

# Predicting Tenant Churn

## For Property Management

...

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
  - Market overview
  - Findings
    - Data overview - EDA
    - Feature Engineering
    - Model evaluation
  - Conclusions and recommendations
-



# Problem Statement

Our churn model will predict, if a tenant rents a unit, the probability that tenant is a churned or a non churned tenant. Armed with this knowledge a property manager can better understand how the tenant mix will change over time and which units are likely to become available soon. This will enable managers to proactively target the high churn candidates and provide incentives for them to extend their contract. In addition, the model can highlight which units likely will need to be filled in the future, so that the vacancy period of a unit can be reduced.

The data we are using for this modeling is anonymized with all the personal info removed to protect the privacy of the tenants.

- This is a classification model.
- We will be evaluating our model based on Accuracy.

# Problem Statement - In business application

## Business Objective

To help Property Managers run their business effectively by predicting whether 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 data from Property Management System to engineer data
- Train the models to predict Churned Tenant.
- Evaluate the model using Accuracy as the criteria



# Property Management - Stats

The property management industry is extremely lucrative and is becoming more and more popular with the rise of renters. Let's take a look at some of the top statistics from 2018 below.

**280K+**

total property  
management companies  
in US

**\$76B+**

property management  
industry revenue

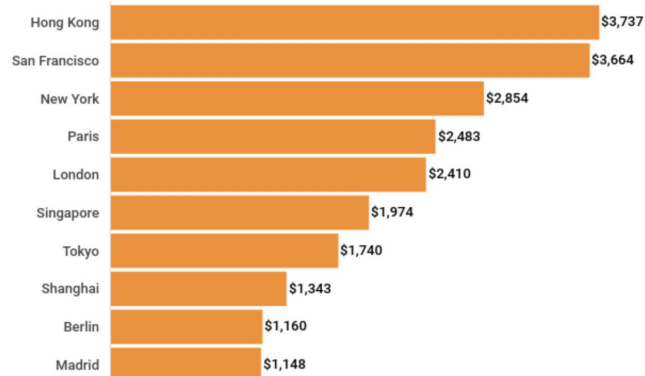
**13%**

percentage of US GDP  
made up by real estate

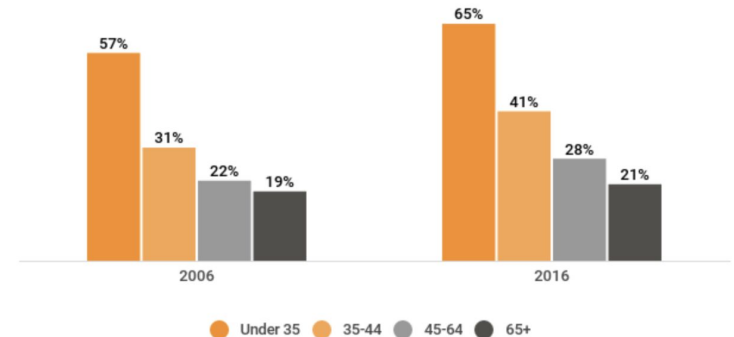
**67%**

percentage of US renters  
aged 35 and under

RENT PRICES AROUND THE WORLD



RENTERS BY AGE IN 2016





## Opportunity for Technology

- **Property management is a multi-billion dollar industry, with plenty of growth opportunities for those companies that are willing to dive head-first into technology and embrace the Millennial market.**
- **With renting at a high not seen in decades, and demand up due to lower inventory, the property management industry is expected to continue to grow for the next several years.**



## Opportunity for Technology

- **New innovations in technology, including the use of AI to increase availability and efficiency are streamlining the industry like never before, making it easier to take care of client needs, and keep property owners up-to-date on the day-to-day operations of their properties.**
- **All signs point to it being a good time to be a property manager.**

# Birds eye view

## Qualitative data 1

- Real Estate Data sets
  - Resident history from Property Management System

## Qualitative data 2

- Time scale
  - Five years of Resident data

## Qualitative data 3

- Baseline Score
  - 0.67

## Qualitative data 4

- Models / Classifiers Explored
  - Logistic Reg, Random Frst, Extra trees.

