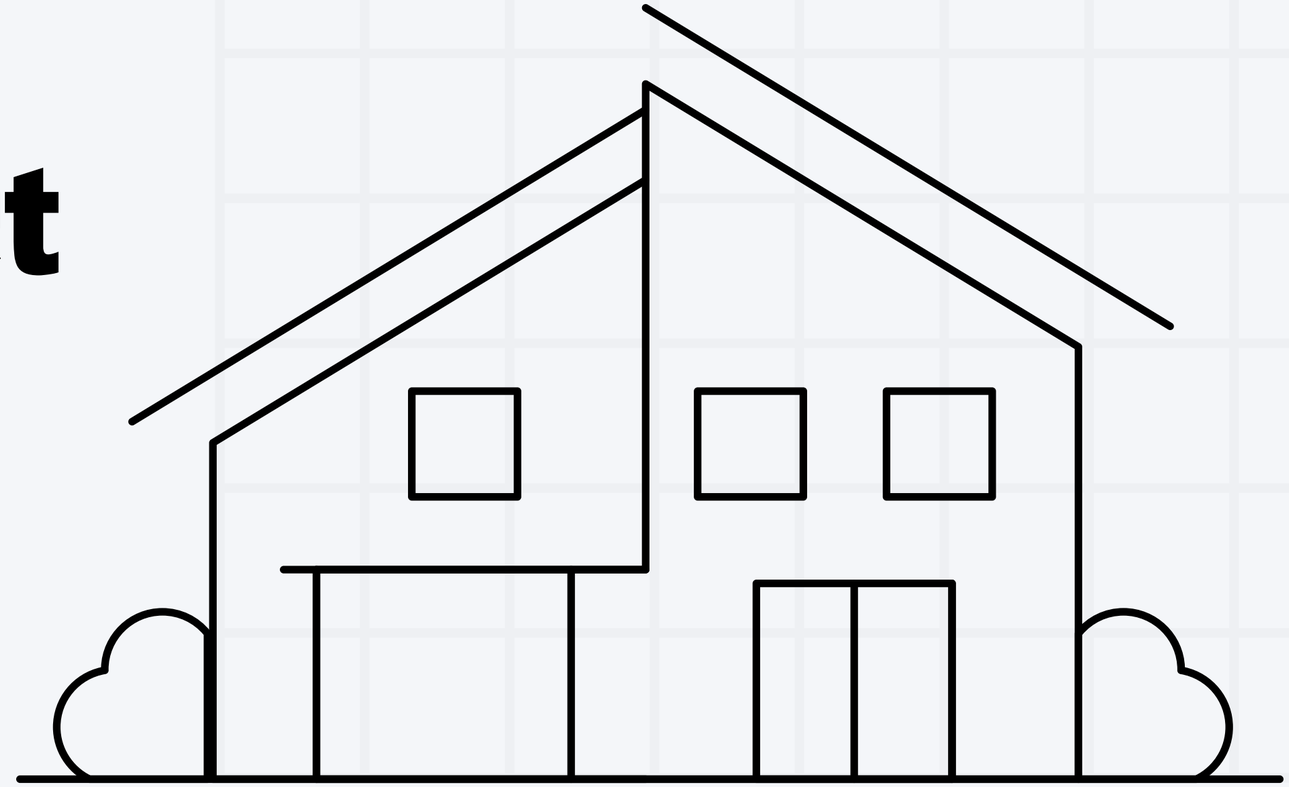
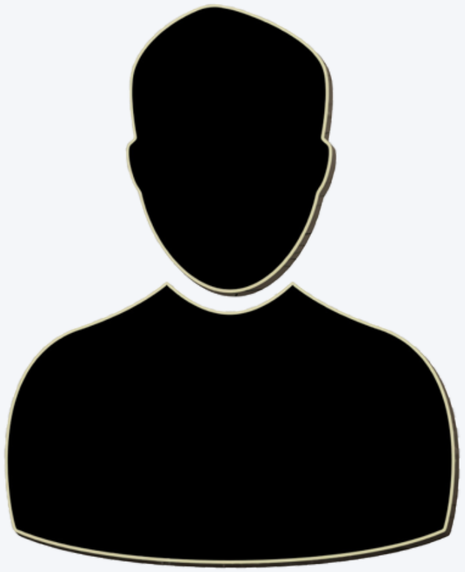
Final Project

