# Tenant Churn

… **Property Management System**

Kathirvel Kumararaja
Mar 13 , 2020

# Kathirvel Kumararaja

## Vice President - Operations, DevJee Inc.



## Credentials:

- Data Scientist with 20 + years of IT consulting experience.
- Holds an Engineering degree and and MBA.
- Experience in handling data analysis for multi-billion dollar capital development projects
  - Burj Khalifa tower in Dubai
  - Pentagon Renovation project, Arlington Virginia.

# Agenda

- Problem Statement
- Bird's eye view
- Findings
  - Data overview - EDA
  - Feature Engineering
  - Model evaluation
- Conclusions and recommendations

# Problem Statement

## Business Objective

To help Property Managers run their business effectively by predicting if a Tenant is Churned or a Non-Churned tenant.

## Challenges

With the thin margins and lots of pro-tenant regulations in place Property Management companies need to have a clear insights on the expected vacancy (tenant churn) to plan their cash flow

## Desired Outcome

- Use date from Property Management System to engineer data
- Train the models to predict Churned Tenant.
- Evaluate the model using **Accuracy** as the criteria

# Birds eye view

| | |
|---|---|
| **Qualitative data  1** | • Real Estate Data sets<br>   ■ Resident history from Property Management System |
| **Qualitative data  2** | • Time scale<br>   ○ Five years of Resident data |
| **Qualitative data  3** | • Baseline Score<br>   ○ 0.67 |
| **Qualitative data  4** | • Models / Classifiers Explored<br>   ○ Logistic Reg, Random Frst, Extra trees. |

# Data Overview - EDA

# EDA - Rent fields
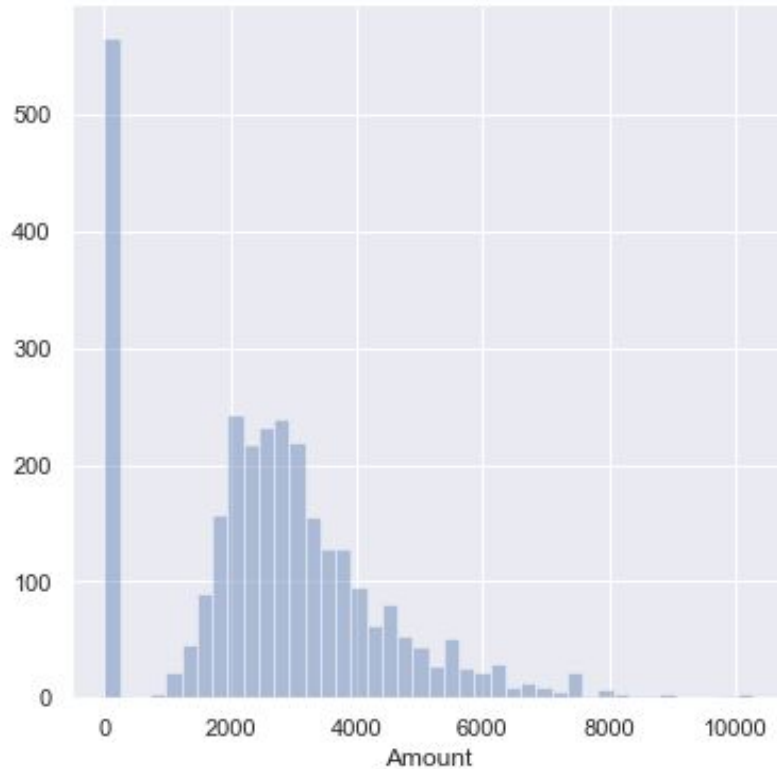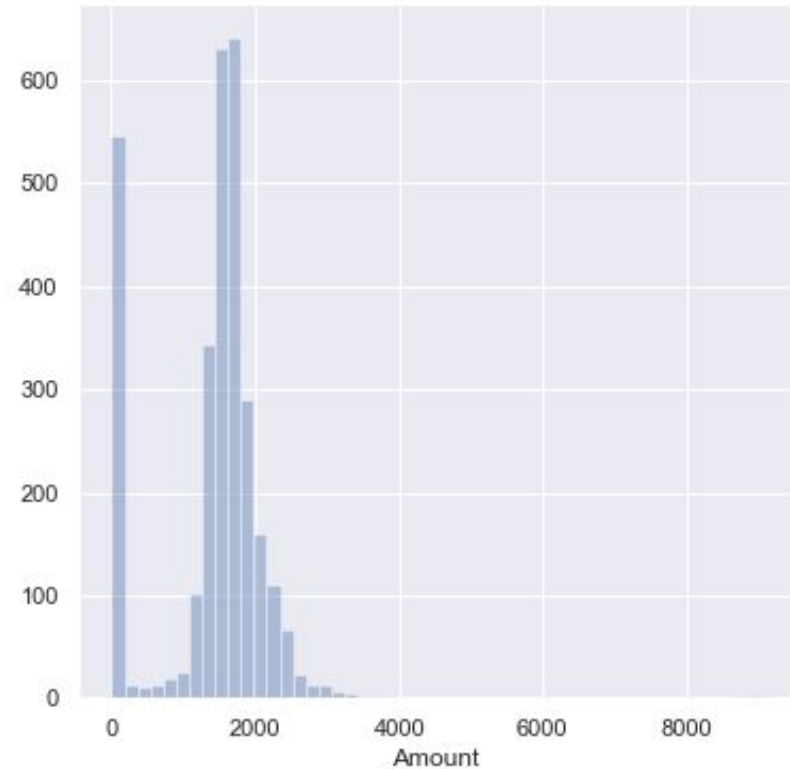
# EDA - Move-Out Monthwise