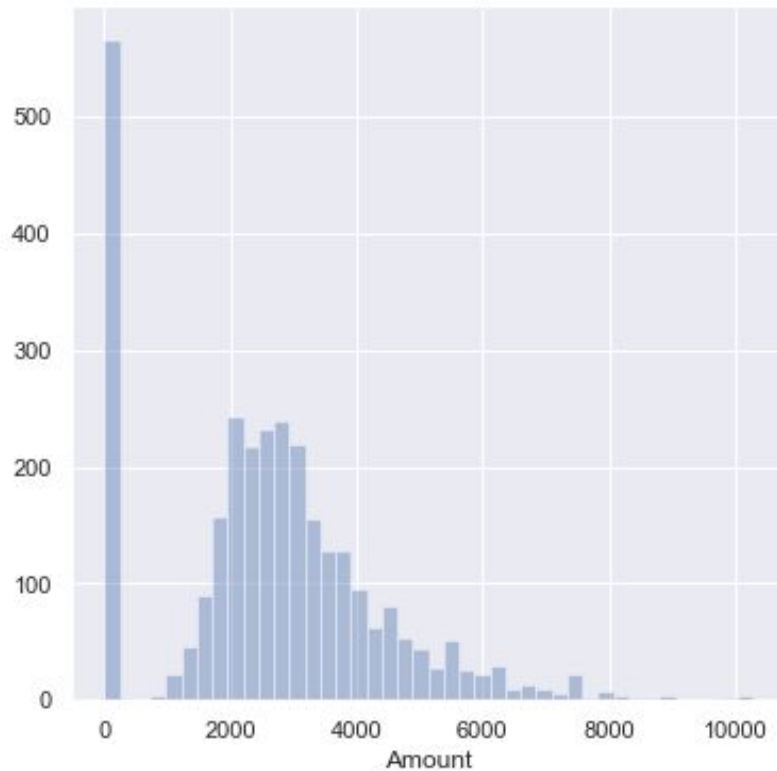


The background is a dark blue to black gradient. It features a grid of small, glowing blue dots that form a perspective view, receding into the distance. Overlaid on this grid are horizontal bands of binary code (0s and 1s) in a lighter blue color. The overall effect is a digital, data-driven aesthetic.

