VISA: Unlocking Customer Insights to Maximize Transaction Value

Introduction:

Visa is a global credit card provider, widely recognized for its secure payment services and strong customer base across the United States. Each day, Visa processes millions of transactions, from routine grocery shopping and fuel purchases to entertainment, travel, and online retail. With a network that spans thousands of merchants and cities, the company holds data on how, where, and when people spend.

As VisaNova enters its next phase of growth, the focus is shifting from acquiring new cardholders to **increasing the value generated from existing customers**. To achieve this, the company wants to better understand customer behavior, spending patterns, and potential risks in order to drive smarter decisions across marketing, risk, and customer experience teams.

The company believes there is significant untapped potential in:

- Understanding which customer segments contribute most to overall transaction value,
- Identifying spending patterns across time, location, and categories,
- Detecting anomalies or early signs of fraud,
- Forecasting future spending behavior based on historical trends,
- And creating tailored strategies for customer engagement, retention, and risk prevention

Objective:

In this project, we aim to analyse Visa data to:

- **Understand the key drivers of Customer Lifetime Value** by breaking it down into the main business objective into smaller, measurable components.
- **Uncover customer spending patterns** through segmentation and behavioural analysis.
- **Investigate** seasonal, geographic, and category-level spending trends.
- **Forecast** transaction behavior using historical data to support planning and future decision-making.

Business Impact:

This analysis helps Visa to:

- **Boost Revenue** by identifying high-value customer segments and top spending categories.
- **Improve Retention** through personalized offers and targeted engagement.
- **Reduce Fraud** with early anomaly detection and risk-based insights.
- **Enhance Planning** via forecasting of customer behavior and seasonal trends.
- **Optimize Experience** by aligning spend patterns with customer preferences.

Dataset Information:

There is one dataset namely

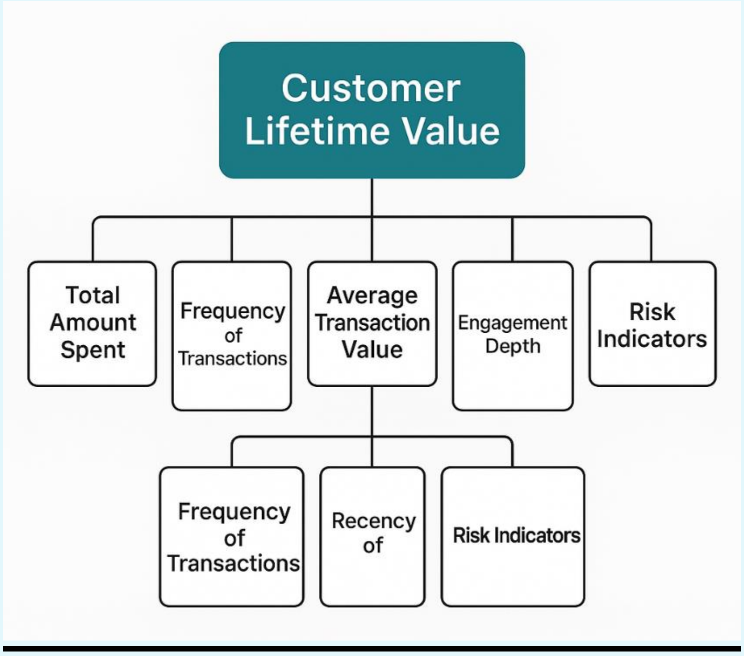
- **Visa:**
    - **Count of Rows:** 1048574
    - **Count of Columns:** 24

Explanation of Data Columns:

- **ID:0 :** Row index or serial number from CSV export.
- **trans_date_trans_time :** Date and time of the transaction in YYYY-MM-DD HH:MM:SS format.

