Retail Sales Analysis Project Report

End-to-End Data Analytics Case Study

Author: Bassam Mohammed

Presentation Overview

1. **Executive Summary**
   A high-level overview of the project's scope and key findings.

2. **Dataset & Data Cleaning**
   Understanding the raw data and the preparation steps involved.

3. **MySQL Analysis & Power BI Dashboards**
   Detailed insights from SQL queries and interactive visualizations.

4. **Key Insights & Recommendations**
   Actionable strategies derived from the data analysis.

5. **Tools & Skills Used**
   Technologies and methodologies employed in the project.

Executive Summary

This project delivers a comprehensive retail analysis, spanning the entire data analytics pipeline from raw data to actionable insights. We focused on transforming complex datasets into clear, data-driven strategies for business growth.

1

Data Preparation

Six raw datasets were meticulously cleaned and standardized using Python (Pandas), addressing issues like encoding problems, null values, and inconsistent naming conventions.

2

SQL Analysis

Ten critical business questions were addressed through optimized SQL queries in MySQL, extracting key metrics on sales, customer behavior, and employee performance.

3

Visualization & Insights

Cleaned datasets were visualized in interactive Power BI dashboards, revealing trends in product sales, customer demographics, and promotional impact to inform strategic decisions.

Dataset & Data Cleaning: Dataset Overview

The analysis was performed on six distinct retail datasets, encompassing a wide range of business operations from transactions to employee data.

Dataset	Records	Description
transactions.csv	**~6.4M**	Individual sales transactions, including product and discount details.
products.csv	**~18K**	Details for each product, such as category, subcategory, and production cost.
customers.csv	**~1.6M**	Customer demographics, including location, gender, and job.
stores.csv	**35**	Store locations and names across various regions.
employees.csv	**404**	Employee data, including their names and roles.
discounts.csv	**181**	Information on promotional events and discount percentages.

Dataset & Data Cleaning: Data Cleaning (Python)

Data cleaning was a critical step, executed using the Pandas library in Python. This process ensured data integrity and consistency for subsequent analysis.

- Removed null values and duplicate entries to enhance data reliability.

- Standardized column names and formatting in Python
- (Country name standardization was performed later in Power BI for reporting clarity)

- Ensured proper character encoding by using utf8mb4 in MySQL for multilingual data support.

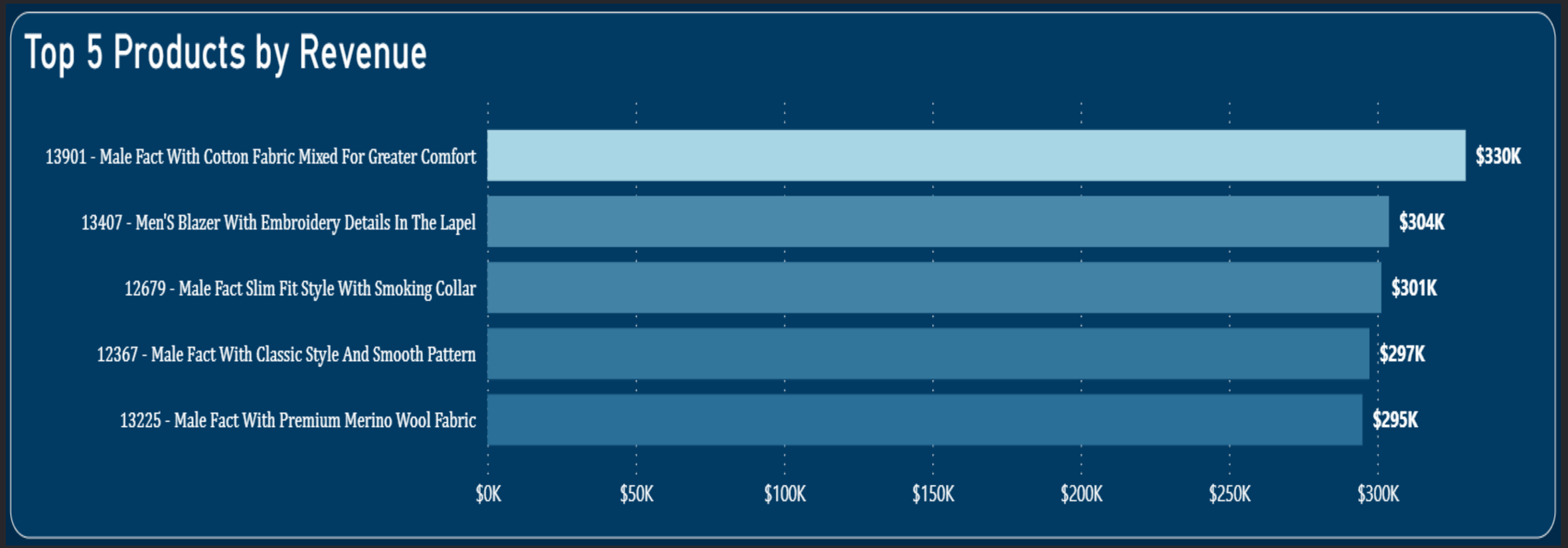
- Outputted cleaned datasets as *_cleaned.csv files, ready for re-import into MySQL.

