Default Prediction and Business Insights - American Express

Consulting Analytics Case Study

2023

Agenda

1. Introduction
2. Data Exploration
3. Model Development
4. Interpretation

Identify and Predict Customers Credit Card Defaulters

Outline of Problem

- Amex is unable to make use of huge data of their customers to make business decisions to create a measurable economic benefit.
- Amex can create economic benefit by reducing credit card defaulters.
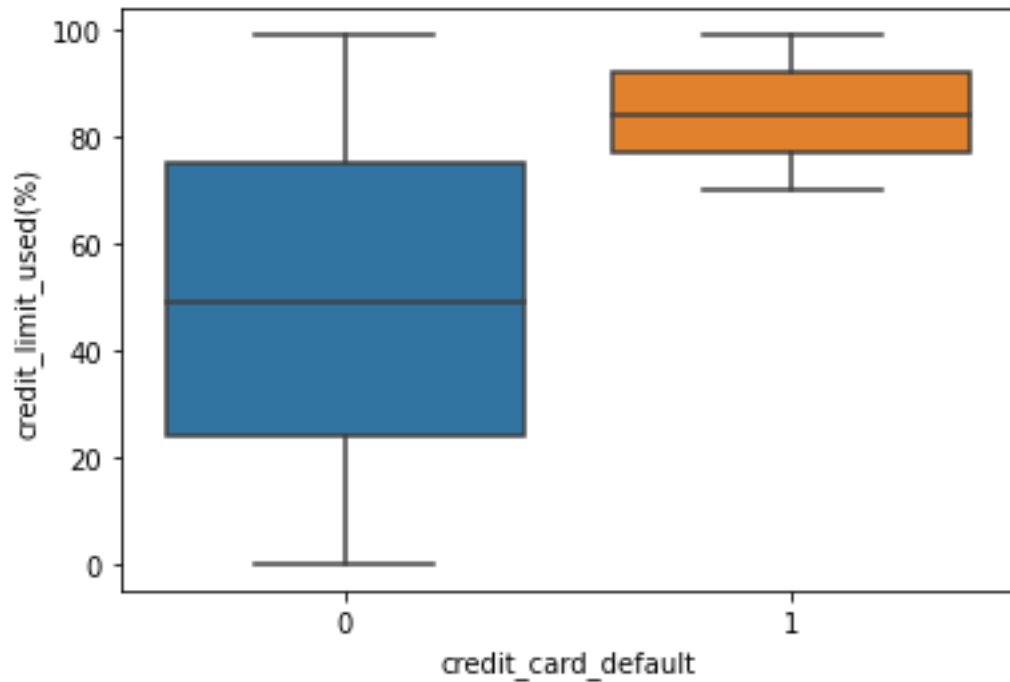
Data Exploration

Data Quality Assessment

|	Issues	Solution
Completeness	owns_cars, no_of_children, no_of_days_employed had missing values	Removed those rows
Validity	Occupation type has unknown value no_of_children, no_of_days_employed, total_family_members, migrant_worker had float data type	Didn't change Changed float to Int

