Customer Shopping Behavior Analysis

This project analyzes 3900 customer purchases to uncover spending patterns, product preferences, and subscription behavior, guiding strategic business decisions for enhanced growth and customer satisfaction.

Project Overview & Dataset Summary

Project Goal

Analyze customer shopping behavior using transactional data to identify key insights into spending patterns, customer segments, product preferences, and subscription behavior.

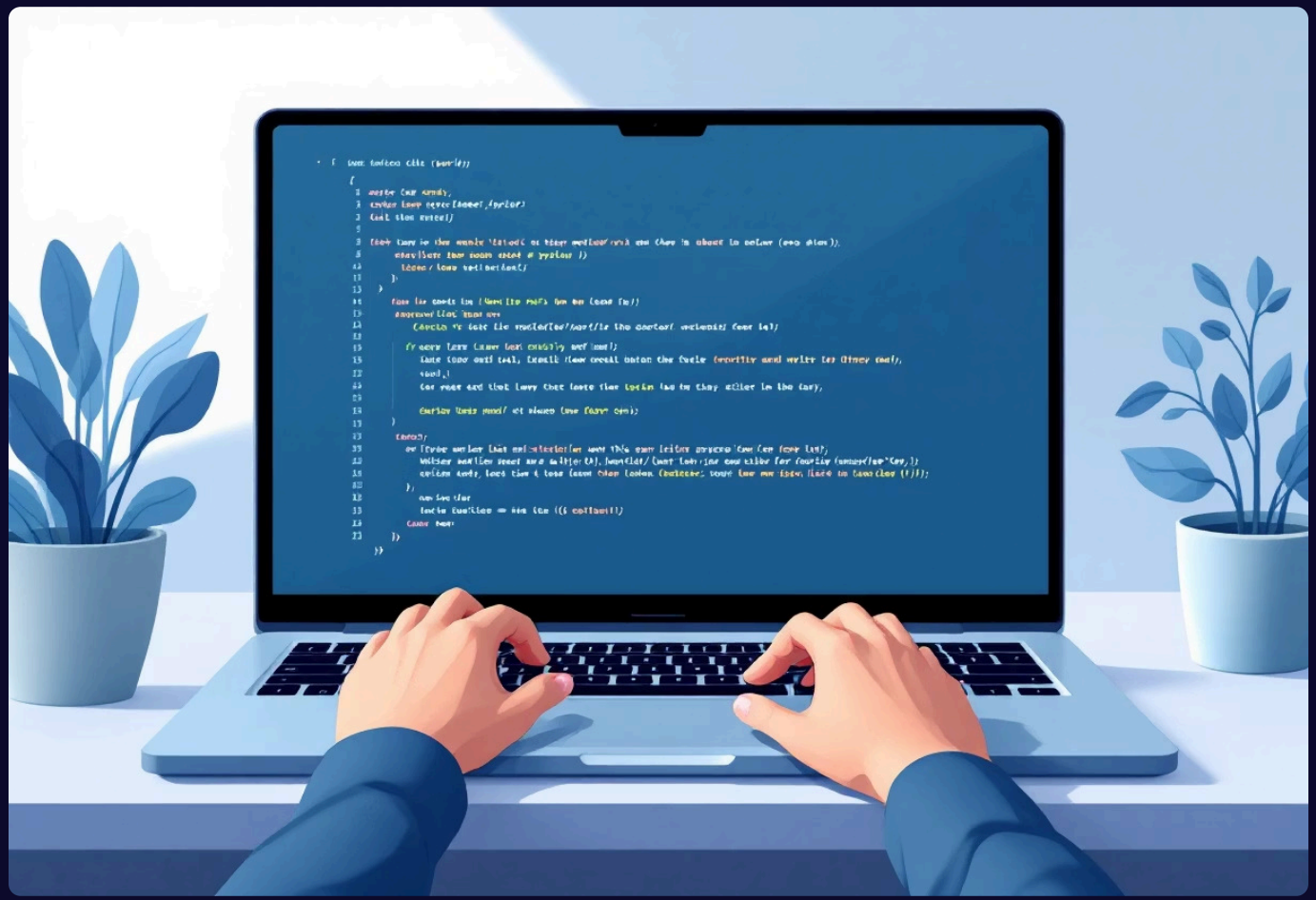
This analysis aims to inform strategic business decisions and optimize marketing efforts.