MySQL Analysis & Power BI Dashboards: Sales Performance

Our initial MySQL analysis focused on understanding core sales performance and product profitability.

Key Questions & Insights

- Q1: Top Products by Revenue
  - Product 13901 generated the highest revenue at $330K, indicating strong market demand.

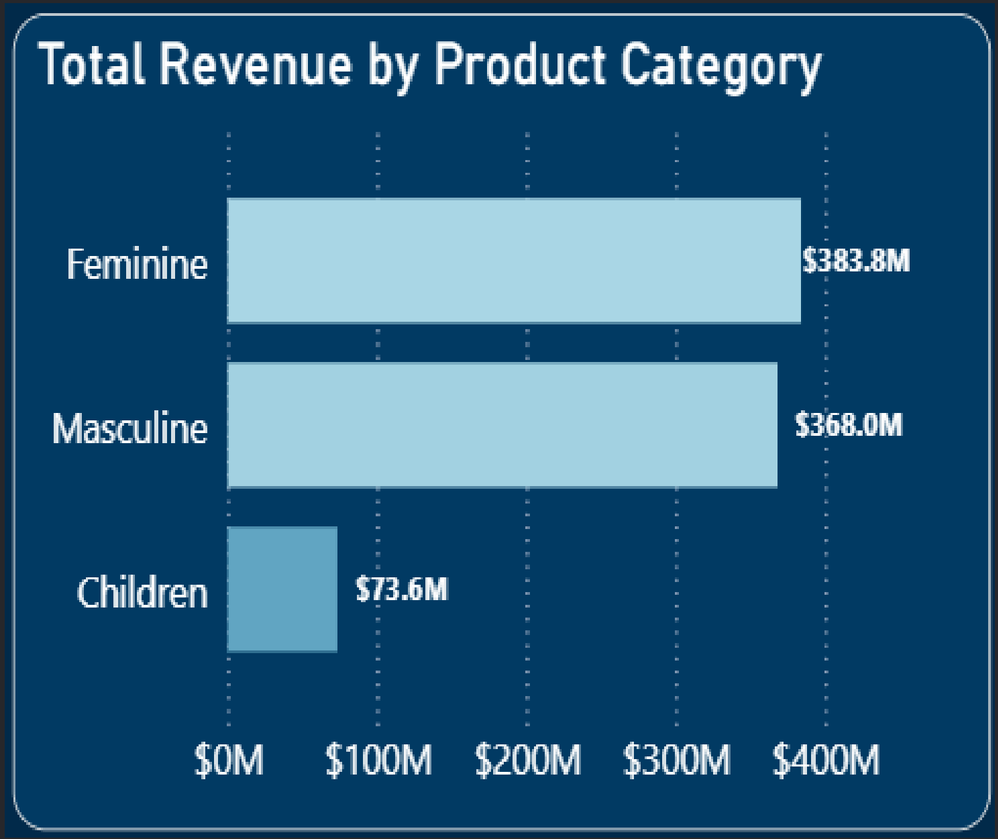


MySQL Analysis & Power BI Dashboards: Sales Performance

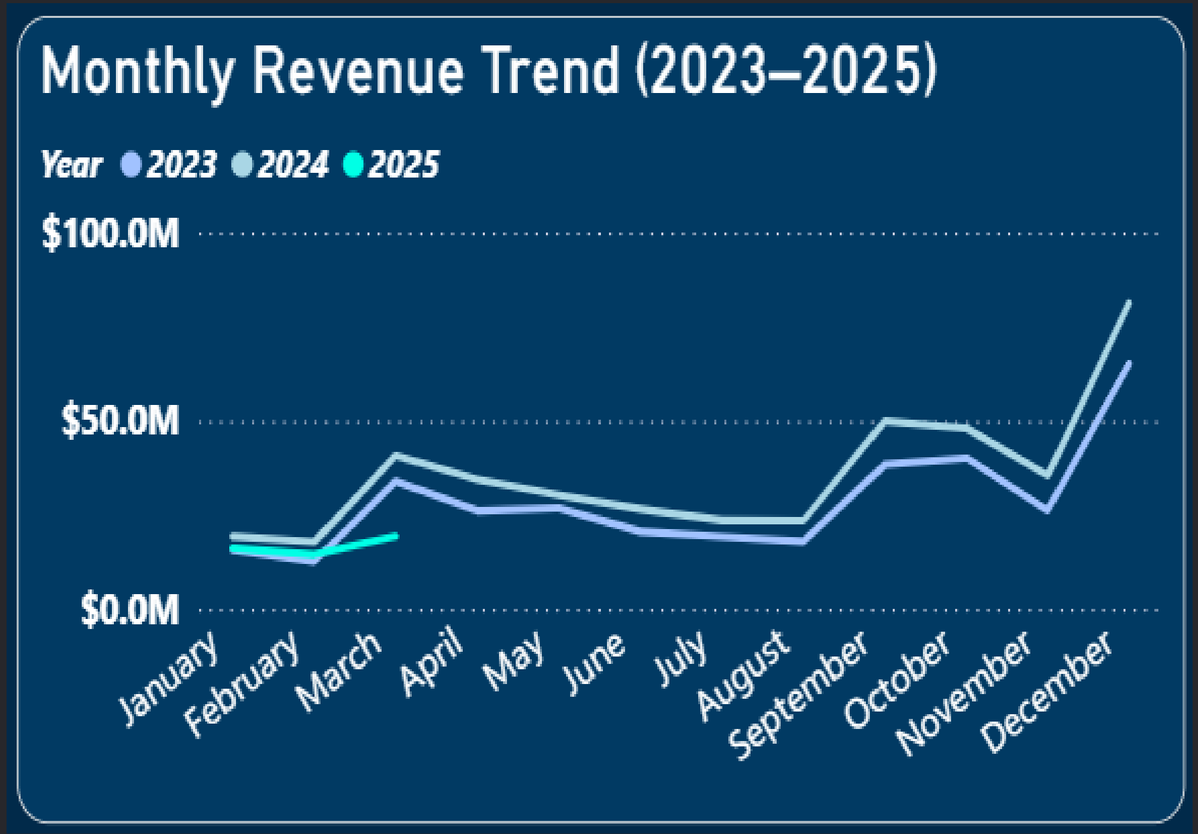
Our initial MySQL analysis focused on understanding core sales performance and product profitability.

Key Questions & Insights

- Q2: Most Profitable Categories
  - The “Feminine" category dominated revenue across all categories.



- Q3: Monthly Revenue Trends
  - Sales consistently peaked in December each year, with elevated trends starting in September and October — indicating strong seasonal demand.



