Data-Driven Credit Risk Management

The Modern Loan Prospecting Intelligence

Imperial Galactic Bank (IGB)

The Imperial Galactic bank (IGB)

- Historically struggling with loan defaults.
- Due to recent hike in competition, they have to increase the volume of loans to survive.
- Furthermore, IGB's management has a mission to increase profits to support the great Emperor Palpatine's new Death Star venture project.

Mission Statement

- Deep dive into IGB's credit usage data.
- Learn the variables associated with defaults.
- Suggest a strategy to mitigate the credit risk.

Help IGB make data-driven decisions on who to prospect and what credit limit to provide.

This practice would also help IGB have a better understanding of their current and potential customers, which would inform their future strategy, including their planning of offering targeted credit products to their customers.

Process

Pre-Processing & EDA	Data Visualization & Modeling	Tool Deployment & Iteration

Pre-Processing & EDA

- Set up an environment
- Data connections
- Data cleaning
- Data wrangling
- Address data anomalies
- Define Target variable
- Checking assumptions

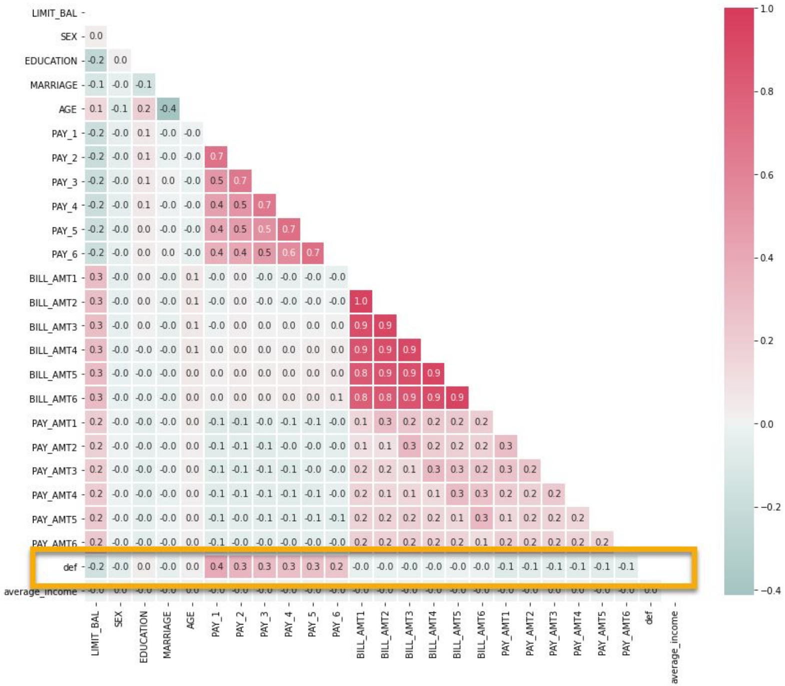
Data Visualization & Modeling

- Correlation Matrix
- Target Variable on
  - Age groups
  - Genders
  - Education levels
  - Marital Status
  - Payment Status
- Classification Model

Tool Deployment & Iteration

- Databricks
- Tableau Server
- Alteryx Server

Which variables are associated with default?

