BENGALURU HOUSE PRICE PREDICTION

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

SUBMITTED TO:
MRS . PRIYA RAINA

SUBMITTED BY:
SUMIT GAWARE (8025340040)
MEGHAVI SISODIYA (8025340044)

INTRODUCTION:

- Bengaluru real estate market is highly dynamic, making accurate house price prediction essential for buyers, sellers, and investors.
- Housing demand is increasing due to IT growth.
- This project leverages machine learning models (Linear Regression, Lasso, Ridge, Random Forest) to predict property prices based on key features like Area, BHK, Bathrooms, Location, and Property Type.
- This project builds a prediction model + full Streamlit app.
- An interactive Streamlit app allows users to input property details, get price predictions, rental estimates, and neighborhood insights for informed decision-making.

DATASET DESCRIPTION

Dataset Size: (1000 , 11)
Following are the parameters in Dataset :

- Area
- Location
- Bhk
- Bath
- Balcony
- Parking
- Furnishing
- Property_Type
- Age
- Price
- Price_Lakhs

SNAPSHOT OF DATASET

The following is head part of dataset:

```
dataset.head(4)
```

	Area	Location	Bhk	Bath	Balcony	Parking	Furnishing	Property_Type	Age	Price	Price_Lakhs
0	2065	Bannerghatta Road	2	3	0	1	Semi-Furnished	Independent House	3	17280000	172.8
1	1539	Yelahanka	3	1	0	1	Unfurnished	Villa	8	9410000	94.1
2	2048	Bannerghatta Road	3	1	2	0	Semi-Furnished	Independent House	10	20300000	203.0
3	1233	Sarjapur Road	3	2	1	2	Fully-Furnished	Apartment	12	9060000	90.6

DEALING WITH CATEGORICAL DATA

1.CATEGORICAL VARIABLES LOCATION, FURNISHING, PROPERTY TYPE WERE ENCODED USING LABELENCODER.
2.CONVERTS TEXT LABELS INTO NUMERICAL VALUES FOR MODEL TRAINING.
3.ENSURES ML MODELS CAN PROCESS NON-NUMERIC DATA WITHOUT ERRORS.

```
# Handling Categorical Data
# We fit on the whole dataset for the dropdowns, but in production, you'd save these encoders
model_df['Location'] = le_loc.fit_transform(model_df['Location'].astype(str))
model_df['Furnishing'] = le_furn.fit_transform(model_df['Furnishing'].astype(str))
model_df['Property_Type'] = le_prop.fit_transform(model_df['Property_Type'].astype(str))
```

DEALING WITH NUMERICAL DATA

1. THE PROJECT HANDLES NUMERICAL DATA SUCH AS AREA, BHK, BATHROOMS, BALCONY, PARKING, AND PROPERTY AGE USING STREAMLIT INPUT WIDGETS.
2. THESE WIDGETS ENSURE THAT ONLY VALID NUMBERS ARE ENTERED.

```
area = st.number_input("Area (Sq. Ft)", min_value=300, max_value=10000, value=1200)
bhk = st.slider("BHK", 1, 6, 2)
bath = st.slider("Bathrooms", 1, 6, 2)
balcony = st.slider("Balcony", 0, 4, 1)
parking = st.selectbox("Parking", [0, 1], format_func=lambda x: "Yes" if x==1 else "No")
age = st.slider("Property Age (Years)", 0, 50, 5)
```

FEATURE ENGINEERING

1. SELECTED IMPORTANT FEATURES AFFECTING HOUSE PRICE FOR MODEL TRAINING.
2. TARGET VARIABLE PRICE_LAKHS IS USED FOR NUMERICAL STABILITY.
3. HELPS THE MODEL FOCUS ON MEANINGFUL ATTRIBUTES LIKE AREA, BHK, BATHROOMS, ETC.

```python
# Feature Selection
X = model_df[['Area', 'Location', 'Bhk', 'Bath', 'Balcony', 'Parking', 'Furnishing', 'Property_Type', 'Age']]
y = model_df['Price_Lakhs'] # Predicting in Lakhs is more numerically stable
```

