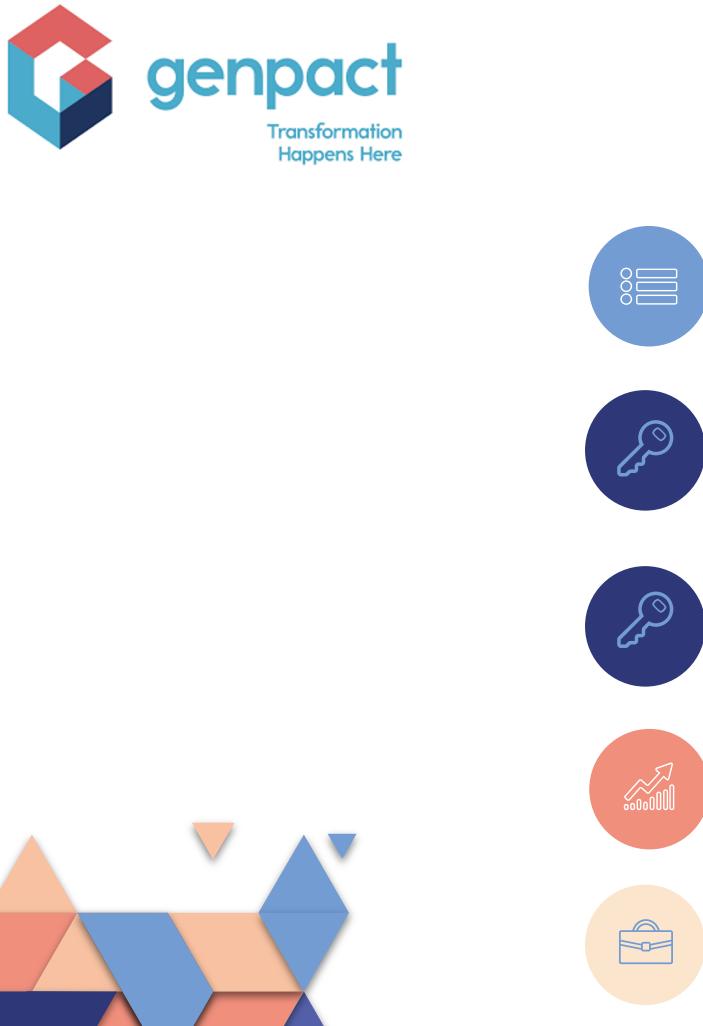


Propensity to Pay Collection Model Utilizing Machine Learning Techniques

Terp F5

Raymond Luo
Ziying Yan
Sophia Hao
Delia Hu
Michelle He



Agenda

Overview & Business Problem

Data Wrangling

Exploratory Data Analysis

Modeling & Results

Business Recommendation

Problem Overview



A strong **correlation** between the past behavior and propensity to pay



Data Overview

Genpact is a global professional services firm delivering digital transformation by putting digital and data to work to create competitive advantage.

