

**Finance Project Report: Predicting Personal Loan Acceptance**

1. Project Overview
2. Data Overview
3. Data Pre-processing

3.1 Data Cleaning

3.2 Data Transformation

3.3 Feature Engineering

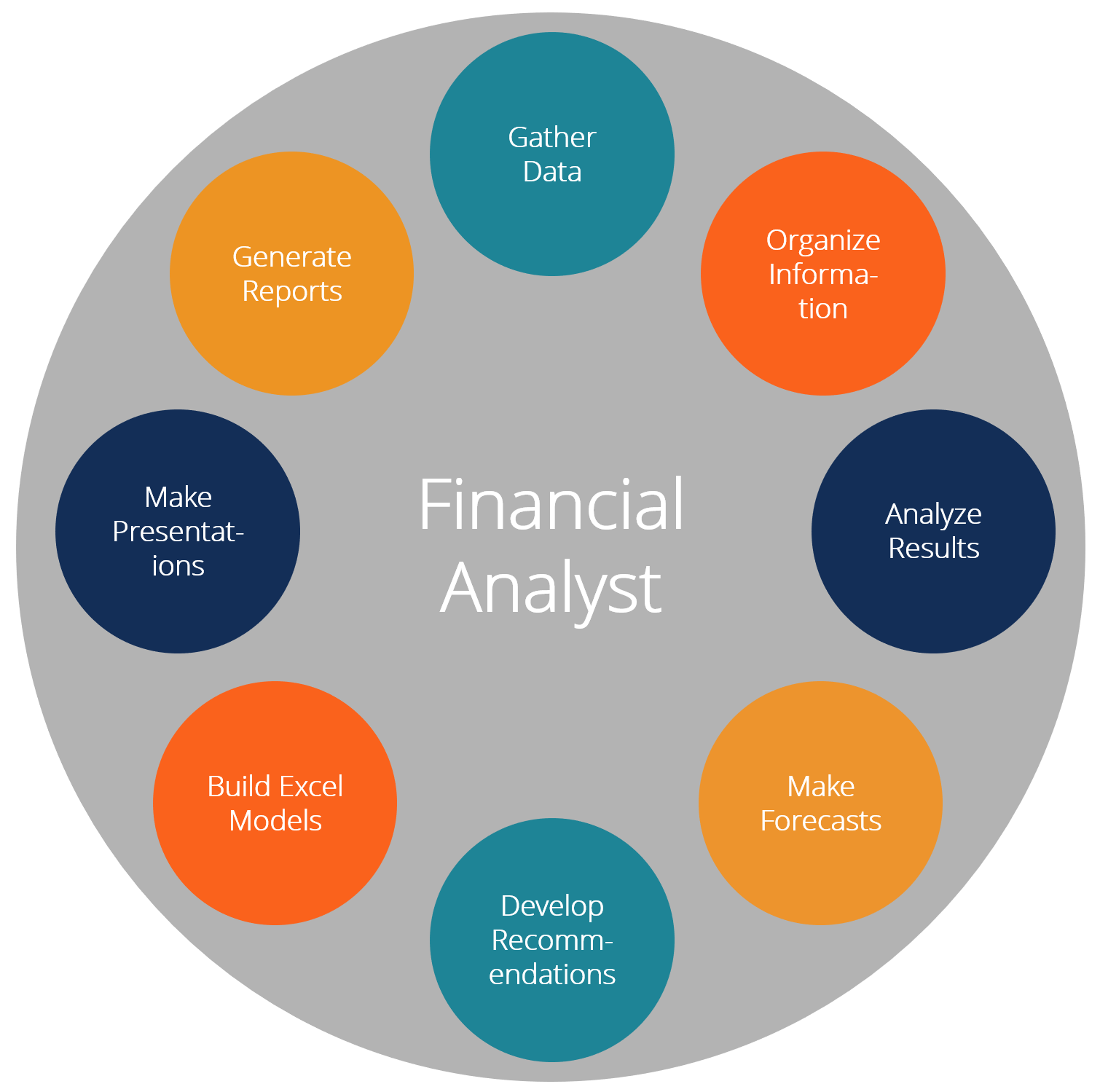
1. Correlation Analysis
2. Exploratory Data Analysis (EDA)

4.1 Distribution of Key Variables

4.2 Visual Insights

6) Feature Selection.

