

# **Market Segmentation On Business Market**

To analyse Real Estate Market in India using Market Segmentation

**Contributors:**  
**Thejaswin S**  
**Ramniwash Kumar**  
**Sameer Sekhar Bashistha**

**GitHub Link :** <https://github.com/thejaswin123/Market-segmentation>

**0.Fermi estimation**

To analyse the Real Estate Market in India using Segmentation analysis and come up with a feasible strategy to enter the market.

### Overview:

Real estate sector is one of the most globally recognized sectors. It comprises four sub sectors - housing, retail, hospitality, and commercial. The growth of this sector is well complemented by the growth in the corporate environment and the demand for office space as well as urban and semi-urban accommodations. The construction industry ranks third among the 14 major sectors in terms of direct, indirect and induced effects in all sectors of the economy.

In India, the real estate sector is the second-highest employment generator, after the agriculture sector. It is also expected that this sector will incur more non-resident Indian (NRI) investment, both in the short term and the long term. Bengaluru is expected to be the most favoured property investment destination for NRIs, followed by Ahmedabad, Pune, Chennai, Goa, Delhi and Dehradun.

## **1. Data Sources (Data Collection)**

source

<https://prsindia.org/policy/analytical-reports/land-records-and-titles-india>

<https://www.magicbricks.com/Property-Rates-Trends/ALL-COMMERCIAL-rates-in-New-Delhi>

<https://rbidocs.rbi.org.in/rdocs/Publications/PDFs/RAPMSA07052015FL6A5F839BAB9B4D0190CE127D04DEF74C.PDF>

<https://www.ibef.org/industry/real-estate-india/infographic>

We have to Convert data from the above websites to csv format.

Importing the Dataset:

We will import the dataset that we need to use. Here, We import the All\_cities\_rental.csv file. It can be imported using the below code

```
df=pd.read_csv("All_cities_rental.csv")
```

## **2. Data Pre-processing (Steps and Libraries used)**

Importing Libraries: Firstly, we will import the libraries for our model, which is part of data pre-processing. The code is given below:

```
/*  
import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
from sklearn.decomposition import PCA  
from kmodes.kprototypes import KPrototypes  
import plotnine  
from IPython.display import display, HTML  
from operator import itemgetter  
from IPython.display import SVG  
import tensorflow as tf  
import os  
import plotly.graph_objects as go  
import time  
import datetime  
from datetime import date  
from plotly.offline import init_notebook_mode, iplot  
import plotly.express as px  
import matplotlib.gridspec as gridspec */
```

we have imported for the performing mathematics calculation.

- Matplotlib is for plotting the graph, and pandas is for managing the dataset.
- Seaborn is a data visualisation library, it is based on matplotlib.
- KModes clustering is one of the unsupervised Machine Learning algorithms that is used to cluster categorical variables. The k-modes algorithm extends the k-means paradigm to cluster categorical data by using
  - (1) a simple matching dissimilarity measure for categorical objects
  - (2) modes instead of means for clusters
  - (3) a frequency-based method to update modes in the k-means fashion to minimise the clustering cost function
- The k-prototypes algorithm, as described in “Clustering large data sets with mixed numeric and categorical values” by Huang (1997), is an extension of k-means for mixed data. The k-prototypes algorithm **combines k-modes and k-means** and is able to cluster mixed numerical / categorical data.
- The Plotly Python library is an interactive, open-source plotting library that supports over 40 unique chart types covering a wide range of statistical, financial, geographic, scientific, and 3-dimensional use-cases.

Libraries used:

- Numpy
- Pandas
- Matplotlib
- ScikitLearn
- Keras
- Seaborn
- Kmodes
- kprototypes

### 3. Data visualisation

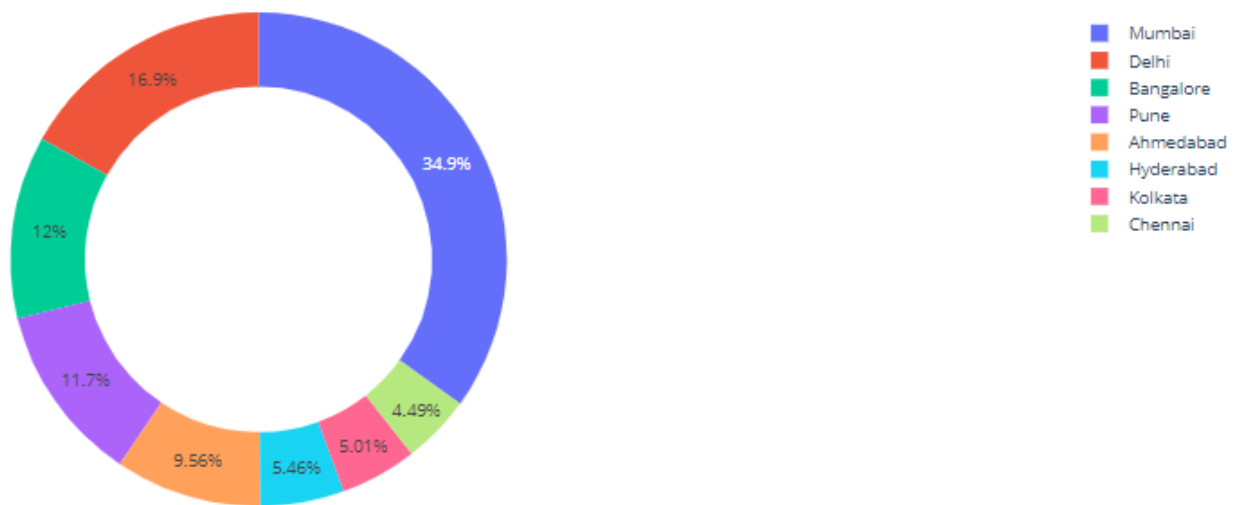
Correlation between the numerical features of the data

```
1 sns.heatmap(data.corr(), annot=True)
```

<AxesSubplot:>



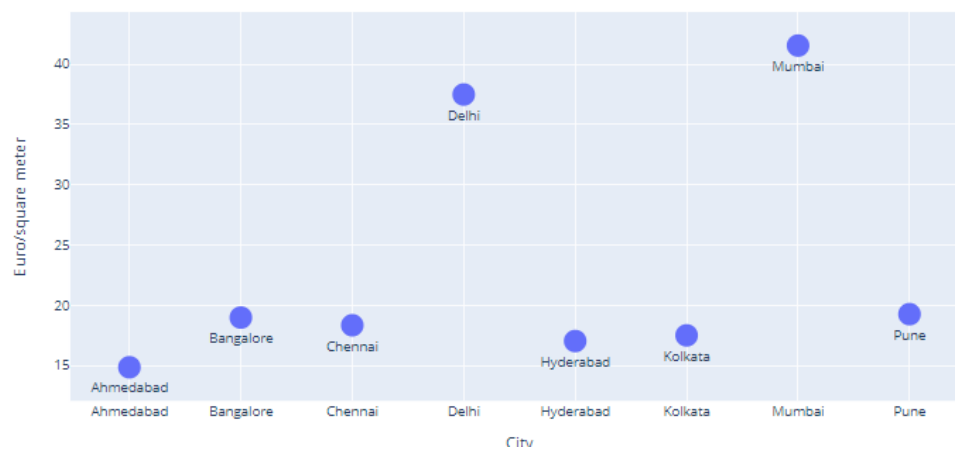
Pie chart of all the city ration distribution in the dataset