Move Out traffic - Monthwise
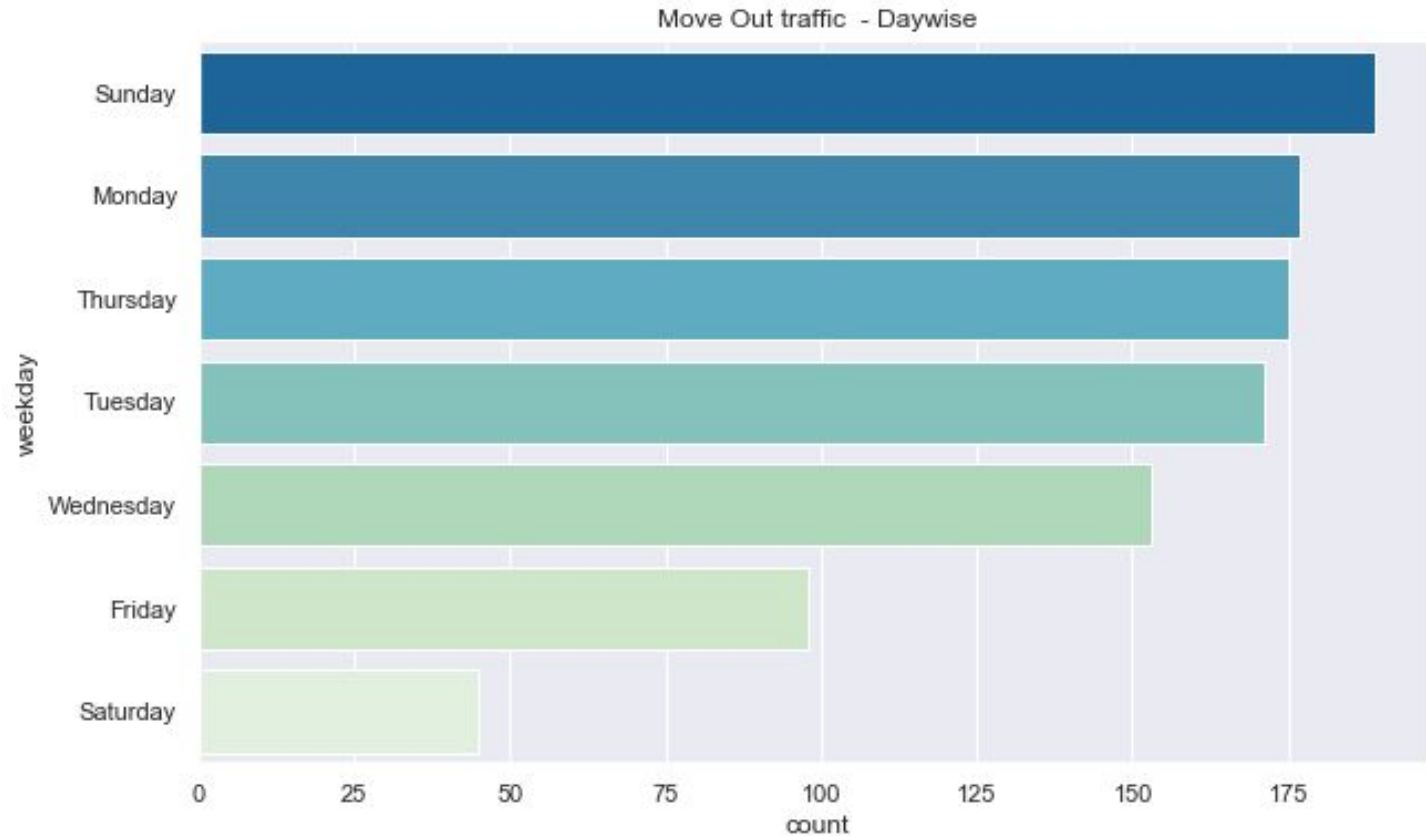
# EDA - Move-Out Monthwise



