

A vibrant night photograph of the Perth city skyline, featuring numerous illuminated skyscrapers reflected in the water of the Swan River in the foreground. The sky is a deep blue, transitioning to a lighter hue near the horizon. The city lights create a colorful reflection on the water's surface.

PERTH LIVING

PREDICTING RENT PRICES USING MACHINE LEARNING

SIDDHARTH DEY

INTRODUCTION

Goal: The project aims to predict the rent prices in Perth using a machine learning algorithm (linear regression model) trained with historic rental price data.

PROBLEM STATEMENT

The background of the slide features a wide-angle photograph of a coastal town. In the foreground, there's a sandy beach with a few people. To the left, a large, ornate yellow building with multiple arched doorways stands near the water. The middle ground shows a dense cluster of houses and trees on a hillside overlooking the ocean. A small, rocky peninsula or breakwater extends into the water on the right side. The water is a vibrant turquoise color. The sky is clear and blue.

Problem Statement: How can we assist homebuyers and realtors in making informed decisions regarding real estate transactions by using a Linear Regression model trained with the help of a comprehensive dataset that explores various spatial, logistic, and geosocial features of houses in the area?

DATA SET AND CLEANING

The chosen dataset was Perth House prices data set which has 33656 rows and 19 columns.

INDEX	COLUMN NAME	COLUMN DESCRIPTION	DATATYPE
1	ADDRESS	Address in df	Object
2	SUBURB	Suburbs within PERTH	object
3	PRICE	Price in AUD	Number (float)
4	BEDROOMS	Number of bedrooms	Number(float)
5	BATHROOMS	Number of bathrooms	Number(float)
6	GARAGE	Number of garage places	Number(float)
7	LAND_AREA	Total land area (m^2)	Number(float)
8	FLOOR_AREA	Internal floor area (m^2)	Number(float)
9	BUILD_YEAR	Year in which the property was built	Number(float)
10	CBD_DIST	Distance from the center of Perth (m)	Number(float)