Data Exploration

Credit_card_default Vs Credit_limit_used

Credit_card_default Vs Credit_score

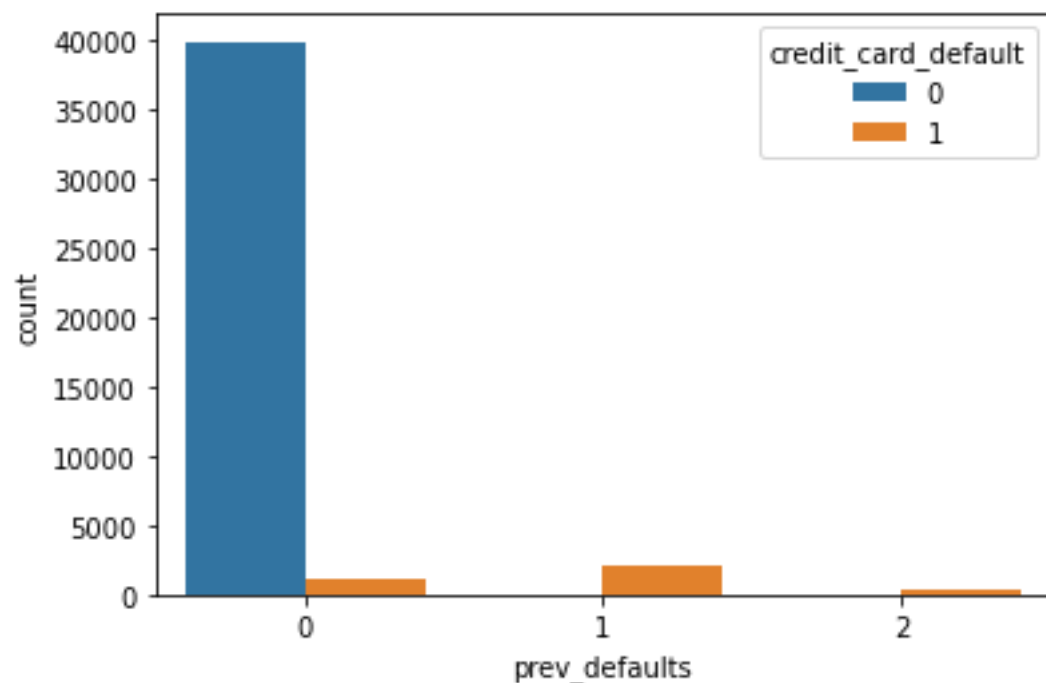Defaulters have used more credit limit than non-defaulters

Defaulters have used less credit score than non-defaulters
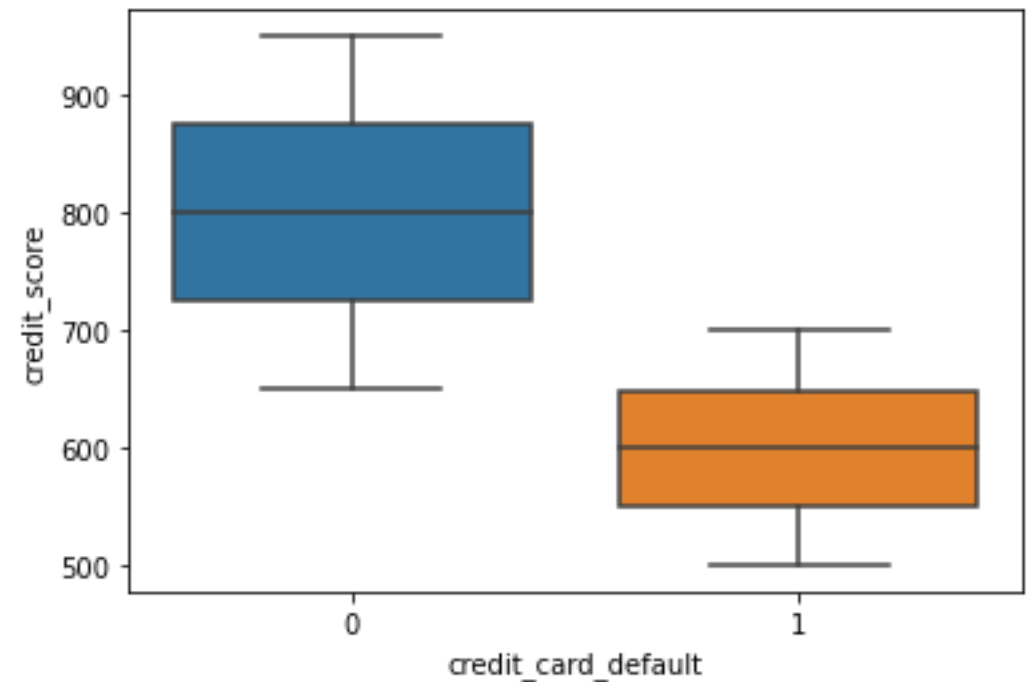
Data Exploration