Move Out traffic - Daywise

T
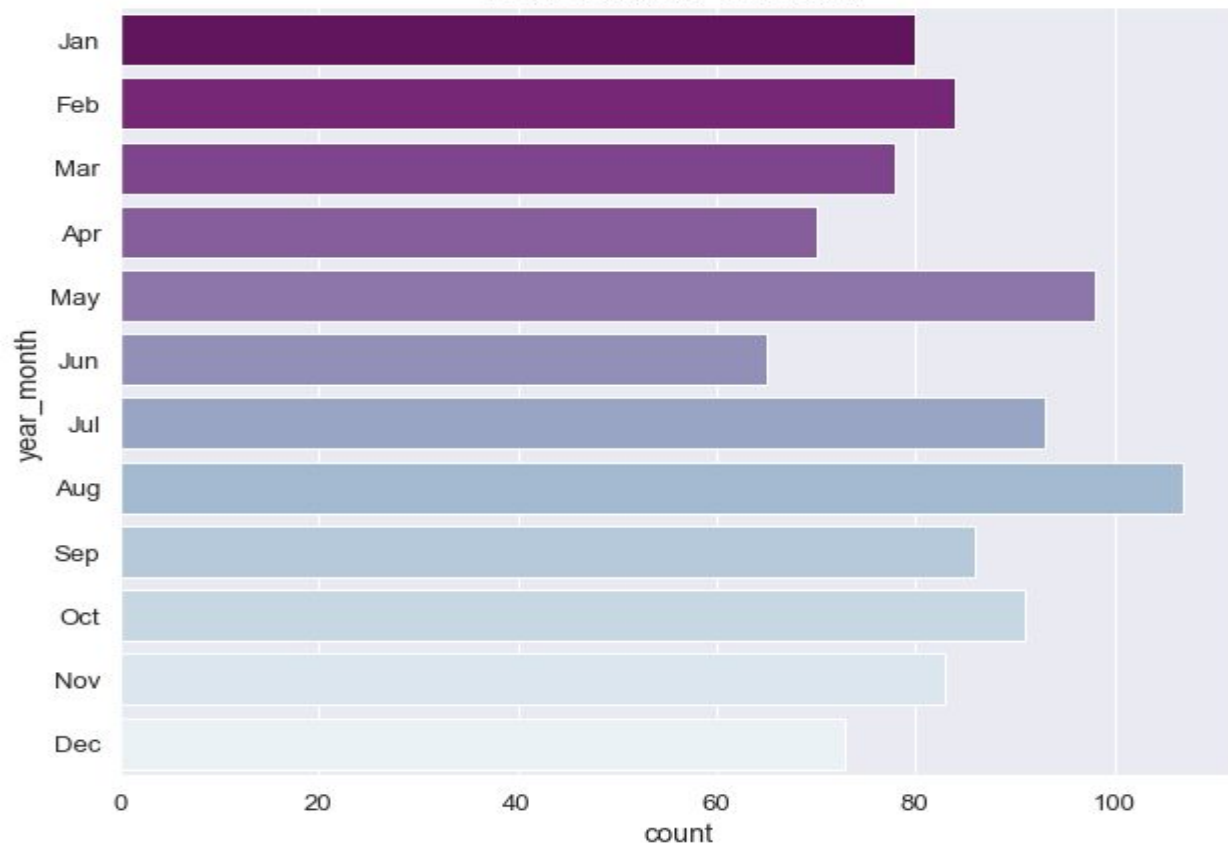
# Model Evaluation

Criteria :

- Main Metric - Testing Accuracy
  - Confusion Matrix
  - ROC with AUC curve
  - Model coefficients

# Model Performance - overview

| Model | Train Score | Test Score |
|---|---|---|
| Base Line | 0.78 | |
| Logistic Regression | 0.78 | 0.77 |
