Loan Default Analysis: Are Unqualified Borrowers Being Targeted by Lenders?

THE iSCHOOL Syracuse University

Rohan Nitin Mahajan, Christy Sato, Chris Smith, Lennart Zeugner

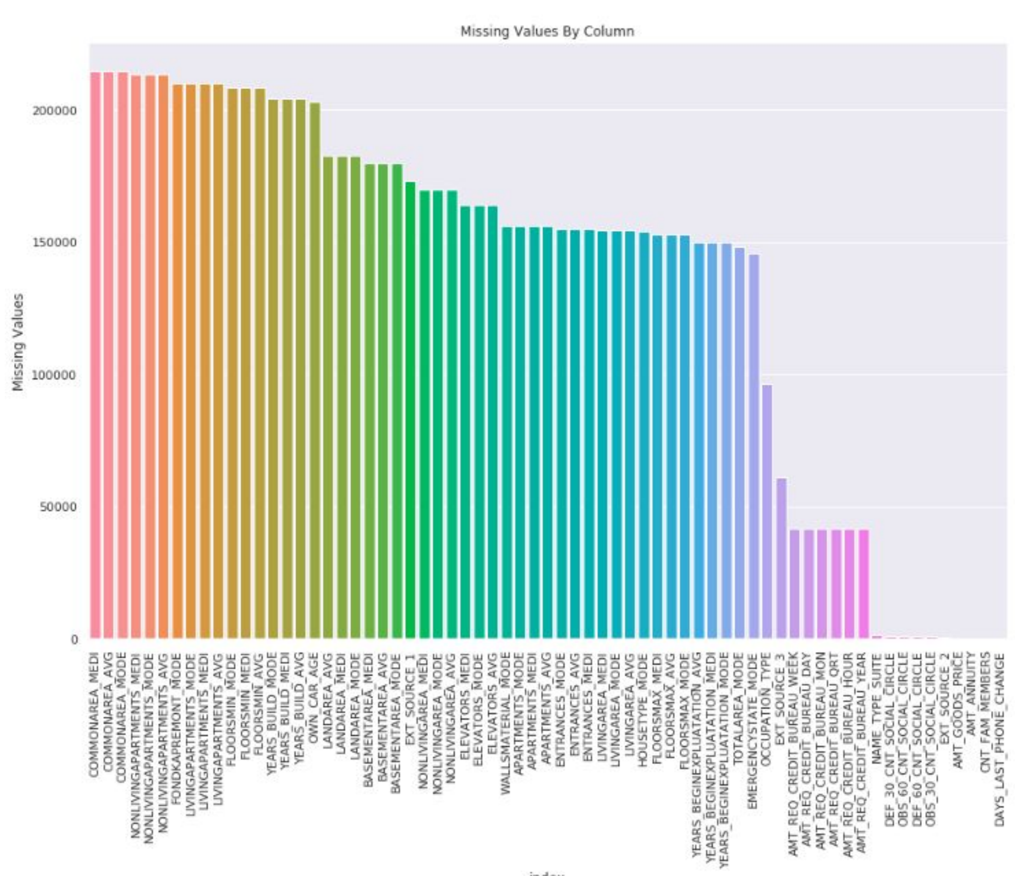
Problem and Objectives

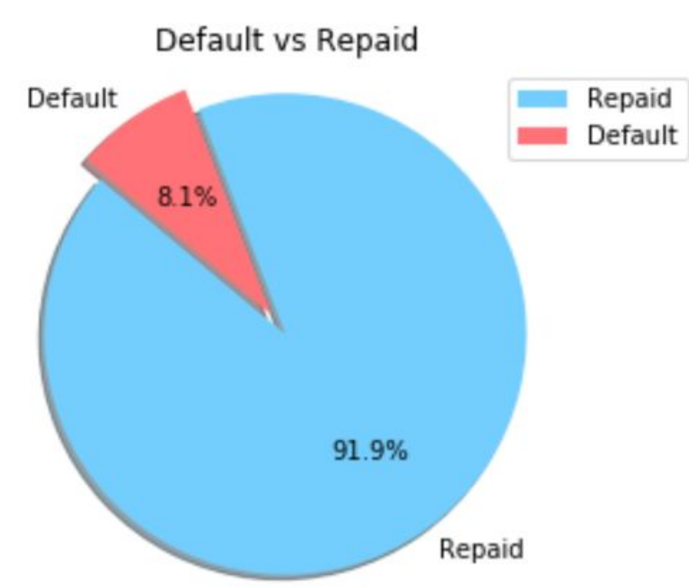
Today, unfortunately, many individuals struggle to get personal loans due to insufficient or non-existent credit histories. Often times this population is taken advantage of by untrustworthy lenders. Clients are misguided into thinking lenders have their best interests in mind, but in fact, are only concerned with boosting the business' bottom line. As a result, these clients are set up to fail on repayments which can lead to their loans going into default.

