EDULYT INDIA

PROJECT REPORT

ON

Data-Driven Retail Insights:

Enhancing Customer Profiling and Transaction Efficiency

Submitted By:

Shubham Gupta

in.sg5447@gmail.com

AI/ML

Amity University Noida

Problem Statement

Sanity Checks – Data Cleaning

- Provide a meaningful treatment where the Credit Card entries are blank.
- Identity where Price is equal to Selling Price even after having a Coupon Code, apply an automatic discount of 5% for those entries.
- Make sure that the return date is after the Purchase Date.
- If the Coupon ID is NULL, make sure that NO discount is given, the Selling Price should be equal to Price.
- Age should be greater than 18 for all the CC holders.
- Transaction ID should be unique for all.

Tasks

- Customer Segmentation Based on Spend in Dollars, based on Swipes, segmentation example below.
- Calculate the spend in terms of Product, State and Payment method.
- Calculate the highest 5 spending in all above categories.
- Give your opinion on return category like customers returning the products belongs to which state, age group, condition, category of the product or is it related to discount.
- Create a profile of customers in terms of timing of their order.
- Which payment method is providing more discount for customers?
- Create a profile for high value items vs low value items and relate that wrt to their number of orders.
- Do you think if merchant provides more discount then can it will lead to increase in number of orders?

TOOLS USED: Power BI and Microsoft Excel.

Link:
https://drive.google.com/drive/folders/14YGKWwdwejDXQPvn7S8kqK4XWU8PXm2_?usp=sharing

Dataset Description

Dataset Description

The dataset for this project, sourced from Credit Banking - 3 copy.xls, comprises two Excel sheets: Project_2 and Customer_Info, designed for retail transaction analysis in the BFSI domain.

- **Project_2 (Transactions)**: Contains 10000+ records of retail transactions with 18 columns, including:
  - Credit_card: Customer ID (links to Customer_Info).
  - Product_ID, P_CATEGORY, CONDTION, Brand: Product details.
  - Price, Selling_price, Coupon_ID: Pricing and discount information.
  - Date, Time, Return_date: Transaction and return timestamps.
  - Payment Method, Transaction ID, Return_ind: Payment and return status.
  - Other identifiers: GTIN, MPN, Merchant_name, M_ID.
  - Key features: Captures purchase behavior, discounts, and returns across categories like electronics, shoes, and luggage.

- **Customer_Info (Customers)**: Includes 124 records with 9 columns, detailing:
  - C_ID: Customer ID (matches Credit_card in Project_2).
  - Email, Name, Mobile_number, Gender, Age: Demographic data.
  - City, State, Address: Geographic information (address masked).
  - Key features: Enables demographic segmentation (e.g., age, gender, location).

