Mortgage Default Prediction

Chadwick

The problem:

Can we predict residential mortgage defaults based on macroeconomic factors?

The data:

CSV from: http://www.creditriskanalytics.net/datasets-private2.html

-pool of residential mortgages backing residential mortgage backed securities (RMBS)
-measured in 2016 and includes recently originated and older mortgages

The Process

Data Cleaning and EDA → Modeling → Refining and Evaluation

Technologies:

-pandas

-numpy

-mplt

-seaborn

-sklearn

Models:

-logistic

-knn

-adaboost/gdboost

-random forest

-bayes

Methods:

-cross-validation

-confusion matrix

-class balance

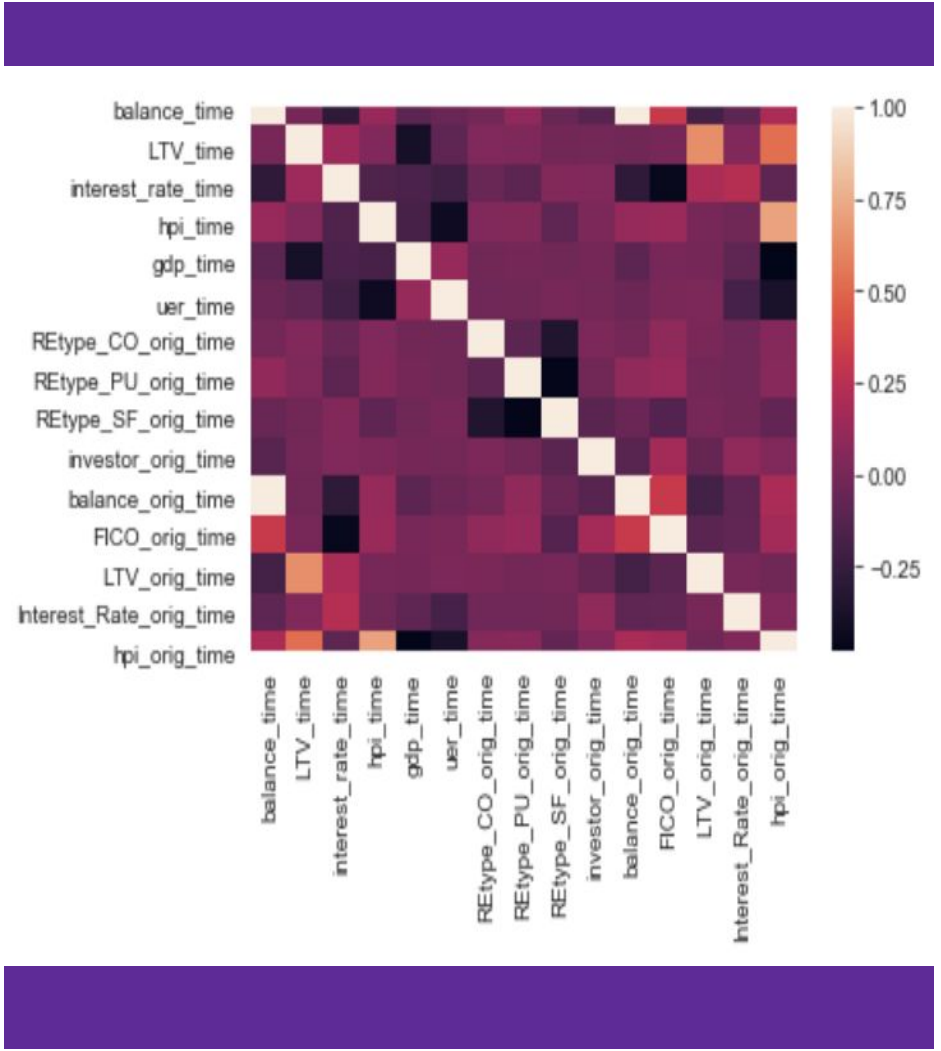
-gridsearch

Variable Selection

-macroeconomic factors and individual loan factors
-gdp, housing pricing index, interest rates, LTV ratio

Correlation

-not significant correlation except for those variables which were measured at loan origination and remeasured in 2016



Model Results

Models Run:
-Log
-KNN
-AdaBoost
-GradientBoost
-Naive Bayes
-Random Forest

Top Models:
-Log
-GradientBoost
-Random Forest

Log: AUC (.76)
TN (.81)

