

Data Analysis of Bank Customer Churn



Humza Khan

Agenda



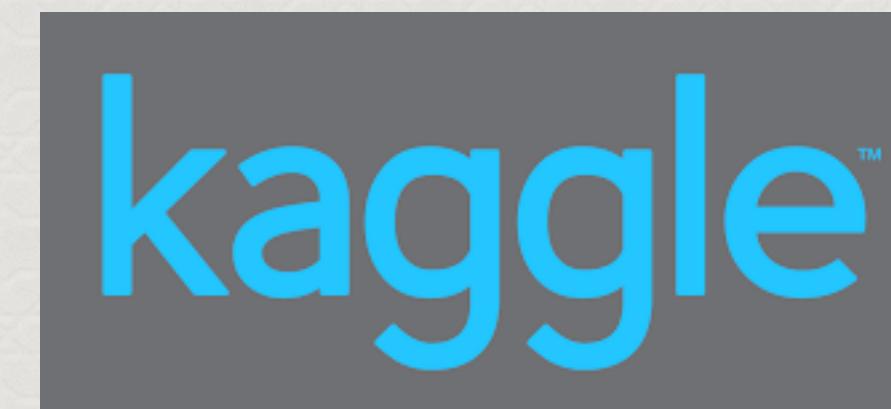
- ◆ Motivation
- ◆ Data Collection
- ◆ Exploratory Data Analysis
- ◆ Feature Engineering
- ◆ Modeling Results
- ◆ Feature Importance
- ◆ Future Improvements

Motivation



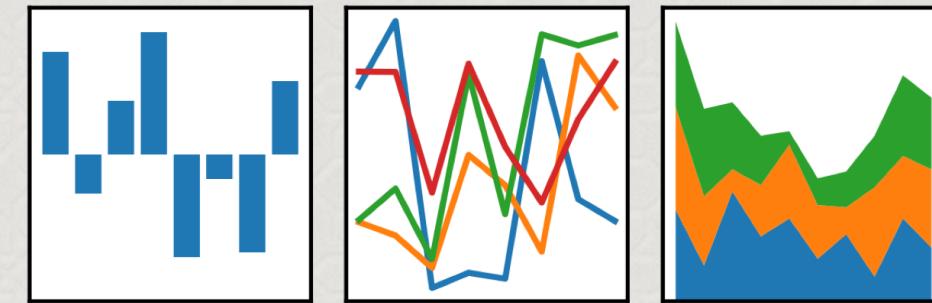
- ◆ How can I keep customers from the bank's perspective?
- ◆ There are customers that tend to leave and others that don't
- ◆ Build a model that can identify customers' characteristics