Presented by Ranger Merah

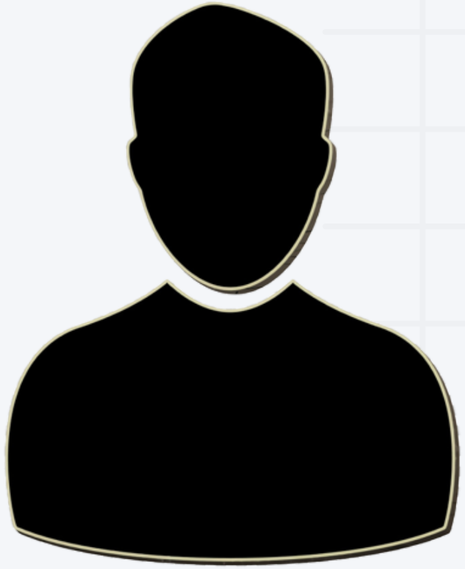


Our Team

Yulia Dwi Rahmawati



Labib Umam Alfaruq



Ahmad Fahmi Abdurrahman

Content

- Problem Understanding
- Dataset Understanding
- Exploratory Data Analyst
- Data Preparation
- Model
- Deployment

Latar Belakang

Perumahan di India bervariasi, mulai dari istana para maharaja, gedung apartemen modern di kota-kota besar, hingga rumah petak di desa-desa terpencil. Ada pertumbuhan luar biasa di sektor perumahan India seiring dengan meningkatnya pendapatan.

Menyewa adalah perjanjian di mana pembayaran dilakukan untuk penggunaan sementara atas suatu barang, jasa, atau properti milik orang lain. Sewa kotor adalah ketika penyewa membayar jumlah sewa tetap dan tuan tanah membayar semua biaya properti yang biasanya dikeluarkan oleh kepemilikan. Menyewa bisa menjadi contoh ekonomi berbagi.

Data Understanding

	Posted On	BHK	Rent	Size	Floor	Area Type	Area Locality	City	Furnishing Status	Tenant Preferred	Bathroom	Point of Contact
0	2022-05-18	2	10000	1100	Ground out of 2	Super Area	Bandel	Kolkata	Unfurnished	Bachelors/Family	2	Contact Owner
1	2022-05-13	2	20000	800	1 out of 3	Super Area	Phool Bagan, Kankurgachi	Kolkata	Semi-Furnished	Bachelors/Family	1	Contact Owner
2	2022-05-16	2	17000	1000	1 out of 3	Super Area	Salt Lake City Sector 2	Kolkata	Semi-Furnished	Bachelors/Family	1	Contact Owner
3	2022-07-04	2	10000	800	1 out of 2	Super Area	Dumdum Park	Kolkata	Unfurnished	Bachelors/Family	1	Contact Owner
4	2022-05-09	2	7500	850	1 out of 2	Carpet Area	South Dum Dum	Kolkata	Unfurnished	Bachelors	1	Contact Owner

- BHK: Number of Bedrooms, Hall, Kitchen.
- Rent: Rent of the Houses/Apartments/Flats.
- Size: Size of the Houses/Apartments/Flats in Square Feet.
- Floor: Houses/Apartments/Flats situated in which Floor and Total Number of Floors (Example: Ground out of 2, 3 out of 5, etc.)
- Area Type: Size of the Houses/Apartments/Flats calculated on either Super Area or Carpet Area or Build Area.
- Area Locality: Locality of the Houses/Apartments/Flats.
- City: City where the Houses/Apartments/Flats are Located.
- Furnishing Status: Furnishing Status of the Houses/Apartments/Flats, either it is Furnished or Semi-Furnished or Unfurnished.
- Tenant Preferred: Type of Tenant Preferred by the Owner or Agent.
- Bathroom: Number of Bathrooms.
- Point of Contact: Whom should you contact for more information regarding the Houses/Apartments/Flats.