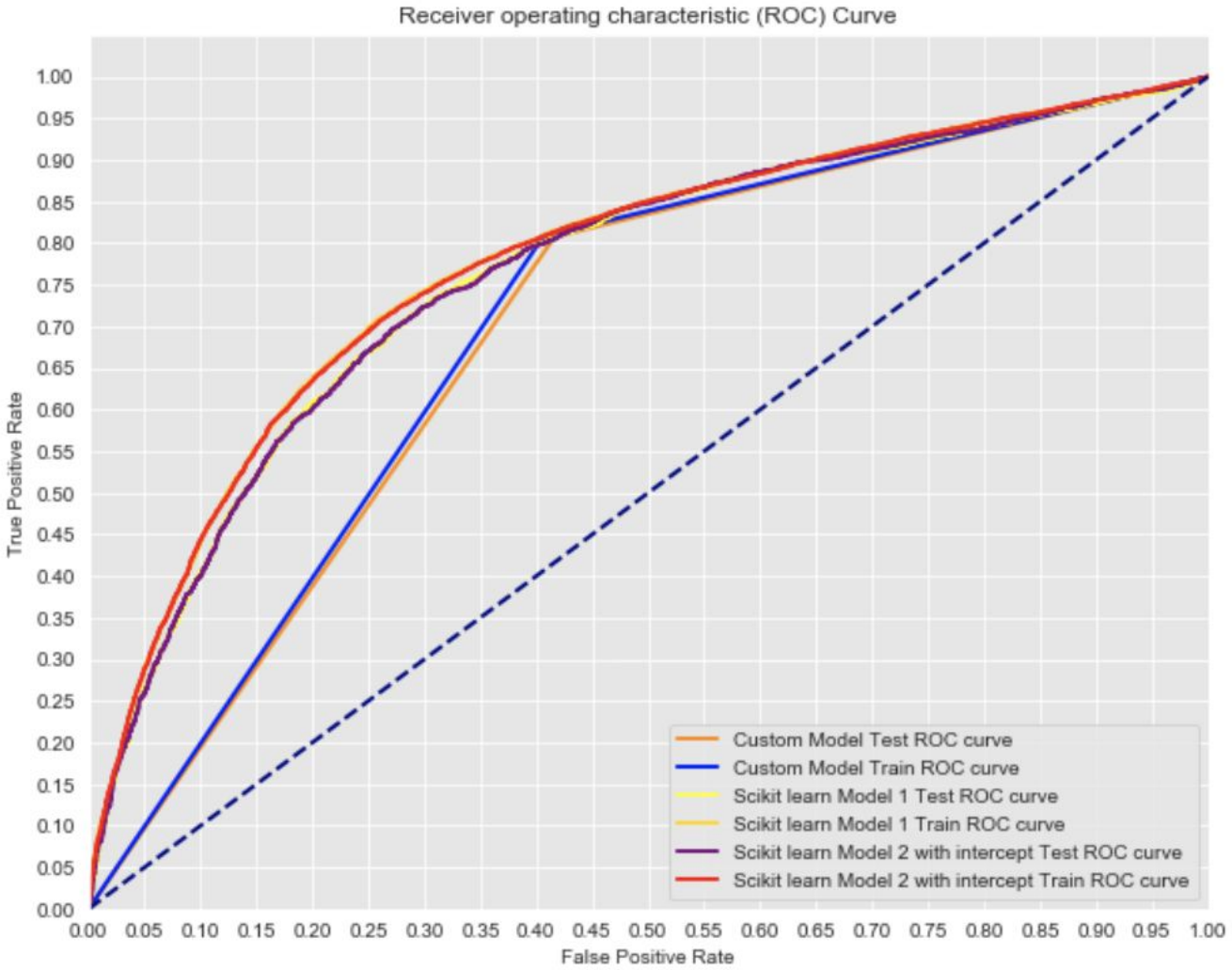
GB: TN (.78)
f1 (.8 and .65)

RF: TN (.76)
f1 (.79 and .63)

Logistic

AUC: .76

True Negative: .81*



Gradient Boost

AUC:

- .72

True Negative:

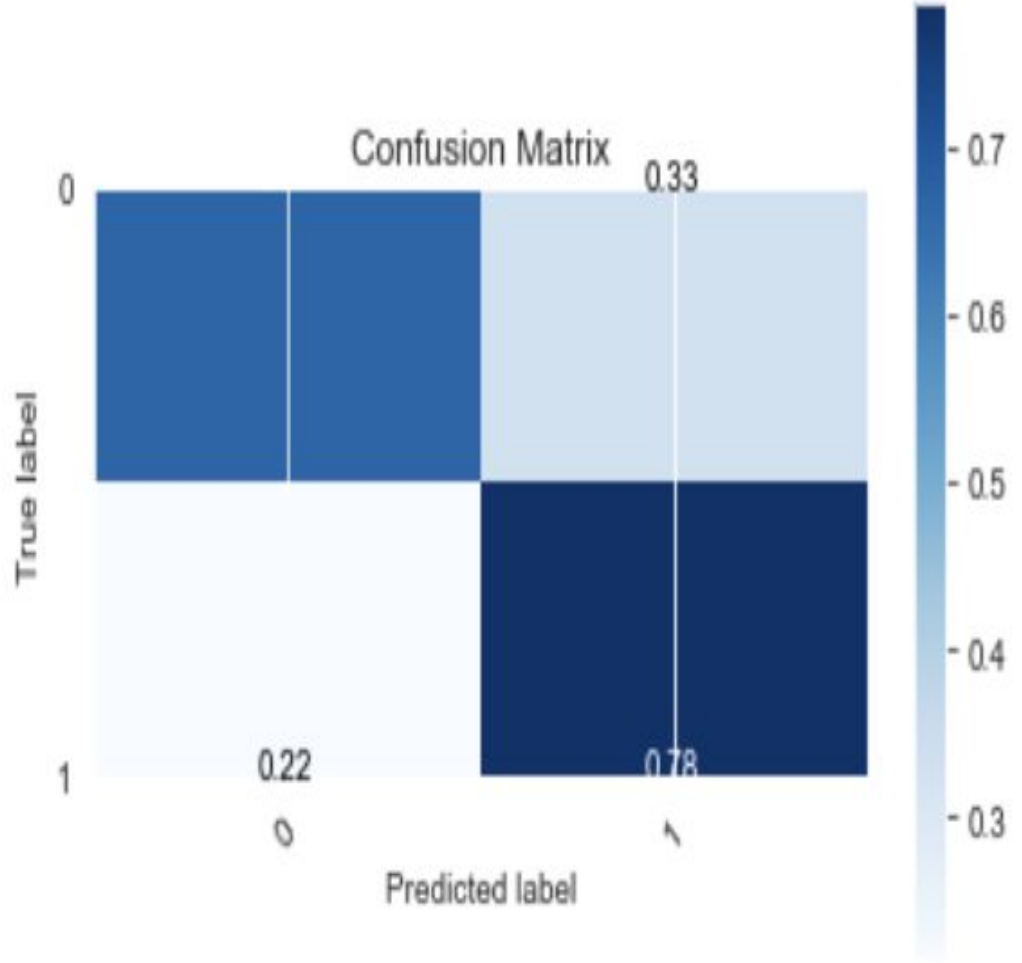
- .78

Accuracy:

- .74

F1:

- .80/.65*



Random Forest

True Negative:

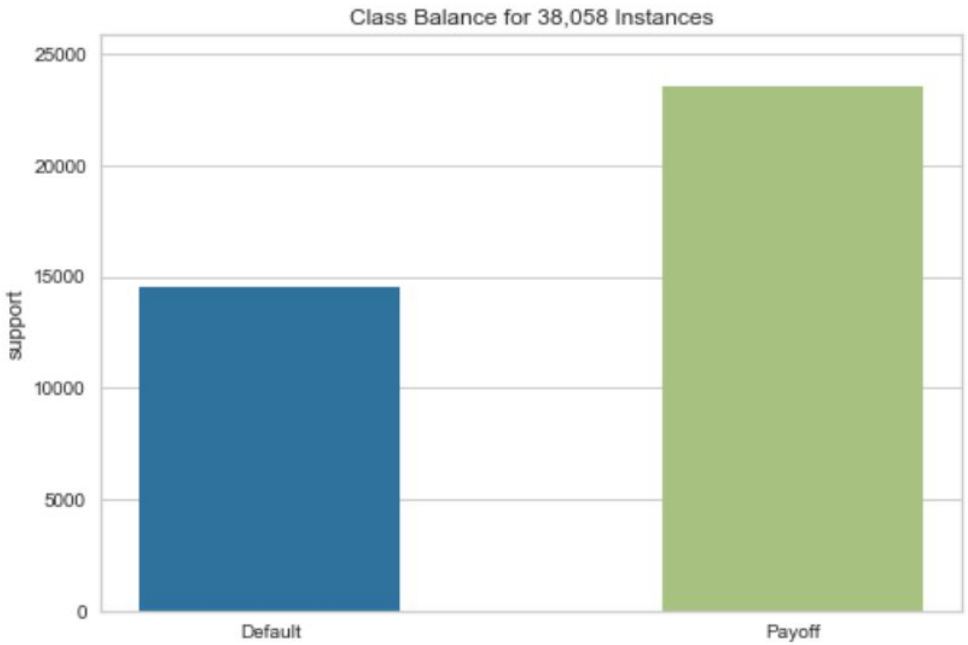
-> .77

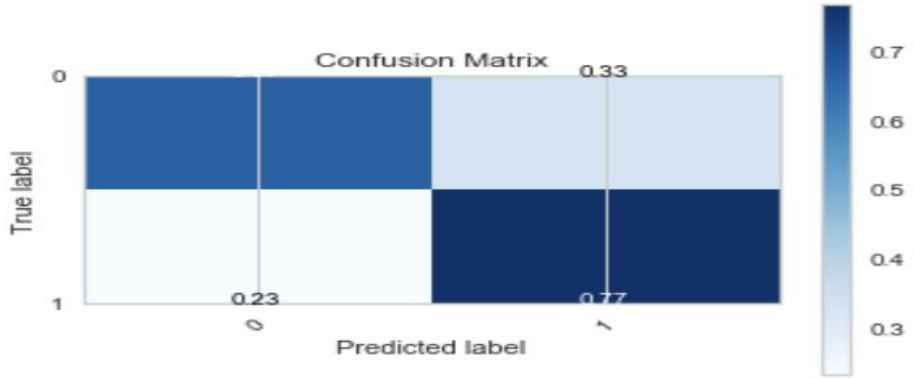
F1:

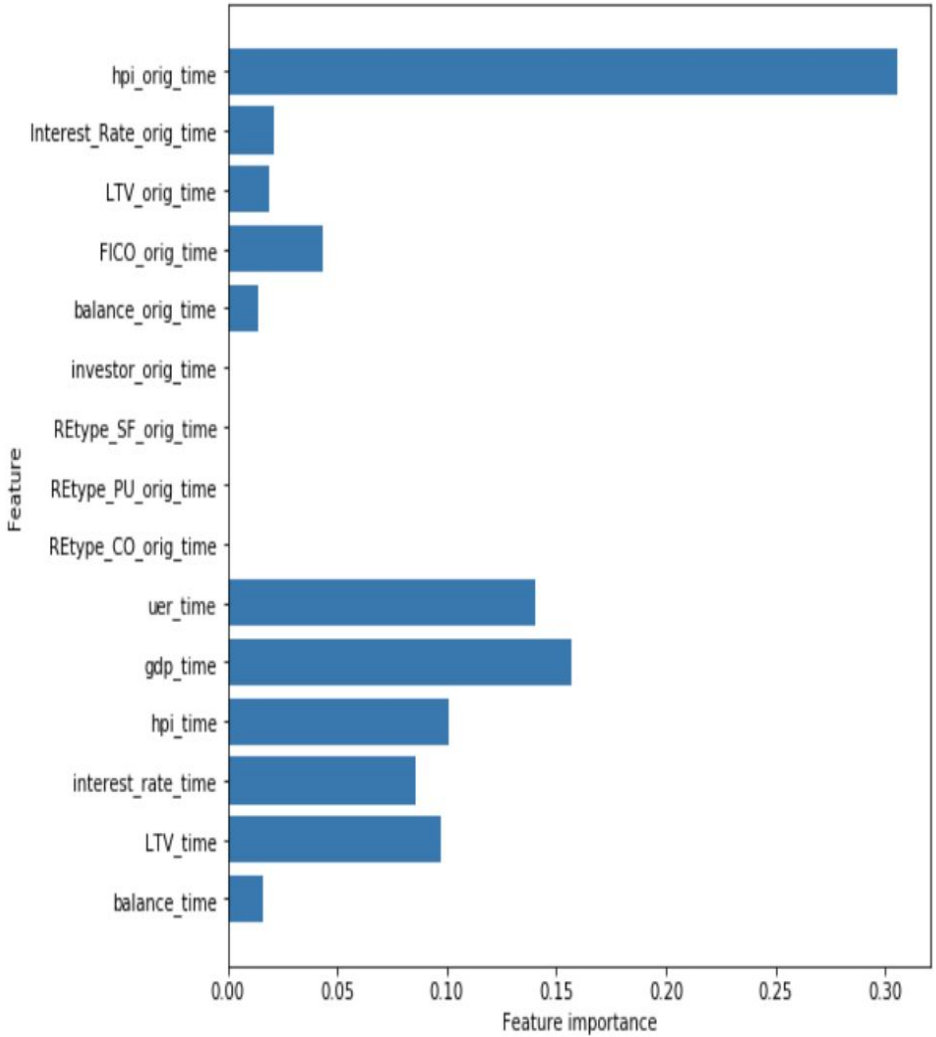
-> .79 / .63

Accuracy:

-> .73







-hpi_orig_time
    -Housing Pricing Index at time of loan origination

-gdp_time
    -GDP growth at measurement time

-interest_rate_time
    -Interest Rate at observation time

-uer_time
    -Unemployment Rate at observation time

-LTV_time
    -Loan to Value ratio at time of measurement

Evaluation and Refining

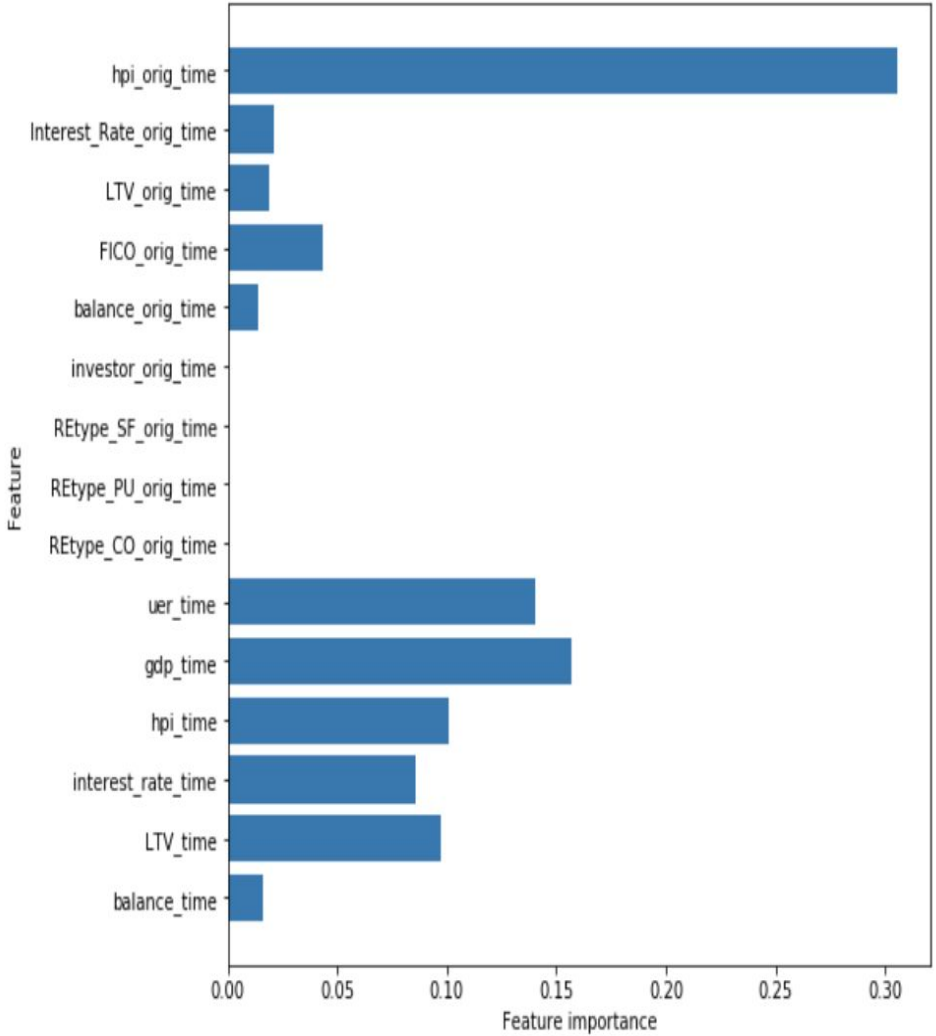
-Scores Checked:
    -F1
    -Accuracy
    -Confusion Matrix