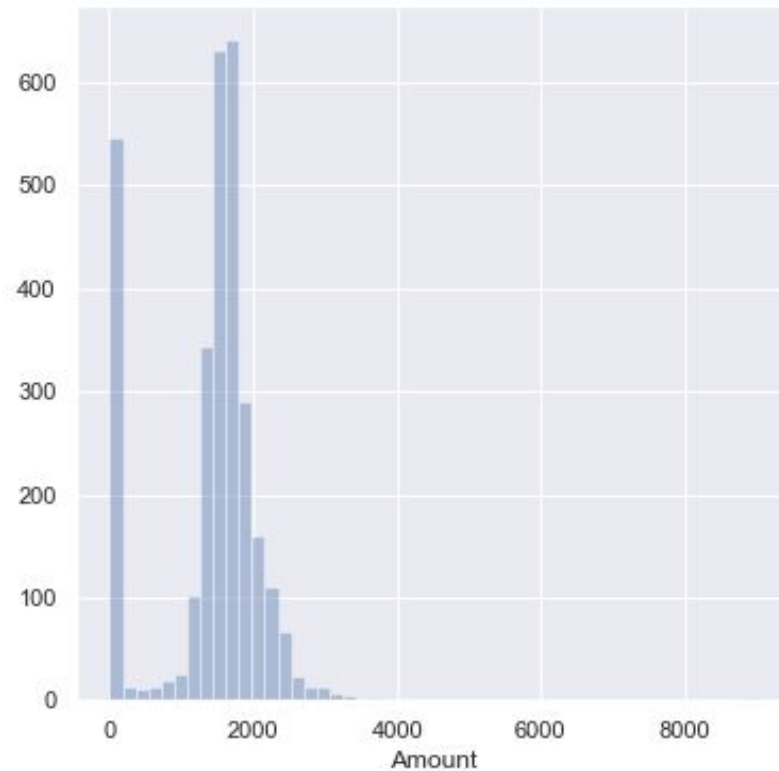
# Data Overview - EDA

# EDA - Rent fields

Distribution of Legal Rent



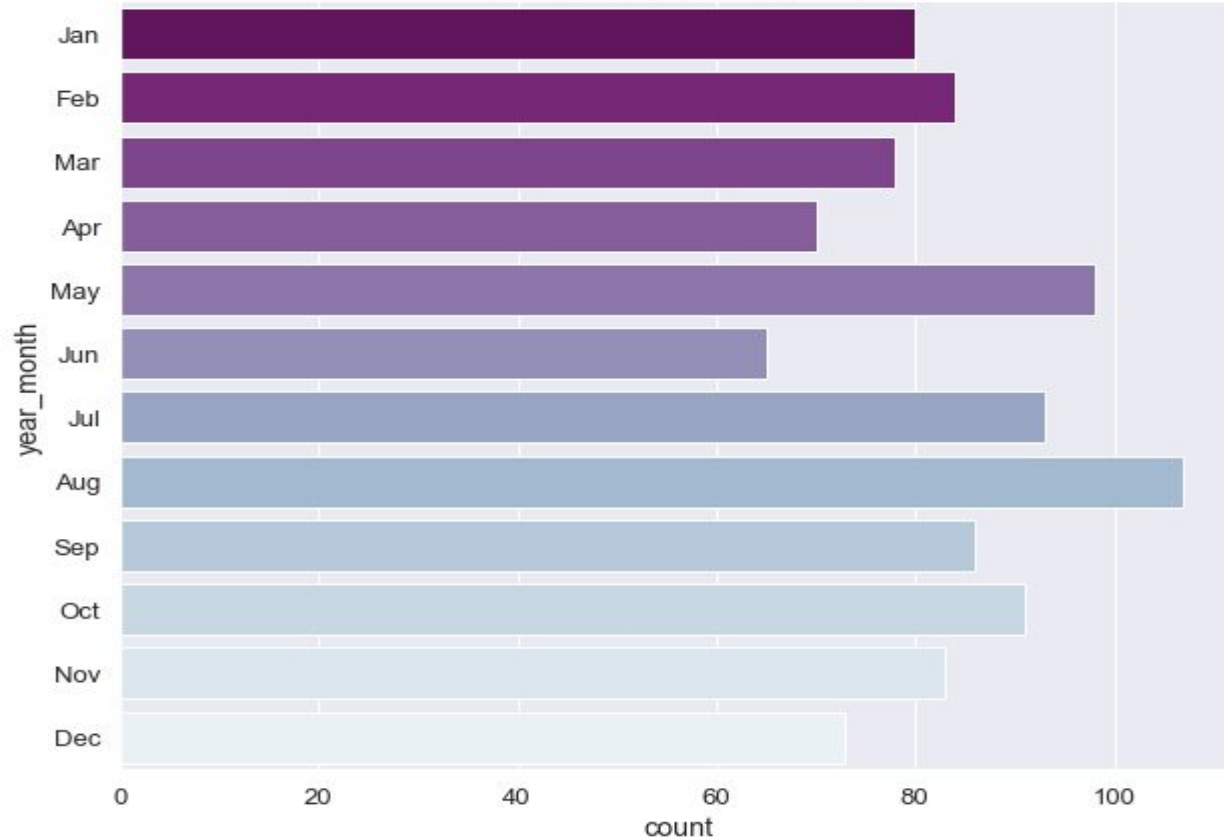
Distribution of Rent





# EDA - Move-Out Monthwise

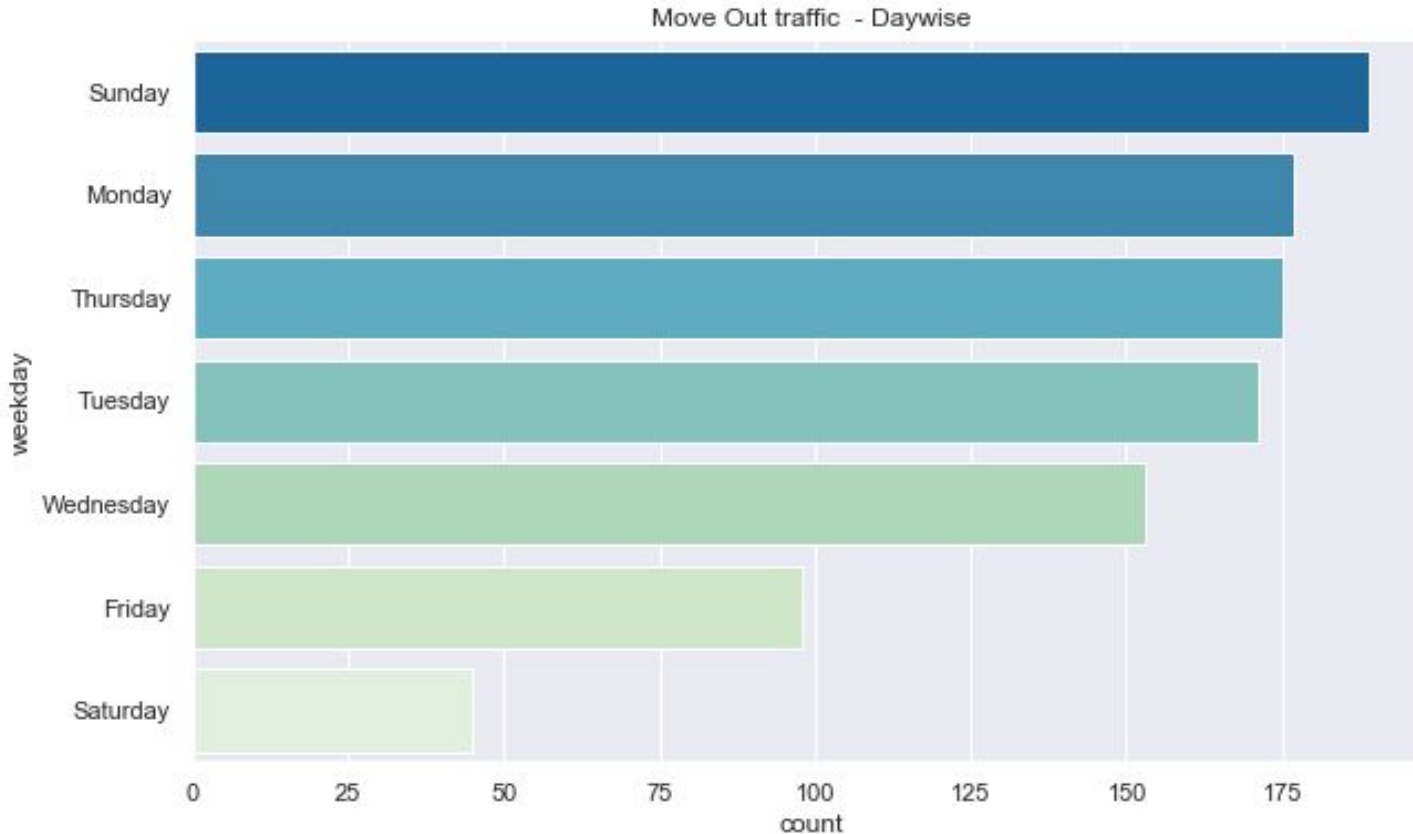
Move Out traffic - Monthwise



## *Monthwise Moveout Traffic Graph Interpretation*

Traffic Graph shows that most of the traffic happens during Summer months (May, Jul, Aug, Oct). I have also noticed that the traffic is more just after the beginning of School Holidays and picks up just before School start month.

