MARAX AI

DataStory

1st September, 2017.

License

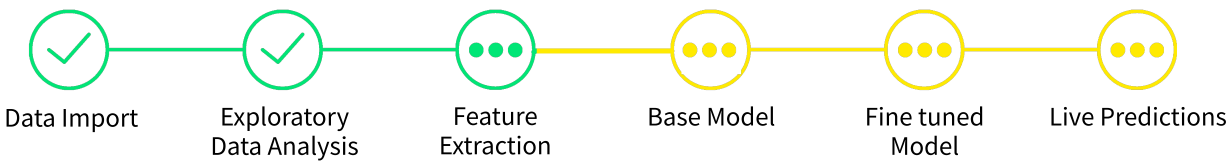
Introduction

Nestaway is a home rental platform for Singles and Families. Nestaway has partnered with Marax AI to predict user behaviour, perform product analytics and eventually generate actions through AI models to retain users and maximize their lifetime value. Marax is starting with predicting which tenants might raise a move out request in the next 30 days so that Nestaway can take targeted action to retain them.

This report consolidates **Exploratory Data Analysis(EDA)** Marax team has performed on the data provided by Nestaway as well as certain *external datasets* which Marax has collected to make the AI models more robust and accurate. EDA was performed to get initial insights on the data in the context of Churn. Eventually leading to advanced Artificial Intelligence/Deep learning models to make highly accurate predictions and recommendations to solve churn for Nestaway.

It also summarizes some interesting insights which will help us create rich features to train/tune our AI models. This exploratory data analysis helped us determine relationships between explanatory variables, check assumptions and detect anomalies which are key to making accurate predictions.

The initial EDA process is the most time consuming part of this pilot and we are ~ 60% through before we go live with predictions. Currently, Marax team is creating features for making accurate predictions through our proprietary AI platform.

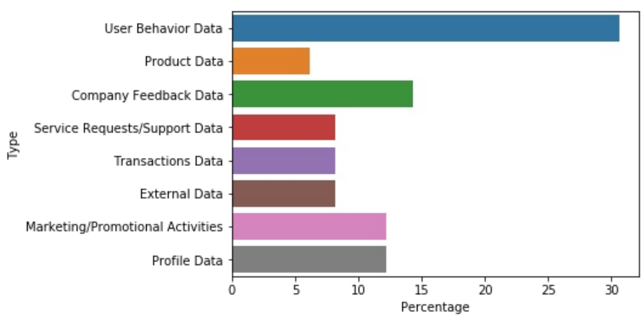


The data and the insights in this report are used to help you understand how your customers behave on Nestaway. However, these are just preliminary observations which are taken into consideration while making prediction/ tuning our AI models for Nestaway
If you decide to use the data, we recommend you use it together with other observations and behavior patterns to understand how Churn for your platform can be reduced.

