

Online Sales Coupon Analysis in Python - Final Report

Introduction

This project analyzes online sales coupon data to understand customer engagement, coupon effectiveness, and product sales trends. The goal is to derive insights into how coupons impact sales and customer behavior.

Data Preprocessing

Datasets Used:

- train.csv – Contains campaign, customer, and coupon redemption data.
- customer_transaction_data.csv – Records sales transactions, including prices and discounts.
- customer_demographics.csv – Provides demographic details of customers.
- coupon_item_mapping.csv – Maps coupons to items.
- campaign_data.csv – Contains details of marketing campaigns.
- item_data.csv – Lists item brands, types, and categories.

Data Cleaning & Merging Steps:

- Handled missing values in customer demographics and campaign data.
- Merged datasets based on common keys.
- Removed duplicate rows and formatted date columns.
- Filtered out invalid or zero-transaction data.

Exploratory Data Analysis (EDA)

Key Findings:

- Grocery category had the highest engagement.
- The top-selling product was from the Grocery category.
- Year-wise coupon usage analysis showed high engagement in 2013.
- Month-wise sales trends indicated coupon usage significantly impacted sales.

Feature Engineering

Newly Created Features:

- Total Discount Given: Sum of coupon and other discounts.
- Final Price After Discount: Selling price adjusted for total discount.
- Estimated Profit (Assumed 20% margin): Profit estimation based on adjusted final price.

Conclusion & Recommendations

Key Findings:

- Coupons significantly boost sales, especially in Grocery and Pharmaceuticals.
- The impact of discounts is noticeable, as sales increased with coupon usage.
- Customer engagement is highest for daily essential categories.