7) Data Flow

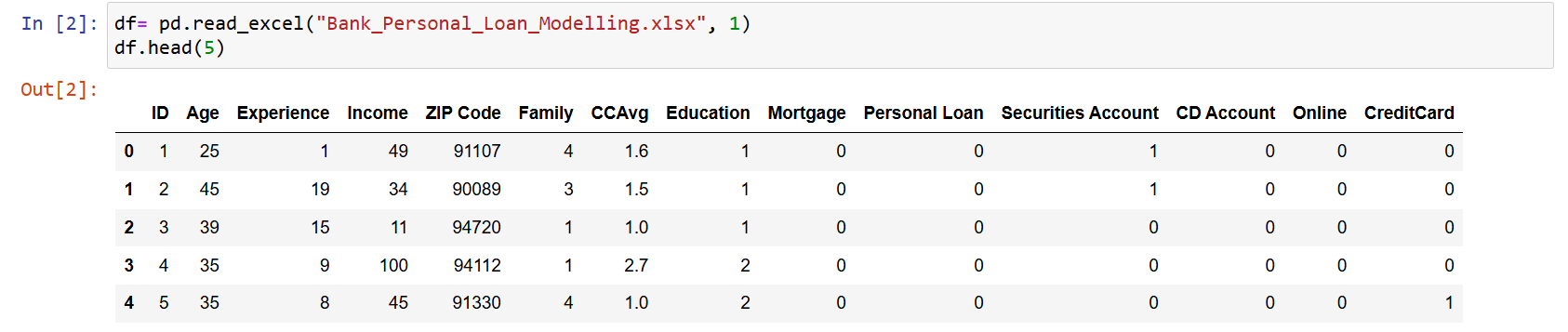


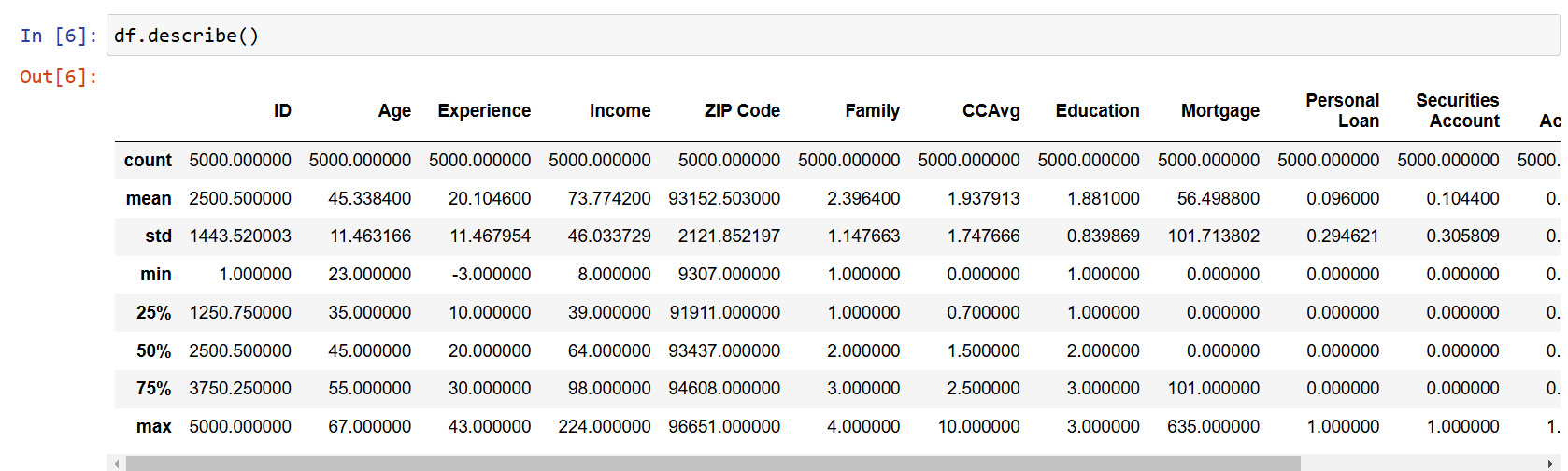
**Project Overview:** The objective of this project is to help Thera Bank increase the success rate of its personal loan campaigns by leveraging demographic, financial, and behavioral data of its customers. Specifically, the goal is to predict whether a customer will accept a personal loan offer based on their attributes.

