Credit card Approval Predication

Presentation by

Akhilesh Maurya

Introduction

- In this project, I applied machine learning techniques to predict credit card application approvals. By analyzing various applicant attributes and building predictive models, the goal is to understand the factors influencing approval decisions and improve the decision-making process.

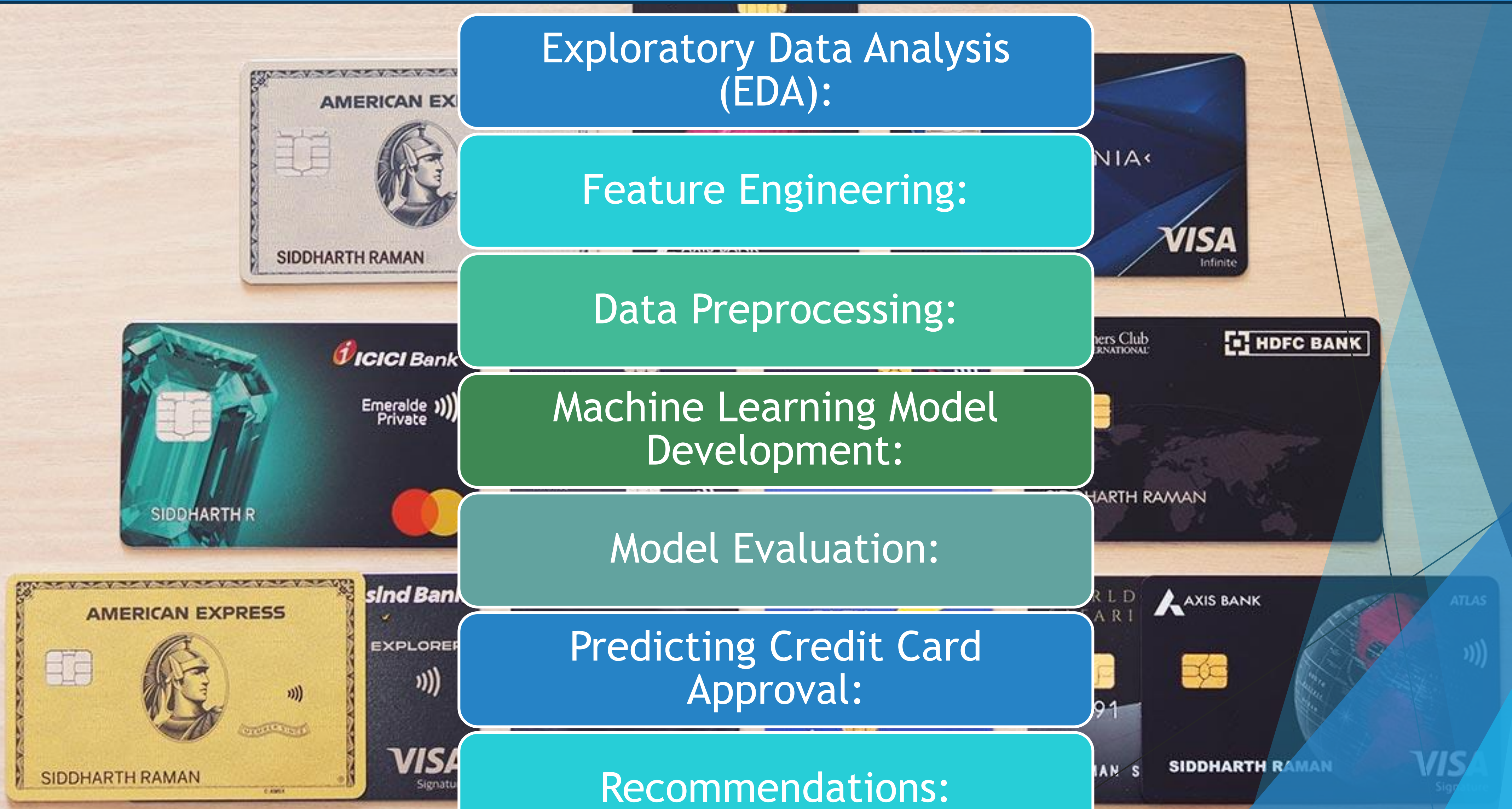
Problem Statement

- The primary objective of this project is to predict the approval or rejection of credit card applications.
- The challenge lies in understanding the key factors influencing credit card approval decisions and
- building a predictive model to assist in the decision-making process.

Dataset Overview

- **ID:** Unique identifier for each record.
- **Gender:** Applicant's gender.
- **Has a car:** Indicates car ownership (0 or 1).
- **Has a property:** Indicates property ownership (0 or 1).
- **Children count:** Number of children.
- **Income:** Applicant's income.
- **Employment status:** Current employment status.
- **Education level:** Highest education level attained.
- **Marital status:** Marital status.
- **Dwelling:** Type of dwelling.
- **Age:** Applicant's age.
- **Employment length:** Duration of current employment.
- **Has a mobile phone:** (0 or 1).
- **Has a work phone:** (0 or 1).
- **Has a phone:** (0 or 1).
- **Has an email:** (0 or 1).
- **Job title:** Applicant's job title.
- **Family member count:** Number of family members.
- **Account age:** Age of the applicant's account.
- **Is high risk (Target Variable):** Indicates if the applicant is high risk (0 for no, 1 for yes).

