# CREDIT CARD DATA ANALYSIS

## PROJECT WORK

<u>*Name*</u> *:  SANDE BHUVANACHANDRA YADAV*

<u>*Project Name*</u> *: Credit card data analysis(using python)*

## Project Summary :

This project analyzes credit card transaction data to uncover trends, derive actionable insights into customer spending behaviors.  Using the Python's powerful data manipulation and visualization libraries, key patterns were identified in transaction, customer segmentation.

## Objectives:

- Perform exploratory data analysis (EDA) on credit card transactions
- Understand customer spending habits and transaction frequency
- Identify patterns related to fraudulent activities

## Key Findings:

- Transaction Trends:   Peak transaction periods and commonly used merchant categories
- Customer Behavior:  High-spending vs Low-spending customer groups based on their Age groups and their purchasing patterns
- Data Quality Issues: Handling missing values, duplicate transactions, and inconsistencies.
- Visualization Insights:  Bar charts and histograms revealing spending distributions

## Conclusion:

This analysis provides insights into credit card transaction patterns, customer behaviors, and potential fraud risks. By leveraging Python for data manipulation and visualization, we uncovered trends that can guide strategic decision-making in financial operations.

# Data Description

### Source of data:

It is a public dataset, which is freely available on Kaggle.com

### Structure of the dataset:

The dataset contains Three tables

1. Customer Acquisition cost

2. Spend cost

3. Repayment cost

**In Customer Acquisition cost contains 8 columns & 100 row**

Shape (100,8)

Sl.No,

Customer,

Age,

City,

Product,

Limit,

Company,

Segment

**In Spend DataFrame contains 1500 rows & 7 columns**

*shape (1500,7)*

Sl.no,

Customer,

Month,

Type,

Amount,

Month_name,

Year

**In Repayment DataFrame contains 1523 rows & 6 columns**

*shape(1523 & 6)*

Sl.no,

Customer,

Month,

Amount,

Unnamed: 4

**Data preprocessing steps:**

- Finding null values
- Handling null values
- Columns names
- Drop unnecessary columns


**Data Cleaning steps:**

- Checking column data types
- Converting data types, based on our required format

# *Exploratory Data Analysis*

- Summary statistics
- Missing Data handling
- Trends and Patterns observed in data
- Key findings from data and project it through visual representations

*Challenges & Solutions*

1. *Handling Large datasets*

   *Challenge:* For this project, transactions generate massive amounts of data, making processing slow and memory-intensive

   *Solution:* Accessing libraries like pandas with optimized functions, leverage chunk processing to handle large files.

2. *Missing or Inconsistent Data*

   *Challenge:* Some transaction records may be incomplete or have errors, leading to inaccurate insights.

   *Solution:* Applying data cleaning techniques like filling missing values, dropping incomplete rows, and ensuring standardization in formatting (dates –object data type to datetime datatype).

3. *Effective Data visualization:*
   *Challenge:* Presenting insights in clear, impactful way can be challenging, especially for large datasets.

   *Solutions:* use Seaborn, and Matplotlib for interactive, well-structured plots to reveal trends in an intuitive way.

*Pandas functions for this Project:*

- .head()
- .shape()
- .describe()
- .info()
- .groupby()
- .pivot_table()
- .merge()
- .to_datetime()
- .strftime
- .loc[]
- .plot(kind = ' ')
- User defined functions
- .reset_index()
- .sort_values()