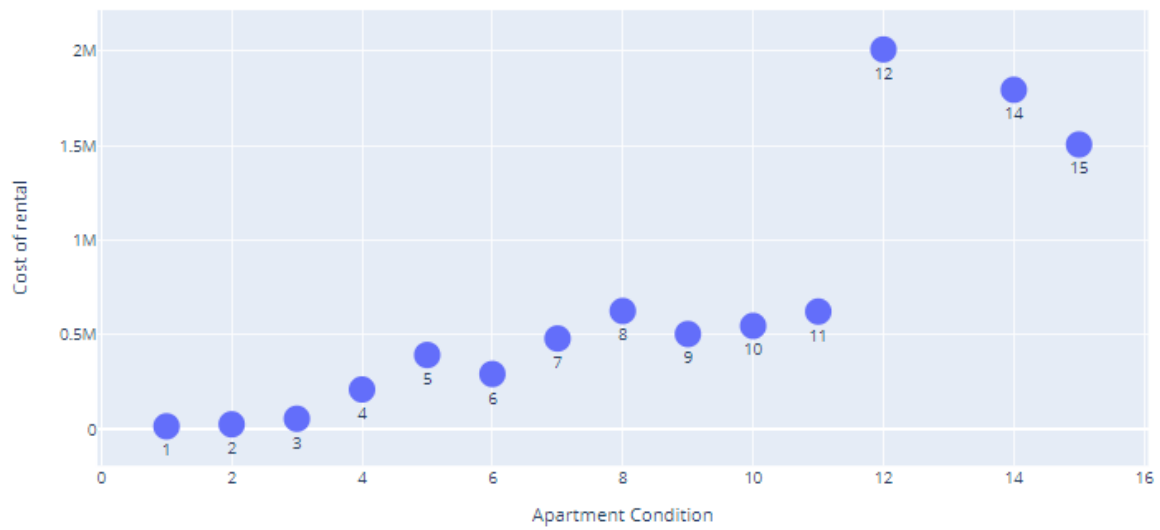
Average rental per month



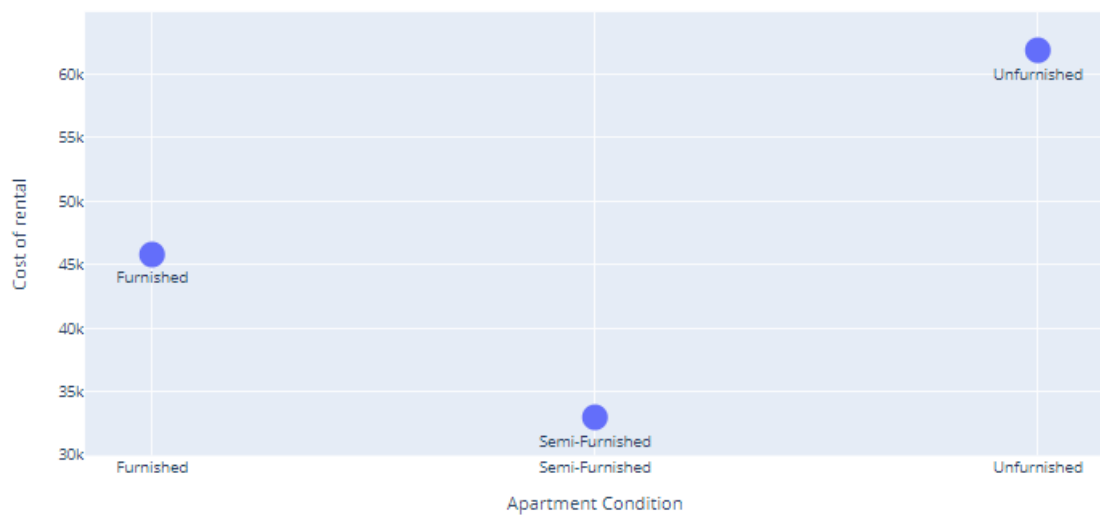
Average rental per month compare by area per square meter



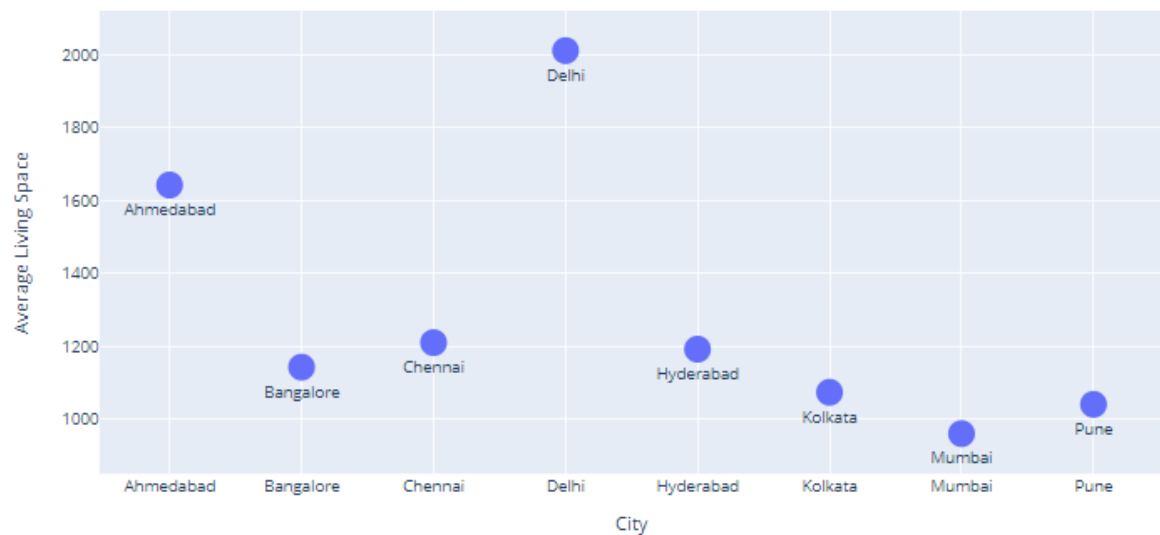
Average rental per month group by bedroom



