genpact

Transformation Happens Here

Fall 2020 Case Competition

Team All Stars
Arkav Banerjee
Nishi Intwala
Kritika Panjiar
Yash Shukla

01

THE PROBLEM

Propensity to pay collections

02

THE MODEL

Data processing and machine learning

03

THE OBSERVATIONS

Key findings and issues

04

THE RECOMMENDATIONS

Business insights and policy suggestions

Problem

Business Overview

Image is integral to business success

Maintaining **steady cash flows** and a **consistent accounts payable system** is vital

This is especially important for **lending businesses**

Propensity to Pay

The **likelihood** of a debtor to settle their account in a timely manner

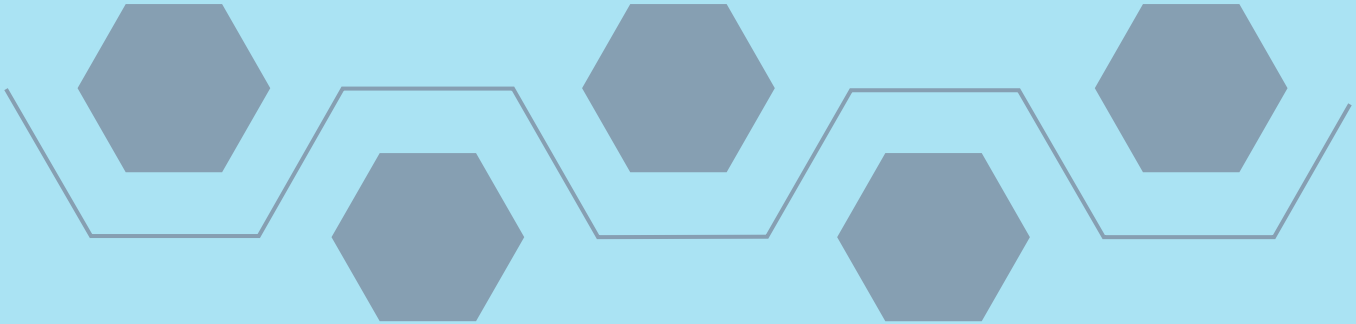
Financial institutions commit **considerable resources** into this

An accurate method would be **invaluable** to firms

Objectives

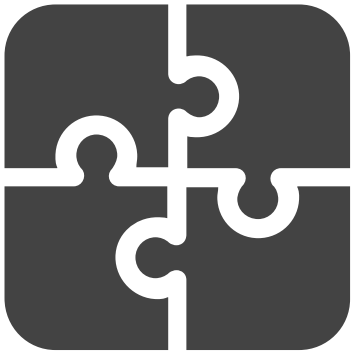
Develop and test a machine learning algorithm to **predict** defaulters