Contents

Data Universe

Data distribution



63M
Data Points

24
Months

56,682
Tenants

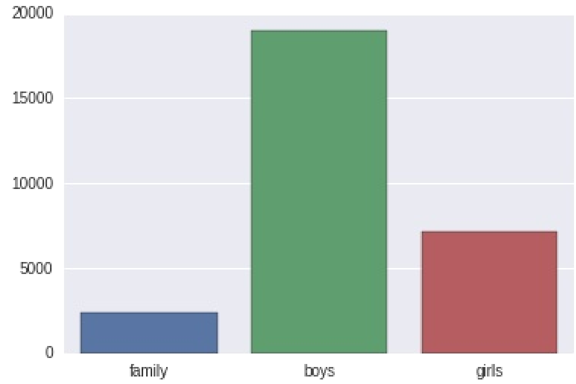
15%

of new users use coupons to sign up

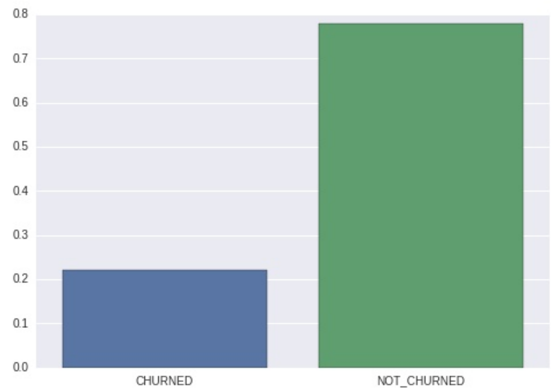
25%

of new users are referred by old nesties

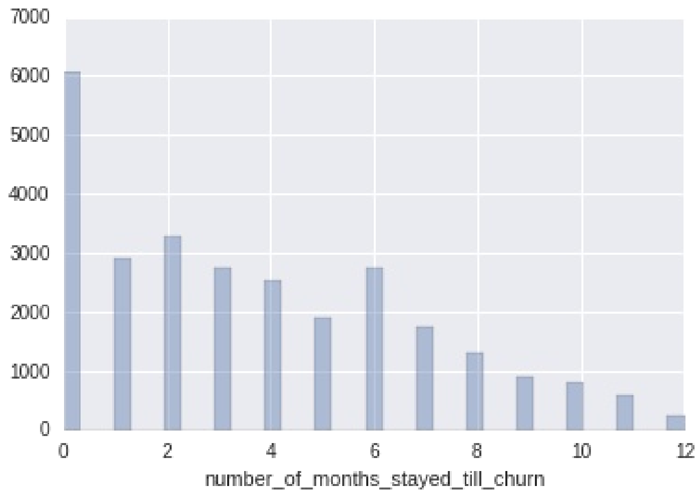
Tenants distribution



Churned vs Not Churned



Duration stay distribution at a Nestaway



9%

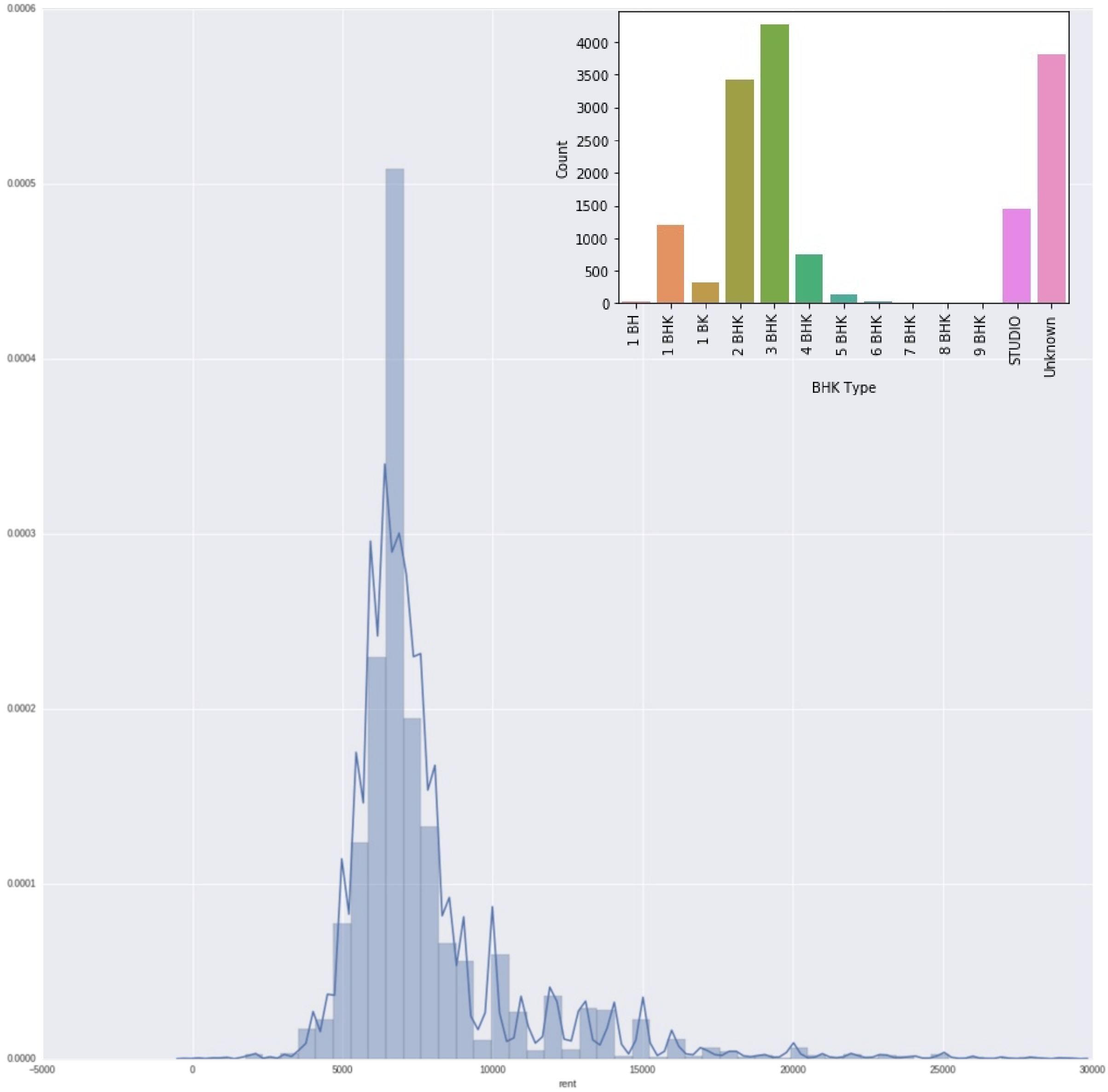
of Tenants leave in the first month of stay

15,427 Houses

63%
of houses have rent ranging from 6k to 8k INR

50%
of houses comprise of 2 and 3BHKs

Rent price distribution **House type distribution**

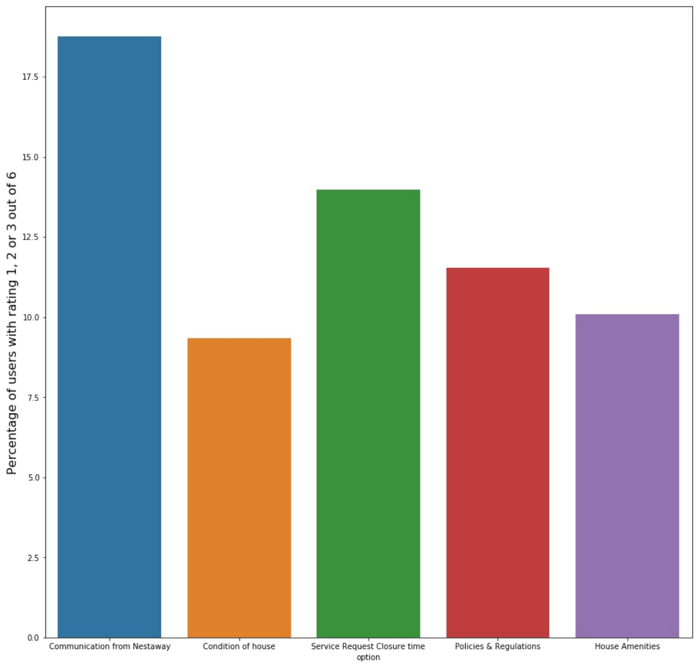


157,789 Feedbacks

39%
of in-app feedbacks have a rating

67%
of tenants have given a feedback

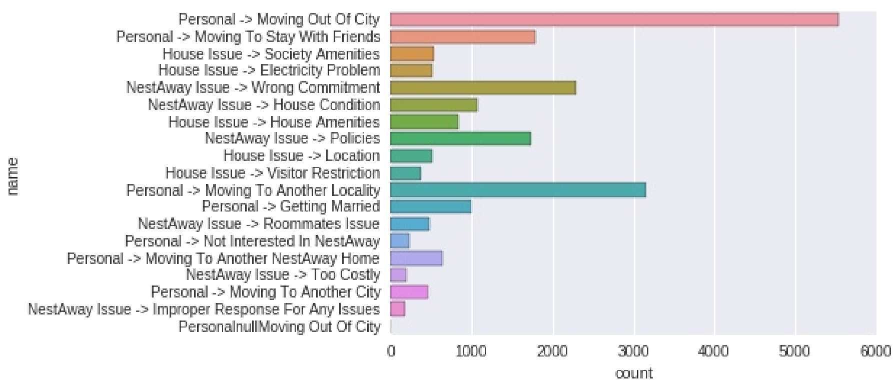
NPS feedback distribution



50%
of tenants left Nestaway due to either Locality or wrong commitments from Nestaway