MySQL Analysis & Power BI Dashboards: Geography & Stores

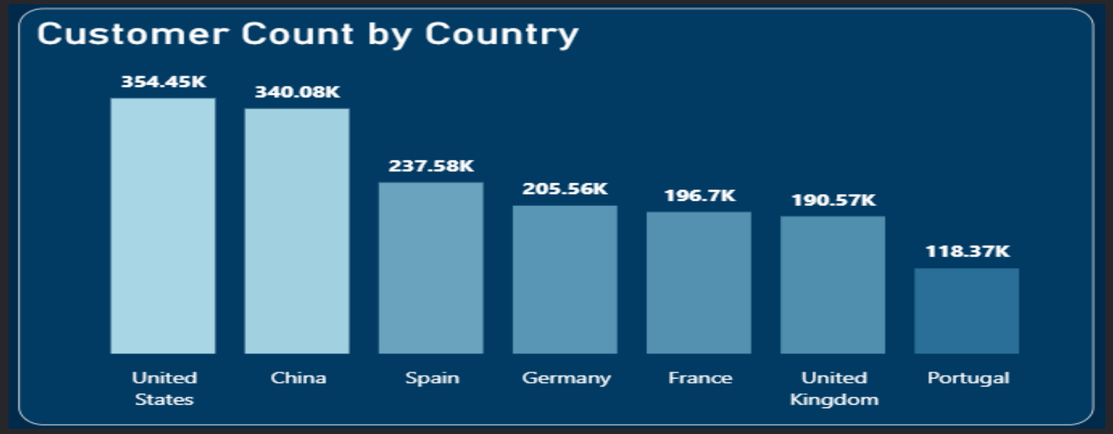
This segment delved into geographical sales performance and customer demographics.

Key Questions & Insights

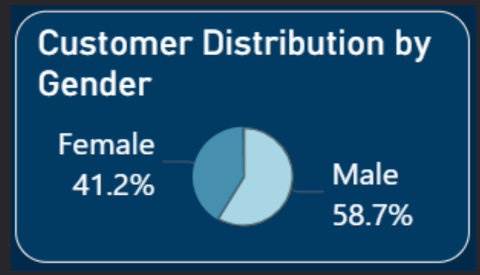
- **Q4: Top Regions by Revenue**
  - Stores located in Asia and North America generated the most significant revenue, highlighting key markets.

Region	Total Revenue
Asia	$606.43M
Europe	$134.42M
North America	$84.51M
Total	**$825.36M**

- **Q5: Countries with Most Customers**
  - The U.S. and China lead in customer count, each with over 340K customers, followed by key European markets like Spain and Germany.



- **Q6: Customer Distribution**
  - Male customers outnumber females, suggesting they could be a key focus for targeted marketing strategies.

