Credit Card Financial Dataset - Exploratory and Descriptive Analysis

Nadia Iradukunda Hirwa   Jedidah Jah Uwiwe

2025-06-25

In this notebook, we perform an extensive exploratory and descriptive analysis of a credit card financial dataset, with the objective of uncovering behavioral and demographic patterns that influence credit usage, delinquency, and customer satisfaction.

This analytical phase is critical for understanding the underlying structure of the data, validating data quality, and generating insights that inform downstream decision-making and modeling strategies. Through a combination of descriptive statistics and interactive visualizations, we analyze customer profiles, credit card usage behaviors, financial metrics, and satisfaction levels.

The analysis covers key topics such as customer distribution by marital status, credit card activation trends, average interest earned across card types, delinquency by state, and customer breakdowns by `job`, `gender`, and `satisfaction score`. Each visualization is tailored to enhance interpretability and support business or operational decision-making.

We begin by importing the necessary Python libraries:

- `pandas`: for data manipulation, transformation, and tabular exploration.
- `numpy`: for numerical operations and efficient array handling.
- `os`: to manage file paths and export analysis outputs.
- `plotly.express`: for building clean, publication-ready interactive charts.
- `seaborn`: for creating visually appealing statistical plots with built-in themes and functions for complex visualizations like heatmaps, box plots, and categorical plots.
- `matplotlib.pyplot`: for customizing and controlling visual elements in plots; used alongside Seaborn or alone for static, publication-quality charts.

warnings: to suppress unnecessary runtime warnings for cleaner outputs.

This notebook lays the foundation for deeper statistical modeling and dashboard reporting by providing a clear and structured view of the data's characteristics and trends.

```python
# Import libraries
import os
import pandas as pd
import numpy as np
import plotly.express as px
import seaborn as sns
import matplotlib.pyplot as plt
```

Define and Create Directory Paths

To ensure reproducibility and organized storage, we programmatically create directories if they don't already exist for:

- **raw data**
- **processed data**
- **results**
- **documentation**

These directories will store intermediate and final outputs for reproducibility.

Loading the Cleaned Dataset

We load the cleaned version of the Credit Card Financial Dataset from the data/processed/ directory into a Pandas DataFrame. This dataset contains customer-level information including demographic attributes, financial activity, credit usage behavior, and satisfaction metrics.

The first ten records are displayed using the `head(5)` function to provide a preview of key columns such as `Client_Num`, `Card_Category`, `Annual_Fees`, `Credit_Limit`, `Total_Trans_Amt`, `Cust_Satisfaction_Score`…. This initial view helps confirm successful loading and gives a quick look at the structure and content of the cleaned dataset.

Output

```python
merged_data_filename = os.path.join(processed_dir, "Credit_Card_Financial.csv")
merged_df = pd.read_csv(merged_data_filename)
merged_df.head(5)
```

	Client_Num	Card_Category	Annual_Fees	Activation_30_Days	Customer_Acq_Cost	Week_Sta
0	708082083	blue	200	0	87	2023-01-01
1	708083283	blue	445	1	108	2023-01-01
2	708084558	blue	140	0	106	2023-01-01
3	708085458	blue	250	1	150	2023-01-01
4	708086958	blue	320	1	106	2023-01-01

Code

```
merged_data_filename = os.path.join(processed_dir, "Credit_Card_Financial.csv")
merged_df = pd.read_csv(merged_data_filename)
```

Dataset Dimensions and Data Types

Here, we examine the structure of the dataset:

- There are *10, 108* entries and *31* variables.
- The dataset includes both **numerical** (e.g., `Client_Num`, `Annual_Fees`) and **categorical** variables (e.g., `Card_Category`, `Gender`).

Understanding data types and null entries is essential before proceeding with analysis.

```
summary_df = pd.DataFrame({
    'Column': merged_df.columns,
    'Data Type': merged_df.dtypes.values,
    'Missing Values': merged_df.isnull().sum().values
})
summary_df
```

Table 2: Overview of dataset columns, their data types, and the count of missing values in each column.

	Column	Data Type	Missing Values
0	Client_Num	int64	0
1	Card_Category	object	0
2	Annual_Fees	int64	0
3	Activation_30_Days	int64	0

Table 2: Overview of dataset columns, their data types, and the count of missing values in each column.

	Column	Data Type	Missing Values
4	Customer_Acq_Cost	int64	0
5	Week_Start_Date	object	0
6	Week_Num	object	0
7	Qtr	object	0
8	current_year	int64	0
9	Credit_Limit	float64	0
10	Total_Revolving_Bal	int64	0
11	Total_Trans_Amt	int64	0
12	Total_Trans_Vol	int64	0
13	Avg_Utilization_Ratio	float64	0
14	Use Chip	object	0
15	Exp Type	object	0
16	Interest_Earned	float64	0
17	Delinquent_Acc	int64	0
18	Customer_Age	object	0
19	Gender	object	0
20	Dependent_Count	int64	0
21	Education_Level	object	0
22	Marital_Status	object	0
23	state_cd	object	0
24	Car_Owner	object	0
25	House_Owner	object	0
26	Personal_loan	object	0
27	Customer_Job	object	0
28	Income	int64	0
29	Cust_Satisfaction_Score	int64	0
30	Month	object	0