| Decision Tree | 1.0 | 0.90 |
| Random Frst. | 1.0 | 0.90 |
| Extra trees | 1.0 | 0.90 |
| Voting clsfr. | 0.99 | 0.87 |

# Model Evaluation

**Extra trees model** produced

## 0.90

accuracy score - better than other machine models.
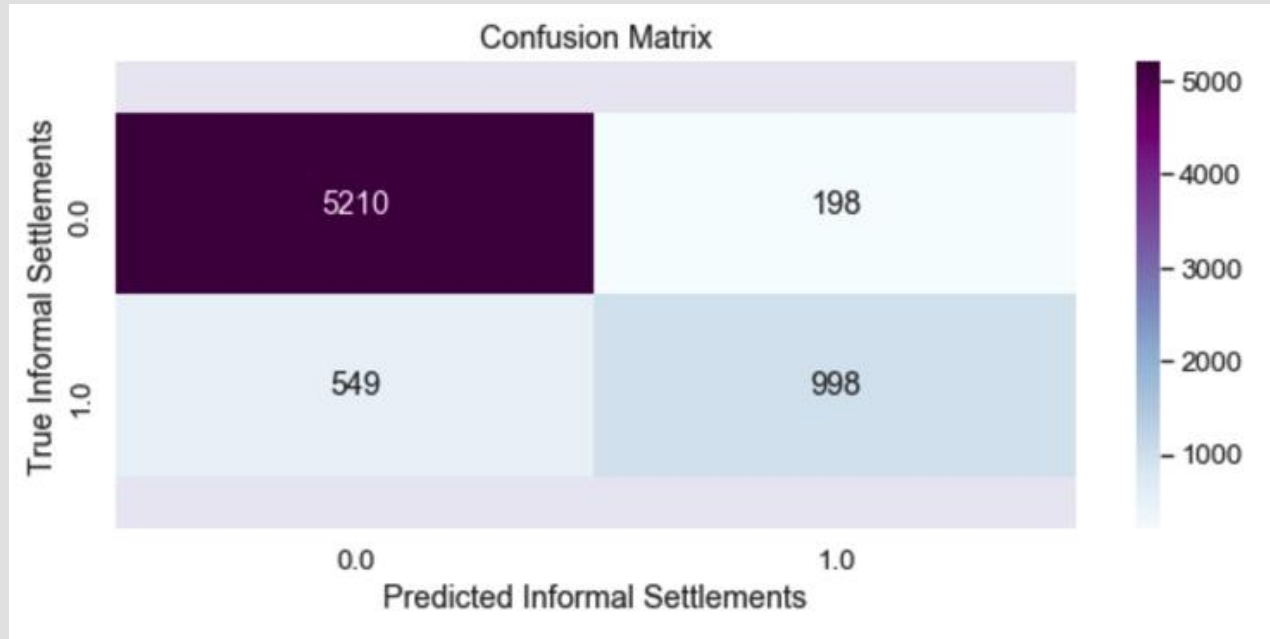
# Model Evaluation - Confusion Matrix