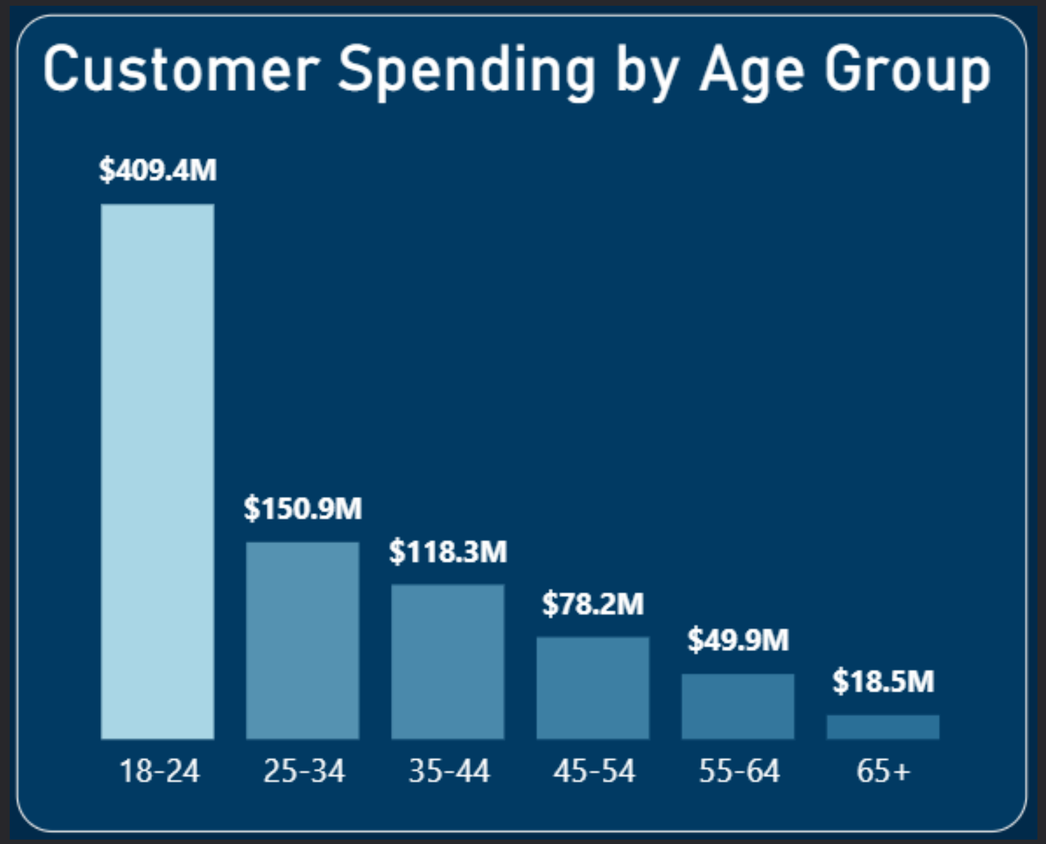


MySQL Analysis & Power BI Dashboards: Customer, Staff & Promotions

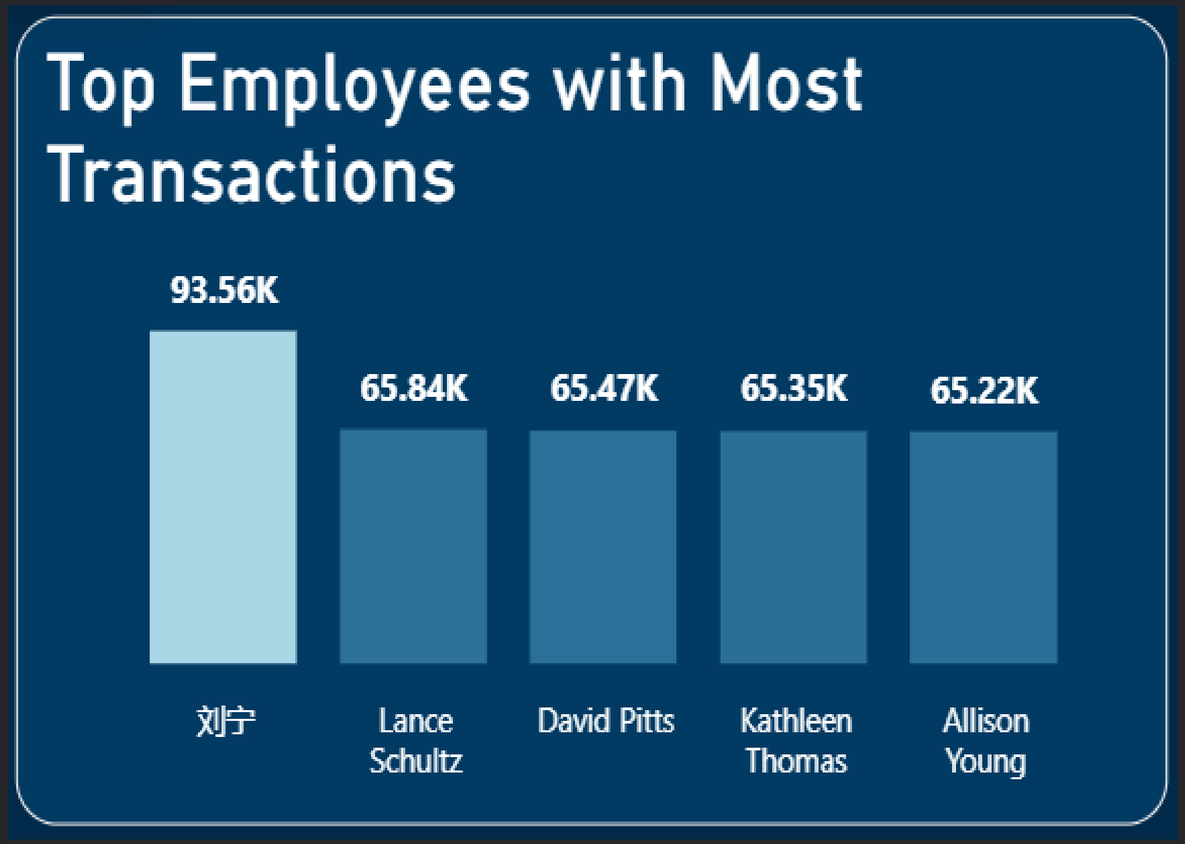
This analysis focused on customer engagement, employee performance, and the direct impact of discount strategies on sales volume.