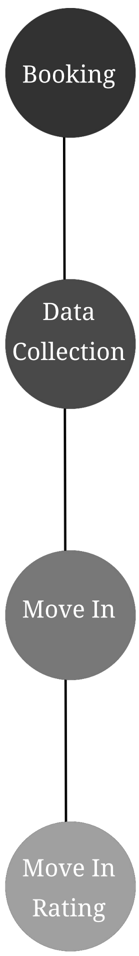
63%
of churned tenants gave 1, 2, 3 star rating

Move-out requests' feedback distribution



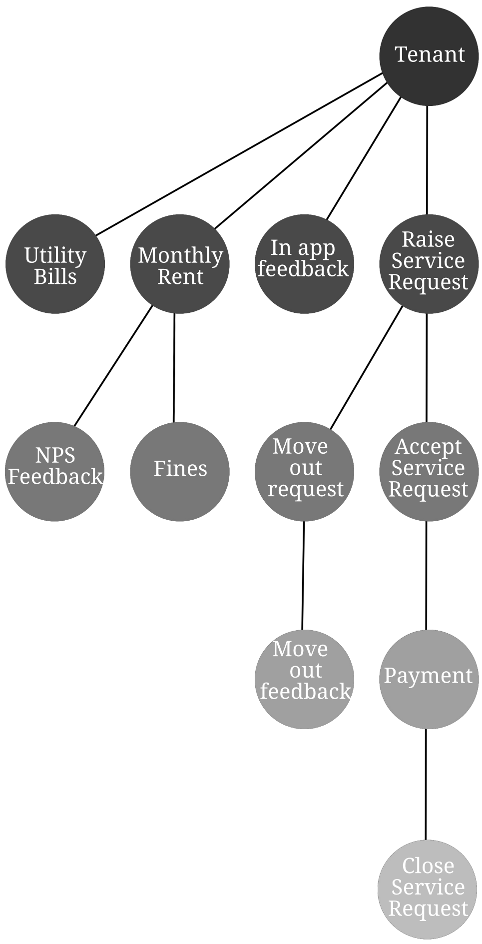
Tenant **Story**

Onboarding



A tenant typically interacts with Nestaway through a mobile application, using which they can pay their monthly rent after which they are asked a NPS feedback on Nestaway. Failure of payment of rent leads to penalties and eventual eviction of the tenant. They can also raise service requests to solve a problem which Nestaway is responsible for. After the service request is resolved, tenant is asked to rate the service request. If a tenant decides to move out of the current house, they can send a move out request with the reason for moving out using the app. The tenant can also at any time give feedback on the Nestaway platform through the app itself.

Living



Data
Queries

Effect of service requests on churn

Effect of house amenities on churn

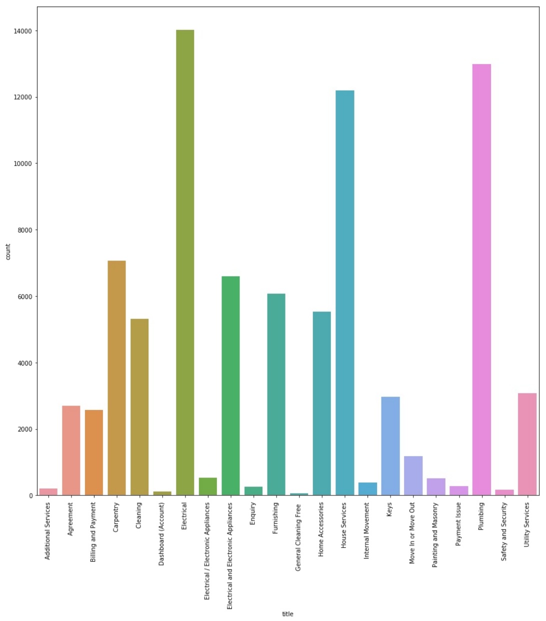
Effect of house locality on churn

Effect of **service requests** on churn

Earlier, we found out that the tenants rated the Service requests and Communications from Nestaway as the areas Nestaway can improve on. We explored Service requests and its relation with Churn and came up with the following findings.