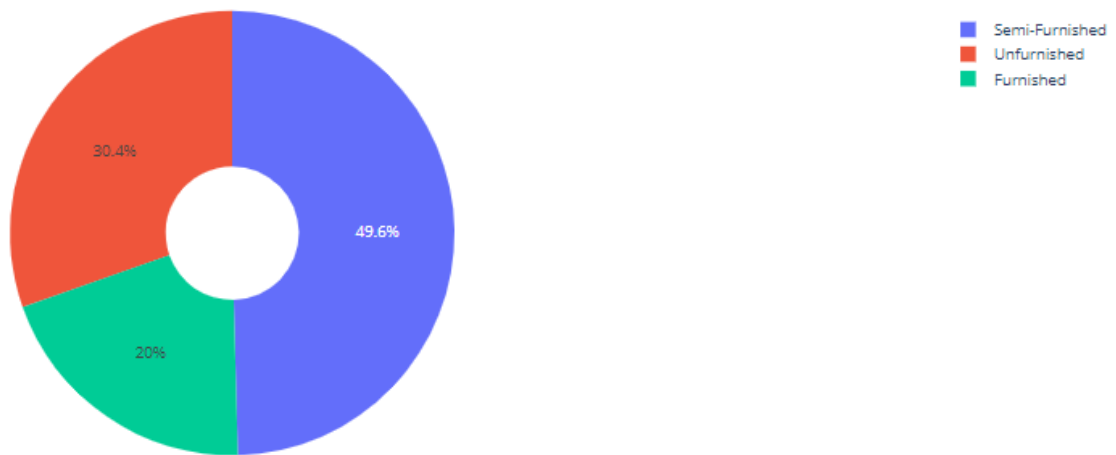
Average rental per month group by furnish\_type



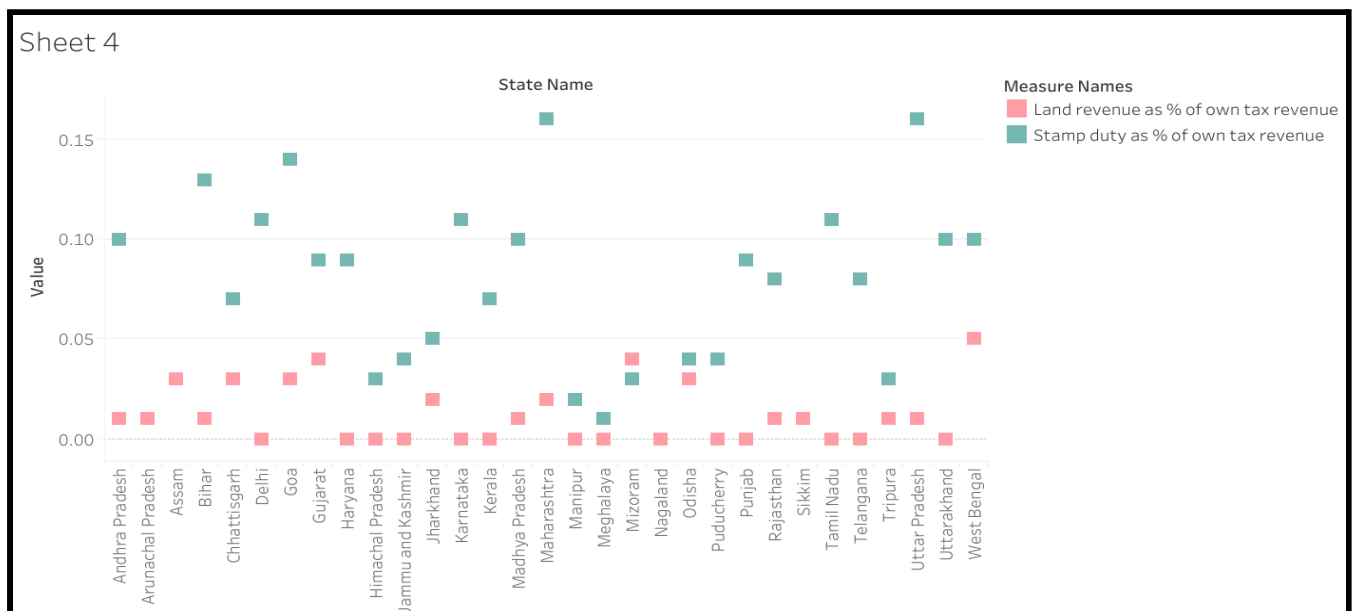
Average living space group by city



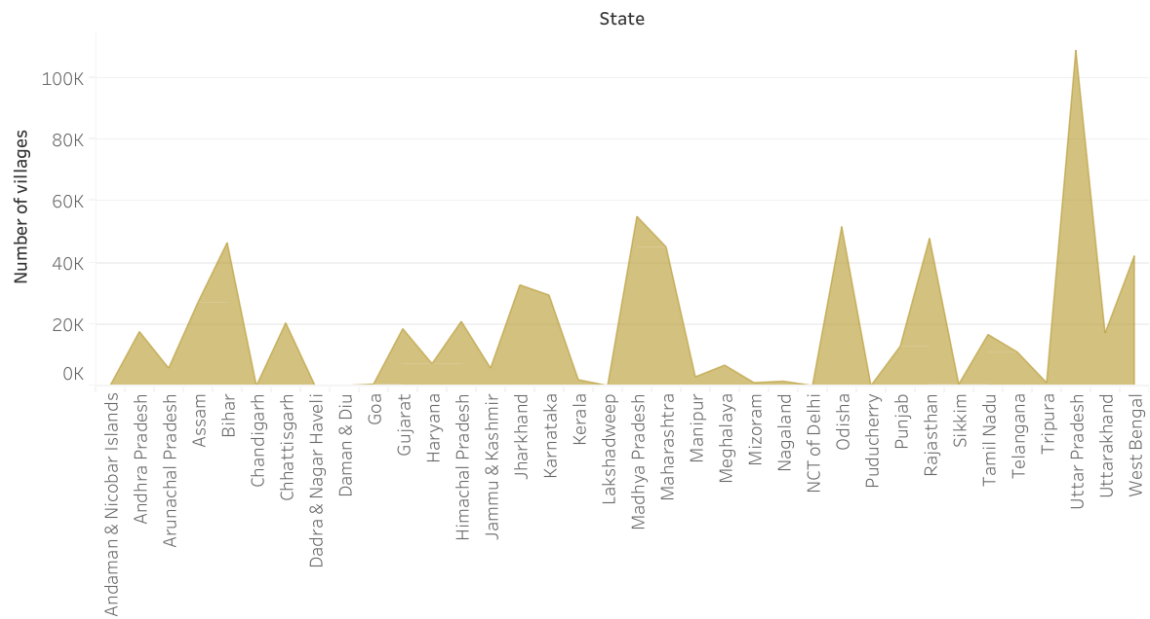
Percentage of the residence is newly constructed or not