Key Questions & Insights

- Q7: Top Spending Age Groups
  - Young adults aged 18–24 drive the highest spending, making them ideal for targeted loyalty programs.



- Q8: Most Active Employees
  - Employee 刘宁 recorded the highest number of transactions, indicating outstanding individual performance and operational efficiency.

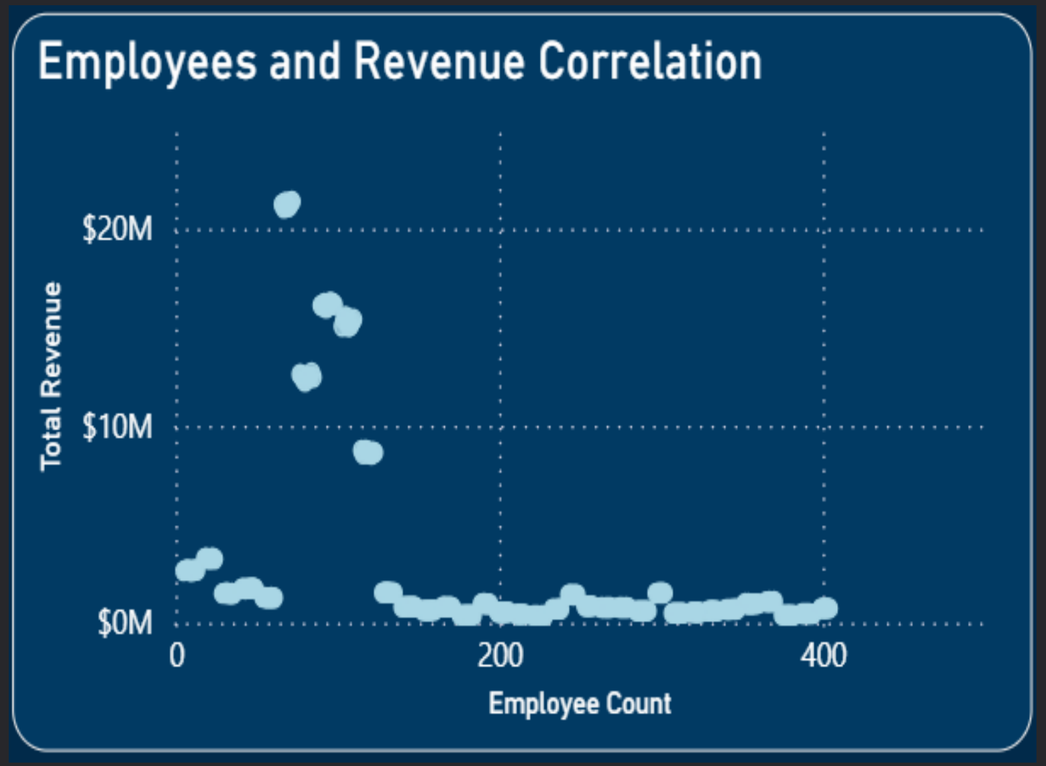


MySQL Analysis & Power BI Dashboards: Customer, Staff & Promotions

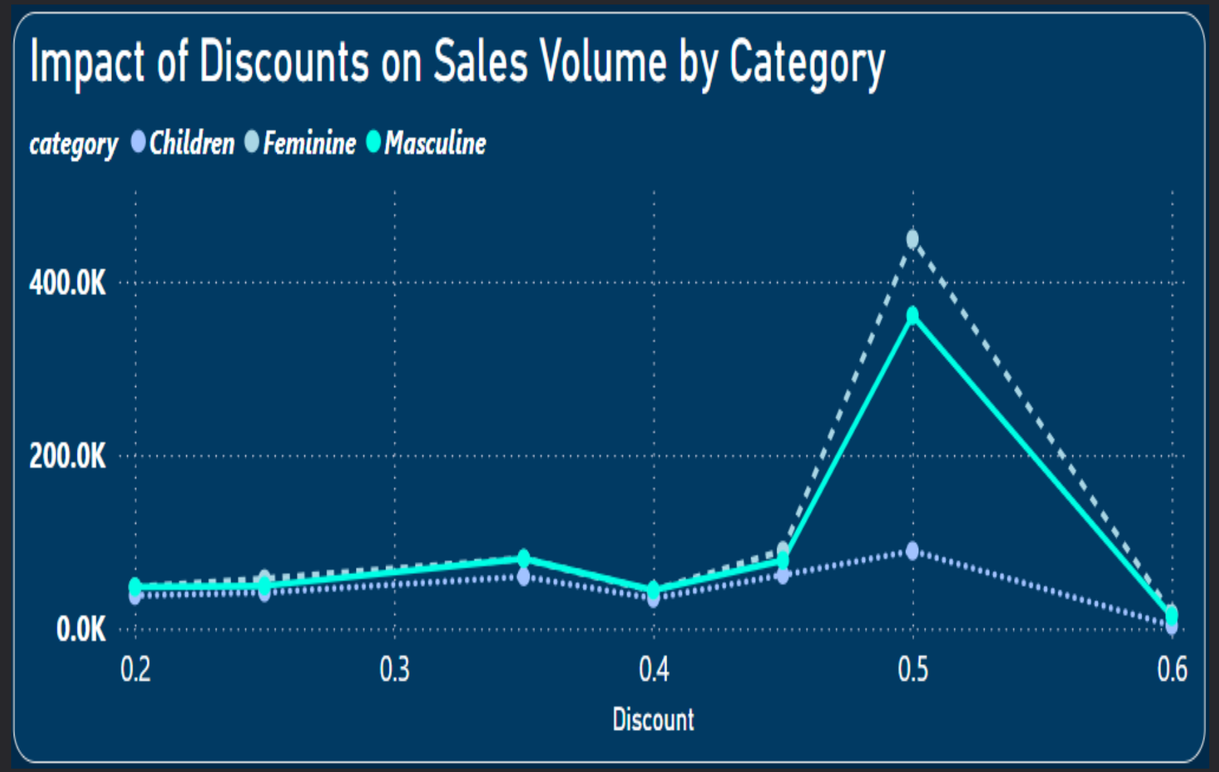
This analysis focused on customer engagement, employee performance, and the direct impact of discount strategies on sales volume.

Key Questions & Insights

- **Q9: Employees per Store vs. Revenue**
  - The chart suggests that higher employee counts don't always translate to greater revenue — most top-revenue-generating stores have a moderate staff size, indicating a potential efficiency threshold.



- **Q10: Discount Impact on Sales Volume**
  - A 50% discount consistently results in significant increases in sales volume in all of the three categories Children, Feminine, and Masculine.