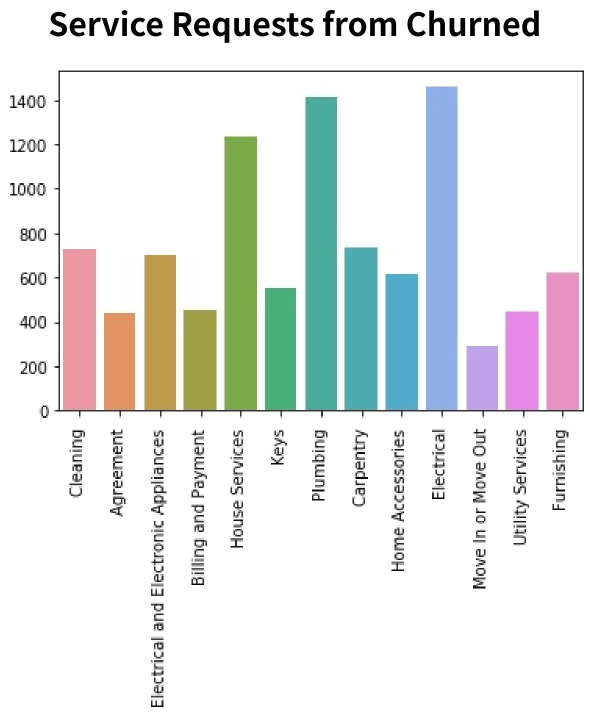
Key findings

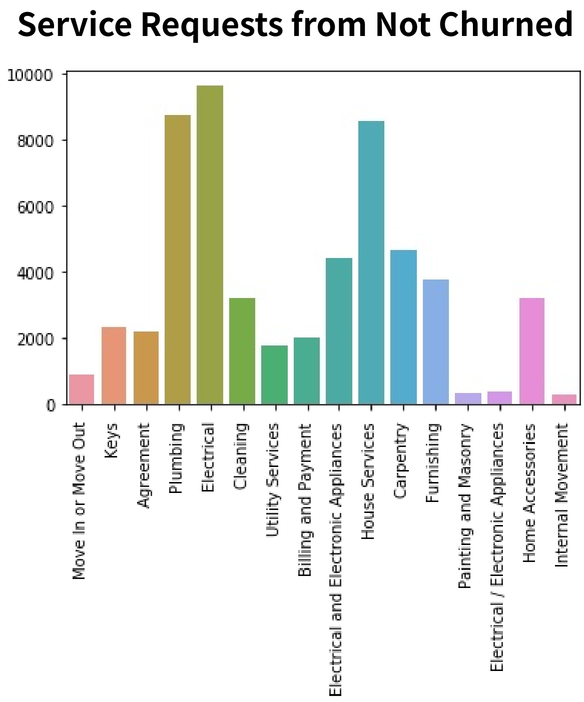
1. Churned tenants *incurred double the service costs.*
2. Churned tenants *raised twice the number of service requests* than current tenants.
3. *House services, Electrical and Plumbing* are the most requested Services in Nestaway



What types of Shared Service requests are tenants raising?

Electricals, House services and Plumbing represent **45.78%** of all service requests raised
13.12 % of the Service requests were shared





What types of requests churned Tenants and Non churned Tenants raise ?

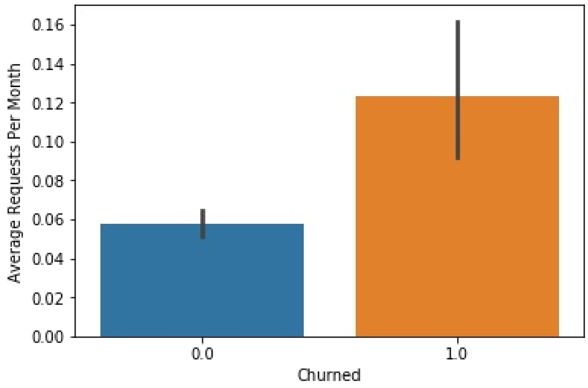
After looking at different categories of service requests, we explored if there is any significant difference between the categories of service requests that churned Tenants are raising with respect to non-churned Tenants.

Service requests count was normalized and plotted against Churned and Non Churned Tenants. It was observed that the *churned tenants and Non Churned tenants raised almost the same types of service requests*.

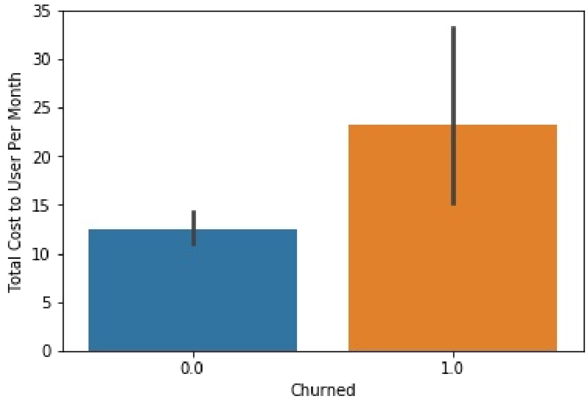
The bar plot shows the number of service requests raised (y axis) by churned(left graph) and non-churned(right graph) Tenants in different categories (x axis)

How many Service Requests do tenants raise and how much do they pay for it?

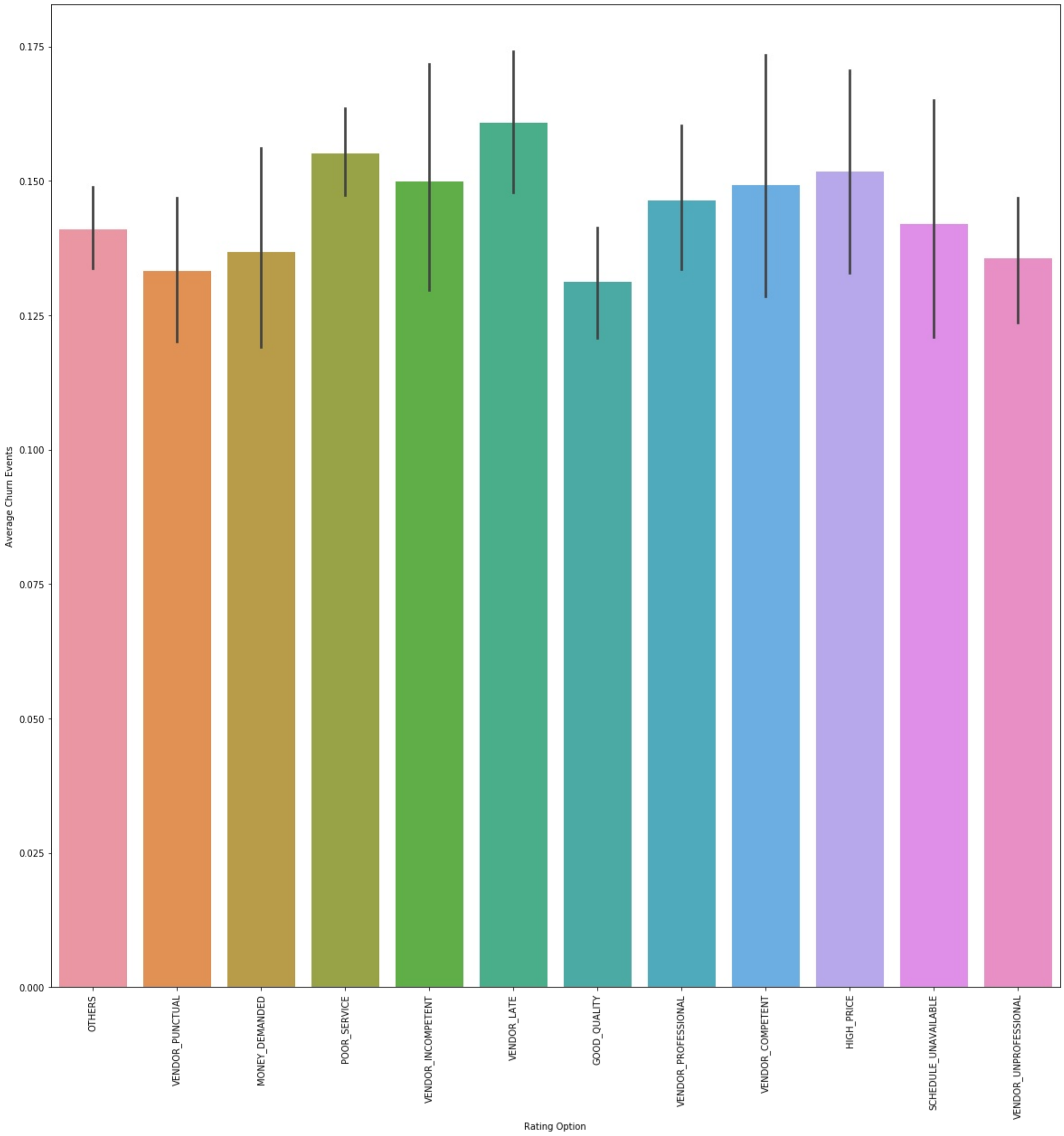
The number of service requests were taken per month as total number of service requests would have been biased towards newer Tenants who had lesser time to have any service requests



