Verizon Case

UC Consulting Team

Danae Vassiliadis

Forough Mofidi

Ankit Gubiligari

Angy Wu Feng

David Zhu

AGENDA

Meet the Team:

Delivery Lead:
Danae Vassiliadis (Data Strategy Sr. Manager)

Project Lead:
Forough Mofidi (Data Strategy Manager)

Data Science Consultant:
Angy Wu Feng

Data Analyst:
David Zhu

Modeling SME:
Ankit Gubiligari

Business Problem

Current State

- Verizon relies on contract-based cell phone sales as a core business component.
- Identifying customers at risk of payment default is a critical concern.
- With continued growth and customer acquisition, predicting default risks becomes vital.
- Proactive risk management supports sustained business expansion and financial stability.

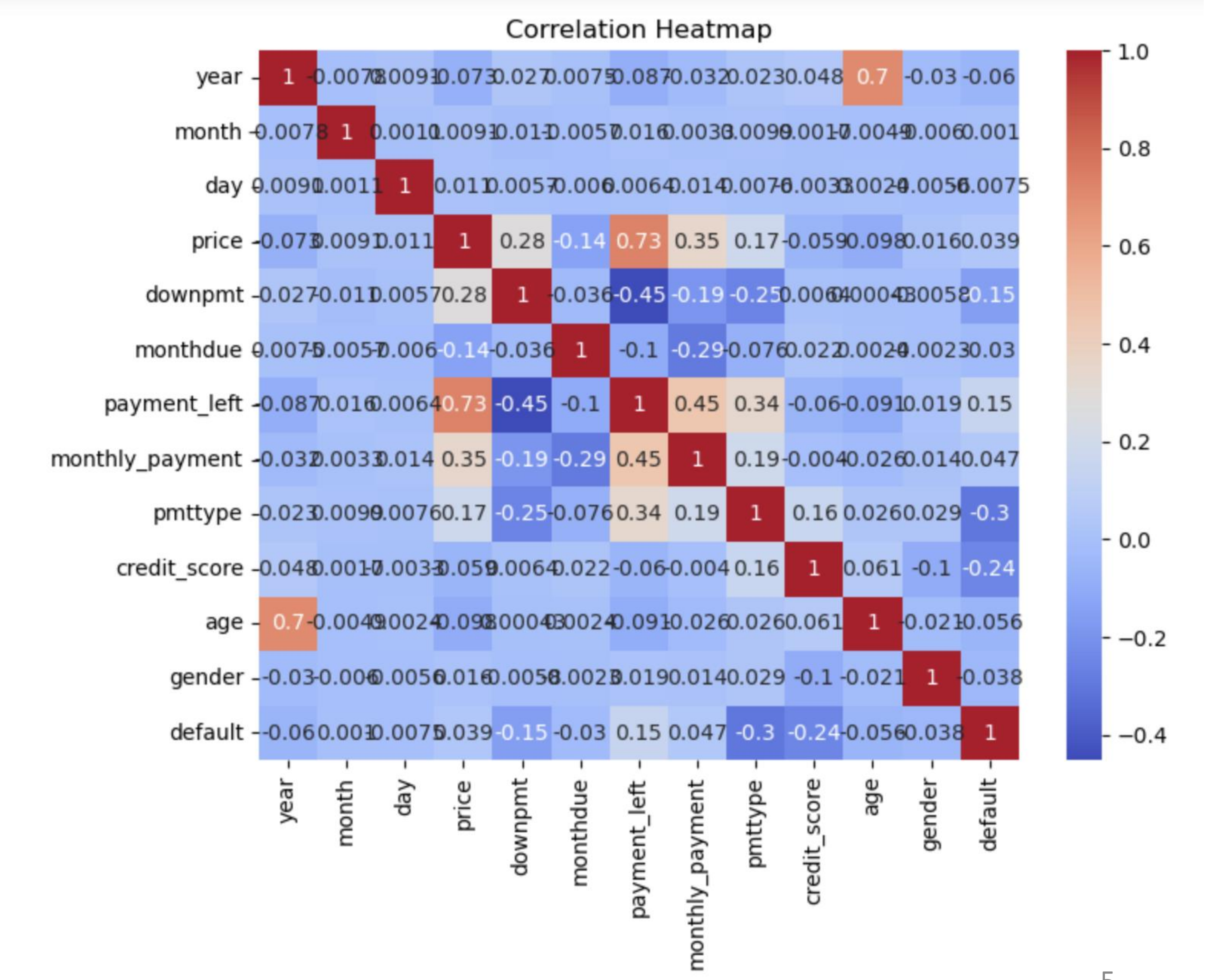
Future State

- Creation of a reliable machine learning model for early identification of payment default risks.
- Components of ML Model:
  - Front-end interface for in-store customers
  - Customer input of details
  - Activation of prediction model with single click
  - Model capable of predicting future customer default status quickly and efficiently.