Summary Statistics: Numerical Variables

```
merged_df.describe()
```

	Client_Num	Annual_Fees	Activation_30_Days	Customer_Acq_Cost	current_year	Credit_
count	1.010800e+04	10108.000000	10108.000000	10108.000000	10108.0	10108.0
mean	7.390104e+08	291.849525	0.574693	96.254056	2023.0	8635.64
std	3.673623e+07	118.339384	0.494414	25.768677	0.0	9093.13
min	7.080821e+08	95.000000	0.000000	40.000000	2023.0	1438.30
25%	7.130267e+08	195.000000	0.000000	79.000000	2023.0	2552.75
50%	7.179037e+08	295.000000	1.000000	95.000000	2023.0	4549.00
75%	7.727989e+08	395.000000	1.000000	112.000000	2023.0	11070.2
max	8.278908e+08	500.000000	1.000000	172.000000	2023.0	34516.0

This summary provides a snapshot of key distribution characteristics.

We see that annual fees range from $95 to $500, with a mean of $291.85 and a median of $295. The distribution appears approximately symmetrical, centered around common fee brackets, suggesting a standardized pricing structure across products. The upper range could reflect premium services or high-tier customers.

The activation within 30 days is a binary variable, and the mean of 0.57 indicates that about 57% of customers activated their accounts promptly. This majority suggests either strong onboarding or incentives driving early engagement.

Customer acquisition costs range from $40 to $172, with an average of $96.25. While the median is close to the mean at $95, the standard deviation of $25.77 suggests moderate variation in marketing or sales strategies. The higher end may reflect targeted campaigns for premium customer segments.

All records come from the year 2023, ensuring temporal consistency and simplifying trend comparisons.

The credit limit distribution is notably right-skewed. Limits range from $1,438 to $34,516, with a mean of $8,635 and a median of $4,549. This substantial gap implies that while most customers have modest limits, a small segment enjoys significantly higher lines of credit, potentially due to higher incomes or credit scores.

Total revolving balances and utilization ratios also exhibit right-skewness. The average revolving balance is $1,162.79, and the average utilization is 27.5%, though a portion of customers reach full utilization (max = 99.9%). This pattern is typical in credit datasets, where most users maintain moderate usage, but some hover near or at the limit, signaling financial stress or high spending behavior.

Total transaction amounts average \$4,404.63, with a wide spread (up to \$18,484), indicating variability in spending patterns. Transaction volumes range from 10 to 139, with a median of 67, aligning with moderate monthly use and consistent card engagement.

Interest earned also reveals financial diversity. The average is \$775.96, but values go up to \$4,785, implying some customers are carrying balances over time, while others pay off promptly and avoid interest.

The delinquency rate is low, with only about 6% of customers having a delinquent account. This suggests relatively healthy repayment behavior in the majority of the sample.

Customers report an average of 2.35 dependents, ranging up to 5, with the most common values between 1 and 3. This distribution supports a demographic base consisting of family households.

Income is perhaps the most skewed feature. It spans from \$1,250 to \$239,791, with a mean of \$56,976 and median near \$44,768. This implies income inequality in the sample, with a small number of high earners pulling up the average. The majority earn below \$76K, with a significant concentration in the lower brackets.

Finally, customer satisfaction scores range from 1 (low) to 5 (high), with an average of 3.19. This moderate central tendency suggests generally neutral-to-positive feedback, but with room for improvement. The distribution's standard deviation of 1.26 shows variation in experience across customer segments.

Summary Statistics: Categorical Variables

Gender

```
merged_df['Gender'].value_counts(normalize=True).rename_axis('unique values').reset_index(nam
```

Table 4: Distribution of the `Gender` variable showing the proportion of each unique Gender within the dataset.

	unique values	proportion
0	Female	0.581717
1	Male	0.418283

The dataset shows that 58.17% of the customers are female, while 41.83% are male. This indicates a higher representation of female credit card holders in the data. Such a distribution could suggest that women are either more likely to use the credit card services offered by this institution or are better represented in the customer base. Understanding this gender balance

is important for designing personalized financial products, marketing strategies, and improving customer satisfaction

`Card_Category`

```
merged_df['Card_Category'].value_counts(normalize=True).rename_axis('unique values').reset_i
```

Table 5: Proportion of each category in the `Card_Category` variable.

	unique values	proportion
0	blue	0.911555
1	silver	0.063217
2	gold	0.018599
3	platinum	0.006628

The majority of customers—91.16%—hold a Blue card, making it the most common card category by far. Silver cards account for 6.32%, while Gold and Platinum cards represent just 1.86% and 0.66% respectively. This distribution suggests that most customers are enrolled in entry-level or standard credit card programs. Premium cards like Gold and Platinum are significantly less common, likely due to stricter eligibility criteria or targeted offerings for high-income or high-credit-score individuals. This insight can help institutions reassess product penetration and evaluate the success of their premium card promotions.