After plotting the number of Service Requests raised by Churned people and non Churned people, it is observed that the *Tenants who Churned raised almost twice the number of Service Requests than Non Churned Tenants.*



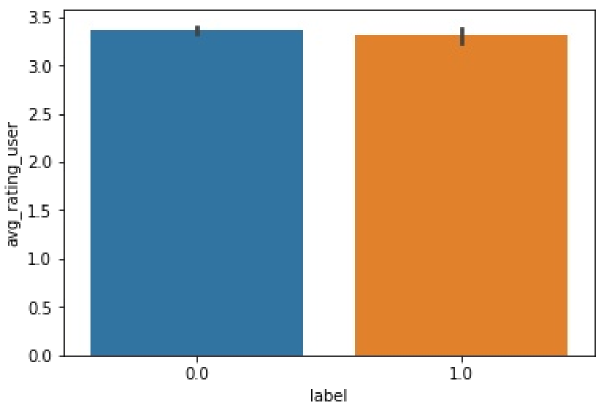
Similarly, we plotted Cost of Service Requests by Churned people and non Churned people on the x axis and the total_cost_user which was normalised we observed that the *churned tenants paid almost twice the cost for Service Requests than Non Churned Tenants*



The Y-axis is the average number of churned users per month associated with these feedbacks. The X-axis is a feedback option of a service request.

Does Service Feedback affect Churn?

On exploring the Service Ratings given by Tenants who churned, a plot was made between the average number of churn events(y axis) vs various rating options of feedbacks(x axis).
It was observed that higher number of average churn events are associated with negative rating options and vice versa



A Feedback Rating vs Churned and Non Churned Tenants was plotted and it was observed with preliminary analysis that there is *no direct correlation between ratings of services and Churned Tenants.*

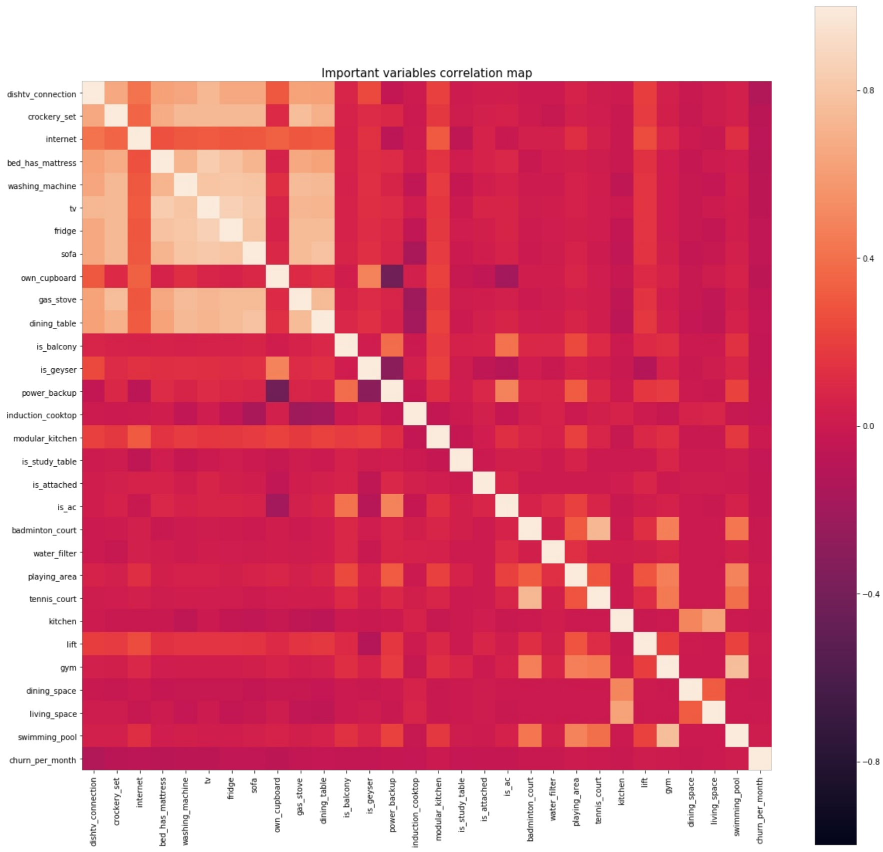
Effect of **house amenities** on churn

Tenants biggest interaction with Nestaway during the stay is the home in which they are staying in. Exploration was done on the House Amenities Tenants have access to and understand if there is Churn Behavior associated with Amenities.

Key findings

1. Relationships of Amenities with each other
2. Amenities Correlation with Tenant Churn

How are amenities related to each other?

