

LOAN DEFAULT

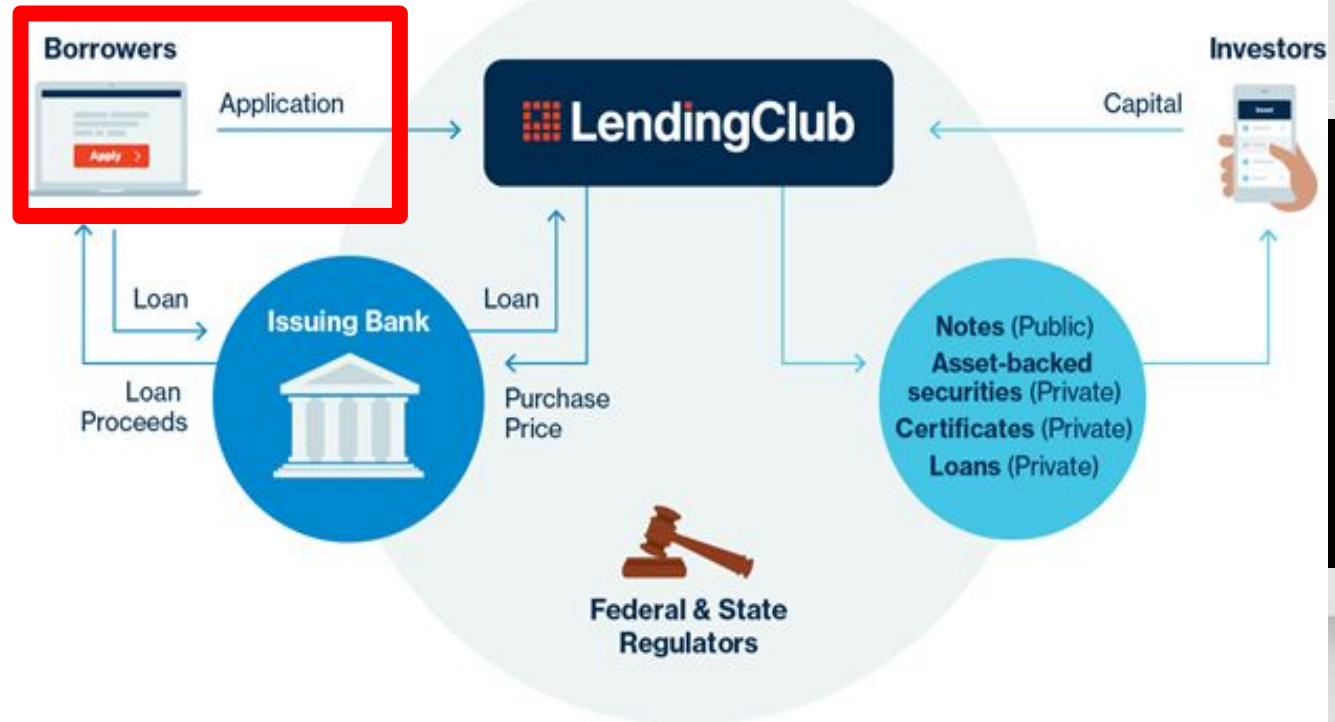
MODELING

FOR LENDING CLUB



Lending Club is a peer to peer lending company based in the United States, in which investors provide funds for potential borrowers and investors earn a profit depending on the risk they take (the borrowers credit score). The company also registers its offerings as securities with the Securities and Exchange Commission (SEC), and to offer loan trading on a secondary market. Lending Club provides the "bridge" between investors and borrowers

LENDING CLUB BUSINESS MODEL



STRATEGY



BUSINESS INTEREST

CURRENT STATE

The status of the
initiative and SWOT
Report

THE FUTURE

Exploring the
possible
enhancements and
goal setting

BUSINESS INTEREST



The client does not want to miss-classify prospective applicants during their pre-screening process



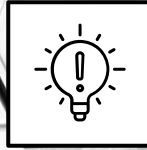
Operational risk management by early detection of applicants that could be considered at risk of default down the line

THREE PILLAR APPROACH



IDENTIFY

The ability to capture and create a risk profile of borrowers



INNOVATE

Creating a product that will be able to manage the risk appetite of Lending Club



LISTEN

Understanding the customer journey to increase revenue and increase customer inclusion