Exploratory Data Analysis

```
df.info()
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4746 entries, 0 to 4745
Data columns (total 12 columns):
 #   Column             Non-Null Count  Dtype 
---  ------             --------------  ----- 
 0   Posted On          4746 non-null   object
 1   BHK                4746 non-null   int64 
 2   Rent               4746 non-null   int64 
 3   Size               4746 non-null   int64 
 4   Floor              4746 non-null   object
 5   Area Type          4746 non-null   object
 6   Area Locality      4746 non-null   object
 7   City               4746 non-null   object
 8   Furnishing Status  4746 non-null   object
 9   Tenant Preferred   4746 non-null   object
 10  Bathroom           4746 non-null   int64 
 11  Point of Contact   4746 non-null   object
dtypes: int64(4), object(8)
memory usage: 445.1+ KB
```

```
df.duplicated().sum()
```

✓ 0.0s

```
0
```

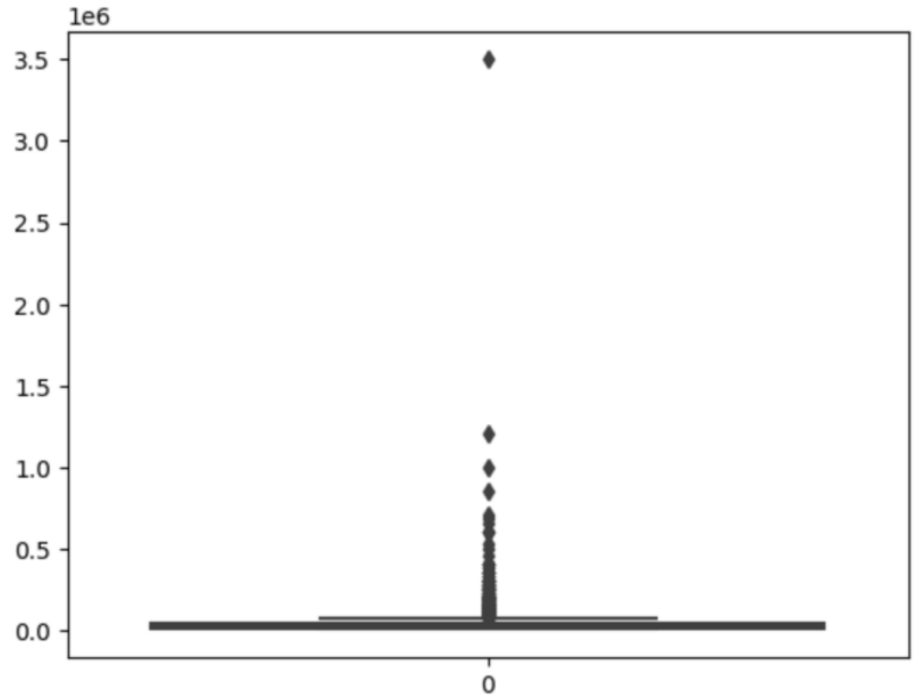
```
df.describe()
```

✓ 0.0s

	BHK	Rent	Size	Bathroom
count	4746.000000	4.746000e+03	4746.000000	4746.000000
mean	2.083860	3.499345e+04	967.490729	1.965866
std	0.832256	7.810641e+04	634.202328	0.884532
min	1.000000	1.200000e+03	10.000000	1.000000
25%	2.000000	1.000000e+04	550.000000	1.000000
50%	2.000000	1.600000e+04	850.000000	2.000000
75%	3.000000	3.300000e+04	1200.000000	2.000000
max	6.000000	3.500000e+06	8000.000000	10.000000