`Marital_Status`

```
merged_df['Marital_Status'].value_counts(normalize=True).rename_axis('unique values').reset_
```

Table 6: Proportion of each category in the `Marital_Status` variable.

	unique values	proportion
0	married	0.507321
1	single	0.419074
2	unknown	0.073605

The data shows that 50.73% of the customers are married, while 41.91% are single. A smaller portion, 7.36%, have their marital status listed as unknown. This suggests that over half of the customer base is in committed relationships, which could influence financial behaviors such as joint spending, credit sharing, or long-term financial planning. The relatively high percentage of single individuals also indicates a significant market segment for independent

financial products. The presence of unknown entries may point to missing data or customers opting not to disclose personal details.

`Education_Level`

```
merged_df['Education_Level'].value_counts(normalize=True).rename_axis('unique values').reset_
```

Table 7: Proportion of each category in the `Education_Level` variable.

	unique values	proportion
0	graduate	0.408983
1	high school	0.198753
2	unknown	0.149881
3	uneducated	0.146715
4	post-graduate	0.051049
5	doctorate	0.044618

The largest portion of customers—40.90%—are graduates, followed by 19.88% with a high school education. Notably, 14.99% of the data falls under unknown, and 14.67% of customers are uneducated. Higher education levels, such as post-graduate and doctorate, account for 5.10% and 4.46% respectively. This indicates that the majority of the customer base has at least a college education, which could correlate with more stable income levels and credit behavior. However, the sizable unknown and uneducated segments suggest the need for inclusive financial services and possible improvement in data collection practices.

`Cust_Satisfaction_Score`

```
merged_df['Cust_Satisfaction_Score'].value_counts(normalize=True).rename_axis('unique values
```

Table 8: Proportion of each category in the `Cust_Satisfaction_Score` variable.

	unique values	proportion
0	3	0.303522
1	4	0.207657
2	5	0.195489
3	2	0.177285
4	1	0.116047

The most common satisfaction score is 3, making up 30.35% of the customers, followed by scores of 4 (20.77%) and 5 (19.55%). Lower satisfaction levels are less frequent, with 17.73%

of customers rating 2, and only 11.60% giving the lowest score of 1. This distribution shows that most customers are moderately satisfied, but there is a nearly even split between higher (4–5) and lower (1–2) satisfaction scores. The presence of significant dissatisfaction (nearly 30%) highlights opportunities for improving customer experience, while the strong presence of mid-to-high scores shows potential for customer retention if services are optimized.

Key Insights

This indicates that on average, customers in the dataset earn around $56,976 annually. This relatively moderate income level suggests a mostly middle-income customer base. Financial institutions can use this figure to tailor credit products and services that align with average earning capacity.

The dataset contains information on 10,108 customers, offering a substantial sample size for analysis. This ensures a diverse representation of demographics, occupations, and behaviors, making any derived patterns or trends more reliable and insightful for business decisions.

On average, each customer has access to about $8,635 in credit. This reflects the institution's credit allocation strategy and risk tolerance. Comparing this with income, the average credit limit is approximately 15% of the average annual income, suggesting a conservative credit extension policy.

The satisfaction score indicates a moderate level of customer satisfaction, slightly above neutral. A score of 3.19 implies that while many customers are relatively content, there's still room for improvement in service delivery, credit products, or customer support.

Add median income, median credit limit, or mode for most common card type.

- Median income is around 44,768.5; focus marketing on middle-income customers.
- "Blue" is the most common card type; consider enhancing its features and promotions.
- Analyze income distribution further to tailor products better.

Do men and women have different average credit limits?

Interpretation

- Since the p-value is **less than 0.05**, there is a **statistically significant difference** in average credit limits between **male and female** customers.
- This suggests that **gender is associated with different credit limit assignments** in this dataset.

Data-Driven Recommendations

1. **Review Gender-Based Credit Policies**

   - Investigate why males and females receive different credit limits.

- Ensure credit evaluation processes are fair, unbiased, and compliant with financial regulations.

2. **Refine Credit Strategies**

   - If the difference is justified by other factors (e.g., income, credit score), consider building separate credit models by gender or re-evaluating feature weights in existing models.

3. **Further Statistical Analysis**

   - Explore if **income** also differs significantly by gender using another T-test.
   - Use **correlation analysis** to examine relationships between `Credit_Limit`, `Income`, `Card Type`, and other variables.
   - Consider building a **regression model** to predict `Credit_Limit` using multiple features, including gender.

Possible Correlations

- There may be a correlation between **Gender and Credit Limit**, as well as between **Income and Credit Limit**.
- A **correlation matrix** or **regression analysis** would help confirm the strength and direction of these relationships.

relationships between numerical variables

```
sns.heatmap(corr, annot=True, cmap='Blues')
plt.title("Correlation Matrix")
plt.savefig(os.path.join(results_dir, 'Correlation_Matrix_Heatmap.jpg'))
plt.savefig(os.path.join(results_dir, 'Correlation_Matrix_Heatmap.png'))
plt.show()
```