CURRENT STATE



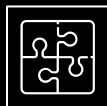
THE DATA

Asset (dataset)
Overview and
Understanding



MODELING

Current state of
the modeling
process



QUICK STATS

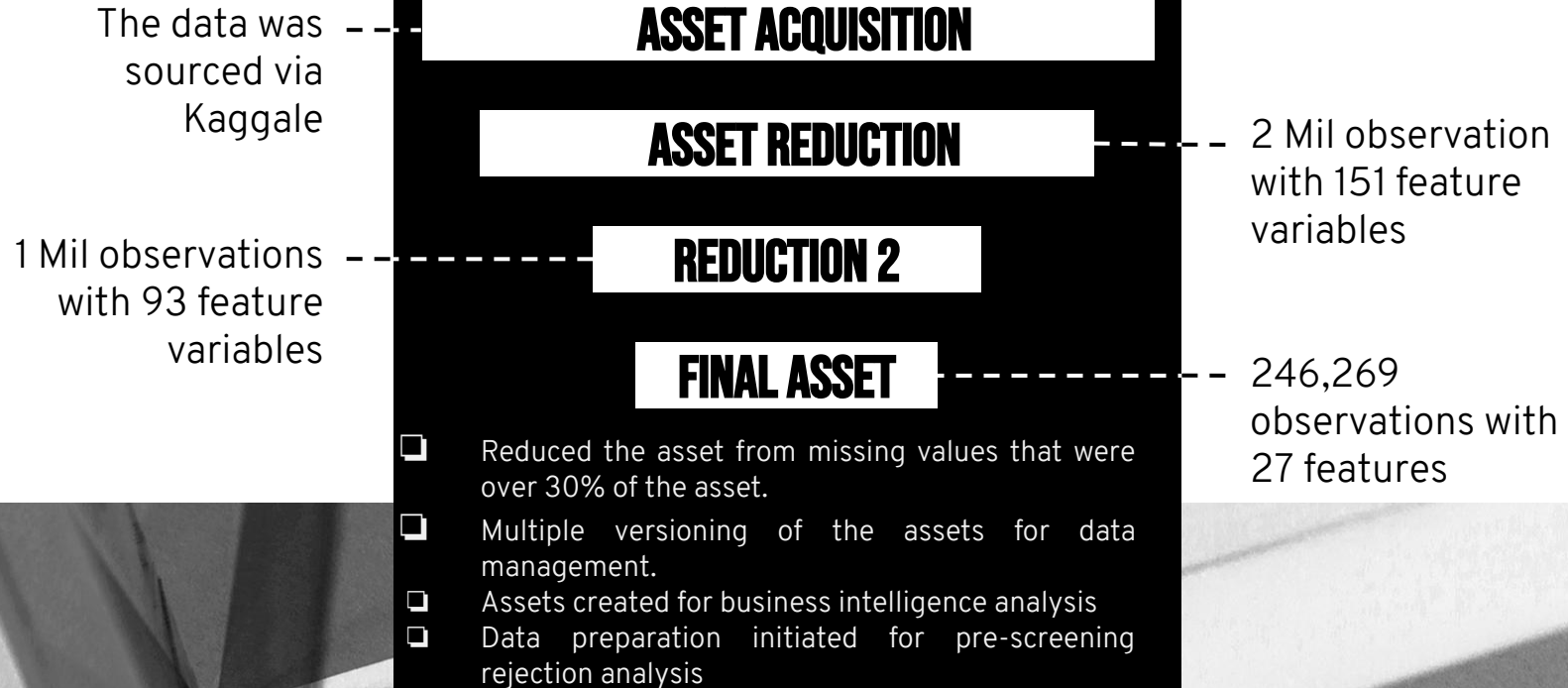
The current
landscape of the
asset



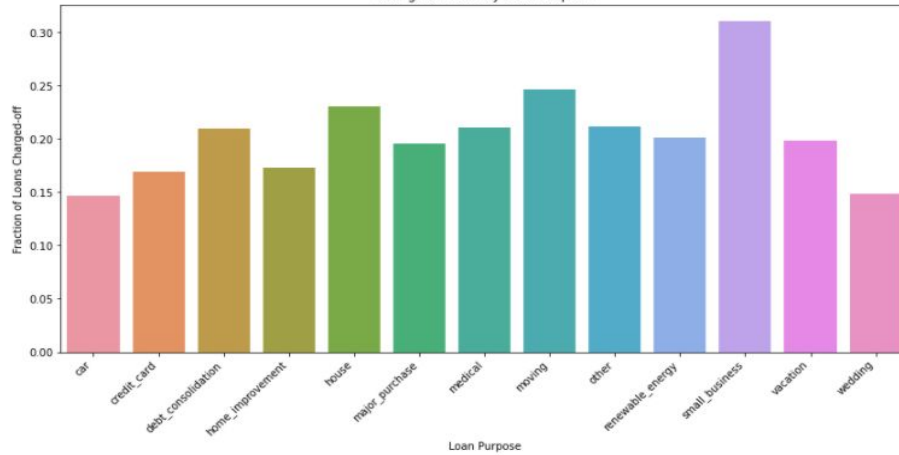
SWOT

Analysis of the
of the initiative

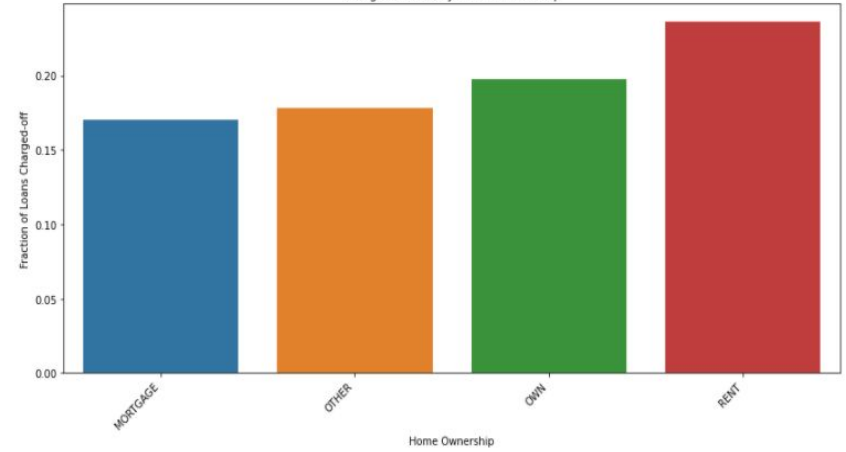
THE DATA



Charge-off Rate by Loan Purpose



Charge-off Rate by Home Ownership



QUICK STATS:



RESPONSE VARIABLE



PURPOSE

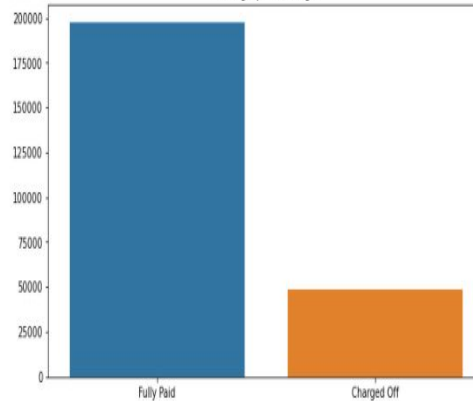


HOME OWNERSHIP

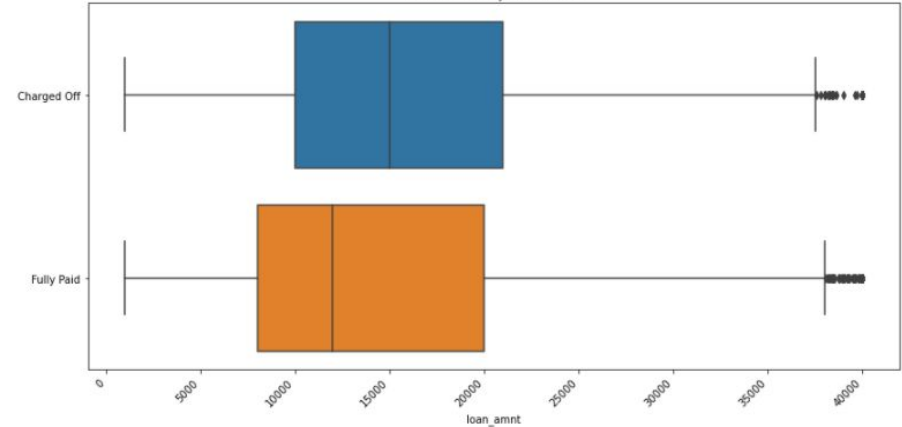


LOAN AMOUNT

Demographic of Target



Loan Amount by Loan Status





SK LEARN



PYCARET

MODELING

Performance Indicators and Model Comparison

We will prioritize Recall, Accuracy, and AUC

Benchmark indicators:

- Recall: should be above .45
- Accuracy: equal or above .60
- AUC: should be equal or above .50/.60, ideally over .70 for better balance in predicting our positive class

For the purpose of this version of the initiative we will consider two out of the three benchmark indicators as a green light for model selection. Also, in consideration, if either one metric stands out we will raise the question, if we should include it to our model selection list.