DATA CLEANING

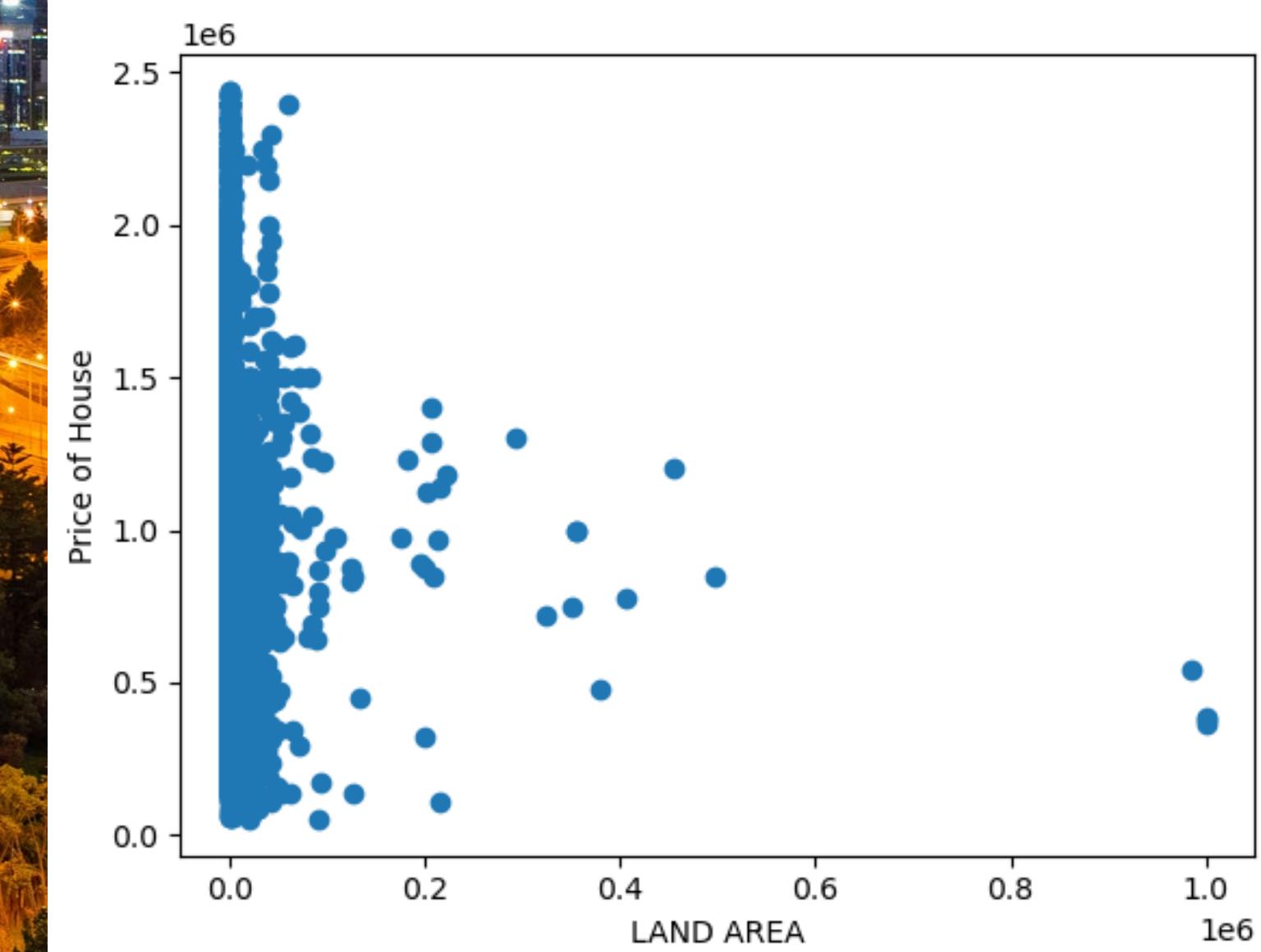
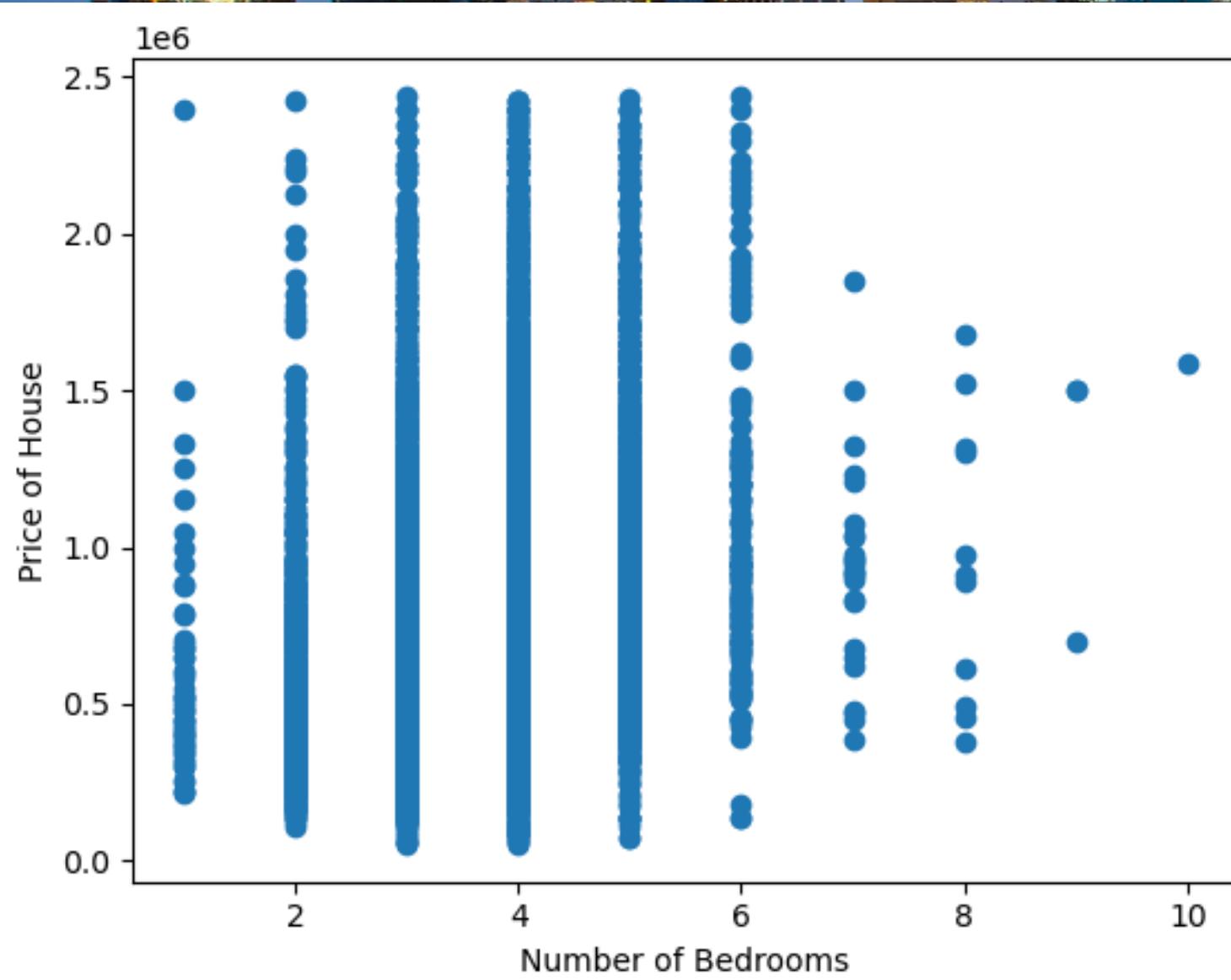
10	CBD_DIST	Distance from the center of Perth (m)	Number(float)
11	NEAREST_STN	The nearest public transport station from the property	Object
12	NEAREST_STN_DIST	The nearest station distance (m)	Number(float)
13	DATE SOLD	Month & year in which the property was sold	Number(float)
14	POSTCODE	Local Area Identifier	Number(float)
15	LATITUDE	Geographic Location (lat) of ADDRESS	Number(float)
16	LONGITUDE	Geographic Location (long) of ADDRESS	Number(float)
17	NEAREST_SCH	Location of the nearest School	Object
18	NEAREST_SCH_DIST	Distance to the nearest school	Number(float)
19	NEAREST_SCH_RANK	Ranking of the nearest school	Number(float)

