

PREDICTING BUSINESS SURVIVAL RATE IN LONDON BOROUGHHS

DATA SCIENCE PROJECT

OVERVIEW – VALUE OF PREDICTION

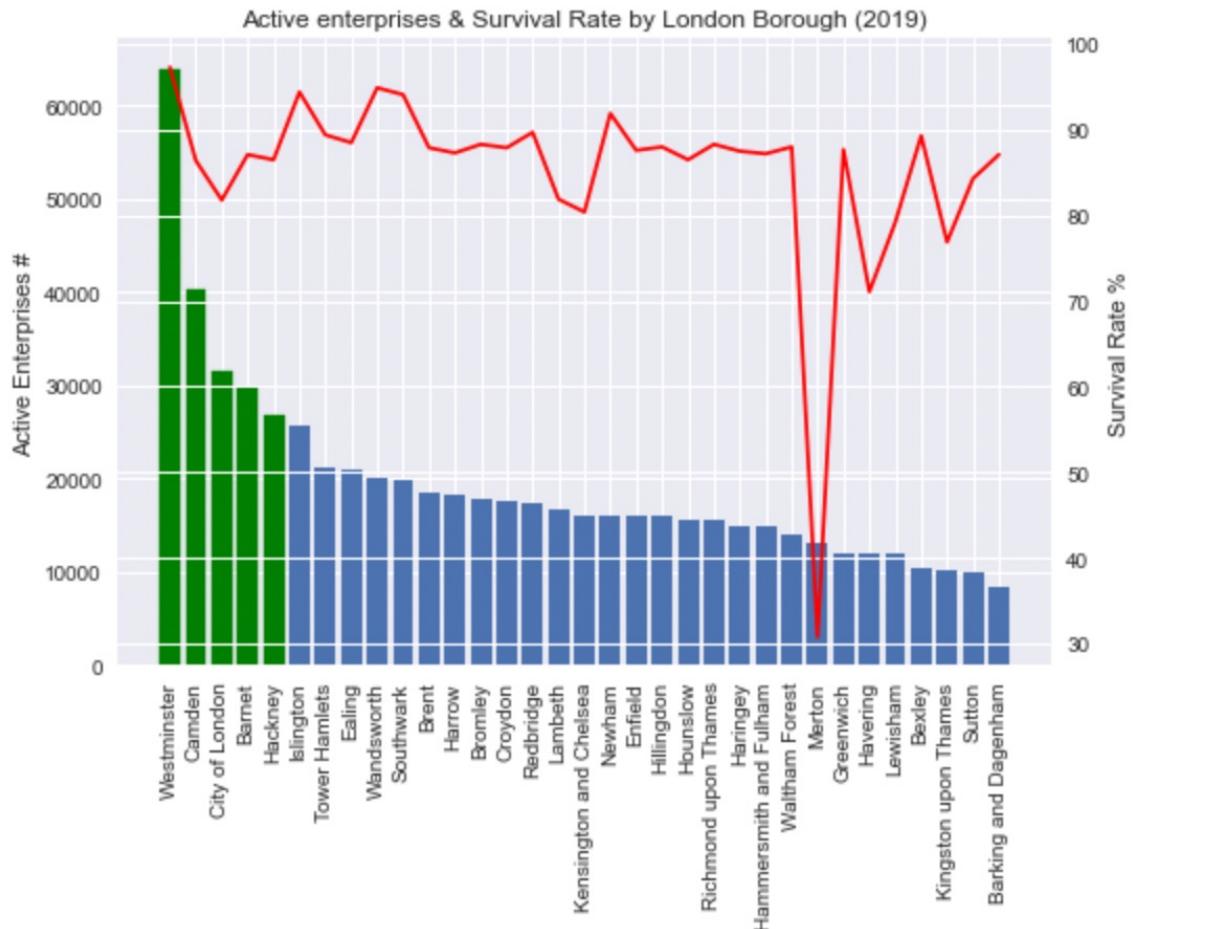
- Newer businesses emerges every year – financing them will need proper due diligence.
- Predicting business survival rate allows for better decision making on whether to finance the new company or not.
 - Provides expected success rate based on historical performance of business category.
 - Plans for capital/resource management.
- Government entities may also be interested in the business trend.

