EDA OF BANKING RISK ANALYSIS USING AI

Chatgpt prompt to perform eda:

Generate Python code using pandas and matplotlib for basic EDA on a dataset named 'banking.csv'. Include data summary, missing value detection, correlation matrix, and two visualizations.

RESULTS :

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load dataset
df = pd.read_csv("Banking.csv")

# 1. Data Summary
print("🔍 Data Summary:\n")
print(df.describe(include='all'))

# 2. Missing Values
print("\n🚨 Missing Values:\n")
print(df.isnull().sum())

# 3. Correlation Matrix (numerical only)
print("\n📈 Correlation Matrix:\n")
correlation = df.corr(numeric_only=True)
```

```
# Plot the correlation heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(correlation, annot=True, cmap="coolwarm", linewidths=0.5)
plt.title("Correlation Matrix")
plt.show()

# 4. Visualization 1: Age Distribution
plt.figure(figsize=(10, 6))
plt.hist(df['Age'], bins=20, color='skyblue', edgecolor='black')
plt.title("Age Distribution")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.grid(True)
plt.show()

# 5. Visualization 2: Estimated Income vs. Credit Card Balance (Scatter Plot)
plt.figure(figsize=(10, 6))
plt.scatter(df['Estimated Income'], df['Credit Card Balance'], alpha=0.5, color='green')
plt.title("Estimated Income vs Credit Card Balance (Scatter Plot)")
```

```
# 6. Visualization 3: Bank Loans Distribution
plt.figure(figsize=(10, 6))
sns.histplot(df['Bank Loans'], bins=20, kde=True, color='salmon')
plt.title("Distribution of Bank Loans")
plt.xlabel("Bank Loans")
plt.ylabel("Frequency")
plt.grid(True)
plt.show()

# 7. Visualization 4: Properties Owned Distribution
plt.figure(figsize=(8, 6))
sns.countplot(x='Properties Owned', data=df, palette='viridis')
plt.title("Distribution of Properties Owned")
plt.xlabel("Number of Properties Owned")
plt.ylabel("Count")
plt.grid(axis='y')
plt.show()
```

```
# 8. Visualization 5: Gender Distribution
plt.figure(figsize=(6, 6))
df['GenderId'].value_counts().plot(kind='pie', autopct='%1.1f%%', colors=['lightblue', 'lightcoral'])
plt.title("Gender Distribution")
plt.ylabel("")
plt.show()
```

```
🔍 Data Summary:

       Client ID              Name          Age   Location ID Joined Bank  \
count       3000              3000  3000.000000   3000.000000        3000
unique      2940              2913          NaN           NaN        2579
top     IND48103  Jeremy Alexander          NaN           NaN  17-03-2019
freq           3                 2          NaN           NaN           4
mean         NaN               NaN    51.039667  21563.323000         NaN
std          NaN               NaN    19.854760  12462.273017         NaN
min          NaN               NaN    17.000000     12.000000         NaN
25%          NaN               NaN    34.000000  10803.500000         NaN
50%          NaN               NaN    51.000000  21129.500000         NaN
75%          NaN               NaN    69.000000  32054.500000         NaN
max          NaN               NaN    85.000000  43369.000000         NaN

       Banking Contact Nationality           Occupation Fee Structure  \
count             3000        3000                 3000          3000
unique              49           5                  195             3
top      Ernest Rivera    European  Associate Professor          High
freq                77        1309                   28          1476
mean               NaN         NaN                  NaN           NaN
std                NaN         NaN                  NaN           NaN
min                NaN         NaN                  NaN           NaN
25%                NaN         NaN                  NaN           NaN
50%                NaN         NaN                  NaN           NaN
75%                NaN         NaN                  NaN           NaN
max                NaN         NaN                  NaN           NaN
```

```
        Loyalty Classification  ...  Bank Deposits  Checking Accounts  \
count                     3000  ...   3.000000e+03       3.000000e+03
unique                       4  ...            NaN                NaN
top                       Jade  ...            NaN                NaN
freq                      1331  ...            NaN                NaN
mean                       NaN  ...   6.715602e+05       3.210929e+05
std                        NaN  ...   6.457169e+05       2.820796e+05
min                        NaN  ...   0.000000e+00       0.000000e+00
25%                        NaN  ...   2.044004e+05       1.199475e+05
50%                        NaN  ...   4.633165e+05       2.428157e+05
75%                        NaN  ...   9.427546e+05       4.348749e+05
max                        NaN  ...   3.890598e+06       1.969923e+06
```