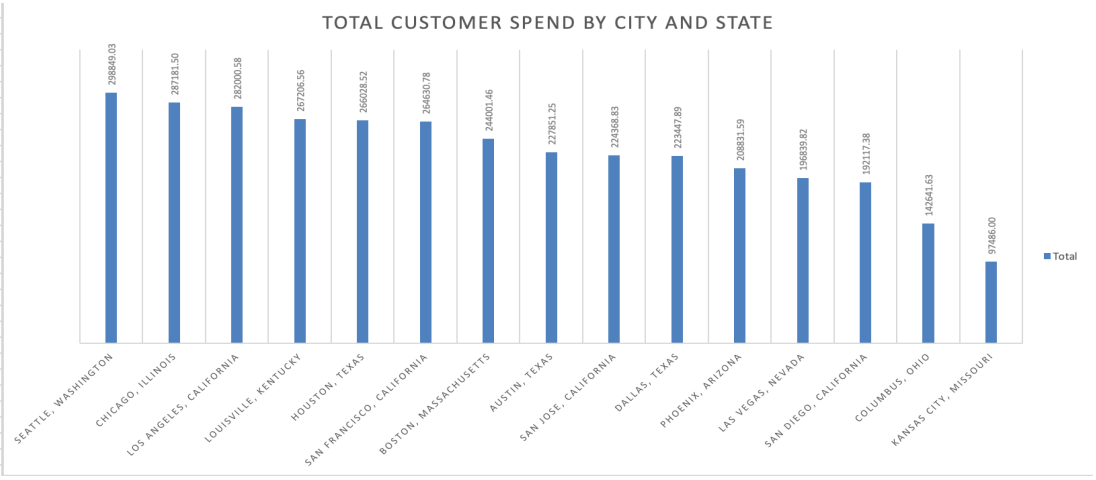
Data Cleaning and Preprocessing

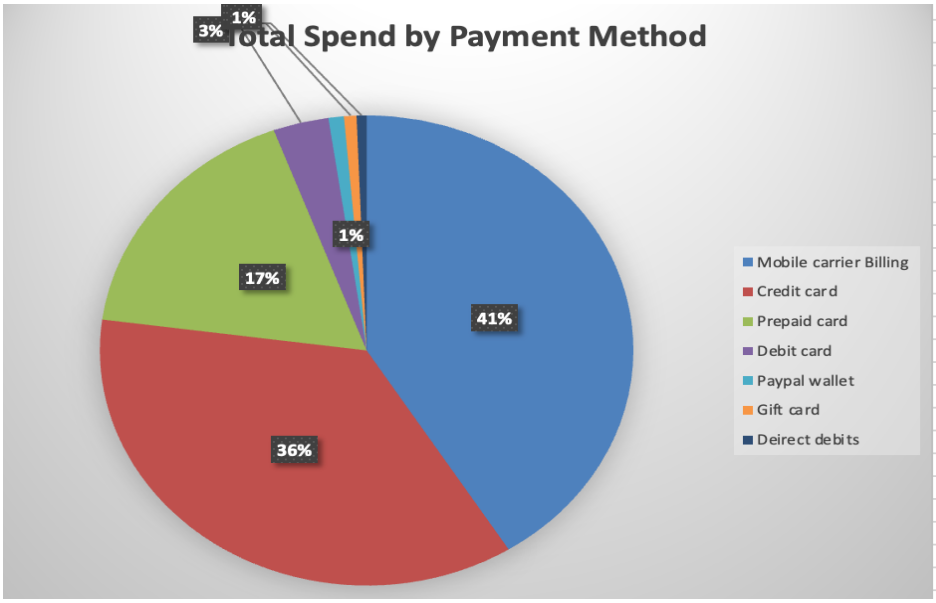
1. **Data Type Validation:**
    - Ensured consistent data types in Project_2:
        - Price, Selling_price, AdjustedPrice, AdjustedSP: Converted to Decimal Number (e.g., $1,484.74).
        - Date, Return_date: Standardized to Date format (YYYY-MM-DD).
        - Credit_card, Coupon_ID, Transaction ID: Set to Text to prevent formatting issues.
    - In Customer_Info, set Age to Whole Number and C_ID to Text.
2. **Handling Duplicates:**
    - Verified Transaction ID uniqueness in Project_2 (162 unique IDs, no duplicates found).
    - Checked for duplicate C_ID in Customer_Info (124 unique IDs, no duplicates).
3. **Merging Datasets:**
    - Performed a left join of Project_2 with Customer_Info on Credit_card = C_ID to enrich transaction data with demographics (e.g., Gender, Age, State).
    - Handled unmatched records (6 blank Credit_card entries) by assigning null values for customer attributes, flagged for further analysis.
4. **Null Value Treatment:**
    - Beyond Coupon_ID and Credit_card, checked for nulls in critical columns (e.g., Price, Selling_price, Date): No missing values found.
    - For optional fields like Return_date (152 nulls, expected for non-returned items), retained nulls as they are valid.
5. **Column Creation for Analysis:**

- Appended AdjustedPrice and AdjustedSP columns to Project_2 using Power Query to support discount and spend analysis.
- Power Query logic:
  - AdjustedPrice: if [Coupon_ID] is null then [Price] else [Selling_price].
  - AdjustedSP: if [Price] = [Selling_price] and [Coupon_ID] is not null then [Selling_price] * 0.95 else [Selling_price].