Data Collection



pandas

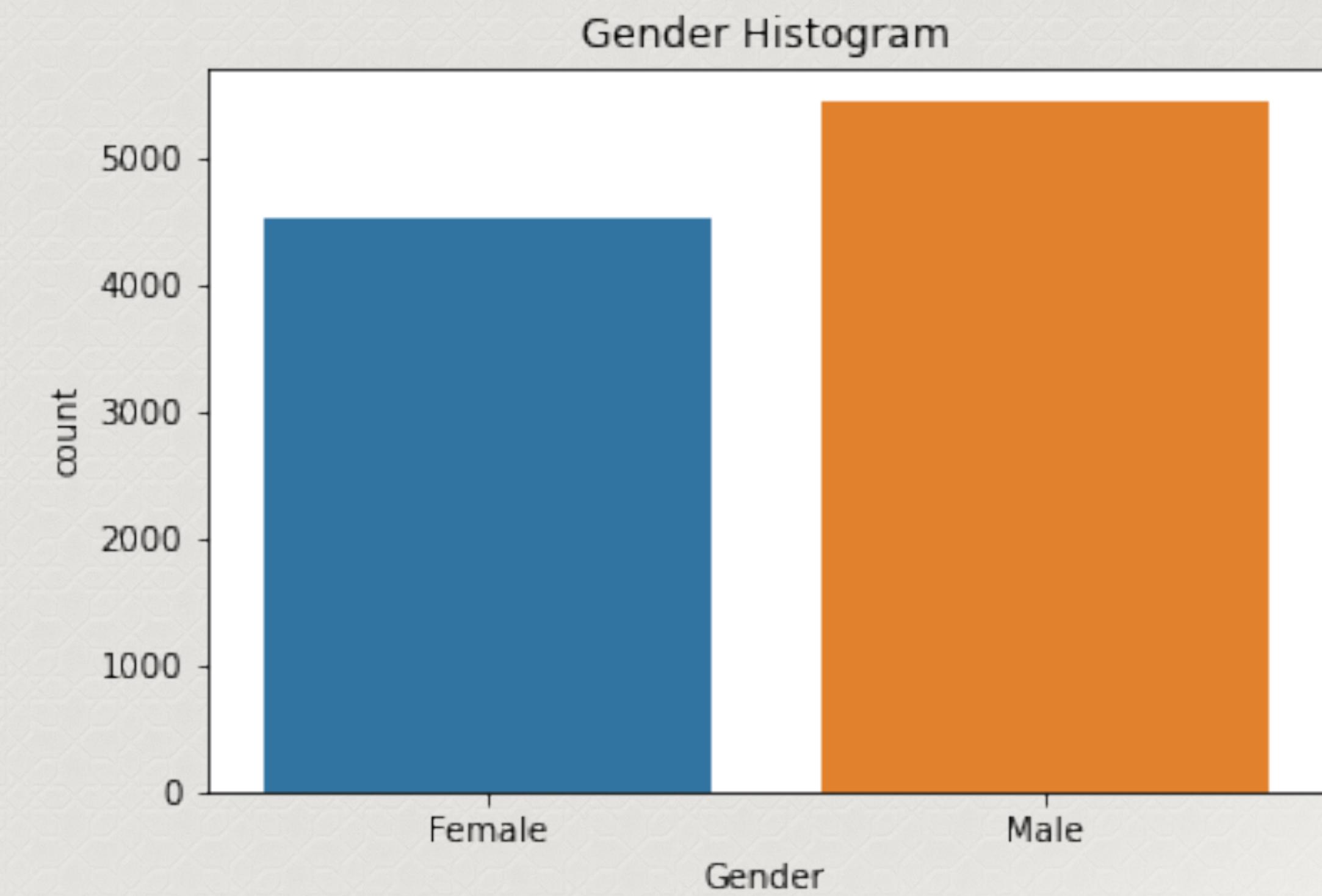
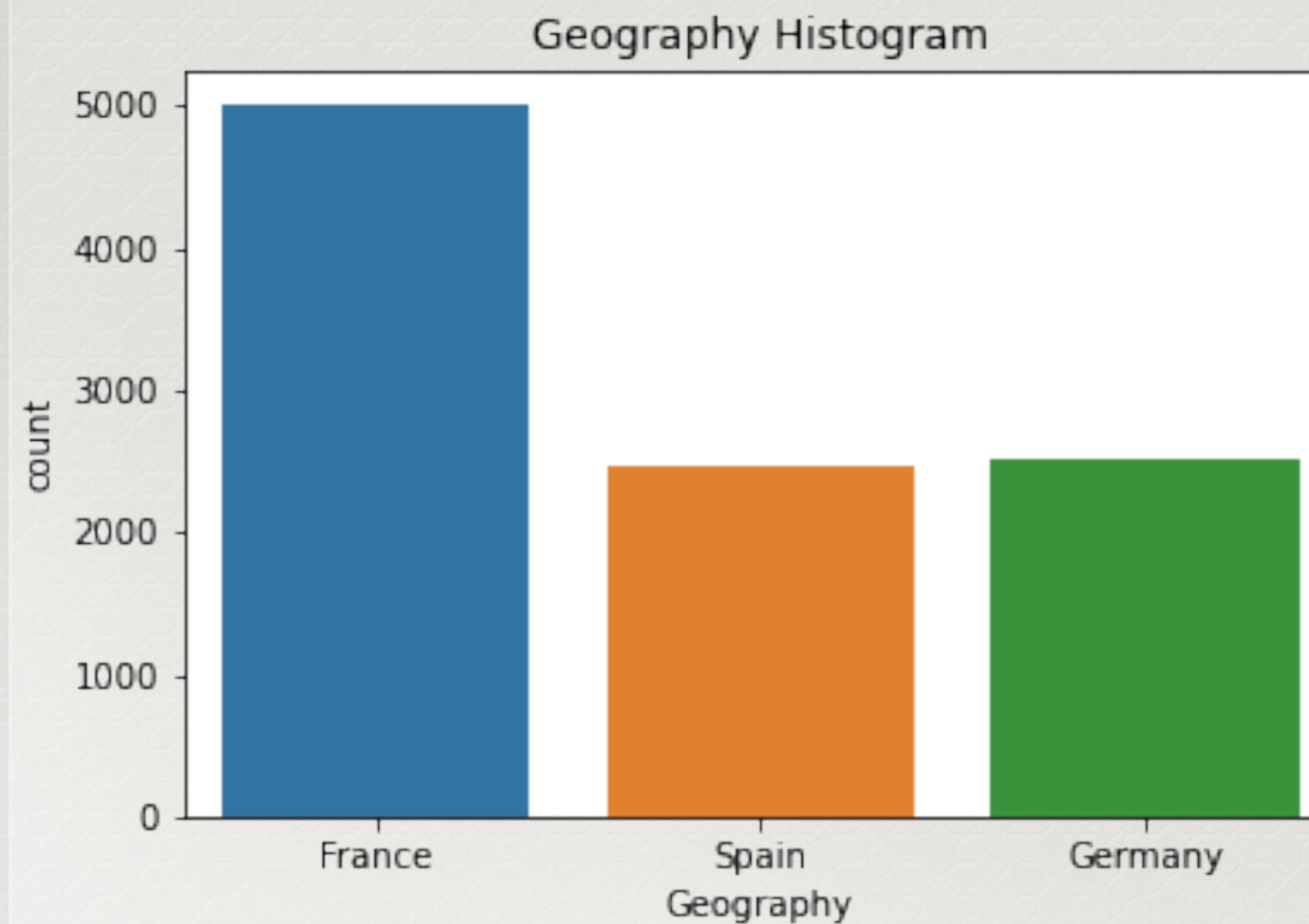
$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Features