Original Data

Transaction Records
45,841 rows; 28 columns

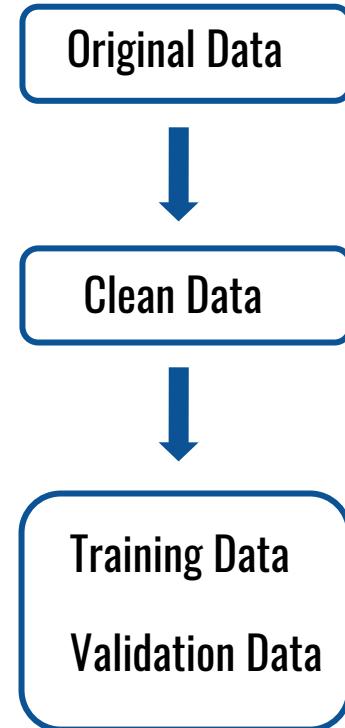


	Cust_Num	Customer.Group	Payment.Method.description	DocumentNo	Doc.Date	Net.due.dt	Pstng.Date	Amount	Clearing.Date	Crlng.doc.	ZIPCODE	Re
1	5039221123	Wholesalers	Direct Debits 2	91225033129	14-Jan-15	28-Feb-15	14-Jan-15	23,672.43	02-Mar-15	80992133504	AX0012	AA
2	5039221123	Wholesalers	Direct Debits 2	91225033130	14-Jan-15	28-Feb-15	14-Jan-15	3,689.47	02-Mar-15	80992133504	AX0012	AA
3	5039221123	Wholesalers	Direct Debits 2	91225033131	15-Jan-15	28-Feb-15	15-Jan-15	366.84	02-Mar-15	80992133504	AX0012	AA
4	5039221116	Wholesalers	Direct Debits 2	91225033132	23-Feb-15	28-Feb-15	23-Feb-15	259.84	02-Mar-15	80992133507	AX0013	AA
5	5039221094	Wholesalers	Direct Debits 2	91225033133	25-Feb-15	02-Mar-15	25-Feb-15	176.21	02-Mar-15	80992133508	AX0014	AA
6	5039221094	Wholesalers	Direct Debits 2	91225033134	26-Feb-15	03-Mar-15	26-Feb-15	352.41	02-Mar-15	80992133508	AX0014	AA
7	5039221094	Wholesalers	Direct Debits 2	91225033135	24-Feb-15	01-Mar-15	24-Feb-15	2,290.65	02-Mar-15	80992133508	AX0014	AA
8	5039221094	Wholesalers	Direct Debits 2	91225033136	25-Feb-15	02-Mar-15	25-Feb-15	526.82	02-Mar-15	80992133508	AX0014	AA
9	5039221094	Wholesalers	Direct Debits 2	91225033137	24-Feb-15	01-Mar-15	24-Feb-15	881.02	02-Mar-15	80992133508	AX0014	AA
10	5039221094	Wholesalers	Direct Debits 2	91225033138	24-Feb-15	01-Mar-15	24-Feb-15	176.21	02-Mar-15	80992133508	AX0014	AA
11	5039221094	Wholesalers	Direct Debits 2	91225033139	25-Feb-15	02-Mar-15	25-Feb-15	1,057.22	02-Mar-15	80992133508	AX0014	AA
12	5039221094	Wholesalers	Direct Debits 2	91225033140	25-Feb-15	02-Mar-15	25-Feb-15	4,052.69	02-Mar-15	80992133508	AX0014	AA
13	5039221115	Wholesalers	Direct Debits 2	91225033141	27-Jan-15	28-Feb-15	27-Jan-15	2,407.40	02-Mar-15	80992133516	AX0015	AA
14	5039221115	Wholesalers	Direct Debits 2	91225033142	19-Jan-15	28-Feb-15	19-Jan-15	536.95	02-Mar-15	80992133516	AX0015	AA
15	5039221115	Wholesalers	Direct Debits 2	91225033143	15-Jan-15	28-Feb-15	15-Jan-15	1,171.46	02-Mar-15	80992133516	AX0015	AA
16	5039221115	Wholesalers	Direct Debits 2	91225033144	19-Jan-15	28-Feb-15	19-Jan-15	1,217.52	02-Mar-15	80992133516	AX0015	AA
17	5039221115	Wholesalers	Direct Debits 2	91225033145	14-Jan-15	28-Feb-15	14-Jan-15	263.53	02-Mar-15	80992133516	AX0015	AA
18	5039221115	Wholesalers	Direct Debits 2	91225033146	14-Jan-15	28-Feb-15	14-Jan-15	4,804.11	02-Mar-15	80992133516	AX0015	AA