- **cc_num :** Anonymized credit card number used for the transaction.
- **merchant :** Name of the merchant or store where the transaction occurred.
- **category :** Type of transaction or spending category.
- **amt :** Amount of the transaction in USD.
- **first :** First name of the cardholder.
- **last :** Last name of the cardholder.
- **gender :** Gender of the cardholder.
- **street :** Street address of the cardholder.
- **city :** City where the cardholder resides.
- **state :** State where the cardholder resides.
- **zip :** ZIP code of the cardholder's address.
- **lat :** Latitude coordinate of the cardholder's location.
- **long :** Longitude coordinate of the cardholder's location.
- **city_pop :** Population of the city where the cardholder resides.
- **Job :** Occupation of the cardholder.
- **Dob :** Date of birth of the cardholder in YYYY-MM-DD format.
- **trans_num :** Unique alphanumeric identifier for each transaction.
- **unix_time :** Unix timestamp representing the time of the transaction.
- **merch_lat :** Latitude coordinate of the merchant's location.
- **merch_long**: Longitude coordinate of the merchant's location.
- **is_fraud**: Binary indicator of whether the transaction was fraudulent. 1 = fraud, 0 = not fraud
- **merch_zipcode** : ZIP code of the merchant's location

KPI Tree:



Data Cleaning and Processing:

Data cleaning and preparation are crucial steps in ensuring the accuracy and integrity of any data analysis process. For the **Visa: Unlocking Customer Insights to Maximize Transaction Value** project, this step ensures the multi-table retail dataset is reliable, consistent, and ready for meaningful exploration and insights. The dataset includes VISA csv file contains various columns. Below are the key steps undertaken:

- **Data Import:**

The CSV files were imported from a shared Google Drive folder using gdown in Google Collab. The file (e.g.,visa.csv) was loaded into a separate DataFrame for analysis.

- **Missing Values:**

A thorough check for null values across dataset was performed using .isna().sum(). The visa_data.csv file had approximately **19,5973 missing values** in the merch_zipcode column. Since the overall dataset has **1 million+ records**, these rows represent a small proportion and were not removed to maintain data integrity.

● **Duplicate Records:**

The DataFrame was checked for duplicate rows using .duplicated().sum(). No duplicate rows where found.

● **Data Type Verification:**

Using .dtypes, we verified the data types of each column. Certain fields such as trans_date_trans_time, dob were originally in string/object format and were converted to datetime using pd.to_datetime() for time-based analysis.

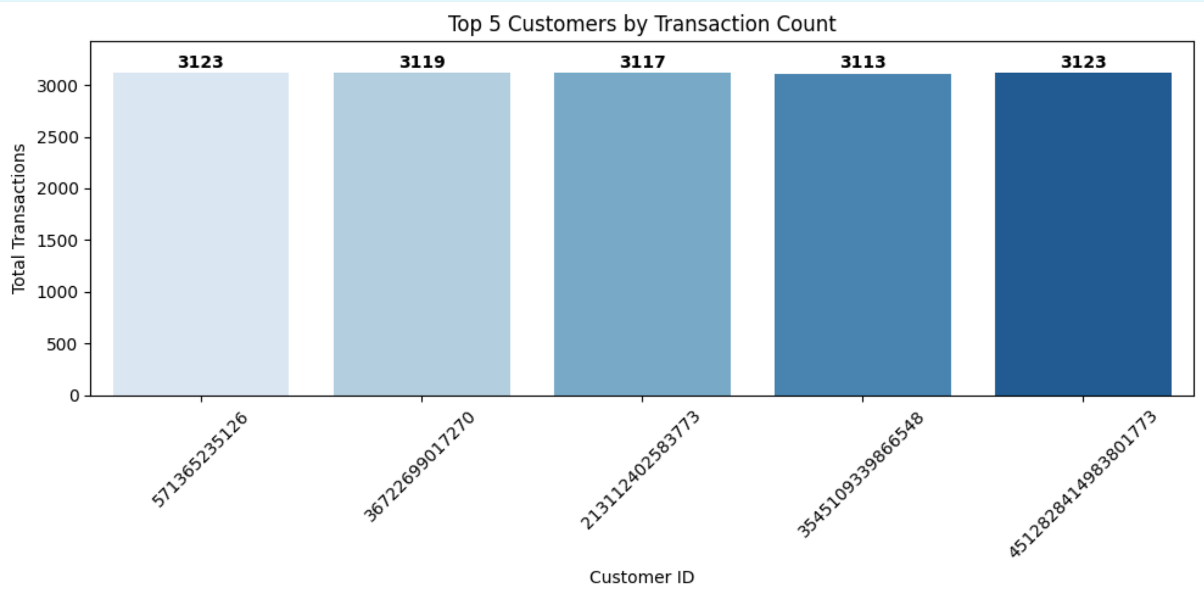
● **Outlier Detection:**

Outliers were examined using the Interquartile Range (IQR) method on key metrics such as Quantity, Price, and Discount. No outliers were found in key metrics.