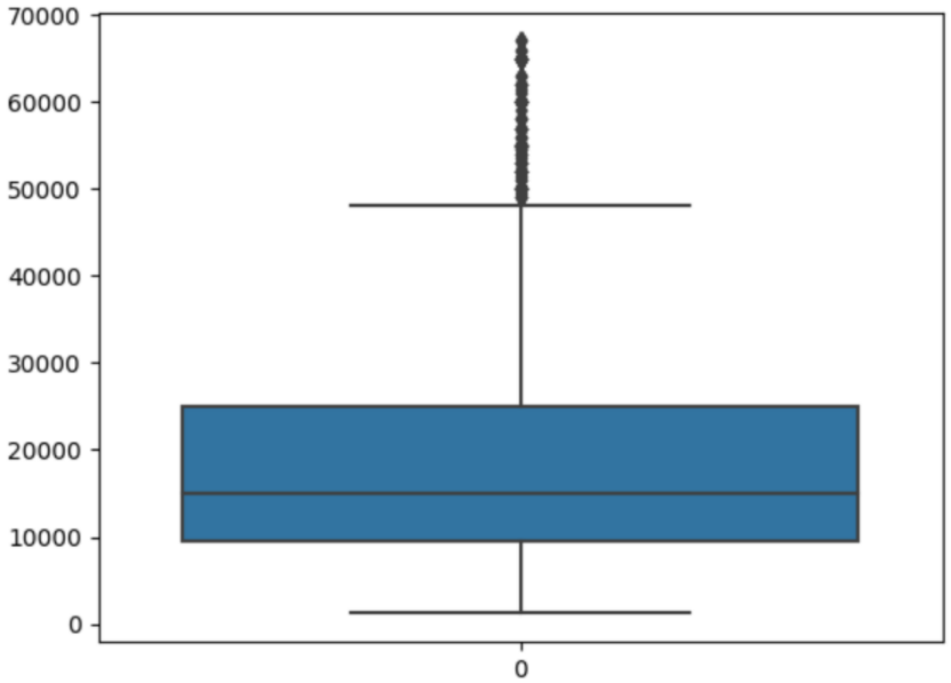
Outlier Identification

```
fig1 = sns.boxplot(df['Rent'])
```
✓ 0.1s



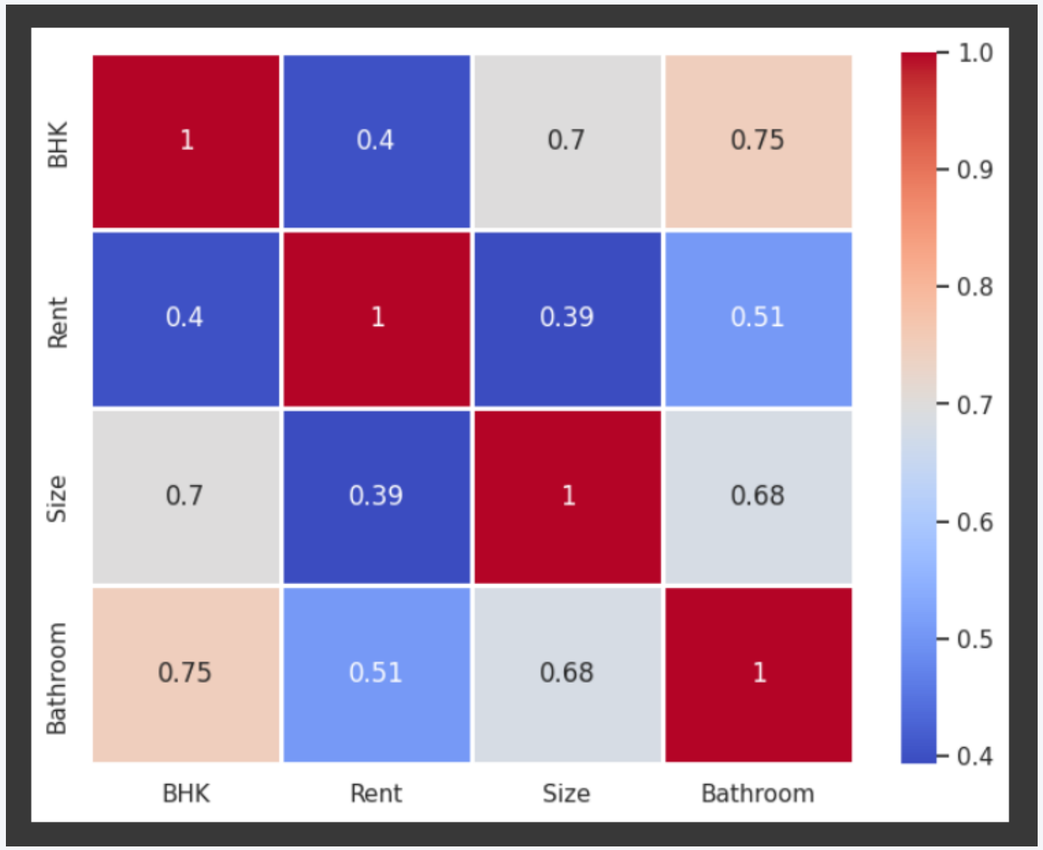

```
persentile25 = df['Rent'].quantile(0.25)
persentile75 = df['Rent'].quantile(0.75)
IQR = persentile75-persentile25
Upper = persentile75+1.5*IQR
Lower = persentile25-1.5*IQR
```
✓ 0.0s

```
fig1 = sns.boxplot(df_new['Rent'])
```
✓ 0.0s



