Problem Set3
Visualization
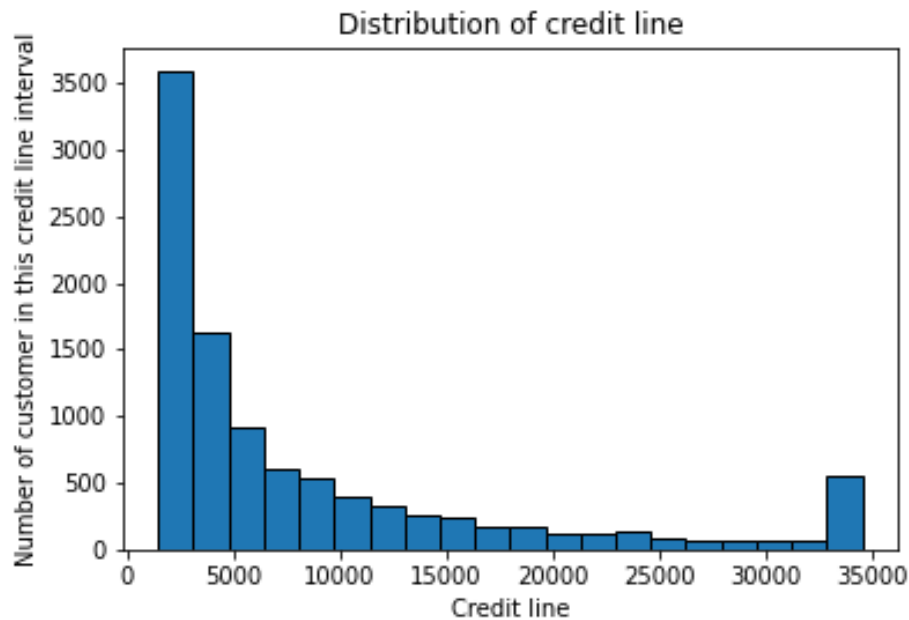
Xinya (Lucia) Yu

Data Used

The dataset is from website https://leaps.analyttica.com/home, containing information of credit card customers. This dataset includes 18 features, which include gender, age, education, credit line, etc.
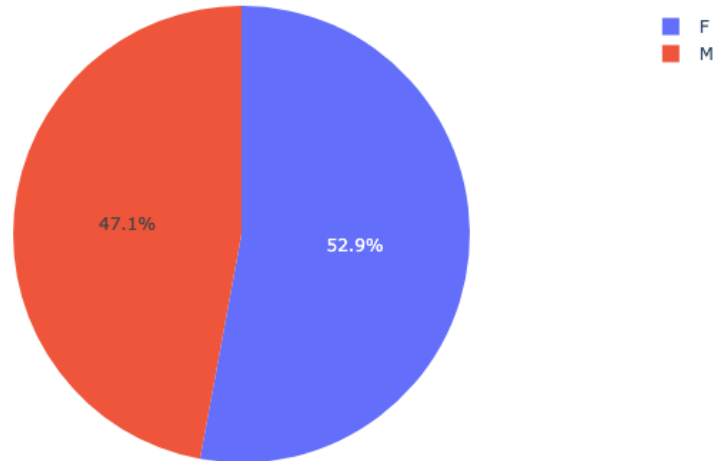
Figures

Figure 1

The number of customers whose credit line is less than 5000 is the largest. With the increase of credit line, the number of customers in the corresponding interval decreases, but the number of customers in the interval of 35,000 suddenly increases
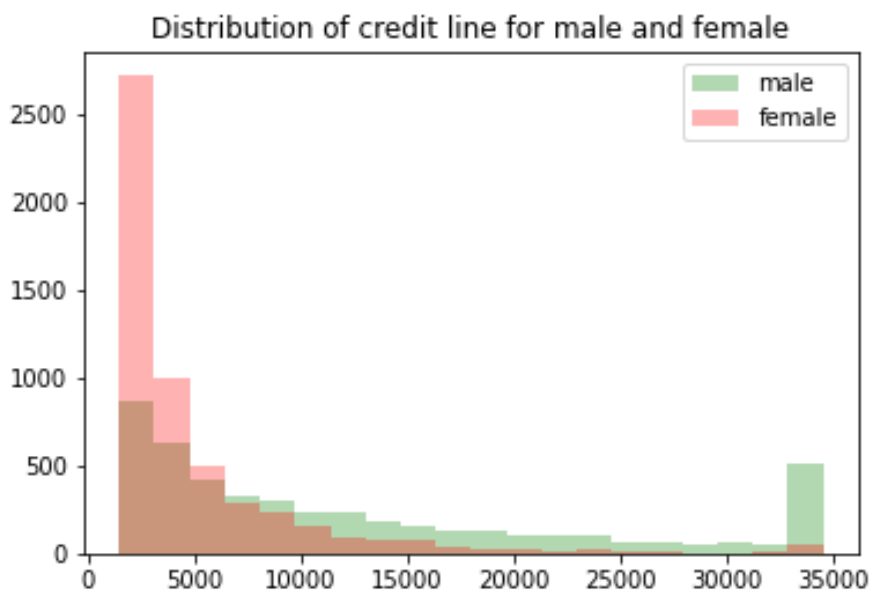
Figure 2

Distribution of gender

The proportion of female and male is nearly even. I wonder whether there is a difference between female customers' and male customers' credit line, so I plot a histogram as follow.

Figure 3

Distribution of credit line for male and female

It's obvious from the figure that female customers tend to have low credit line, but most of the customers who have high credit lines are male.