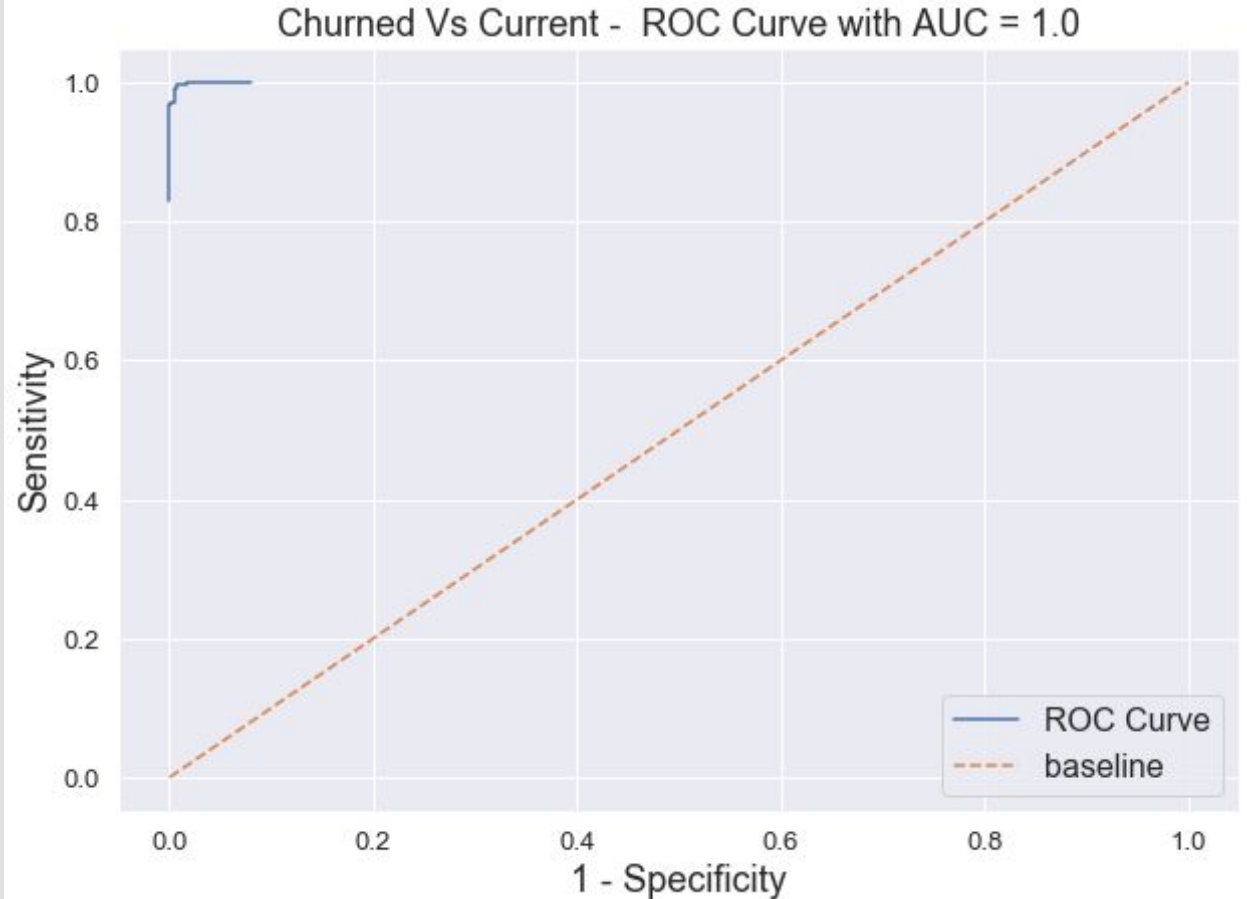- Accuracy = 0.90

- TPR % 83.6%
($TPR=TP/(TP+FN)$)

