# How to Prevent a Car Loan Default Crisis
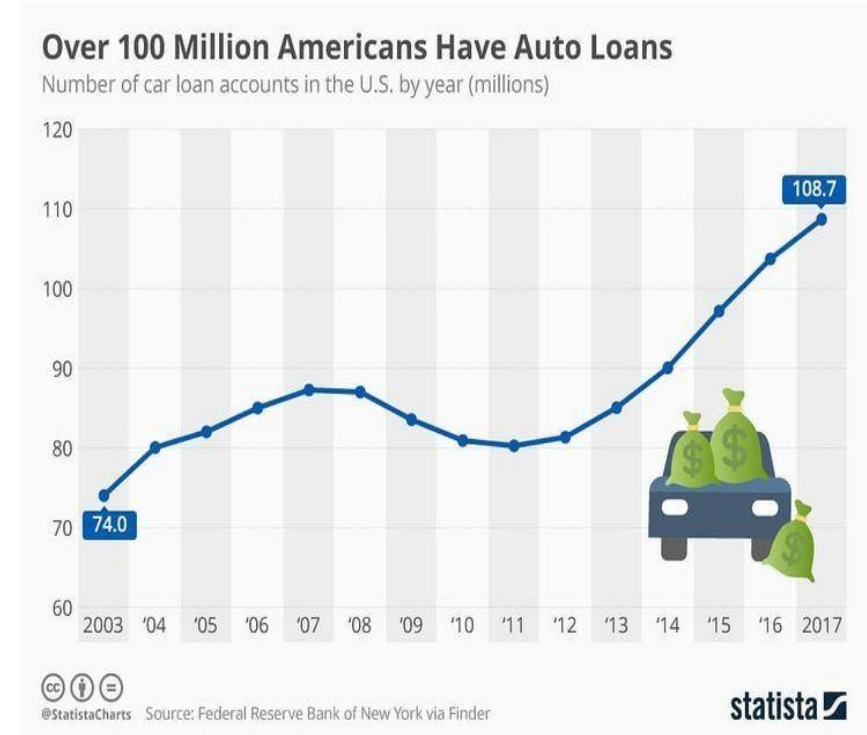
By Michael Scognamiglio
nyc-mnhtn-ds-080320

# Presentation Structure

➔ What is this Project about and What is the Business Case?

➔ What's the Dataset and Process like?

➔ What did you found during EDA and statistical testing?

➔ What was the modeling process like?

➔ What did you find through Modeling?

➔ What's next?

# What is this Project Anyway?

- Car Loan Defaults
- America has over $1 trillion dollars in car loan debt
- Build a model that can accurately flag whether or not a loan is trouble
- Provide Insights through EDA and stat testing what factors should be looked at the closest
- Analyze final model to see if any additional insights can be gleaned



**Over 100 Million Americans Have Auto Loans**
Number of car loan accounts in the U.S. by year (millions)

74.0 ... 108.7 (2003–2017)