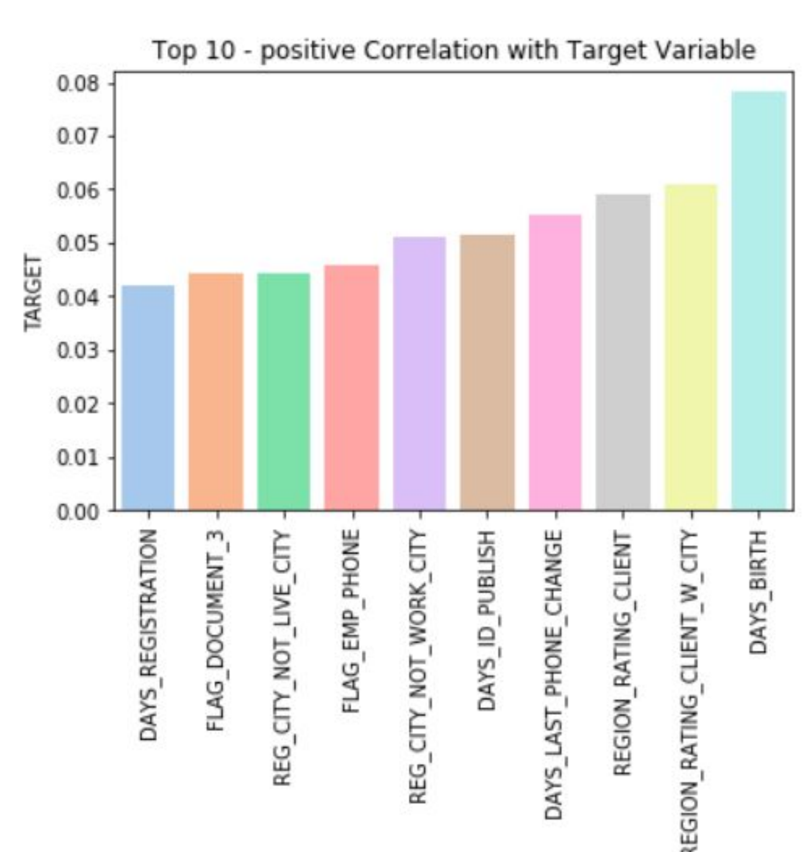
To provide a more positive loan experience, the analysts will use a variety of alternative data to predict the client's repayment abilities. By understanding consumer financial status and habits, suggestions will be implemented and loans that best fit their situation to lead the clients to be successful will be recommended.