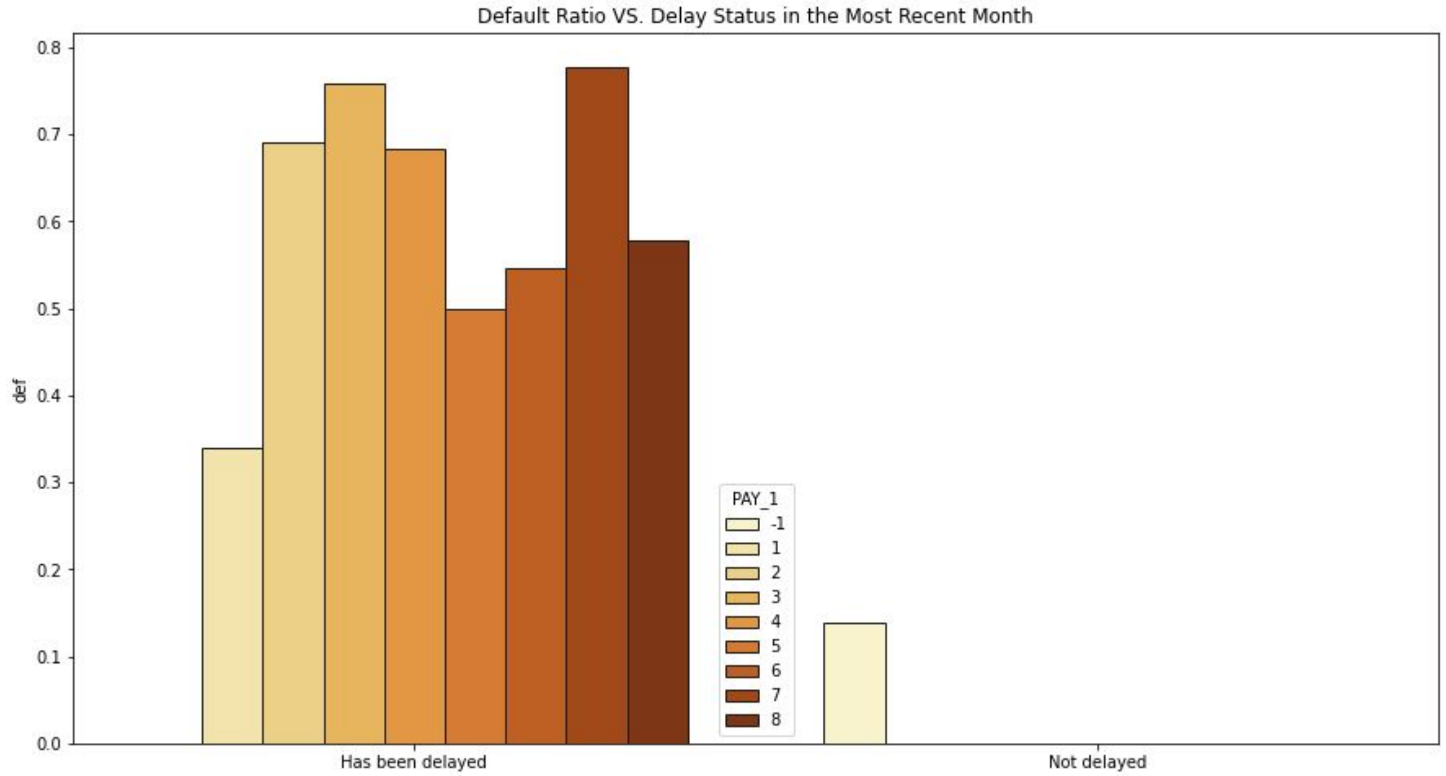


Negative correlation between the credit limit (LIMIT_BAL) and default, the target variable (def) - Borrowers with good credit history are given higher credit limit by banks. **The higher the credit limit for a borrower, the less likely the borrower will default.**