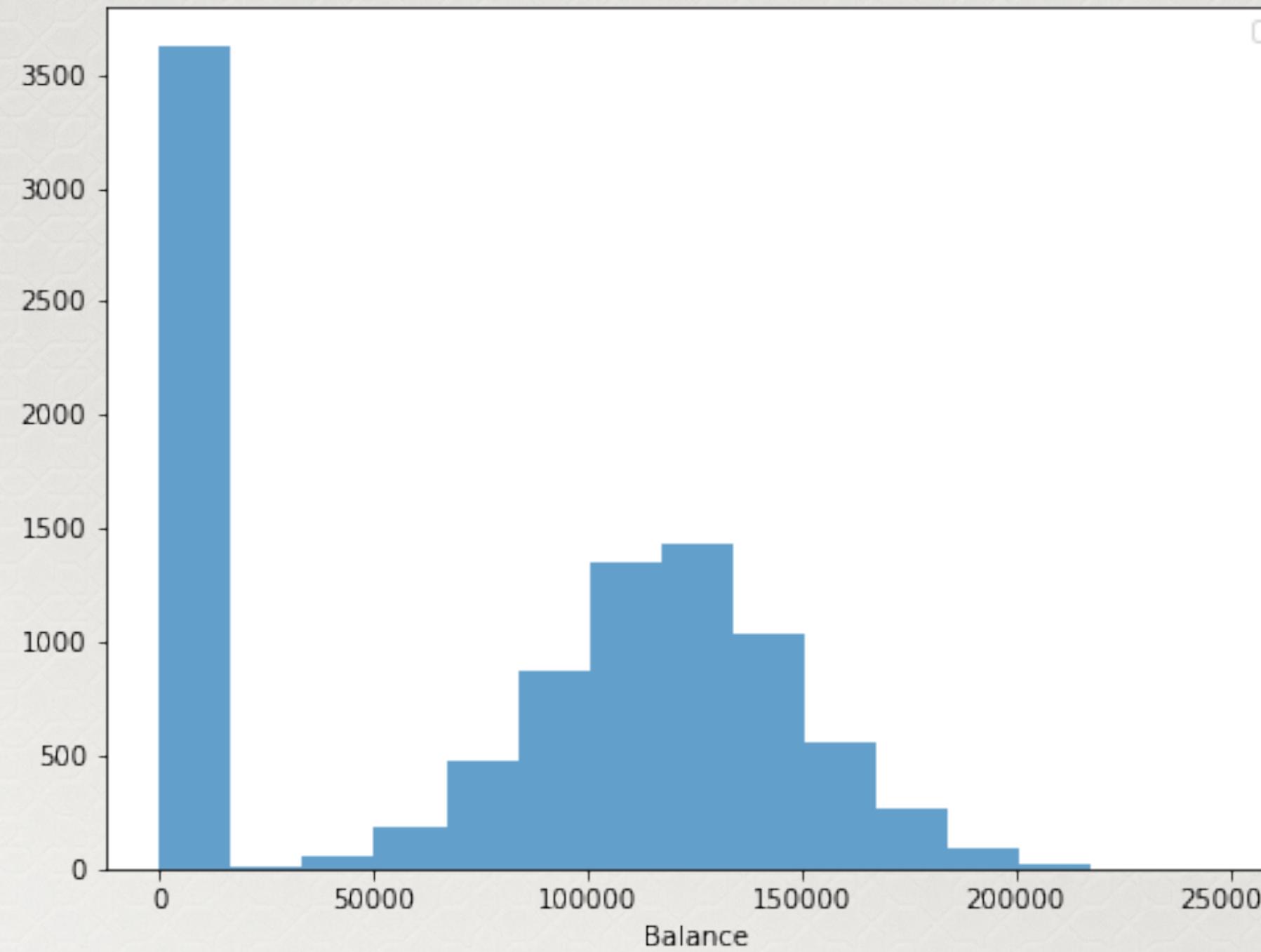
CustomerID	Estimated Salary	CreditScore	Geography
Gender	Age	Tenure	Balance
NumofProducts	HasCrCard	IsActiveMember	Exited