GridSearch used to find optimal parameters

Cross Validation

Imbalance Check

Changing parameters to find optimal iteration of model



-hpi_orig_time
    -Housing Pricing Index at time of loan origination

-gdp_time
    -GDP growth at measurement time

-interest_rate_time
    -Interest Rate at observation time

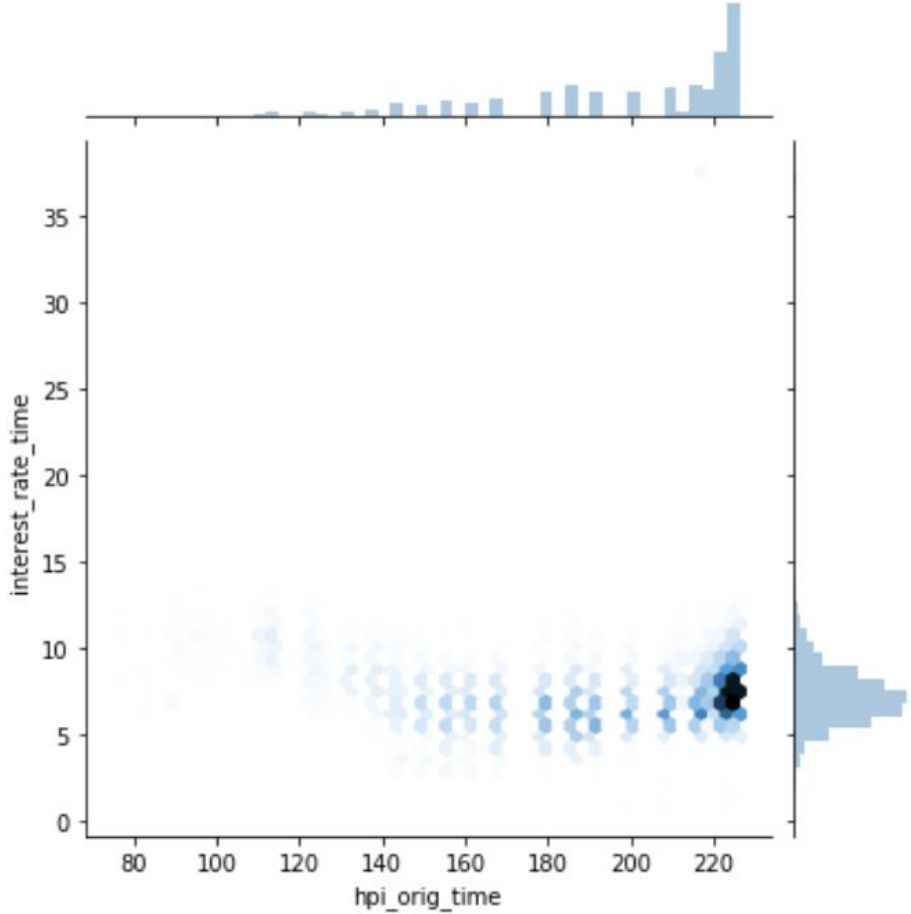
-uer_time
    -Unemployment Rate at observation time