Feature Selection (1/2)

- Correlation heatmap shows variables with high association to one another (the darker the more directly related)

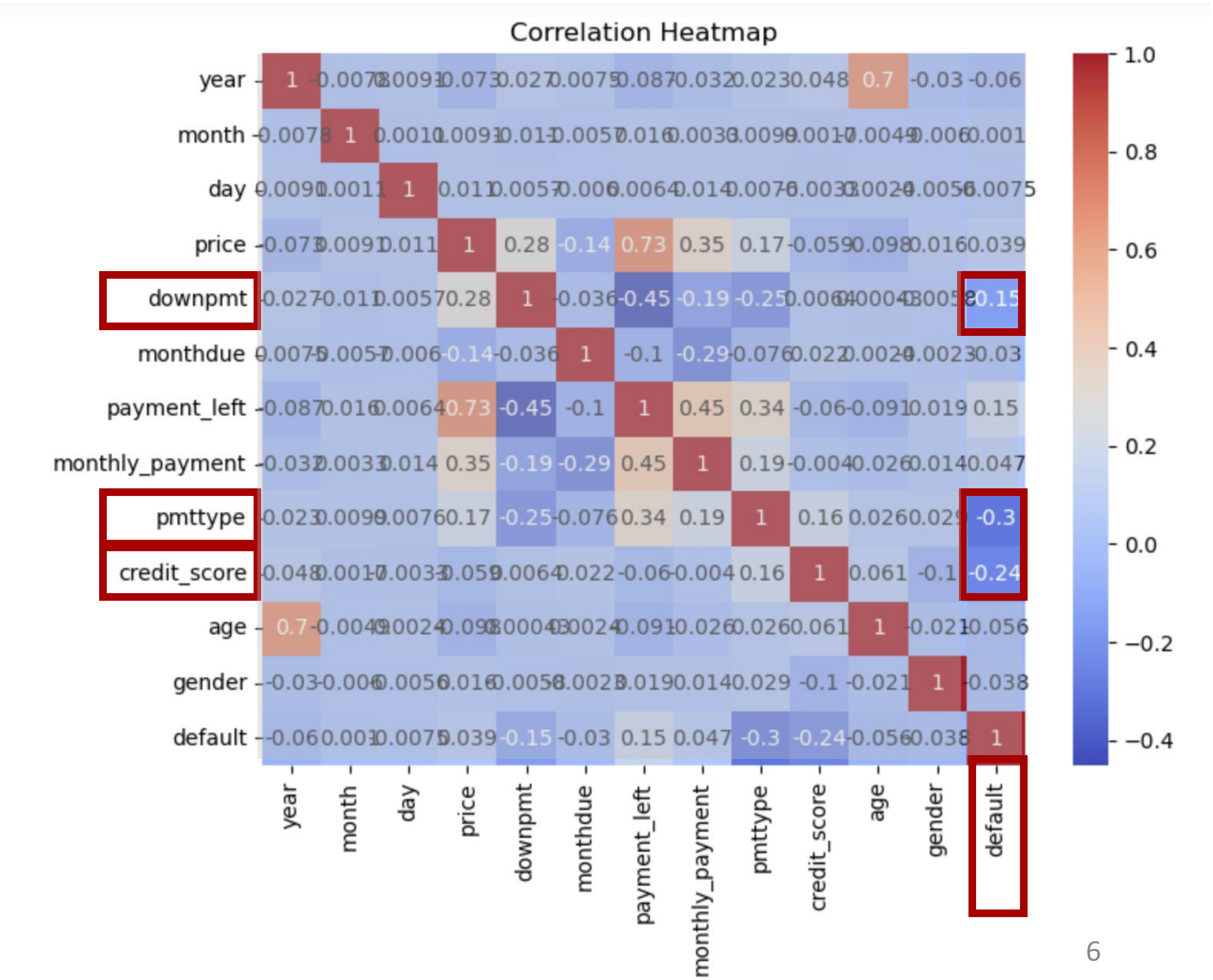
- Features most associated with the propensity for a customer to default
  - Payment Type (1- credit card, 3- store gift card, 4- debit, 5-cash, 2- NA)
  - Credit Score
  - Down Payment



Unearthing Significance in Relationships

Features most associated with the propensity for a customer to default:

- Payment Type (1: Credit Card, 3: Store Gift Card, 4: Debit, 5: Cash, 2: N/A)
- Credit Score
- Down Payment