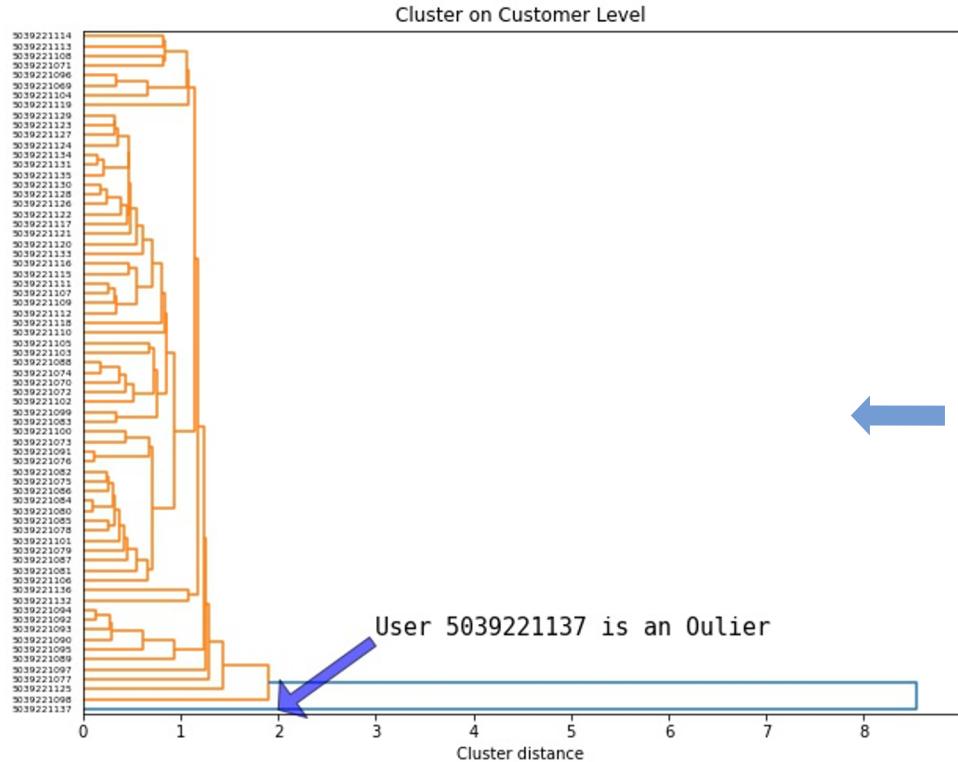
Data Wrangling

Cleaning

- Delete unnecessary columns
- Group categories by frequency and correlation
- Cluster customers and find out outlier.



Data Wrangling



Cleaning

- Delete unnecessary columns
- Group categories by frequency and correlation
- Cluster customers and find out outlier.

Data Wrangling

Cleaning

- Delete unnecessary columns
- Group categories by frequency and correlation
- Cluster customers and find out outlier.

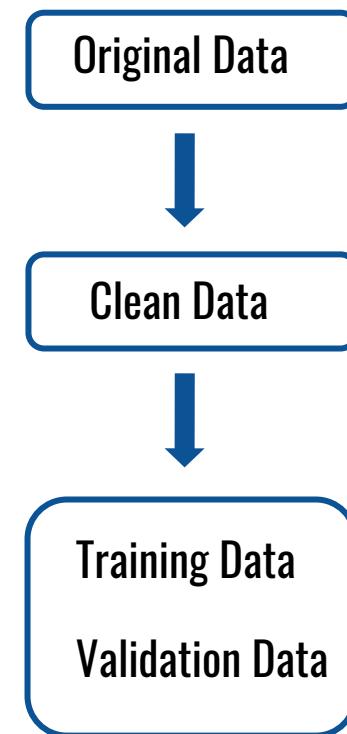
Generate new variables

- Weekday of due date and document date
- Season of due date
- Dynamic customer-based frequency and mean of delay days and amount