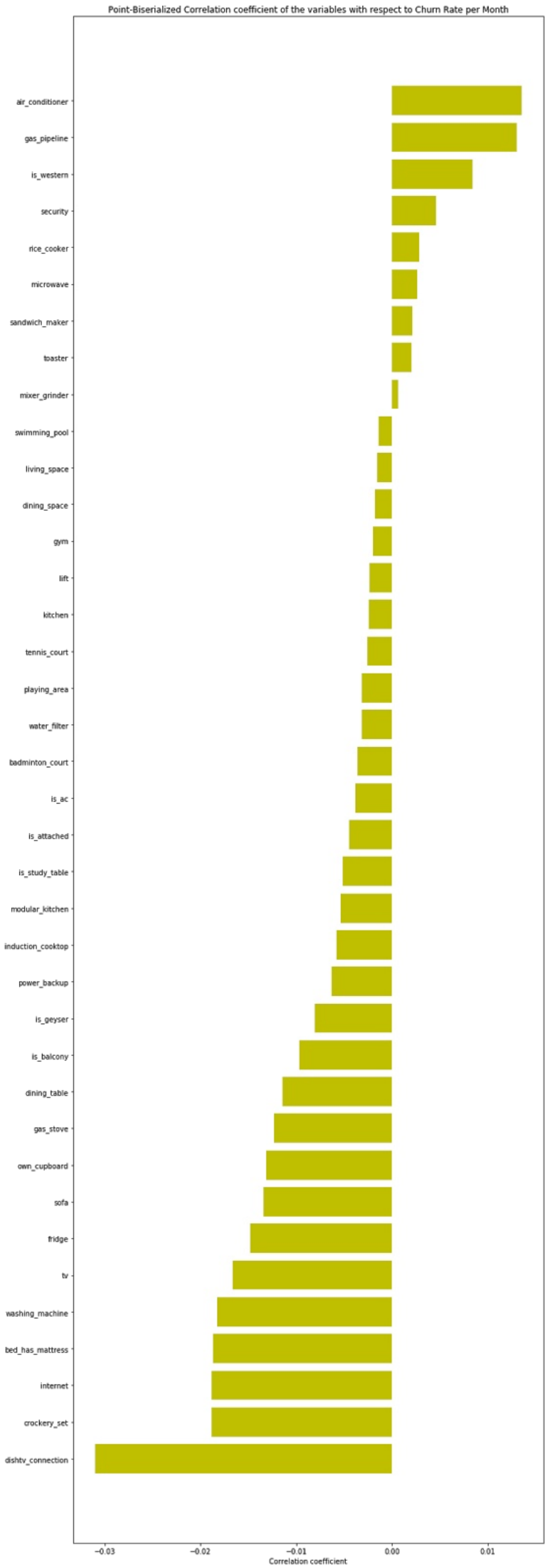


This heatmap shows the correlation of the presence or absence of various amenities with each other. The high positive correlation seen(light brown) indicates positive correlation.

This plot helps us understand the relationship between the various amenities.

Eg. We can see that houses that have badminton courts also have gyms, swimming pools and tennis courts. More intricate details can be derived by doing bi and multivariate analyses.
The highly correlated features also indicate the important features for further stages of testing.

How are having certain amenities correlated with Tenant Churn Behavior?



A plot was made to understand correlation of presence of various amenities with churn per month. Since the amenities were a binary variable and the churned Tenants per month was continuous, we used *Point-biserial correlation coefficient*. Churn rate was calculated as Churned Tenants upon Length of the house on the platform.

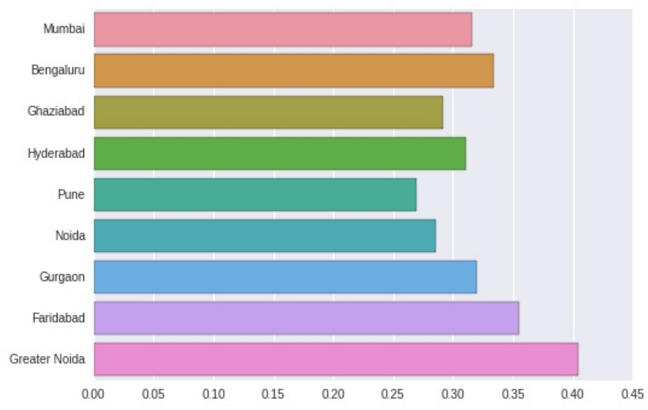
We can see that important amenities are negatively correlated with churn. This indicates that the absence of these amenities was seen in houses associated with more frequent churn.

Effect of **house locality** on churn

One of the main reasons why tenants send a move out request was the locality of their house.
We decided to further explore the localities of Nestaway tenants in Bangalore.