RESULTS

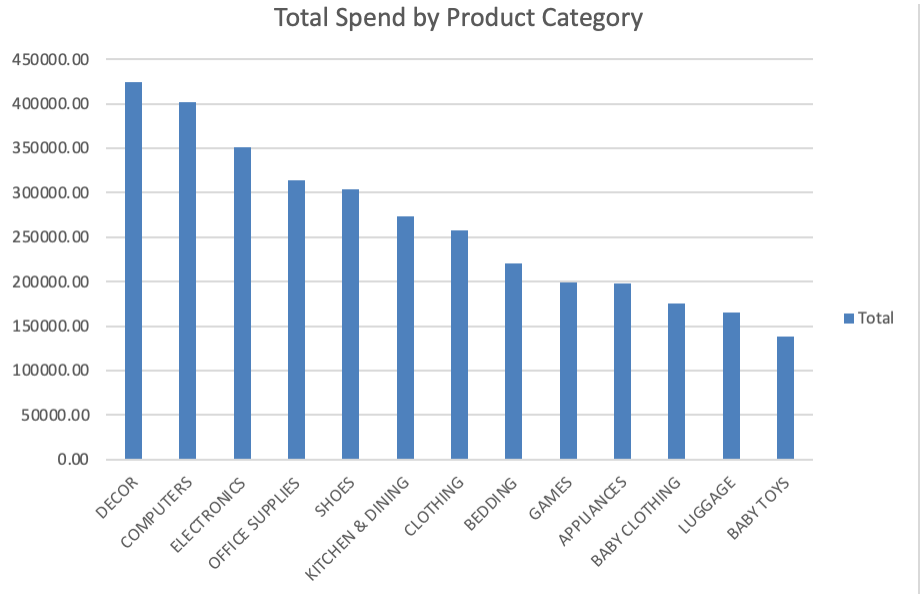
City, State	Sum of Final_SP
Seattle, Washington	298849.03
Chicago, Illinois	287181.50
Los Angeles, California	282000.58
Louisville, Kentucky	267206.56
Houston, Texas	266028.52
San Francisco, Californ	264630.78
Boston, Massachusett	244001.46
Austin, Texas	227851.25
San Jose, California	224368.83
Dallas, Texas	223447.89
Phoenix, Arizona	208831.59
Las Vegas, Nevada	196839.82
San Diego, California	192117.38
Columbus, Ohio	142641.63
Kansas City, Missouri	97486.00
Grand Total	**3423482.83**



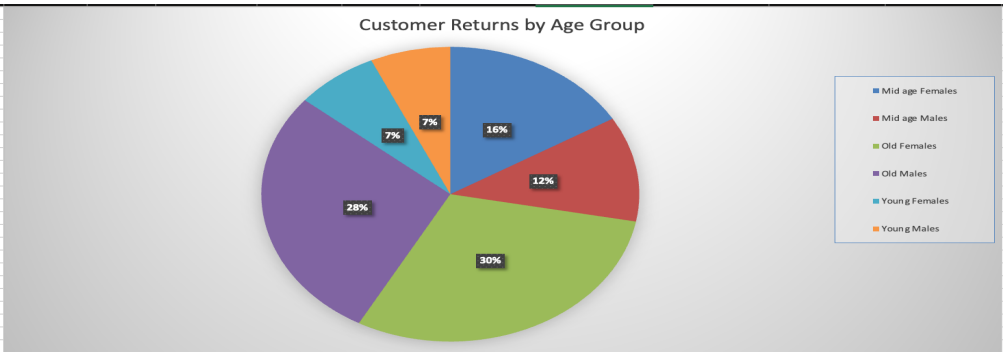


Payment Method	Sum of Final_SP
Mobile carrier Billing	1405743.10
Credit card	1233696.32
Prepaid card	590059.12
Debit card	115366.64
Paypal wallet	31992.28
Gift card	25956.51
Deirect debits	20668.86
Grand Total	**3423482.83**

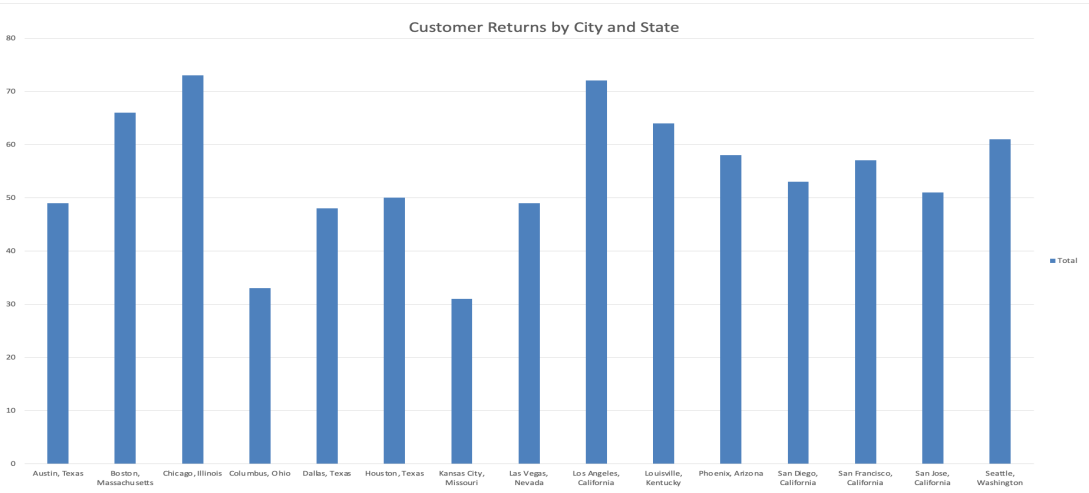
Product Category	Total Amount Spent
DECOR	424417.40
COMPUTERS	401952.39
ELECTRONICS	351174.17
OFFICE SUPPLIES	313715.04
SHOES	304033.81
KITCHEN & DINING	273261.63
CLOTHING	258033.39
BEDDING	220577.77
GAMES	198631.04
APPLIANCES	198110.78
BABY CLOTHING	175766.69
LUGGAGE	165193.82
BABY TOYS	138614.90
Grand Total	**3423482.83**