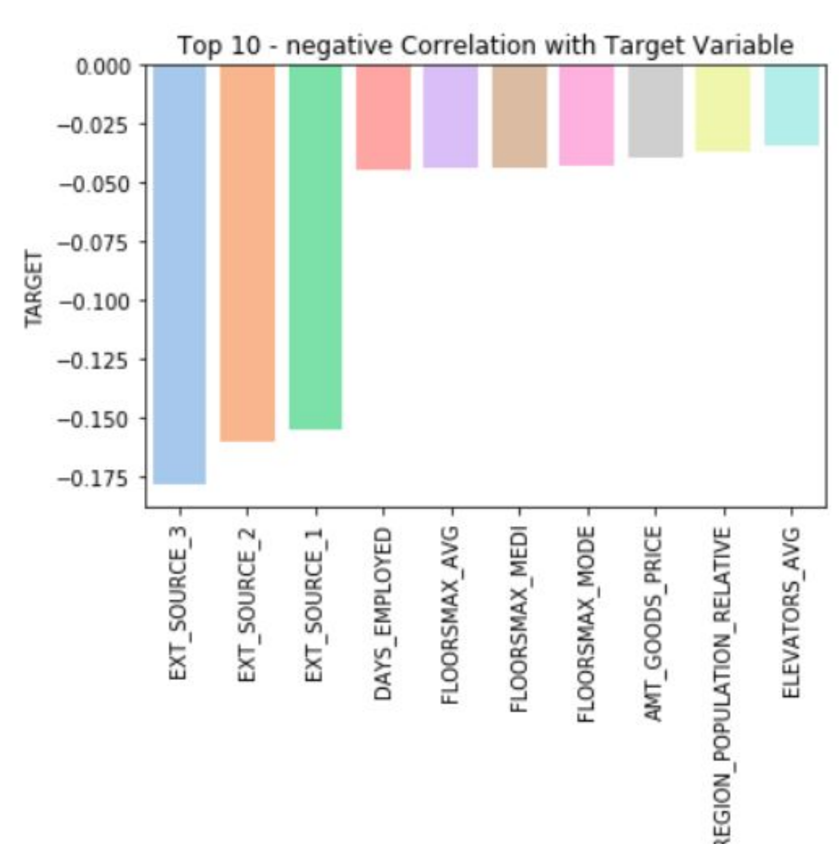
Data Description

The dataset consists of 120 features and 307,511 observations. With about 51 columns with 50% or more of missing values and 91.9% of repaid loans while only having 8.1% of loans defaulted, the dataset is unbalanced and needs preprocessing. By taking a closer look at the loan target variable, and focus on its top positive and negative correlations, both feature engineering and dimensionality reduction are intuitive next steps as models are created.