MODEL SELECTION

1. USERS CAN SELECT FROM LINEAR, LASSO, RIDGE, AND RANDOM FOREST MODELS.
2. SELECTED MODEL IS TRAINED ON THE TRAINING DATASET FOR PRICE PREDICTION.
3. PROVIDES FLEXIBILITY TO COMPARE PERFORMANCE OF DIFFERENT ALGORITHMS.

```
model_choice = st.selectbox("Choose Algorithm", ["Linear Regression", "Lasso Regression", "Ridge Regression", "Random Forest'
```

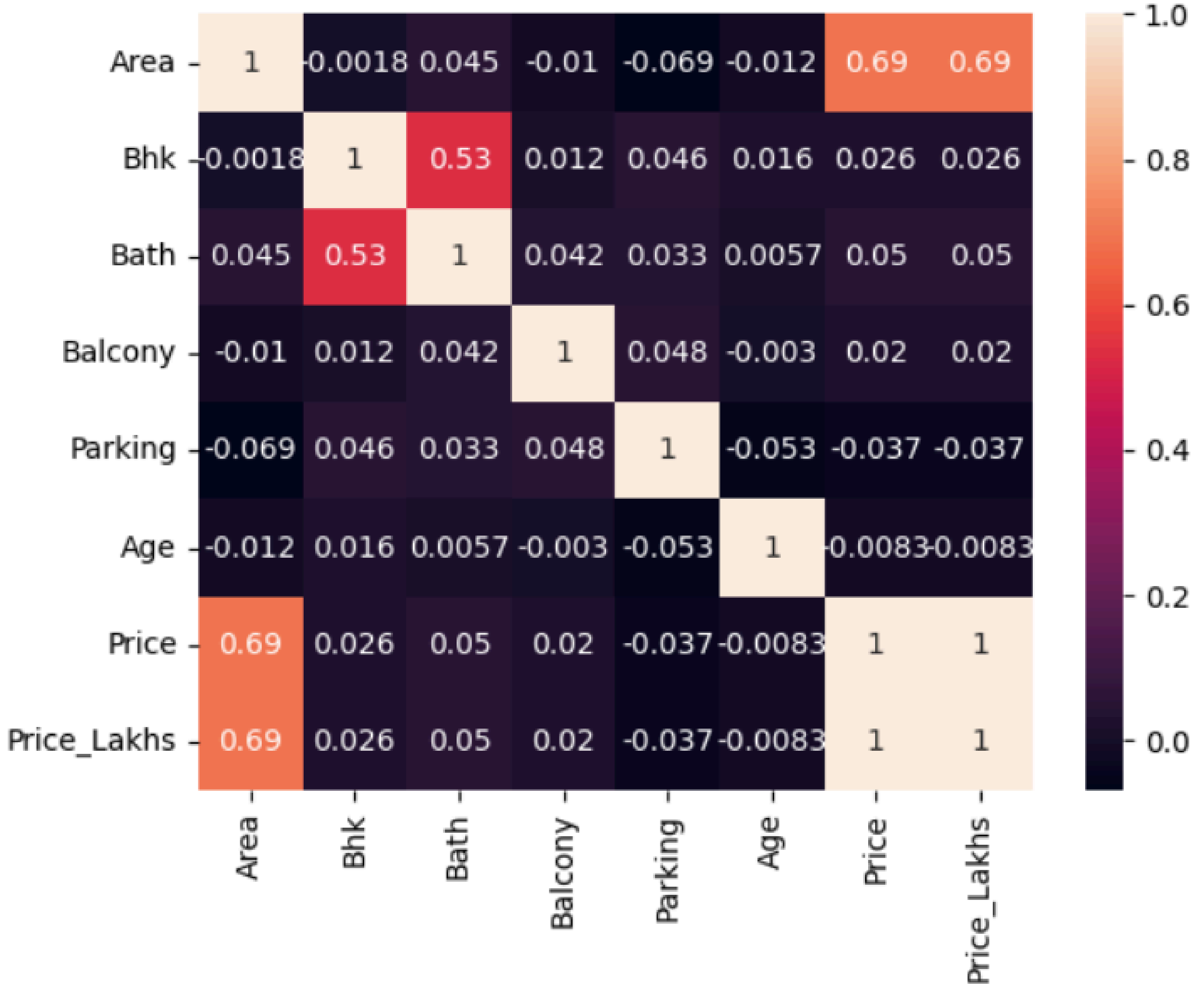
```
# Model Training (On the fly for this demo project)
if model_choice == "Linear Regression":
    model = LinearRegression()
elif model_choice == "Lasso Regression":
    model = Lasso(alpha=0.1)
elif model_choice == "Ridge Regression":
    model = Ridge(alpha=1.0)
else:
    model = RandomForestRegressor(n_estimators=50, random_state=42)

model.fit(X_train, y_train)
```

