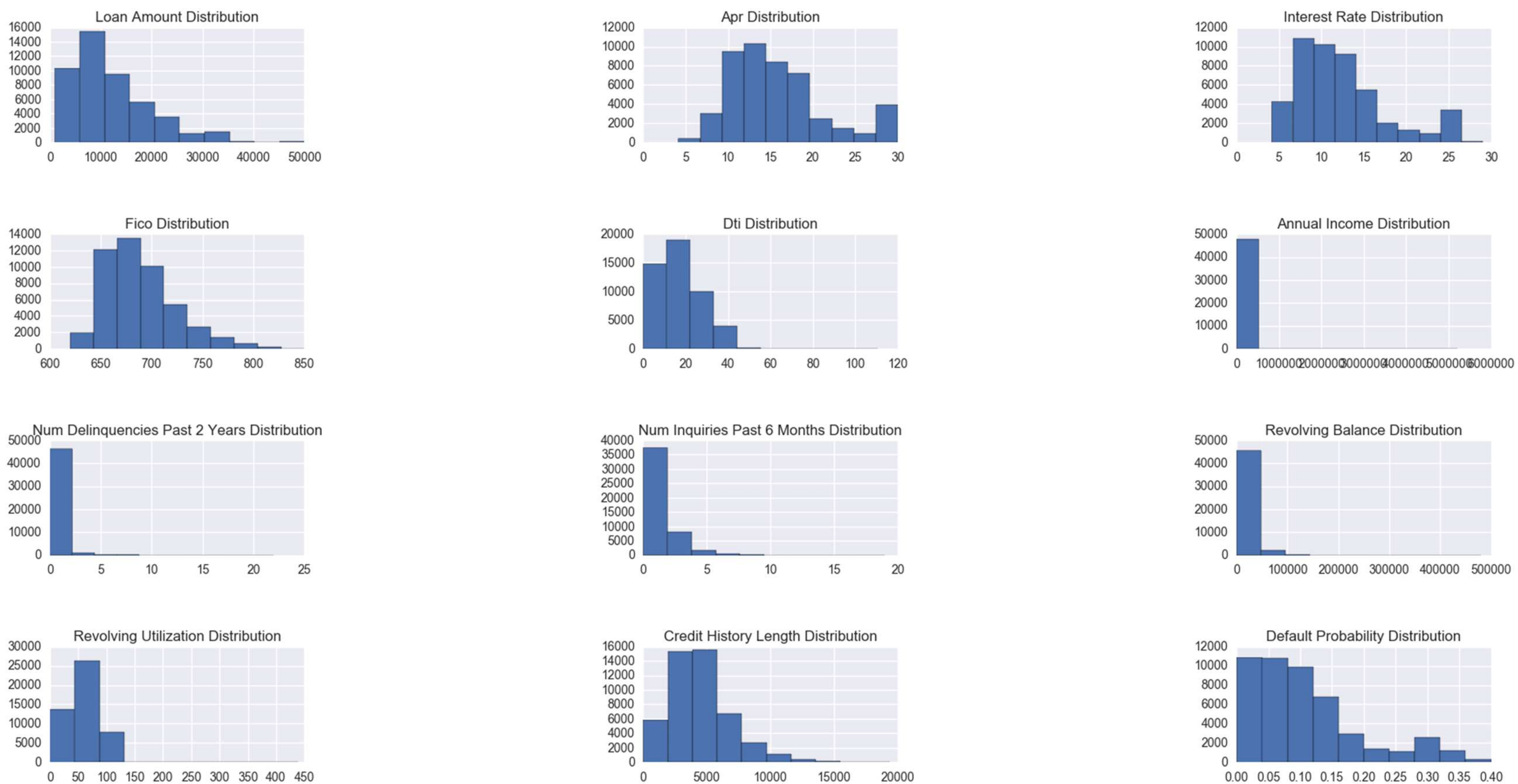
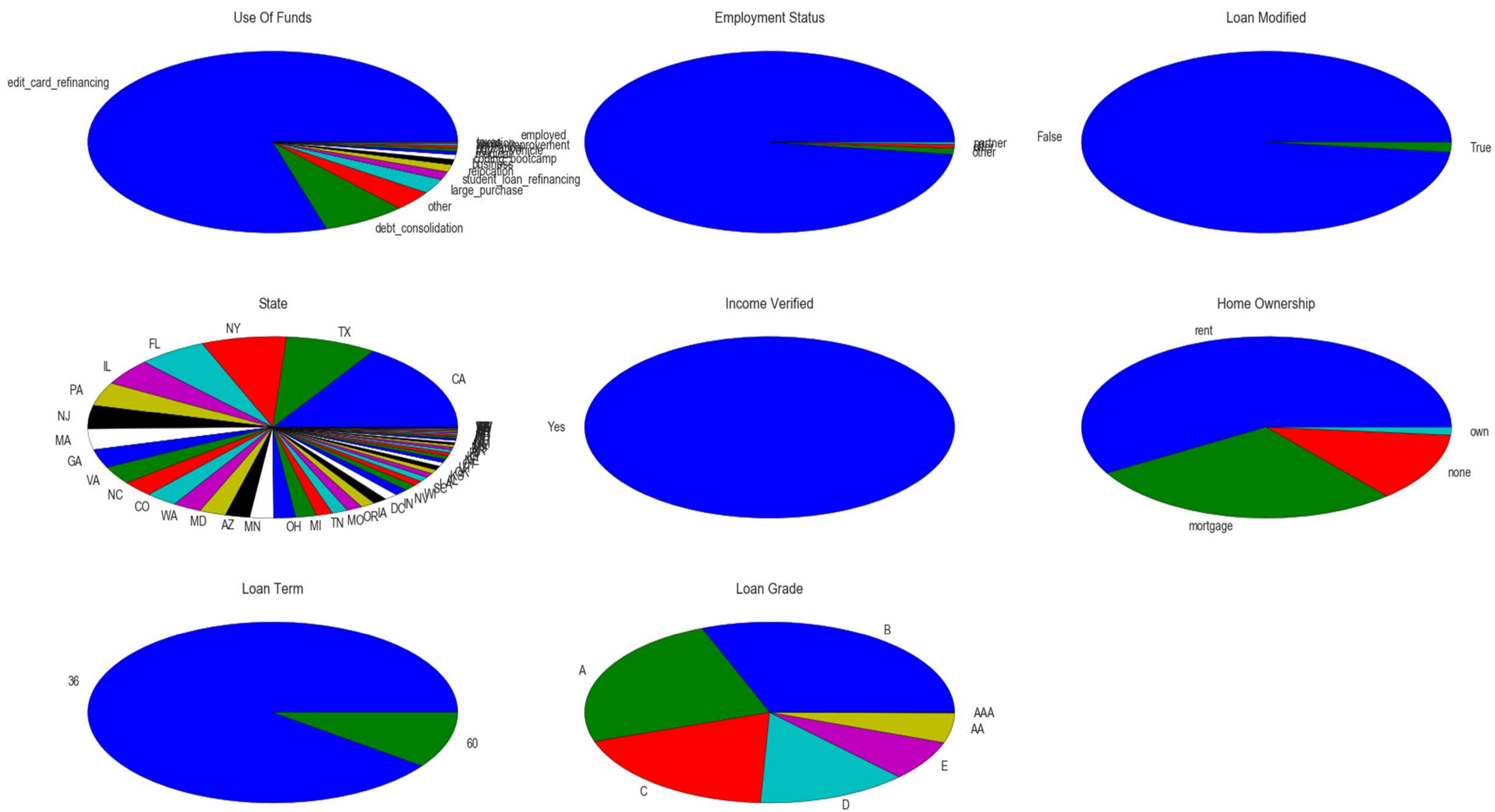
Sample Originator EDA/DD & Feature Identification