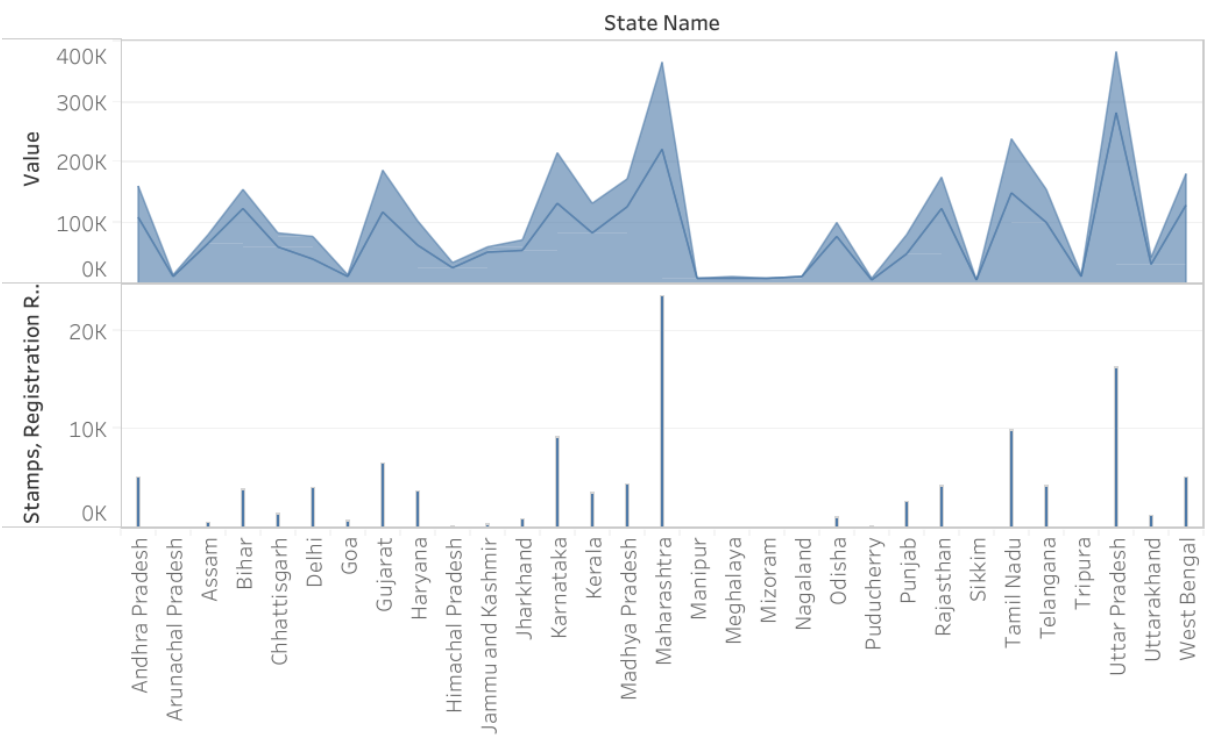
Visualisation of land revenue and Stamp duty of all the states