```
max                        NaN  ...   3.890598e+06       1.969923e+06

        Saving Accounts  Foreign Currency Account  Business Lending  \
count      3.000000e+03               3000.000000      3.000000e+03
unique              NaN                       NaN               NaN
top                 NaN                       NaN               NaN
freq                NaN                       NaN               NaN
mean       2.329084e+05              29883.529993      8.667598e+05
std        2.300078e+05              23109.924010      6.412303e+05
min        0.000000e+00                 45.000000      0.000000e+00
25%        7.479440e+04              11916.542500      3.748251e+05
50%        1.640866e+05              24341.190000      7.113147e+05
75%        3.155750e+05              41966.392500      1.185110e+06
max        1.724118e+06             124704.870000      3.825962e+06
```

```
🚨 Missing Values:

 Client ID                    0
 Name                         0
 Age                          0
 Location ID                  0
 Joined Bank                  0
 Banking Contact              0
 Nationality                  0
 Occupation                   0
 Fee Structure                0
 Loyalty Classification       0
 Estimated Income             0
 Superannuation Savings       0
```

```
 Amount of Credit Cards       0
 Credit Card Balance          0
 Bank Loans                   0
 Bank Deposits                0
 Checking Accounts            0
 Saving Accounts              0
 Foreign Currency Account     0
 Business Lending             0
 Properties Owned             0
 Risk Weighting               0
 BRId                         0
 GenderId                     0
 IAId                         0
 dtype: int64
```

```
Correlation Matrix:

                          Age  Location ID  Estimated Income  \
Age                        1.000000    -0.007763         -0.001682
Location ID               -0.007763     1.000000         -0.014235
Estimated Income          -0.001682    -0.014235          1.000000
Superannuation Savings    -0.023504    -0.002113          0.374802
Amount of Credit Cards    -0.004275    -0.010115         -0.038399
Credit Card Balance        0.003431    -0.015764          0.298527
Bank Loans                 0.004773    -0.019554          0.329926
Bank Deposits             -0.010725    -0.032282          0.260332
Checking Accounts         -0.002896    -0.030547          0.291412
Saving Accounts            0.001205     0.003305          0.261299
Foreign Currency Account -0.024935    -0.011679          0.306999
Business Lending           0.000129     0.000477          0.328531
```

```
Saving Accounts            0.001205     0.003305          0.261299
Foreign Currency Account -0.024935    -0.011679          0.306999
Business Lending           0.000129     0.000477          0.328531
Properties Owned           0.002229    -0.035176         -0.008175
Risk Weighting            -0.001198     0.007579          0.664726
BRId                      -0.011292     0.013299          0.018563
GenderId                  -0.012727     0.025892         -0.036112
IAId                      -0.005061     0.000232         -0.009528