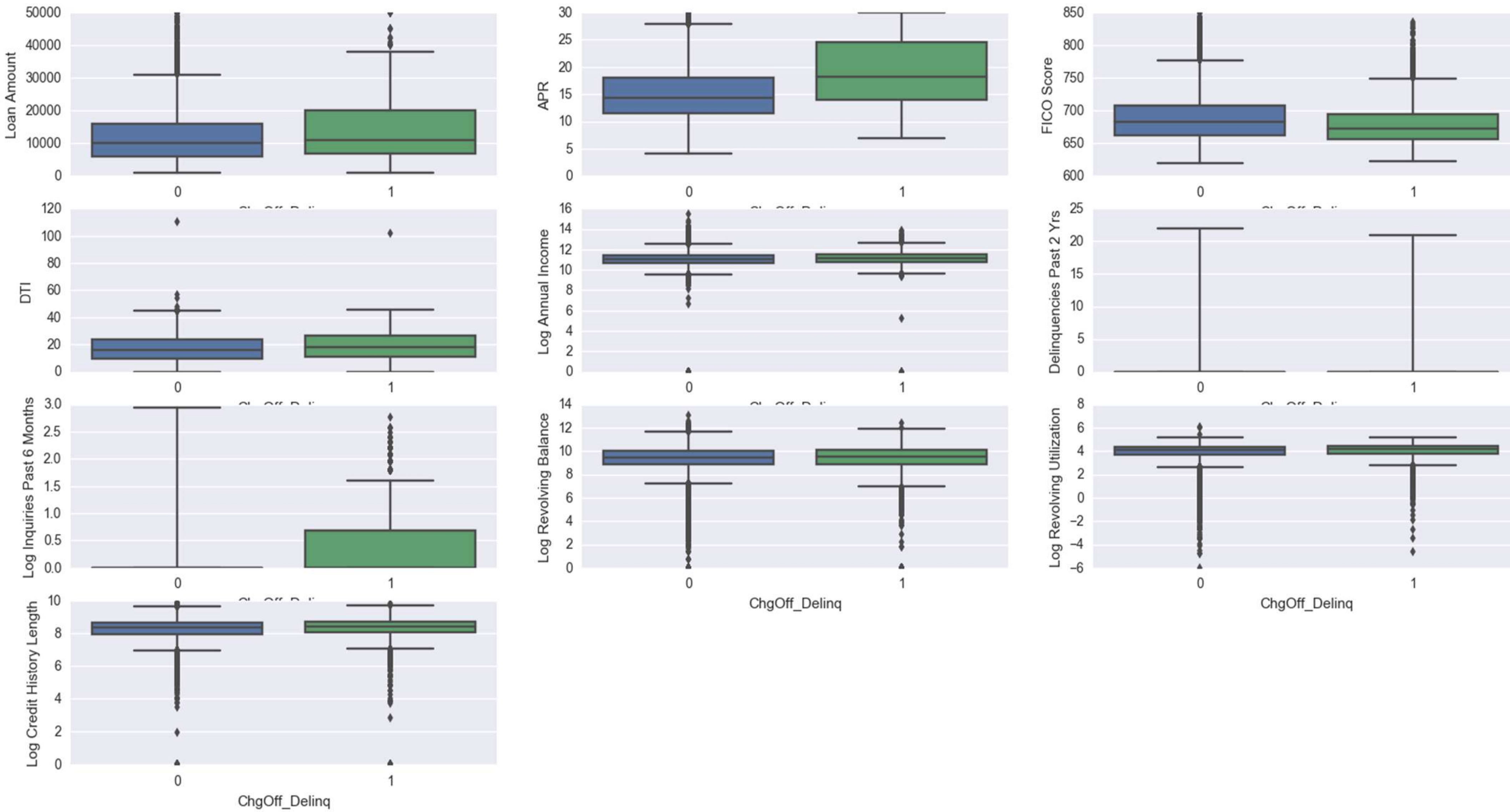
Univariate Visualization: Continuous Variables