**Data Overview:** The dataset, "Bank\_Personal\_Loan\_Modelling.xlsx," contains data for 5000 customers, with the following key variables:

* **Demographic Information**: Age, Experience, Income, Family, Education, etc.
* **Bank Relationship**: Mortgage, Credit Card, Securities Account, CD Account, etc.
* **Target Variable**: Personal Loan (whether the customer accepted the loan or not).

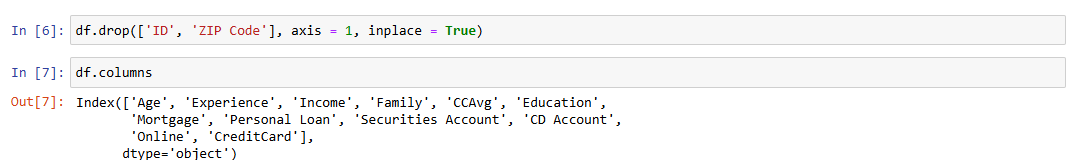
Only 9.6% (480) of customers accepted the personal loan offer in the past campaign, which highlights the challenge of predicting loan acceptance.

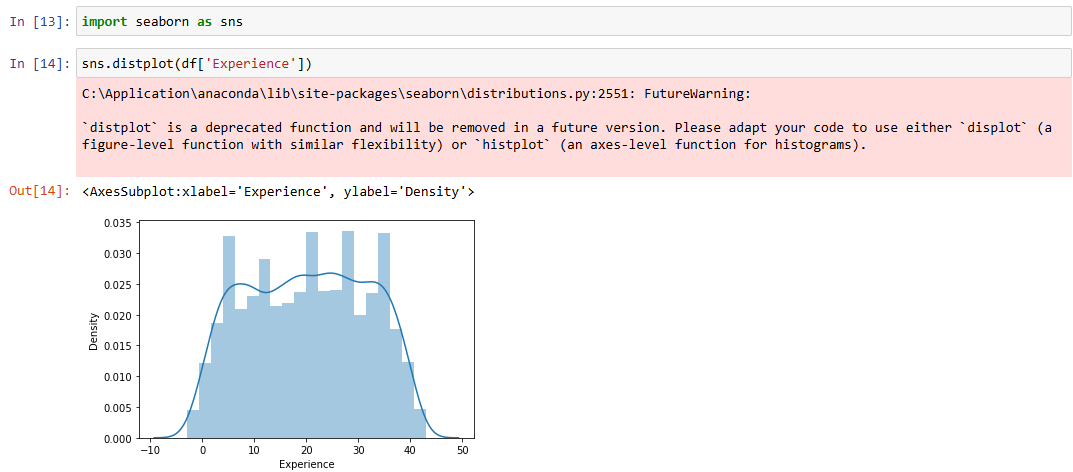


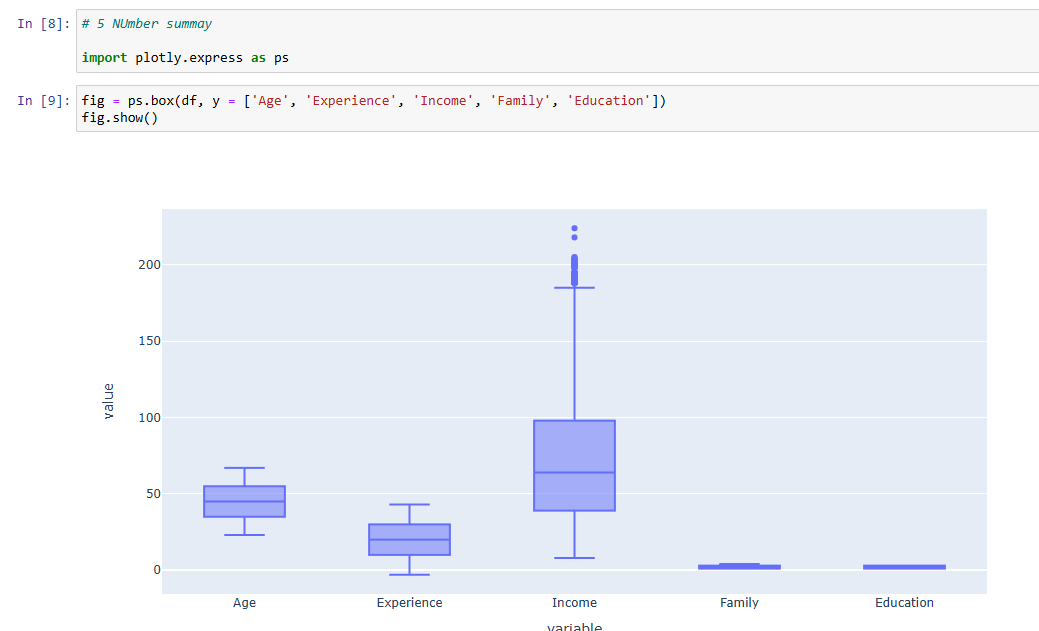


### Data Pre-processing:

1. **Data Cleaning:**
   * The columns 'ID' and 'ZIP Code' were removed as they were not relevant for predictive analysis.
   * Missing values: No missing values were found in the dataset.
   * Outliers: The 'Experience' column contained negative values, which were replaced with the mean of the column (20.10 years).