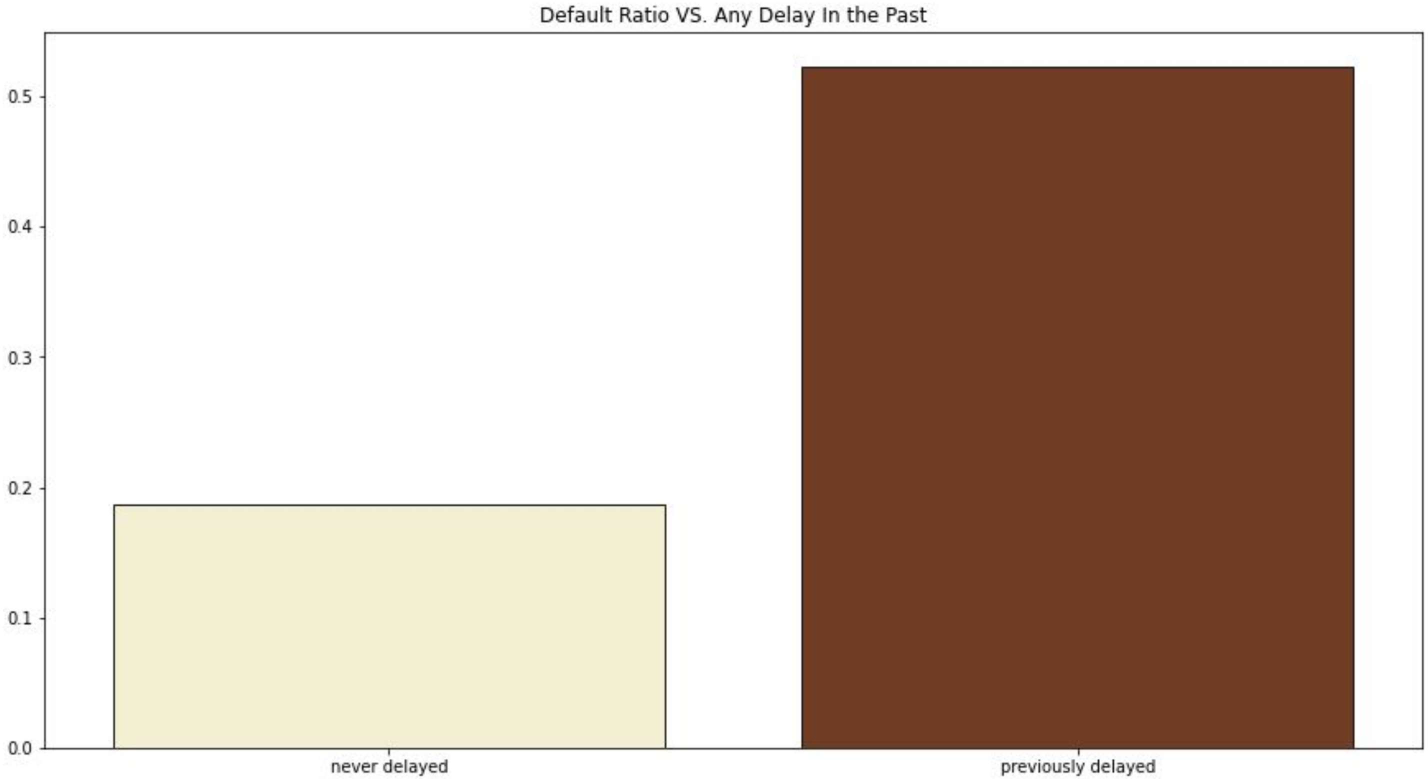
Positive correlation between delayed payment duration in previous months (PAY_X) and the target variable (def). This implies **the longer the payment delayed in previous months, the more likely the borrower will default in the following month.**

The positive correlation becomes **stronger as you move from old payment period to recent payment period**.

Payment Status in <u>Most Recent Month</u> Matters



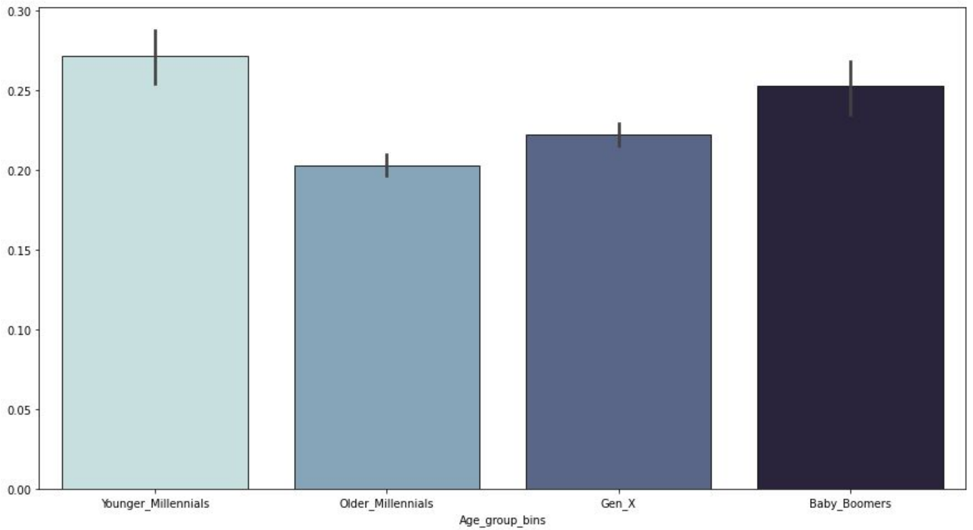
Delayed Only Once? Strong Indication of Default