Dataset Details

Our dataset comprises 3900 rows and 20 columns, detailing customer demographics, purchase specifics (item, category, amount, season, size, color), and shopping actions (discounts, promo codes, previous purchases, frequency, review ratings, shipping type).

Only 37 missing values were found in the 'Review Rating' column.

Data Collection & Initial Review



Data Sourcing

The raw transactional data was sourced from Kaggle, providing a rich foundation for our analysis. It was then loaded into Python for initial Exploratory Data Analysis (EDA).

```
import pandas as pd
import numpy as np
df=pd.read_csv('shopping_trends
.csv')
```

This initial step ensures data integrity and prepares it for subsequent cleaning and transformation.

Data Cleaning & Preparation

01

Missing Data Handling

We checked data types and null values. Missing values in the 'Review Rating' column were imputed using the median rating of each product category to maintain data quality.

02

Column Standardization

Column titles were converted to SNAKE CASE (e.g., 'purchase_amount_usd') for improved readability and coding efficiency, ensuring consistency across the dataset.

03

Feature Creation

New features were engineered, including 'age_group' by binning customer ages and 'purchase_frequency_days' derived from purchase data, enriching our analytical capabilities.

04

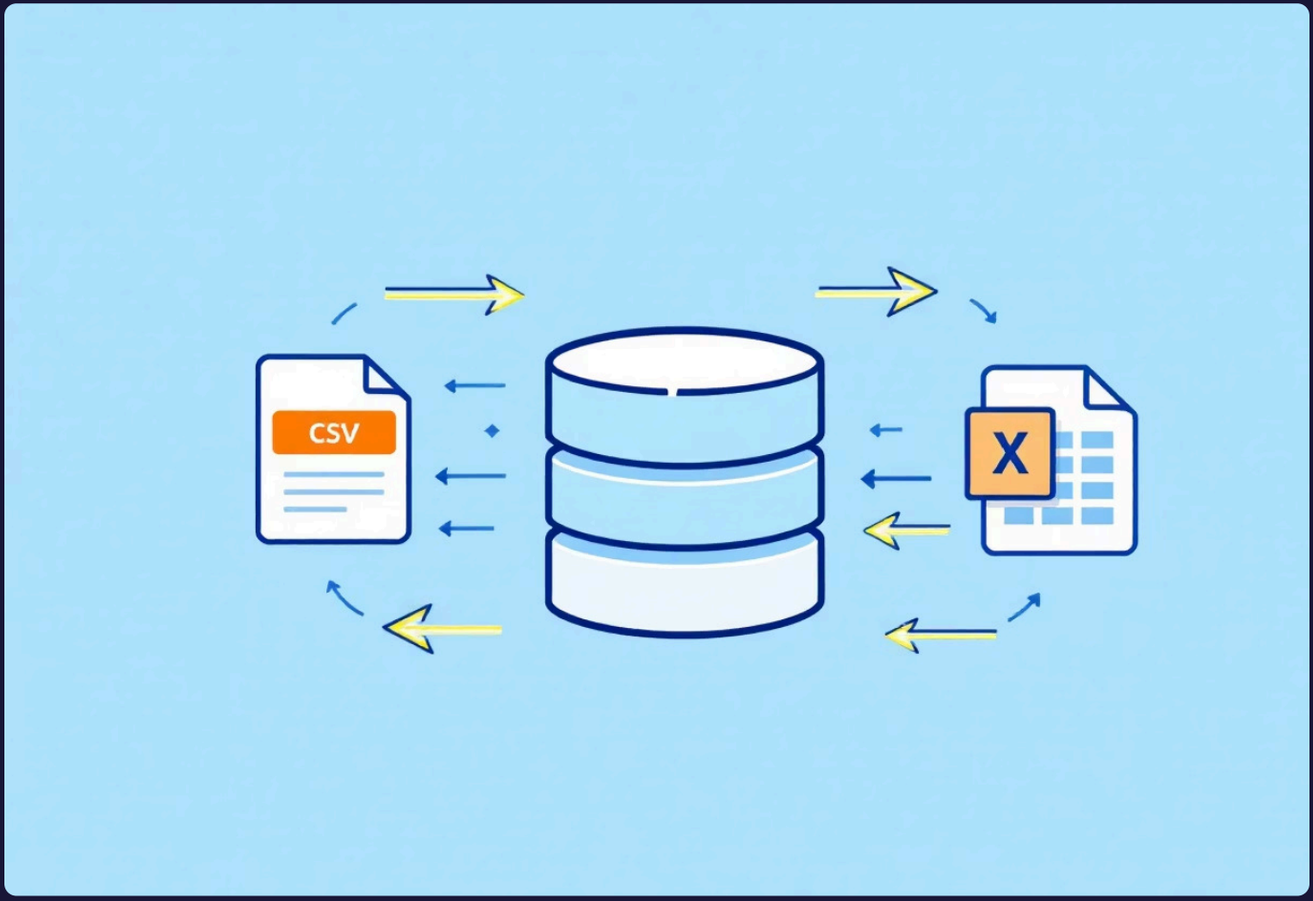
Data Consistency Check

Redundancy between 'discount_applied' and 'promo_code_used' was identified. The 'promo_code_used' column was subsequently dropped to streamline the dataset.

Data Export & SQL Analysis

Data Export

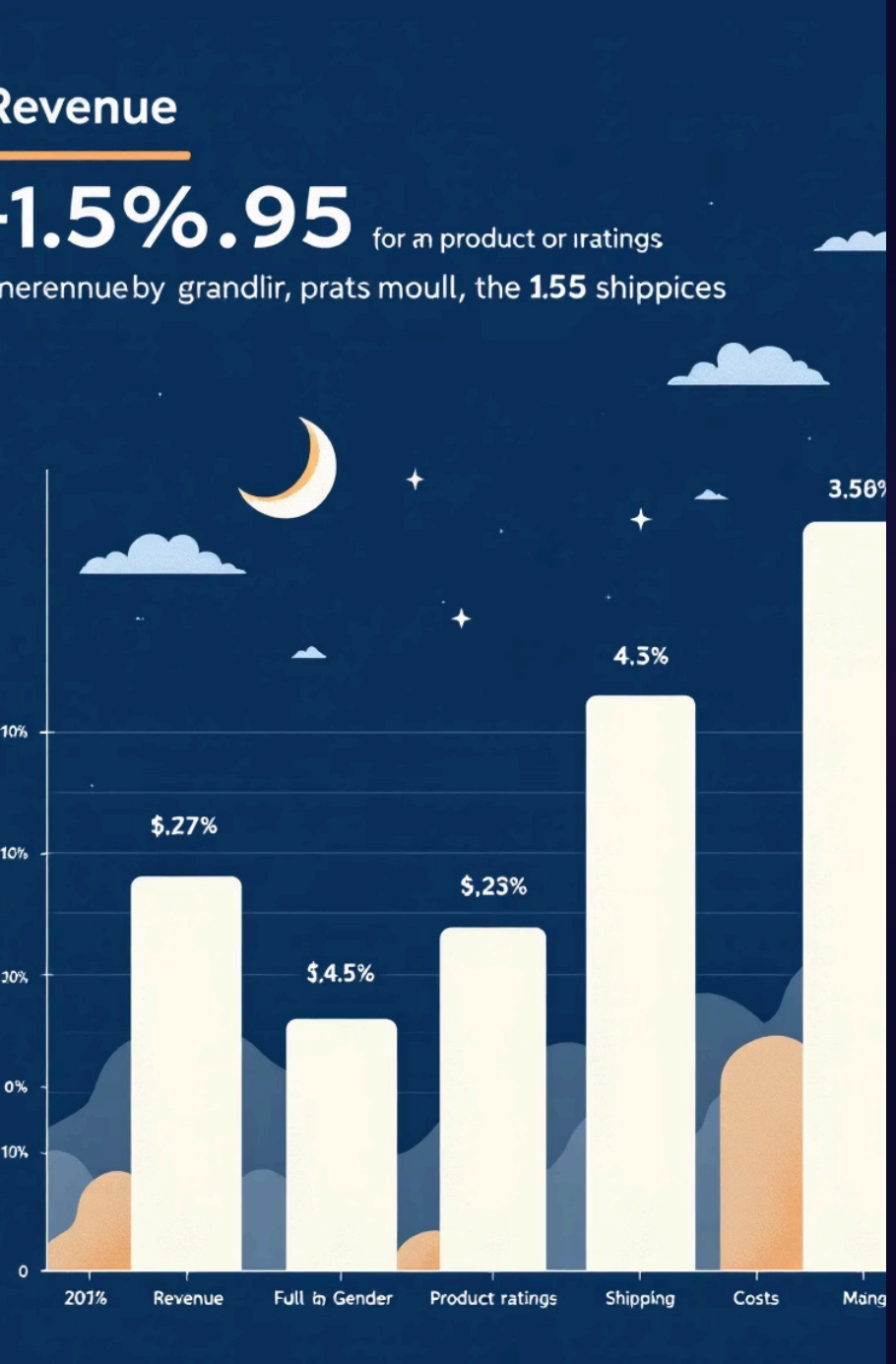
The cleaned and prepared data was saved into CSV and Excel formats, then exported to MySQL for structured query analysis. This ensures the data is accessible for advanced querying and reporting.