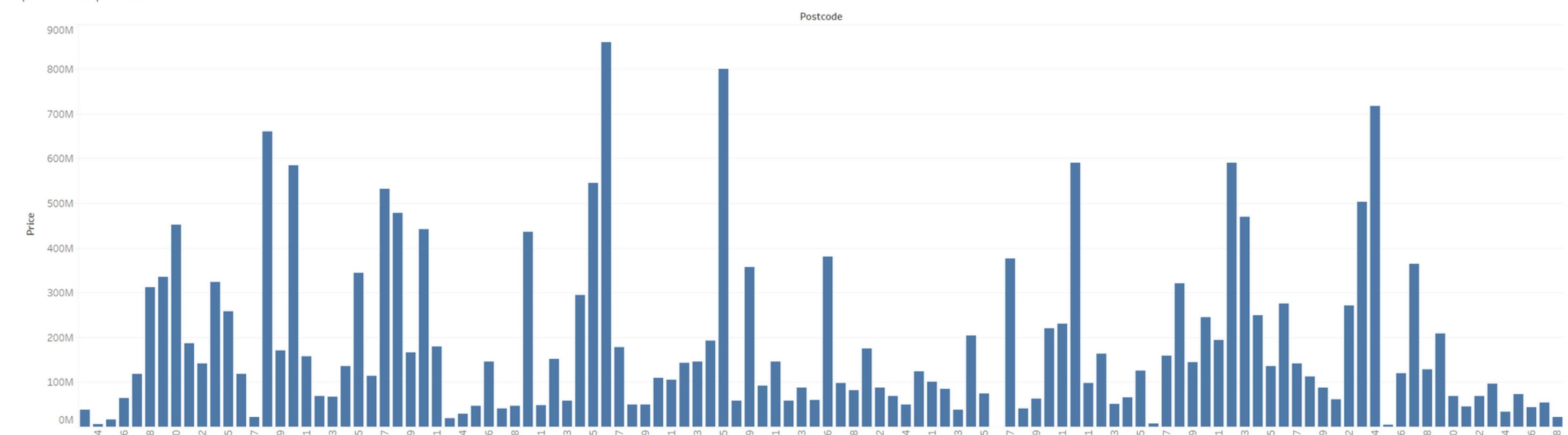
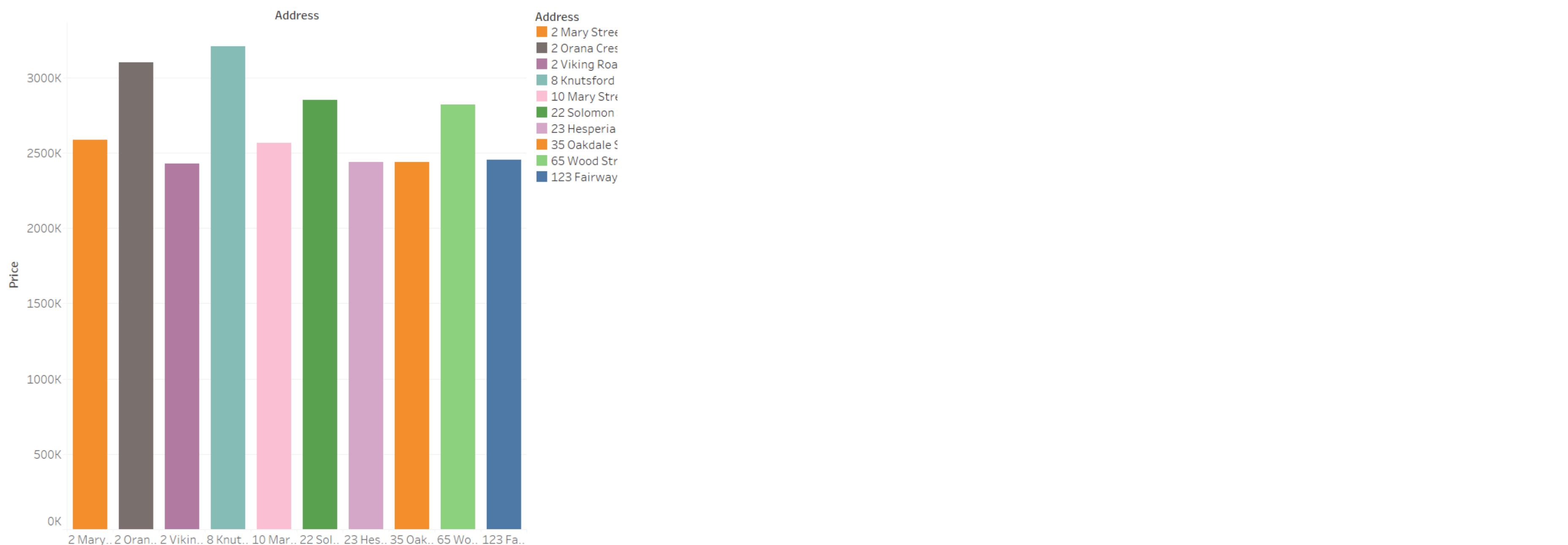
```
: df.describe()
```