Age Group vs Default Rate

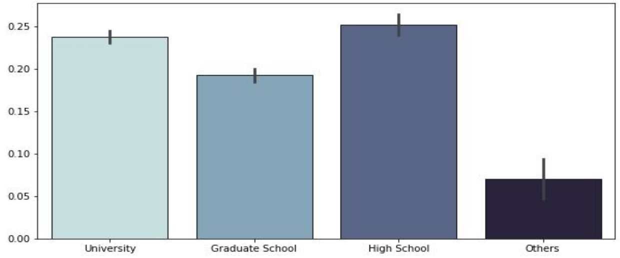
- There is a **U-Shape** in the default distribution across the age groups.

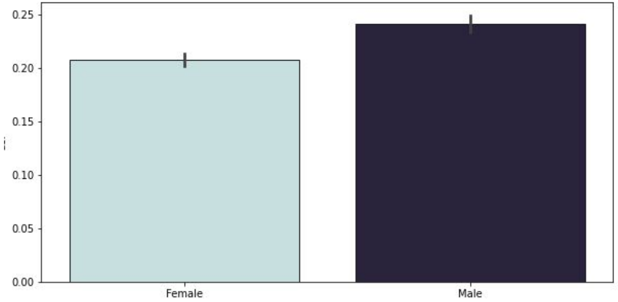
- The Older_Millennials group **(between age of 25 and 34) shows the lowest default rate**.

