysis

Day 3: Revenue Analysis & Reporting

DAY-1: PYTHON BASICS & DATA HANDLING TASKS → SKILLS GAINED

Exercise 1: Variables and Data Types → Data Handling

Exercise 2: Mathematical Operations → Computation

Exercise 3: Lists → Data Structuring

Exercise 4: Dictionaries → Mapping

Exercise 5: Conditionals → Logic

Exercise 7: Pandas DataFrame Creation and Manipulation → Data Manipulation

Exercise 8: NumPy Array Operations → Array Operations

Exercise 9: Matplotlib Basic Plotting → Data Visualization

Exercise 10: Seaborn Basic Plotting → Statistical Plots

DAY 2: DATA CLEANING & EXPLORATORY DATA ANALYSIS

Data Cleaning :

- Checked for missing values and handled them
- Standardized data types & formats
- Corrected inconsistent entries
- Removed duplicates
- Ensured data integrity for analysis

EDA Major Insights :

- Most expensive product identified in each category
- Average age varies across regions
- Popular product categories differ by region
- Transaction value comparison shows gender-wise trends

Few Important insights from EDA

Product Price Distribution

Product Price Distribution
Average and median prices ($248.33 and $249.00) are nearly identical, showing a balanced price range across products.

What Do Men and Women Buy?
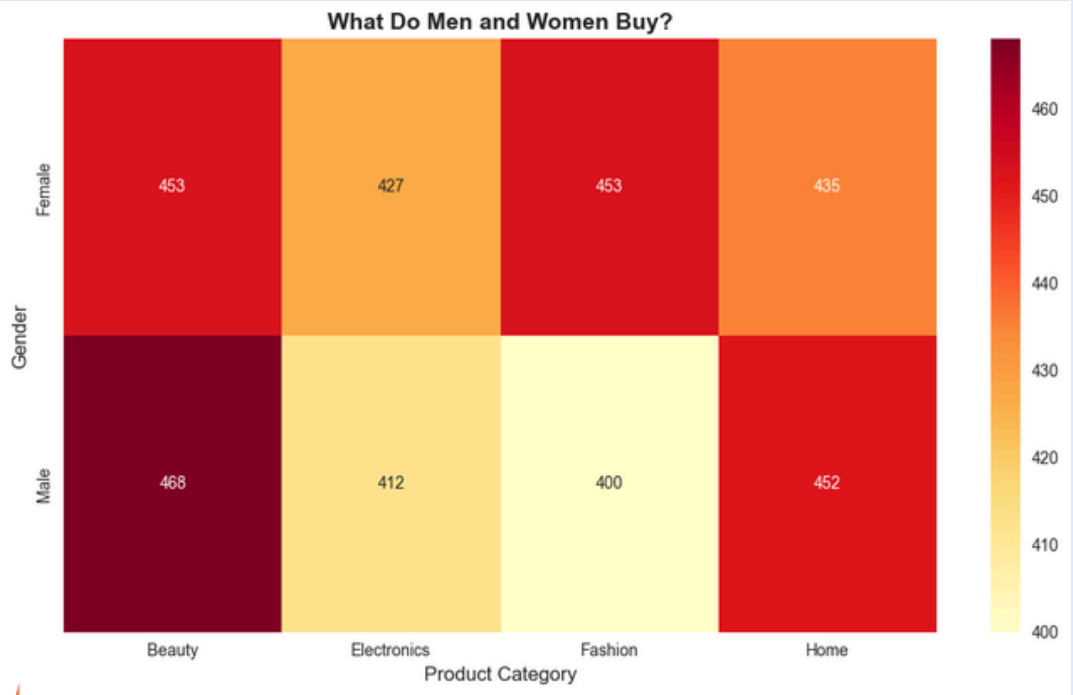
Category Purchases by Gender
Gender trends show Beauty popular among Males and Fashion among Females, guiding targeted marketing strategies.

Which Age Groups Spend the Most?
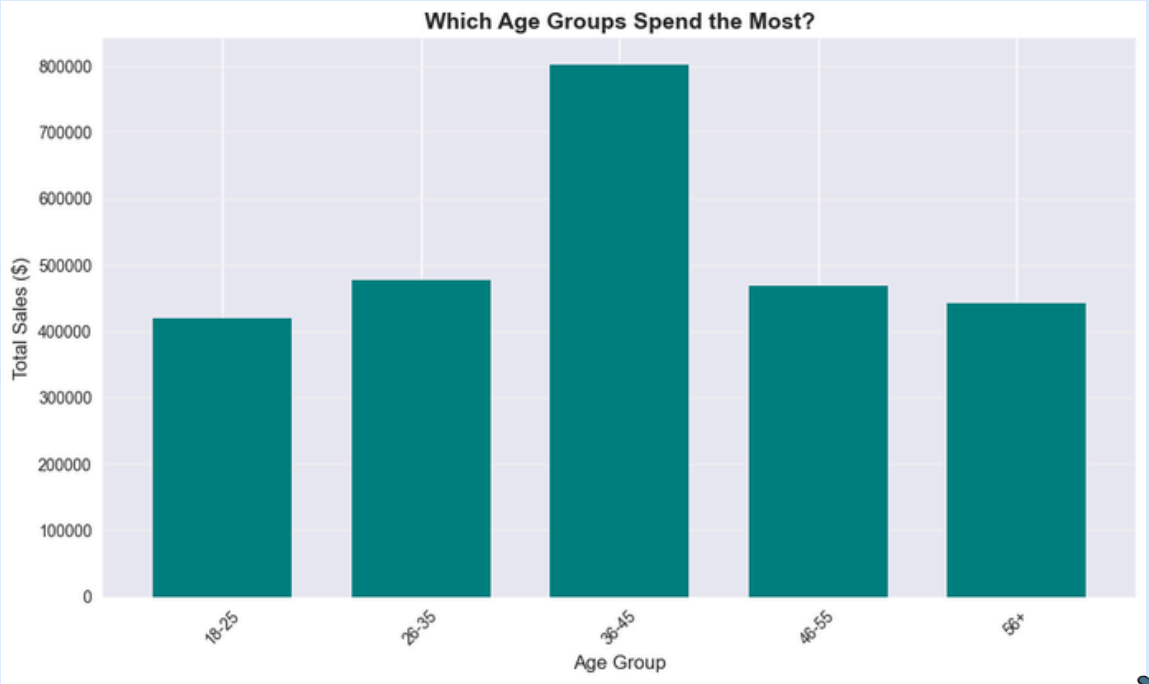
Total Sales by Age Group
The 36–45 age group records the highest spending, highlighting a key target segment for marketing and inventory focus.