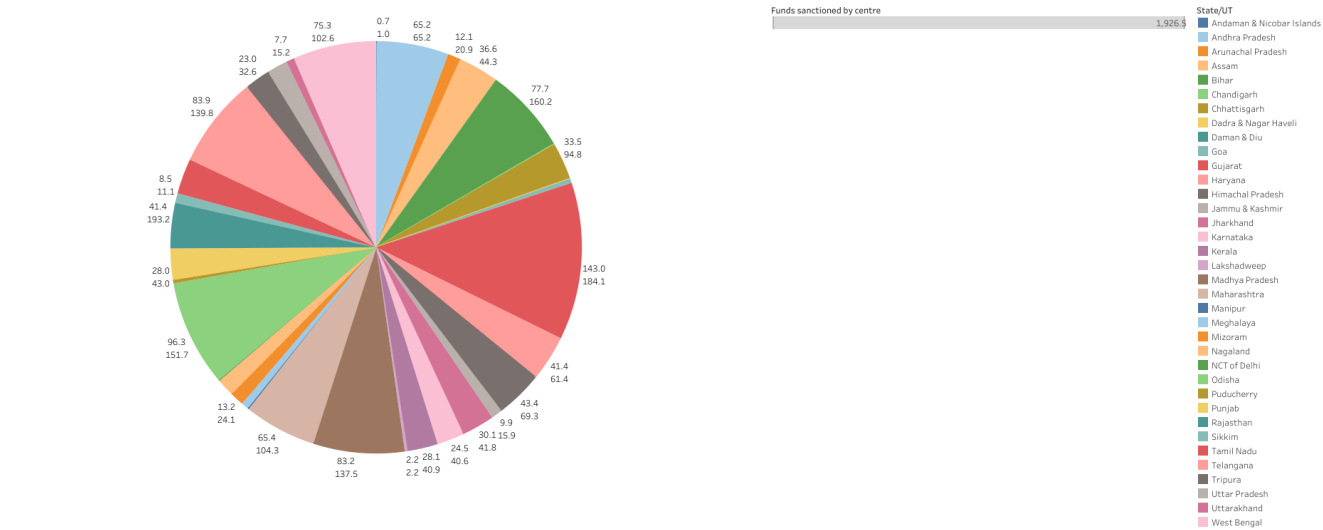
Sheet 1



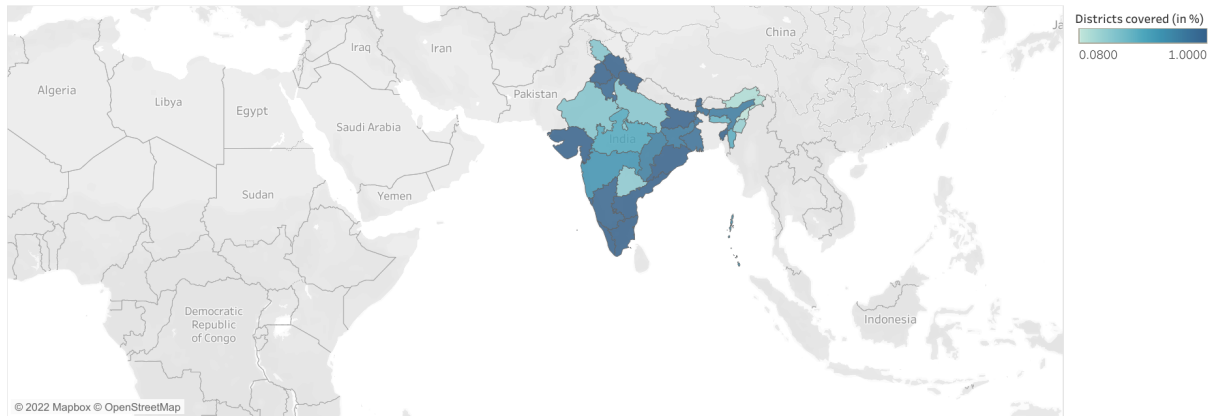
Sheet 3



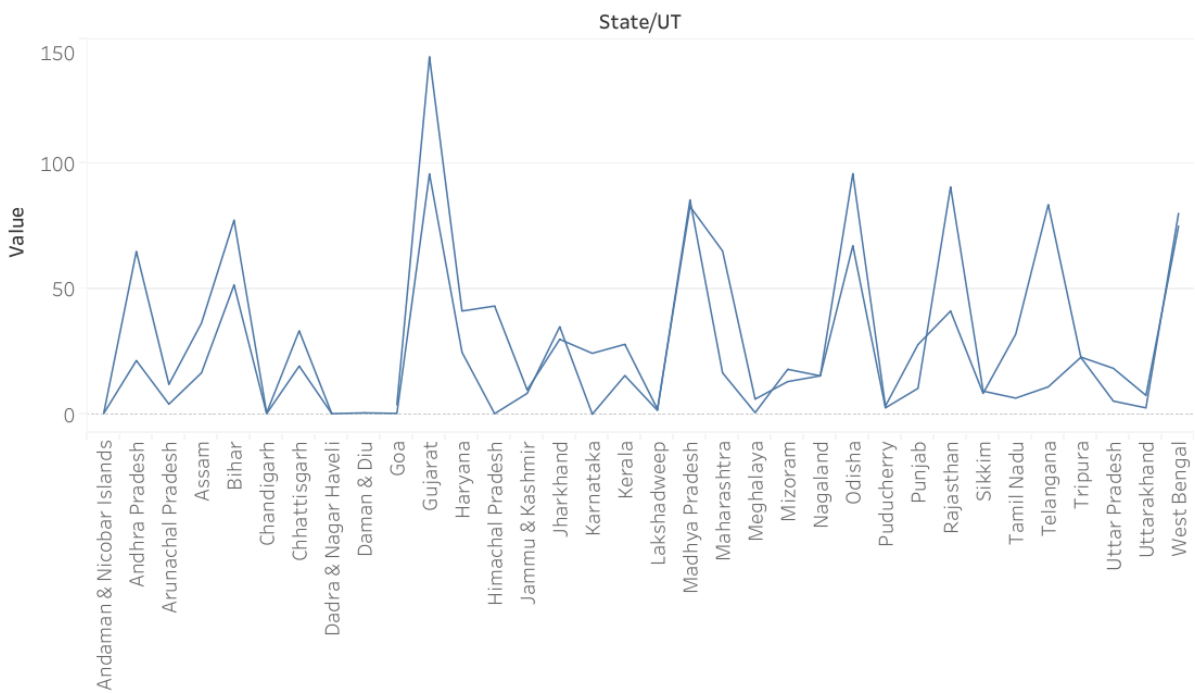




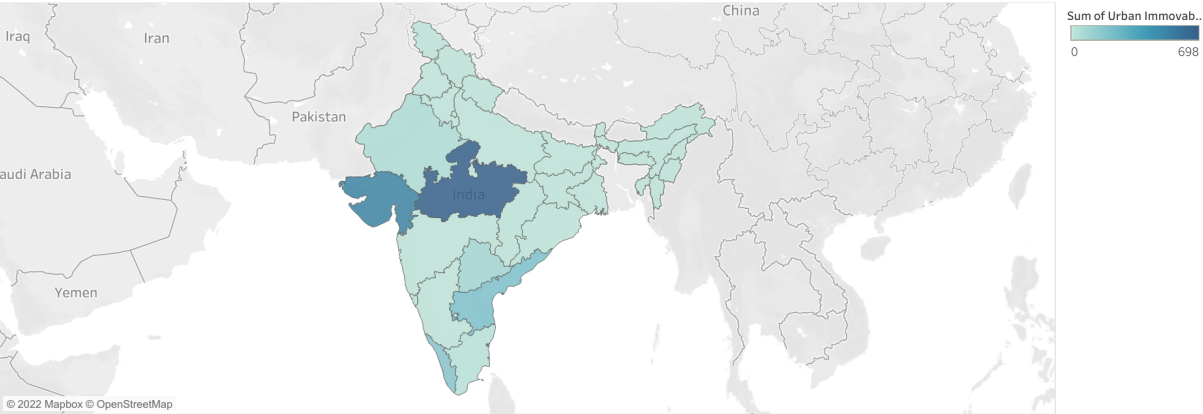
Sheet 3



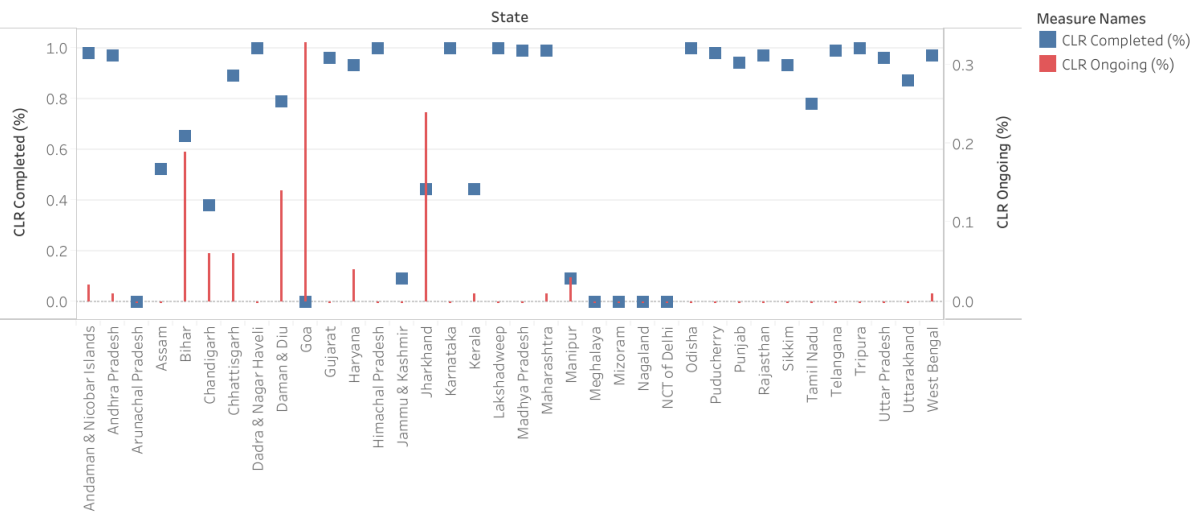
Sheet 4



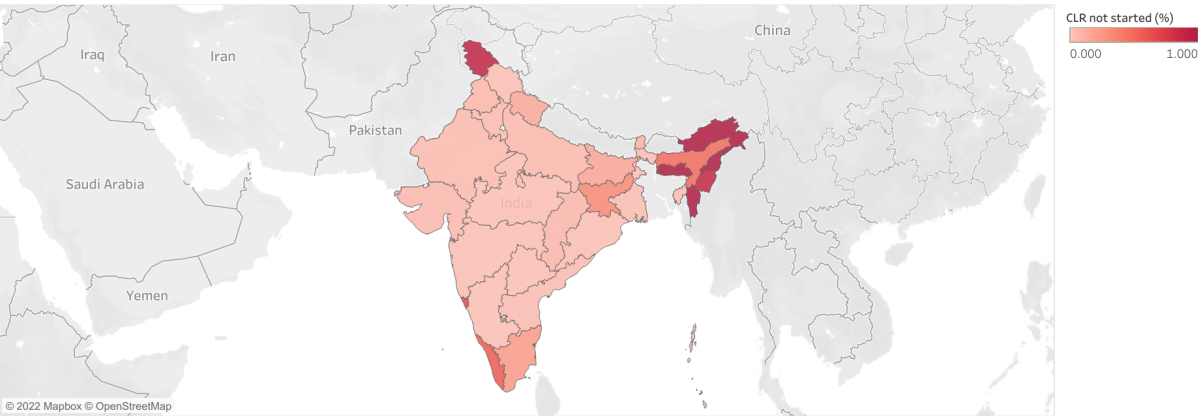
Sheet 1



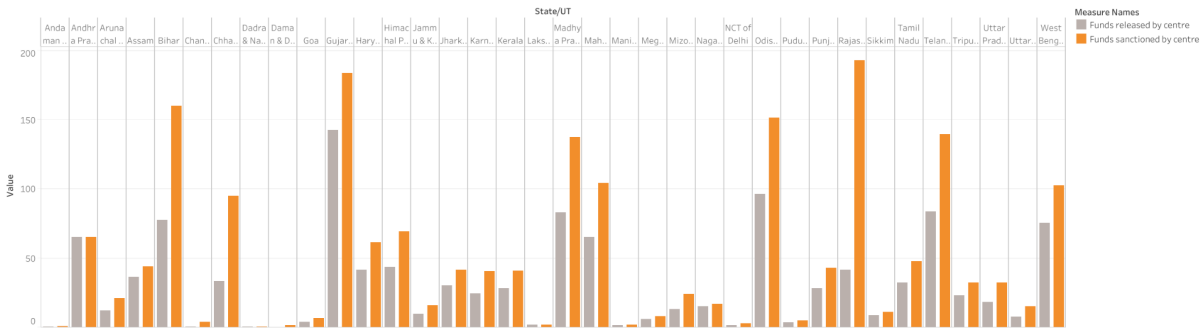
Sheet 3



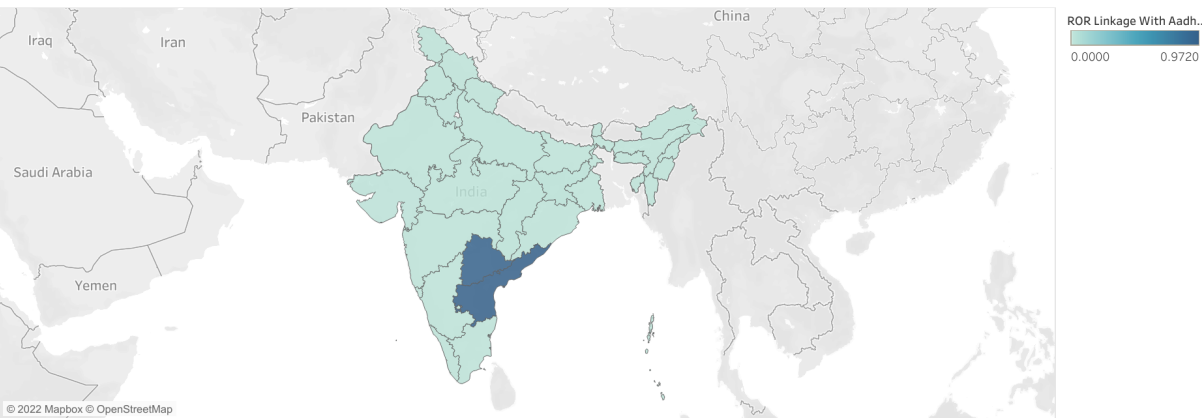
Sheet 4