FEATURE CORELATION MATRIX

```
corelation = dataset.select_dtypes(include=['number']).corr()
sns.heatmap(corelation, xticklabels=corelation.columns, yticklabels=corelation.columns, annot=True)
```

```
<Axes: >
```



HOMEPAGE

Navigation

Go to

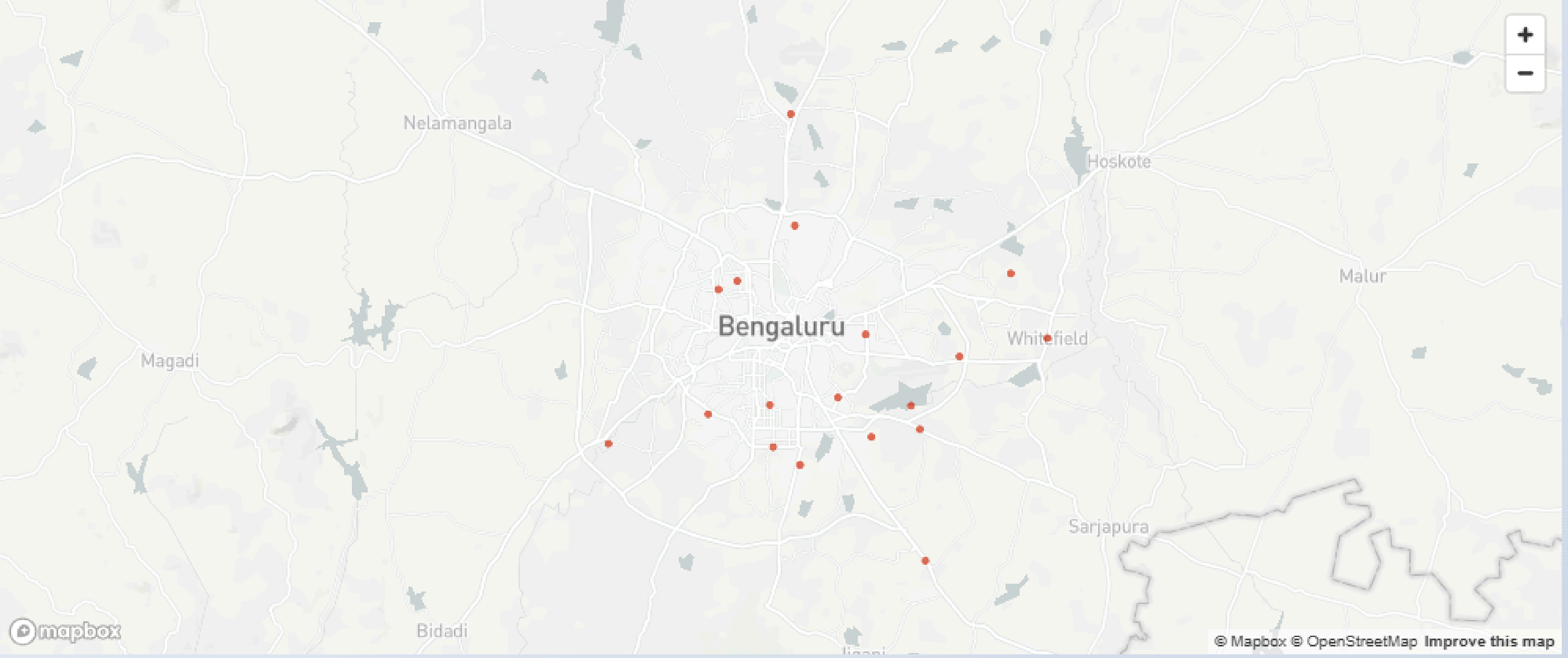
- Dashboard & Analysis
- Price Predictor
- Calculators

Created by Sumit Gaware and Meghavi Sisodiya

Bengaluru House Price Prediction

Total Properties	Avg Price (Lakhs)	Locations	Avg Area (sqft)
1000	₹123.73	20	1649