# EDA - Move-Out Day-wise



*Daywise Moveout  
Traffic Graph  
Interpretation*

The above graph of  
Daywise Moveout  
Traffic Graph shows  
that most of the  
traffic happens  
mostly during  
Sunday & Mondays.



# Model Evaluation

## Criteria :

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

# Model Performance - overview

Model	Train	Test Score
Base Line	0.67	
Logistic Regression	0.92	0.92
Decision Tree	0.96	0.95
Random Frst.	1.0	0.95
Extra trees	1.0	0.95
Neural Net	0.96	0.94



# Model Evaluation

Extra trees model produced

0.95

accuracy score - better than other machine models.



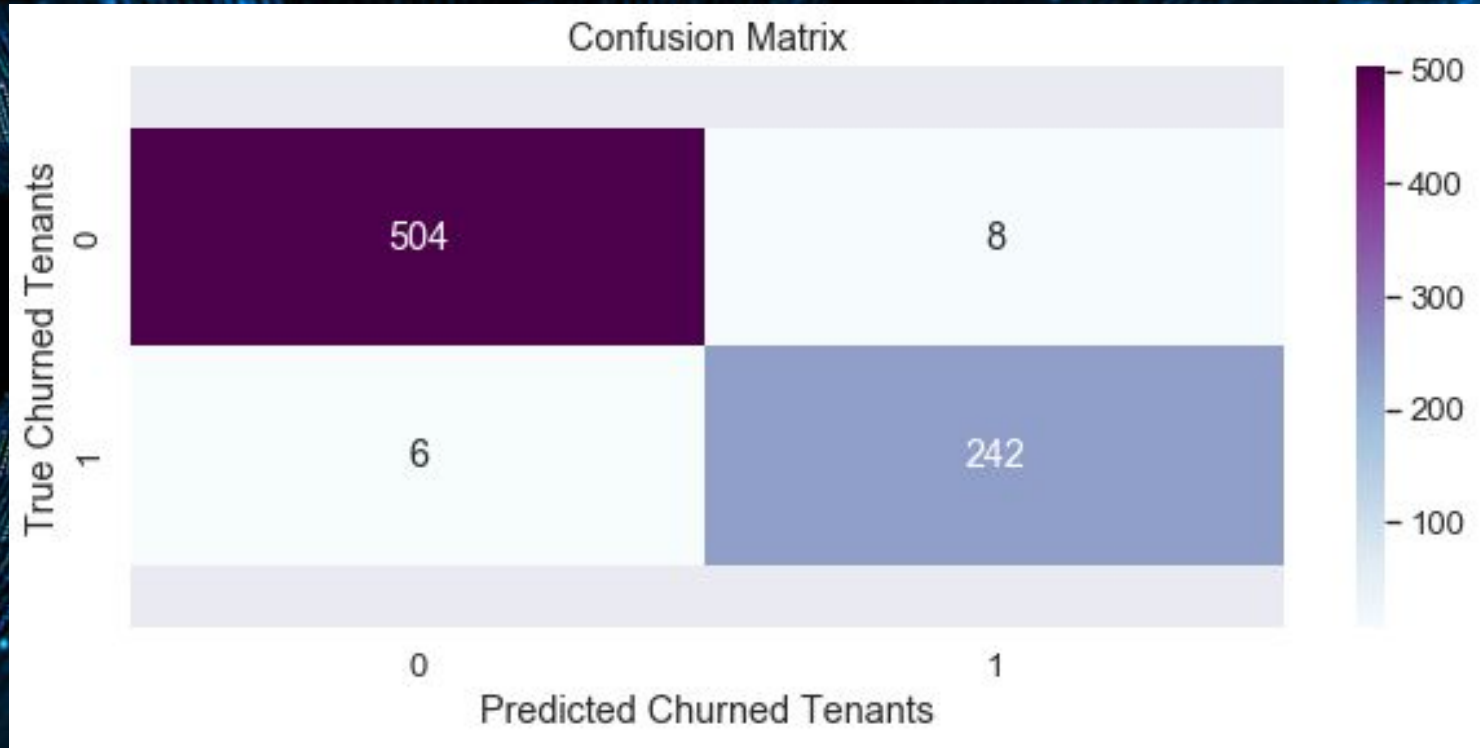


# Model Evaluation - Confusion Matrix

- Accuracy = 0.90

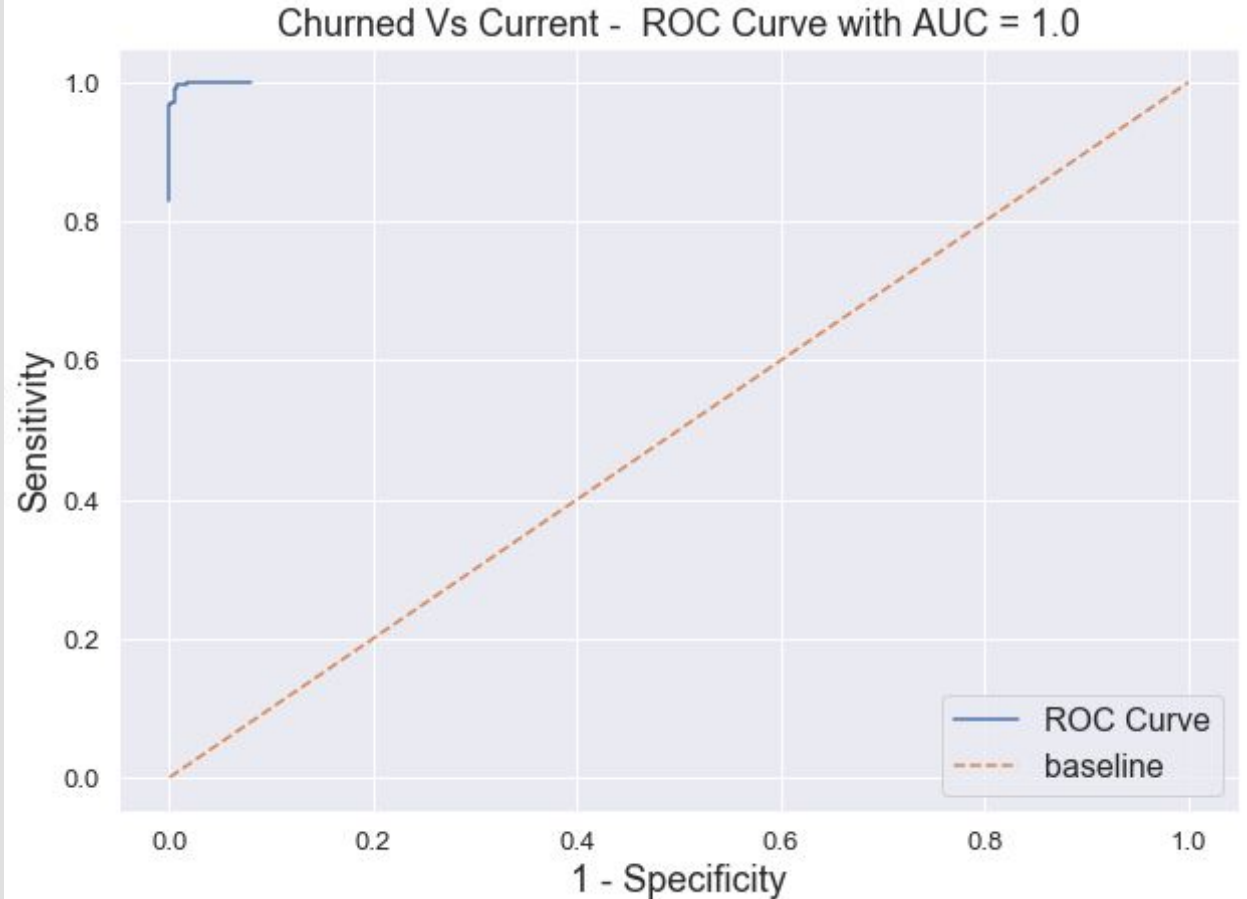
- TPR % 98%  
( $TPR = TP / (TP + FN)$ )