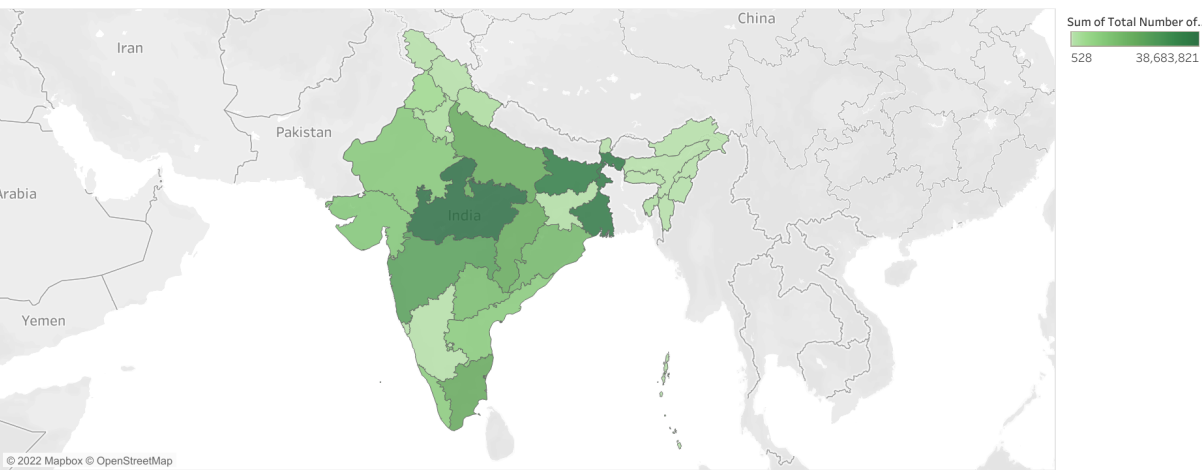
Sheet 1



Sheet 2



Sheet 1



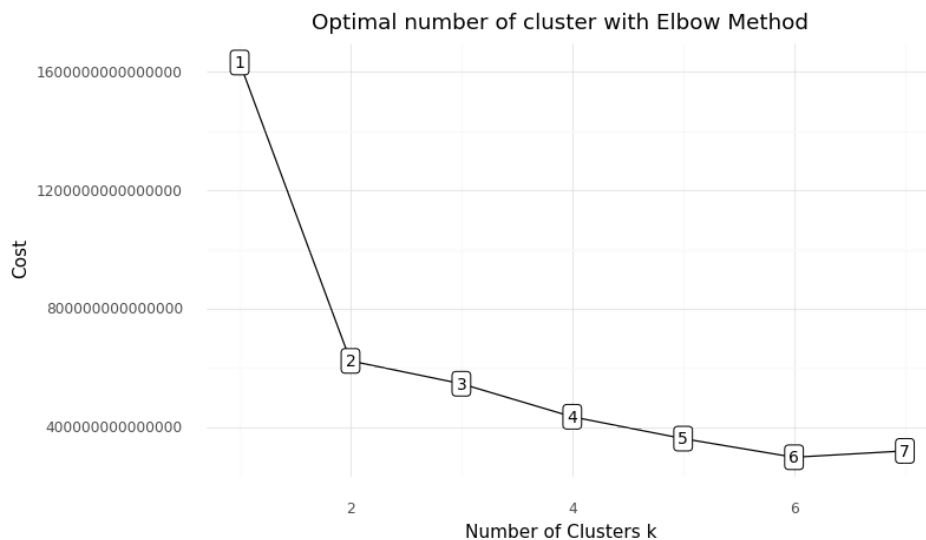
## 4. Segment Extraction (ML techniques used)

### K-Prototype Clustering

We considered the k-prototypes algorithm which is a mixture of k-means and k-modes. The reason to use the k-prototypes algorithm was that it can handle both numerical and categorical features simultaneously and the data had most of the columns categorical.