***Approaches to solve business problems:***

1. For missing values and age below 18:
    imputed those values/records with mean of their column
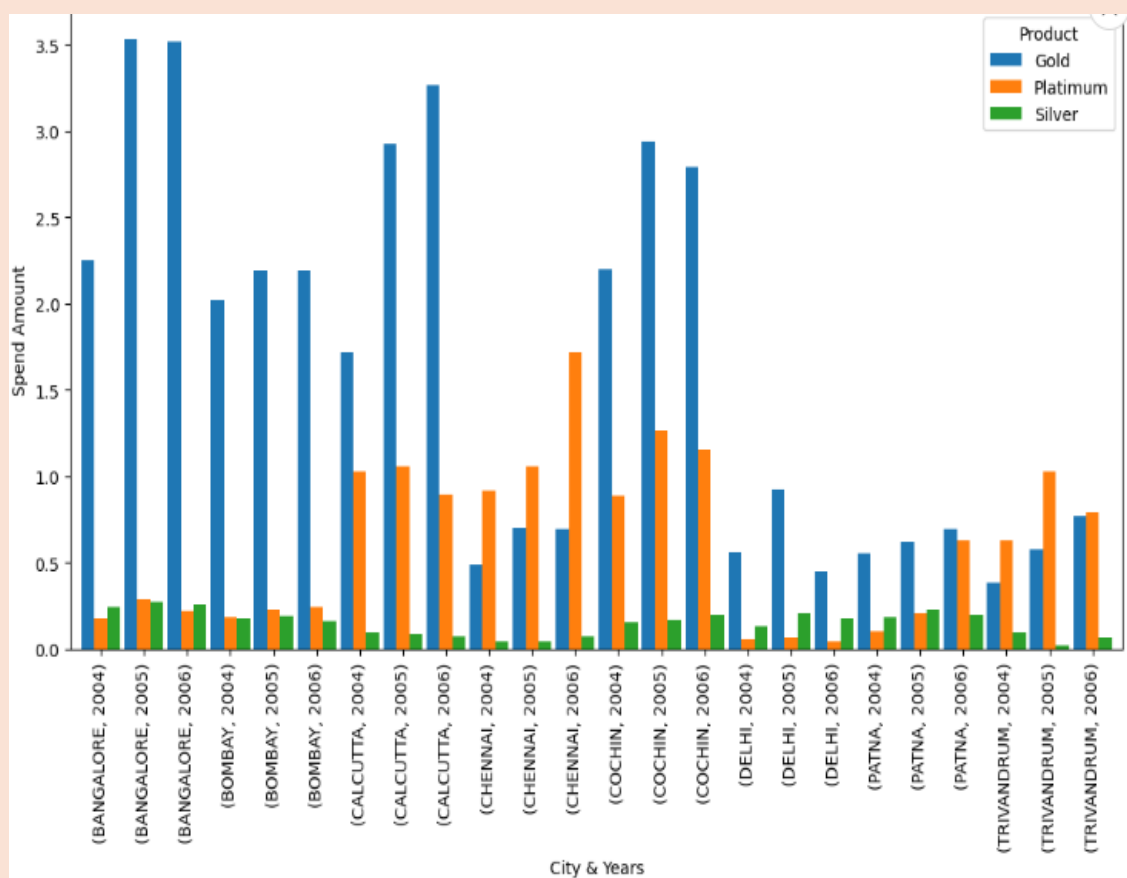
2. For excess of Spend amount:

imputed those records with 50% of customers limit amount

3. If Repayment more than limit amount

 imputed those records with 50% of customers limit amount

# <u>*Key Insights*</u>

The following are the some different behavior based on Products over period of time, on different cities.



From the above visual representation, we conclude, majority amount spent for Gold over a period of time.

# Total spend on Monthly wise over different cities



Monthly comparison of total spends, city wise
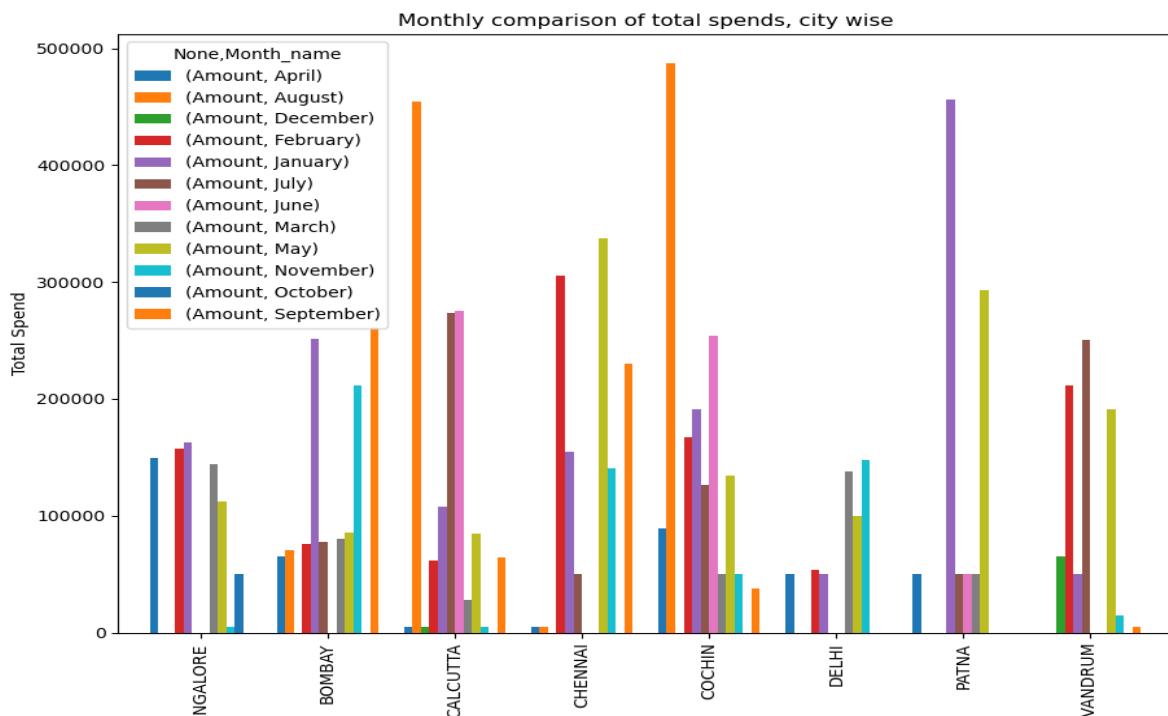
## Comparision of Total Spend Amount on different years on Air Tickets