SQL Data Analysis

Structured analysis in MySQL addressed key business questions, including revenue by gender, top 5 products by revenue, average shipping costs, subscriber vs. non-subscriber spending, and top discounted items.

For example, male customers generated $112,428 in revenue, while females generated $52,365.



Key SQL Insights: Revenue & Products

Revenue by Gender

Male customers contributed $112,428 in total revenue, significantly more than female customers at $52,365.

Top 5 Products

Gloves (3.89 avg. rating), Sandals (3.85), Boots (3.82), Hat (3.81), and Shorts (3.80) are the top revenue-generating products.

Shipping Costs

Average purchase amounts vary by shipping type: Express ($61.34) and Standard ($58.21).

Key SQL Insights: Subscriptions & Discounts

Subscription Impact

Subscribers (759 customers) have an average spend of *$59.12* generating **$44,869** in total revenue.

Non-subscribers (2010 customers) have a similar average spend of *$59.66* but higher total revenue at **$119,924.**

Discounted Rate Purchases

Sneakers (51.40%)

Jewelry (50.00%)

Hat (49.57%)

Sweater (49.56%)

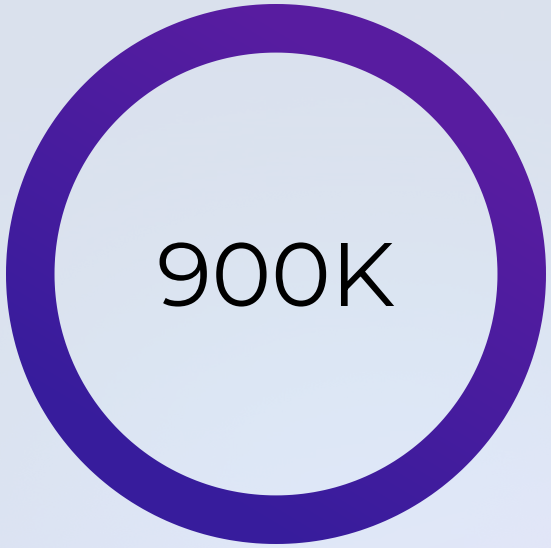
Coat (47.97%)