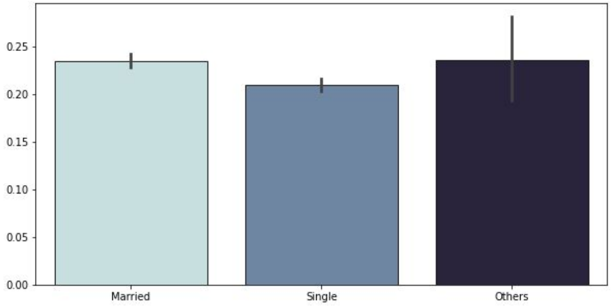


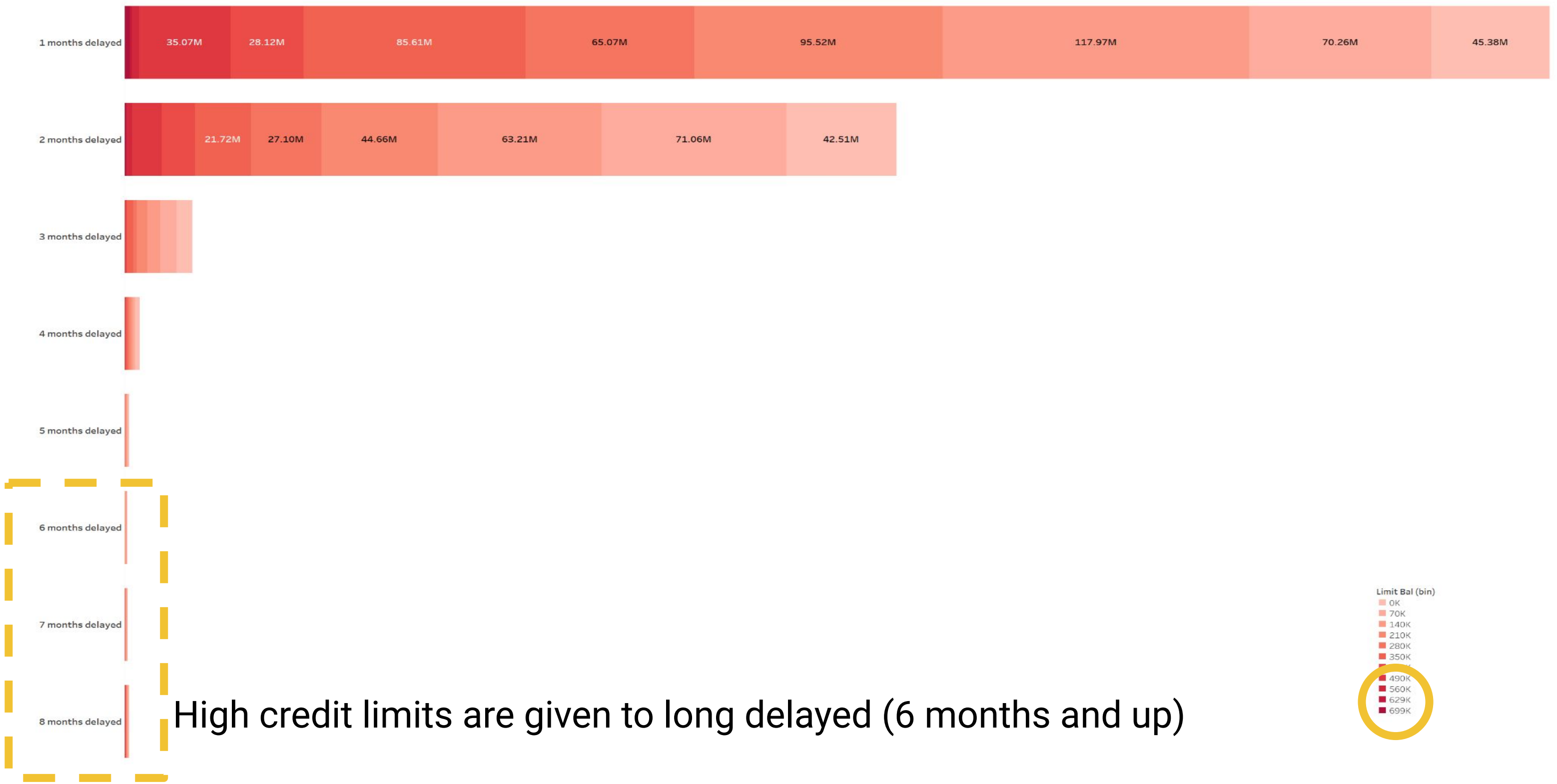
Other Variables
vs
Default Rate

- Male borrowers have higher default rate than Female borrowers.
- Married borrowers have higher default rate than single borrowers.
- Borrowers with a more advanced education background have a lower default rate.