How Old Are Our Customers?
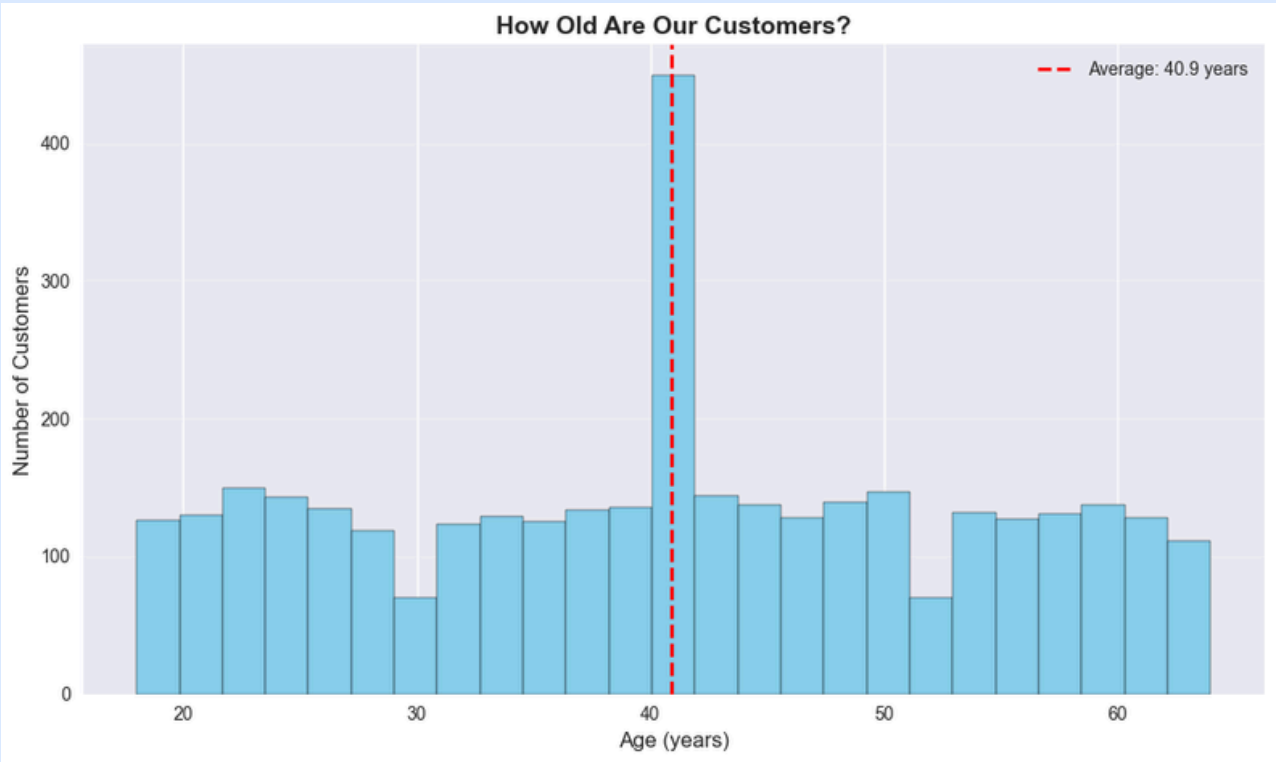
Customer Age Distribution
Average customer age is 40.9 years, with most customers concentrated around this range—ideal for audience targeting.

Most Popular Payment Methods
Card is the top payment choice, followed by Cash and Bank Transfer—insightful for improving payment systems.