	PRICE	BEDROOMS	BATHROOMS	GARAGE	LAND_AREA	FLOOR_AREA	BUILD_YEAR	CBD_DIST	NEAREST_STN_DIST	POSTCODE
count	1.920800e+04	19208.000000	19208.000000	19208.000000	19208.000000	19208.000000	19208.000000	19208.000000	19208.000000	19208.000000
mean	6.998658e+05	3.676020	1.861776	2.183205	2441.914879	187.532330	1988.922741	18327.112922	4188.389681	6086.237245
std	3.872944e+05	0.750726	0.587704	1.221165	18446.099139	72.985315	21.907656	10937.511971	4172.353928	64.491313
min	5.200000e+04	1.000000	1.000000	1.000000	61.000000	1.000000	1870.000000	1300.000000	46.000000	6003.000000
25%	4.380000e+05	3.000000	2.000000	2.000000	494.000000	134.000000	1977.000000	10100.000000	1600.000000	6030.000000
50%	5.850000e+05	4.000000	2.000000	2.000000	675.000000	177.000000	1995.000000	15800.000000	3000.000000	6065.000000
75%	8.500000e+05	4.000000	2.000000	2.000000	809.000000	228.000000	2005.000000	24300.000000	5100.000000	6150.000000
max	2.440000e+06	10.000000	7.000000	50.000000	999999.000000	849.000000	2017.000000	56900.000000	34300.000000	6558.000000