                          Superannuation Savings  Amount of Credit Cards
Age                                    -0.023504               -0.004275
Location ID                            -0.002113               -0.010115
Estimated Income                        0.374802               -0.038399
Superannuation Savings                  1.000000               -0.039416
```

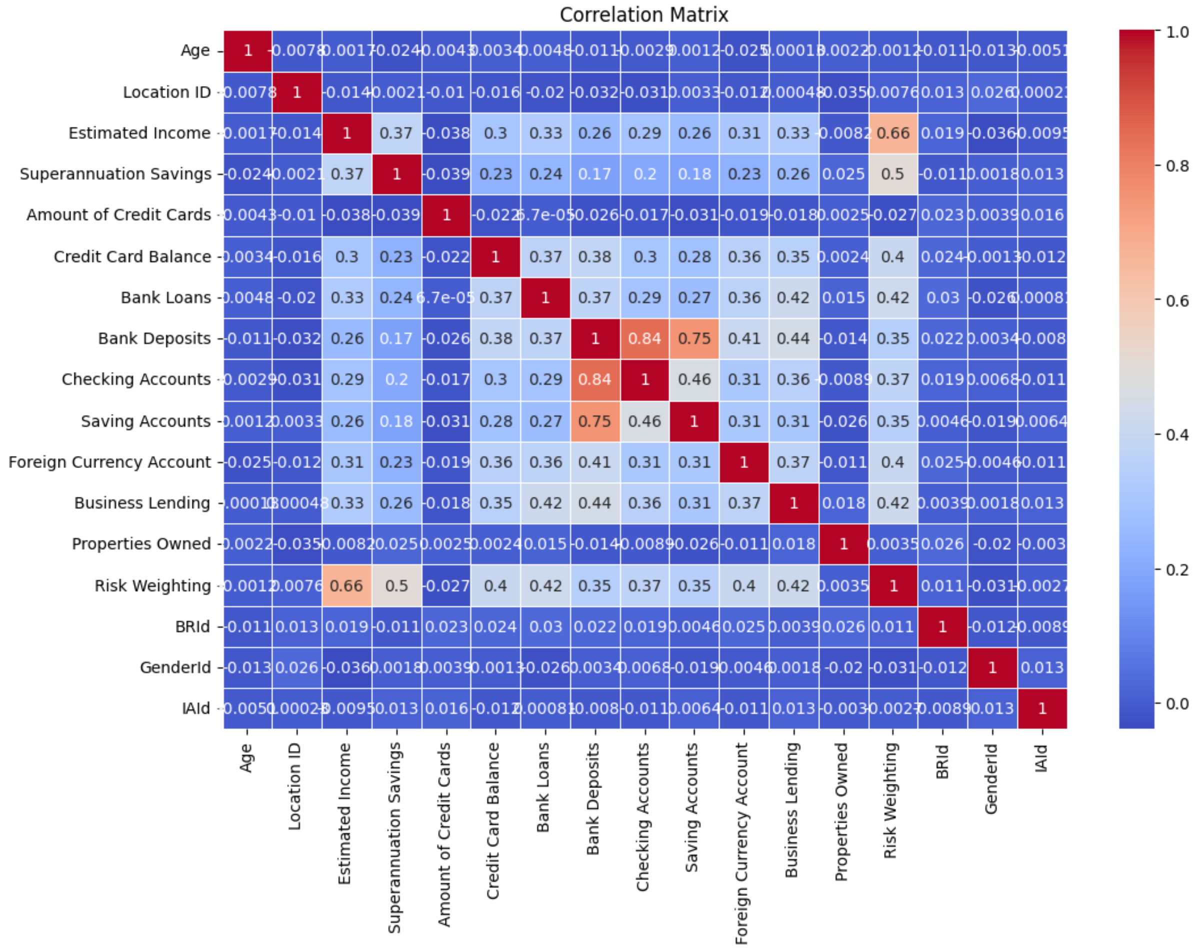
```
Age                                    -0.023504               -0.004275
Location ID                            -0.002113               -0.010115
Estimated Income                        0.374802               -0.038399
Superannuation Savings                  1.000000               -0.039416
Amount of Credit Cards                 -0.039416                1.000000
Credit Card Balance                     0.225562               -0.021644
Bank Loans                              0.241457                0.000067
Bank Deposits                           0.174084               -0.025614
Checking Accounts                       0.198188               -0.017295
Saving Accounts                         0.177132               -0.031438
Foreign Currency Account                0.228103               -0.018598
Business Lending                        0.264919               -0.018311
Properties Owned                        0.024669                0.002476
Risk Weighting                          0.499640               -0.027378
BRId                                   -0.010602                0.022977
```

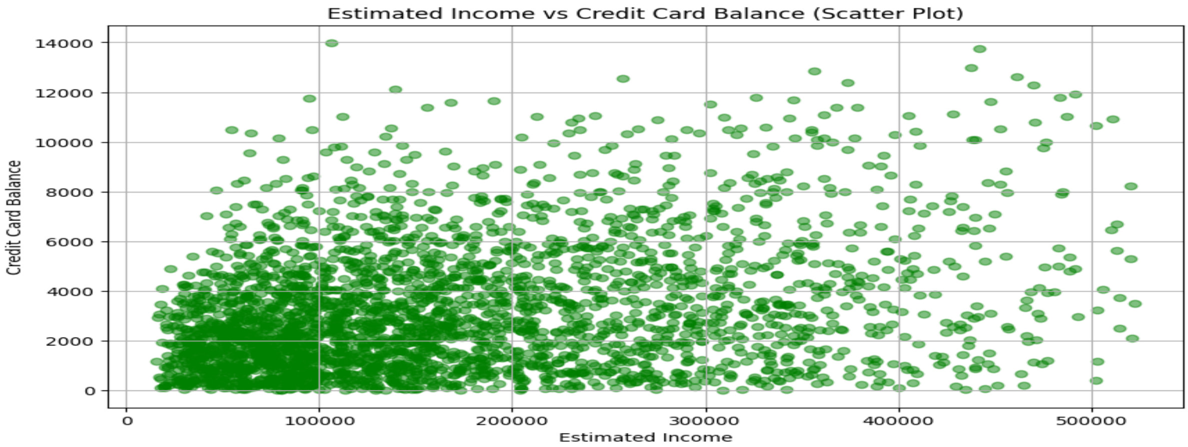
```
                          Credit Card Balance  Bank Loans  Bank Deposits  \
Age                                  0.003431    0.004773      -0.010725
Location ID                         -0.015764   -0.019554      -0.032282
Estimated Income                     0.298527    0.329926       0.260332
Superannuation Savings               0.225562    0.241457       0.174084
Amount of Credit Cards              -0.021644    0.000067      -0.025614
Credit Card Balance                  1.000000    0.369509       0.383877
Bank Loans                           0.369509    1.000000       0.373155
Bank Deposits                        0.383877    0.373155       1.000000
Checking Accounts                    0.298672    0.292082       0.844278
Saving Accounts                      0.284696    0.268882       0.754744
Foreign Currency Account             0.357720    0.364391       0.406347
Business Lending                     0.351063    0.417095       0.441298
Properties Owned                     0.002354    0.014687      -0.013553
```

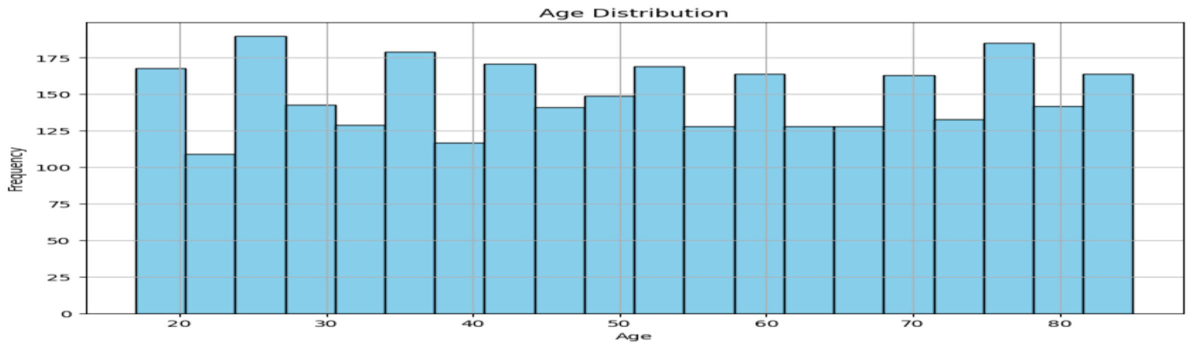
```
                          Checking Accounts  Saving Accounts  \
Age                               -0.002896         0.001205
Location ID                       -0.030547         0.003305
Estimated Income                   0.291412         0.261299
Superannuation Savings             0.198188         0.177132
Amount of Credit Cards            -0.017295        -0.031438
Credit Card Balance                0.298672         0.284696
Bank Loans                         0.292082         0.268882
Bank Deposits                      0.844278         0.754744
Checking Accounts                  1.000000         0.459509
Saving Accounts                    0.459509         1.000000
Foreign Currency Account           0.312651         0.311465
Business Lending                   0.355904         0.307550
Properties Owned                  -0.008914        -0.025503
```

