NETAJI SUBHAS UNIVERSITY OF TECHNOLOGY

COMPUTER HARDWARE SOFTWARE WORKSHOP

COCSC19

SUBMITTED BY:

YASH CHAUHAN- 2021UCS1628

KHUSHI KHARKE- 2021UCS1617

ADITYA YADAV- 2021UCS1640

POWER-BI TASK

1. Explore Power View, Power Query

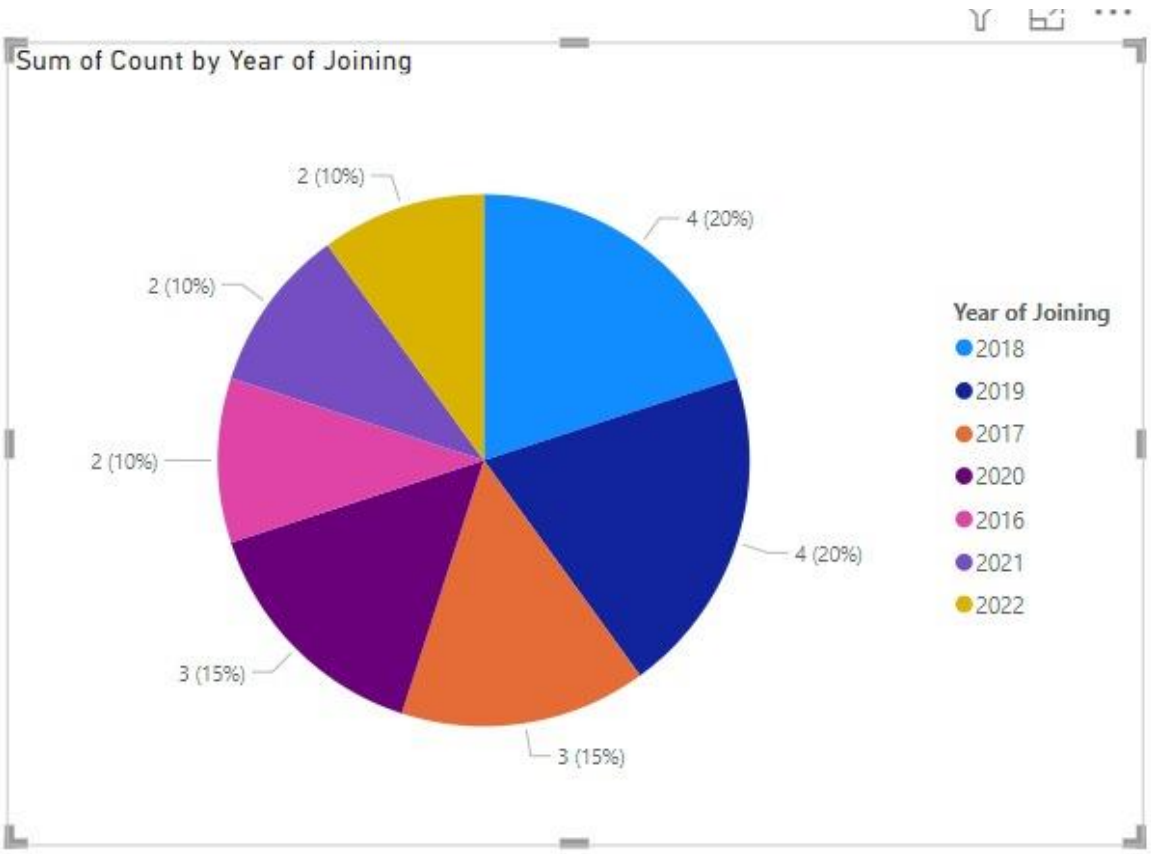
(Create a table Employee(empid, gender, department, salary, country, year_of_joining) connect to Employee data file.
Remove missing gender and department values.
Extract year_of_joining column and visualize number of employees w.r.t year of experience in the company.
Perform self-join using Power Query.
Aggregate salary with gender and Visualize using Pie chart)

	Gender	Department	Salary	Country	year_of_joining	Custom
1001	Male	Sales	50000	USA	2018	1
1002	Female	HR	55000	Canada	2019	1
1003	Male	IT	60000	UK	2017	1
1004	null	Marketing	48000	USA	2020	
1005	Male	Finance	70000	Australia	2016	1
1006	Female	Operations	52000	Germany	2021	1
1007	Male	Sales	48000	Canada	2019	1
1008	null	IT	62000	USA	2017	
1009	Male	Finance	65000	UK	2020	1
1010	Female	null	53000	Australia	2018	
1011	Male	Marketing	58000	Germany	2019	1
1012	null	Operations	54000	Canada	2022	
1013	Male	IT	63000	UK	2018	1
1014	Female	null	67000	USA	2019	
1015	Male	HR	52000	Australia	2021	1
1016	Female	Sales	49000	Germany	2020	1
1017	Male	null	60000	Canada	2017	
1018	Female	Marketing	51000	UK	2018	1
1019	Male	Finance	68000	USA	2022	1
1020	Female	null	64000	Australia	2016	

= Table.SelectColu

	Year of Joining
1	2018
2	2019
3	2017
4	2020
5	2016
6	2021
7	2019
8	2017
9	2020
10	2018
11	2019
12	2022
13	2018
14	2019
15	2021
16	2020
17	2017
18	2018
19	2022
20	2016

	Year of Joining	Count
1	2018	4
2	2019	4
3	2017	3
4	2020	3
5	2016	2
6	2021	2
7	2022	2