MODEL SELECTION

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
qda	Quadratic Discriminant Analysis	0.1983	0.5000	0.9997	0.1982	0.3308	-0.0000	-0.0017	5.6410
nb	Naive Bayes	0.4492	0.6587	0.8148	0.2391	0.3697	0.0911	0.1478	1.9780
lr	Logistic Regression	0.6008	0.6640	0.6449	0.2809	0.3910	0.1584	0.1883	7.8160
svm	SVM - Linear Kernel	0.6143	0.0000	0.4326	0.2928	0.2586	0.0780	0.0972	18.2330
knn	K Neighbors Classifier	0.5768	0.5273	0.4242	0.2139	0.2844	0.0283	0.0316	6.6440
dt	Decision Tree Classifier	0.7023	0.5508	0.2997	0.2722	0.2853	0.0979	0.0981	5.1950
ada	Ada Boost Classifier	0.7717	0.6637	0.2057	0.3651	0.2630	0.1403	0.1484	16.9730
gbc	Gradient Boosting Classifier	0.7904	0.6920	0.1745	0.4295	0.2481	0.1509	0.1717	73.2410
et	Extra Trees Classifier	0.7944	0.6776	0.1140	0.4299	0.1802	0.1059	0.1368	54.2830
catboost	CatBoost Classifier	0.8036	0.7162	0.1131	0.5215	0.1858	0.1239	0.1718	85.6450
xgboost	Extreme Gradient Boosting	0.8023	0.7105	0.1113	0.5058	0.1824	0.1195	0.1647	94.7230
lightgbm	Light Gradient Boosting Machine	0.8039	0.7163	0.0891	0.5320	0.1527	0.1015	0.1551	6.2130
rf	Random Forest Classifier	0.8012	0.6943	0.0760	0.4906	0.1316	0.0828	0.1305	42.8570
ridge	Ridge Classifier	0.7768	0.0000	0.0493	0.2195	0.0804	0.0084	0.0115	2.1820
lda	Linear Discriminant Analysis	0.7768	0.6699	0.0486	0.2180	0.0795	0.0077	0.0106	11.7650

BASELINE MODEL

	ACCURACY	AUC	RECALL
LOGISTIC REGRESSION	61%	67%	64%
NAIVE BAYES CLASSIFIER	47%	65%	79%
LIGHT GRADIENT BOOSTING MACHINE	80%	72%	10%

OPTIMIZED MODEL

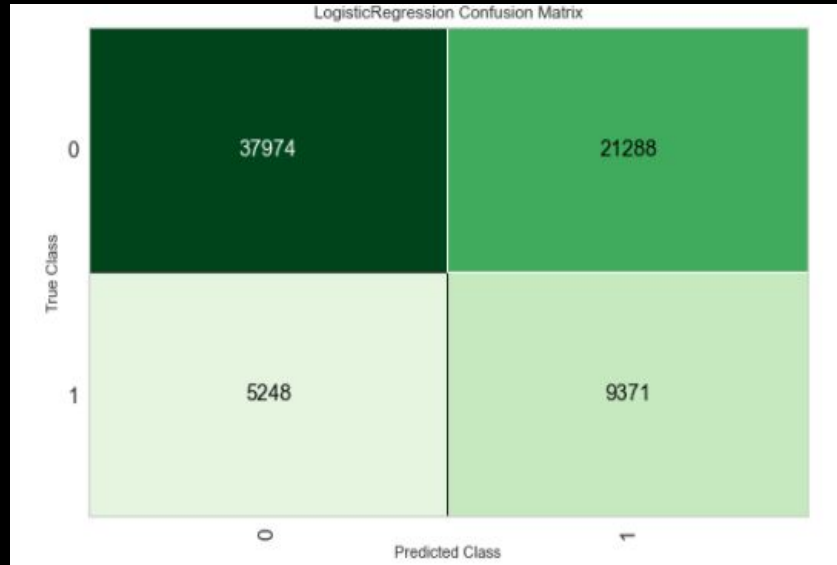
	ACCURACY	AUC	RECALL
LOGISTIC REGRESSION	60%	66%	65%
NAIVE BAYES CLASSIFIER	35%	59%	85%
LIGHT GRADIENT BOOSTING MACHINE	78%	68%	19%