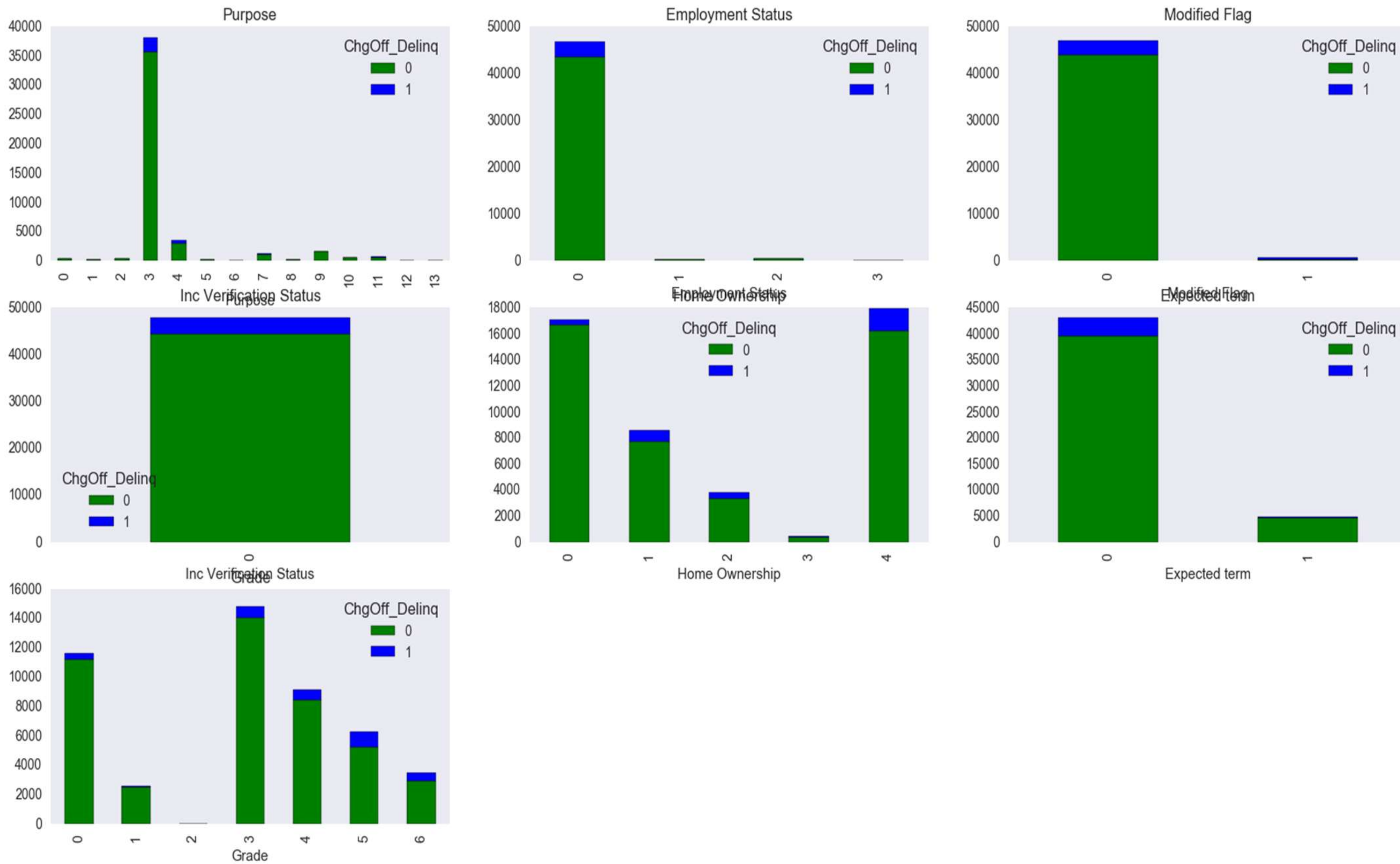
Univariate Visualization: Categorical