Most Popular Payment Methods

- MobileMoney — 826
- BankTransfer — 882
- Cash — 889
- Card — 903
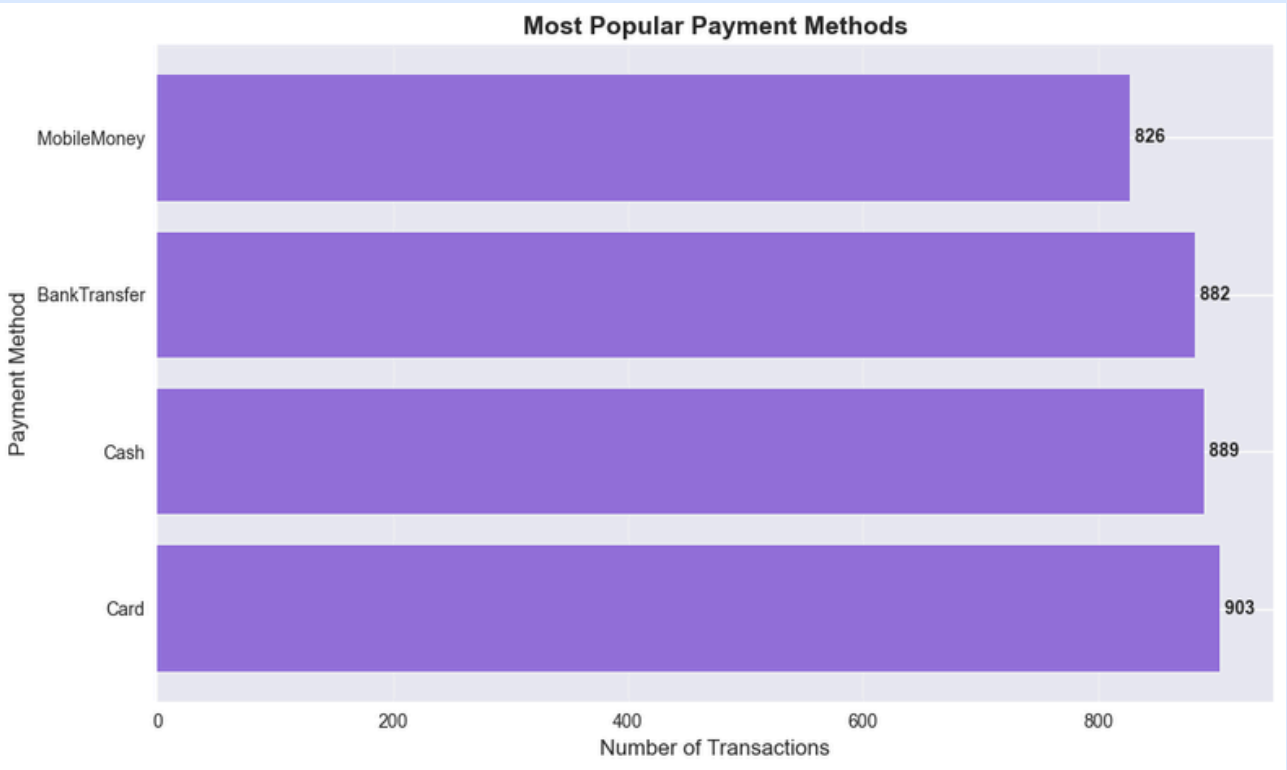
Major Insight 1 : Most expensive product identified in each category
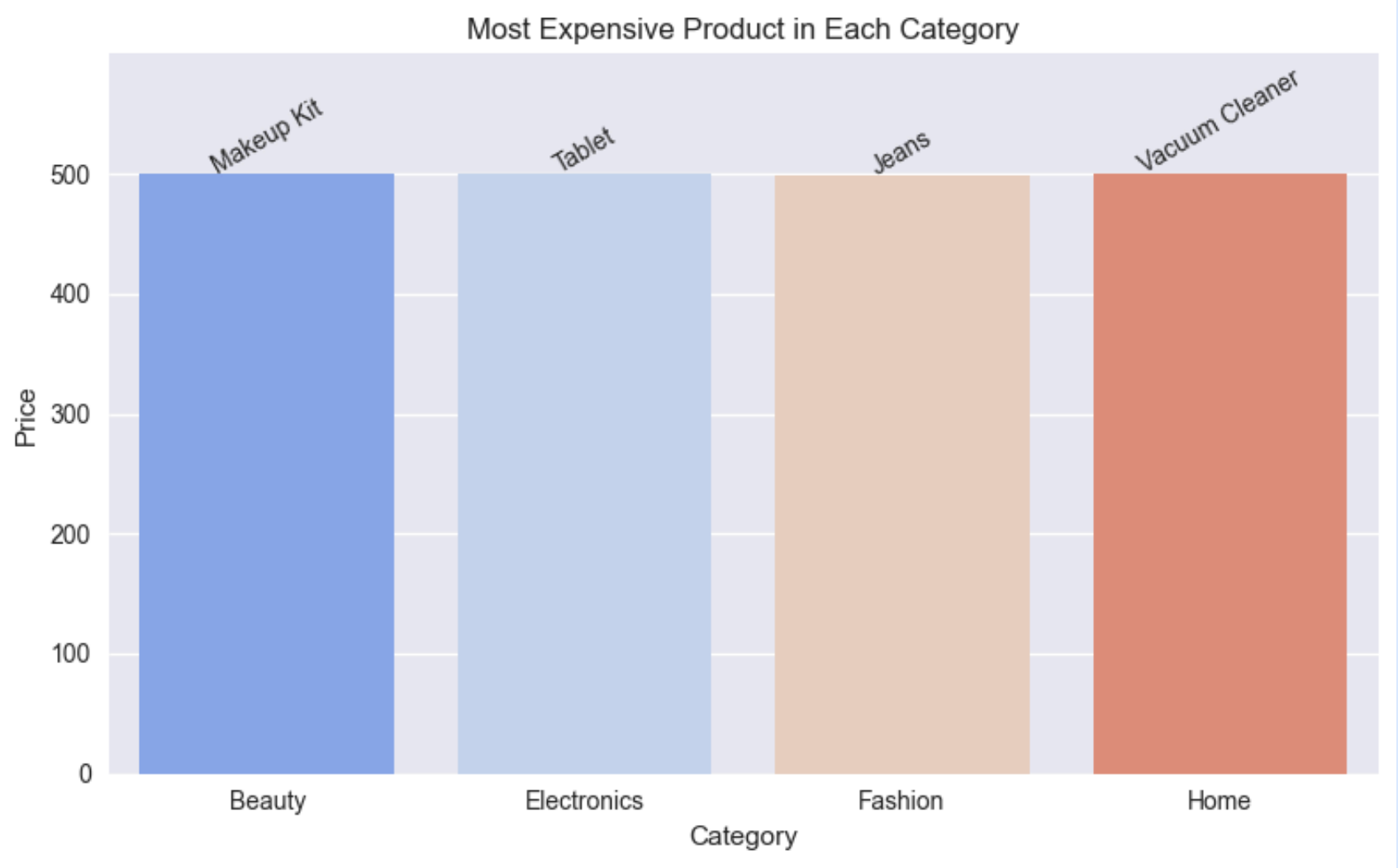
HISTOGRAM