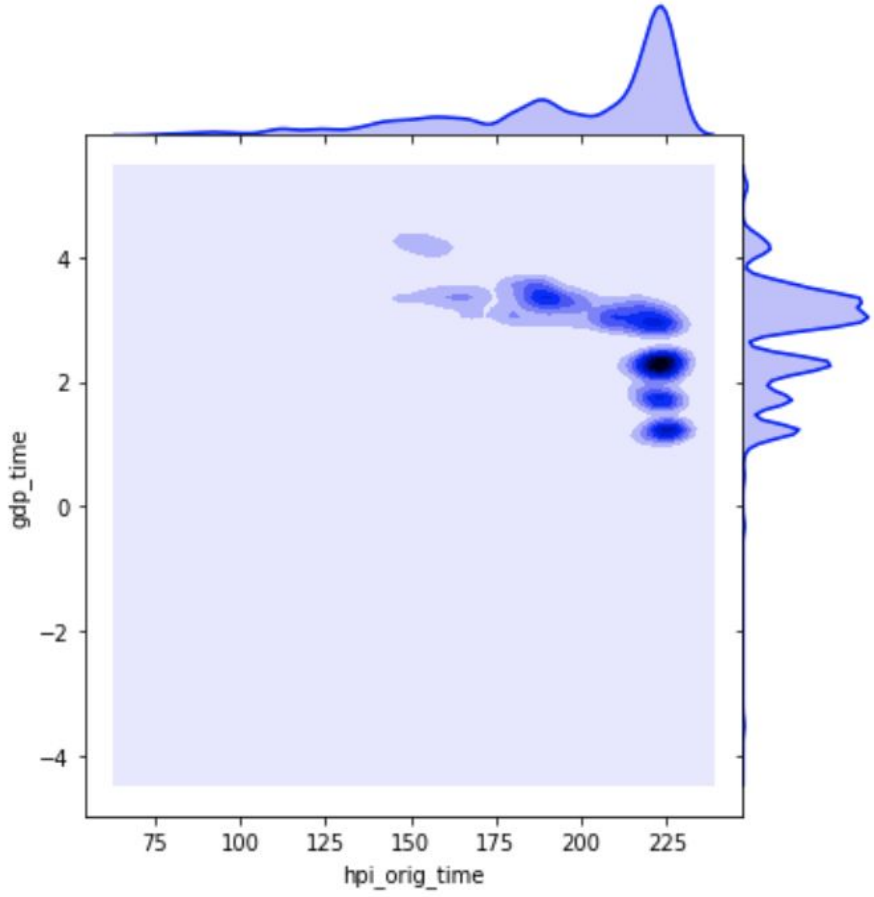
-LTV_time
    -Loan to Value ratio at time of measurement

What can feature importance tell us?

How do the Housing Pricing Index, interest rates, changes in GDP, and the unemployment rate affect loan defaults?

	hpi_orig_time	interest_rate_time	gdp_time	uer_time	LTV_time
hpi_orig_time	1.000000	-0.088708	-0.481889	-0.373782	0.526782
interest_rate_time	-0.088708	1.000000	-0.166542	-0.205196	0.138000
gdp_time	-0.481889	-0.166542	1.000000	0.117291	-0.385634
uer_time	-0.373782	-0.205196	0.117291	1.000000	-0.083424
LTV_time	0.526782	0.138000	-0.385634	-0.083424	1.000000


interest_rate_time
hpi_orig_time
0
5
10
15
20
25
30
35
80
100
120
140
160
180
200
220


gdp_time
hpi_orig_time
−4
−2
0
2
4
75
100
125
150
175
200
225

Usefulness of the Model

-used for predicting if a residential mortgage will default or not

-can be used on a large scale as an economic indicator

-can be compared to current and past trends to see how defaults relate to economic strength

-since it backs RMBS instruments, can be used as an input to predict market movements

Who can use this model?

-the borrower

-the RMBS issuer

-investors

-short side of CDS

-banks, although not as useful

Improvement to the Process

-can use more model types

-can refine models and refit with different parameters

-ability to predict when default will occur, either by calendar or by age of the mortgage