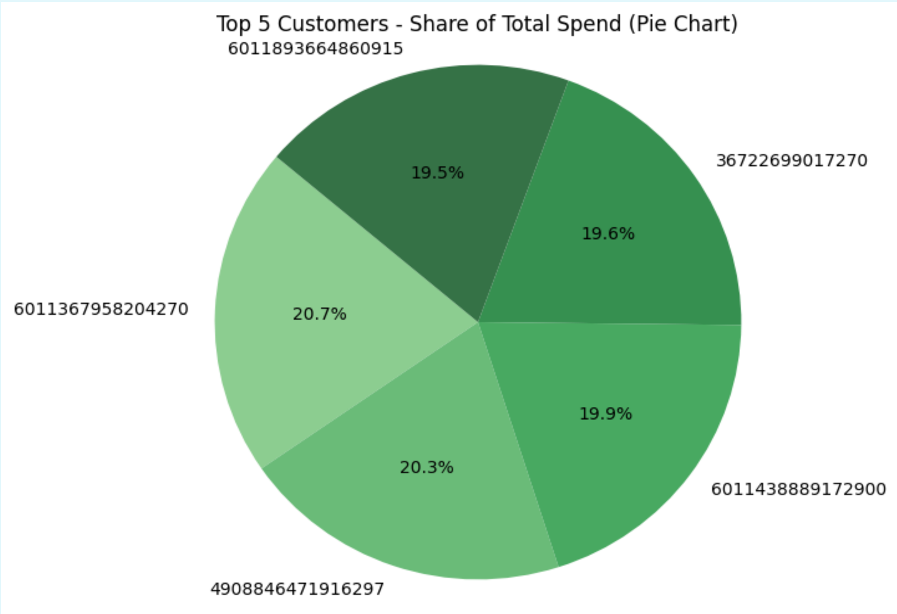
Data Overview:

After cleaning and processing the data, it was reviewed for completeness. After this, key aspects of the dataset were analysed to gain a clearer understanding and a comprehensive overview of the underlying patterns.

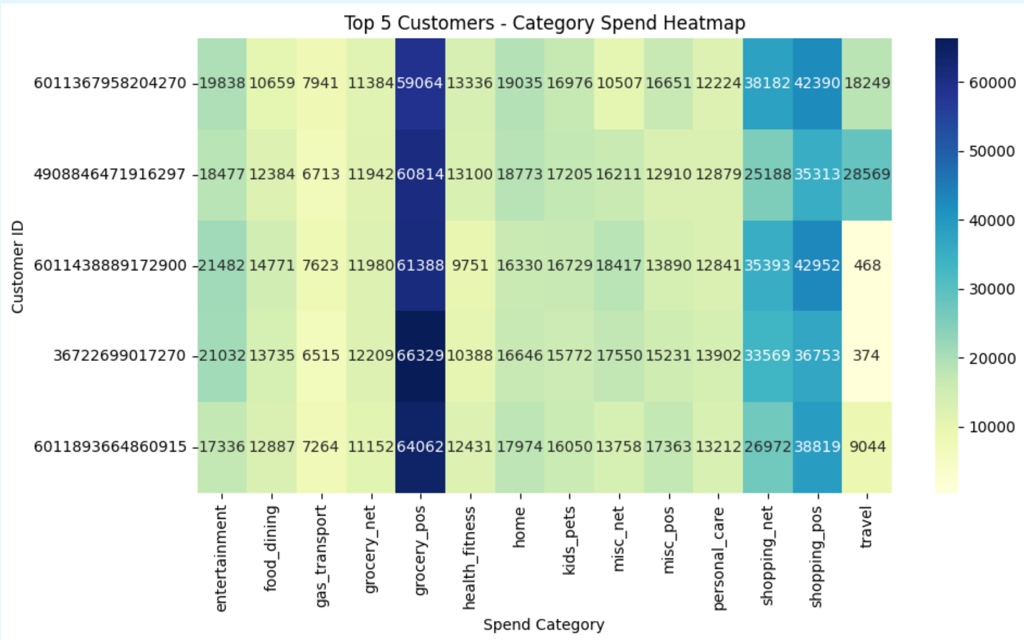
● **Total Transactions for Top 5 Customers**