DATA ACQUISITION & TRANSFORMATION

- Three main datasets used for exploratory analysis:
 - London Business Rate - period: 2012-2019; source: Wikipedia, UK GOV Data.
 - London Crime Data - period: 2018-2020; source: Kaggle.
 - London Venue - period: 2021; source: Foursquare.
- One final dataset used for modelling (joined). Due to limited information, there are only 66 rows & 135 features
- Note: only 2-years worth of data can be overlapped.

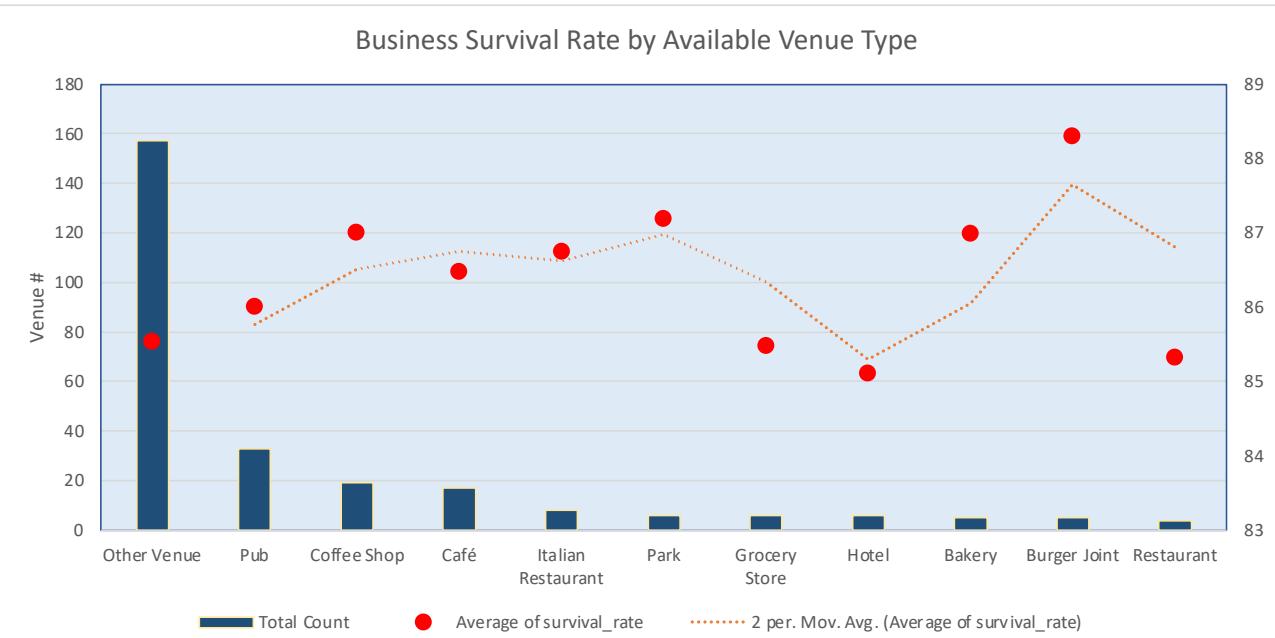
USING BUSINESS SURVIVAL RATE AS TARGET VARIABLE

- TARGET VARIABLE AVAILABLE ON BOROUGH LEVEL.
- INVERSE OF DEATH RATE.