Steps



- Exploratory Data Analysis (EDA):
- Feature Engineering:
- Data Preprocessing:
- Machine Learning Model Development:
- Model Evaluation:
- Predicting Credit Card Approval:
- Recommendations:

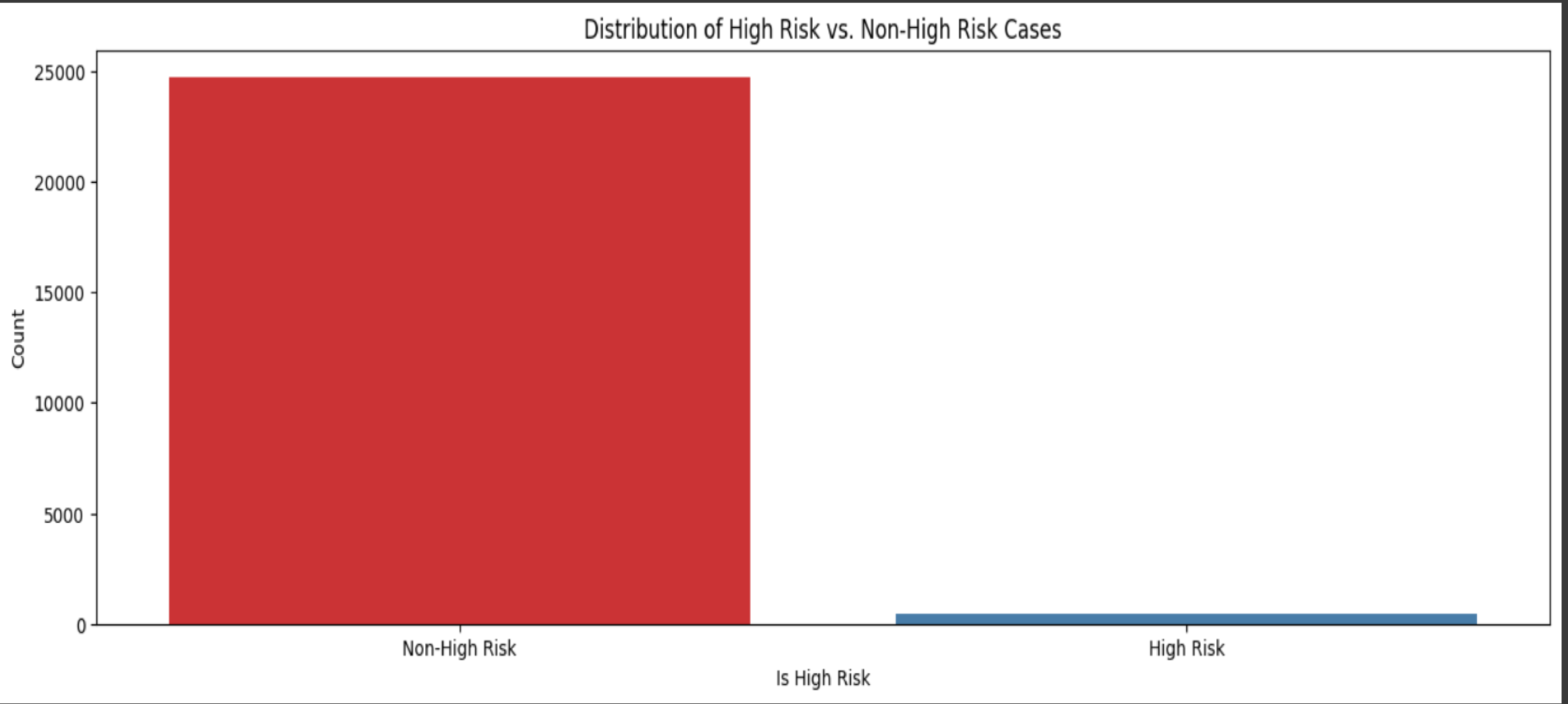
1. Exploratory Data Analysis (EDA):

	ID	Gender	Has a car	Has a property	Children count	Income	Employment status	Education level	Marital status	Dwelling	Age	Employment length	Has a mobile phone	Has a work phone	Has a phone	Has an email	Job title	Family member count	Account age	Is high risk
0	5037048	M	Y	Y	0	135000.0	Working	Secondary / secondary special	Married	With parents	-16271	-3111	1	0	0	0	Core staff	2.0	-17.0	0
1	5044630	F	Y	N	1	135000.0	Commercial associate	Higher education	Single / not married	House / apartment	-10130	-1651	1	0	0	0	Accountants	2.0	-1.0	0
2	5079079	F	N	Y	2	180000.0	Commercial associate	Secondary / secondary special	Married	House / apartment	-12821	-5657	1	0	0	0	Laborers	4.0	-38.0	0
3	5112872	F	Y	Y	0	360000.0	Commercial associate	Higher education	Single / not married	House / apartment	-20929	-2046	1	0	0	1	Managers	1.0	-11.0	0
4	5105858	F	N	N	0	270000.0	Working	Secondary / secondary special	Separated	House / apartment	-16207	-515	1	0	1	0	NaN	1.0	-41.0	0

```
df.shape
```