Key Findings:

- Uniform Price: All top products (Beauty, Electronics, Fashion, Home) priced at $500.
- Price Leaders: Makeup Kit, Tablet, Jeans, Vacuum Cleaner.
- Price Ceiling: $500 may be the maximum price limit in the dataset.

Recommendations:

- **Check Price Cap: Verify if $500 is a data or market limit.**
- **Analyze Performance: Study sales and profitability of $500 products.**
- **Explore Premium Products: Consider introducing higher-priced items in high-demand categories.**

ROW BAR CHART

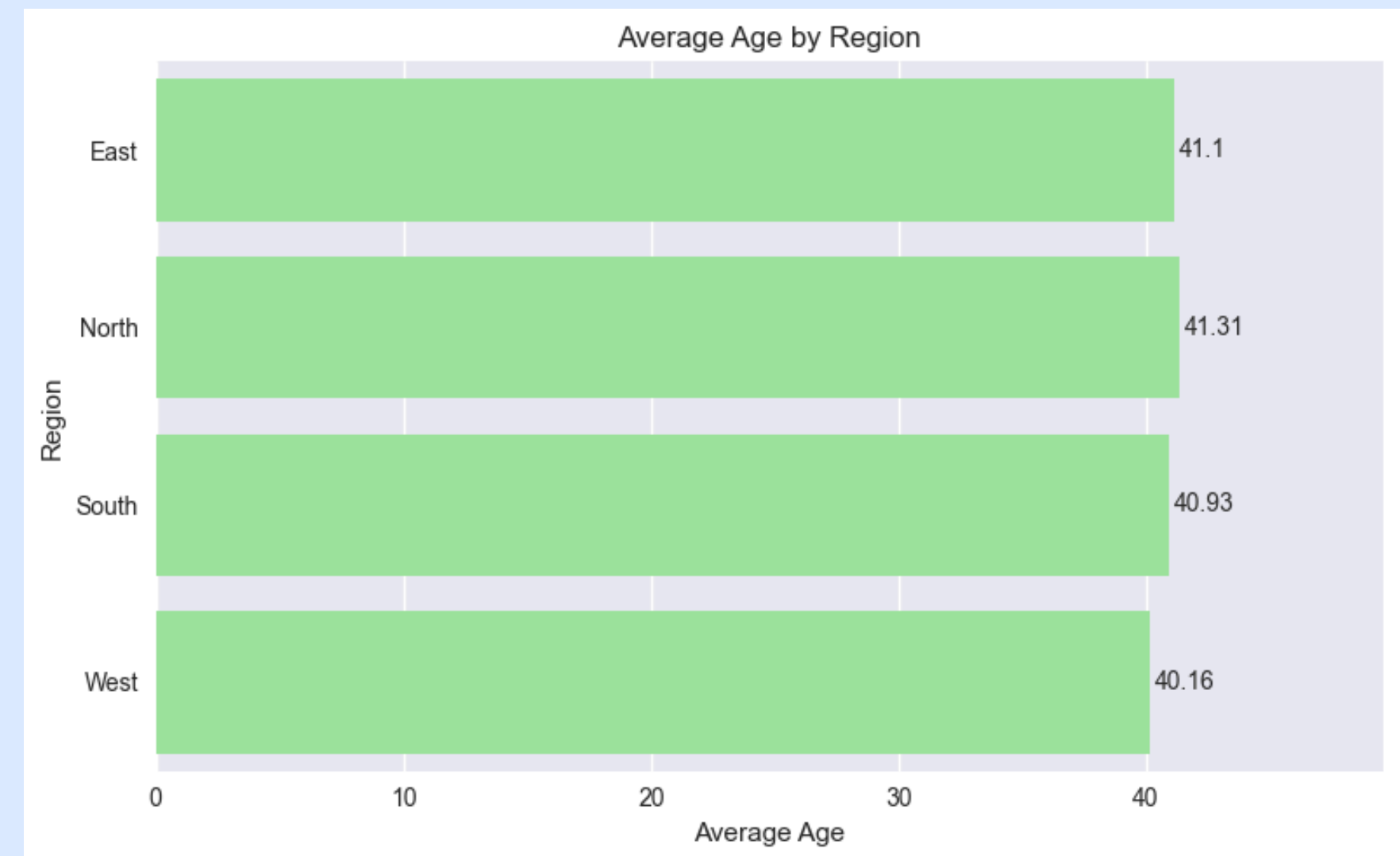## Major Insight 2 : Average age varies across regions

Key Findings:

- Minimal Difference: Average age is similar across regions (40.16 – 41.31 years).
- North Oldest, West Youngest: North = 41.31, West = 40.16.
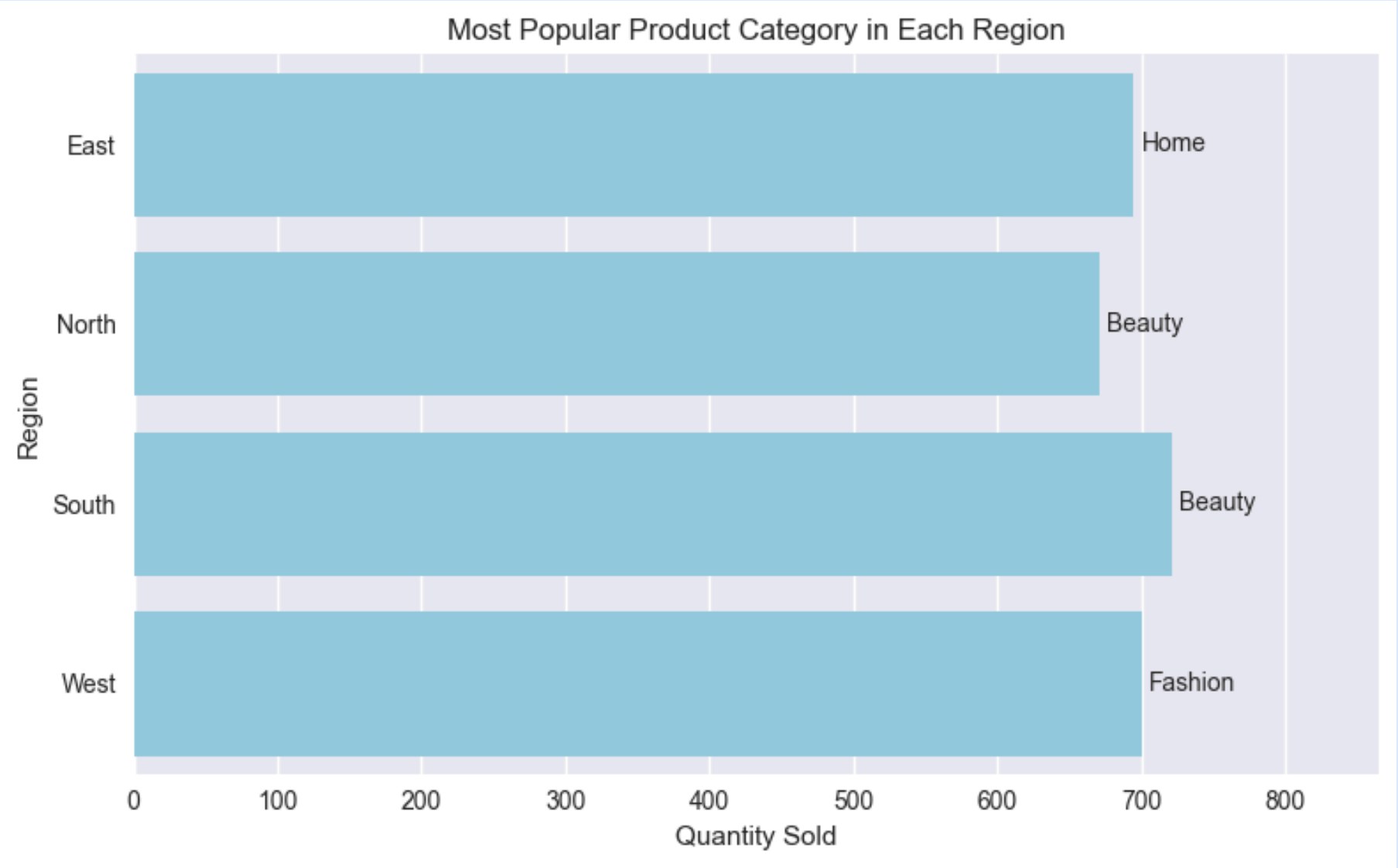- Region Not a Majors Factor: Small variance; age differences are negligible regionally.

Recommendations:

- **Treat Regions as Homogeneous: No major adjustments needed based on age.**
- **Drill Down Locally: Check sub-regions or cities for meaningful differences.**
- **Focus on Other Factors: Use income, education, or behaviors for segmentation instead of region.**

Average Age by Region

Region	Average Age
East	41.1
North	41.31
South	40.93
West	40.16

Major Insight 3 : Popular product categories differ by region

Most Popular Product Category in Each Region

ROW BAR CHART

Key Findings:

- Beauty Leads Overall: Most popular in North & South; South has highest sales (~725 units).
- Regional Variation: East → Home, West → Fashion; volumes high and consistent (675–725 units).
- East Lowest Volume: Home category in East has lowest top-category sales (~675 units).

Recommendations:
- Expand Beauty Success: Apply North & South strategies to East & West.
- Support Regional Leaders: Optimize inventory and promotions for Home (East) & Fashion (West).
- Boost East Performance: Investigate bottlenecks and improve sales of leading products.

BAR CHART

Key Findings:

- Males Higher ATV: $766.31, ~$40.69 above females.
- Females Contribute Significantly: $725.62, still a high-value segment.
- Small Difference: Only ~5.6% higher for males; both genders drive revenue.