Correlation Data

```
plt.figure(figsize=(8,6))
ax = sns.heatmap(df_new.corr(), cmap = "coolwarm", annot=True, linewidth=2)
```



Data Preprocessing

```
[24] def feature_floor_split(r: pd.core.series.Series):
         if ' out of ' in r['Floor']:
             floor, max_floors = r['Floor'].split(' out of ')
             return [
                 -1 if 'Basement' in floor else int(floor)
                 , int(max_floors)
             ]
         # This accounts for the values where instead of having a floor like "1 out of 3", it just has "1".
         return [int(r['Floor'].split(' out of ')[0]), None]

     def feature_is_basement(r: pd.core.series.Series) -> str:
         if 'Basement' in r['Floor']:
             return 1
         return 0
```

```
df_new['Floor'] = df_new['Floor'].str.replace('Ground', '0')
df_new
```

✓ 0.0s

	Posted On	BHK	Rent	Size	Floor	Area Type	Area Locality	City	Furnishing Status	Tenant Preferred	Bathroom	Point of Contact
0	2022-05-18	2	10000	1100	0 out of 2	Super Area	Bandel	Kolkata	Unfurnished	Bachelors/Family	2	Contact Owner
1	2022-05-13	2	20000	800	1 out of 3	Super Area	Phool Bagan, Kankurgachi	Kolkata	Semi-Furnished	Bachelors/Family	1	Contact Owner
2	2022-05-16	2	17000	1000	1 out of 3	Super Area	Salt Lake City Sector 2	Kolkata	Semi-Furnished	Bachelors/Family	1	Contact Owner
3	2022-07-04	2	10000	800	1 out of 2	Super Area	Dumdum Park	Kolkata	Unfurnished	Bachelors/Family	1	Contact Owner
4	2022-05-09	2	7500	850	1 out of 2	Carpet Area	South Dum Dum	Kolkata	Unfurnished	Bachelors	1	Contact Owner

```
df_new['Basement'] = df_new.apply(lambda row: feature_is_basement(row), axis=1)
df_new['Floor Number'] = df_new.apply(lambda row: feature_floor_split(row)[0], axis=1)
df_new['Building Floor'] = df_new.apply(lambda row: feature_floor_split(row)[1], axis=1)
df_new.head(5)
```