Dummy Coding

14 independent variables

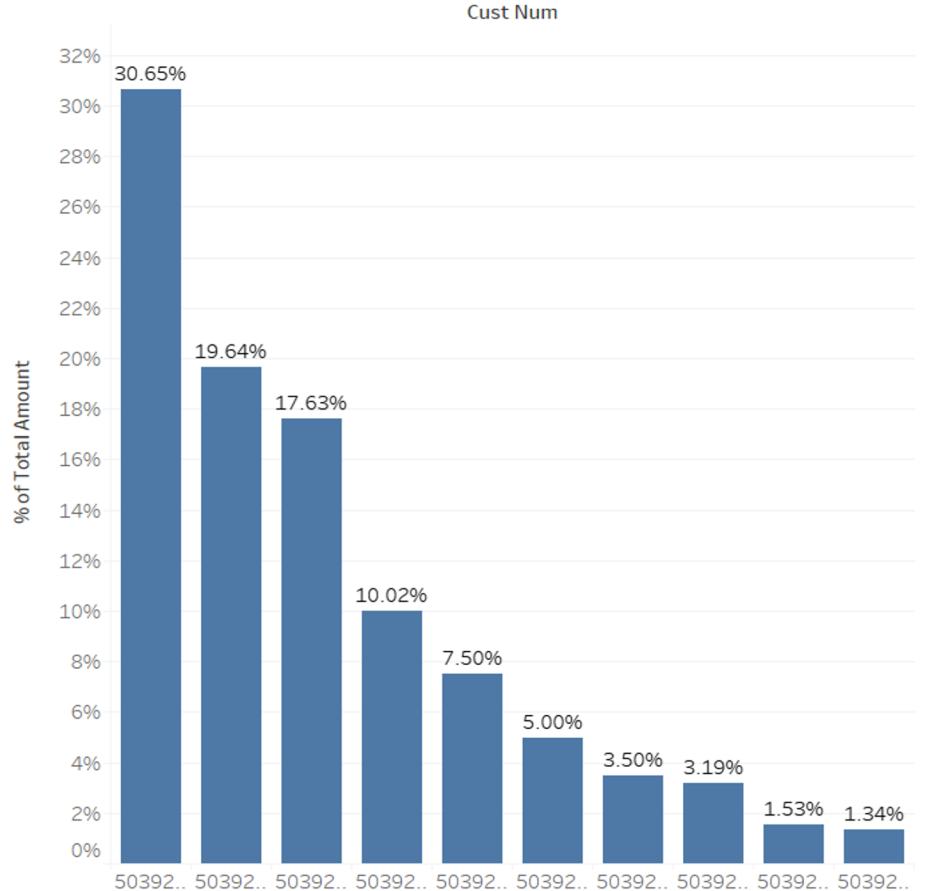
(35 variables after dummy coding)



Exploratory Data Analysis

Customers' contribution to total amount

- Top 4 customers contributed 77.94% of the total amount.

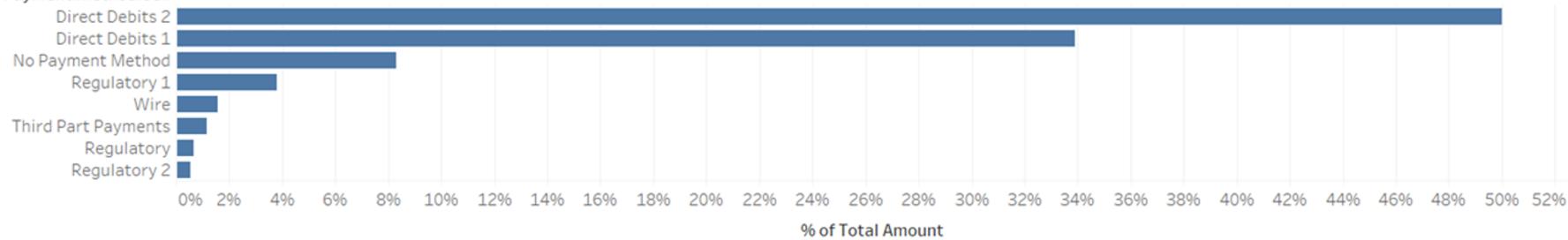


Payment methods' impact on amount

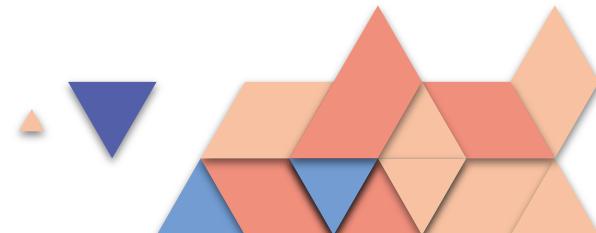
- 84% of the payments were from different forms of Direct Debits.
- Direct Debit

<distribution of amount---payment>

Payment.Method.de...