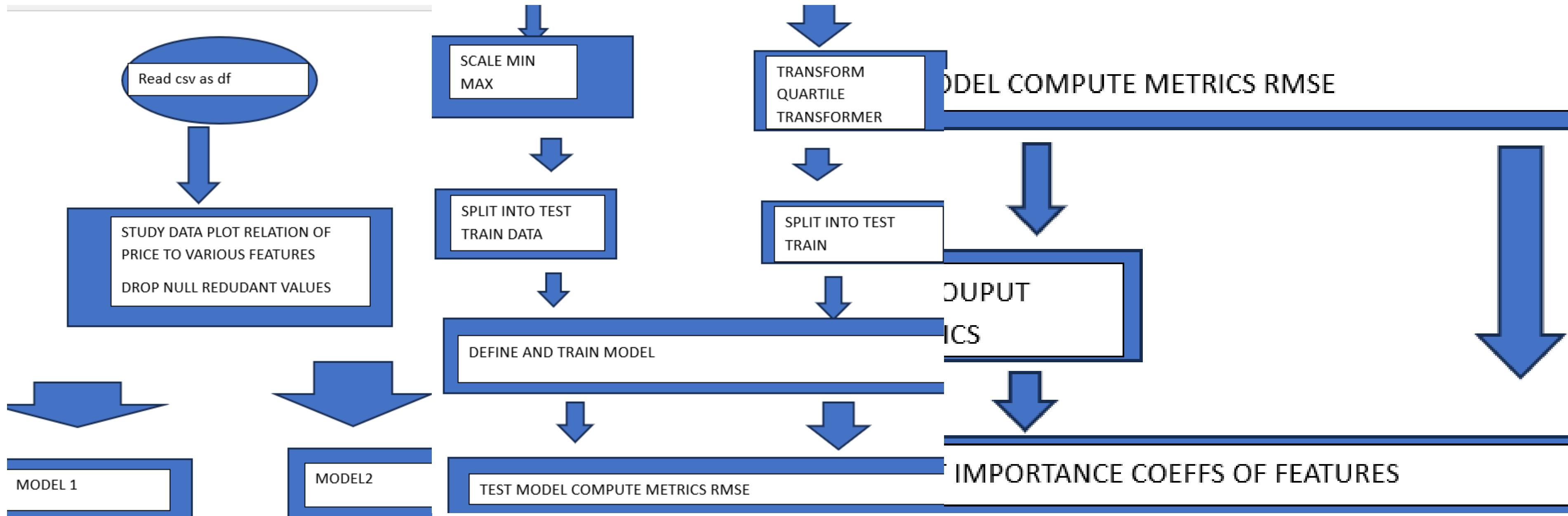
Windows aktivieren

VISUALIZATIONS OF THE INITIAL DATA SET

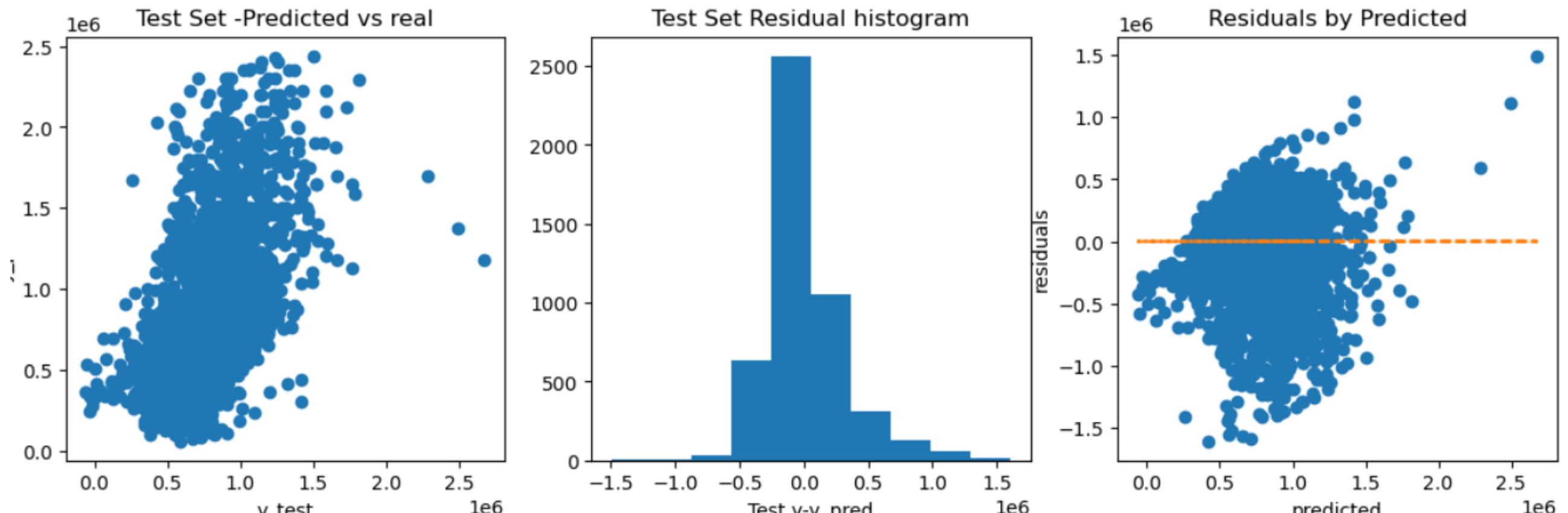