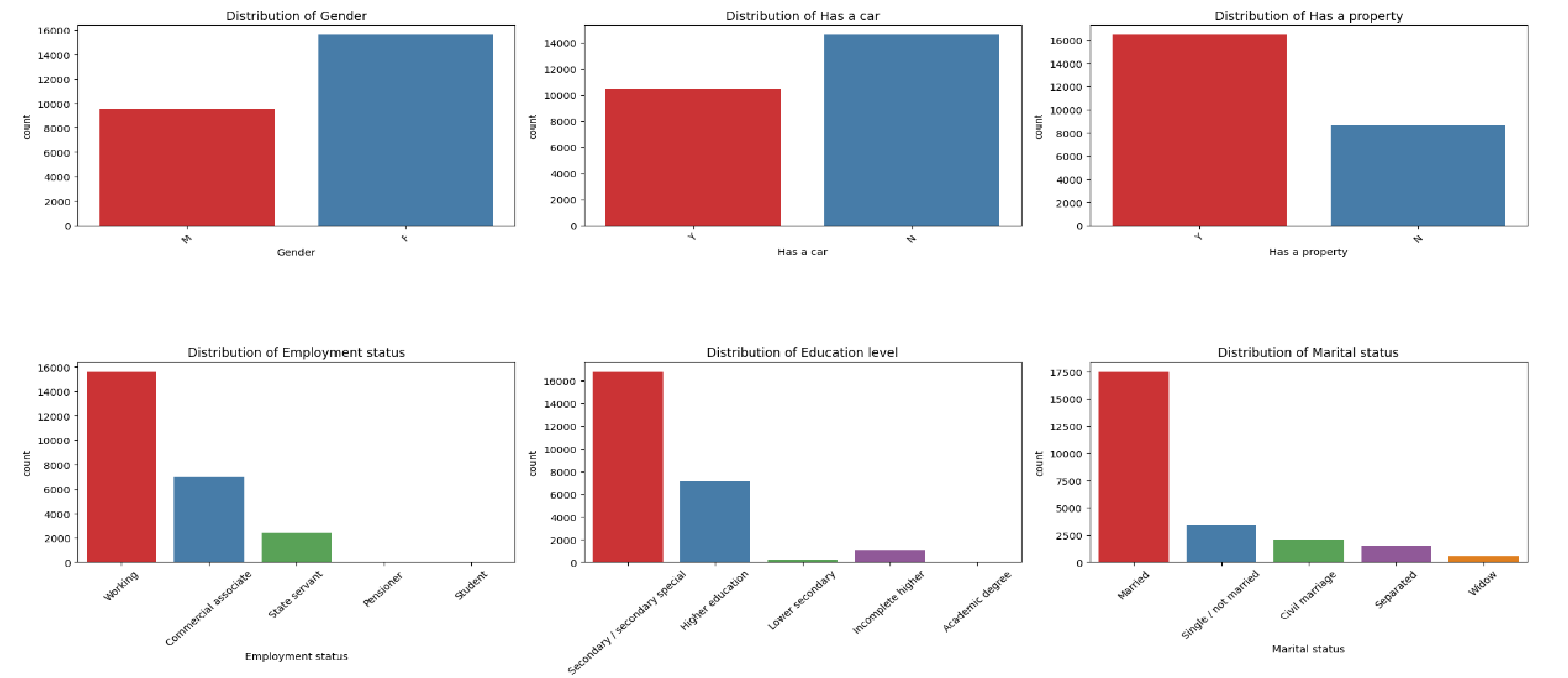
```
(36457, 20)
```

1. Exploratory Data Analysis (EDA):



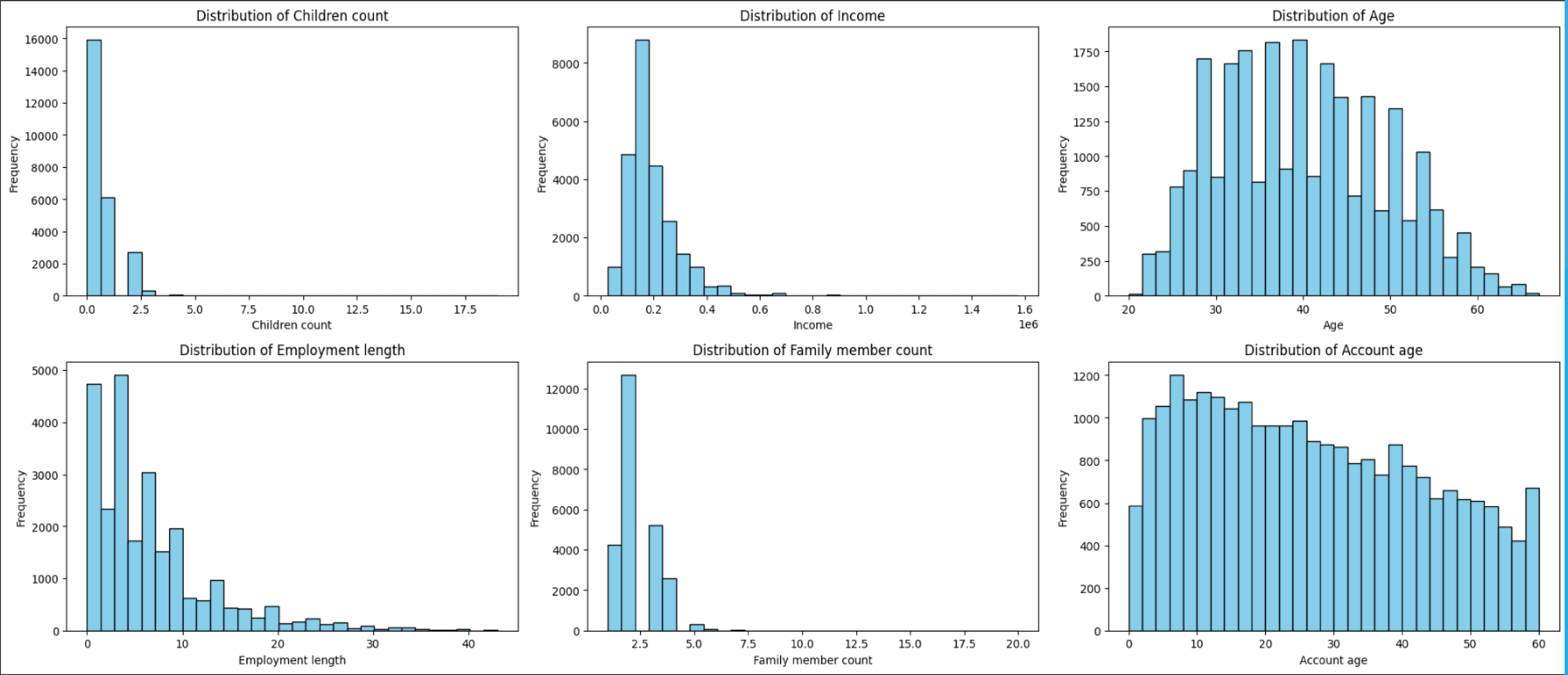
1. Exploratory Data Analysis (EDA):