@StatistaCharts   Source: Federal Reserve Bank of New York via Finder

statista

# Dataset and Process

- Dataset was obtained from Financial Institutions Data from India
- Compiled into large Kaggle Dataset
- 200,000 rows and 60 initial features
- Initial Steps involved Data Cleaning and Data Exploration
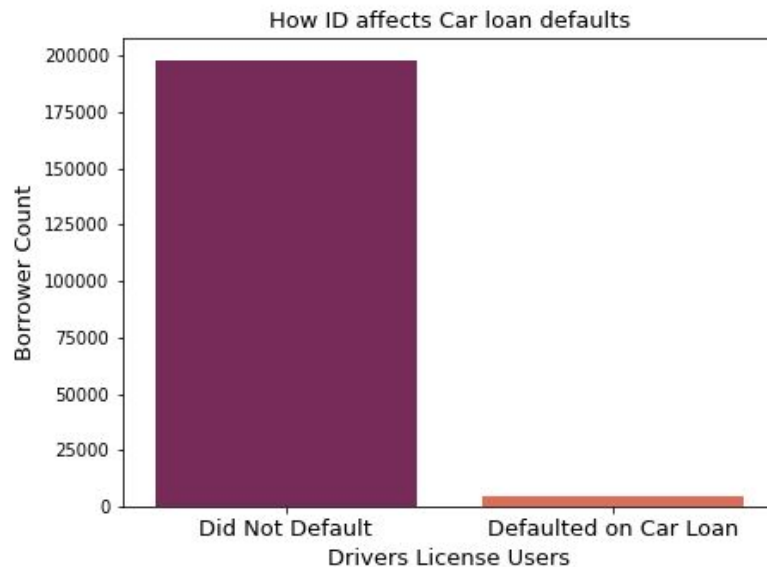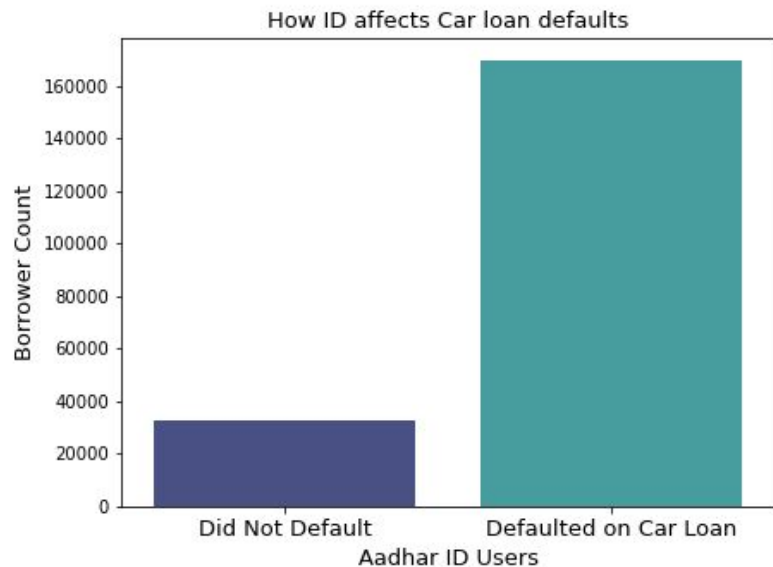- Major issue was Class imbalance about 5/1
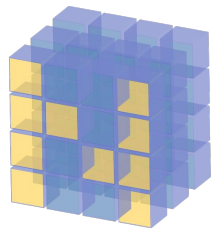
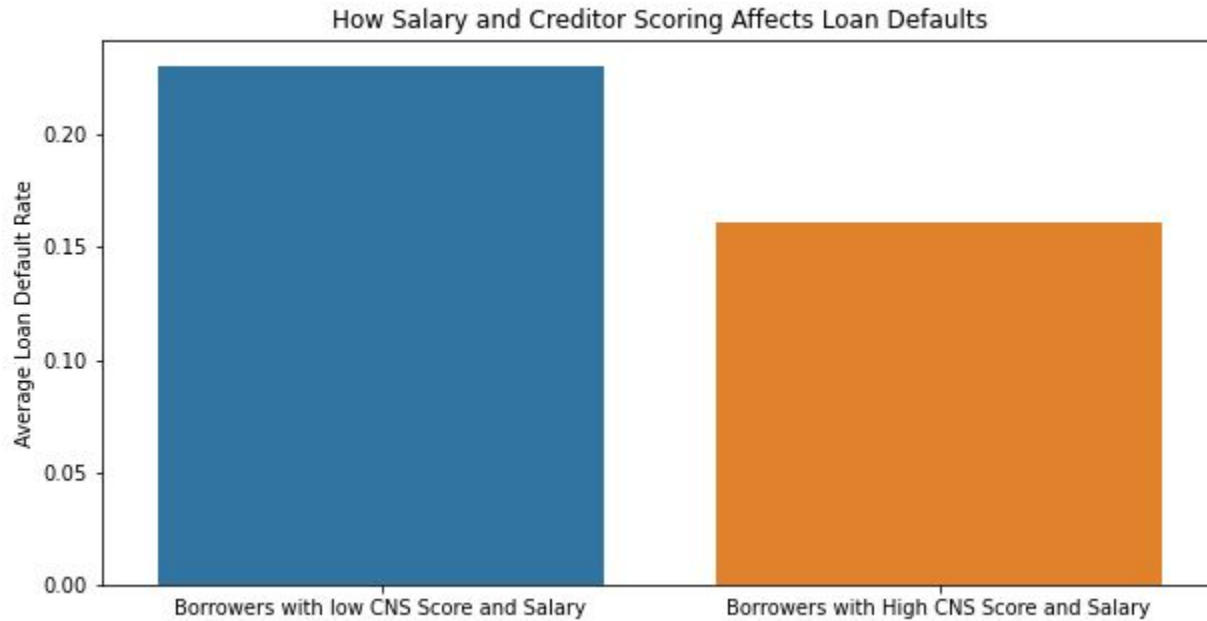# EDA and Statistical Testing Findings