Provide **insights** and **recommendations** based on analysis



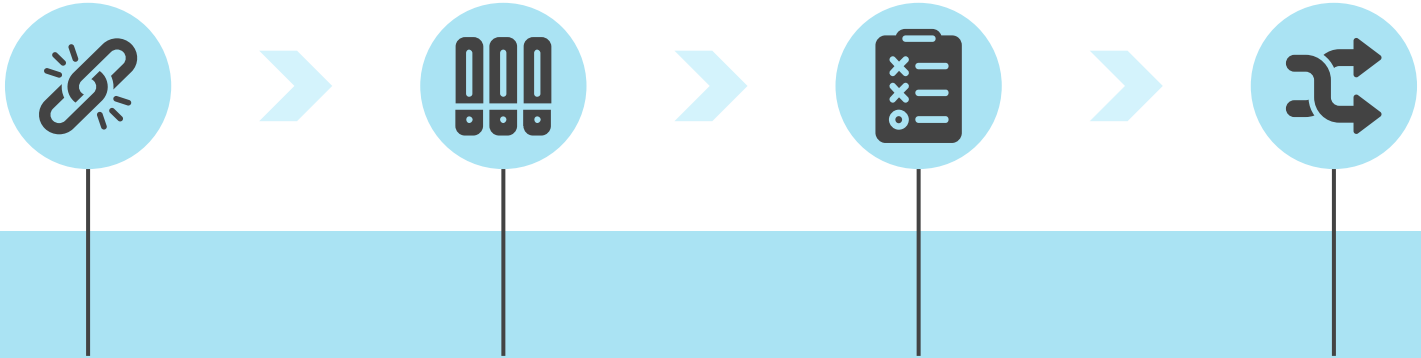
Project Walkthrough

- Data import and exploratory analysis
- Data processing and feature engineering
- Modeling and evaluation
- Result analysis and recommendations



Data and Modeling

Data Processing



STEP 1

Outlier Removal

STEP 2

Duplicate Removal

STEP 3

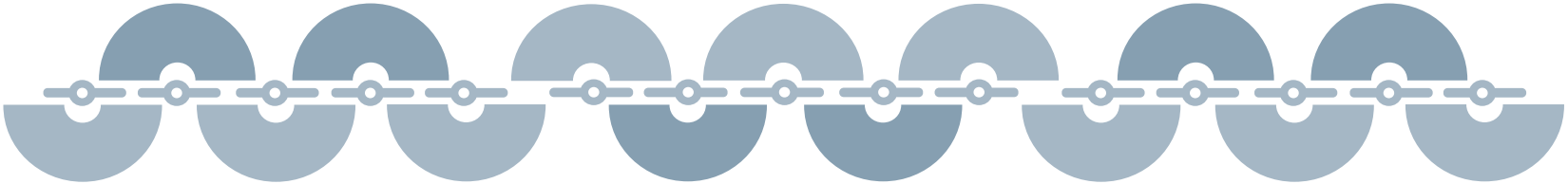
Cleaning Datetime Columns

STEP 4

Changing Data Types

Feature Engineering

- Days of Week
- Weekend Indicator
- Public Holidays



- Dropping irrelevant columns
- Creating Target Variable
- Extracting Year, Month , Day

Data Modelling

- Created Base model
- Logistic Regression, SVM, Decision Trees, Bagging and Boosting
- Final Model: **Random Forest**
- Performed Hyperparameter Tuning
- Generated classification report