Items that recorded the highest number of purchases where discounts were applied

Power BI Dashboard: Visualizing Customer Behavior

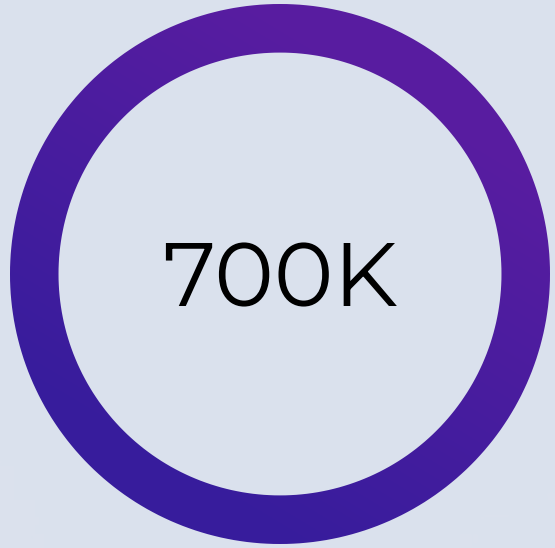
This Power BI report provides a comprehensive analysis of customer behavior, highlighting key purchasing patterns, revenue drivers, and customer segments. It offers actionable insights for marketing, sales, and customer engagement strategies.

Dashboard Highlights: Sales & Demographics



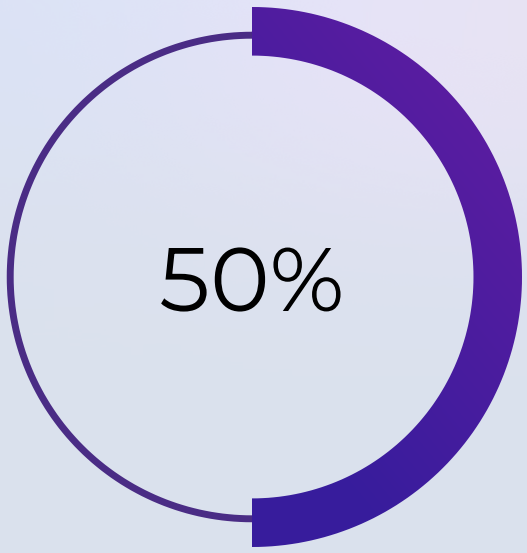
Clothing Revenue

Clothing generates the highest revenue.



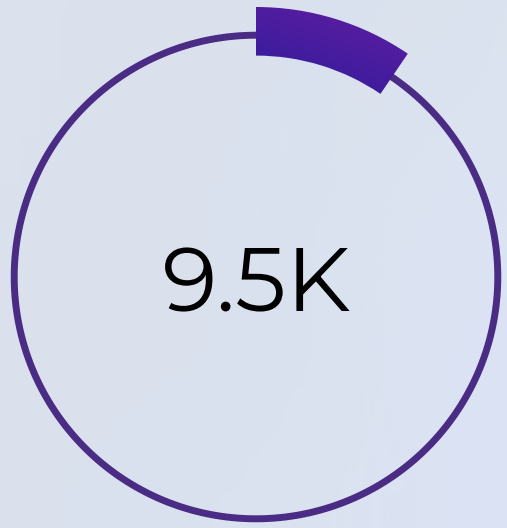
Accessories Revenue

Accessories are a significant revenue driver.



Customer Split

Customers are split evenly between subscribers and non-subscribers.



Adult Sales

Adults contribute the most to overall sales.

The dashboard reveals that Clothing and Accessories are top revenue categories, with Young Adults leading in sales. The even split between subscribers and non-subscribers indicates a strong opportunity for subscription growth.

Business Recommendations

Boost Subscriptions

Promote exclusive benefits to encourage more customers to subscribe and enhance retention.

Customer Loyalty Programs

Implement reward systems for repeat buyers to foster loyalty and move them into the 'Loyal' segment.

Review Discount Policy

Strategically balance sales boosts from discounts with maintaining healthy profit margins.

Product Positioning

Highlight top-rated and best-selling products in marketing campaigns to maximize their visibility and appeal.

Targeted Marketing

Focus marketing efforts on high-revenue age groups and customers who prefer express shipping for optimized impact.