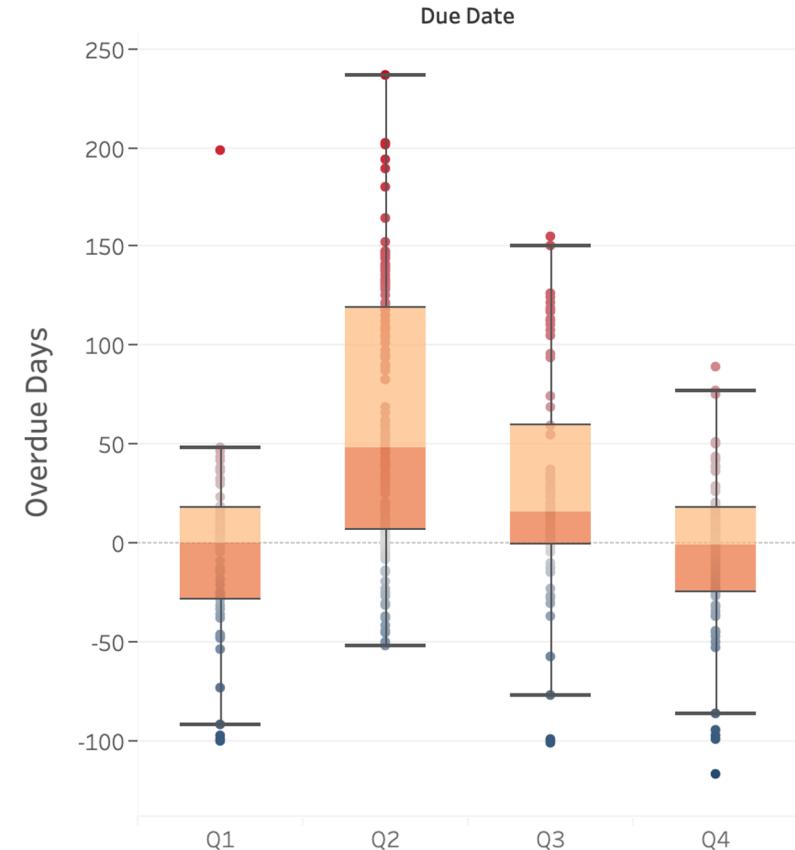


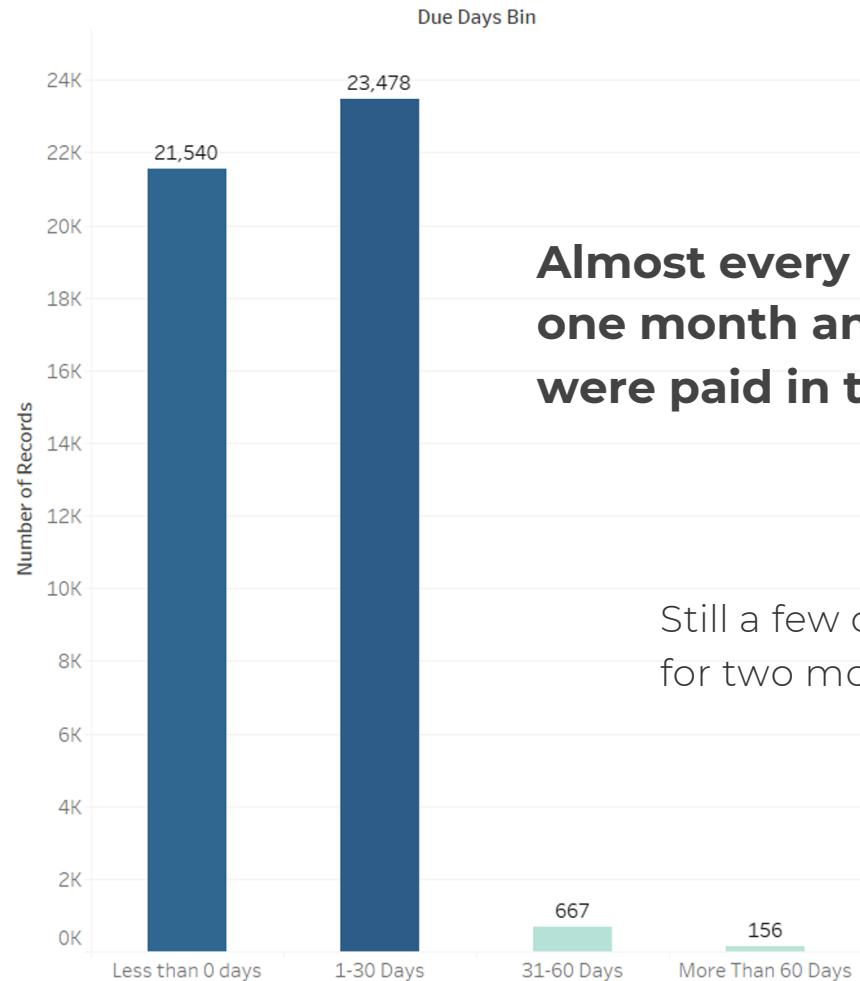
% of Total Amount for each Payment.Method.description.