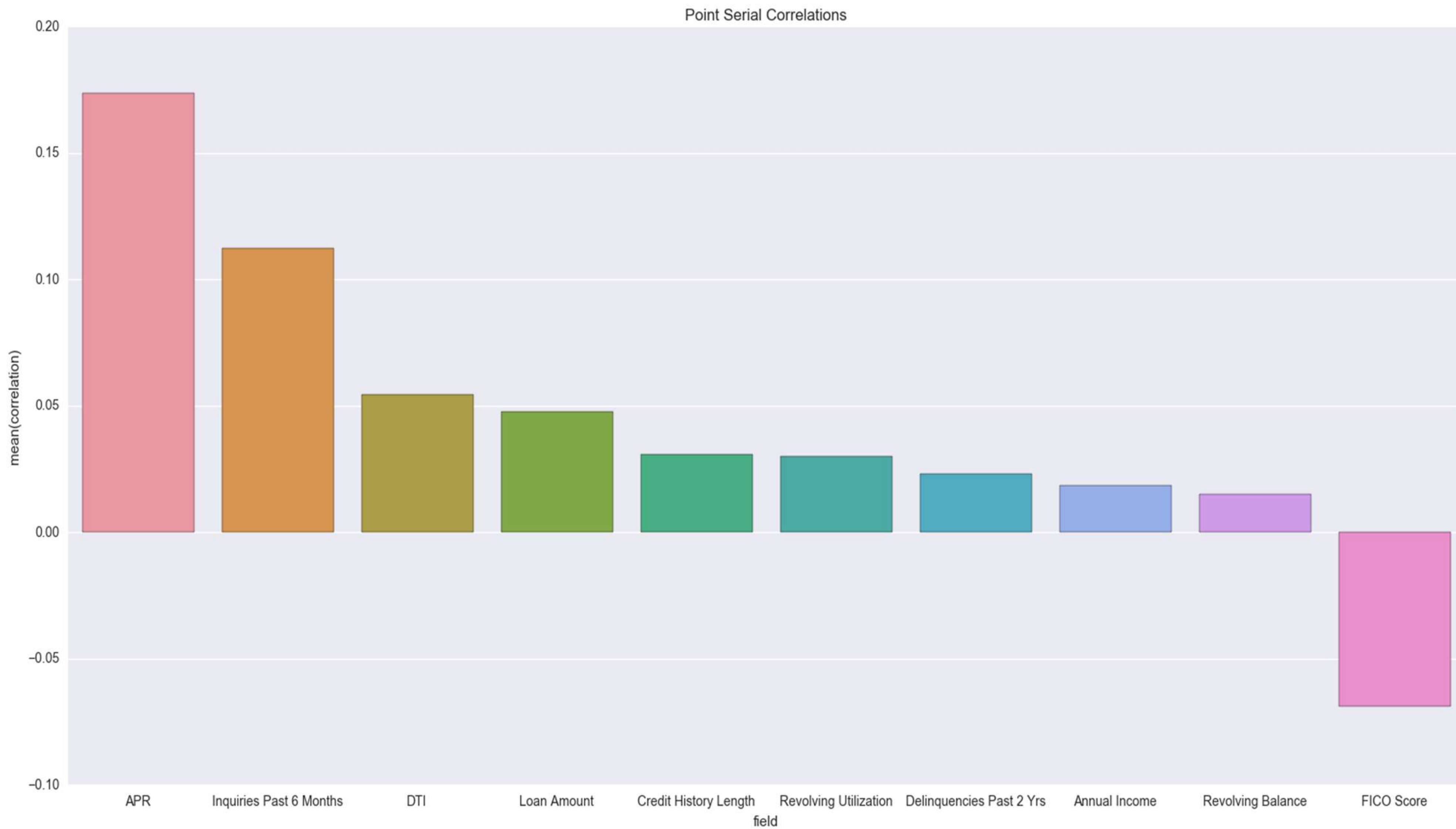
Bivariate: versus Charge-offs/Past Due: Continuous Variables