BEST MODEL PERFORMANCE

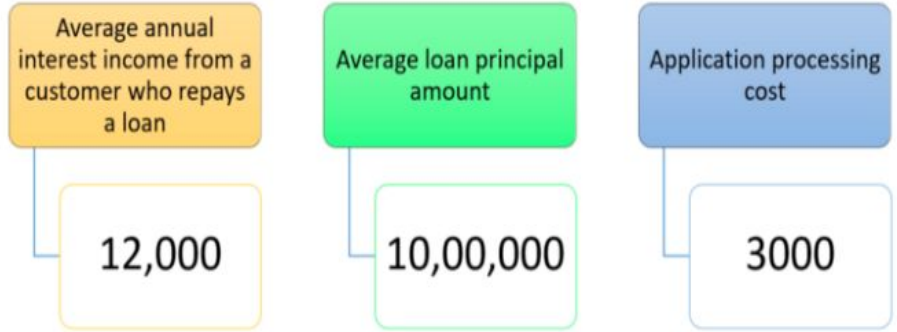
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	Logistic Regression	0.6408	0.6948	0.6410	0.3057	0.4139	0.1994	0.2279

	loan_amnt	int_rate	installment	emp_length	annual_inc	loan_status	dti	earliest_cr_line	Label	Score
105317	15000	6	457.97	4	60000.0	0	24	2004	0	0.7566
85341	16000	9	335.57	10	80000.0	0	10	2000	0	0.7024
119062	11200	8	353.35	1	40040.0	0	21	2010	0	0.7631
190777	20000	6	616.82	10	100000.0	0	16	2002	0	0.7081
121272	4300	10	140.33	4	45000.0	0	1	2007	0	0.7538
140039	4800	17	172.50	7	30912.0	0	15	1991	0	0.5979
94536	35000	29	1121.43	10	98000.0	1	22	2004	1	0.8369
205149	28000	8	879.88	10	75000.0	0	16	1987	1	0.5004
229335	10000	6	304.59	5	50000.0	0	11	1994	0	0.9489
79818	20000	7	622.04	3	106000.0	0	21	1976	0	0.6348

BUSINESS TRANSLATION: NET REVENUE FUNCTION



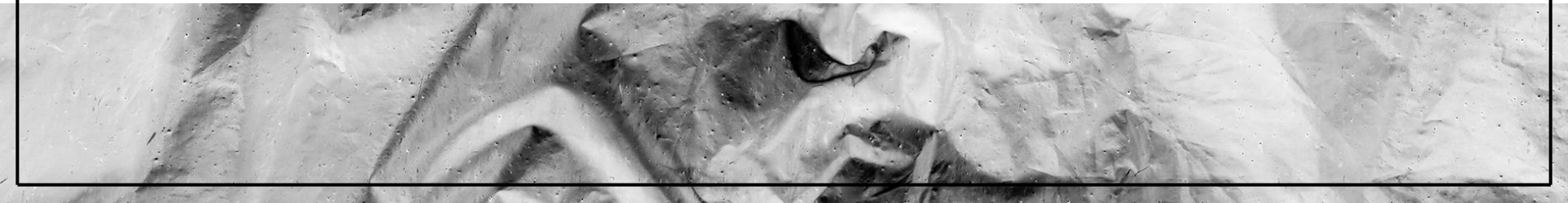
$$\begin{aligned}
 \text{Net revenue} = & (\# \text{ of TN} \times \text{Average annual interest income per customer}) \\
 & - (\# \text{ of FP} \times \text{False positive cost per customer}) \\
 & - (\# \text{ of FN} \times \text{False negative cost per customer}) \\
 & - (\# \text{ of TP} \times \text{True positive cost per customer})
 \end{aligned}$$



False positive cost	False negative cost	True positive cost	True negative revenue
<ul style="list-style-type: none"> interest revenue lost by predicting a customer who will repay as a defaulter, thereby not giving him the loan 	<ul style="list-style-type: none"> principal amount lost by sanctioning loan to a customer who will default 	<ul style="list-style-type: none"> not sanctioning loan to a customer who will default, thereby forfeiting the application processing fee 	<ul style="list-style-type: none"> Average annual interest income from a customer who repays a loan
-12,000	-10,00,000	-3000	12000