- **Total Amount Spent by top 5 Customers**

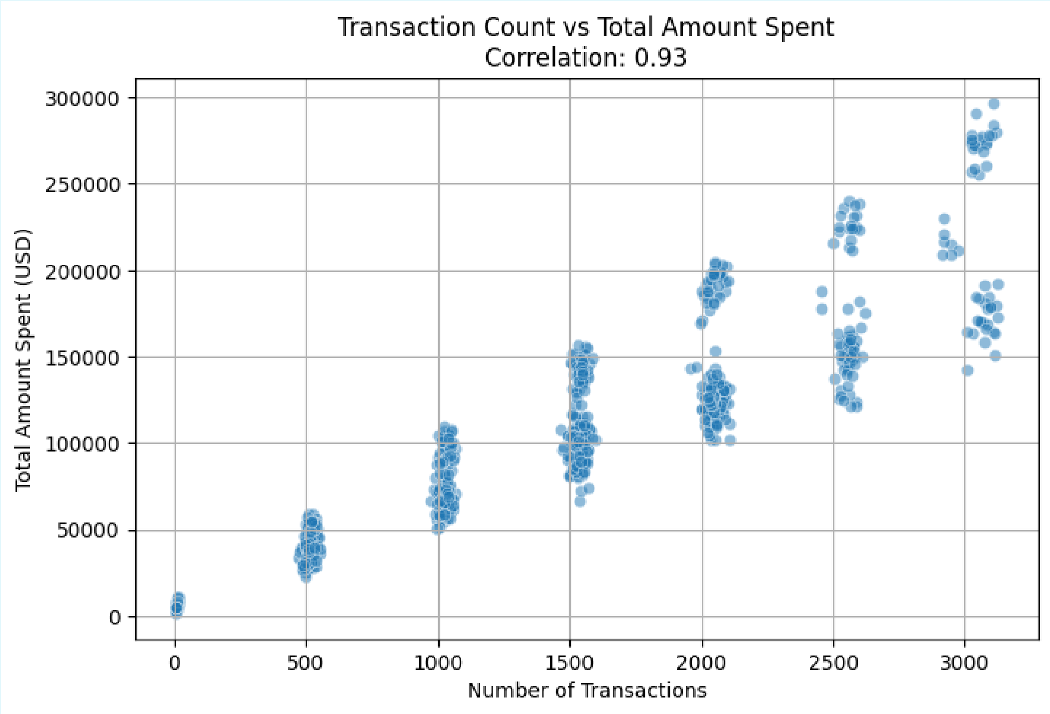


- **Category Spend for Top 5 Customers**



Exploratory Data Analysis:

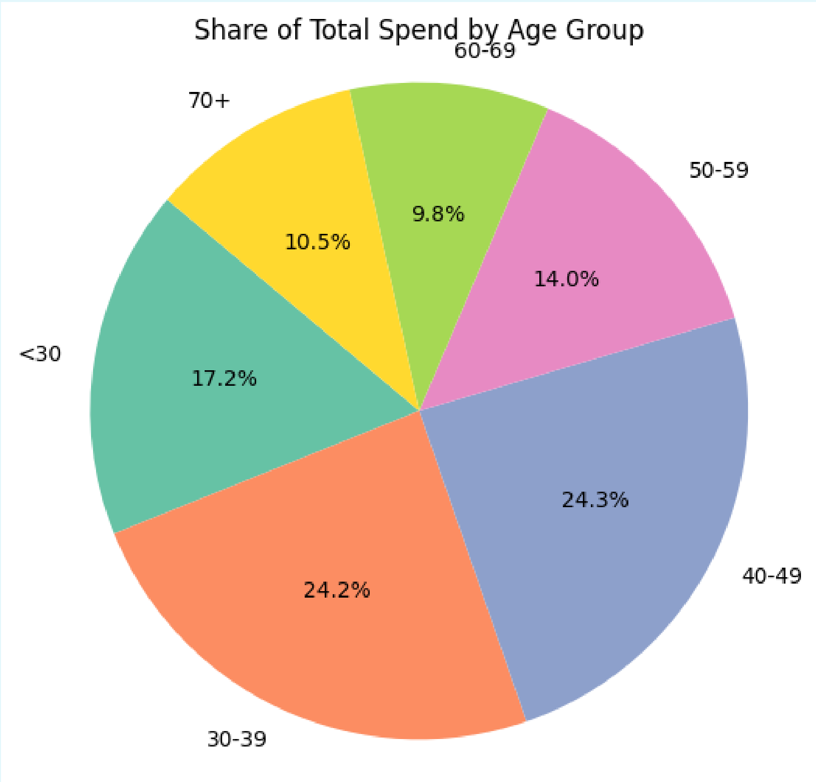
Hypothesis 1: Customers with higher transaction frequency contribute more to overall spend



Analysis: High-frequency customers are most valuable. Focus on retention, rewards, and loyalty campaigns for these segments.

Recommendation: Focus on High-Frequency, High-Spend Customers and Launch loyalty rewards, tiered cashback, and exclusive offers for high-frequency users.

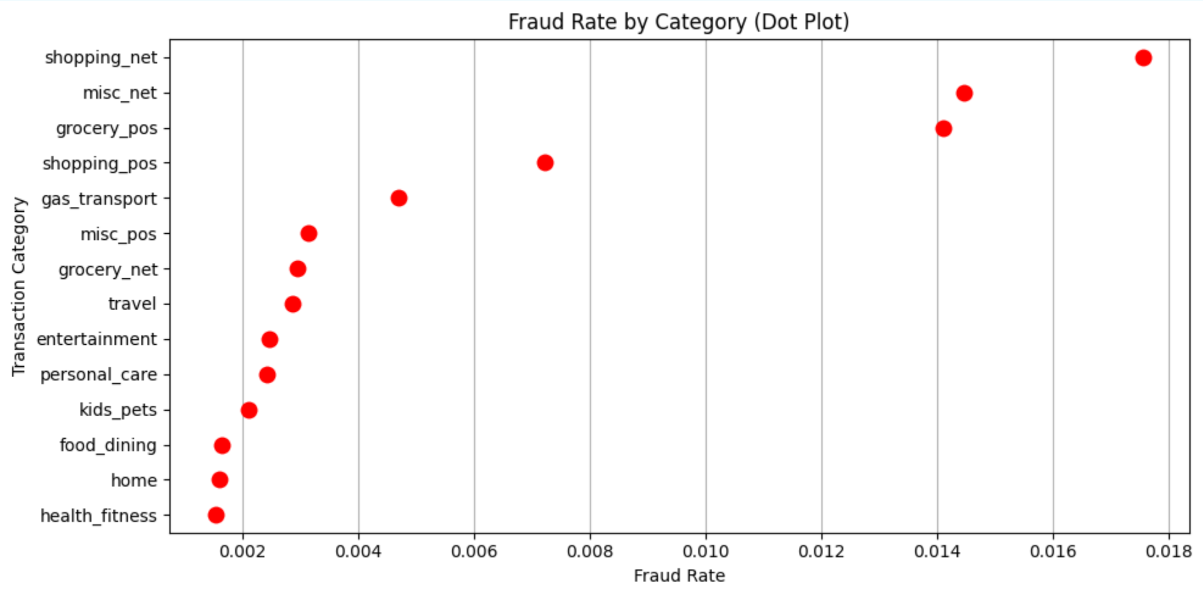
<u>Hypothesis 2:</u> Customers aged 30–50 contribute the most to total transaction value



Analysis: Mid-life customers are the most valuable. They are ideal targets for premium offers, travel cards, and upsell strategies.