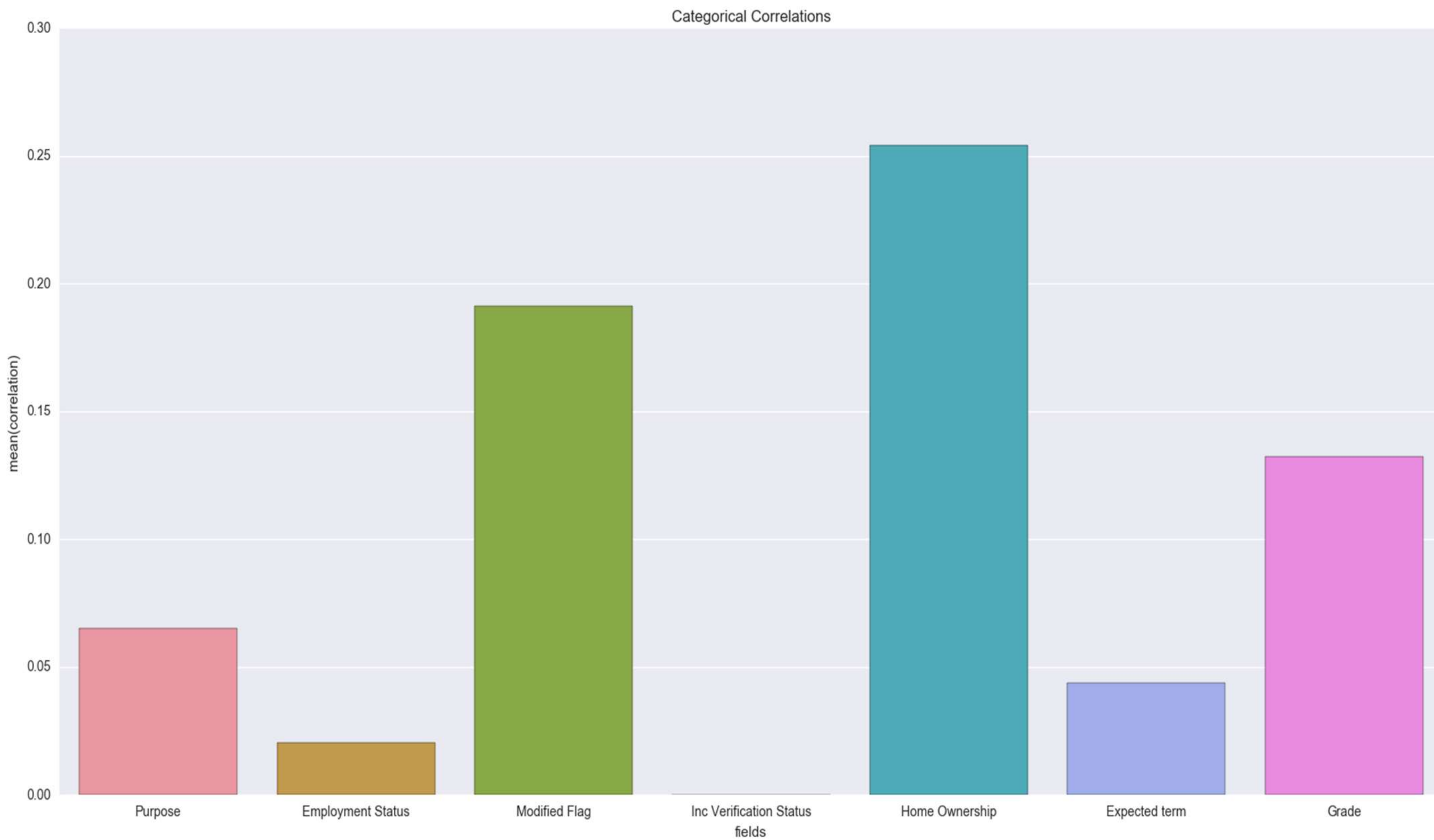
Bivariate versus Charge-offs/Past Due: Categorical Variables