# EDA and Statistical Testing Findings



How Salary and Creditor Scoring Affects Loan Defaults

# Model Selection

| Model | F1 Score | Recall Score |
|---|---|---|
| Logistic Regression | .654 | .90 |
| KNN | .550 | .56 |
| Decision Tree | .660 | .93 |
| Random Forest | .550 | .66 |
| XGBoost | .550 | .65 |
| Voting Classifier | .654 | .79 |

**All values are Test set scores.**
**Recall Score was prioritized as evaluation metric to minimize false negatives**
**(predicted won't default but in reality did default)**
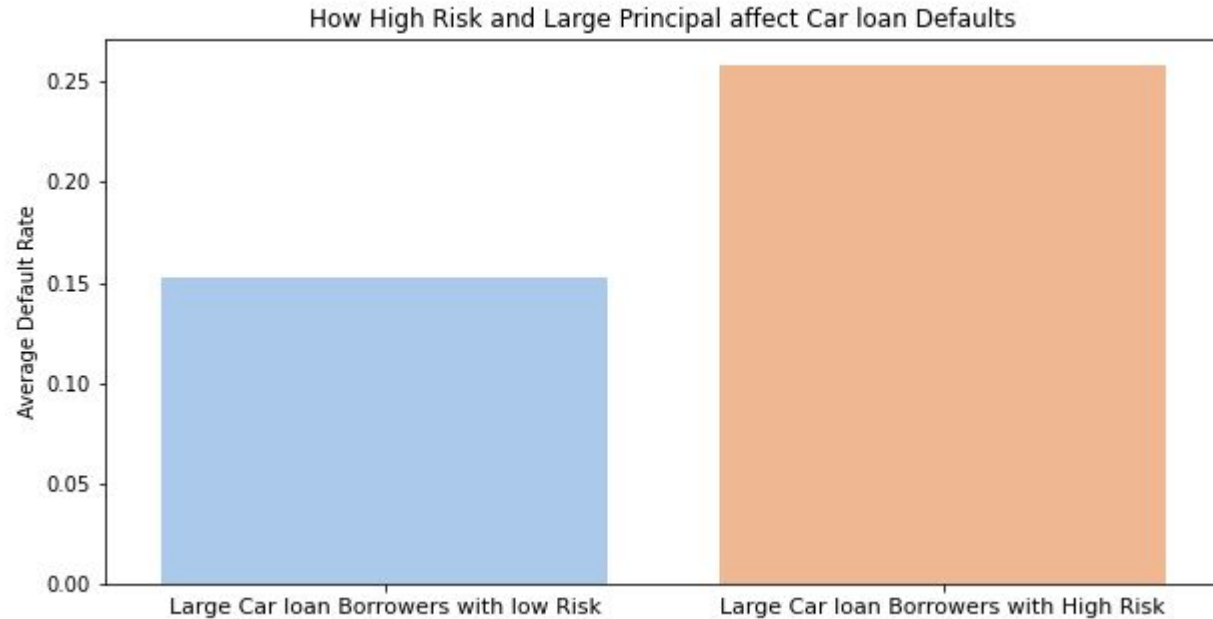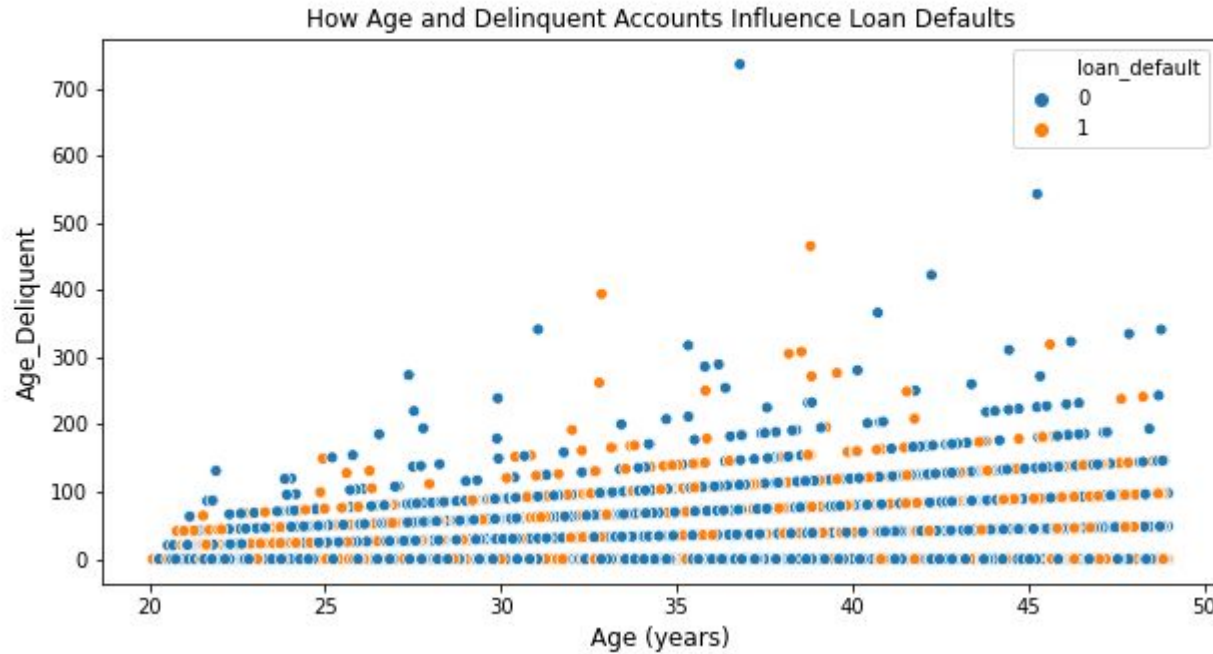
# Final Model
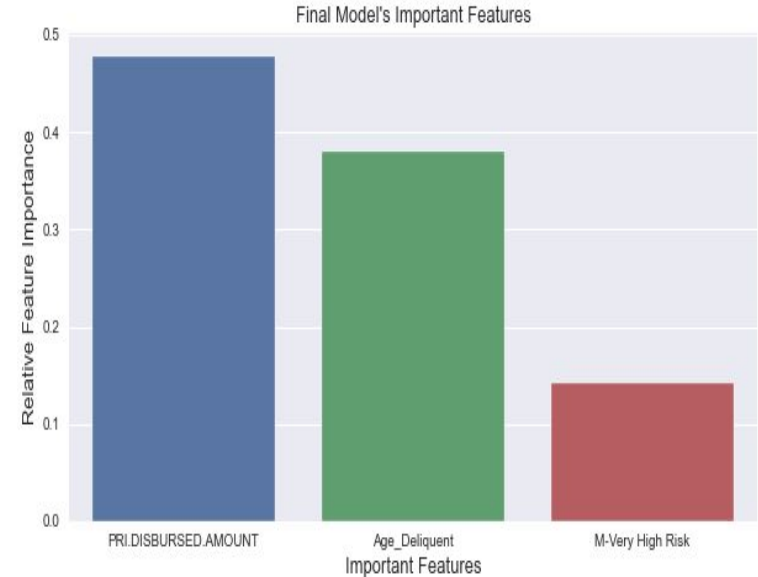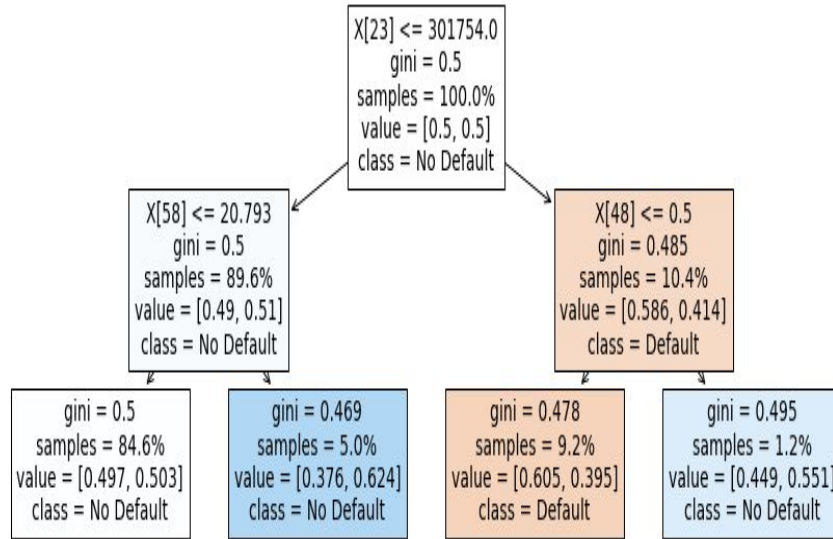