The k-prototypes algorithm combines the “**means**” of the numerical part and the “**modes**” of the categorical part to build a new hybrid Cluster Center “**prototype**”. Based on the “prototype”, it builds a Dissimilarity Coefficient formula and the Cost Function applicable to the mixed-type data.

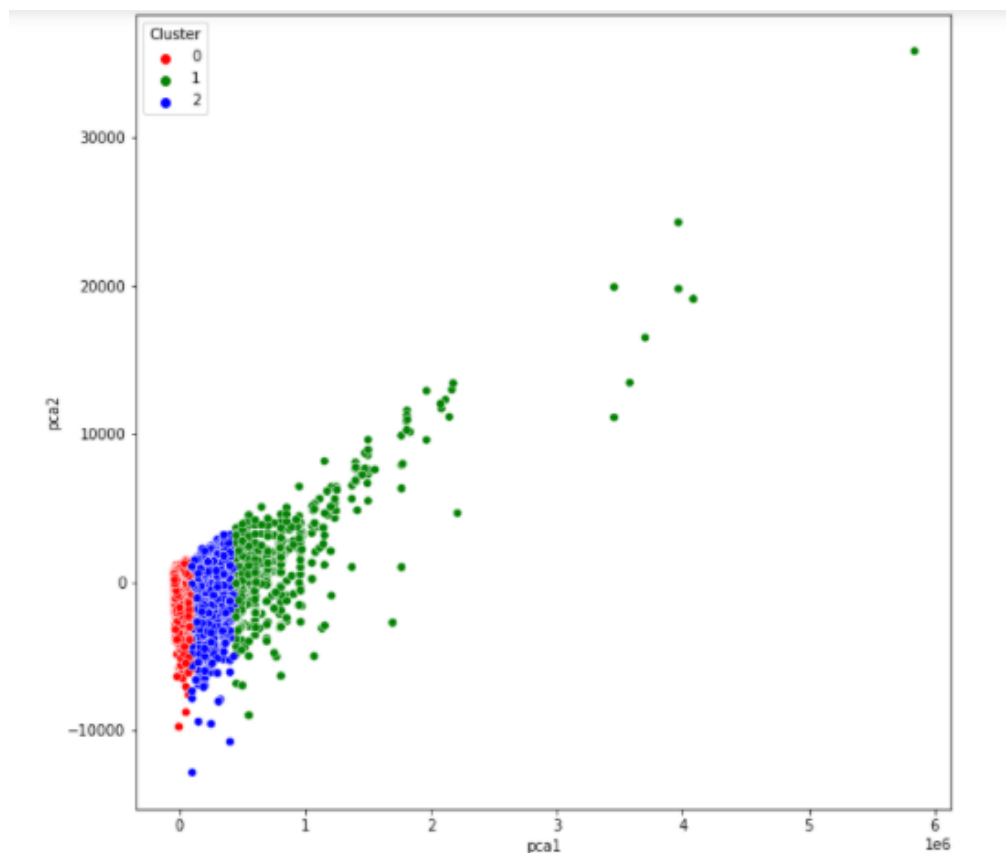
We are using the Elbow method to determine the optimal number of clusters for K-Prototype clusters. Instead of calculating the within the sum of squares errors (WSSE) with Euclidian distance, K-Prototype provides the cost function that combines the calculation for numerical and categorical variables. We can look into the Elbow to determine the optimal number of clusters.



## 5. Profiling and describing potential segments:

By segmenting the dataset using the K-Prototype algorithm we have found that the ideal number of cluster is 3.

The visualisation of the cluster is done by PCA (Principal Component Analysis) which is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

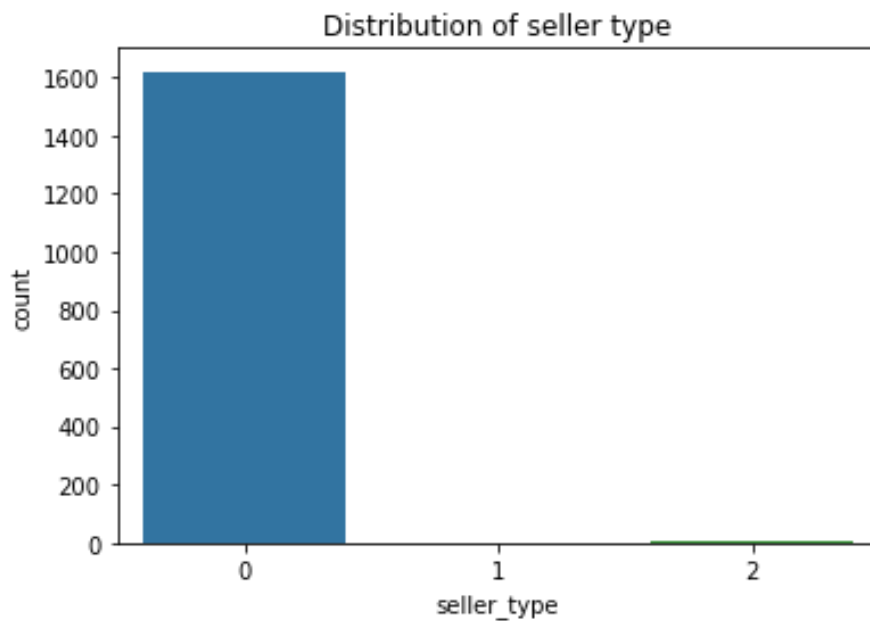


The red Cluster (0) can be targeted to the people who consider buying from an agent and want property with unfurnished BHK.

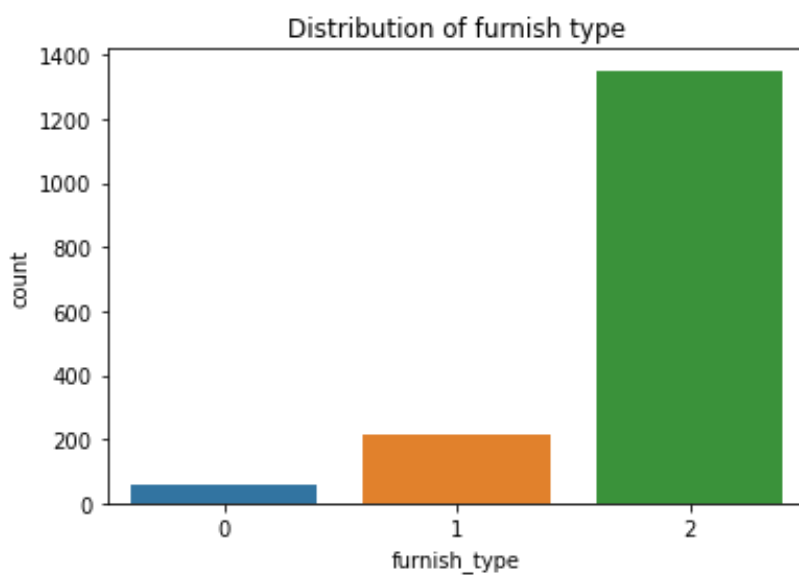
The green cluster (1) can be targeted to the people who want to buy the property from agents and the owner themselves where they need an Apartment and Independent floor type of property.