97.6%

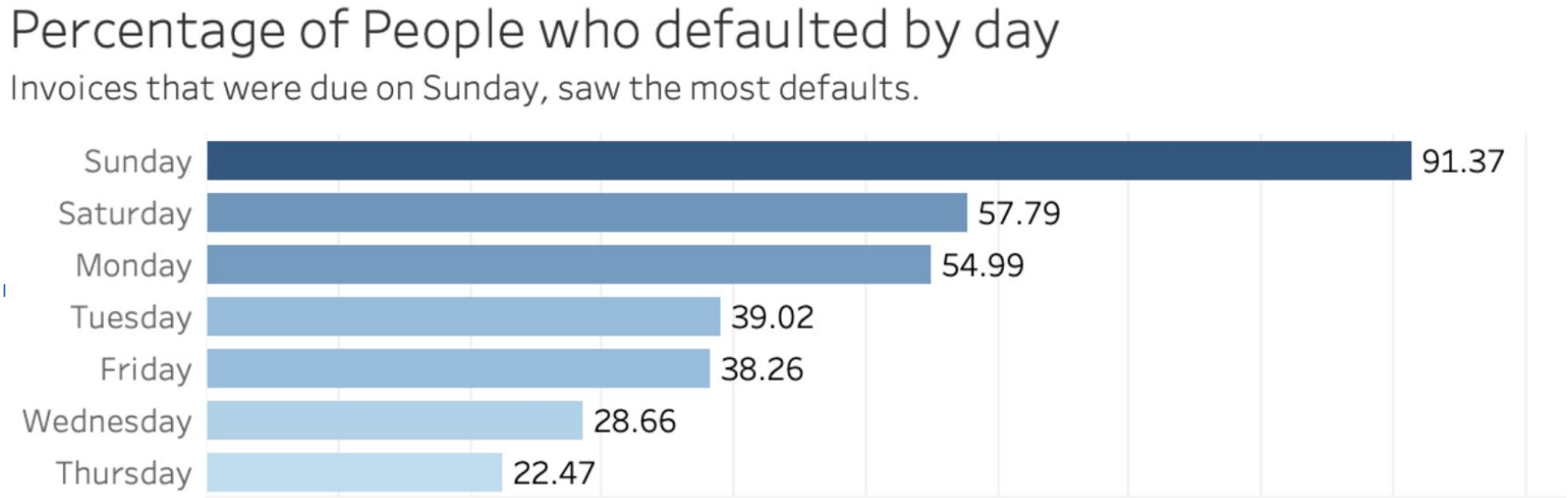
accuracy

Classification Report

	Precision	Recall	F1-score
Did not Default	0.98	0.98	0.98
Default (0-30 days)	0.98	0.98	0.98
Default (30-60 days)	0.99	0.93	0.96
Default (60+ days)	0.95	0.81	0.88