Recommendations:

- **Boost Female ATV: Use bundling, upselling, or complementary items.**
- **Maintain Male Performance: Ensure high-demand products are well-stocked and promoted.**
- **Cross-Analyze Volume: Consider total transactions by gender to optimize resource allocation.**

Major Insight 4 : Transaction value comparison shows gender-wise trends

Average Transaction Value by Gender
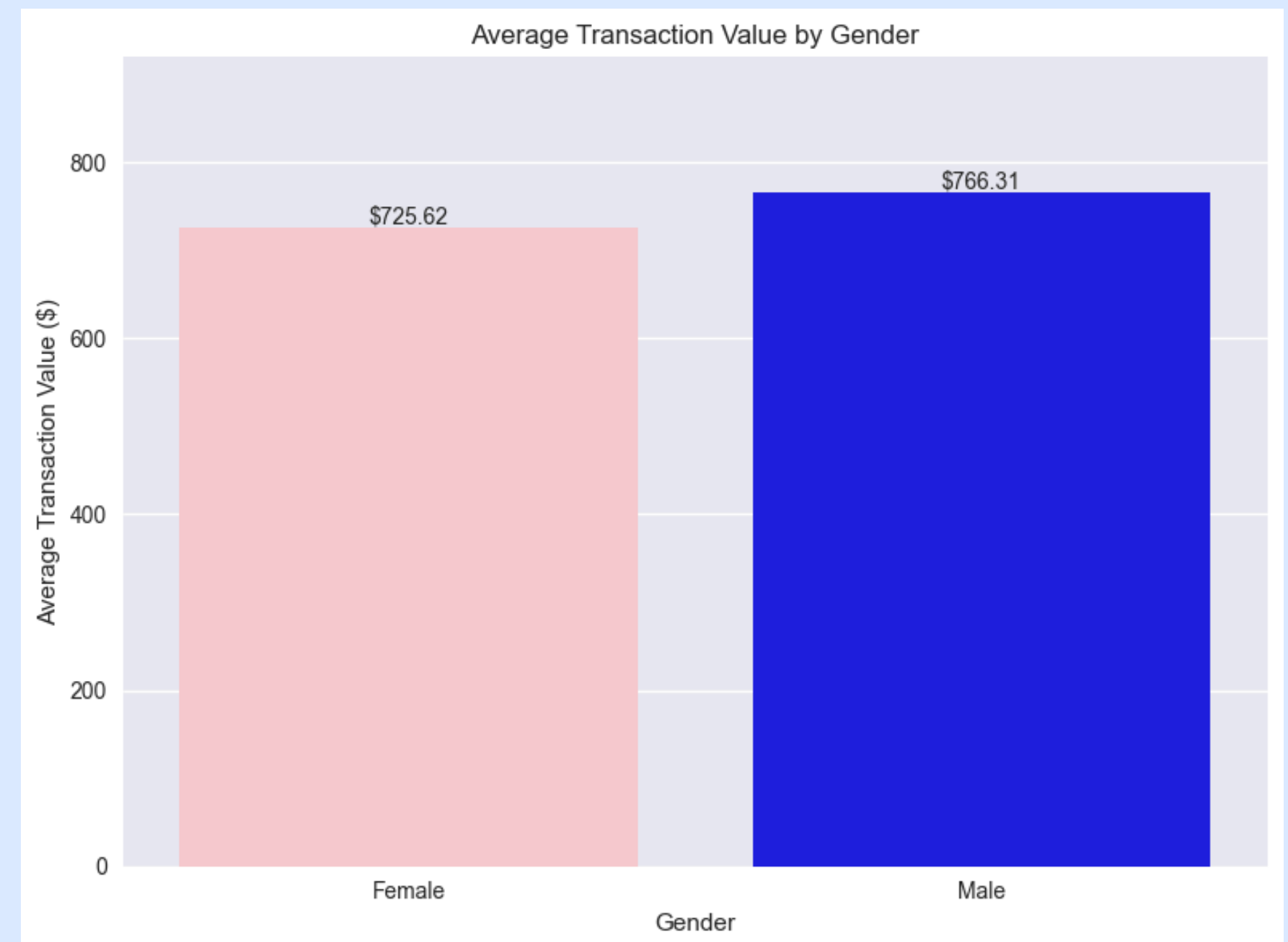
Day-3 → Major Insight 5 : Revenue Distribution Across Regions