The blue cluster (3) can be targeted to the people who need semi-furnished type property with less number of bathrooms and bedrooms.

### Cluster 0 (red)

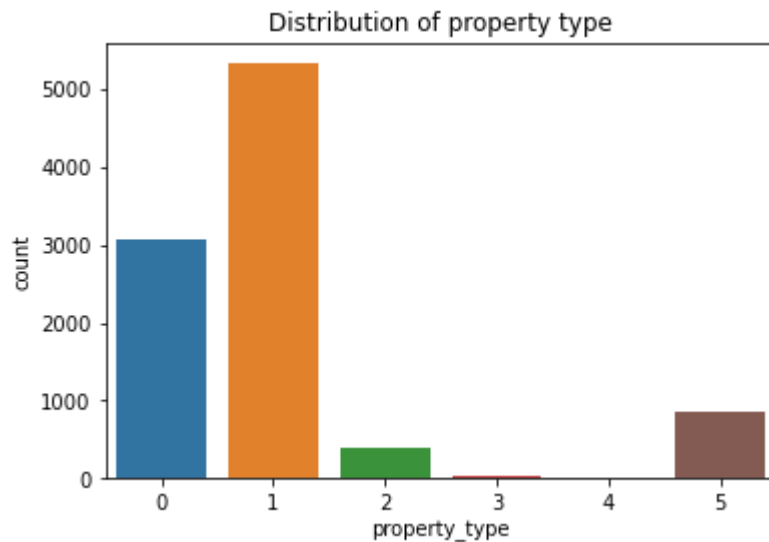


Maximum property in this cluster are of seller type 0 i.e They are being sold by Agents.



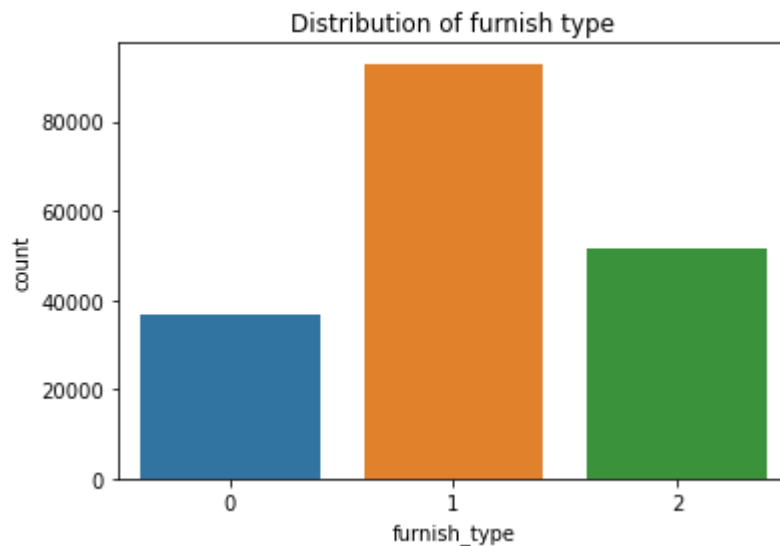
Here maximum number of property are unfurnished type.

### Cluster 1(Green)



Almost 90 percent of the total property in this cluster are of property type 0 & 1. I.e Apartment and Independent floor

### Cluster 2 (Blue)



**Almost 50 percent of the total property are of semi-furnished type in this cluster.**

### Visualising demographics

Demographic segmentation is a market segmentation technique where an organisation's target market is segmented based on demographic variables such as age, gender, education, income, etc. It helps organisations understand who their customers are so that their needs can be addressed more effectively.

Instead of reaching an entire market, companies can use demographic segmentation to focus their time and resources on those segments that have customers who are most likely to make purchases, and are therefore most valuable to them.

Here are some segments of audience to target based on demographics.

### **Students**

- Students look for rental properties that give them the opportunity to live with many people in order to save costs. They also tend to look for the houses near the nightlife and shops.

### **Qualified and working professionals**

- Generally, they look for at least one plus bedroom properties and reasonable value for money.
- As they are more mature than students and in full-time employment, there is a level of reliability and security renting to this target group.

### **Families**

- This type of renter tends to need two or more bedrooms, outdoor area, parking space and storage unit for belongings. They are happy with unfurnished homes as they have their own furniture.
- They are extremely stable and generally rent for long-term periods.

## **6. Customising the Market Mix**

A marketing mix includes multiple areas of focus as part of a comprehensive marketing plan. The term often refers to a common classification that began as the four Ps: **product, price, placement, and promotion**.

**Product:** One of the key decisions an organisation needs to make when developing the product dimension of the marketing mix, is to specify the product in view of customer needs. Often this does not imply designing an entirely new product, but rather modifying an existing one. Other marketing mix decisions that fall under the



product dimension are: naming the product, packaging it, offering or not offering warranties, and after sales support services.

**Price:** Typical decisions an organisation needs to make when developing the price dimension of the marketing mix include setting the price for a product, and deciding on discounts to be offered.

**Placement:** The key decision relating to the place dimension of the marketing mix is how to distribute the product to the customers

**Promotion:** Typical promotion decisions that need to be made when designing a marketing mix include: developing an advertising message that will resonate with the target market, and identifying the most effective way of communicating this message. Other tools in the promotion category of the marketing mix include public relations, personal selling, and sponsorship.

## **Early market entry in accordance with Innovation Adoption Life:**

The **technology adoption lifecycle** is a sociological model that describes the adoption or acceptance of a new product or innovation, according to the demographic and psychological characteristics of defined adopter groups. The process of adoption over time is typically illustrated as a classical normal distribution or "bell curve". The model indicates that the first group of people to use a new product is called "innovators", followed by "early adopters". Next come the **early majority** and **late majority**, and the last group to eventually adopt a product are called "Laggards" or "phobics." For example, a phobic may only use a cloud service when it is the only remaining method of performing a required task, but the phobic may not have an in-depth technical knowledge of how to use the service.

Occupation and family status are some of the main demographic segments which directly influence the online property rental service market.

Similarly Lifestyle, Social Status are some of the main psychological factors which influence the online property rental service market.

The maximum tenants are the people who are working professionals and want to live near the area where they work. Therefore the best location in India which is most suitable to create the early market will be the places like suburb, because it will cost much less for them to live without sacrificing time to travel to their job location. Also the area where it is safe to live.