SURVIVAL RATE
TENDS TO BE
HIGHER FOR FOOD
& BEVERAGES
RELATED CATEGORY

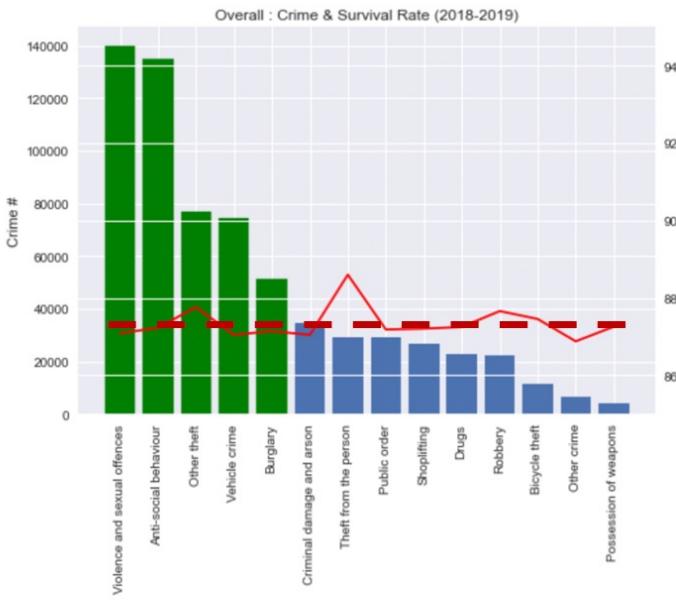
OTHER VENUES ARE
CONSOLIDATED SUM OF REST OF
VENUE CATEGORIES



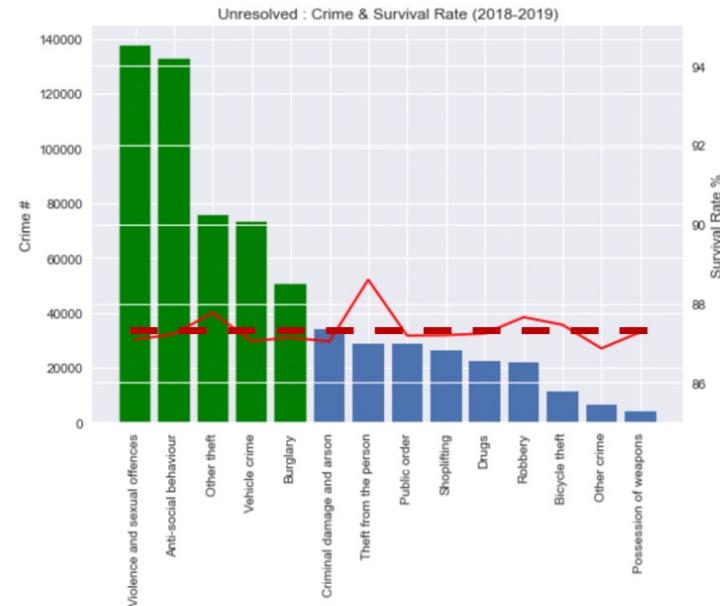
SURVIVAL RATE IS
BETTER IN AREA
WHERE CRIMES ARE
UNRESOLVED

NO SIGNIFICANT
DIFFERENTIATION
OBSERVED WHEN IT
COMES TO CRIME TYPE.