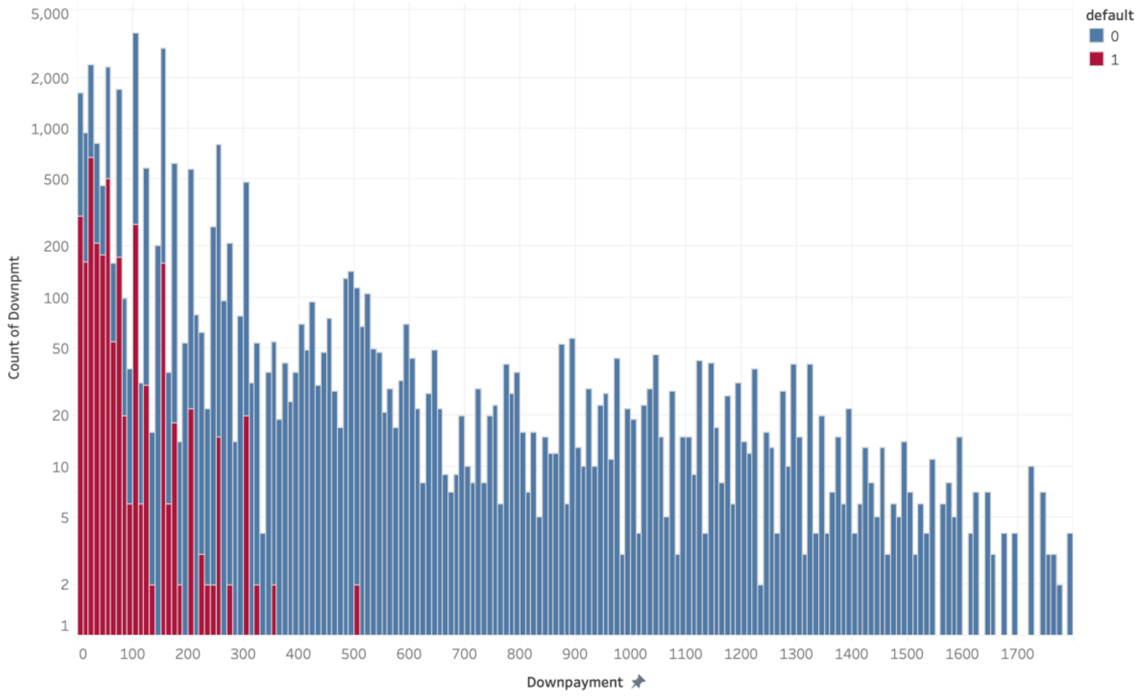


Feature Selection (2/2)

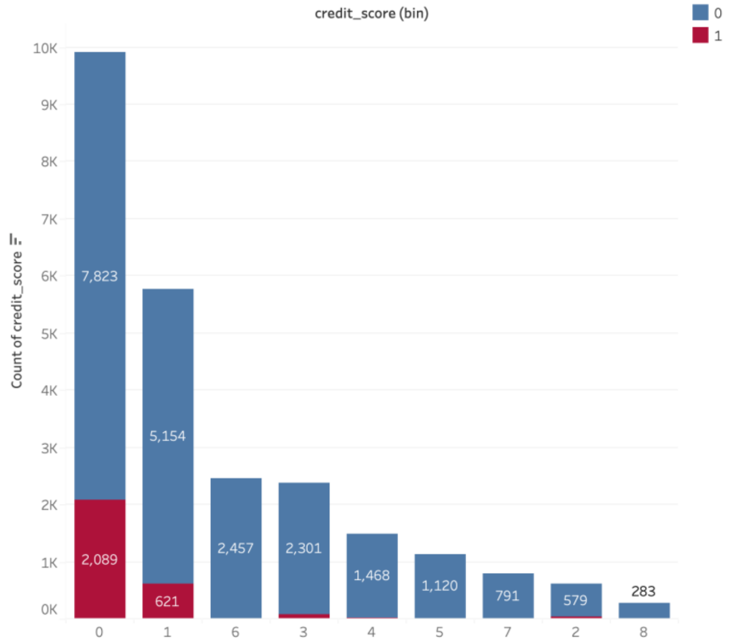
88% of Customers who defaulted (based on the historical data) had:

- Paid with a gift card
- Put < 300$ as a Down Payment
- Had a Credit score bucket of 0 or 1

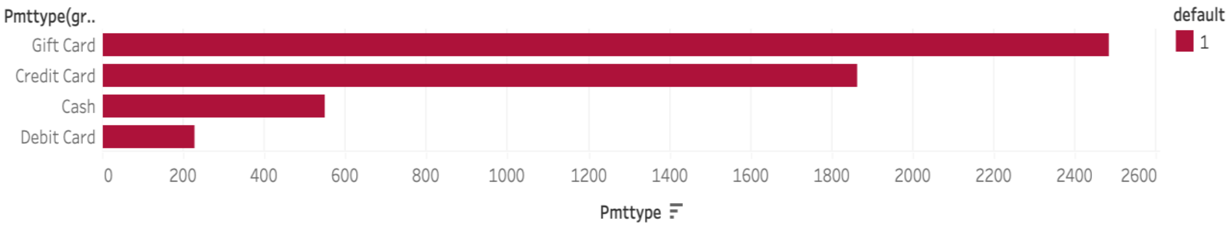
DownPayment by default



default rate by Credit Score



Payment type - default rate



Model Selection

Preliminary Criteria:

- 88% of customers who defaulted had: Payment type of 1, Down payment < 300$, Credit score bucket of 0 or 1
- Filtering based on this criteria would save ~$2.5 million debt from customers who defaulted, but would ultimately result in a loss of ~$5 million in profit from the customers who didn't default

Random Forest:

- Classification Model for prioritizing precision over recall and aiming to minimize rejecting customers who won't end up defaulting (false positives)
- Captures hidden non-linear relationships in data due to the complexity of predicting default rates
- Mitigates impact of outliers in data through aggregation of multiple weak learners (decision trees)

Why: Rejecting customers who will not default has larger cost on business over time than the loss of customers that will default.

Primary Foci: Risk analysis, increase customer attrition, maximize profit margin, demographic and financial customer targeting to ultimately reduce default status.

Random Forest Model Performance

Overview The model has a high accuracy (94%) but lower precision (67%) and recall (50%) for Defaulted Customers. It suggests that while the model is generally accurate, it may struggle to correctly identify instances of True Defaults.

Approved, Not Defaulted 8493	Rejected, Not Defaulted 279
Approved, Defaulted 584	Rejected, Defaulted 578

Approved Default- False Negative
Approved Not-default- True Negative
Rejected Default- True Positive
Rejected Not-default – False Positive

50%

Approved Customers who will Default

The false negative rate: risk factor

97%

Rate of approved Customers who will not default

True negative rate: good to be as high as possible

3.2%

Rejected Customers who will not default

False positive rate: Minimizing this number is key for greatest profit

Future Enhancements:

- Feature Engineering - Scaling or normalizing numerical features.
- Data Balancing - Use oversampling or undersampling to balance the dataset
- Hyperparameter Tuning - Experiment with different hyperparameters
- Threshold adjustment - Control the trade-off between false positives and false negatives.

Business Value

Maximize Profit

- Adjusting customer retention strategy
- Cost saving on acquiring new customers
- Profit for Customers approved who did not default (spread over 36-months) = $242,500,000
- Loss of profit from rejecting customers who did not defaults = -$8,000,000

Customer Targeting

- Identifying high-value customers with a high potential for long-term revenue
- Targeted advertising initiation for higher tier credit score
- Providing personalized service plans, bundles, or loyalty rewards that match individual customer needs
- Down payment amount customization
- Age group targeting and offers

Value Estimation

242,500,000 + – 8,000,000=

Projected Profit of $234,500,000 per 1 million applicants

Appendix

Breakdown of Estimated Values

Customer Base: 1 million

Approval Rate based on random forest model : 8493+584/ 9934 = 91.4%

Declined Rate based on random forest model : 100- 91.4= 8.6%

True Nagtive Rate: TN/TN+FP = 8493/8493+ 578 = 97%

Default Rate of Approved Applicants (based on random forest model false negative rate) : 584/584+8493 = 6.4%

*Hypothetical Default Rate of Declined Applicants based on random forest model (true positive rate = tp/tp+fn) : 50%

FPR = Fp/TN+FP = 279/279+8493 = 3.2%

FNR: Loss of False negatives (approved but defaulted) = 50%* 1 million customers *(-$1,000) = -$500,000,000

Prevent Loss of True positive (rejected defaults) = 50%* 1 million customers * $1,000 = $500,000,000

Profit for TN (approved and didn't default) (spread over 36-months) = 97% * 1 million customers * $250 = $242,500,000

Lost Profit for FP (rejected but didn't default) (spread over 36 months) = 3.2% * 1 million customers * -$250 = -$8,000,000

Model Performance

Classification Report for Random Forest Model:

	precision	recall	f1-score	support
0	0.94	0.97	0.95	8772
1	0.67	0.50	0.57	1162
accuracy			0.91	9934
macro avg	0.81	0.73	0.76	9934
weighted avg	0.91	0.91	0.91	9934

Confusion Matrix for Random Forest Model:

```
[[8493  279]
 [ 584  578]]
```

Cost Breakdown Structure:

Project Role	Hours/Wk	Hourly Salary	Total Individual Cost
Delivery Lead	8	$250/hr	$98,000/yr
Project Lead	20	$200/hr	$196,000/yr
Data Science Consultant	40	$100/hr	$196,000/yr
Strategy Analyst	40	$100/hr	$196,000/yr
Data Analyst	40	$100/hr	$196,000/yr
2 Subject Matter Advisors	4	$250/hr	$98,000/yr
Strategy Manager	2	$150/hr	$14,700/yr
Total Project Cost		$994,700	