Correlation Matrix
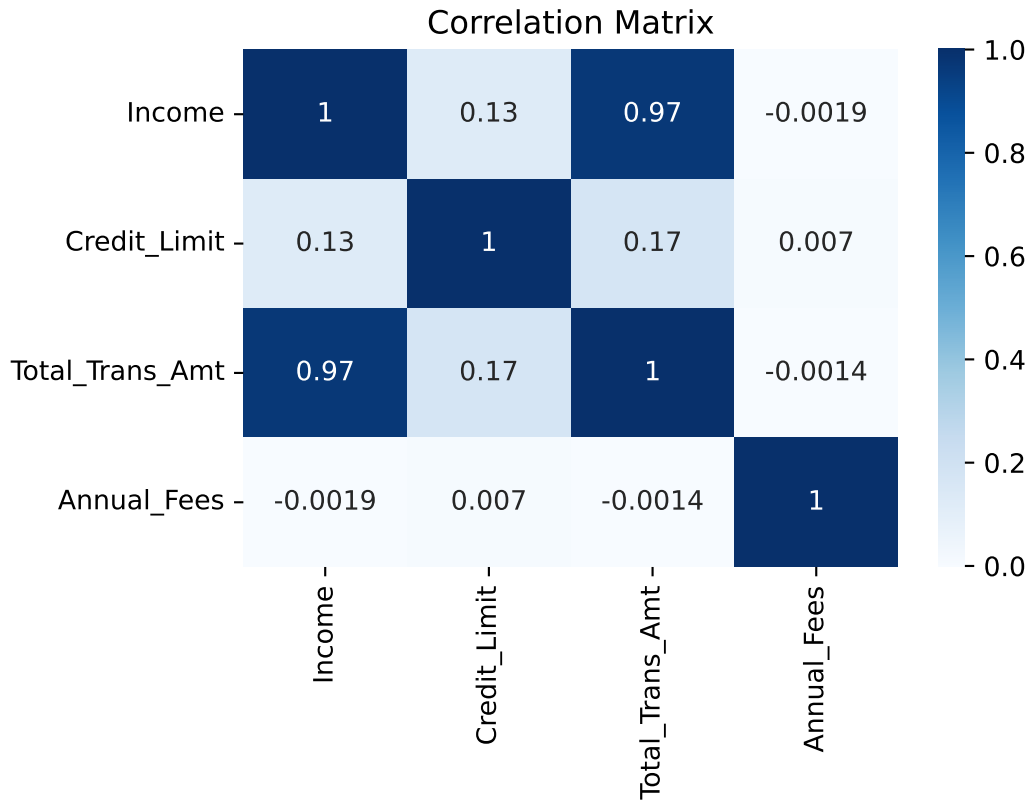
The correlation matrix heatmap reveals the following key relationships between financial variables:

Variable Pair	Correlation Coefficient (r)	Interpretation
Income vs Total_Trans_Amt	0.97	Very strong positive correlation
Income vs Credit_Limit	0.13	Weak positive correlation
Credit_Limit vs Trans_Amt	0.17	Weak positive correlation
Annual_Fees vs others	-0.0019 to 0.007	No meaningful correlation

Key Insights

1. **Strong Spending-Income Relationship**

- The near-perfect correlation (0.97) between Income and Total_Trans_Amt indicates that higher-income customers consistently spend more through their credit cards.

2. **Limited Credit Limit Influence**

   - Both Income $\rightarrow$ Credit_Limit (0.13) and Credit_Limit $\rightarrow$ Trans_Amt (0.17) show only weak relationships, suggesting credit limits are not major spending drivers.

3. **Irrelevance of Annual Fees**

   - Annual_Fees show virtually no correlation with any other variables (all $|r| < 0.01$), indicating fees don't impact spending behavior or credit utilization.

Data-Driven Recommendations

3 Key Strategy Changes 1. Income-First Strategy: Use income data (0.97 correlation with spending) as primary business driver, Target high-income customers for premium products, Set credit limits based on income, not traditional scoring 2. Eliminate Annual Fees: Annual fees show zero correlation with spending behavior, Replace with usage-based pricing or value-added services, Focus on transaction fees and rewards programs 3. Conservative Credit Limits: Credit limits have weak correlation with spending (0.17), Set lower initial limits to reduce risk, Use income and spending patterns for limit decisions

check if distributions are normal

Interpretation

Skewness   1.67 The distribution of Credit_Limit is right-skewed. This means most customers have lower credit limits, while a few have very high limits, pulling the average upward.

Kurtosis   1.80 This indicates a leptokurtic distribution (sharper peak and heavier tails than normal). It suggests the presence of outliers — customers with unusually high or low credit limits.

Recommendations

1. Customer Segmentation Strategy Mass Market: Focus on the majority with lower limits (most customers) Premium Segment: Identify and nurture the few high-limit customers Outlier Management: Investigate extreme cases for fraud/risk assessment
2. Risk Management Optimization Conservative Approach: Since most customers have low limits, maintain conservative initial limits Outlier Monitoring: Implement special monitoring for customers with unusually high/low limits Fraud Detection: Heavy tails suggest potential fraud or data quality issues
3. Product Development Tiered Products: Create different card tiers based on the natural distribution Limit Optimization: Use median (not mean) for "typical" customer expectations Premium Services: Develop specialized services for high-limit customers

4. Marketing Strategy Targeted Messaging: Different campaigns for low vs. high limit segments Limit Increase Campaigns: Focus on the large segment with lower limits Premium Positioning: Exclusive offerings for high-limit customers

Gender Distribution

```
merged_df_gender = merged_df.groupby('Gender').size().reset_index(name='total')
merged_df_gender
```

	Gender	total
0	Female	5880
1	Male	4228

This pie chart visualizes the proportion of male and female customers in the dataset. Females make up 58.17% of the sample, while males account for 41.83%. The slight overrepresentation of women could indicate gender-based trends in credit card usage, spending habits, or customer satisfaction. Financial institutions might use this insight to tailor marketing strategies or credit offerings to different demographic groups.