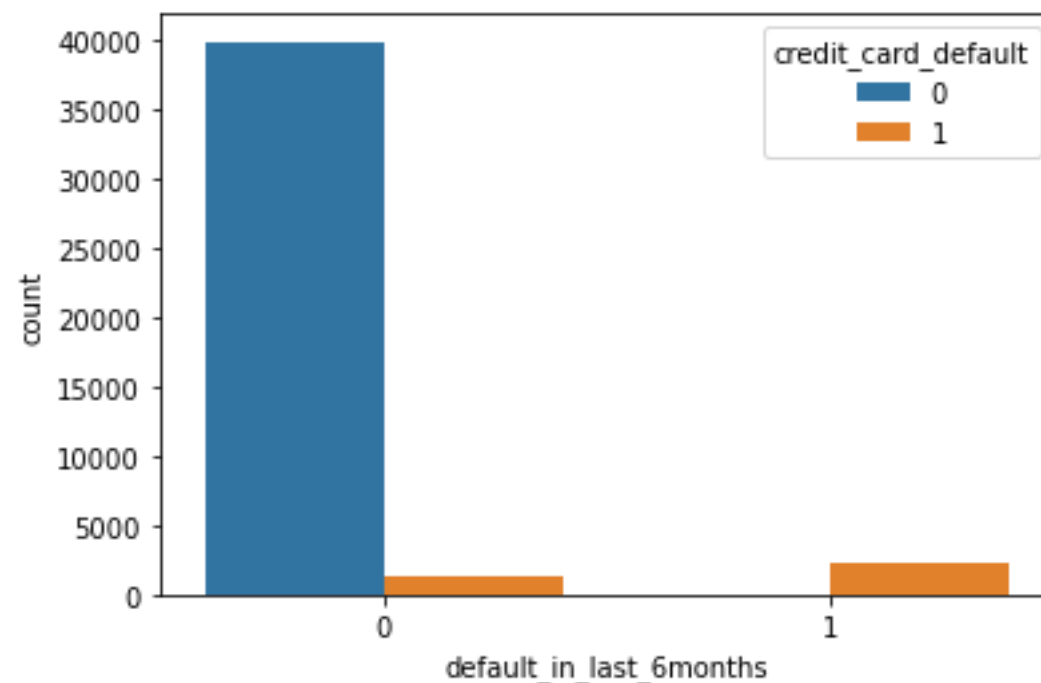Credit_card_default with previous defaults

People who have defaulted earlier tend to default in future

Credit_card_default with default in last six months

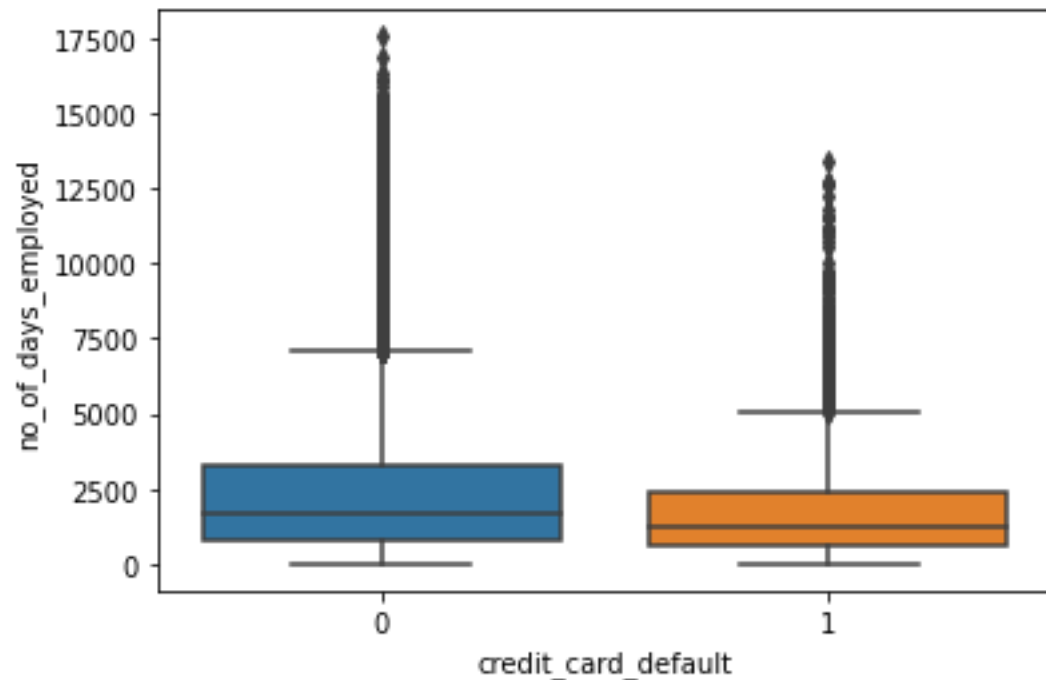People who have defaulted in within 6 months tend to default in future