Distribution of Categorical Variables



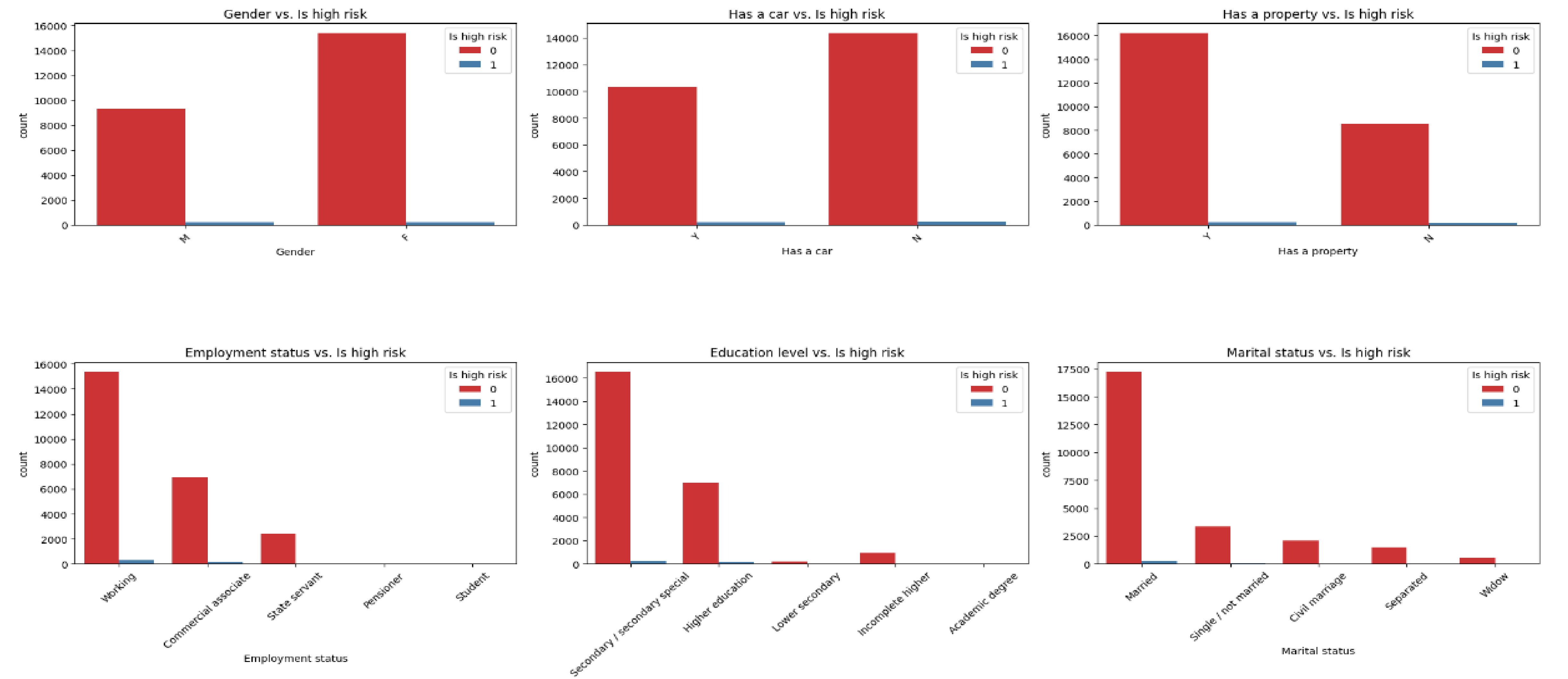
1. Exploratory Data Analysis (EDA):