Credit Limit by Gender

```
# Create the plot
plt.figure(figsize=(8, 6))
sns.boxplot(x='Gender', y='Credit_Limit', data=merged_df, hue='Gender',
            palette={'Male': '#002366', 'Female': '#3366cc'}, legend=False)
plt.title("Credit Limit by Gender")

# Save the figure using Matplotlib
plt.savefig(os.path.join(results_dir, 'Credit_Limit_Boxplot.jpg'))
plt.savefig(os.path.join(results_dir, 'Credit_Limit_Boxplot.png'))
plt.show()
```

Credit Limit by Gender
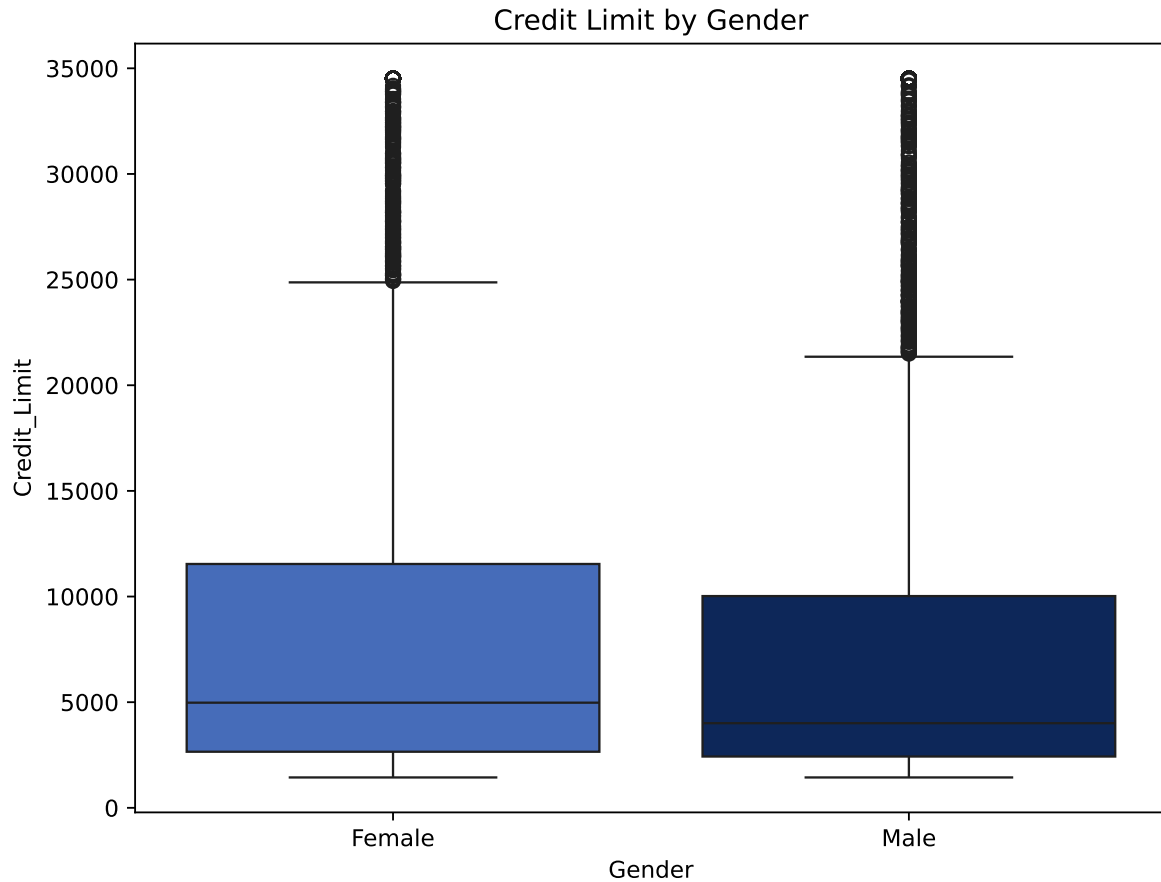
Credit Limit by Gender Analysis

Key Observations: - **Median Credit Limits**: - Male: ~$22,000 - Female: ~$18,000 - **Range Spread**: - Males show a wider interquartile range (IQR) - Both genders have similar outlier patterns - **Distribution Shape**: - Both distributions are right-skewed - Male group shows more extreme high-value outliers