Exploratory Data Analysis

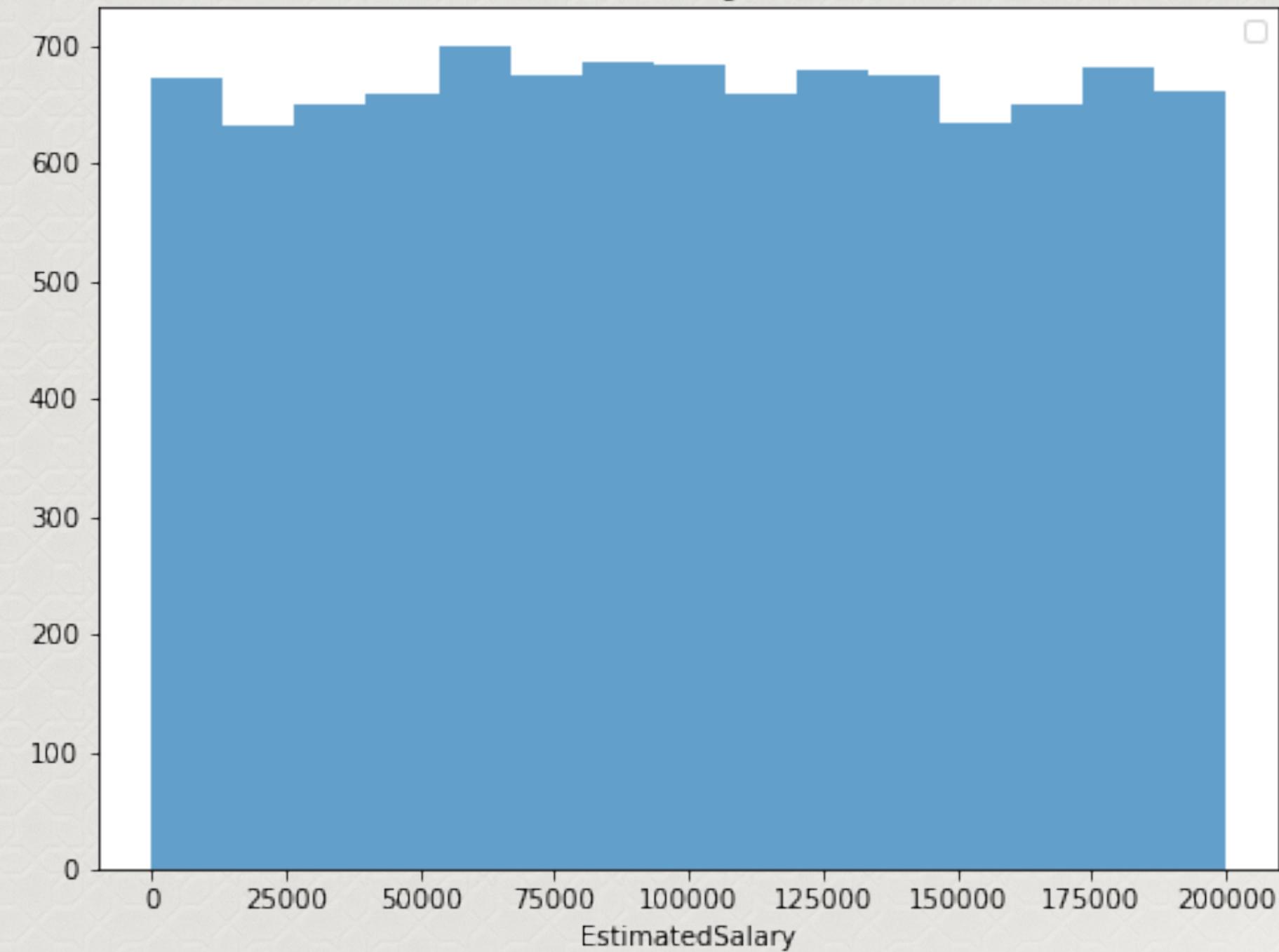


Exploratory Data Analysis

Balance Distribution



Estimated Salary Distribution



Modeling Results Before Feature Engineering

Models	AUROC	Precision	Recall
Random Forest	0.8525	0.7652	0.4644
Gradient Boosting	0.8699	0.8174	0.4644
K-Nearest Neighbors	0.7523	0.6222	0.3440
Decision Tree	0.7710	0.6473	0.4644
Logistic Regression	0.7748	0.5891	0.1867