Distribution of Numerical Variables



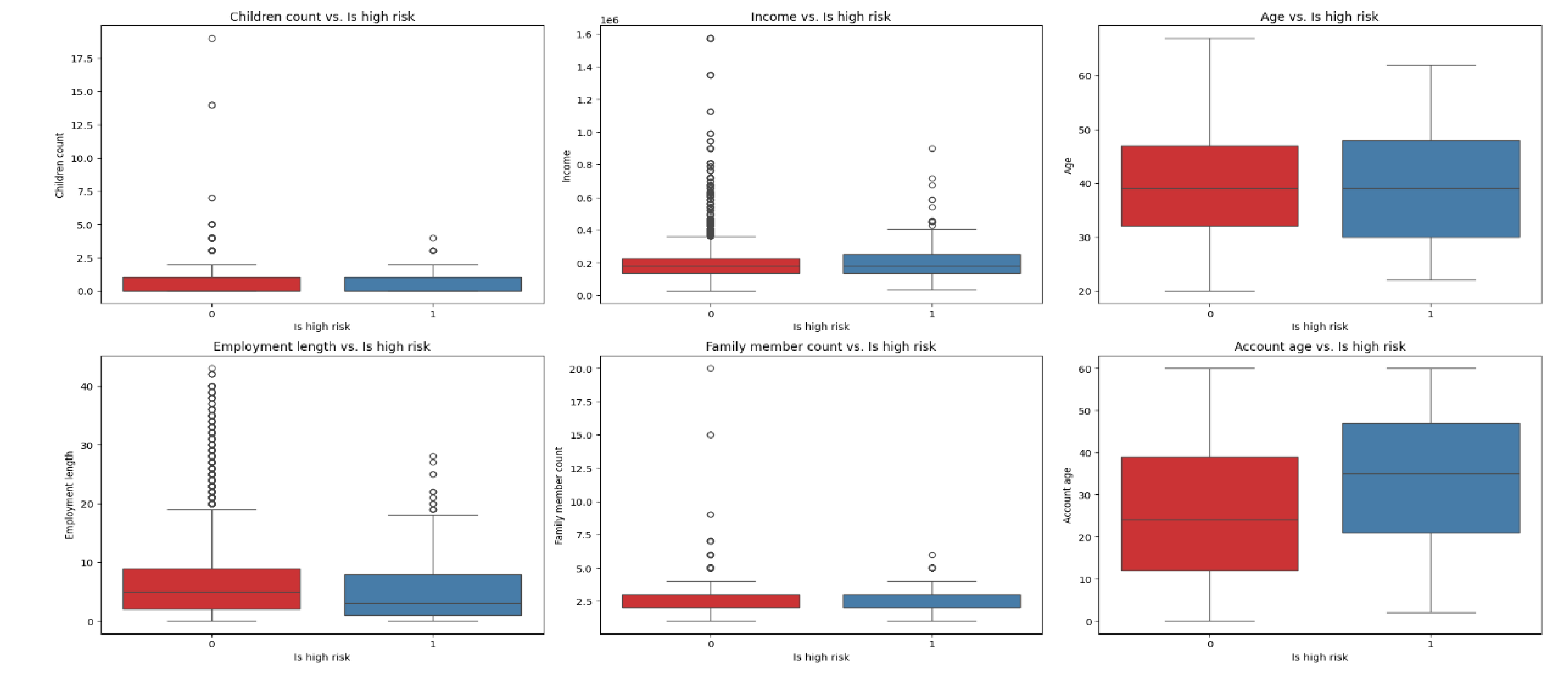
1. Exploratory Data Analysis (EDA):

Distribution of Categorical vs Target Variables



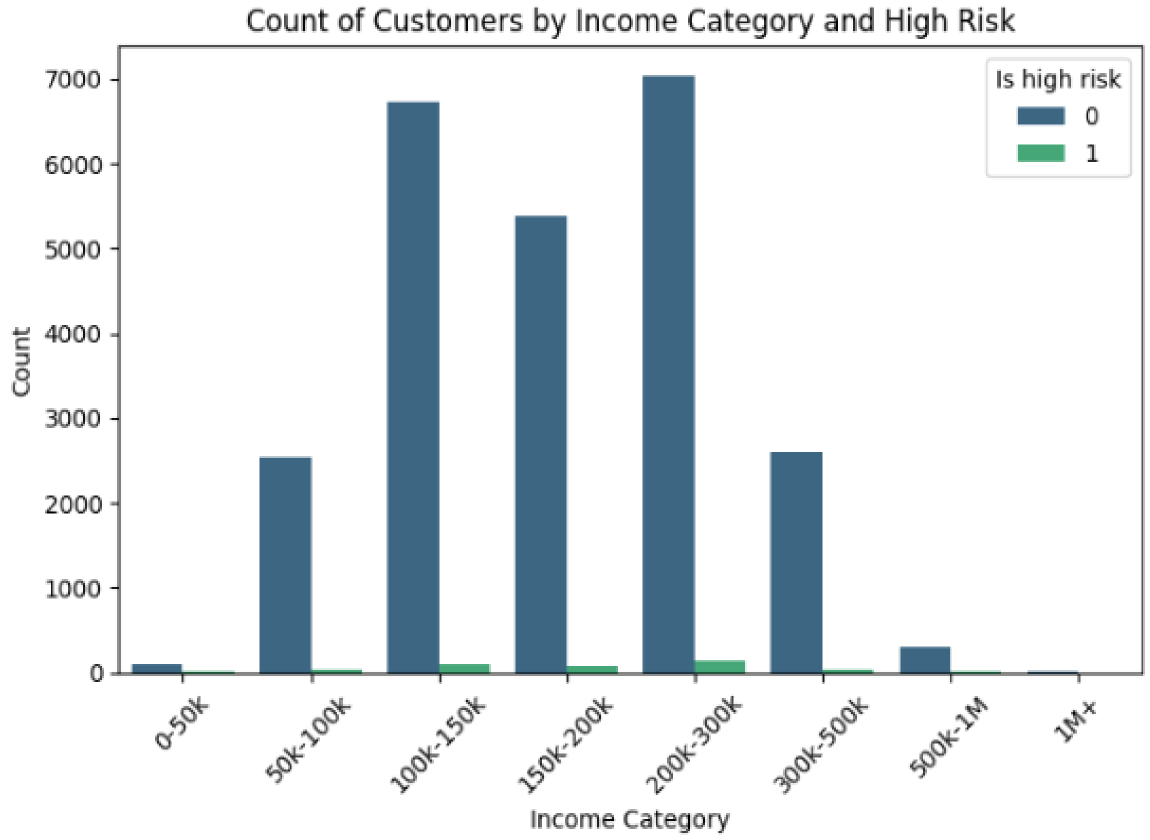
1. Exploratory Data Analysis (EDA):