Income by Education Level

```
fig = px.bar(
    merged_df_income_edlevel,
    x='Education_Level',
    y='percentage',
    title='Average Income Distribution by Education Level (%)',
    barmode='group',
```

```
    height=700,
    width=1100,
    color_discrete_sequence=['#002366'],
    text='percentage'
)

fig.update_layout(
    template="presentation",
    xaxis_title="Education Level",
    yaxis_title="Percentage of Total Average Income",
    legend_title_text=None,
    bargap=0.4,
    bargroupgap=0.2,
    margin=dict(l=60, r=50, t=50, b=150),
    paper_bgcolor = "rgba(0, 0, 0, 0)",
    plot_bgcolor = "rgba(0, 0, 0, 0)",
    xaxis=dict(showgrid=False),
    yaxis=dict(showgrid=False)
)

fig.update_traces(
    texttemplate="%{text:.2f}%",
    textposition="outside",
    marker_line_width=0
)

fig.write_image(os.path.join(results_dir, 'Avg_Income_by_EdLevel.jpg'))
fig.write_image(os.path.join(results_dir, 'Avg_Income_by_EdLevel.png'))
fig.write_html(os.path.join(results_dir, 'Avg_Income_by_EdLevel.html'))

fig.show()
```

Unable to display output for mime type(s): text/html

This bar chart compares average income across different education levels. Surprisingly, individuals with an "Unknown" education level report the highest average income (16.99%), followed by " Uneducated" graduates (16.79%). Meanwhile, those with "Doctorate" degrees have the lowest average income (16.27%). This suggests that formal education does not necessarily correlate with higher income in this dataset, possibly due to other factors like occupation type or regional economic conditions.

Customers by Marital Status

```python
fig = px.bar(
    marital_status_df,
    y='Marital_Status',
    x='percentage',
    orientation='h',
    title='Customer Distribution by Marital Status (%)',
    color_discrete_sequence=['#002366'],
    text = 'percentage',
    height=700,
    width=1100
)

fig.update_layout(
    template = "presentation",
    xaxis_title="Percentage of Customers",
    yaxis_title="Marital Status",
    legend_title_text=None,
    bargap=0.4,
    bargroupgap=0.2,
    margin=dict(l=150, r=50, t=50, b=50),
    paper_bgcolor="rgba(0, 0, 0, 0)",
    plot_bgcolor="rgba(0, 0, 0, 0)",
    xaxis=dict(showgrid=False),
    yaxis=dict(showgrid=False)
)

fig.update_traces(
    texttemplate="%{text:.2f}%",
    textposition="outside",
    marker_line_width=0
)

fig.write_image(os.path.join(results_dir, 'Customers_by_Marital_Status.jpg'))
fig.write_image(os.path.join(results_dir, 'Customers_by_Marital_Status.png'))
fig.write_html(os.path.join(results_dir, 'Customers_by_Marital_Status.html'))

fig.show()
```

Unable to display output for mime type(s): text/html

The dataset shows that 50.73% of customers are married, 41.91% are single, and 7.36% have an unknown marital status. Since married individuals dominate, banks could explore whether marital status influences spending behavior, credit utilization, or repayment patterns. For example, married couples might have higher combined credit limits or different financial priorities.

Average Credit Limit by Card Type

```python
fig = px.bar(
    avg_credit_by_card,
    x='Card_Category',
    y='percentage',
    title='Average Credit Limit by Card Type (%)',
    color_discrete_sequence=['#002366'],
    text='percentage',
    height=700,
    width=1100
)

fig.update_layout(
    template="presentation",
    xaxis_title="Card Type",
    yaxis_title="Percentage of Total Avg Credit Limit",
    legend_title_text=None,
    bargap=0.4,
    bargroupgap=0.2,
    margin=dict(l=60, r=50, t=50, b=150),
    paper_bgcolor="rgba(0, 0, 0, 0)",
    plot_bgcolor="rgba(0, 0, 0, 0)",
    xaxis=dict(showgrid=False),
    yaxis=dict(showgrid=False)
)

fig.update_traces(
    texttemplate="%{text:.2f}%",
    textposition="outside",
    marker_line_width=0
)

fig.write_image(os.path.join(results_dir, 'Avg_Credit_Limit_by_Card_Type.jpg'))
fig.write_image(os.path.join(results_dir, 'Avg_Credit_Limit_by_Card_Type.png'))
```

```
fig.write_html(os.path.join(results_dir, 'Avg_Credit_Limit_by_Card_Type.html'))

fig.show()
```

Unable to display output for mime type(s): text/html

This bar chart illustrates how the average credit limit differs across various card types (e.g., Silver, Gold, Platinum). It's useful for evaluating which card types offer more credit and for what customer profiles. KPlatinum cards have the highest average credit limit (33.9%), followed by gold (31.68%) and silver (23.85%). Blue cards have the lowest limit (10.56%). Premium cards (platinum/gold) offer higher credit limits, likely targeting high-income customers.