MODELLING REPRESENTATION



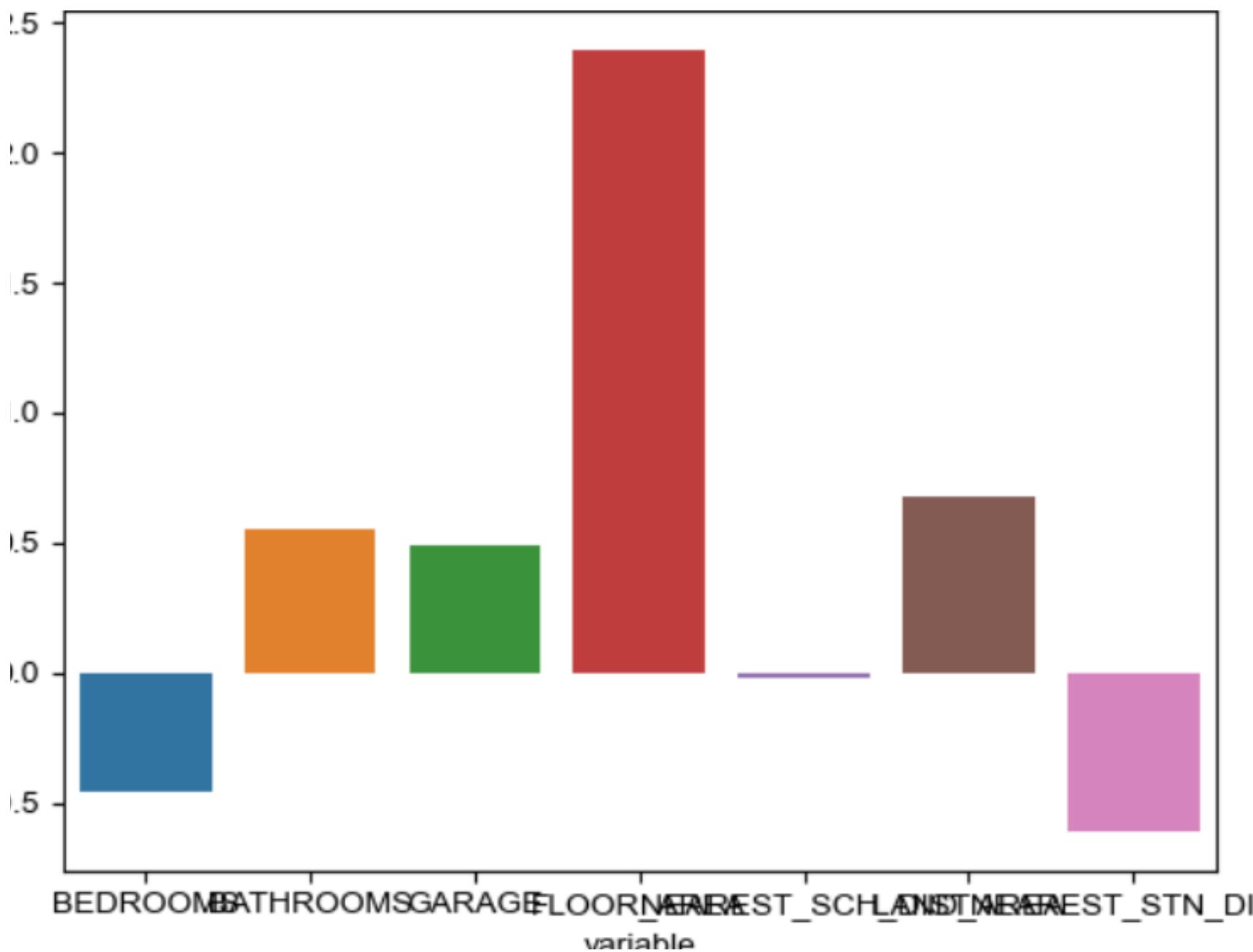
MODEL 1 RESULTS



```
rmse = np.sqrt(mse)
print("RMSE of model is ", rmse)
```