\$10 BILLION

POTENTIAL NET REVENUE



INITIATIVE SWOT ANALYSIS



STRENGTHS

- ❑ Business Analysis
- ❑ Initiative Workflow and Process Management
- ❑ Leveraging External dependencies, when needed



WEAKNESSES

- ❑ Computing Power and Resources
- ❑ More Efficient Model Process Enhancement



OPPORTUNITY

- ❑ Modeling Process and Workflow Management Enhancements
- ❑ Business Intelligence Dashboard
- ❑ Quantify Potential Revenue (gain and loss)
- ❑ RPA (robotic process automation)



THREATS

- ❑ Resource Miss-Management
- ❑ Deadline Awareness

FUTURE PRIORITY GOALS

GOAL 1 (P1)

Begin rejected applicant analysis report for business stakeholder

GOAL 2 (P2)

User acceptance testing (UAT) of Tableau business intelligence dashboards for key stakeholders

GOAL 3 - LOW PRIORITY

Front-end user interface for model deployment

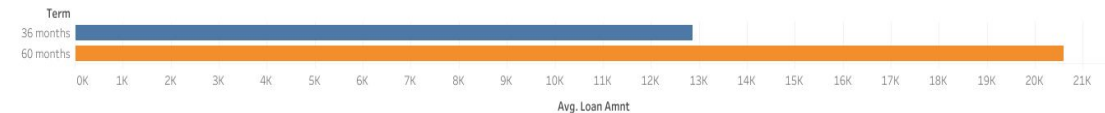
TABLEAU ANALYTICS DASHBOARD

BUSINESS INTELLIGENCE FEATURES:

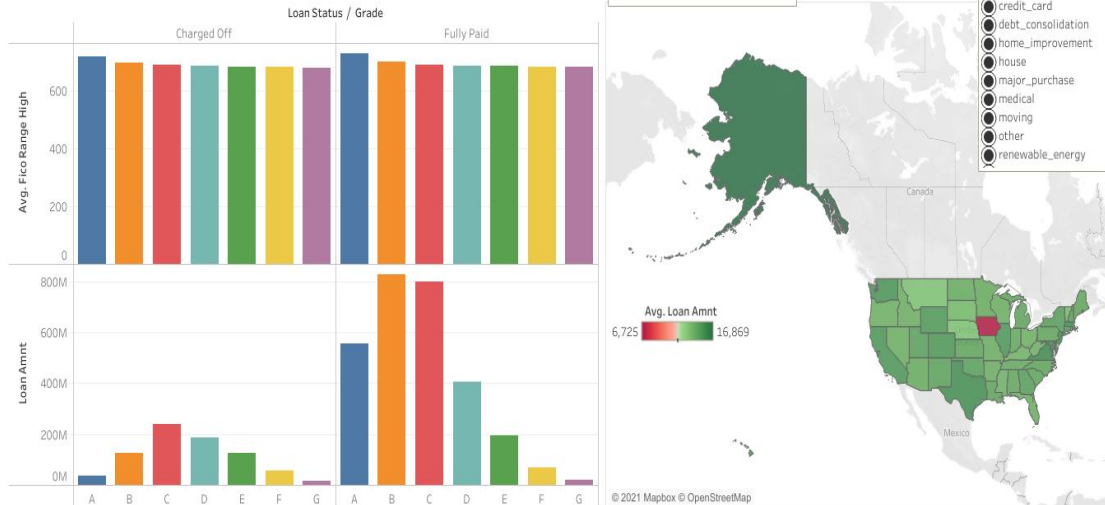
- ☐ BUSINESS PROFILE
- ☐ CUSTOMER ANALYTICS
- ☐ APPLICATION PRE-SCREEN
 - ☐ APPLICANT REJECTION ANALYTICS
 - ☐ INCLUSION OPTIMIZATION ANALYSIS
- ☐ PROJECT BENCHMARK ANALYSIS

Dashboard 1 (2)

Sheet 4



Sheet 1





THANK YOU

CHRISTIAN CORRALES

Business Intelligence Analyst
(Role title for academic project)

corrales.christian1228@gmail.com

<https://github.com/ccorrales1228/Capstone>

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.