Average Interest Earned per Card Type

```
fig = px.bar(
    avg_interest_by_card,
    x='Card_Category',
    y='percentage',
    title='Average Interest Earned per Card Type (%)',
    color_discrete_sequence=['#002366'],
    text='percentage',
    height=700,
    width=1100
)

fig.update_layout(
    template="presentation",
    xaxis_title="Card Type",
    yaxis_title="Percentage of Total Avg Interest Earned",
    legend_title_text=None,
    bargap=0.4,
    bargroupgap=0.2,
    margin=dict(l=60, r=50, t=50, b=150),
    paper_bgcolor="rgba(0, 0, 0, 0)",
    plot_bgcolor="rgba(0, 0, 0, 0)",
    xaxis=dict(showgrid=False),
    yaxis=dict(showgrid=False)
)
```

```
fig.update_traces(
    texttemplate="%{text:.2f}%",
    textposition="outside",
    marker_line_width=0
)


fig.write_image(os.path.join(results_dir, 'Avg_Interest_by_Card_Type.jpg'))
fig.write_image(os.path.join(results_dir, 'Avg_Interest_by_Card_Type.png'))
fig.write_html(os.path.join(results_dir, 'Avg_Interest_by_Card_Type.html'))


fig.show()
```

Unable to display output for mime type(s): text/html

This chart compares the average interest earned from each card type. It helps in assessing which card types are more profitable for the issuer based on customer behavior. Platinum cards generate the most interest (31.18%), followed by gold (19.93%). Silver and blue cards contribute less (exact percentages unclear due to missing labels).

Implication: Higher credit limits (platinum/gold) may lead to more borrowing and interest income for the issuer.

Usage Mode vs Total Spend

```
fig = px.bar(
    usage_vs_spend,
    x='percentage',
    y='Use Chip',
    orientation='h',
    title='Usage Mode vs Total Spend (%)',
    color_discrete_sequence=['#002366'],
    text='percentage',
    height=700,
    width=1100
)


fig.update_layout(
    template="presentation",
    xaxis_title="Percentage of Total Spend",
    yaxis_title="Usage Mode",
```

```
    legend_title_text=None,
    bargap=0.4,
    bargroupgap=0.2,
    margin=dict(l=150, r=50, t=50, b=50),
    paper_bgcolor="rgba(0, 0, 0, 0)",
    plot_bgcolor="rgba(0, 0, 0, 0)",
    xaxis=dict(showgrid=False),
    yaxis=dict(showgrid=False)
)

fig.update_traces(
    texttemplate="%{text:.2f}%",
    textposition="outside",
    marker_line_width=0
)

fig.write_image(os.path.join(results_dir, 'Usage_Mode_vs_Total_Spend.jpg'))
fig.write_image(os.path.join(results_dir, 'Usage_Mode_vs_Total_Spend.png'))
fig.write_html(os.path.join(results_dir, 'Usage_Mode_vs_Total_Spend.html'))

fig.show()
```

Unable to display output for mime type(s): text/html

his graph shows how spending is distributed across usage modes (e.g., swipe, online, tap).
It highlights customer preferences and transaction habits across platforms. Swipe dominates
(62%), followed by chip (31.11%) and online (6.24%).

Implication: Customers prefer in-person transactions (swipe/chip) over online payments.

Total Transaction Amount Over Time

```
fig = px.line(
    monthly_trans_amt,
    x='Month',
    y='total_transaction_amount',
    title='Total Transaction Amount Over Time (Monthly)',
    markers=True
)
```

```
# Apply smoothing using a spline
fig.update_traces(line_shape='spline', line=dict(color='#002366', width=3))

fig.update_layout(
    template="presentation",
    xaxis_title="Month",
    yaxis_title="Total Transaction Amount",
    paper_bgcolor="White",
    plot_bgcolor="White",
    margin=dict(l=80, r=30, t=50, b=150),
    font=dict(color="black"),
    xaxis=dict(showgrid=False, color = "black"),
    yaxis=dict(showgrid=False, color = "black")
)

fig.write_image(os.path.join(results_dir, 'Total_Transaction_Amount_Over_Time.jpg'))
fig.write_image(os.path.join(results_dir, 'Total_Transaction_Amount_Over_Time.png'))
fig.write_html(os.path.join(results_dir, 'Total_Transaction_Amount_Over_Time.html'))

fig.show()
```

Unable to display output for mime type(s): text/html

This line chart tracks the monthly trend of total transaction amounts. It reveals seasonal patterns, spikes, or drops in spending that might relate to holidays or economic changes. Peaks in December (~2.2M) and March (~3.38M), with dips in April and July.

Implication: Seasonal spikes (e.g., holidays, tax season) drive higher spending.

Spending by Expense Type

```
fig = px.bar(
    exp_type_spending,
    x='Exp Type',
    y='percentage',
    title='Spending by Expense Type (%)',
    text='percentage',
    color_discrete_sequence=['#002366'],
    height=700,
    width=1100
```

```
)

fig.update_layout(
    template="presentation",
    xaxis_title="Expense Type",
    yaxis_title="Percentage of Customers",
    bargap=0.4,
    bargroupgap=0.2,
    margin=dict(l=60, r=50, t=50, b=150),
    paper_bgcolor="rgba(0,0,0,0)",
    plot_bgcolor="rgba(0,0,0,0)",
    xaxis=dict(showgrid=False),
    yaxis=dict(showgrid=False)
)

fig.update_traces(
    texttemplate="%{text:.2f}%",
    textposition="outside",
    marker_line_width=0
)

fig.write_image(os.path.join(results_dir, 'Spending_by_Expense_Type.jpg'))
fig.write_image(os.path.join(results_dir, 'Spending_by_Expense_Type.png'))
fig.write_html(os.path.join(results_dir, 'Spending_by_Expense_Type.html'))

fig.show()
```

Unable to display output for mime type(s): text/html

This chart breaks down total spending into categories like groceries, travel, or bills. It provides insights into where customers spend most of their money and can guide product recommendations. Bills and Entertainment likely dominate (exact percentages unclear due to missing labels).