Feature Engineering

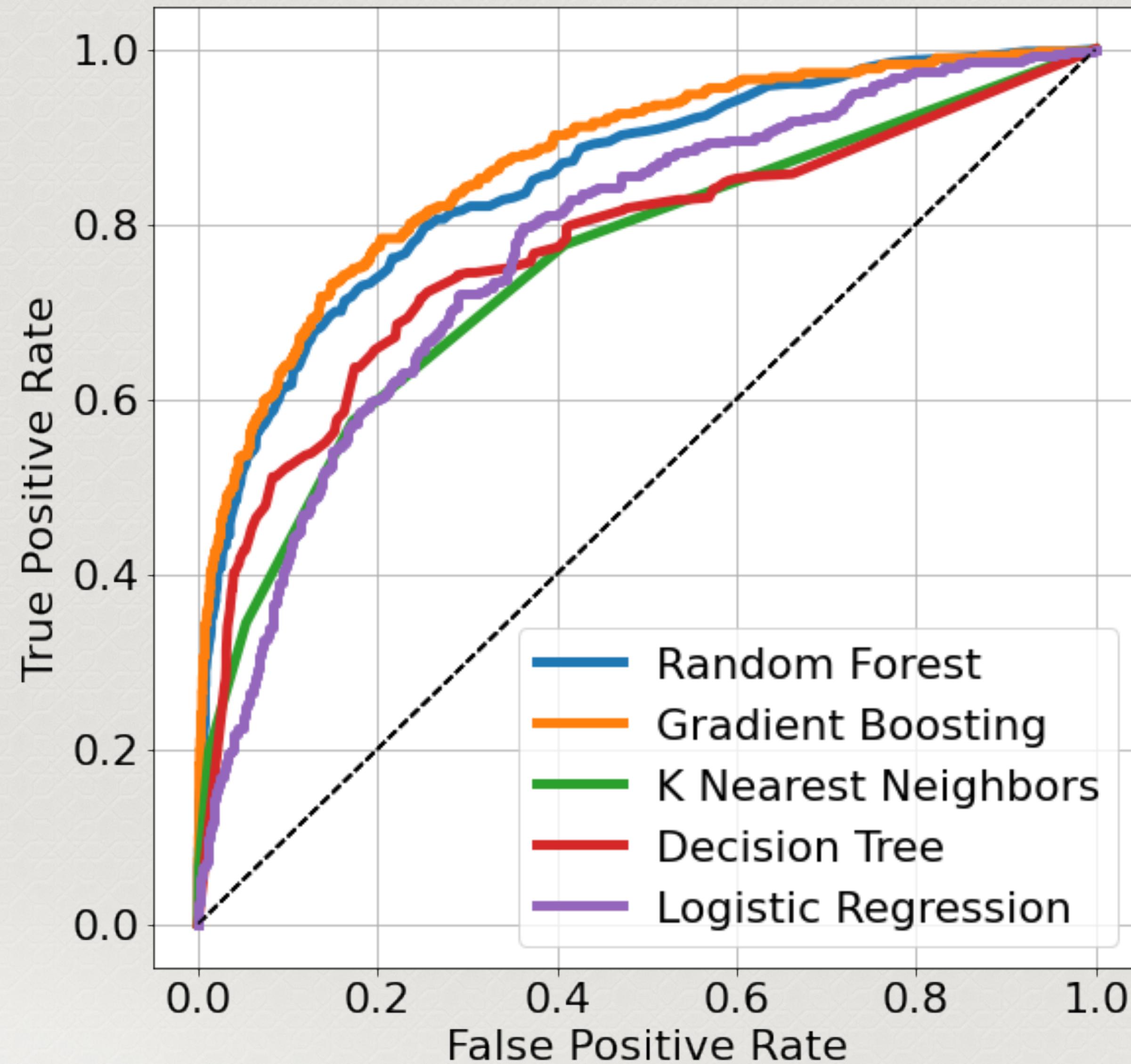
- ◆ Credit Score Table
- ◆ Number of Products by Tenure
- ◆ Estimated Salary by Product
- ◆ Tenure by Age

Category	Score Range
0	<300
1	300-499
2	500-599
3	600-660
4	661-780
5	781-850
6	>850

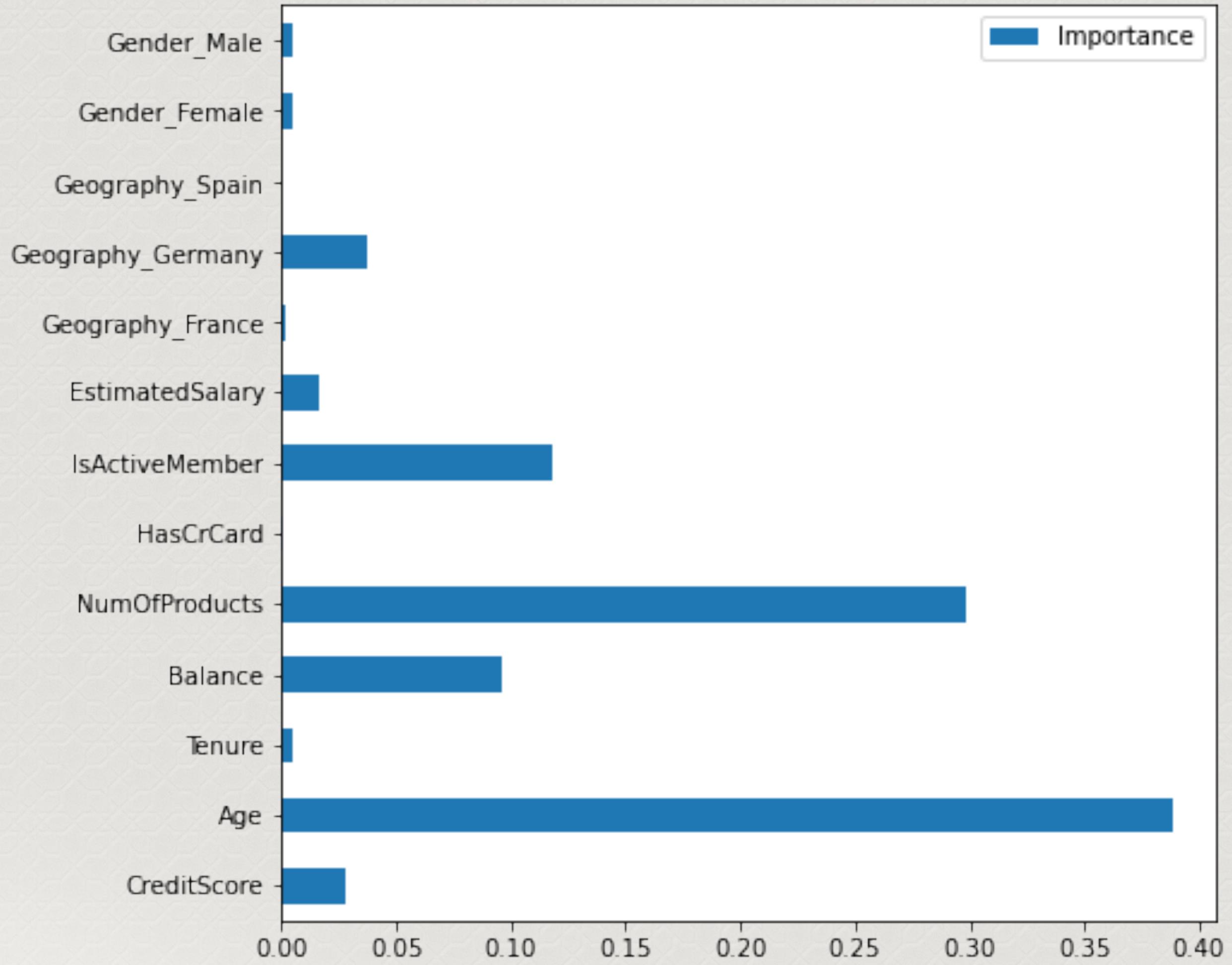
Modeling Results After Feature Engineering