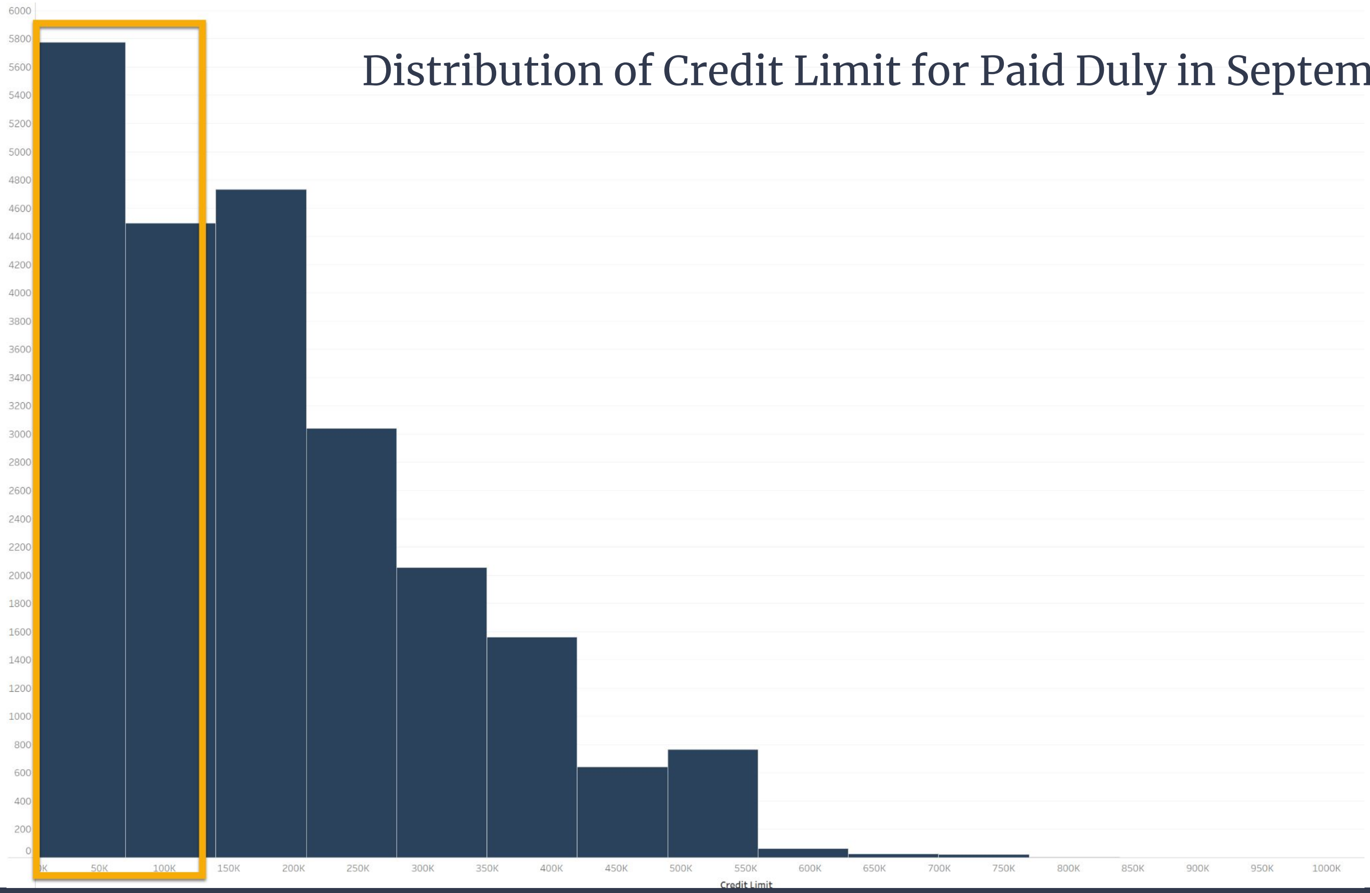






Delayed Loans in September By Credit Limit

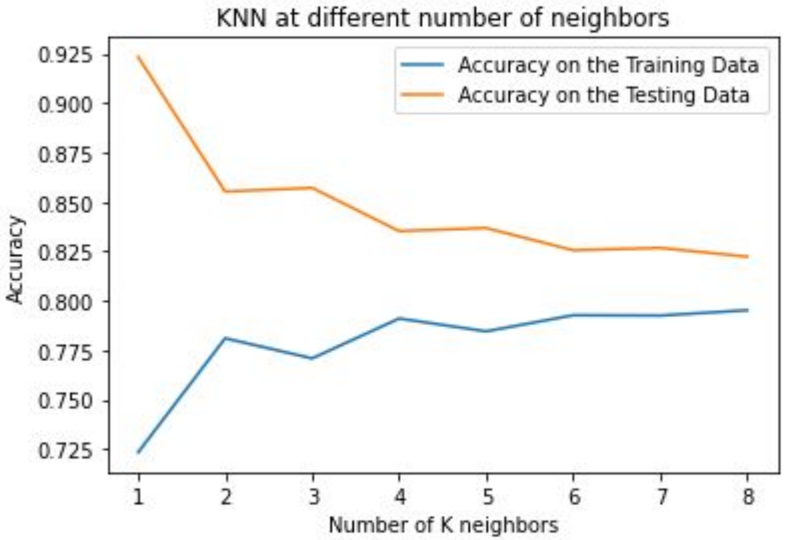
Distribution of Credit Limit for Paid Duly in September



Under-credited Borrowers with Good Payment Behavior

Predictive Classification Model

- KNN (k-nearest neighbors algorithm) model to predict the target default classification.
- Initial Accuracy Score of ~73% at K = 1
- Accuracy Score improved to ~80% after training the model with k=8



Summary

Improve data maintenance process for better analysis

- Data accuracy on billed amounts and paid amounts
- Miscategorized: 0 and -2 on PAY

Enhanced **credit risk management**

- Severe credit concentration was not shown
- Reallocate credit if concentrated on specific demographics (i.e. very young or very old)
- Reduce credit limit on consistently delayed borrowers

Loan prospecting target

- Low risk demographics (i.e. older millennials)
- Under-credited borrowers with good behavior
- Rely on the Default Classification Model