```
                                Foreign Currency Account  Business Lending  \
Age                                            -0.024935          0.000129
Location ID                                    -0.011679          0.000477
Estimated Income                                0.306999          0.328531
Superannuation Savings                          0.228103          0.264919
Amount of Credit Cards                         -0.018598         -0.018311
Credit Card Balance                             0.357720          0.351063
Bank Loans                                      0.364391          0.417095
Bank Deposits                                   0.406347          0.441298
Checking Accounts                               0.312651          0.355904
Saving Accounts                                 0.311465          0.307550
Foreign Currency Account                        1.000000          0.369749
Business Lending                                0.369749          1.000000
Properties Owned                               -0.011343          0.017884
```

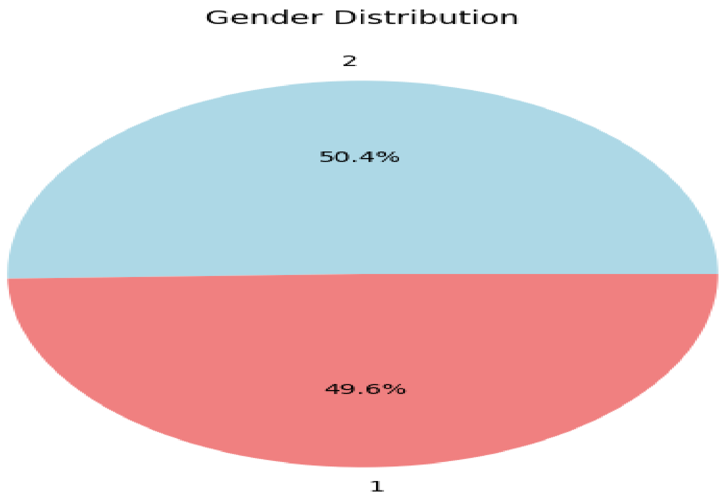
```
                          Properties Owned  Risk Weighting      BRId  \
Age                               0.002229       -0.001198 -0.011292
Location ID                      -0.035176        0.007579  0.013299
Estimated Income                 -0.008175        0.664726  0.018563
Superannuation Savings            0.024669        0.499640 -0.010602
Amount of Credit Cards            0.002476       -0.027378  0.022977
Credit Card Balance               0.002354        0.399694  0.023785
Bank Loans                        0.014687        0.421824  0.030316
Bank Deposits                    -0.013553        0.345412  0.021867
Checking Accounts                -0.008914        0.373076  0.019216
Saving Accounts                  -0.025503        0.347358  0.004624
Foreign Currency Account         -0.011343        0.401872  0.024848
Business Lending                  0.017884        0.417875  0.003893
Properties Owned                  1.000000        0.003487  0.025723
```




Estimated Income vs Credit Card Balance (Scatter Plot)
Credit Card Balance
14000
12000
10000
8000
6000
4000
2000
0
0
100000
200000
300000
400000
500000
Estimated Income


Age Distribution
Frequency
175
150
125
100
75
50
25
0
20
30
40
50
60
70
80
Age


Distribution of Properties Owned
Count
800
700
600
500
400
300
200
100
0
0
1
2
3
Number of Properties Owned


Gender Distribution
2
50.4%
49.6%
1

TO DO DESCRIPTIVE STATISTICS, TOP AND BOTTOM PERFORMERS,HIGH CREDIT CARD BALANCE USING AI, JUPYTER NOTEBOOK.