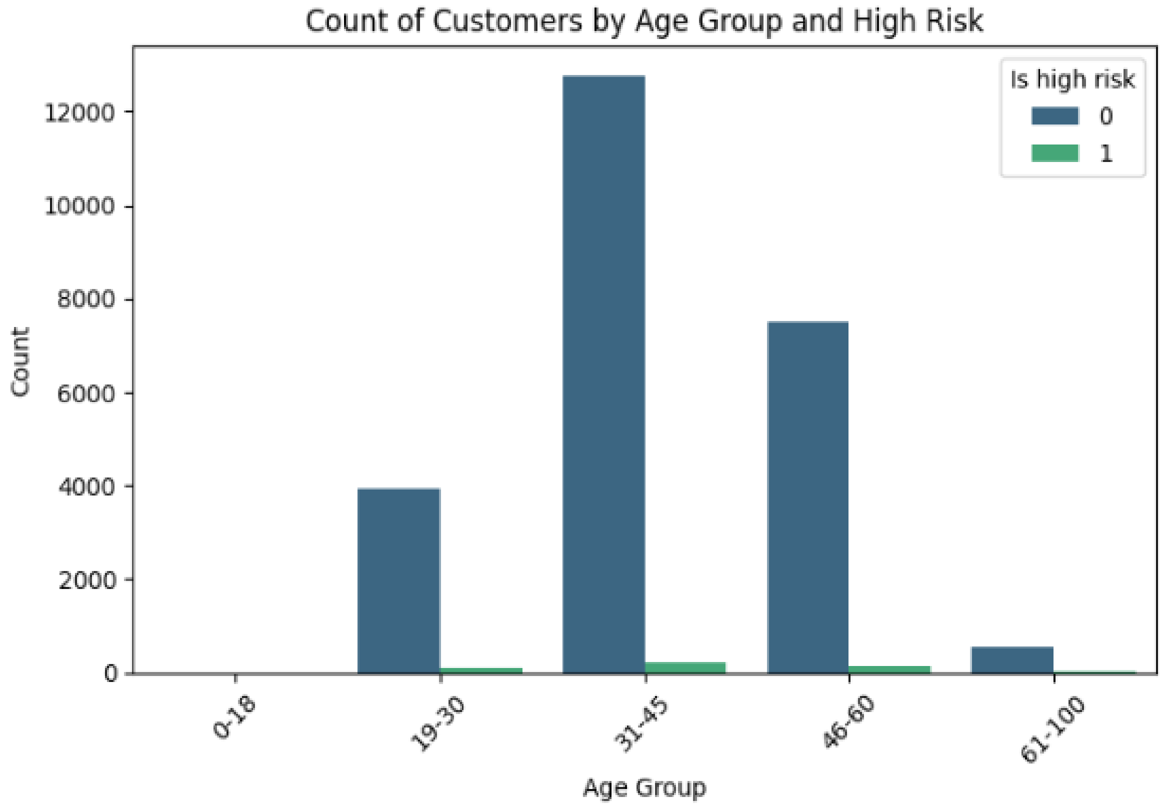
Distribution of Numerical vs target Variables



2. Feature Engineerig

Creating 2 New columns





3. Data Preprocessing

Encoding Catgorical to Numerical

Gender	Has a car	Has a property	Children count	Income	Age	Employment length	Has a mobile phone	Has a work phone	...	Job title_Security staff	Job title_Waiters/barmen staff	Age Group_31-45	Age Group_46-60	Age Group_61-100	I Category
1	1	1	0	135000.0	44	8	0	0	...	0	0	1	0	0	
0	1	0	1	135000.0	27	4	0	0	...	0	0	0	0	0	
0	0	1	2	180000.0	35	15	0	0	...	0	0	1	0	0	
0	1	1	0	360000.0	57	5	0	0	...	0	0	0	1	0	
0	1	1	0	135000.0	36	10	0	1	...	0	0	1	0	0	

4. Machine Learning Model Development

```
X = df_encoded.drop(columns=['Is high risk', 'ID'])
y = df_encoded['Is high risk']
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
y_train.value_counts()
```

```
Is high risk
0    17464
1      293
Name: count, dtype: int64
```