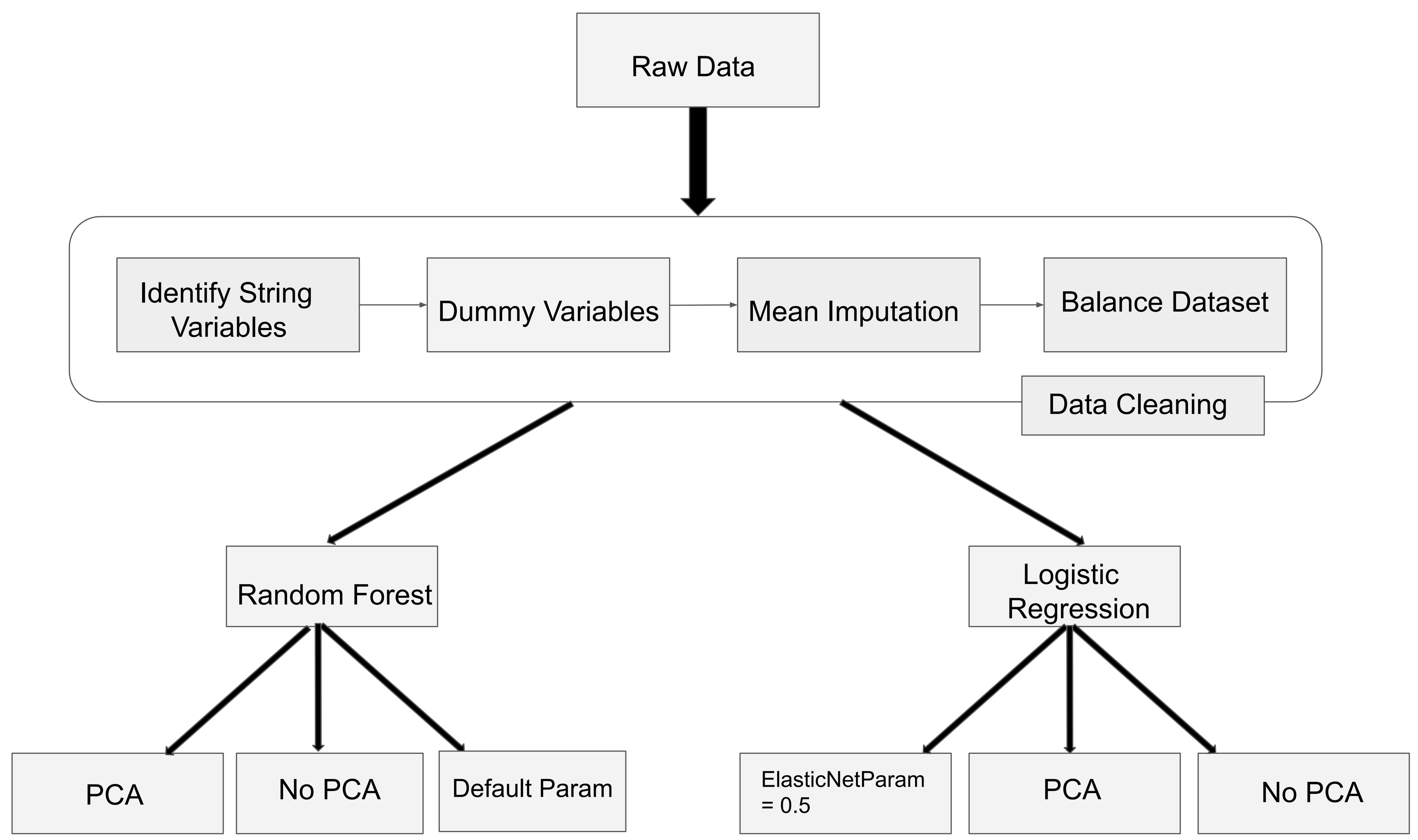


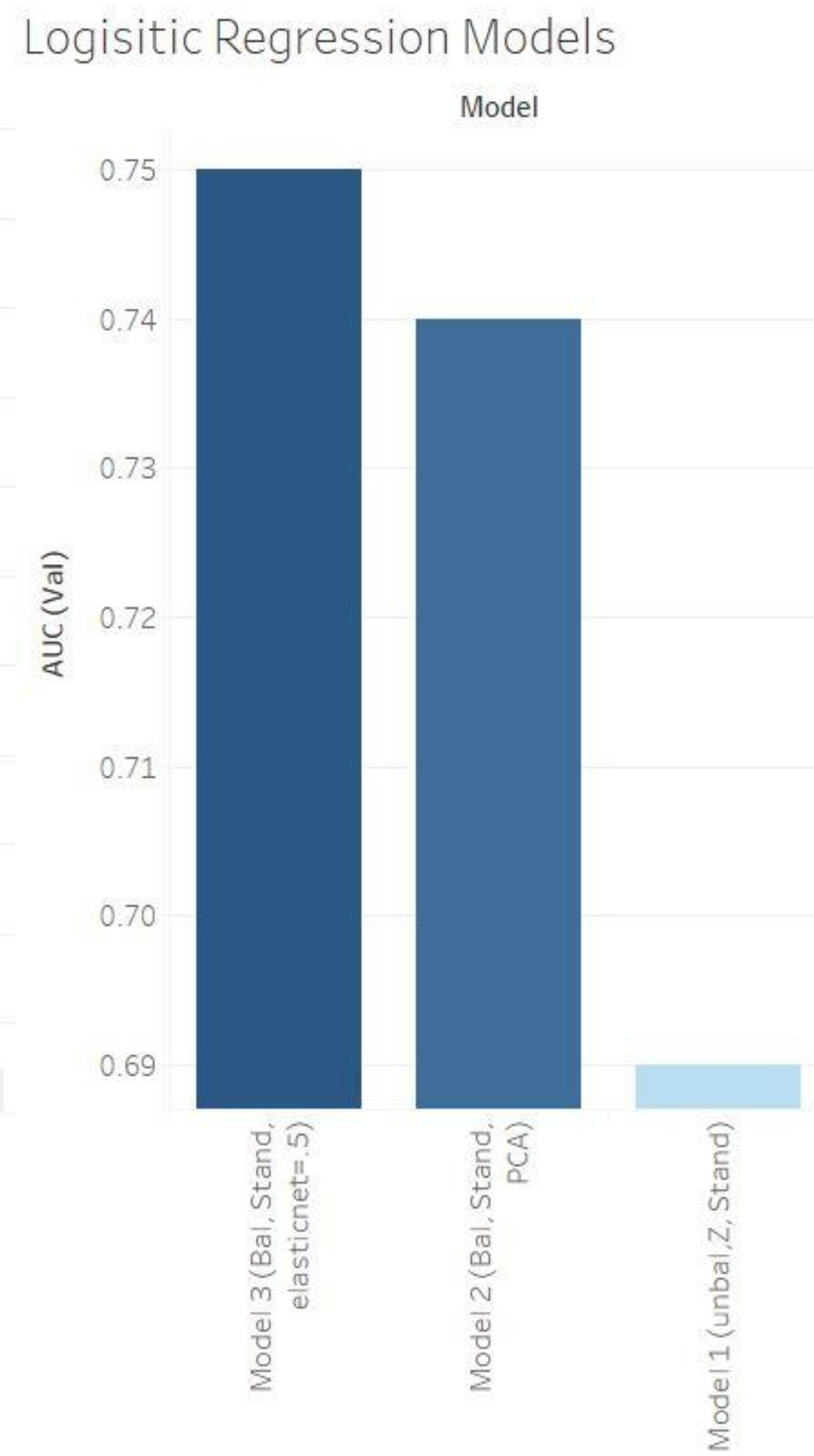






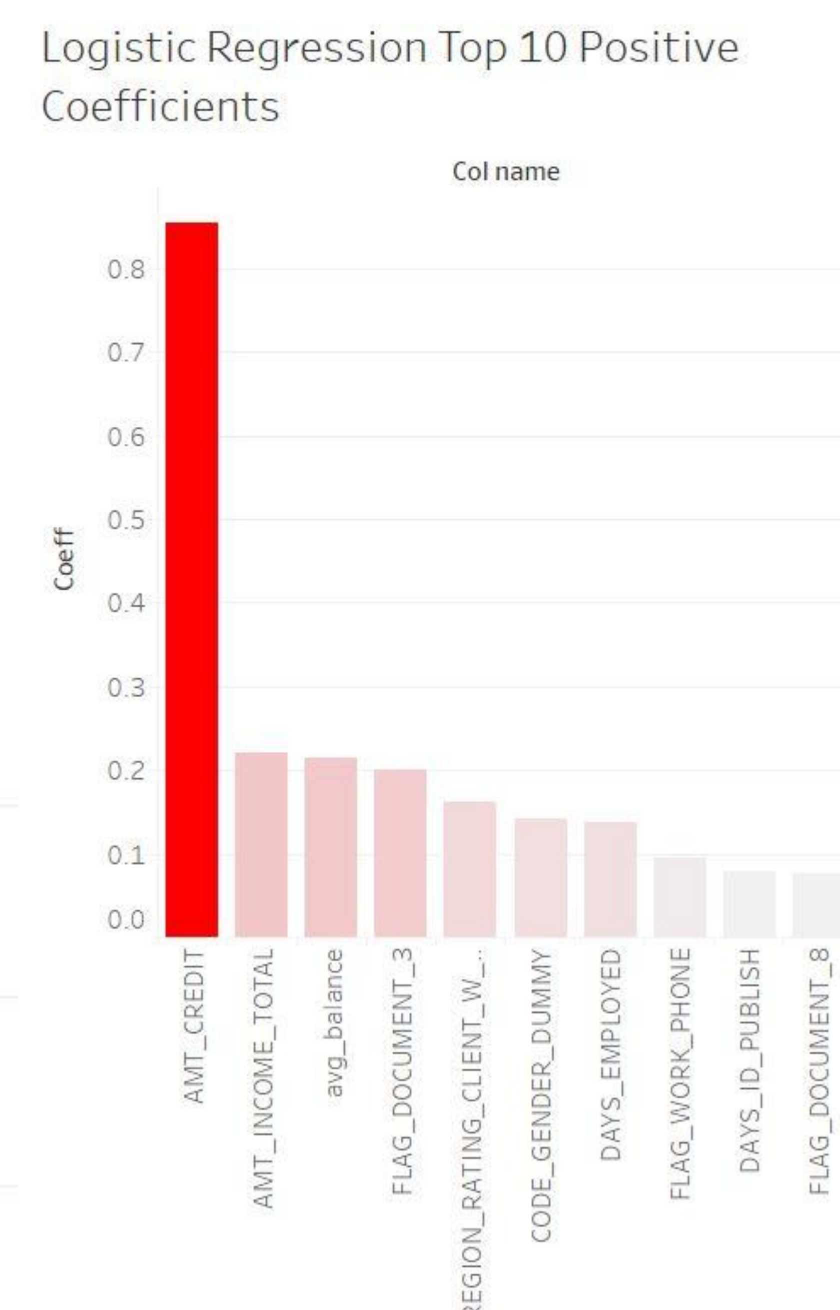
Data Flow

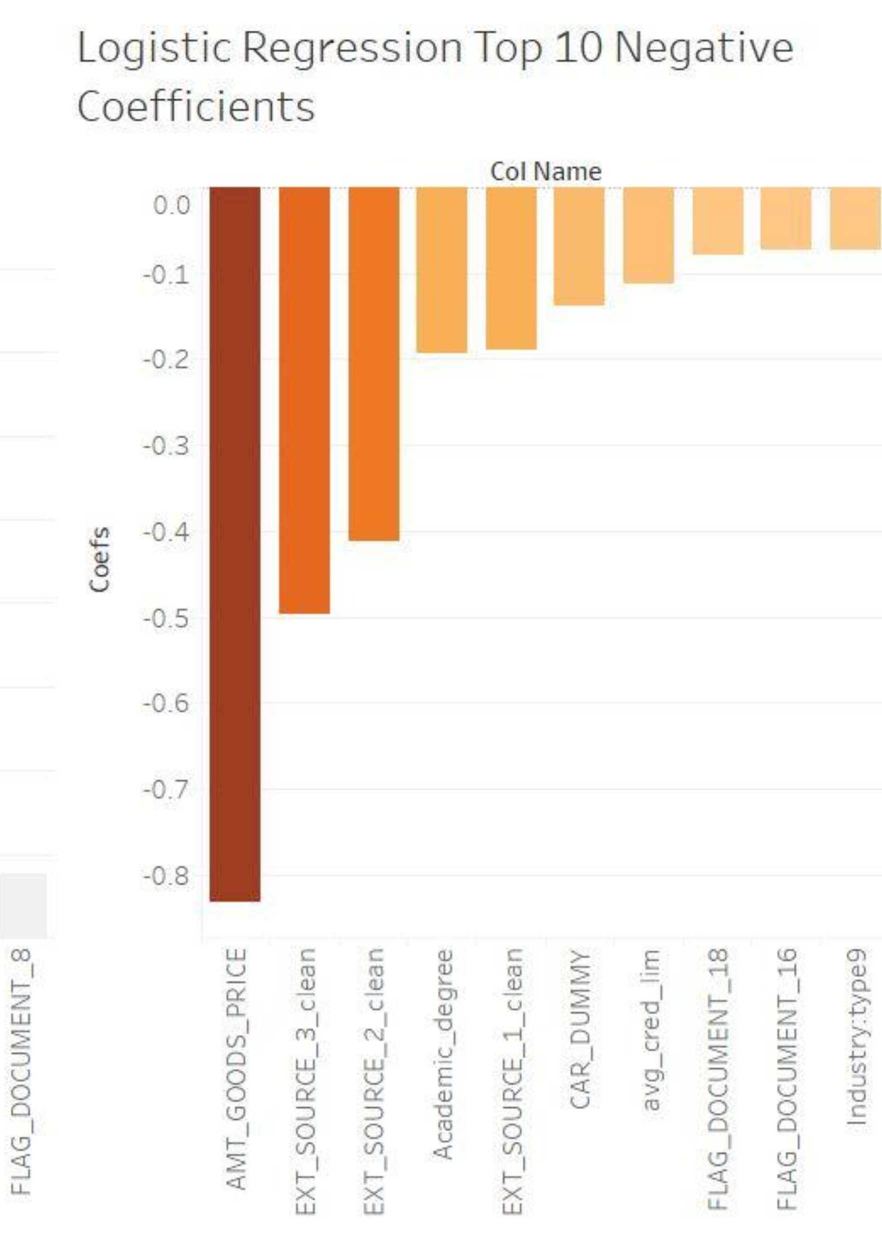




Logistic Regression Model Coefficients - LR Model 3

LR Model 3 Performance Metrics	Label	
Metric	0	1
Precision	0.69	0.68
Recall	0.68	0.69
AUC	0.749	





Conclusion

By analysing the coefficients of the final model, several intuitive conclusions can be made:

- Those borrowers that have larger amount of credit are more likely to default
- This can be reiterated with the values of average credit balance, amount income, and days employed appearing in the top positive coefficients
- It then makes sense to see academic degree, amount consumer loan, and ownership of a car in the negative coefficients
- Therefore, we can then conclude that educated, employed, asset owning, well qualified borrowers are more likely to default on their loans
- This disproves our initial research question as the aforementioned qualities likely describe borrowers that are qualified to receive loans
- Other findings of note
  - Supplying work phone on application
  - How days before application did client switch identification
  - Males more likely to default
  - Borrowers working in financial industry are less likely to default