



The Business Question of Credit One

Use the data to predict the probability of default

Presented by Wanyun Ho

Business Questions

The number of customers who have defaulted on loans is increasing. Credit One could risk losing business.

Is there a better way to decide how much credit to allow the customers to use?

What are the key factors to decide one's probability of default? Can we get a better scoring method from the historical data?

Data Science Process

1. Collect the Data

- Pull out the data set of customers from the open sauce mySQL.
- Extract data into Python dataframe and Excel format.

2. Process the Data

- Examine data: Understand the column and data type.
- Clean the data: Drop the duplicate, missing values, errors.

3. Explore the Data

- Split and plot the data in different ways.
- Test the correlation of variables.

4. Perform In-Depth Analysis

- Create the predictive models
- Evaluate the models.
- Revise the models and go back to explore the data if needed.

5. Communicate Results

- Answer the business problems.
- Visualize the findings in simple way.
- Communicate the results to non-technical people.