Recommendation: Target the 30–50 Age Group with Mid-Life Campaigns: Design campaigns that appeal to professionals, working parents, and commuters and promote travel, finance, and lifestyle benefits to this segment.

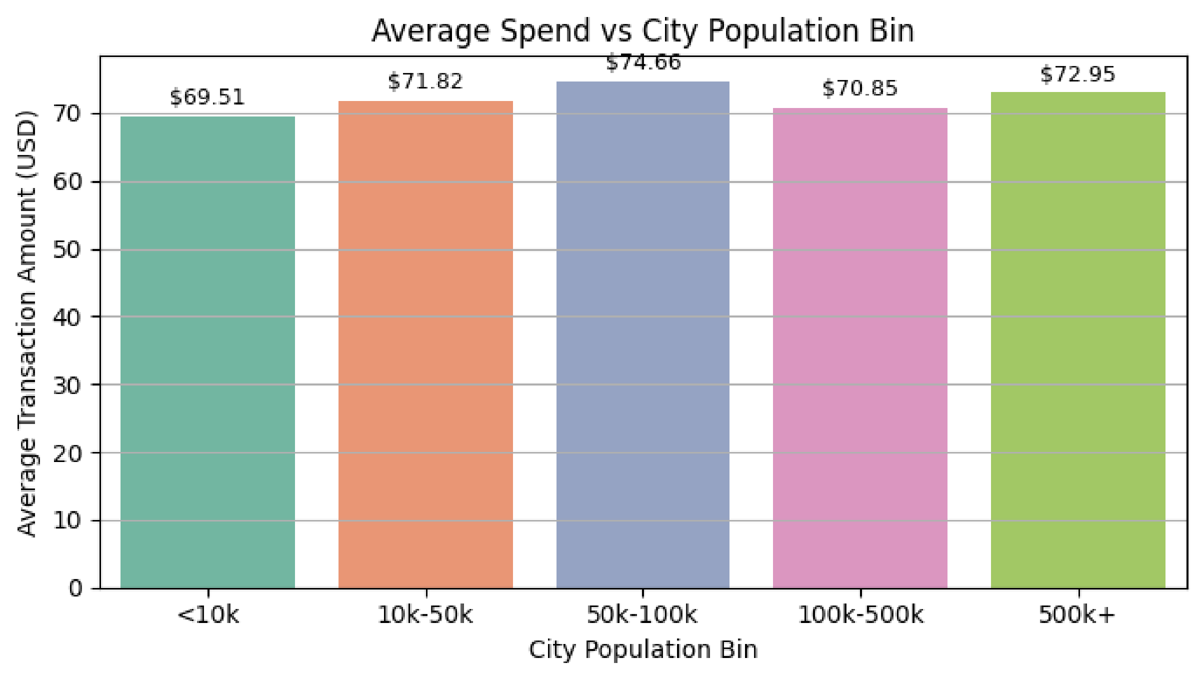
<u>Hypothesis 3:</u> Online and travel-related categories generate higher fraud rates



Analysis: These categories (travel, online shopping) need stronger fraud detection and customer authentication. Risk alerts can reduce fraud exposure.

Recommendation: Prioritize Urban Customers (City Pop > 500K): Geotarget campaigns toward metros and urban hubs and Offer premium card upgrades and high-credit-limit features in these regions.