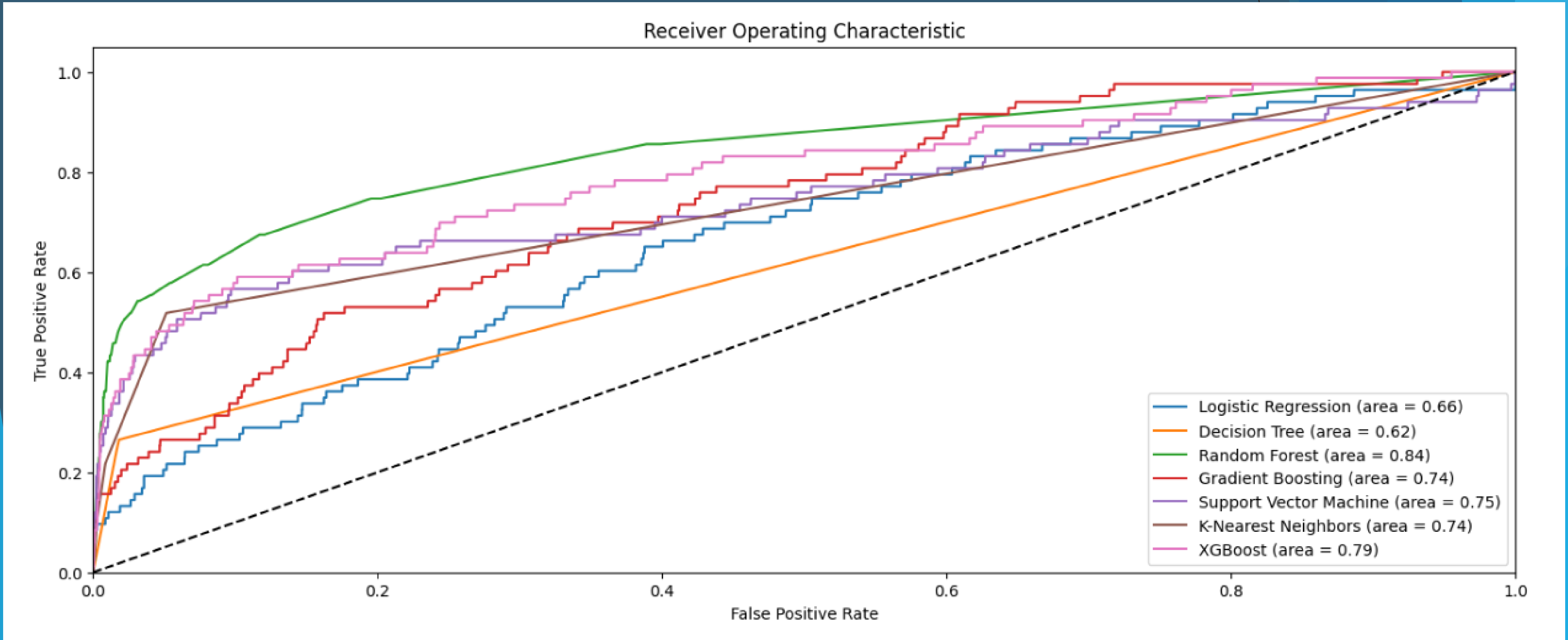
```
models = {
    'Logistic Regression': LogisticRegression(max_iter=1000, random_state=42),
    'Decision Tree': DecisionTreeClassifier(random_state=42),
    'Random Forest': RandomForestClassifier(random_state=42),
    'Gradient Boosting': GradientBoostingClassifier(random_state=42),
    'Support Vector Machine': SVC(probability=True, random_state=42),
    'K-Nearest Neighbors': KNeighborsClassifier(),
    'XGBoost': xgb.XGBClassifier(eval_metric='logloss', use_label_encoder=False, random_state=42)
}
```

5. Model Evaluation

```
Model: Random Forest
Confusion Matrix:
[[4341   16]
 [  65   18]]

Classification Report:
              precision    recall  f1-score   support

           0       0.99      1.00      0.99      4357
           1       0.53      0.22      0.31        83

    accuracy                           0.98      4440
   macro avg       0.76      0.61      0.65      4440
weighted avg       0.98      0.98      0.98      4440

=======================================================
Model: Gradient Boosting
Confusion Matrix:
[[4355    2]
 [  79    4]]

Classification Report:
              precision    recall  f1-score   support

           0       0.98      1.00      0.99      4357
           1       0.67      0.05      0.09        83

    accuracy                           0.98      4440
   macro avg       0.82      0.52      0.54      4440
weighted avg       0.98      0.98      0.97      4440

=======================================================
Model: Support Vector Machine
Confusion Matrix:
[[4357    0]
 [  83    0]]
```