RMSE of model is 306783.0421154133

MODEL 1 RESULTS

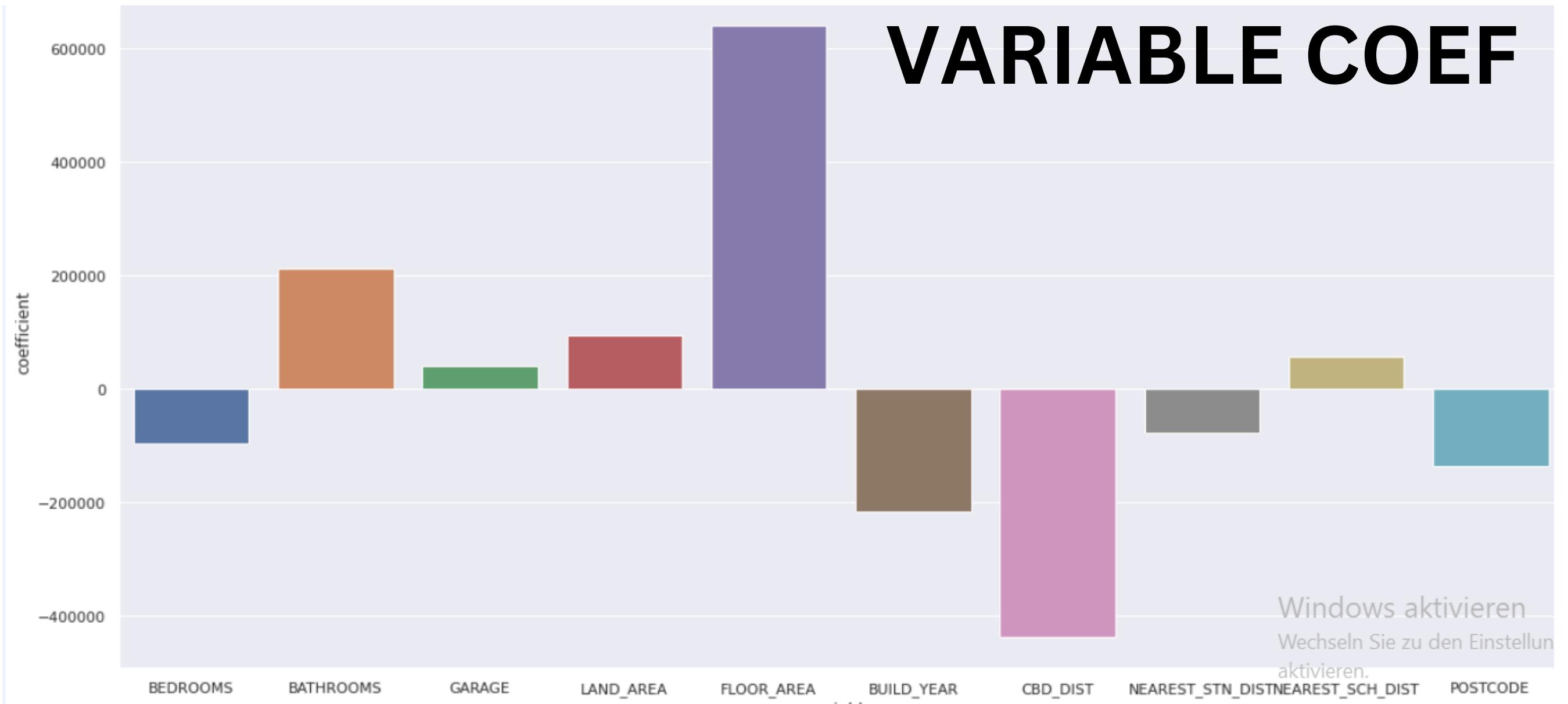


MODEL ZWEI" - RENTAL PRICES

```
# features = [  
#     'BEDROOMS', 'BATHROOMS', 'GARAGE', 'LAND_AREA', 'FLOOR_AREA',  
#     'BUILD_YEAR', 'CBD_DIST', 'NEAREST_STN_DIST', 'NEAREST_SCH_DIST',  
#     'POSTCODE',]  
  
input_area_data = np.array([2, 1, 1, 60, 45, 2007, 4000, 2000, 5000, 6061])  
predicted_price_aud = predict_price(input_area_data)  
print('Predicted Rent Price (AUD): ', predicted_price_aud)
```

Predicted Rent Price (AUD): 401654.8831787494

VARIABLE COEF



```
mse = mean_squared_error(y_test, y_pred)
print('Mean Squared Error: ', mse)
print('RMSE = ', np.sqrt(mse))
```

Mean Squared Error: 1.1882150884946357e+20

RMSE = 10900527916.090282



THANK YOU

SIDDHARTH DEY