Comprehensive 2x2 Regional Revenue Dashboard Visualization
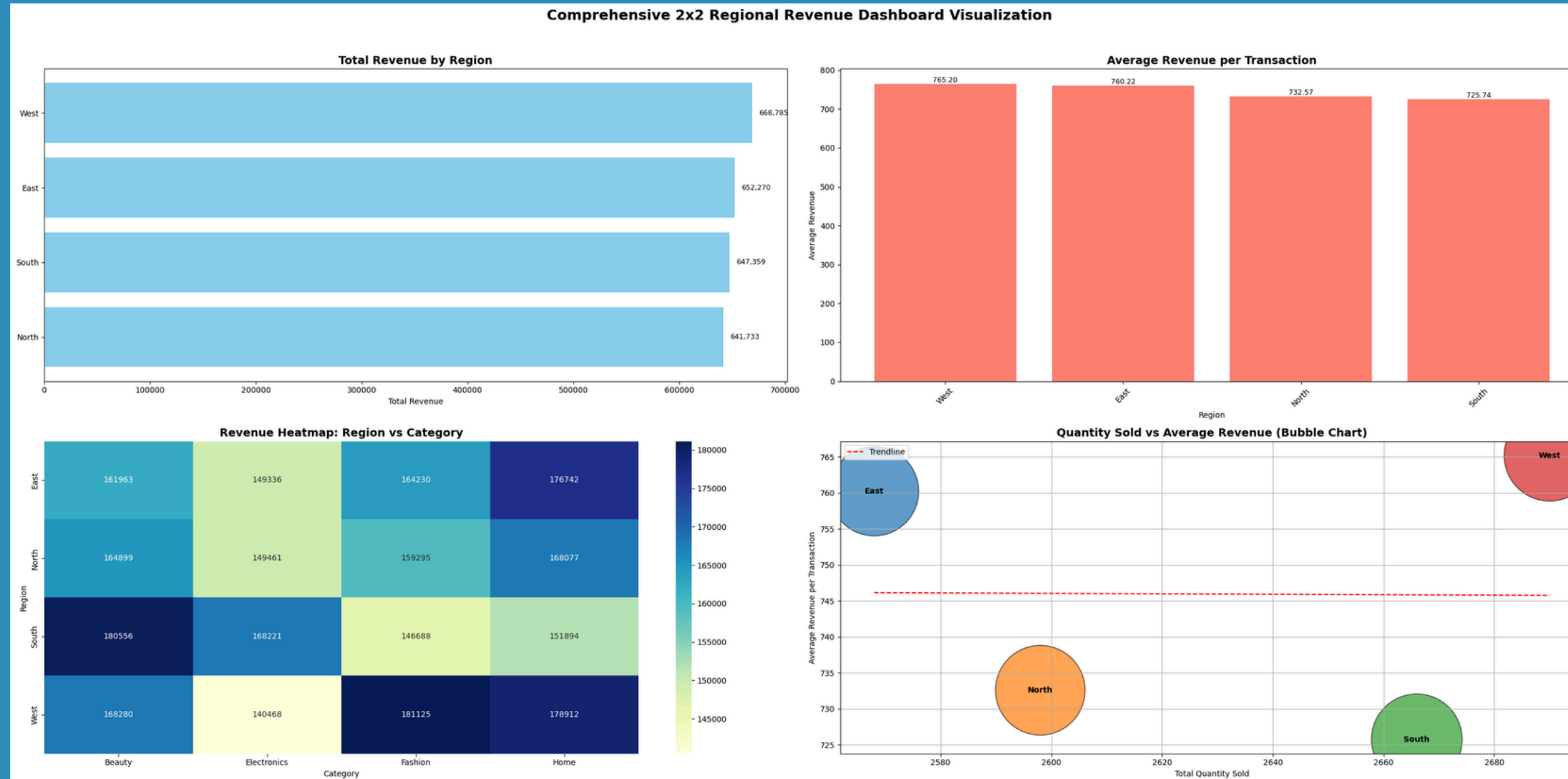
Key Findings:
- **West & East Lead: Highest total revenue; East has highest average revenue per transaction.**
- **West High Volume, Lower Value: West drives revenue via quantity; East combines high volume & value.**
- **Home Category Dominates: Highest revenue in most regions; Electronics underperforms.**

Recommendations:
- **Boost West Value & East Volume: Upselling in West; increase sales volume in East.**
- **Focus on Home & Analyze Electronics: Prioritize inventory & promotion for Home; investigate Electronics.**
- **Standardize East Best Practices: Apply East's successful mix, pricing, and service strategies to other regions.**

Major Insight 6 : Revenue Distribution per Transaction Across Regions

Revenue Distribution per Transaction Across Regions
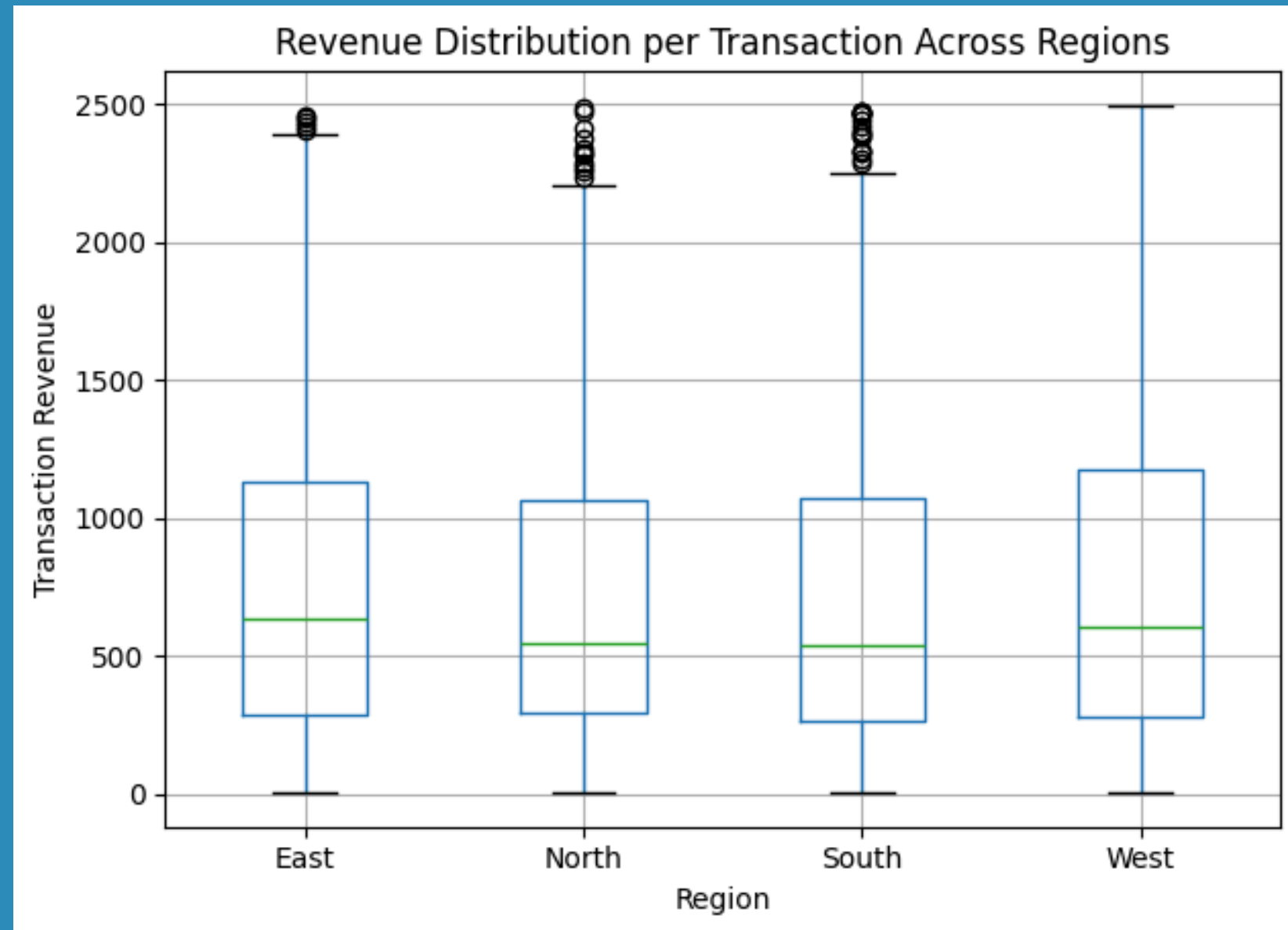
box plot