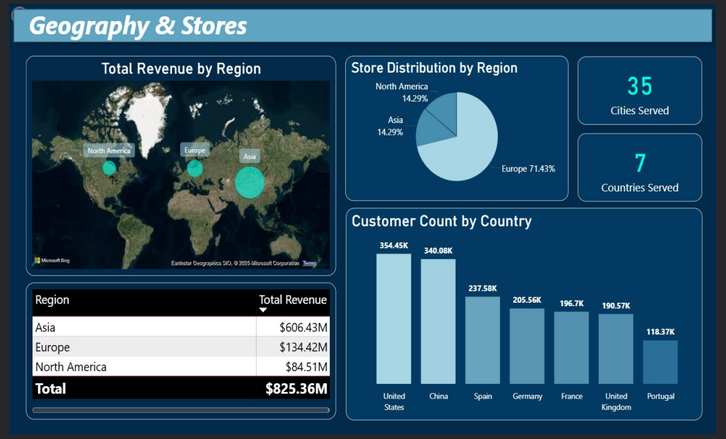


Power BI Dashboards Overview

Our Power BI dashboards deliver a clear, interactive snapshot of key retail metrics, making it easy for stakeholders to analyze performance and make informed decisions.

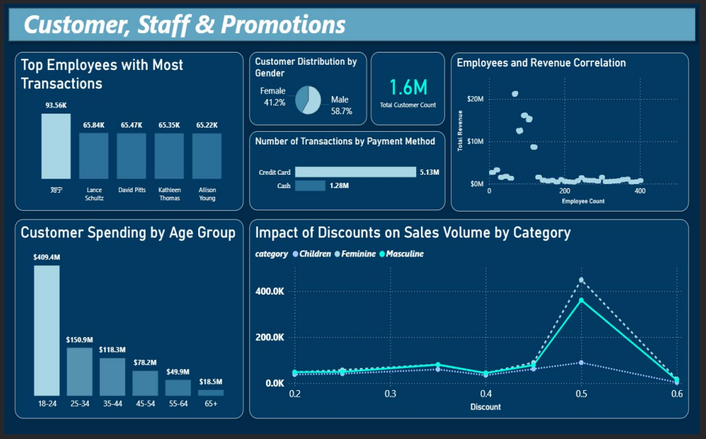
Page 1: Sales Performance

Visualizes product performance and category sales trends with interactive cards and charts.



Page 2: Geography & Stores

Offers insights into geographical sales distribution using maps and detailed demographic breakdowns.



Page 3: Customer, Staff & Promotions

Highlights customer spending behavior, staff performance, and the impact of promotions on sales.

Key Insights & Recommendations

Based on our comprehensive analysis, we've identified several key insights and formulated actionable recommendations to optimize retail performance.

Promotional Focus

50% discounts proved most effective in boosting sales across Feminine, Masculine, and Children categories. Future promotions should prioritize this range to maximize revenue.

Product Investment

Prioritize inventory and marketing on top-performing products like the "Male Fact With Cotton Fabric..." and "Men's Blazer...", which are driving the highest revenue.

Talent Retention

Implement incentive programs to retain **high-performing employees** like "刘宁" and “Lance Schultz”, who significantly contribute to transaction volumes.

Targeted Marketing

Develop specific loyalty programs and marketing campaigns tailored to the **18–24 age group**, identified as the highest-spending demographic.

Strategic Expansion

Explore opportunities to expand store presence or optimize operations in **high-performing regions** like China and the USA to leverage existing market strengths.

Tools & Skills Used

To achieve comprehensive data analysis and impactful visualizations, we leveraged a suite of powerful tools and honed essential skills.

Python

Employed for robust data cleaning, extensive preprocessing, and feature engineering to prepare diverse datasets for analysis.

MySQL

Utilized for efficient data storage, complex SQL queries, and managing relational databases to retrieve specific insights.

Power BI

Implemented for designing interactive dashboards and dynamic reports, enabling intuitive data exploration and visualization.

PowerPoint

Used for crafting compelling presentations and executive summaries, effectively communicating key findings and recommendations.

- **Key Skills Developed:**
  - Advanced SQL for complex data retrieval and manipulation.
  - Proficient Data Cleaning techniques to ensure data accuracy and reliability.
  - Effective use of Joins to integrate disparate data sources.
  - Strategic Data Visualization for clear and concise reporting.
  - Applied Business Intelligence principles to drive actionable insights.

Thank You

Github: https://github.com/Doodad7

LinkedIn: https://www.linkedin.com/in/bassam-mohammed-ba3284315/