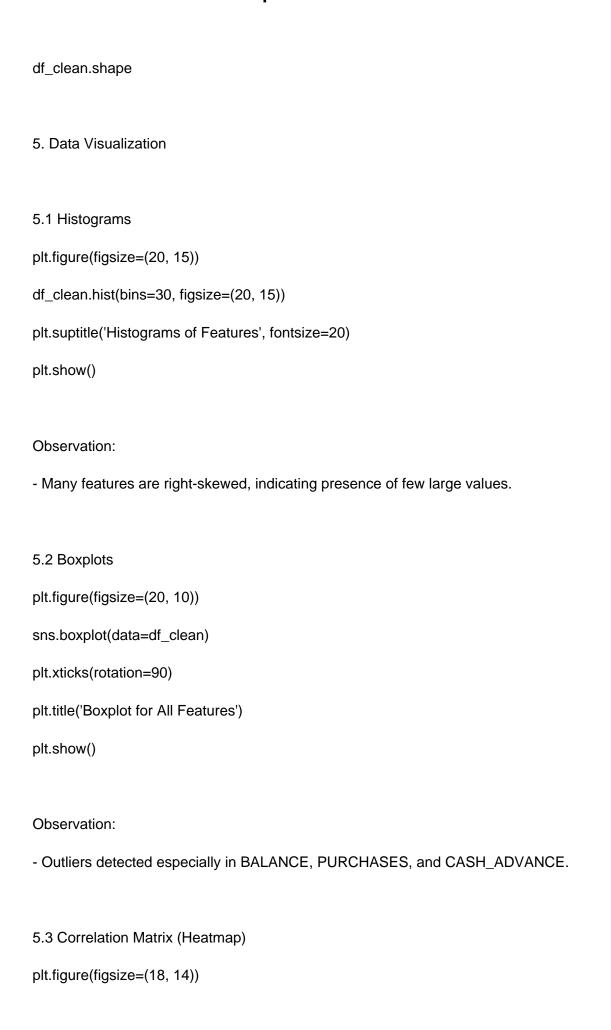
CREDIT CARD CUSTOMER DATASET - EXPLORATORY DATA ANALYSIS (EDA) 1. Import Libraries import pandas as pd import numpy as np import matplotlib.pyplot as plt import seaborn as sns sns.set(style='whitegrid') 2. Load Dataset df = pd.read_csv('CC GENERAL.csv') df.head() 3. Basic Exploration df.info() df.describe() df.isnull().sum() 4. Data Cleaning

df_clean = df.dropna()



```
corr_matrix = df_clean.corr()
sns.heatmap(corr_matrix, annot=False, cmap='coolwarm')
plt.title('Correlation Heatmap', fontsize=18)
plt.show()
```

Observation:

- PURCHASES and CREDIT_LIMIT are moderately correlated.
- BALANCE also shows correlation with CREDIT_LIMIT.

5.4 Pairplot

```
sample_columns = df_clean.columns[1:6]
sns.pairplot(df_clean[sample_columns])
plt.suptitle('Pairplot of Selected Features', y=1.02)
plt.show()
```

Observation:

- Some linear relationships between BALANCE and CREDIT_LIMIT.
- Wide spread in variables like PURCHASES_FREQUENCY.
- 6. Summary of Findings
- Dataset mostly contains continuous numerical variables.
- Significant right skew in several features.
- Presence of notable outliers.
- Some moderate positive correlations among financial features.

Recommendations:

- Outliers may need treatment before modeling.
- Standardization/Normalization recommended.
- Dimensionality reduction (PCA) could help.
- Further unsupervised learning (e.g., clustering) can be explored.