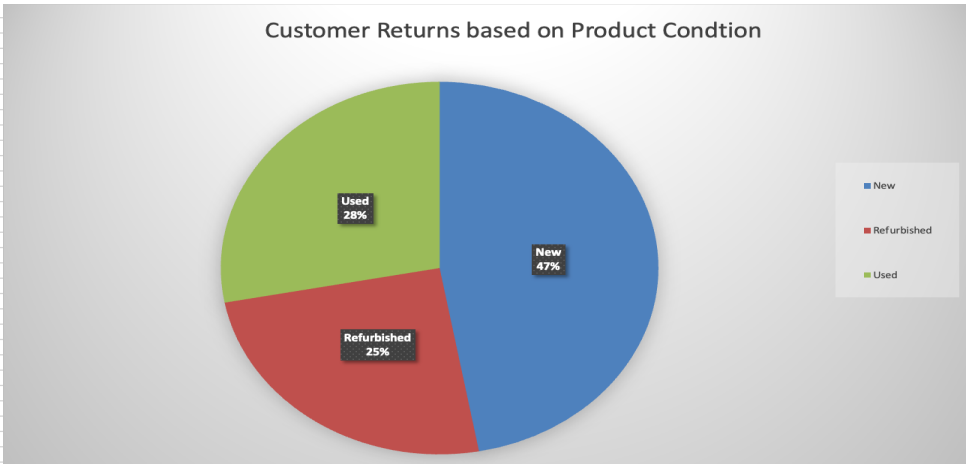
Return_ind	1
Age Group	**Number of Customers**
Mid age Females	134
Mid age Males	94
Old Females	245
Old Males	228
Young Females	58
Young Males	56
Grand Total	**815**



Return_ind	1
City, State	**Number of Customers**
Austin, Texas	49
Boston, Massachusetts	66
Chicago, Illinois	73
Columbus, Ohio	33
Dallas, Texas	48
Houston, Texas	50
Kansas City, Missouri	31
Las Vegas, Nevada	49
Los Angeles, California	72
Louisville, Kentucky	64
Phoenix, Arizona	58
San Diego, California	53
San Francisco, California	57
San Jose, California	51
Seattle, Washington	61
Grand Total	**815**



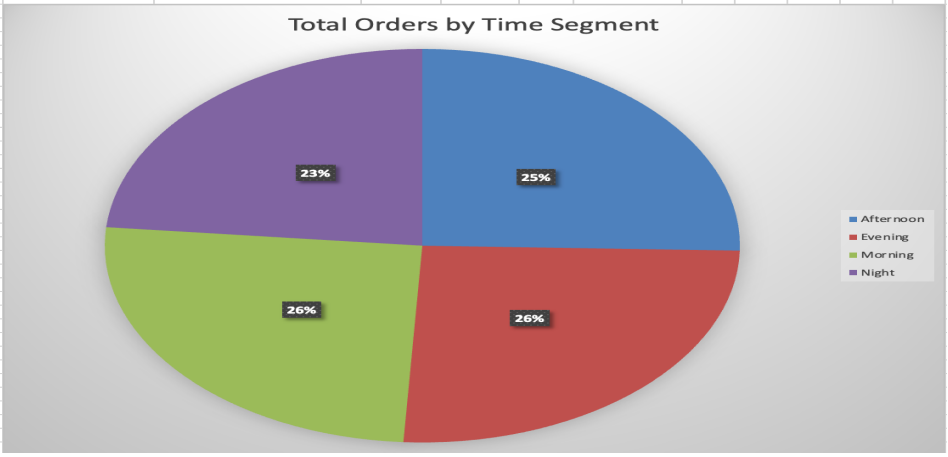
Return_ind	1
Product Condition	**Number of Customers**
New	384
Refurbished	203
Used	228
Grand Total	**815**