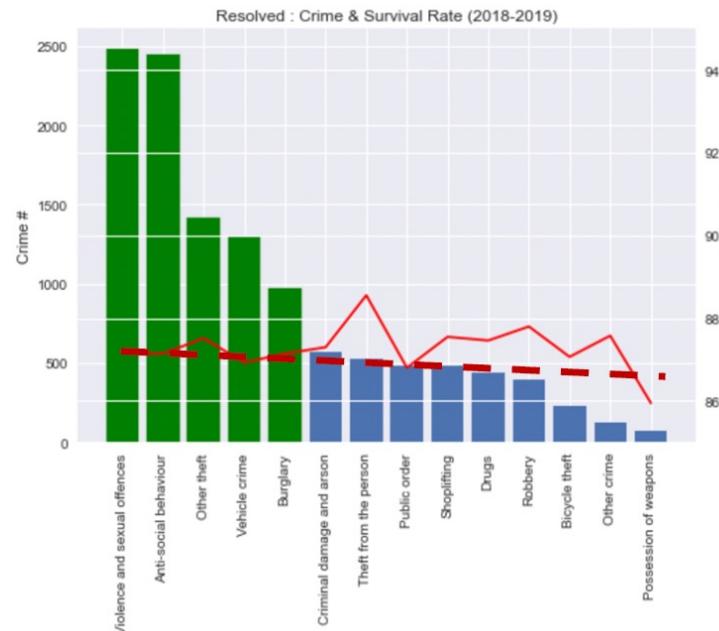
OVERALL



UNRESOLVED



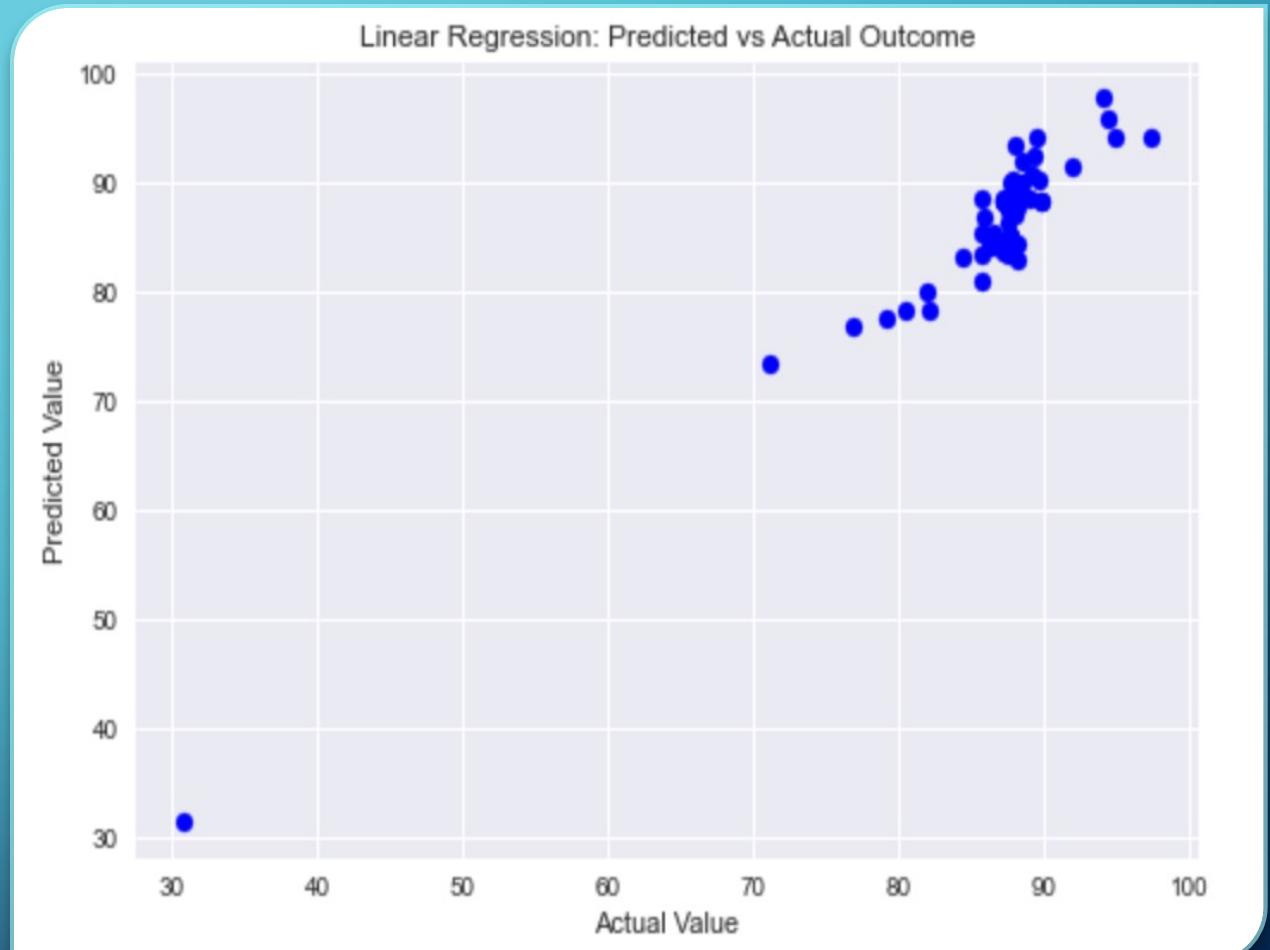
RESOLVED



REGRESSION MODEL PERFORMANCE: MULTIPLE LINEAR REGRESSION

EVALUATION METRIC:

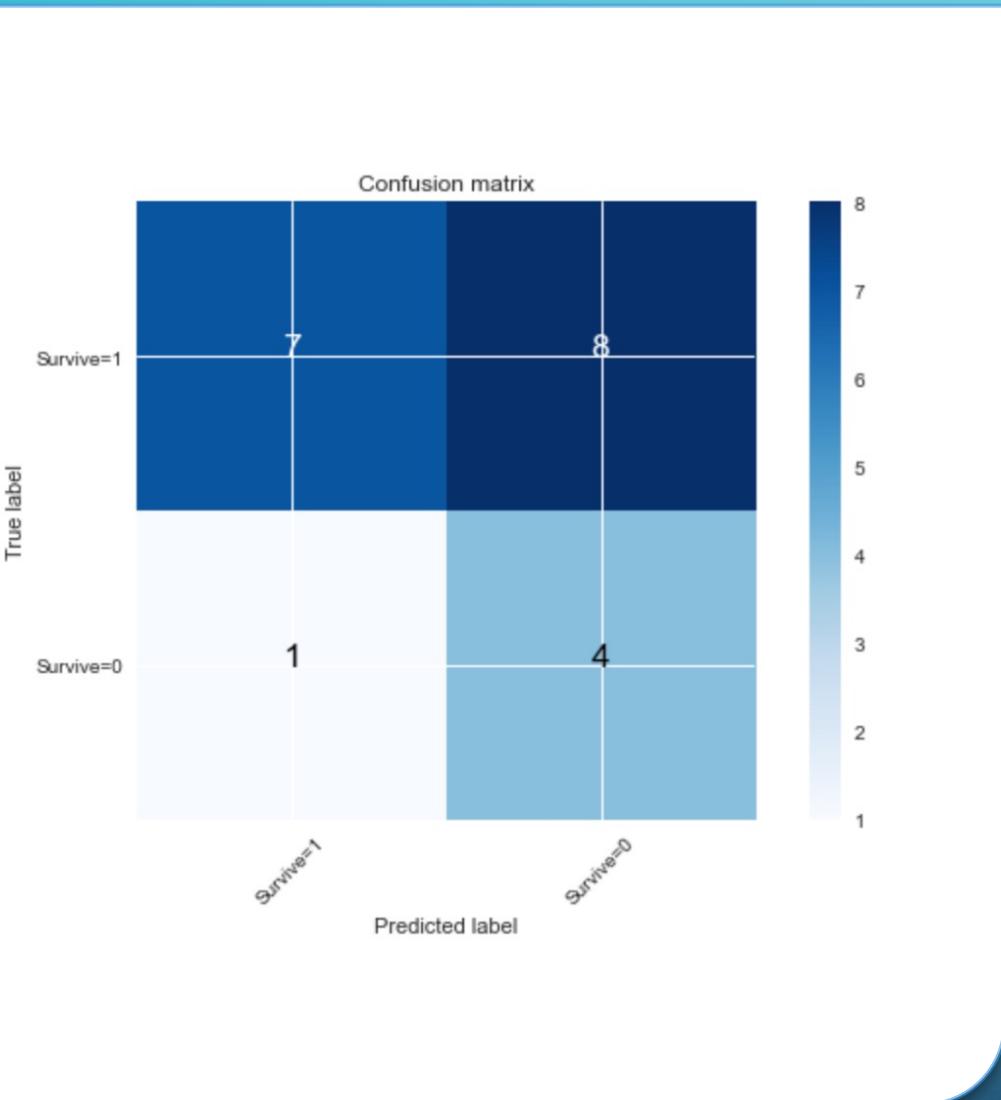
- Residual Sum of Squares : 138.3
- Mean Square Error: 0.76
- R-Squared: 0.759



CLASSIFICATION MODEL PERFORMANCE: LOGISTIC REGRESSION

EVALUATION METRIC:

- Log Loss: 0.684
- Jaccard Index: 0.308



CONCLUSION

- Regression model fares better than classification model.
- Prediction can be further improved with availability of data – mainly business survival rate on monthly basis instead of yearly.
- Follow-up / Future projects include:
 - Deep-dive into venue category using hierarchical clustering.
 - Further analysis and investigation into crime data – application of clustering before predicting survival rate