- TNR% 90.9%
($TNR=TN/(FP+TN)$
)

https://predictfavelas.github.io/kumar_predictions

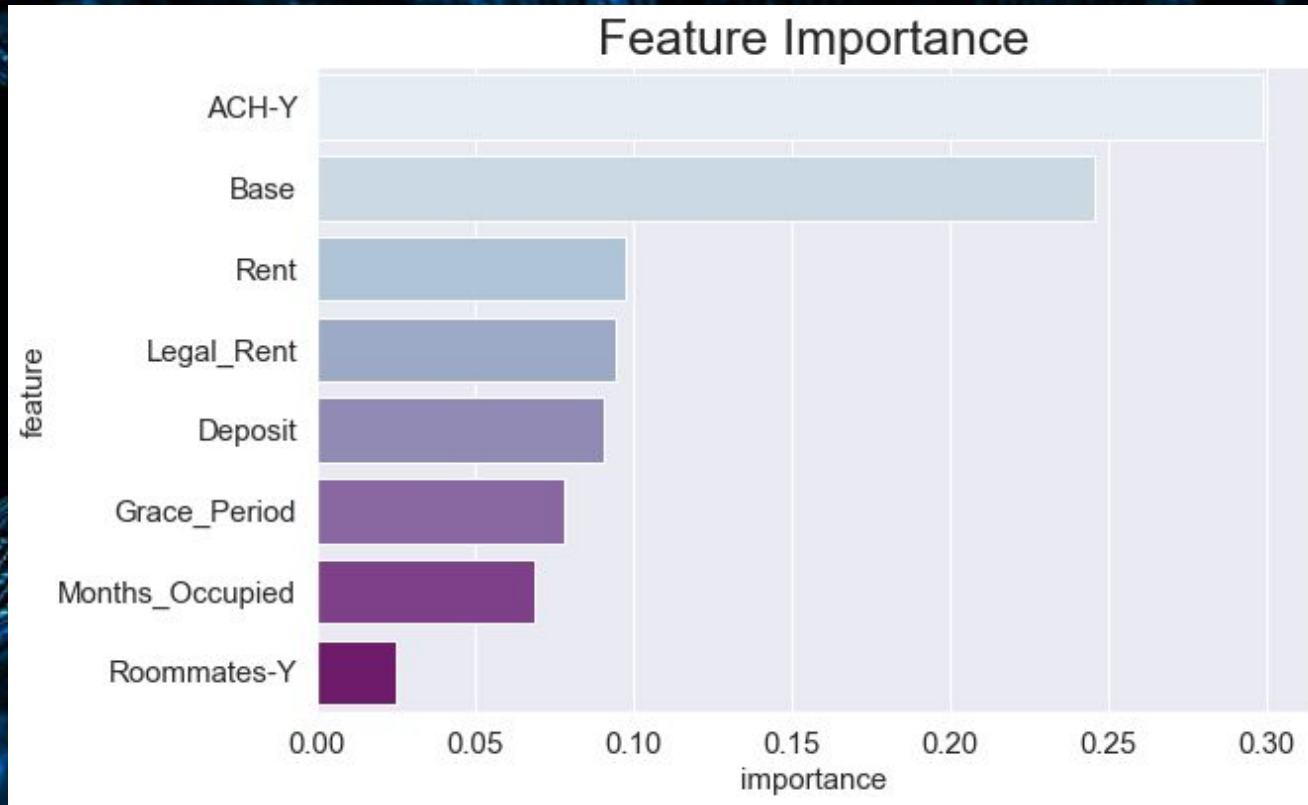# Model Evaluation - ROC AUC Curve

- ROC AUC of close to 1

- Positive and Negative classes are perfectly separated



Churned Vs Current -  ROC Curve with AUC = 1.0

# Model Evaluation - Feature Importance (Extra trees)



Feature Importance

**Extra Trees Model**

- **Tenants signing up for Auto debit of their rent is the most important feature of the model**
- **Late fee calculation percentage is the second most important feature.**