Key Findings:

- **West Leads: Highest median (~$600) and largest IQR, indicating high variability.**
- **Other Regions Similar: East, North, South have similar median and spread (~$550–$600).**
- **High-Value Outliers Present: All regions show sporadic transactions above $2,500.**

Recommendations:

- **Leverage West Insights: Analyze products/customers to replicate high transaction values in other regions.**
- **Standardized Upselling: Apply a unified strategy to East, North, South to raise median revenue.**
- **Convert Outliers: Implement VIP/high-spender programs to increase top-end transaction frequency.**

MAJOR INSIGHT 7 : REVENUE TREND OVER TIME BY REGION

Revenue Trend Over Time by Region
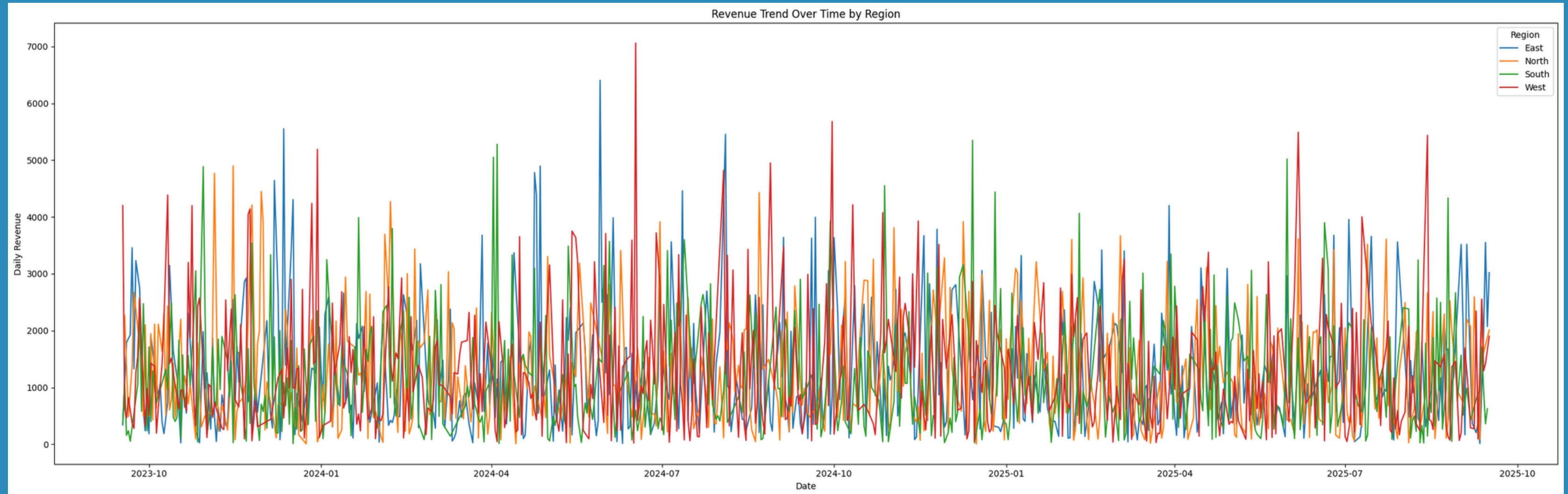
Key Findings:
- **High Volatility: Daily revenue swings from ~$0 to $7,000; no clear long-term trend.**
- **South Peaks Highest: South region reaches maximum daily revenue (~$7,000).**
- **Inter-Regional Correlation: Revenue spikes occur simultaneously across regions, suggesting common external factors.**

Recommendations:
- **Manage External Factors: Align promotions and events to stabilize revenue trends.**
- **Replicate South Success: Analyze South's peak strategies and apply to other regions.**
- **Increase Baseline Revenue: Implement consistent engagement to reduce low-revenue days.**

CONCLUSION

In this capstone project, I conducted a comprehensive analysis of the Sales_Transactions_3500.csv dataset, building foundational Python skills, performing data cleaning, and exploring key insights through EDA. I identified top-performing products, revenue patterns by region, and customer trends by gender, as well as high-value categories and transaction distributions. The analysis revealed opportunities to optimize marketing, inventory, and pricing strategies to maximize both average and peak transaction values. These insights provide a clear foundation for data-driven decision-making and targeted strategies to drive revenue growth across regions and product categories.

Thank you, Excelerate! for this opportunity to apply data analysis skills, explore sales data, and gain valuable hands-on experience.

EXcelerate

LOW SKILL HIGH