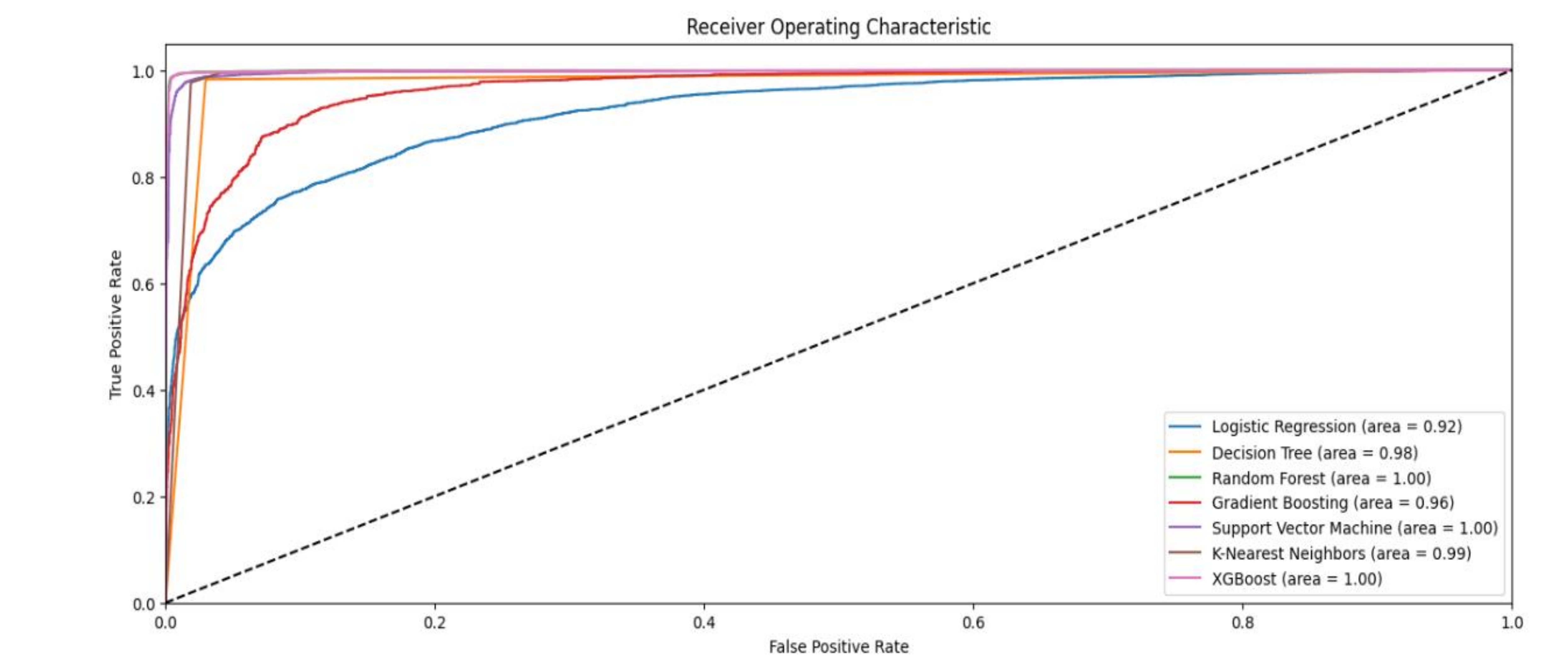
5. Model Evaluation

data is imbalanced

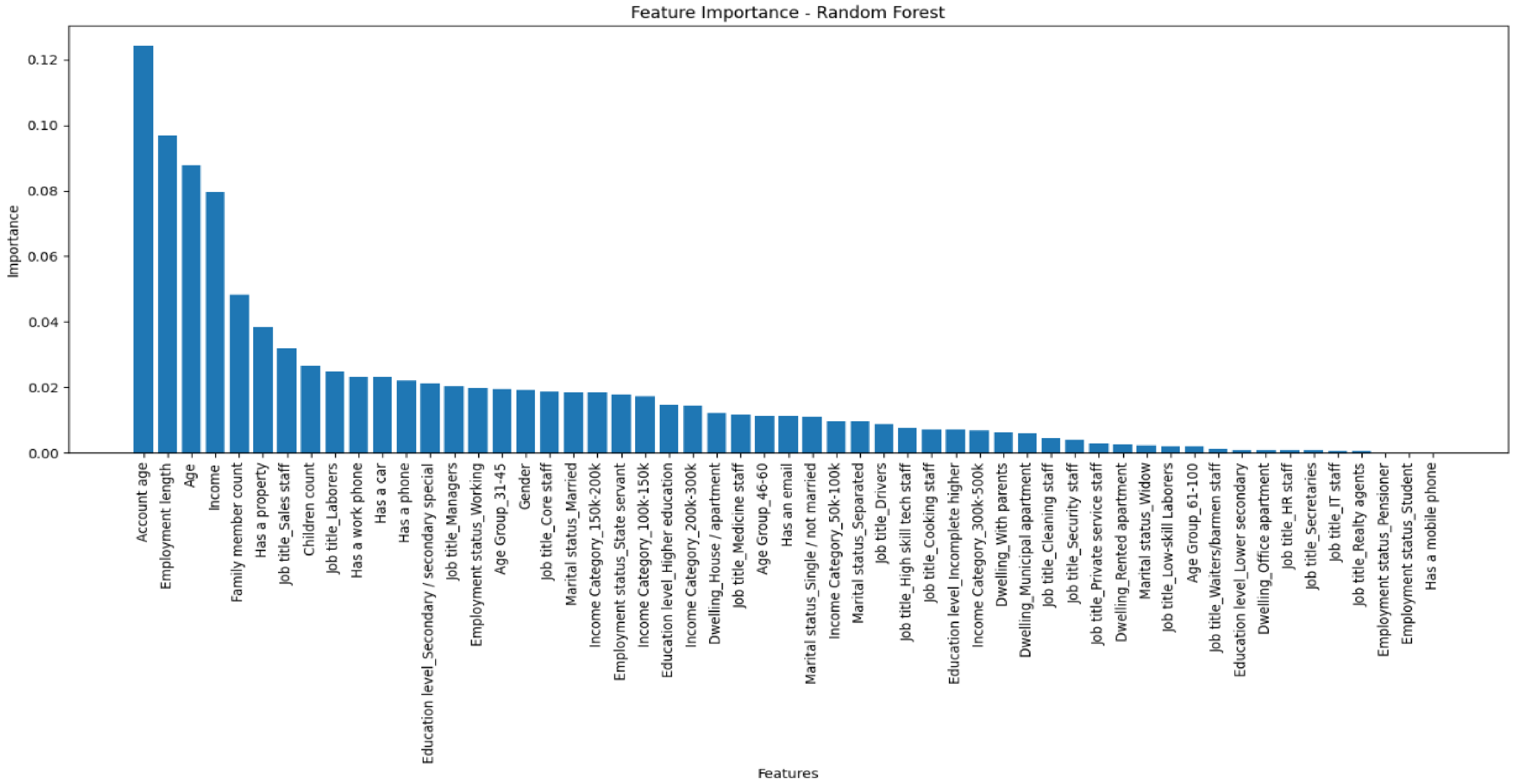


6. Making data balanced

I Apply Smote



7. Feature Importance



8. Conclusion

- Our Random Forest model achieved an impressive 99% accuracy in predicting credit card approvals.
- Key factors influencing approval include income, employment length, age, marital status, and credit history.
- These insights highlight the importance of stable income and long-term employment in the approval process.
- By focusing on these critical factors, financial institutions can better assess applicant risk, improve decision-making, and enhance the credit approval process.