✓ 0.0s Python

	Posted On	BHK	Rent	Size	Floor	Area Type	Area Locality	City	Furnishing Status	Tenant Preferred	Bathroom	Point of Contact	Basement	Floor Number	Building Floor
0	2022-05-18	2	10000	1100	0 out of 2	Super Area	Bandel	Kolkata	Unfurnished	Bachelors/Family	2	Contact Owner	0	0	2.0
1	2022-05-13	2	20000	800	1 out of 3	Super Area	Phool Bagan, Kankurgachi	Kolkata	Semi-Furnished	Bachelors/Family	1	Contact Owner	0	1	3.0
2	2022-05-16	2	17000	1000	1 out of 3	Super Area	Salt Lake City Sector 2	Kolkata	Semi-Furnished	Bachelors/Family	1	Contact Owner	0	1	3.0
3	2022-07-04	2	10000	800	1 out of 2	Super Area	Dumdum Park	Kolkata	Unfurnished	Bachelors/Family	1	Contact Owner	0	1	2.0
4	2022-05-09	2	7500	850	1 out of 2	Carpet Area	South Dum Dum	Kolkata	Unfurnished	Bachelors	1	Contact Owner	0	1	2.0

```
[27] df_new = df_new.drop(columns = ['Posted On', 'Floor'])
     df_new.isnull().sum()
```

```
BHK                 0
Rent                0
Size                0
Area Type           0
Area Locality       0
City                0
Furnishing Status   0
Tenant Preferred    0
Bathroom            0
Point of Contact    0
Basement            0
Floor Number        0
Building Floor      4
dtype: int64
```

```
[28] df_new['Building Floor'] = df_new['Building Floor'].fillna(0)
     df_new['Building Floor'].value_counts()
```

```
4.0     893
3.0     884
2.0     852
5.0     407
1.0     322
7.0     147
6.0      86
8.0      69
14.0     51
10.0     47
20.0     45
12.0     45
22.0     32
```

Label Encode

```
tmp_data = df_new.copy()
from sklearn.preprocessing import LabelEncoder
cols = ('Area Type', 'Area Locality', 'City','Furnishing Status','Tenant Preferred','Point of Contact')
encoders = {}

for c in cols:
    lbl = LabelEncoder()
    lbl.fit(list(tmp_data[c].values))
    tmp_data[c] = lbl.transform(list(tmp_data[c].values))
    encoders[c] = lbl
```

✓ 0.0s

```
[30] tmp_data
```

	BHK	Rent	Size	Area Type	Area Locality	City	Furnishing Status	Tenant Preferred	Bathroom	Point of Contact	Basement	Floor Number	Building Floor
0	2	10000	1100	2	196	4	2	1	2	2	0	0	2.0
1	2	20000	800	2	1359	4	1	1	1	2	0	1	3.0
2	2	17000	1000	2	1568	4	1	1	1	2	0	1	3.0
3	2	10000	800	2	473	4	2	1	1	2	0	1	2.0
4	2	7500	850	1	1685	4	2	0	1	2	0	1	2.0