Overdue delay days vary based on Due date

- Customers whose contract due date is in Q2 were less likely to pay their bills on time .



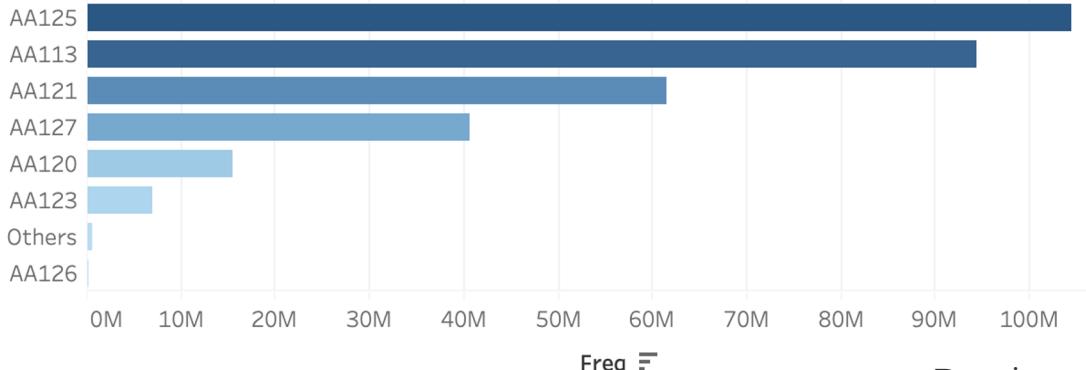


Almost every contract was fulfilled in one month and nearly half of contracts were paid in time!

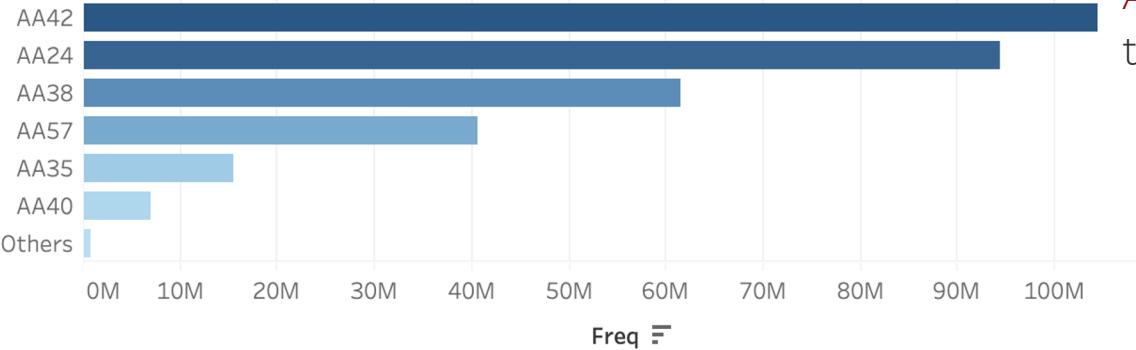
Still a few contracts were paid over one month even for two months.