Correlation with Charge-offs/Past Due: Continuous Variables



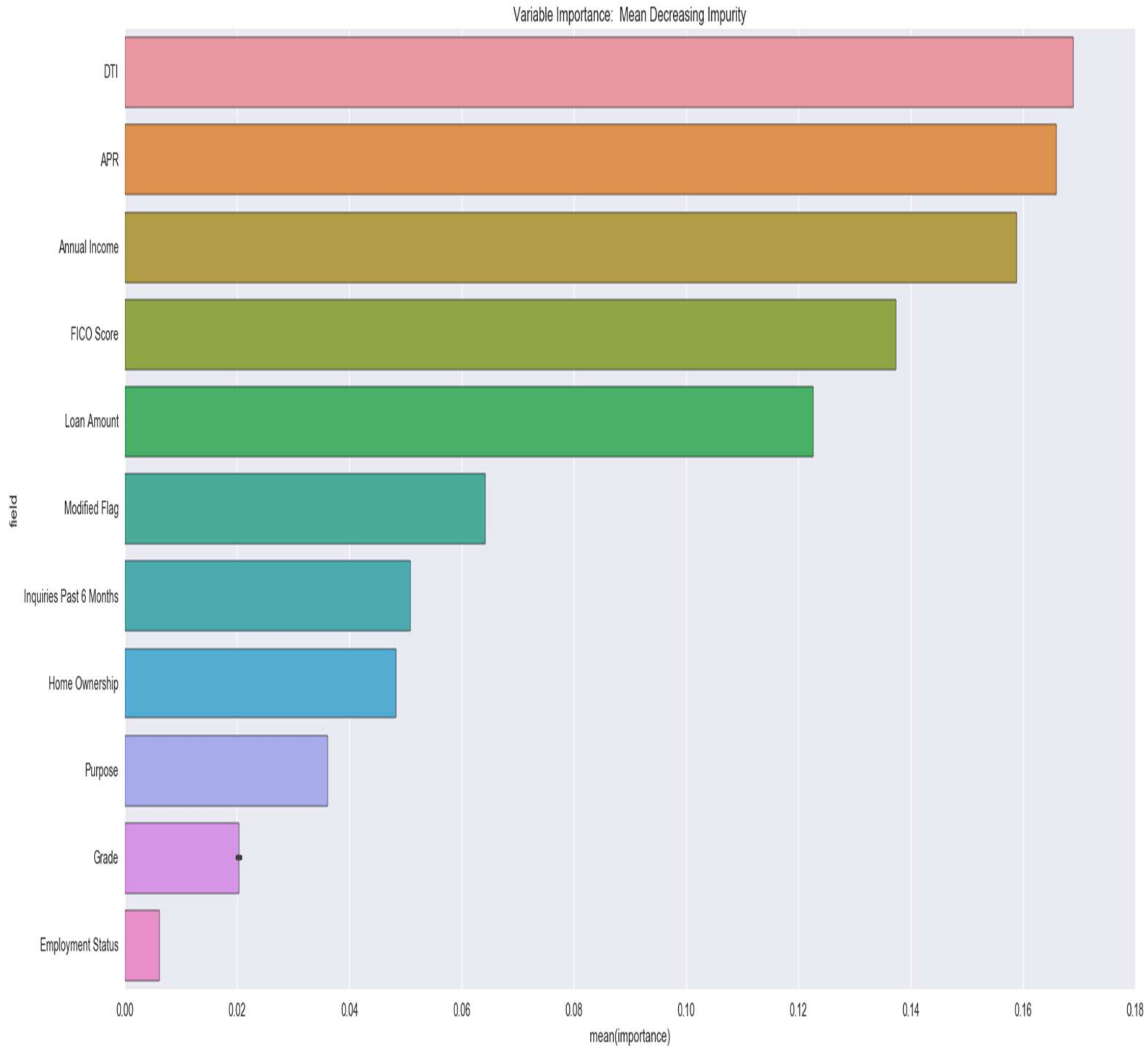
Correlation with Charge-offs/Past Due: Categorical Variables



Feature Importance

- Gini Importance scores and correlations guide which variables are calibrated to credit model.