Model

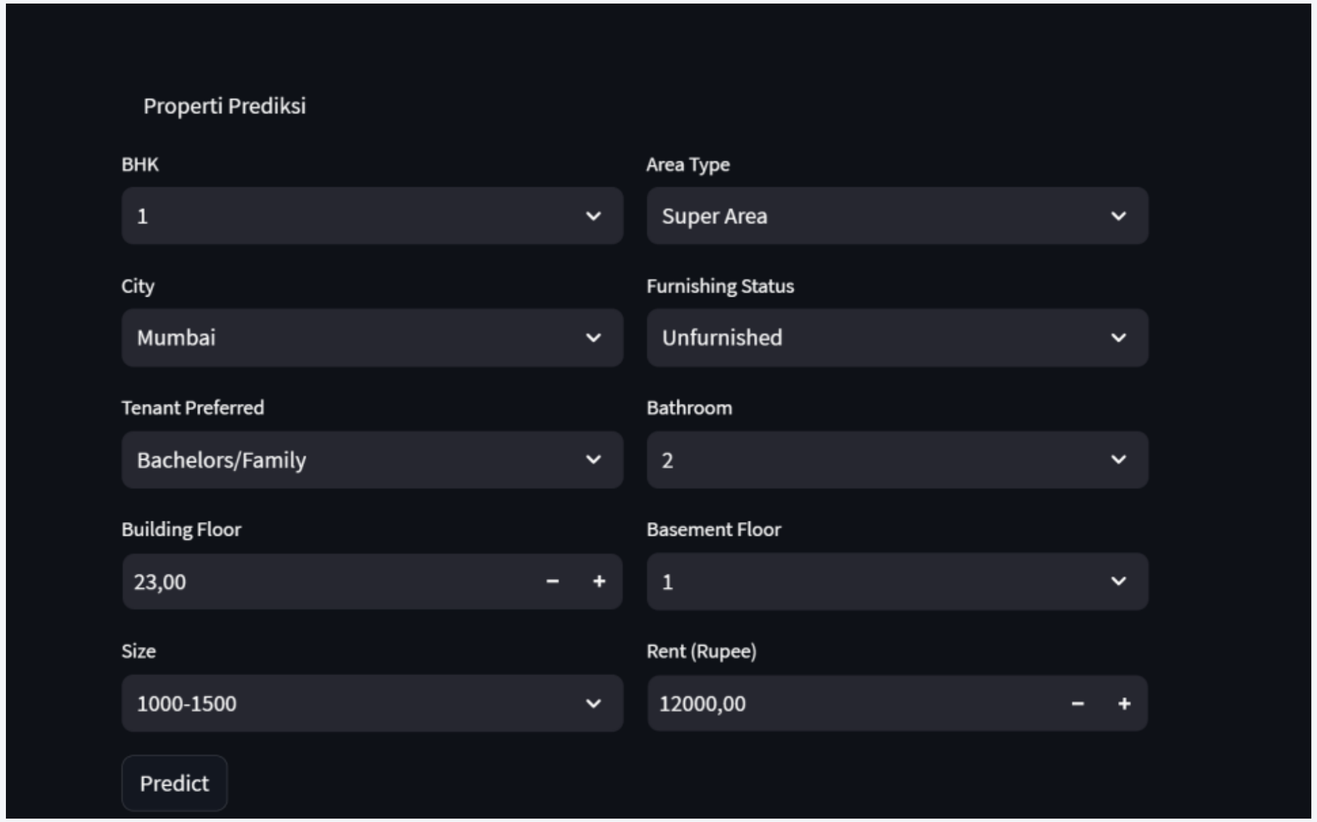
```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

np.random.seed(10)

# splitting data 80:20
X_train, X_test, y_train, y_test = train_test_split(
    tmp_data.loc[:, tmp_data.columns != 'Rent']
    , tmp_data['Rent']
    , test_size=0.2
    , random_state=1
)
```

```
from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn import metrics

KNN = KNeighborsClassifier(n_neighbors=3)
RFC = RandomForestClassifier (n_estimators = 7, criterion = 'entropy', random_state = 7)
SVC = SVC()
LR = LogisticRegression()

for clf in (RFC, KNN, SVC, LR):
    clf.fit(X_train, y_train)
    y_pred = clf.predict(X_train)
    print ('Accuracy Score Of ',
           clf.__class__.__name__,
           "=", 100*metrics.accuracy_score(y_train,
                                           y_pred))
```

✓ 8.0s

```
Accuracy Score Of  RandomForestClassifier = 96.18343195266273
Accuracy Score Of  KNeighborsClassifier = 37.721893491124256
Accuracy Score Of  SVC = 8.49112426035503
Accuracy Score Of  LogisticRegression = 6.7455621301775155
```

Deployment


Properti Prediksi
BHK
1
Area Type
Super Area
City
Mumbai
Furnishing Status
Unfurnished
Tenant Preferred
Bachelors/Family
Bathroom
2
Building Floor
23,00
Basement Floor
1
Size
1000-1500
Rent (Rupee)
12000,00
Predict

Thank You