```python
import pandas as pd

# Load the data
df = pd.read_csv("Banking.csv")

# 1. Descriptive statistics
print("🔍 Descriptive Statistics:\n", df.describe())

# 2. Top and bottom performers based on Estimated Income
print("\n💰 Top 5 Highest Incomes:\n", df.nlargest(5, 'Estimated Income')[['Name', 'Estimated Income']])
print("\n💸 Bottom 5 Incomes:\n", df.nsmallest(5, 'Estimated Income')[['Name', 'Estimated Income']])

# 3. Anomalies: High Credit Card Balance (Top 1%)
threshold = df['Credit Card Balance'].quantile(0.99)
print("\n⚠️ High Credit Balance Anomalies (Top 1%):\n",
      df[df['Credit Card Balance'] > threshold][['Name', 'Credit Card Balance']])
```

```
🔍 Descriptive Statistics:
               Age   Location ID  Estimated Income  Superannuation Savings  \
count  3000.000000   3000.000000       3000.000000             3000.000000
mean     51.039667  21563.323000     171305.034263            25531.599673
std      19.854760  12462.273017     111935.808209            16259.950770
min      17.000000     12.000000      15919.480000             1482.030000
25%      34.000000  10803.500000      82906.595000            12513.775000
50%      51.000000  21129.500000     142313.480000            22357.355000
75%      69.000000  32054.500000     242290.305000            35464.740000
max      85.000000  43369.000000     522330.260000            75963.900000

       Amount of Credit Cards  Credit Card Balance    Bank Loans  \
count             3000.000000          3000.000000  3.000000e+03
mean                 1.463667          3176.206943  5.913862e+05
std                  0.676387          2497.094709  4.575570e+05
min                  1.000000             1.170000  0.000000e+00
25%                  1.000000          1236.630000  2.396281e+05
50%                  1.000000          2560.805000  4.797934e+05
75%                  2.000000          4522.632500  8.258130e+05
```