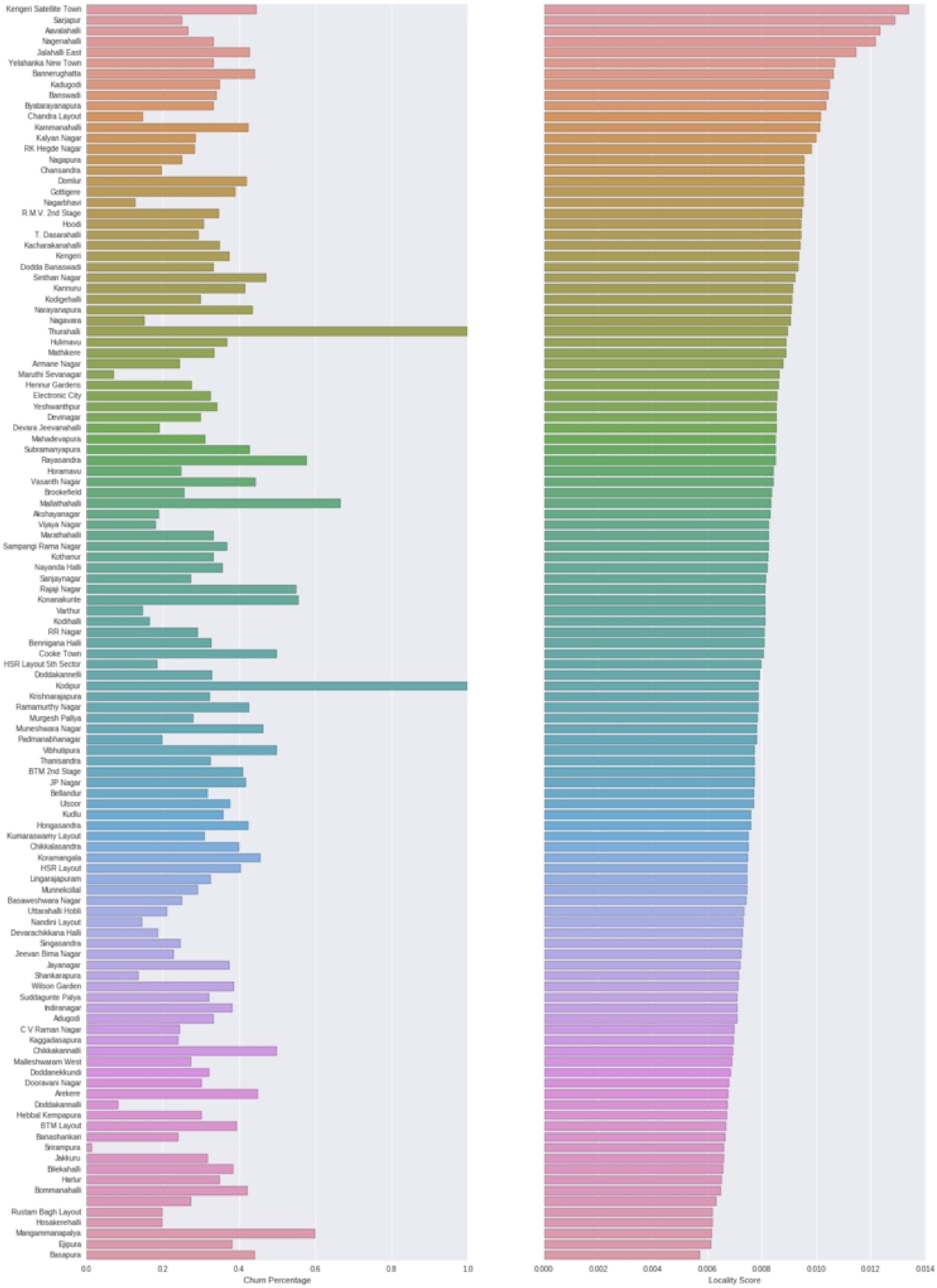
Key findings

1. Tenants who live around *higher restaurant popularity index* are less likely to be Churn.
2. *Locality scores* around Bangalore.

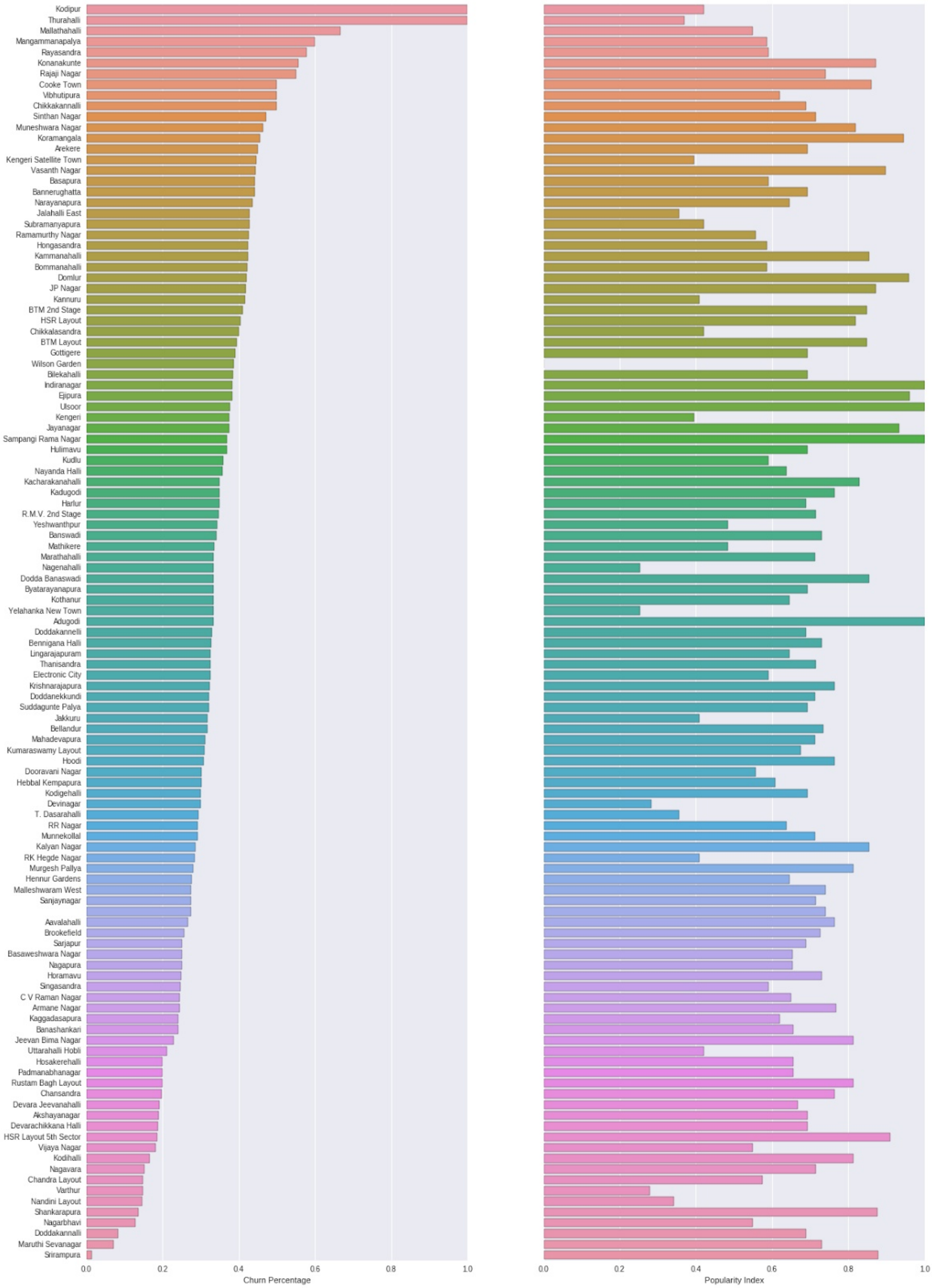


Which cities have the most people Moving out?