Models	AUROC Before FE	AUROC After FE
Random Forest	0.8525	0.8515
Gradient Boosting	0.8699	0.8675
K-Nearest Neighbors	0.7523	0.7433
Decision Tree	0.7710	0.7671
Logistic Regression	0.7748	0.7896

ROC Curve



Gradient Boosting - Feature Importance

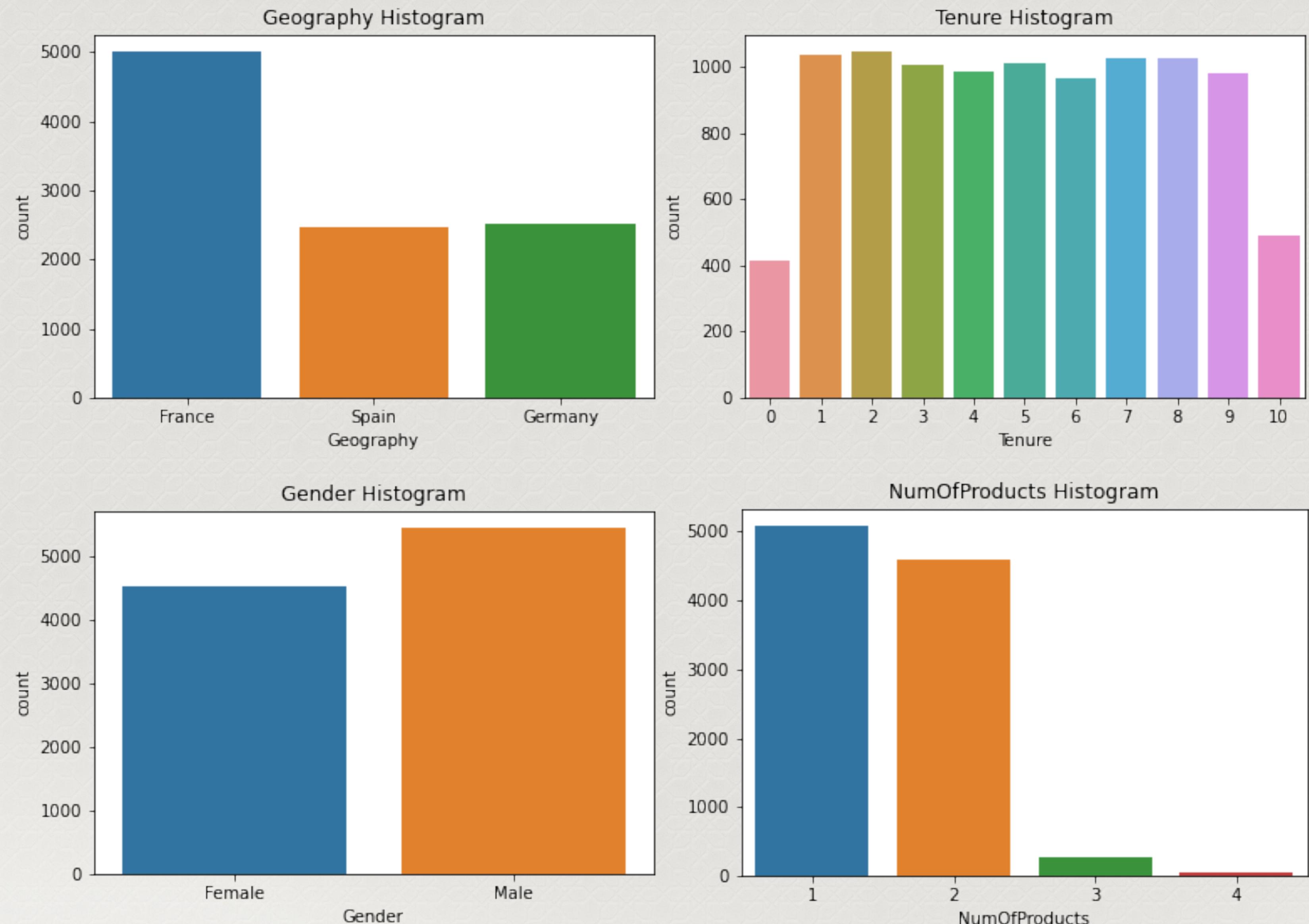


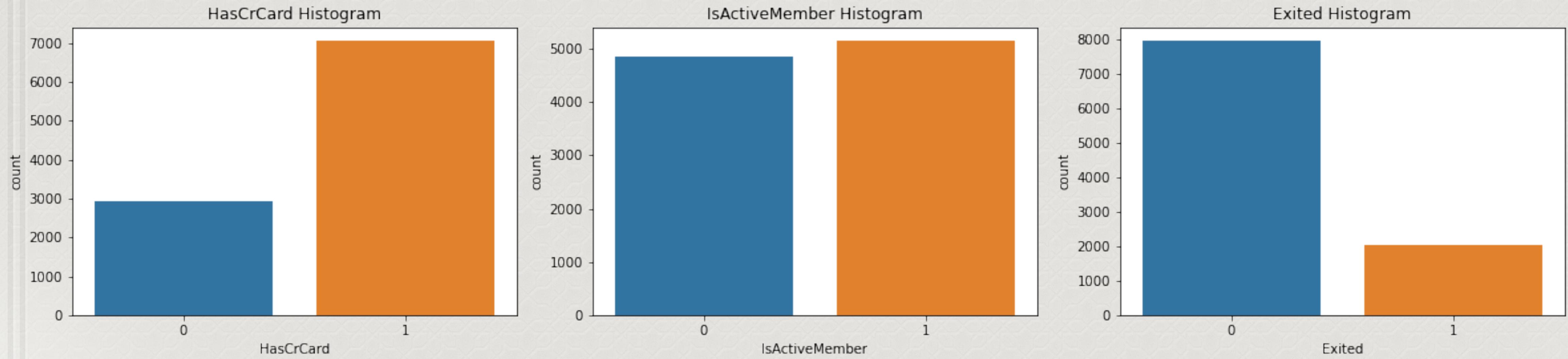
Next Steps



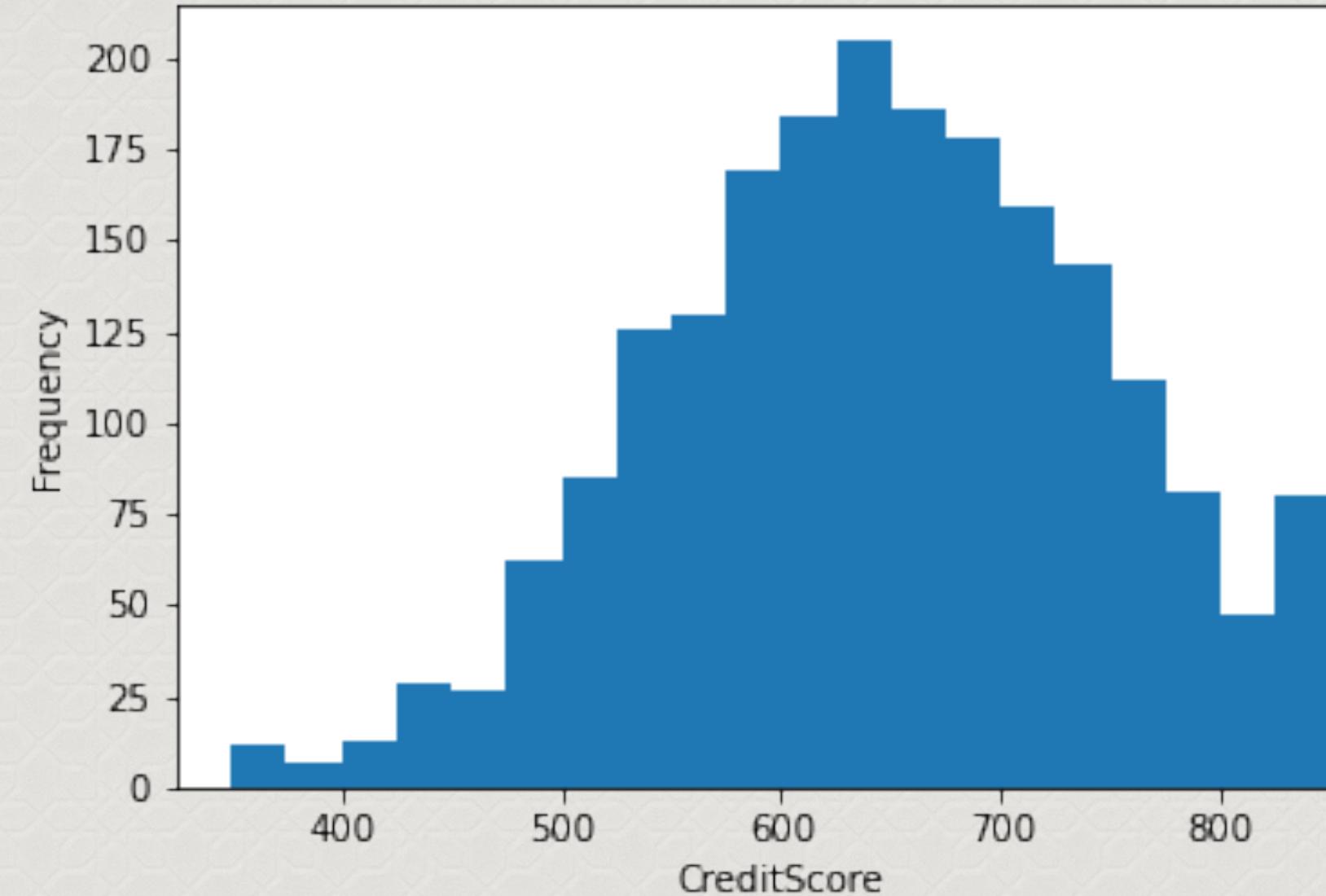
- ◆ More feature engineering
- ◆ PCA - Principal Component Analysis