- TNR% 97%  
( $TNR = TN / (FP + TN)$ )

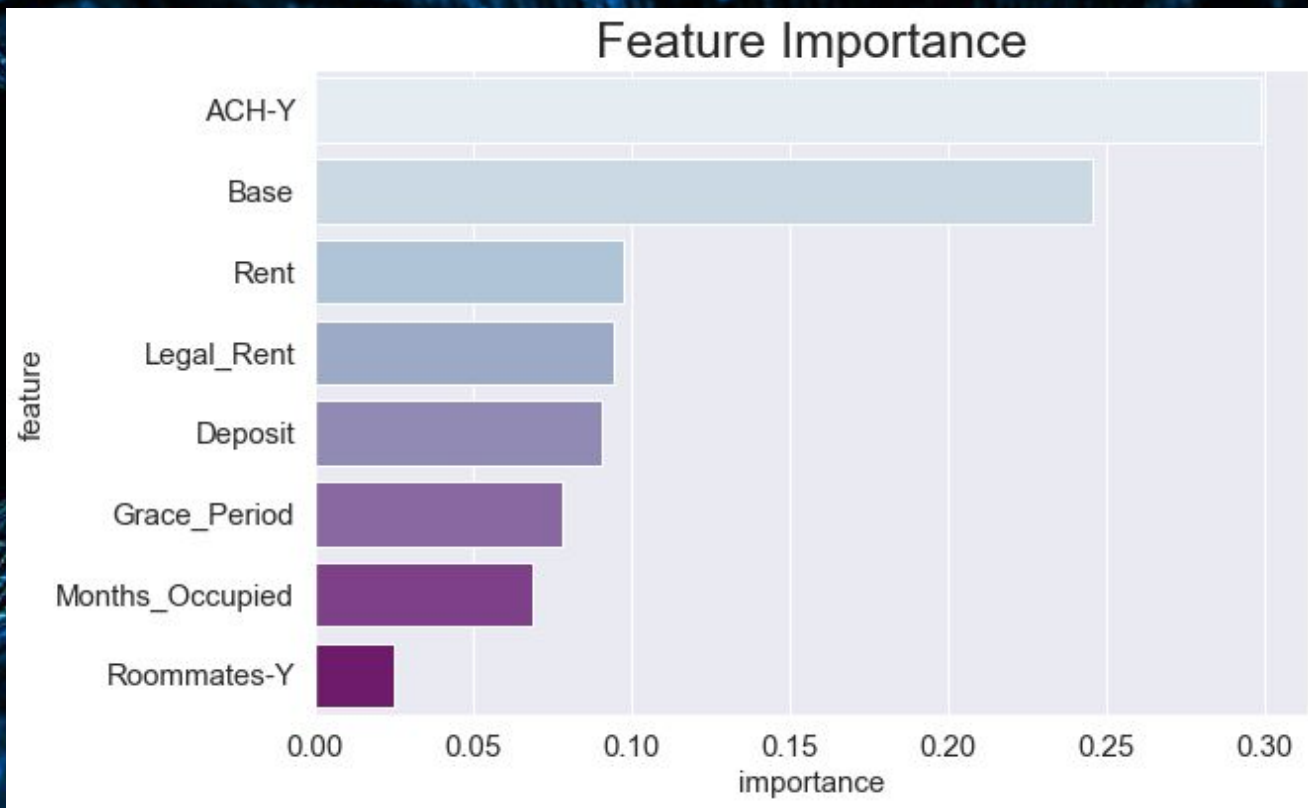


# Model Evaluation - ROC AUC Curve

- ROC AUC of close to 1
- Positive and Negative classes are perfectly separated



# Model Evaluation - Feature Importance (Extra trees)

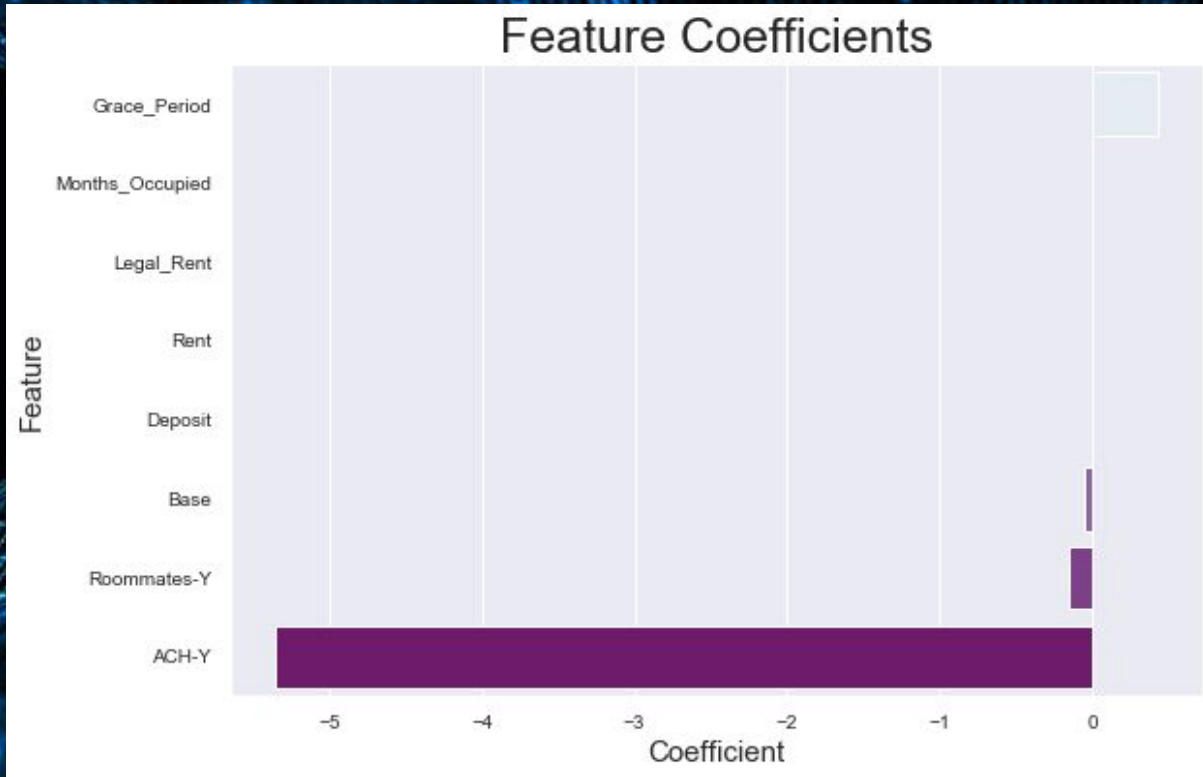


## Extra Trees Model

- 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



- We can infer from this findings that Low rent paying tenant tends to stay longer in a Unit that is he is less likely to churn / move-out.
- Another interesting findings is that the tenants who signed up for electronic payments tend to stay longer.

# 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