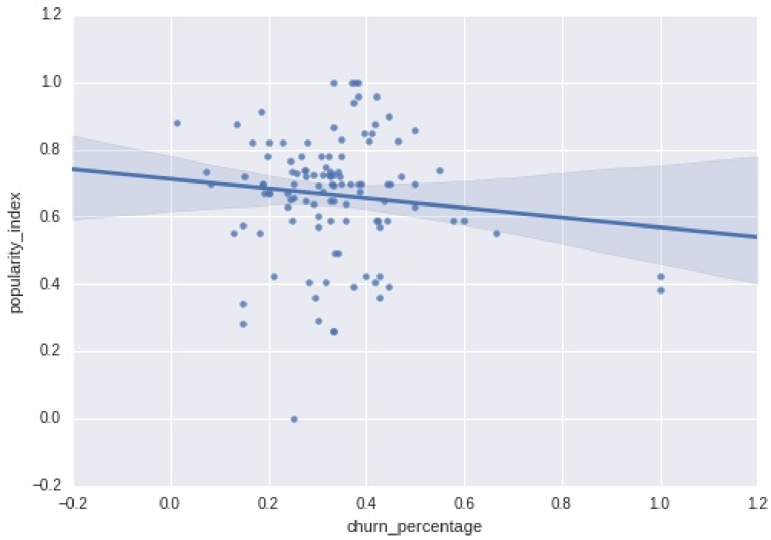
This plot shows us the percentage of users who Churn out from Nestaways in each City. We decided to further drill down on the City and chose Bangalore for our exploration.



Locality scores were created based on nearby places for entertainment and needs. The plot shows the number of Churned percentage of each locality and their corresponding locality score



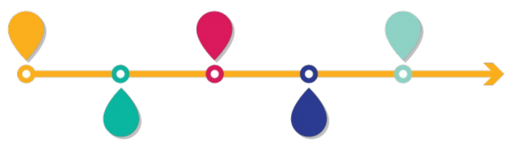
A location's restaurant index and churn percentages were compared to understand Tenant Churn behavior based on the number of restaurants in a locality



To further understand the relation between the Churn Percentage of houses and the Restaurant Popularity index of the locality of the house - it is observed that *houses with less tenant Churn rates have usually higher Popularity index Scores*

Challenges

One of the goals of Exploratory data analysis is to make sure the data which goes into training the model is high quality and less noisy. This section summarizes the data quality and sparsity problems we came across while analysing Nestaway's dataset and suggestions for further data quality improvement, enrichment and collection.



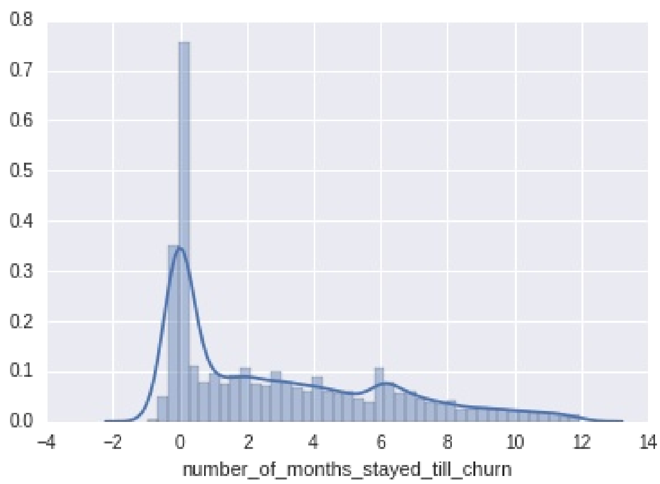
Preserving log of changes in data over time

For example, In service requests, as of now, we can't extract temporal behavior of Service requests because the sequence of statuses is lost.

Anamolies

5% of the results were negative number of months while calculating the number of months a tenant stays in a Nestaway by taking move_in_date and move_out_date data from tenancy_histories table.

Clarification on accuracy of these data points are necessary as it might affect the accuracy of predictions.



Possible Missing Data

During exploration, the size of data points was inconsistent and these data points along with other suggestions are listed in the table

Data point	Remark	Number of Rows
utility_bills	Not updated	115
bathroom_checklists	Not updated	268
bedroom_checklists	Not updated	284
Distinct ids in rooms	Doesn't match with number of rooms in bed_furnishings	9551
more_data in bookings	Not a good practice to have a document in RDBMS	
nach_registrations	Does not have profile id for mapping to tenants	

Null Values

Table	Percentage Nulls	Percentage Nulls	Percentage Nulls
citrus_transaction_logs	move_in_date: 74%	actual_move_in_date: 95%	booking_id: 86%
city_living_costs	multiple city living costs for each city		
feedbacks	reason: 97%	comment: 77%	
beds	move_in_date: 93%		
citrus_transaction_logs	move_in_date: 74%	is_mattress: 100%	cupboard: 100%
houses	booking_type: 53%	tenant_type: 53%	house_group_id: 55%
room_furnishings	is_ac_point: 100%	ac_rent_included: 95%	
tenant_profiling_details	ideal_weekend: 90%	stay_length_city: 90%	
profiles	gender: 96%	marital_status: 96%	city: 95%
house_living_costs	links to city_living_costs which has multiple city living costs associated with it		
move_requests	reasons: 46%		
parkings	no relations with any user id or house id	two_wheeler_count: 100%	four_wheeler_count: 100%
rooms	area: 100%		
tenant_preferences	gender: 89%	qualification: 100%	

Completeness of data is important for getting a complete picture of Tenants. These are the gaps in the current dataset.

Data Recommendations

Having more of high quality data relevant to the task i.e. Tenant Churn prediction generally provides significant AI model performance gain.

Data requested: Tenants Profile data - age, gender, job related data
Reason: Can give more insights into reasons of churn

Data requested: Turn around time of Service requests
Reason: Churn behavior associated with it

Data requested: Tenant Preferences and religious beliefs
Reason: To learn tenant synergies with other tenants in the same Nestaway home

Data requested: App behaviour data like bookings, exploring other houses etc.
Reason: To learn intent of Tenant Moving out

Next steps

This preliminary data exploration forms the initial base work that is required before Marax AI platform is trained to make predictions for Nestaway.

Marax brings in other relevant external datasets that will be used along with Nestaway's data to make precise predictions and eventually personalized recommendations. This exploration also surfaces some challenges we have faced while analysing your data.

As a next step, while our team starts training models for Nestway - we would like to have a follow up conversation about Nestaway's data so that we can incorporate them in the most efficient way to train our AI models and make highly accurate Tenant churn predictions.