-quantify losses

-fit model to different mortgage pools, based on riskiness

-compare defaults at different points in time

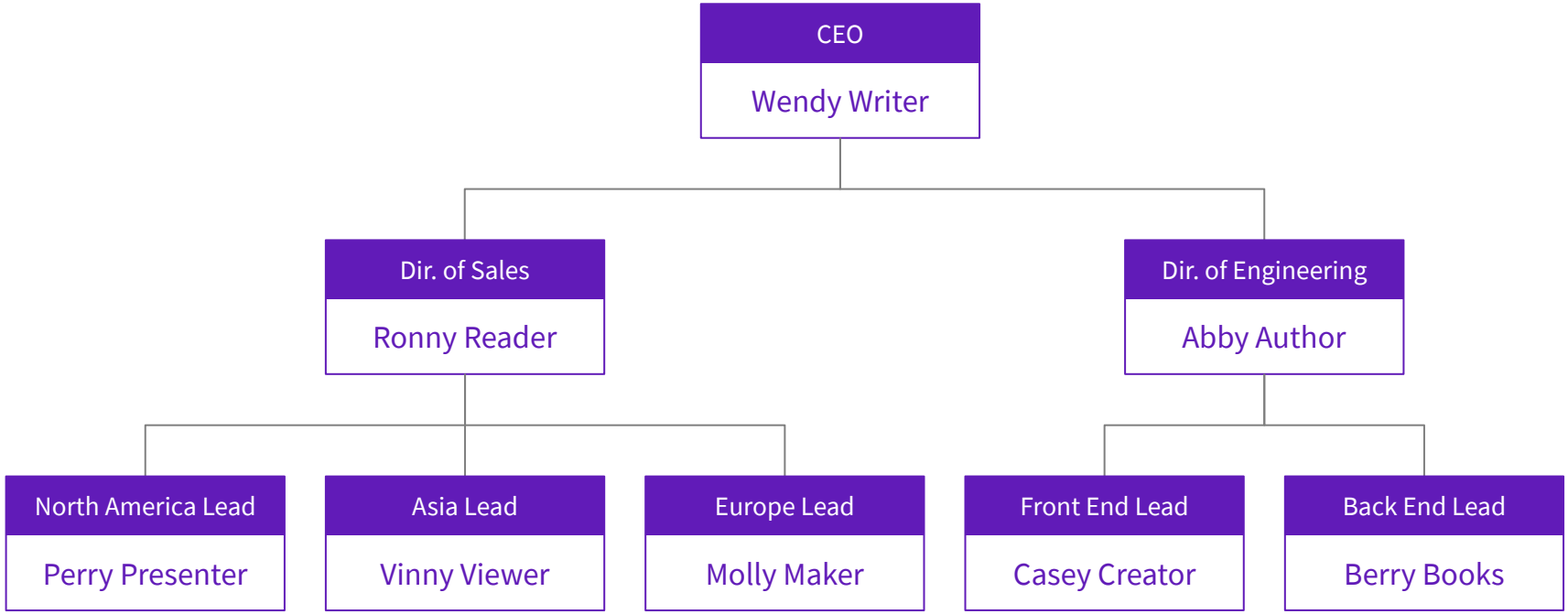
This slide intentionally left blank

brought to you by Wicked productions

Impact

XX% sales increase

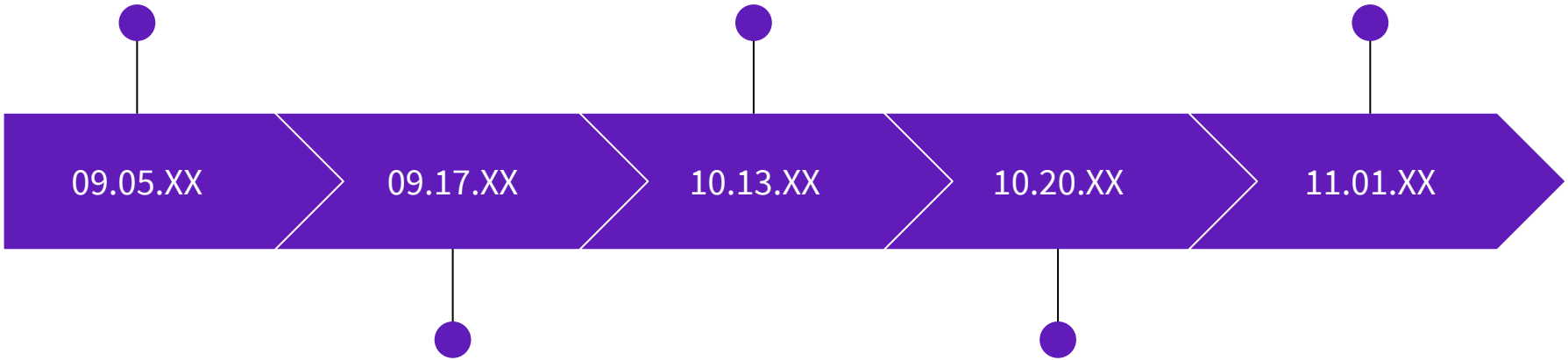
The team


CEO
Wendy Writer
Dir. of Sales
Ronny Reader
Dir. of Engineering
Abby Author
North America Lead
Perry Presenter
Asia Lead
Vinny Viewer
Europe Lead
Molly Maker
Front End Lead
Casey Creator
Back End Lead
Berry Books

Lorem ipsum dolor sit amet, consectetur adipiscing elit

Lorem ipsum dolor sit amet, consectetur adipiscing elit

Lorem ipsum dolor sit amet, consectetur adipiscing elit



Lorem ipsum dolor sit amet, consectetur adipiscing elit

Lorem ipsum dolor sit amet, consectetur adipiscing elit