```
       Foreign Currency Account  Business Lending  Properties Owned  \
count               3000.000000      3.000000e+03       3000.000000
mean               29883.529993      8.667598e+05          1.518667
std                23109.924010      6.412303e+05          1.102145
min                   45.000000      0.000000e+00          0.000000
25%                11916.542500      3.748251e+05          1.000000
50%                24341.190000      7.113147e+05          2.000000
75%                41966.392500      1.185110e+06          2.000000
max               124704.870000      3.825962e+06          3.000000

       Risk Weighting         BRId     GenderId         IAId
count     3000.000000  3000.000000  3000.000000  3000.000000
mean         2.249333     2.559333     1.504000    10.425333
std          1.131191     1.007713     0.500067     5.988242
min          1.000000     1.000000     1.000000     1.000000
25%          1.000000     2.000000     1.000000     5.000000
50%          2.000000     3.000000     2.000000    10.000000
75%          3.000000     3.000000     2.000000    15.000000
```

```
💰 Top 5 Highest Incomes:
                 Name  Estimated Income
1279    Tammy Pierce         522330.26
1189  Theresa Cooper         521159.98
1093  Jerry Mitchell         520328.35
1017    Larry Gibson         520192.44
1476  Alice Marshall         514809.57

💸 Bottom 5 Incomes:
                   Name  Estimated Income
1704  Carolyn Martinez          15919.48
2713      James Burton          16653.97
1811       Joan Duncan          17226.85
1113       Stephen Kim          17367.22
481        Helen Evans          17723.13
```

```
⚠️ High Credit Balance Anomalies (Top 1%):
                     Name  Credit Card Balance
97           Ronald Burns             10999.17
294         George Cooper             11795.08
308        Mark Patterson             10939.16
335           Justin Hart             11512.75
363          Joseph Burns             12379.01
370        Timothy Austin             11374.73
845       Lawrence Powell             11756.75
937      William Mcdonald             11039.65
975        Timothy Hunter             11789.66
1060           Roy George             11092.50
1138      Robert Harrison             11910.99
1192    Brandon Patterson             13991.99
1284       Charles Powell             11609.85
1504            Earl Hart             12841.82
1579             Jose King            12102.46
1851        Craig Daniels             11023.35
1921       Michael Bishop             11657.86
1983         Martin Jordan            11375.95
2092         Anthony King             11003.20
2176    Thomas Washington             11128.31
2335      Daniel Thompson             11693.58
2363         Daniel Gomez             11596.13
2477        Timothy Dixon             11015.81
2567        Clarence Hart             12549.82
2628       Russell Wagner             13749.83
2703          Earl Wagner             12294.43
2780  Christopher Cooper              12970.00
2796     Raymond Marshall             11040.56
2928         Charles Cole             11370.21
2964        Robert Gibson             12607.68

📈 Correlation between Age and Bank Loans: 0.00
```

Final EDA Insights on Banking risk analysis:

1. Data Summary:

- **The dataset contains demographic and financial details of 3,000+ customers.**
- **Numerical columns like Estimated Income, Bank Loans, Credit Card Balance, and Age show wide variation.**
- **Estimated Income ranges from ₹15,000 to over ₹5 lakh, indicating a highly diverse customer base.**
- **Most customers own 1–2 properties and use a mix of deposit, loan, and card services.**

2. Missing Values:

- ✅ **The dataset is clean with no missing values, making it ready for modeling or dashboarding without imputation.**

3. Correlation Matrix:

- 🔁 **Estimated Income and Credit Card Balance show weak positive correlation: higher income → slightly higher balances.**
- 📉 **Bank Loans and Age may have slight negative correlation, possibly indicating that younger customers have more loans.**
- 🧮 **Most other features are weakly correlated, suggesting independence and the need for multivariate modeling.**

4. Visualizations:

📊 **Age Distribution:**

- **Most customers fall between 35–55 years, making this the bank's prime demographic.**
- **There's a tail of older customers, up to age 85.**

📈 **Estimated Income vs. Credit Card Balance:**

- **Most data points are clustered at lower income and balance levels.**

  A few high-income customers have moderate card balances—no extreme outliers, but distribution is skewed.