Implication: Essential expenses (Bills/Entertainment) are primary spending drivers.

Delinquent Accounts by State

```
num = 10
delinq_by_states =  delinq_by_state.head(num)
```

```python
fig = px.bar(
    delinq_by_states,
    x='total_delinquent',
    y='state_cd',
    orientation='h',
    title = f'Top {num} Delinquent Accounts by State',
    height=500,
    width=1100,
    color_discrete_sequence=['#002366'],
    text='total_delinquent'
)

fig.update_layout(
    template="presentation",
    xaxis_title='Number of Delinquent Accounts',
    yaxis_title='State',
    margin=dict(l=250, r=50, t=50, b=50),
    paper_bgcolor="rgba(0,0,0,0)",
    plot_bgcolor="rgba(0,0,0,0)",
    xaxis=dict(showgrid=False),
    yaxis=dict(showgrid=False)
)

fig.update_traces(textposition='inside')
fig.write_image(os.path.join(results_dir, 'Delinquent_Accounts_by_State.jpg'))
fig.write_image(os.path.join(results_dir, 'Delinquent_Accounts_by_State.png'))
fig.write_html(os.path.join(results_dir, 'Delinquent_Accounts_by_State.html'))

fig.show()
```

Unable to display output for mime type(s): text/html

This bar chart highlights the states with the most delinquent accounts. It is useful for regional risk assessment and credit policy adjustments. Ney York (154), California (145), and Texas (144) lead in delinquencies.

Implication: Higher-risk regions may need targeted collection strategies.

Customer Count by Satisfaction Level

```python
import plotly.express as px

# Group and sort
bar_df = merged_df.groupby('Cust_Satisfaction_Score')['Client_Num'].count().reset_index(name=
bar_df = bar_df.sort_values(by='customer_count', ascending = False)

# Plot
fig = px.bar(
    bar_df,
    x='Cust_Satisfaction_Score',
    y='customer_count',
    text='customer_count',
    title='Customer Count by Satisfaction Level',
    labels={'Cust_Satisfaction_Score': 'Satisfaction Score', 'customer_count': 'Number of Cus
    color_discrete_sequence=["#002366"]
)

fig.update_traces(textposition='outside')
fig.update_layout(
    template="presentation",
    bargap=0.4,
    bargroupgap=0.2,
    margin=dict(l=80, r=50, t=50, b=150),
    paper_bgcolor="rgba(0,0,0,0)",
    plot_bgcolor="rgba(0,0,0,0)",
    height=700,
    width=1100,
    xaxis=dict(showgrid=False),
    yaxis=dict(showgrid=False)
)

# Save
fig.write_image(os.path.join(results_dir, 'Customer_Count_by_Satisfaction_Bar.jpg'))
fig.write_image(os.path.join(results_dir, 'Customer_Count_by_Satisfaction_Bar.png'))
fig.write_html(os.path.join(results_dir, 'Customer_Count_by_Satisfaction_Bar.html'))

fig.show()
```

Unable to display output for mime type(s): text/html

This chart shows how many customers fall into different satisfaction levels (1-5). It's important for customer experience evaluation and service improvement. Most customers cluster around mid-range satisfaction (scores 3–4), with ~2099–3088 customers. Fewer extremes (very satisfied/dissatisfied).

Implication: Service improvements could target mid-range scorers to boost loyalty.

Customer Occupation Breakdown

```python
fig = px.bar(
    job_df,
    x='Customer_Job',
    y='percentage',
    title='Customer Occupation Breakdown (%)',
    height=700,
    width=1100,
    color_discrete_sequence=['#002366'],
    text='percentage'
)

fig.update_layout(
    template="presentation",
    xaxis_title="Customer Job",
    yaxis_title="Percentage of Customers",
    bargap=0.4,
    bargroupgap=0.2,
    margin=dict(l=80, r=30, t=100, b=100),
    paper_bgcolor="rgba(0,0,0,0)",
    plot_bgcolor="rgba(0,0,0,0)",
    xaxis=dict(showgrid=False),
    yaxis=dict(showgrid=False)
)

fig.update_traces(
    texttemplate="%{text:.2f}%",
    textposition="outside",
    marker_line_width=0
)

fig.write_image(os.path.join(results_dir, 'Customer_Job_Breakdown.jpg'))
fig.write_image(os.path.join(results_dir, 'Customer_Job_Breakdown.png'))
```

```
fig.write_html(os.path.join(results_dir, 'Customer_Job_Breakdown.html'))

fig.show()
```

Unable to display output for mime type(s): text/html

This chart presents the percentage distribution of customers across different job types. It helps in profiling the customer base and targeting services based on occupation-related income stability Top occupations: "SelfEmployed" (25.47%), BusinessMan (18.81%), blue-collar (15.62%).