Frequency of City and Region

Region F



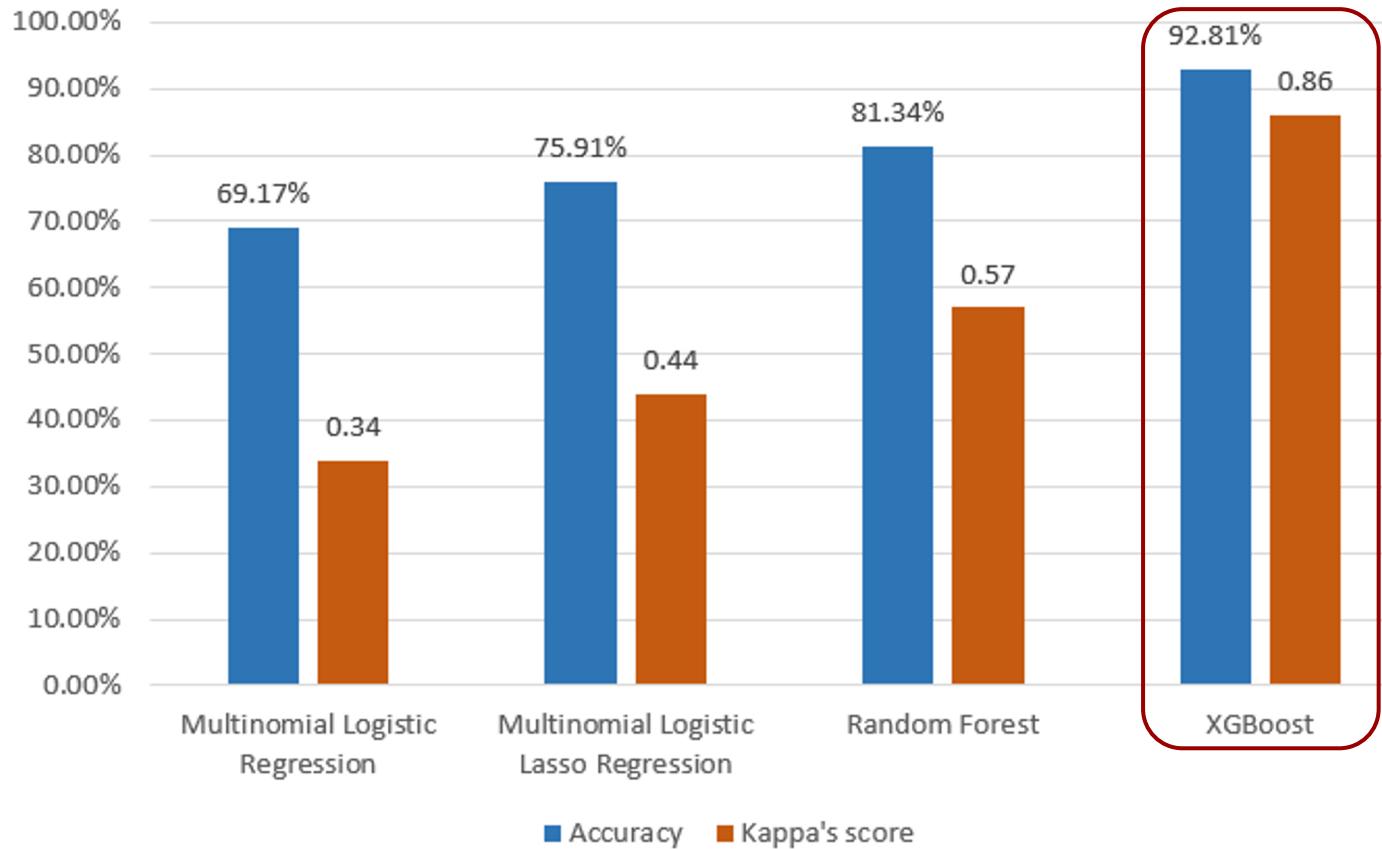
City F



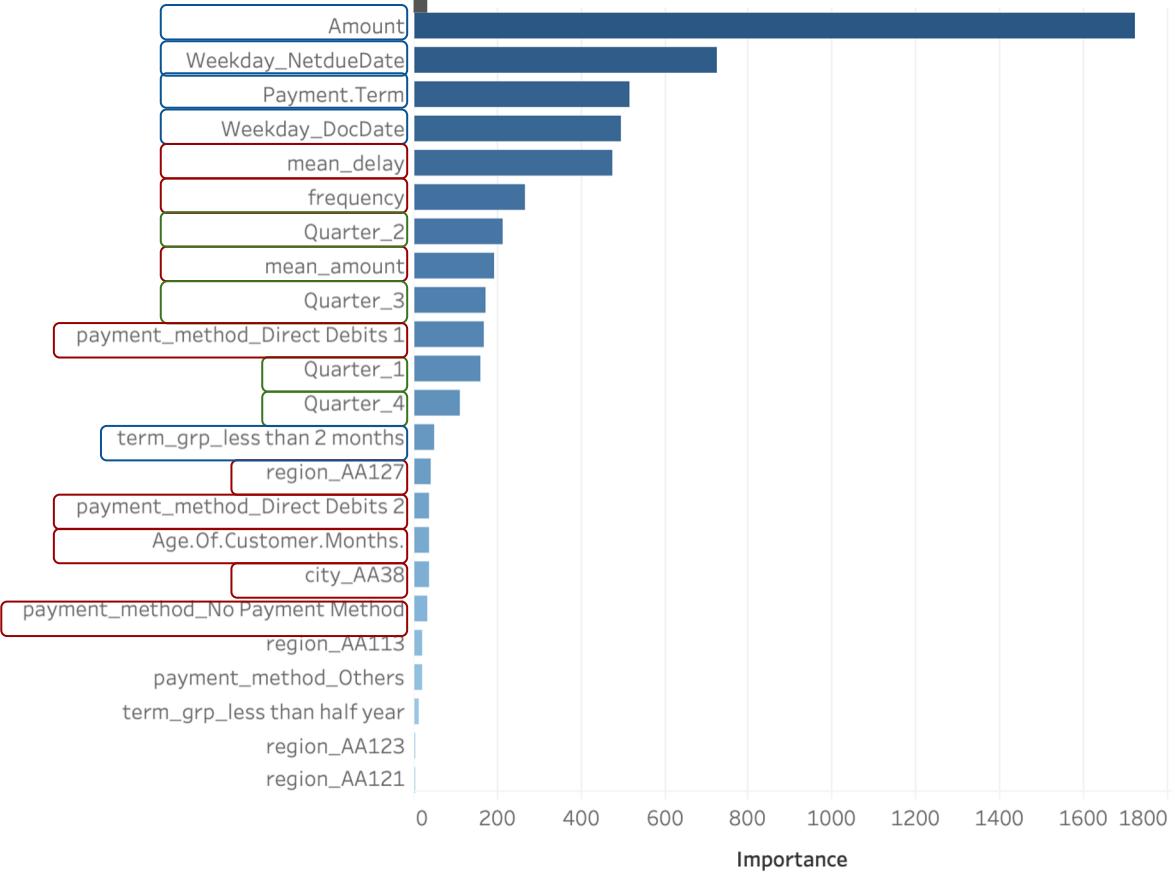
Region AA125 and City AA42 has highest transaction frequency.

Modelling & Insights

Model Performance



Feature Importance



Managerial Implication #1

Choose a nice day to make contract and also set the due date properly.

- Always **sign contract** and **set the due date** on **weekdays**. People can hardly remember to pay for the service on weekends.
- Cyclic variation has impact on propensity to pay, especially pay more attention to **Quarter 2**.



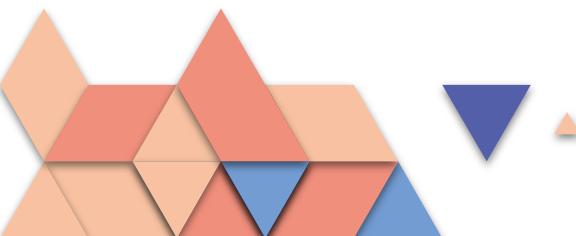
Managerial Implication #2

Focus on customer's behavior to predict propensity to pay, especially on his or her over due delay days, contract amount and demographics.

- Take efforts to encourage customers to **pay in advance**, thus decreasing average delay days.
- **Region AA127 and City AA38** are two locations that are important to delays days predictions.
- **Set aside a small team** focusing on collecting more information about customers to study their paying pattern



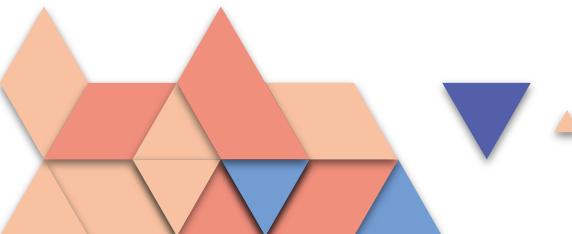
CUSTOMIZED SERVICES



Managerial Implication #3

Make sure decide on a good contract! **Payment term** and **payment methods** is critical to whether the customer will pay on time!

- **Remind customers with long term contract of paying** since they are more likely to pay overdue.
- **Direct debit** is one prior payment method for customer to use.





The Next Step

Our **Machine Learning classifier** is an **effective predictor**, but further statistical modelling needs to be done to provide inference of the **relationship between variables**.

Areas of further investigation:

- The different weekdays effects on paying intention
- The payment methods effect on decision making
- The regional research that focus on specific locations
- The optimal payment term to influence paying
- The customer profile that shows personal pattern