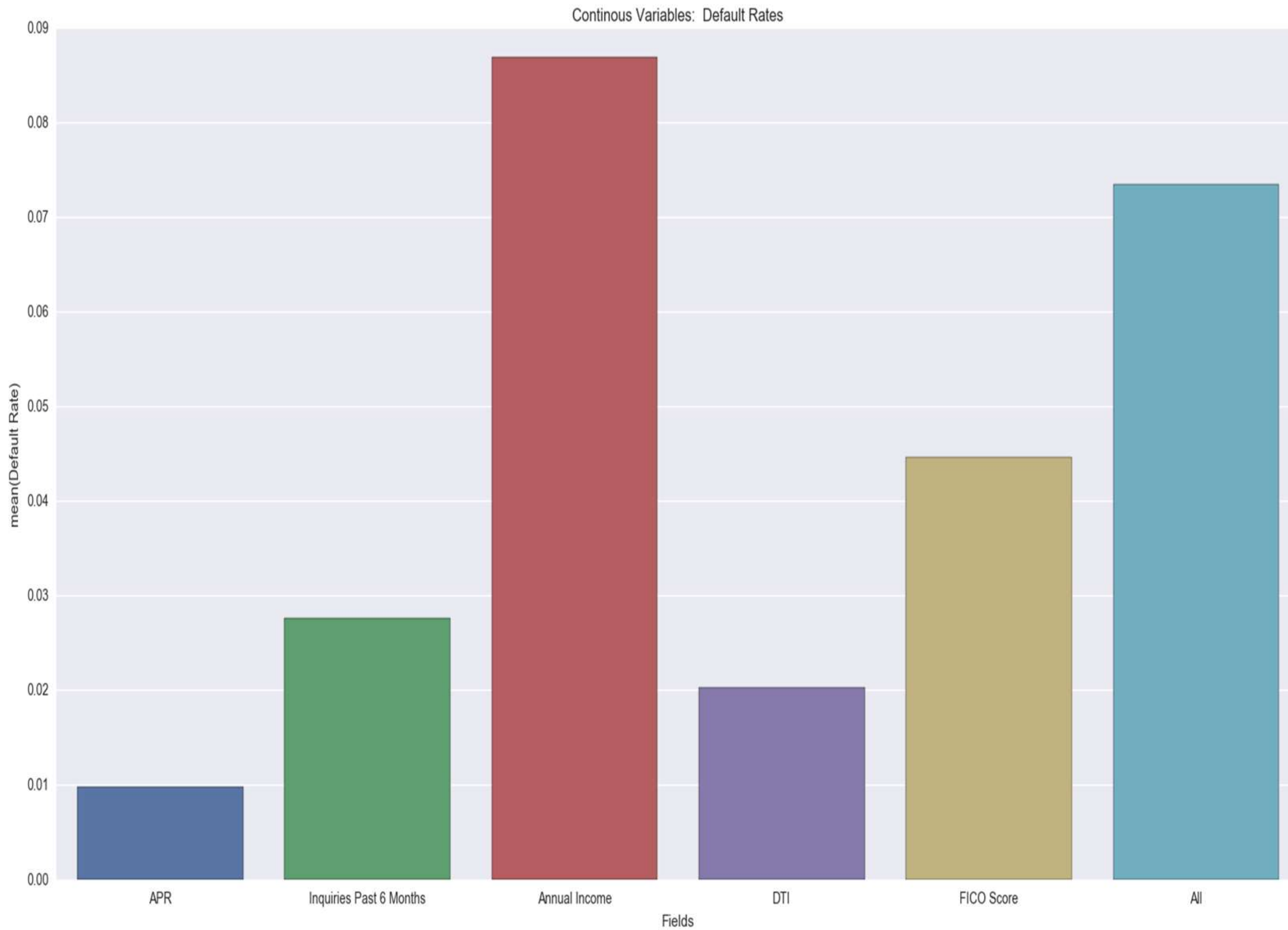
- Feature engineering identifies modifications or augmentation to data for higher explanatory power.