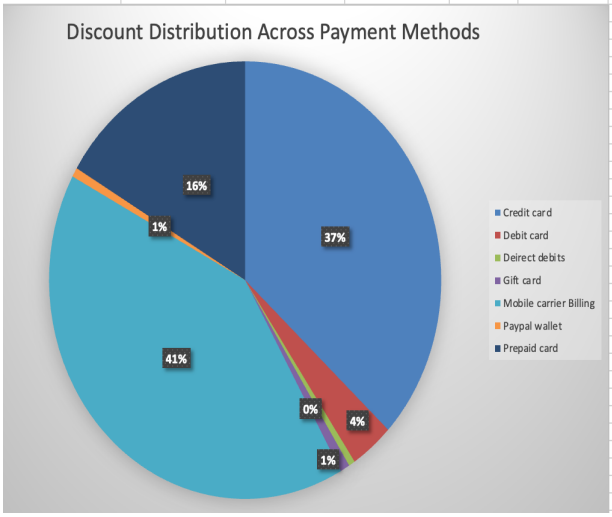
Return_ind	1
Product Category	**Number of Customers**
APPLIANCES	47
BABY CLOTHING	33
BABY TOYS	37
BEDDING	40
CLOTHING	64
COMPUTERS	86
DECOR	102
ELECTRONICS	82
GAMES	67
KITCHEN & DINING	71
LUGGAGE	47
OFFICE SUPPLIES	61
SHOES	78
Grand Total	**815**

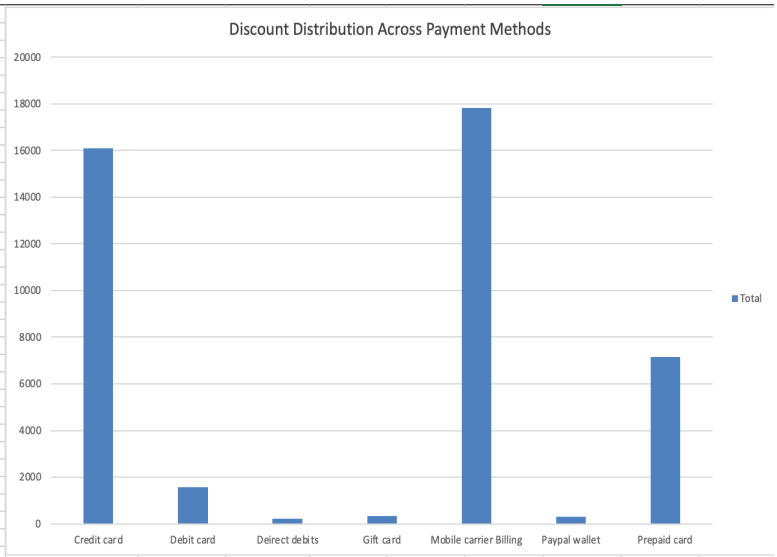
Customer	Count of Transaction ID
1093	4
Afternoon	2
Evening	1
Night	1
1174	7
Evening	1
Morning	4
Night	2
1177	6
Afternoon	1
Evening	1
Morning	3
Night	1
1256	15
Afternoon	4
Evening	4
Morning	4
Night	3
1412	7
Afternoon	2
Evening	1
Morning	3
Night	1
1548	15
Afternoon	6
Evening	5
Morning	1
Night	3
1692	6
Afternoon	1
Evening	2
Morning	2
Night	1
1739	3
Morning	2
Night	1
1779	5
Afternoon	1
Evening	1

TimeSegment	Total Orders
Afternoon	334
Evening	336
Morning	336
Night	309
Grand Total	**1315**



Payment Method	Sum of Discount
Credit card	16081.861
Debit card	1558.018
Deirect debits	227
Gift card	343.7325
Mobile carrier Billing	17810.6985
Paypal wallet	299
Prepaid card	7154.869
Grand Total	**43475.179**





Item Value	Total Value	Number of Orders
High Valued	2613564.87	690
Low Valued Item	853393.14	625
Grand Total	**3466958.01**	**1315**