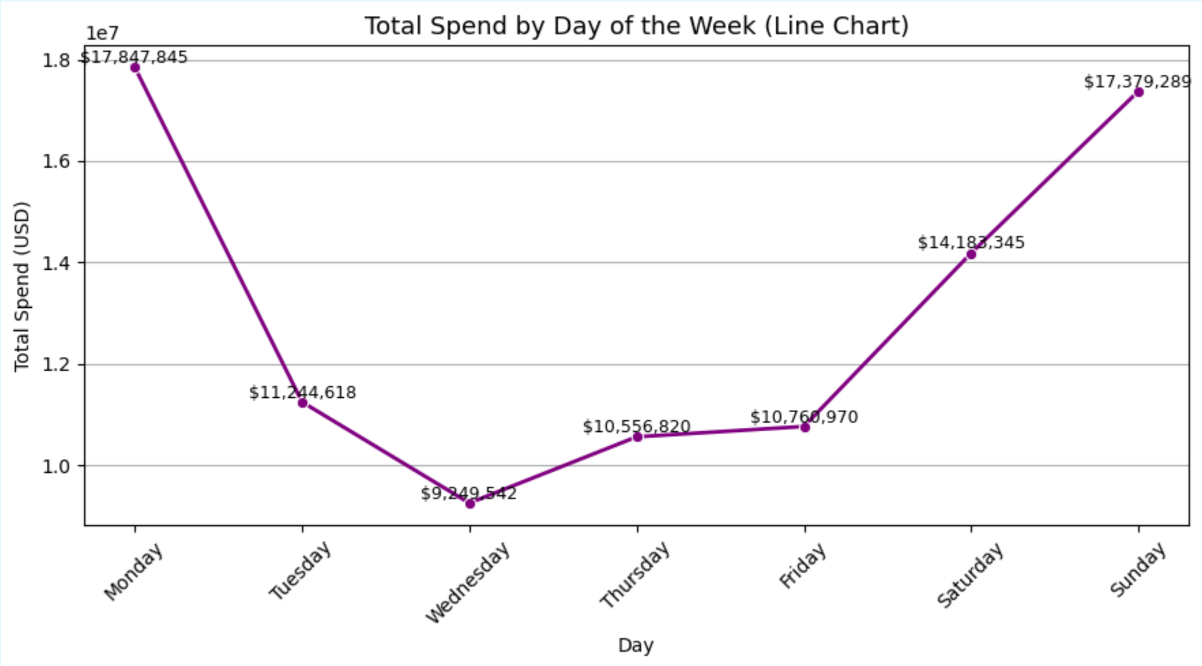
<u>Hypothesis 4:</u> Customers living in high population cities tend to spend more



Analysis: As city population increases, average spend per transaction also increases.

Recommendation: Prioritize Urban Customers (City Pop > 500K): Geotarget campaigns toward metros and urban hubs and Offer premium card upgrades and high-credit-limit features in these regions.

<u>Hypothesis 5:</u> Customers Spend More on Weekends



Analysis: Customers spend more on weekends (Saturday & Sunday), likely driven by leisure, shopping, and entertainment. This confirms that marketing offers, reward campaigns, and promotions should be timed for weekends to maximize impact. But still Monday is on highest peak Monday–Friday for bill payment offers, cashback, and recharge promotions

Recommendation: Schedule Offers and Communications for Weekdays: Launch bill-pay rewards, EMI offers, or cashback campaigns early in the week and Use transaction reminders on Sunday evenings to drive Monday activity.

Closing Recommendation:

- Prioritize high-frequency customers with loyalty tiers and personalized retention offers.
- Target 30–50 age group with mid-life benefits like insurance, EMI, and travel perks.
- Focus on urban customers (pop > 500k) through location-based premium campaigns.
- Time offers around Monday, leveraging peak weekday spending for higher engagement.
- Segment users using RFM modeling and tailor outreach by CLV segment.
- Enhance fraud detection in travel, online_pos, and entertainment categories.
- Use correlation insights (txn_count, city_pop, age) to boost modeling accuracy.
- Personalize campaigns by job and behavior to drive engagement and relevance.

Limitations of the Analysis:

1. **Limited Customer Demographics** - Apart from age, gender, and job, there's no data on income, education, marital status, etc., which could improve segmentation.
2. **Fraud Labels Are Binary and Static** - Fraud analysis is limited by the binary nature of the is_fraud flag — we lack detailed fraud type, severity, or resolution status.
3. **No Marketing Attribution -** We cannot link transactions back to marketing campaigns or promotions to measure campaign ROI or offer effectiveness.
4. **Time Period Unknown -** Without knowing the full time range of the dataset (e.g., 1 month vs. 1 year), it's hard to project trends or model seasonality accurately.

Tools and Libraries:

1. **Libraries Used**: NumPy, Pandas, Matplotlib, Seaborn
2. **Environment**: Google Collab for data processing and analysis

*********End*********