Analysis

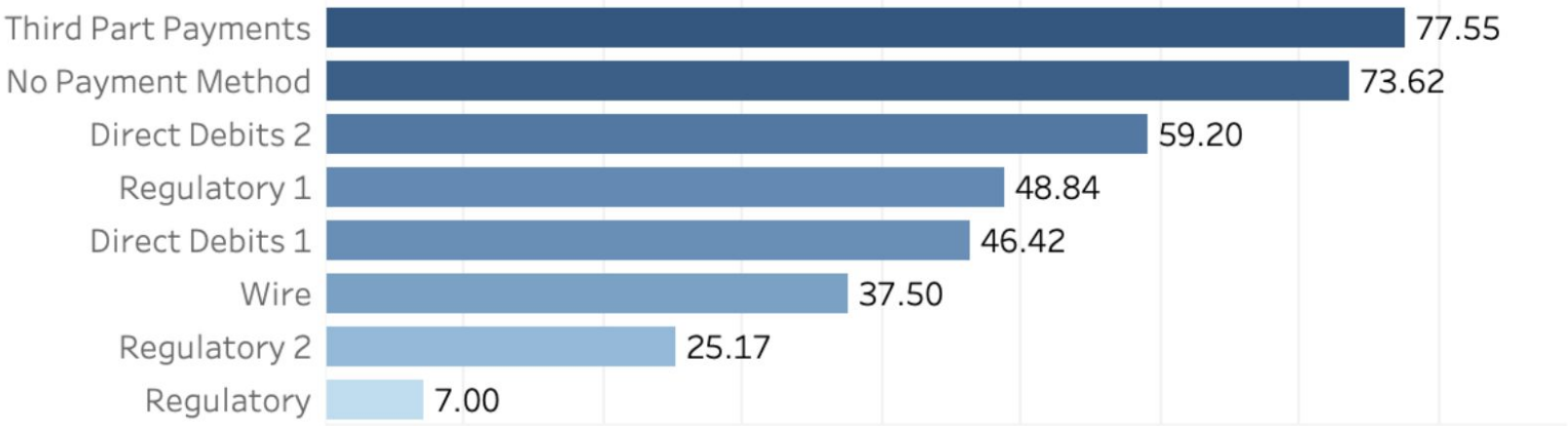
Observation



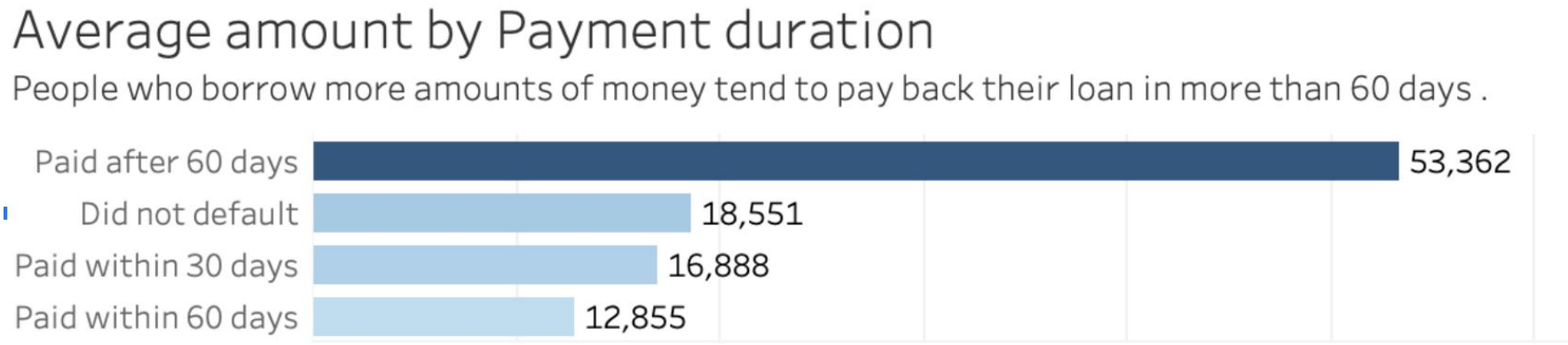
Observation

Percentage defaults by Payment methods

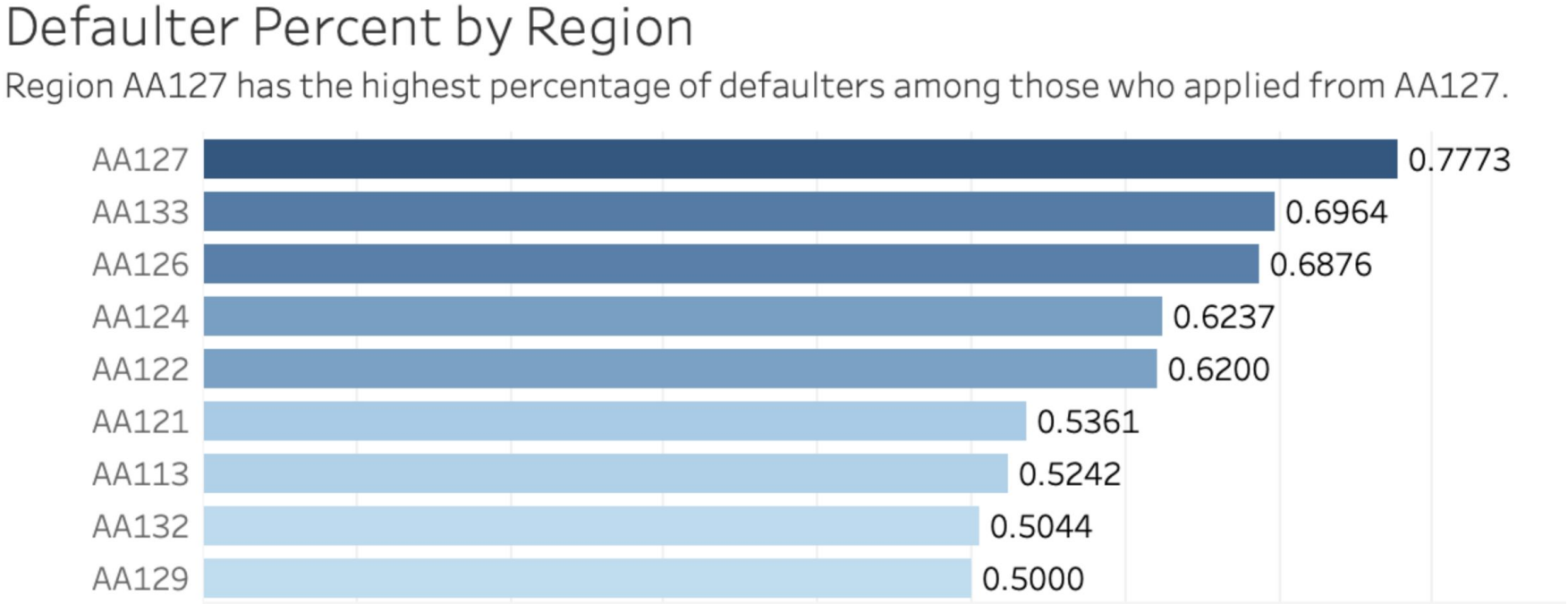
The maximum defaulters were from Third Party Payments or the payment methods were not recorded.