Data Exploration

Credit_card_default Vs no_of_days_employed

From the data it can be inferenced that defaulters didn't had stable job

It is being observed through analysis Low-skill Laborers, Waiters/barmen staff,security staff, cooking staff,drivers, Laborers are at high risk while IT staff, secretaries, Accountants are at low risk.

Model Development

- Removed **outliers** from **number of working days** and **net yearly income**.
- Various statistical techniques like Pearson's,**, Kendall's,** etc. have been applied to find out the **correlation** with credit card default.
- The label of certain columns has been changed, like **gender, owns_car, owns_house,** etc., using the **Label encoder.**
- Since it is an imbalanced dataset, the **SMOTE** method has been applied.
- At last, data is trained and tested on the random **forest classifier.**
- As a result**, Accuracy of 98.496 %** has been achieved with **recall of 97.75%** and **precision of 99.24%**

Confusion Matrix
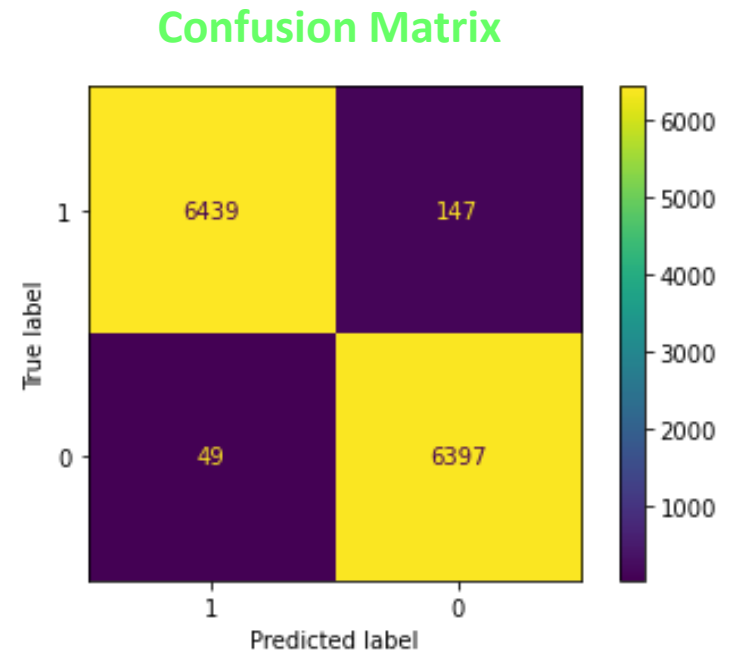
Interpretation

- It has been observed that defaulters have used 78 to 85 percent of their credit limit.
- And they have a credit score between 550 and 650.
- And Previously defaulted at least once specifically within 6 months
- Moreover, they didn't have stable jobs.

For future scope, it is advised to avoid unknown values in occupation type, which can give an idea about a specific occupation.

THANK YOU

Team name – Dataquest
Members – Tapan , Sachin, Ayush
College – G.B pant engineering college