# Model Evaluation - Model Coefficients



Feature Coefficients
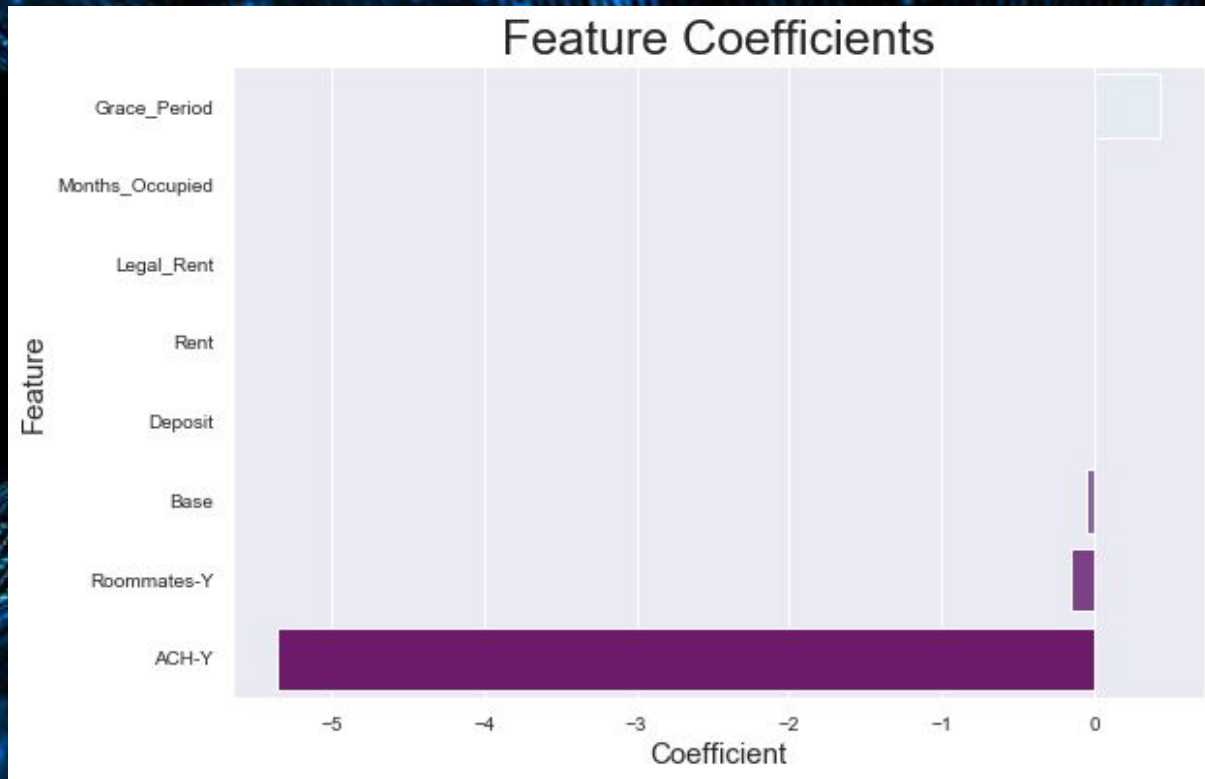
- **Distance to closest apartments is the Sao Paulo dataset biggest indicator of no favela**

- **Distance to closest store listing is biggest indicator of favelas in the census tract**

# Conclusion

- Extra trees model performed the best (at 100% accuracy).
- Our model will help to differentiate the Churned and Non Churned tenants for the Property Managers.
- Would like to get more data from Social Media, Work Order Review to aid the predictive models.
- Time based predictions to be done using Generalized Linear Models.
- Develop an API between Machine Learning Models and the Property Management Application.

# Thanks - Questions?

## The Team

Kathirvel Kumararaja        Mahdi - Instructor        Jeremy - TA        GA DSI 10 Cohort