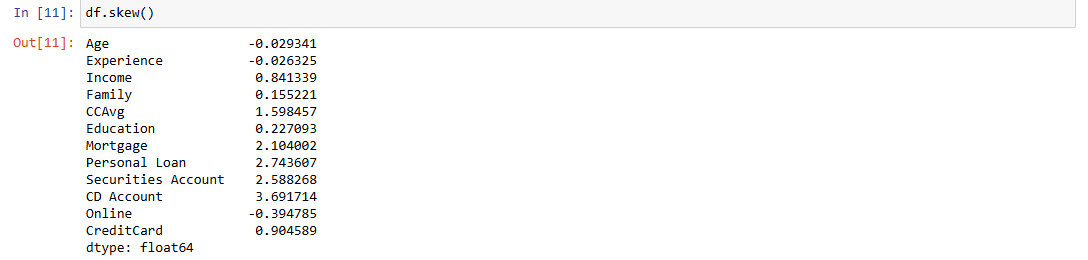


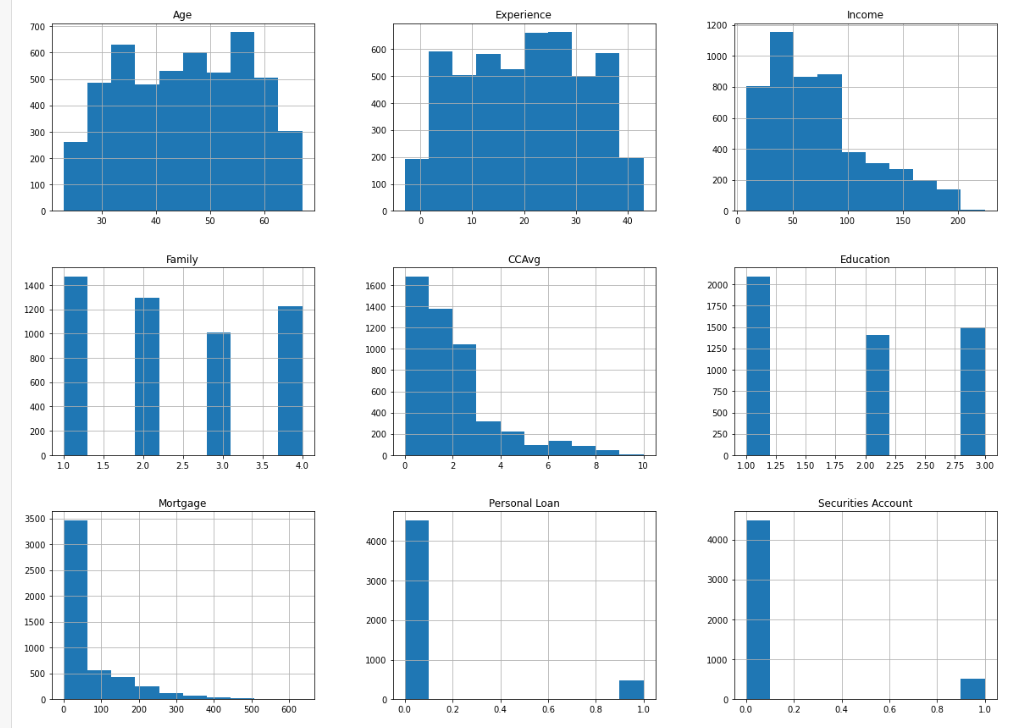




**Data Transformation:**

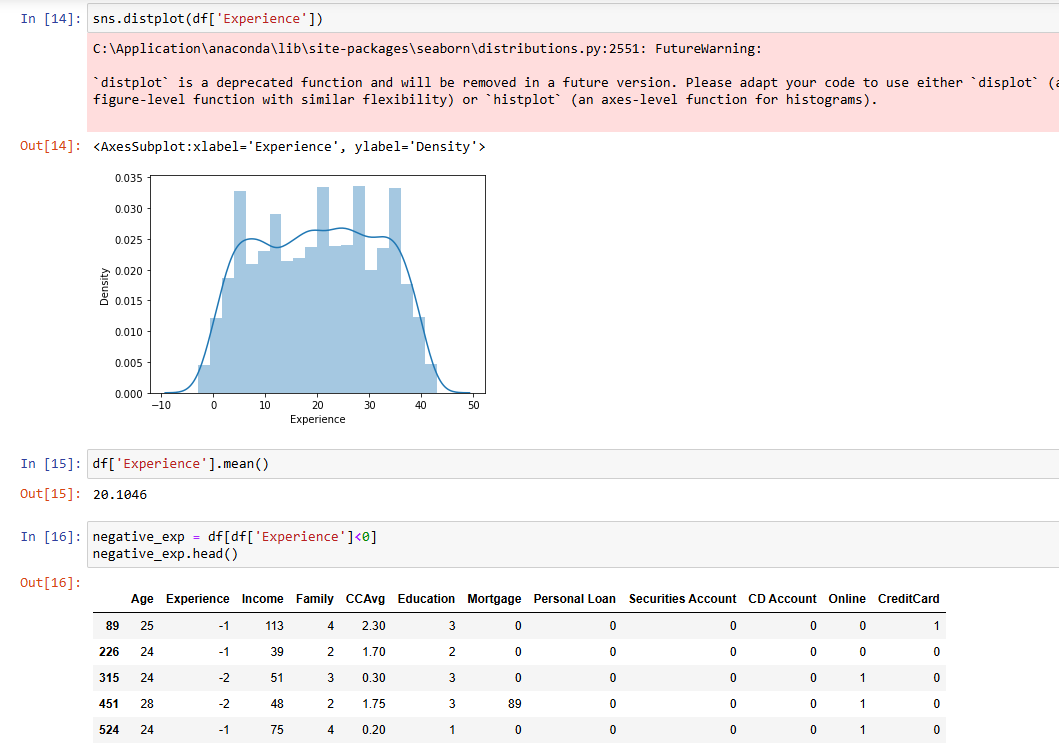
* **Log Transformation**: A log transformation was applied to 'Income' and 'CCAvg' (Credit Card Average) to reduce skewness and normalize the distribution.
* **Power Transformation**: Applied the Yeo-Johnson method to 'Income' for normalization.
* **Categorical Encoding**: The 'Education' column was converted into a categorical variable (Undergraduate, Graduate, Professional Person).