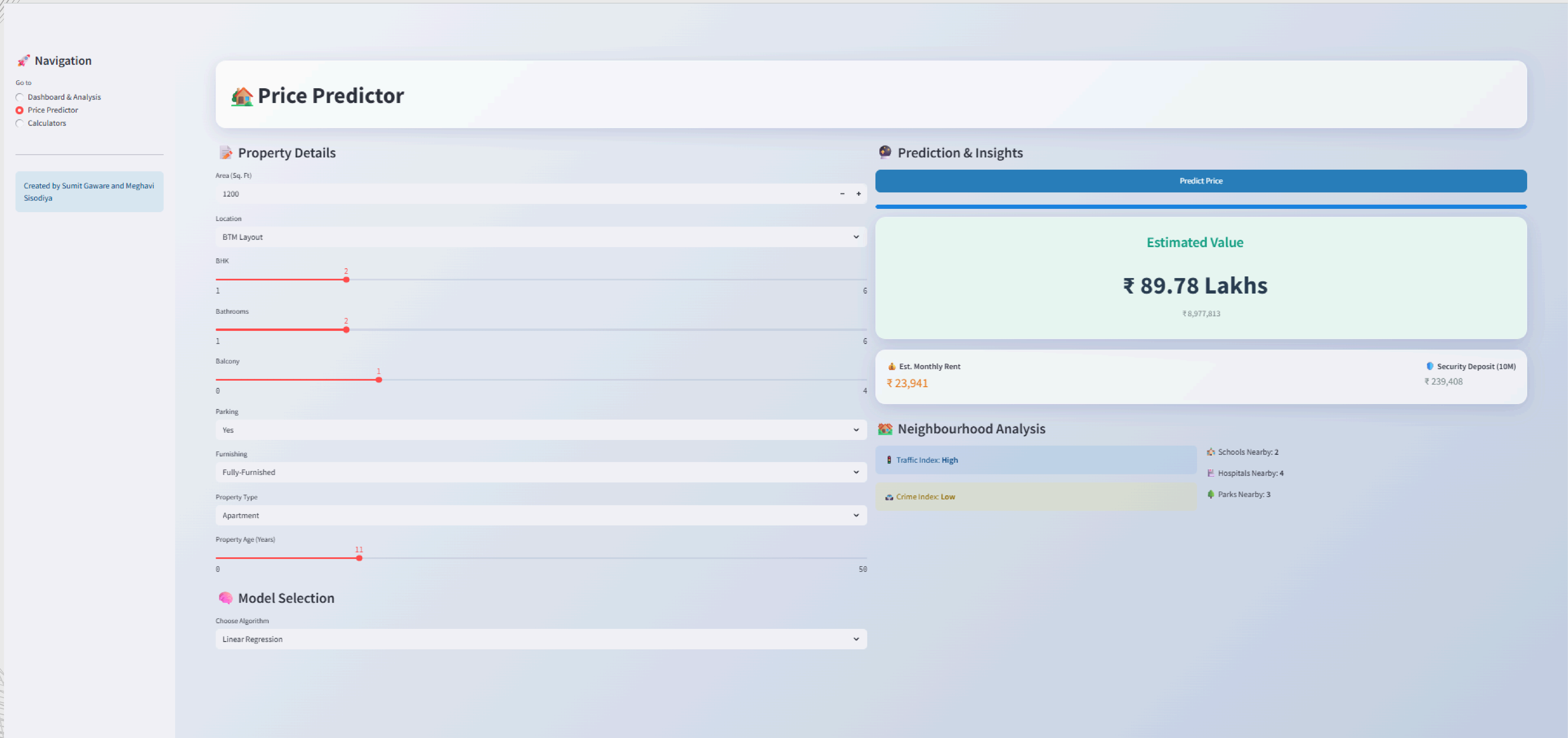
Property Distribution Map



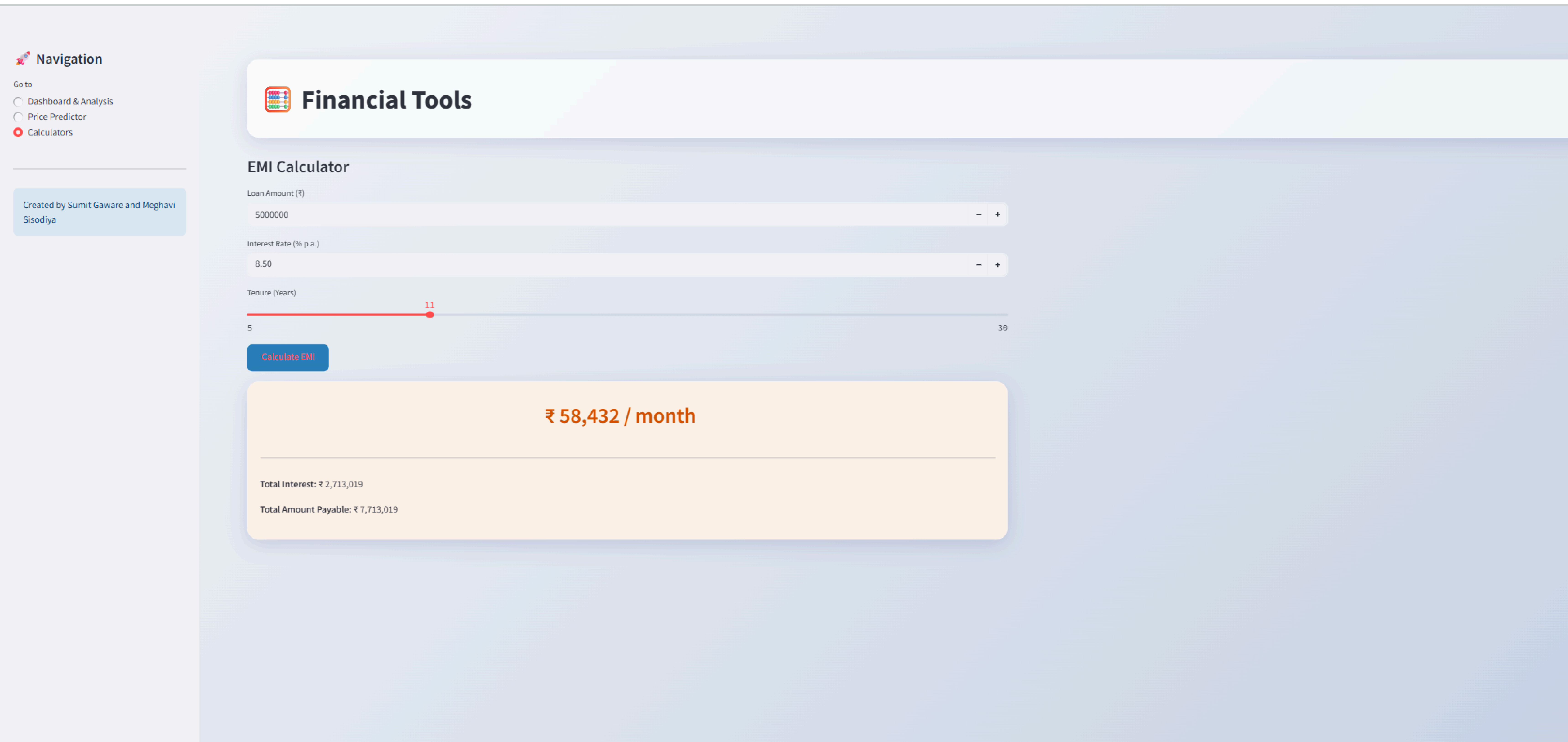
Locality Summary

Location	Price_Lakhs	Area
JP Nagar	154.6	1796.7
Rajajinagar	138.1	1765.5
Malleshwaram	135.3	1728.4
Koramangala	130.6	1701.7
Electronic City	128.6	1674.2
Indiranagar	127.9	1704.0
Banashankari	127.3	1667.8
BTM Layout	126.5	1662.8
Jayanagar	124.1	1632.4
Marathahalli	123.8	1633.8

PRICE PREDICTOR



EMI CALCULATOR



CONCLUSION

The Bengaluru House Price Prediction project successfully demonstrates how machine learning can be applied to real-world real estate problems. By analyzing a comprehensive dataset of Bengaluru property listings, the project identified key factors that influence housing prices, such as location, area, number of rooms, bathrooms, and amenities. Through systematic preprocessing, exploratory data analysis, and feature engineering, the dataset was transformed into a structured format suitable for modeling.

THANK YOU