Appendix

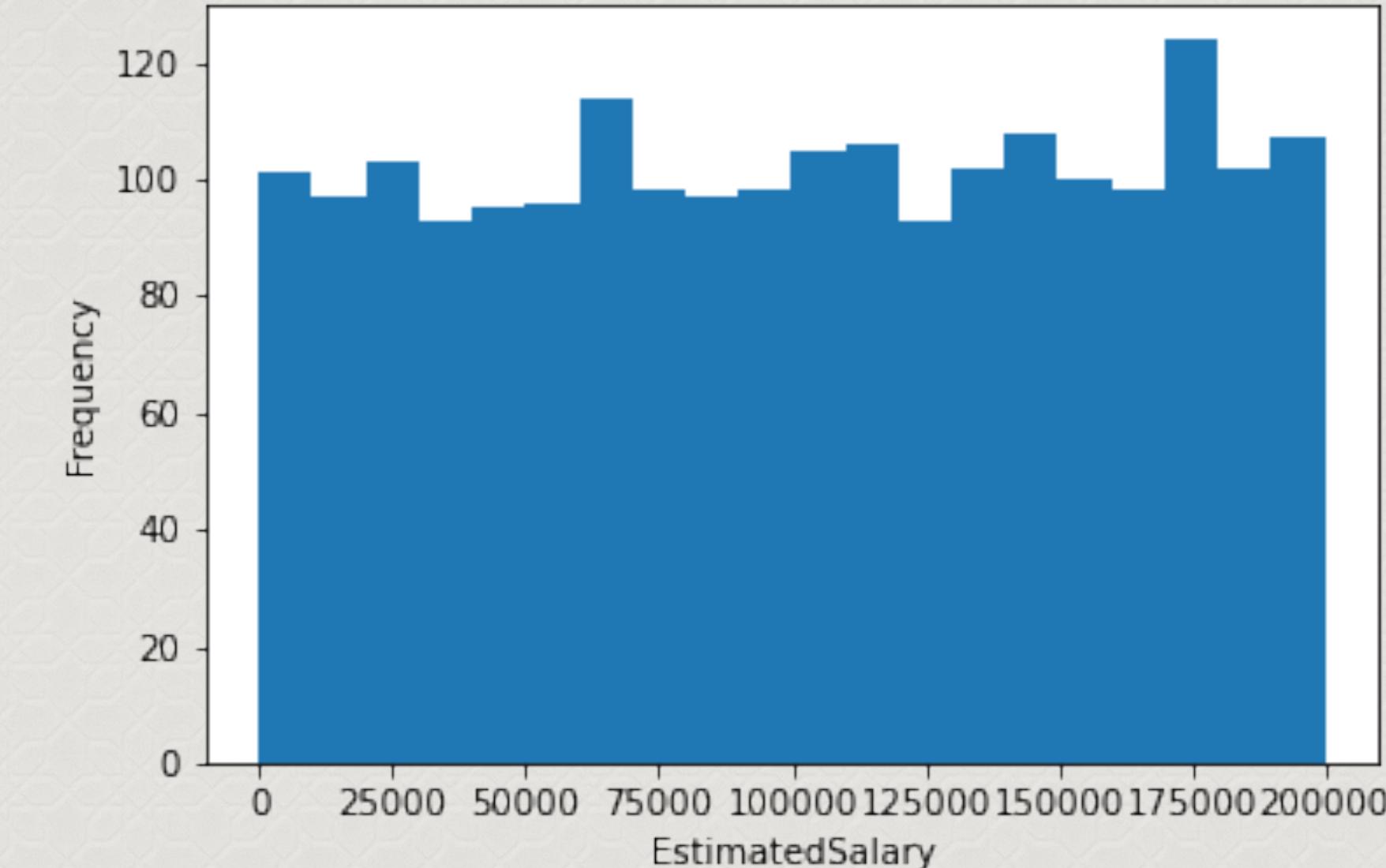




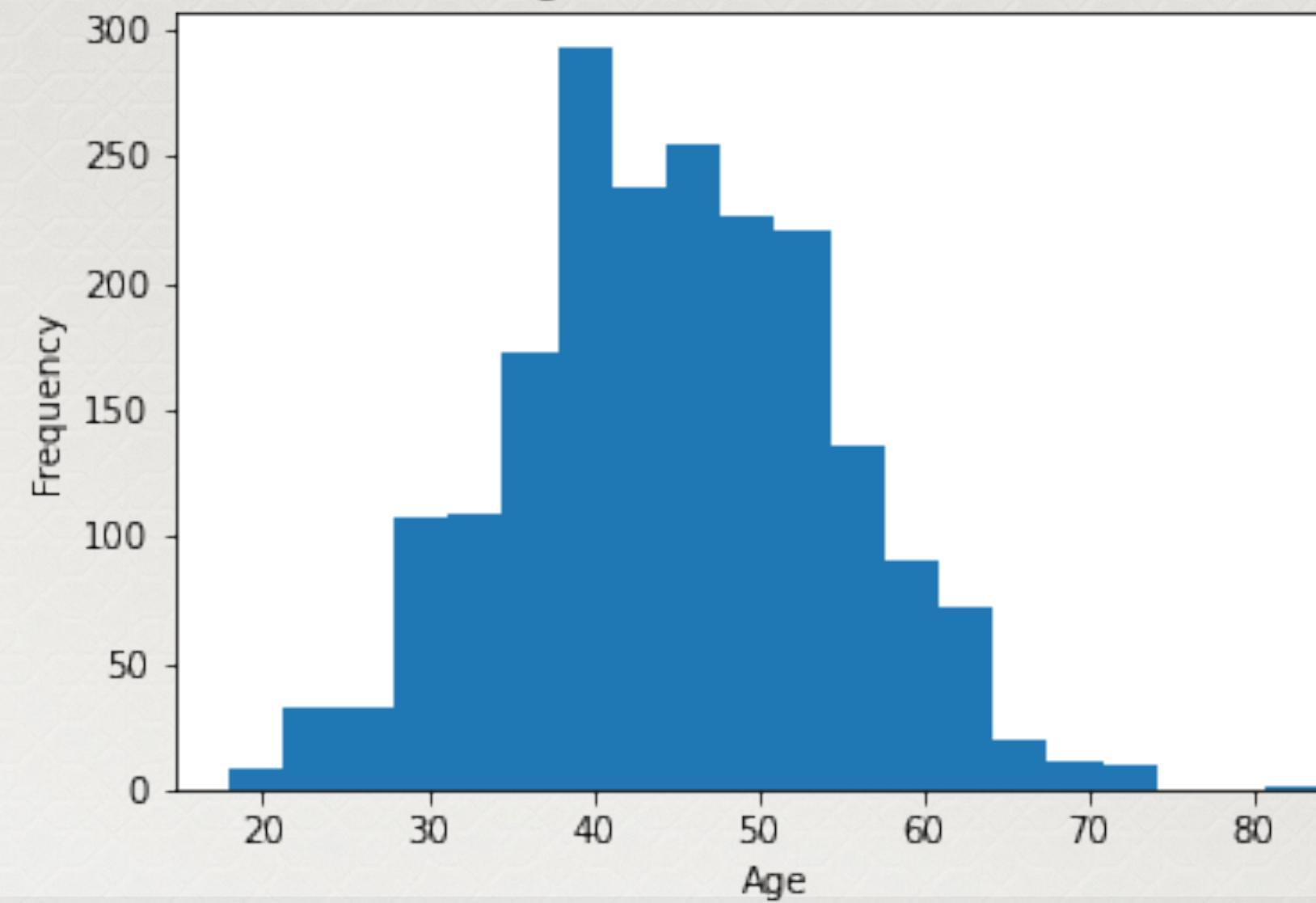
CreditScore Variable Distribution



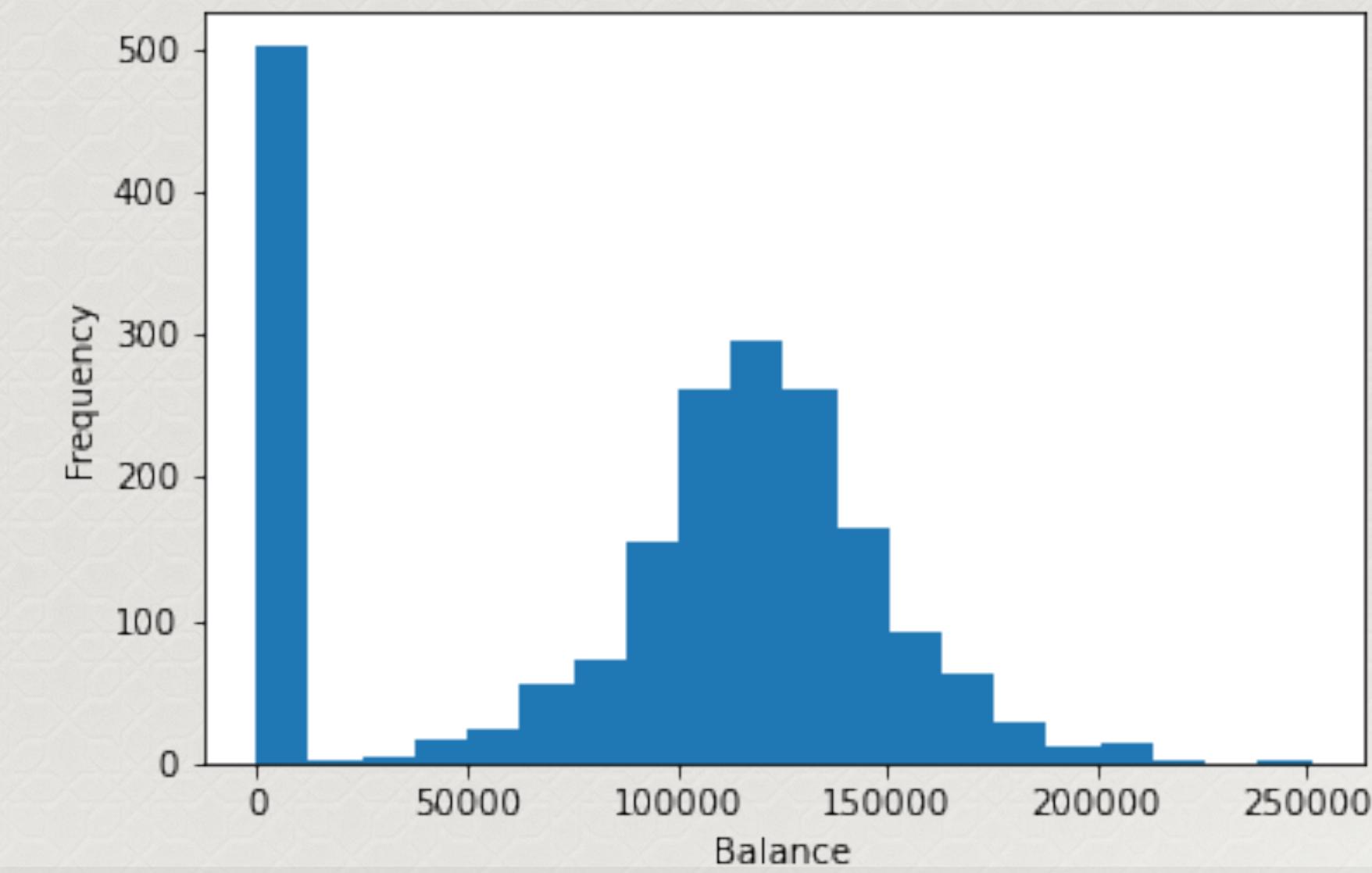