Observation

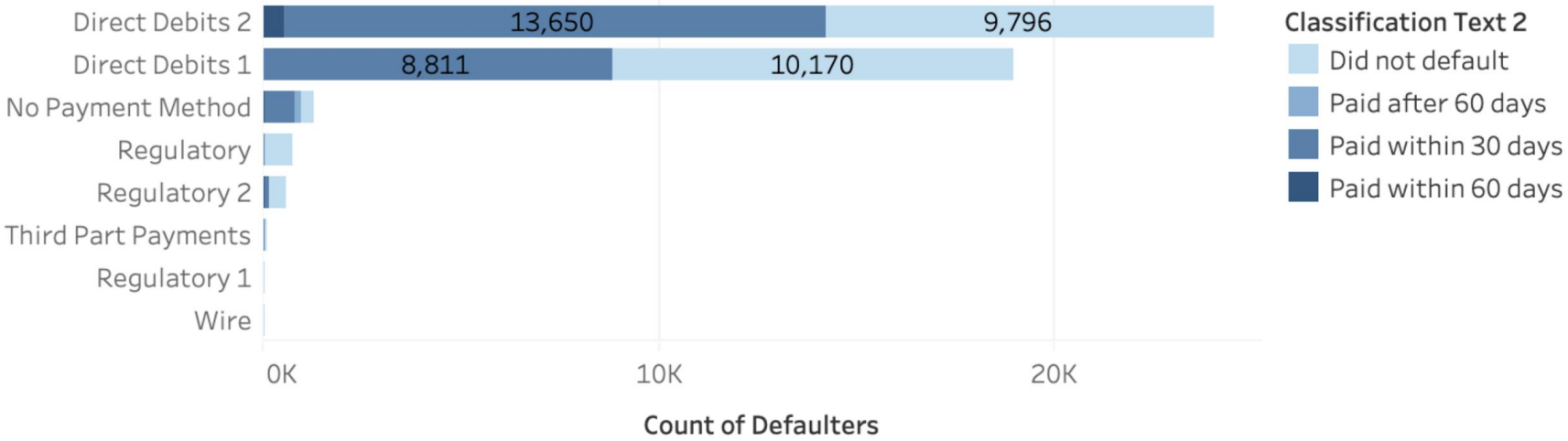


Observation



Observation

Customer classification based on duration of default status and payment method



Recommendations

The Sunday Problem

Sunday due dates should be **moved exclusively** to weekdays

Administrative costs are likely incurred for even a **single day** of default

Payment Type - 2 Culprits

Third party payments and **no payment methods** are most problematic

Customers wanting to choose these options should be **vetted** more thoroughly

Alternative payment options should be offered and promoted

Higher Gain, Higher Loss

60+ days default instances have very high **relative average loan amounts**

Lost expected revenues of these magnitudes can be **crippling**

Besides being more strict with high amount loans, we suggest adding a **clause** to the loan contract that incurs **heavy fines** for defaulting 60+ days

It’s All About the Location

Region is the **highest importance variable** in our prediction model

Further investigation is necessary to distinguish if this is an **in-house** problem or an **endemic** one

Given the **arbitrary** nature of data, more analysis is required to deliver precise policy suggestions

<30-Day Settlements

An exceedingly **large majority** of defaults were settled within **30 days overdue**

Clients predicted to settle debts under 30 days overdue should have their due date **shifted forward** by **30 days**

Future scopes include **finer binning** of day categories and predicting the **exact overdue periods**

To Conclude

Defaulters will remain an issue, but data and analysis can help us get ahead of them