We clearly observe 2005 has the year where majority of customers spend most of the amount on purchasing air tickets.
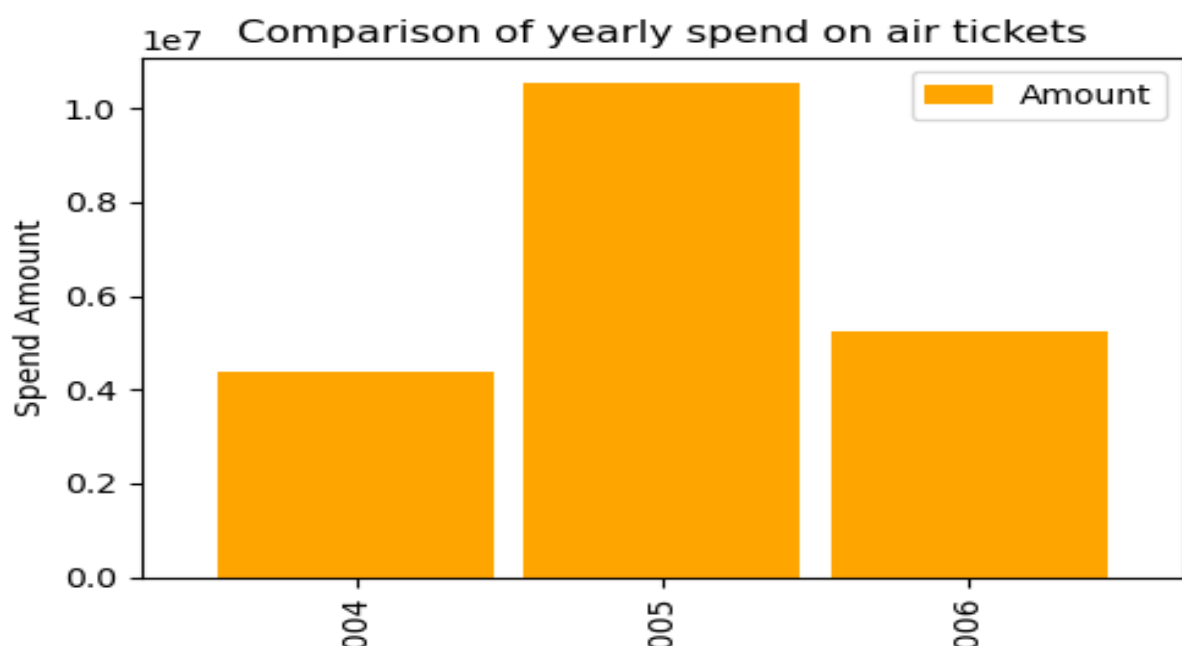
# Comparison of monthly spend for each product (look for any seasonality that exists in terms of spend)

There are Three products which are Gold, Platinum, Silver

From the below graphical representation, we clearly obeserve

In the Month of January the spend amount was high when compard to other products,

- January is the new month for every year,
- January is the festival Season, and majority companies will give offers,
- So January will always lead on top interms of spend
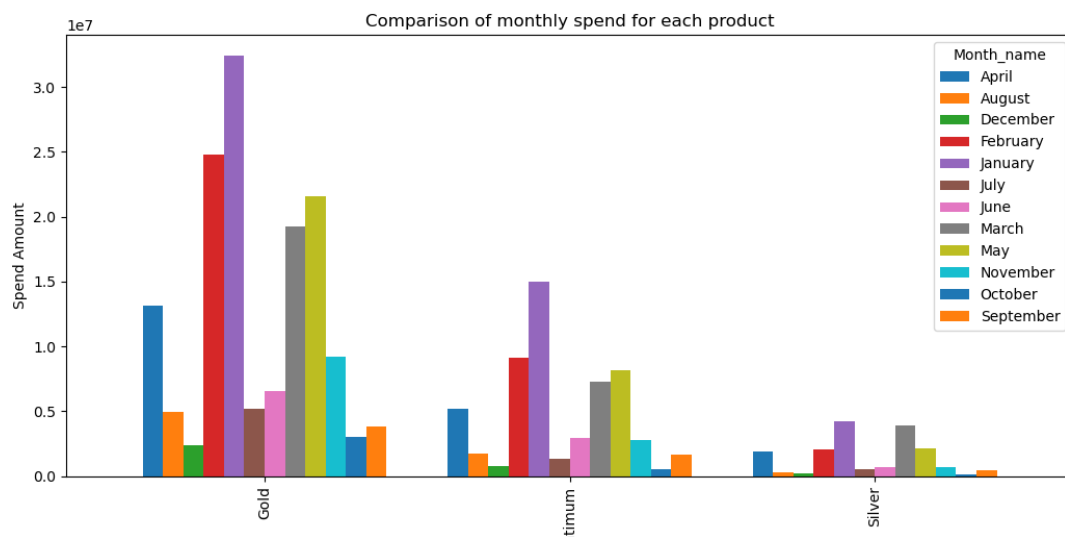


print(" End of the Report")

print(" This is the report generated by SANDE BHUVANACHANDRA YADAV")

print("Thank you")