| Decision Tree | .660 | .93 |
|---|---|---|

# Model Conclusions



How High Risk and Large Principal affect Car loan Defaults

# Model Conclusions



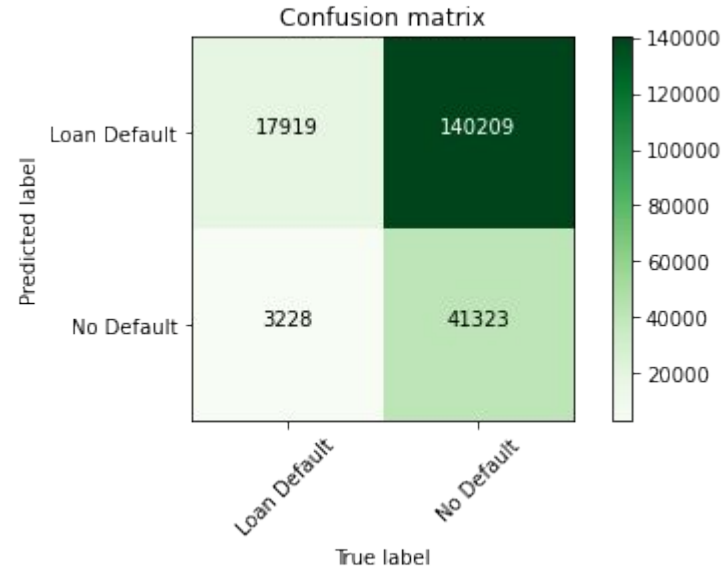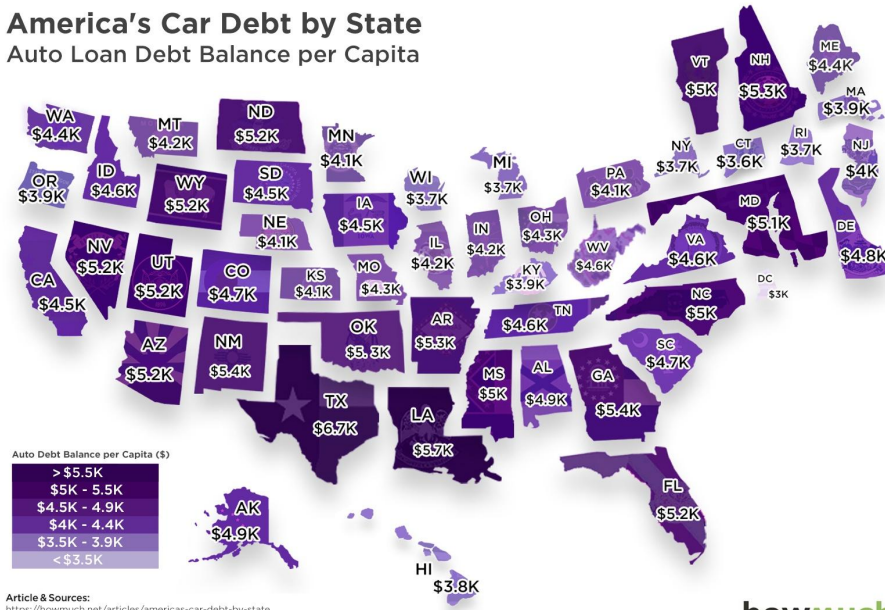How Age and Delinquent Accounts Influence Loan Defaults

# Next Steps

- Final model did a great job of minimizing false negatives.

- However that came at a cost of high Type I Error (false positives).

- For this business case, this was acceptable because false negatives are far worse.

- In the future, I want to explore models that can optimize the ratio of these parameters better.



Confusion matrix

# Any Questions??



**America's Car Debt by State**
Auto Loan Debt Balance per Capita

WA $4.4K
OR $3.9K
ID $4.6K
MT $4.2K
WY $5.2K
ND $5.2K
SD $4.5K
MN $4.1K
NE $4.1K
NV $5.2K
UT $5.2K
CO $4.7K
KS $4.1K
IA $4.5K
MO $4.3K
WI $3.7K
IL $4.2K
IN $4.2K
MI $3.7K
OH $4.3K
KY $3.9K
WV $4.6K
PA $4.1K
NY $3.7K
VT $5K
NH $5.3K
ME $4.4K
MA $3.9K
RI $3.7K
CT $3.6K
NJ $4K
MD $5.1K
DE $4.8K
VA $4.6K
NC $5K
TN $4.6K
SC $4.7K
CA $4.5K
AZ $5.2K
NM $5.4K
OK $5.3K
AR $5.3K
MS $5K
AL $4.9K
GA $5.4K
TX $6.7K
LA $5.7K
FL $5.2K
DC $3K
AK $4.9K
HI $3.8K

Auto Debt Balance per Capita ($)
> $5.5K
$5K - 5.5K
$4.5K - 4.9K
$4K - 4.4K
$3.5K - 3.9K
< $3.5K

**Article & Sources:**
https://howmuch.net/articles/americas-car-debt-by-state
New York Federal Reserve - https://newyorkfed.org

howmuch.net

Thanks.