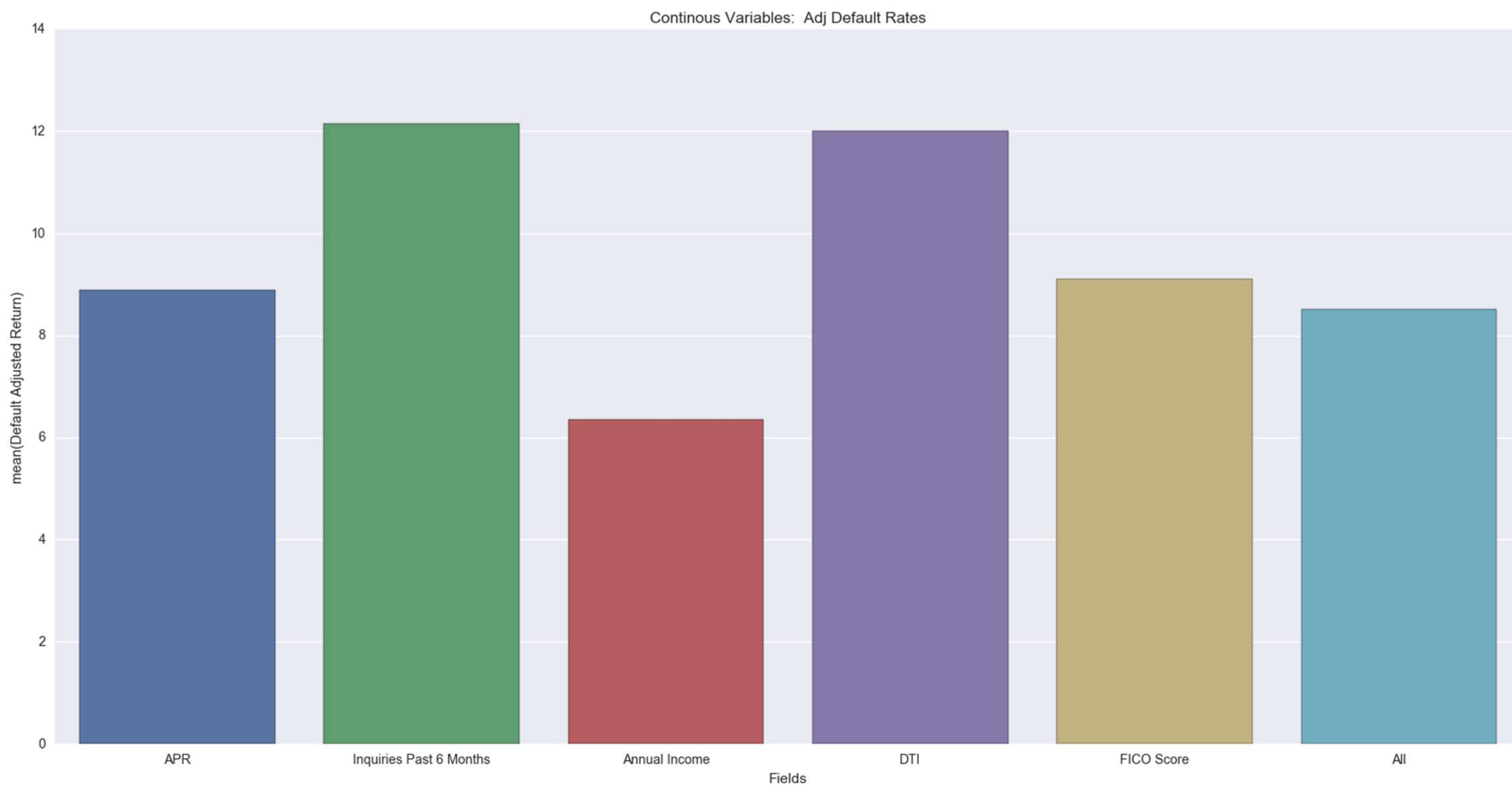
Continuous Variables : Default Rates in Top Quantile:

- We test if each default rate changes significantly for different percentiles and random cuts of data. In this case, annual income does not make the cut.

- The top quantile for annual income actually has a higher default rate than the overall set



Choosing Levels: Top-Quantile



Choosing Levels: Refinement

- Simple quantiles for each variable may limit set of acceptable loans.
- Using FICO as an example, we look at acceptable ranges that will not sacrifice returns.
- Test for multiple time periods