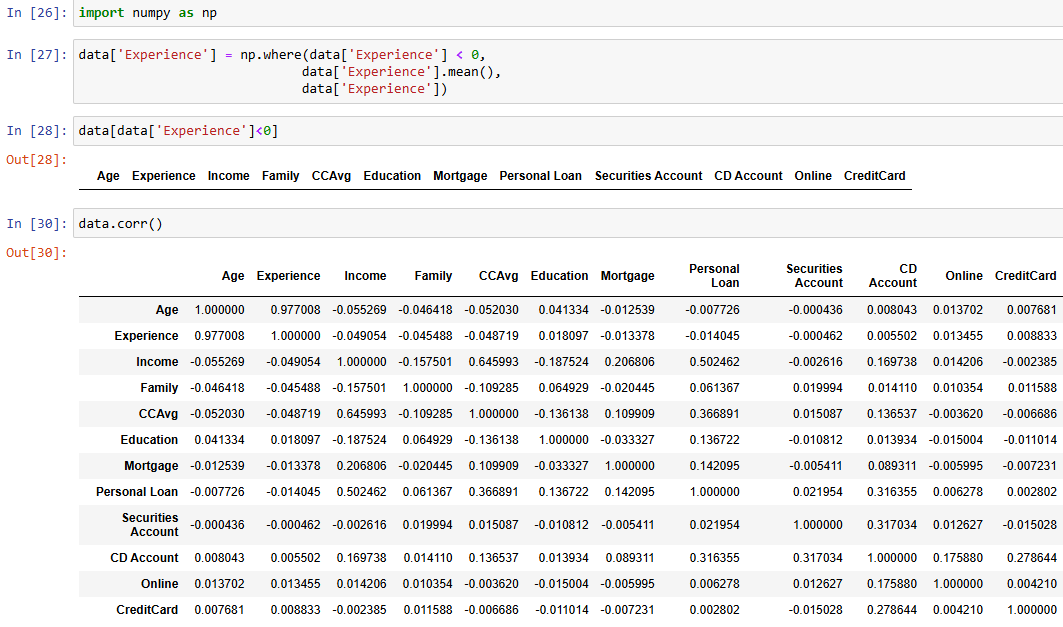
**Feature Engineering:**

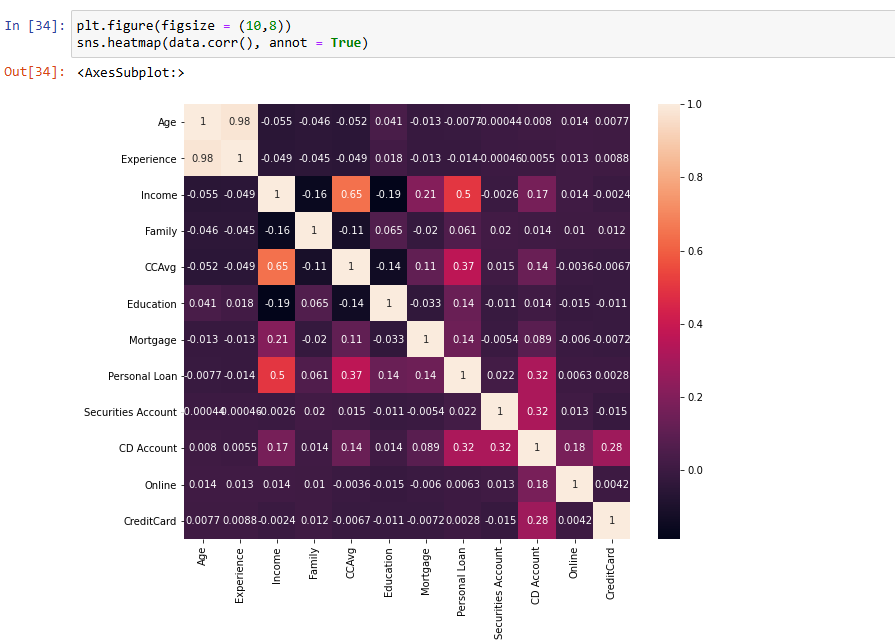
* **Account Holder Category**: A new feature was created by combining 'Securities Account' and 'CD Account' to classify customers into four categories:
  + Holds Securities & Deposit
  + Holds only Securities Account
  + Holds only Deposit Account
  + Does not hold Securities & Deposit Account.



**Correlation Analysis**:

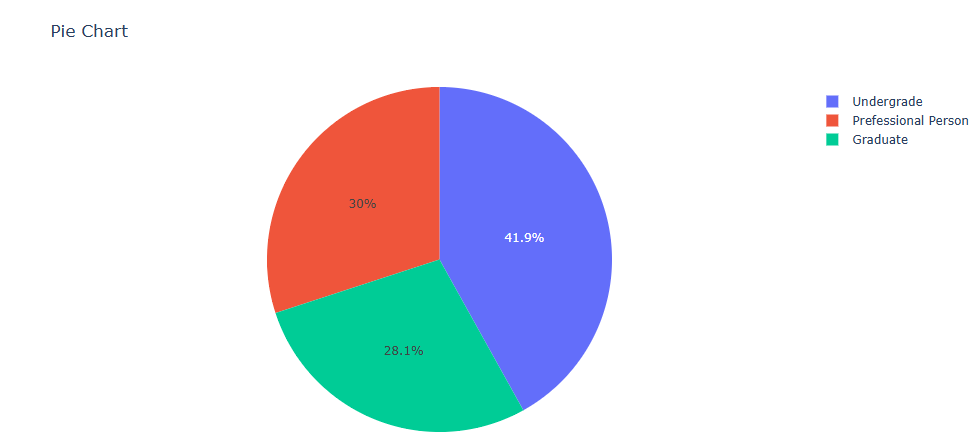
* A correlation matrix was plotted to analyze the relationships between variables. Features such as 'Income' and 'Credit Card Average' showed stronger correlations with the target variable, 'Personal Loan'.





### Exploratory Data Analysis (EDA):

1. **Distribution of Key Variables**:
   * Age, Income, and Credit Card Average (CCAvg) showed significant variation, with skewness observed in the distribution of Income and CCAvg.
   * A comparison of the income distribution for customers who accepted and did not accept the loan revealed that customers who accepted the loan generally had higher incomes.
2. **Visual Insights**:
   * **Box Plots**: Used to identify outliers and trends across categories like Education, Mortgage, and Income.
   * **Pie Charts**: Used to show the distribution of educational backgrounds and account holder categories.
   * **Count Plots**: Illustrated the relationship between customer segments (e.g., Online Banking, Credit Card Usage) and loan acceptance.



### Feature Selection:

The most relevant features identified through EDA and correlation analysis include:

* **Age**
* **Income**
* **CCAvg (Credit Card Average)**
* **Education (Converted to 'EDU')**
* **Account Holder Category**
* **Mortgage**
* **Online Banking Status**

These features were selected to build predictive models.