EstimatedSalary Variable Distribution



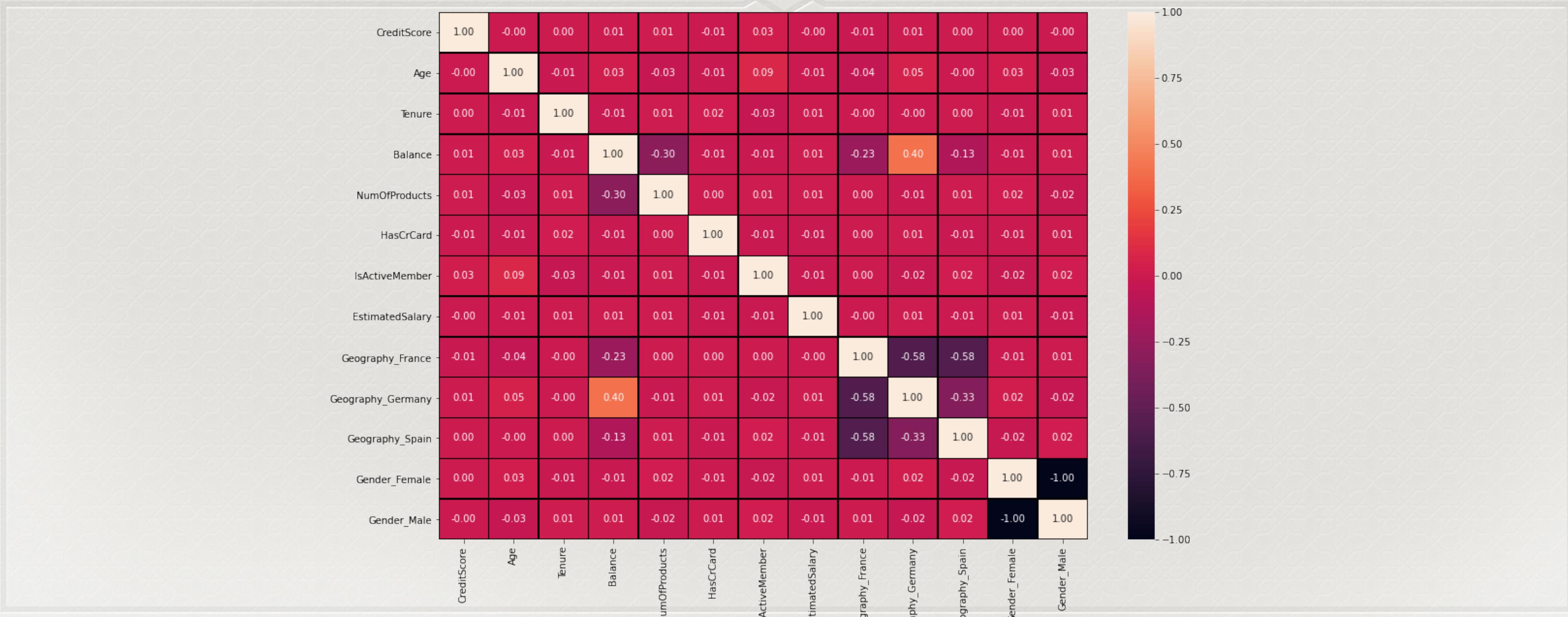
Age Variable Distribution



Balance Variable Distribution



Before FE Heatmap



After FE Heatmap