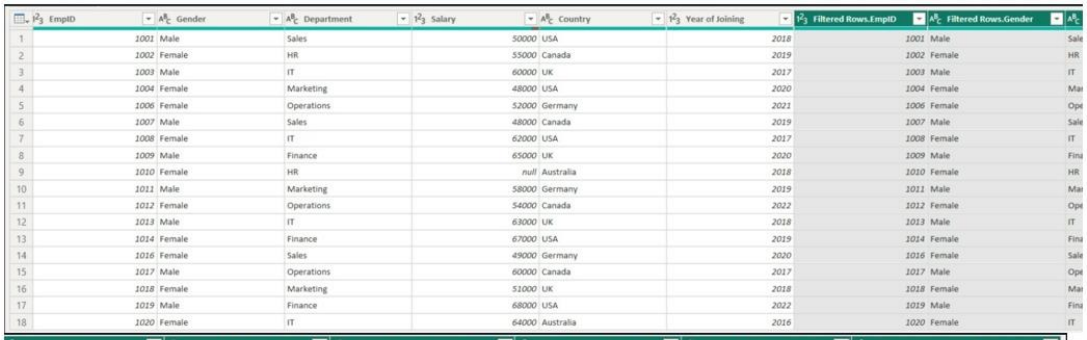


Sheet1

EmpID	Gender	Department	Salary	Country	Year of Joining
1001	Male	Sales	50000	USA	2018
1002	Female	HR	55000	Canada	2019
1003	Male	IT	60000	UK	2017
1004	Female	Marketing	48000	USA	2020
1006	Female	Operations	52000	Germany	2021

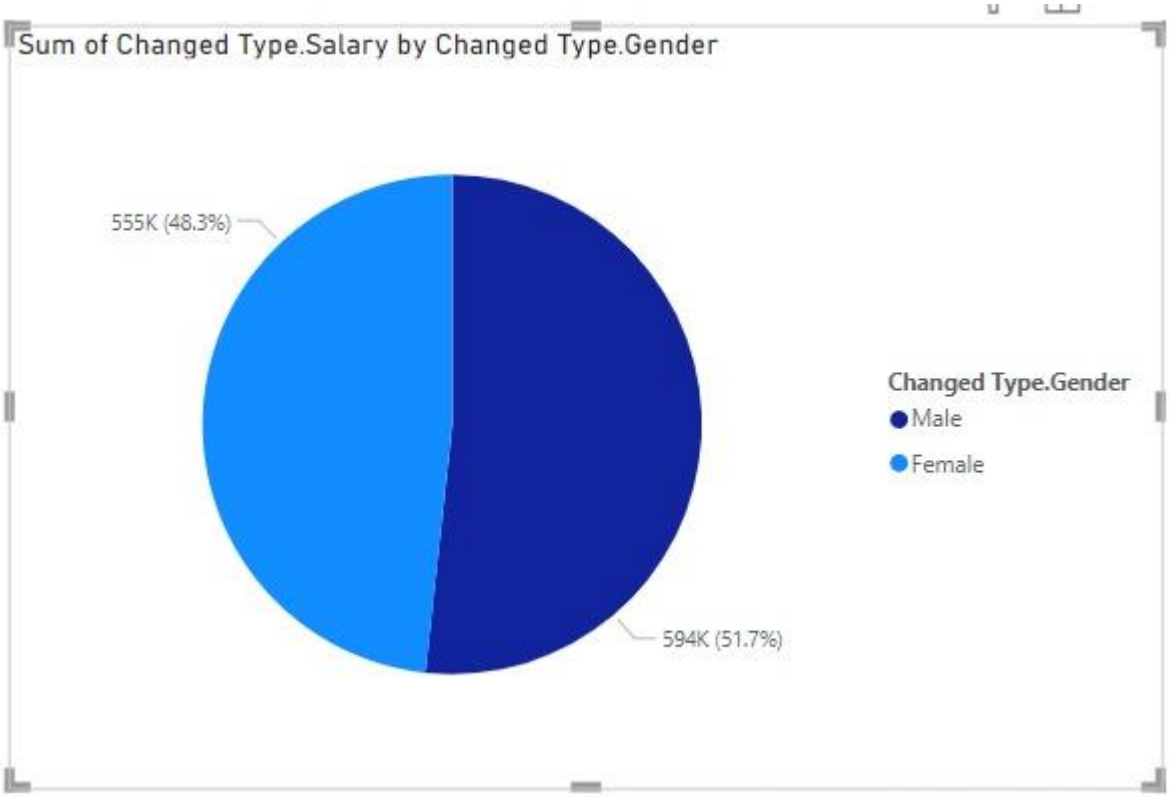
Sheet1 (Current)

EmpID	Gender	Department	Salary	Country	Year of Joining
1001	Male	Sales	50000	USA	2018
1002	Female	HR	55000	Canada	2019
1003	Male	IT	60000	UK	2017
1004	Female	Marketing	48000	USA	2020
1006	Female	Operations	52000	Germany	2021



	EmpID	Gender	Department	Salary	Country	Year of Joining	Filtered Rows.EmpID	Filtered Rows.Gender	
1	1001	Male	Sales	50000	USA	2018	1001	Male	Sale
2	1002	Female	HR	55000	Canada	2019	1002	Female	HR
3	1003	Male	IT	60000	UK	2017	1003	Male	IT
4	1004	Female	Marketing	48000	USA	2020	1004	Female	Mar
5	1006	Female	Operations	52000	Germany	2021	1006	Female	Ope
6	1007	Male	Sales	48000	Canada	2019	1007	Male	Sale
7	1008	Female	IT	62000	USA	2017	1008	Female	IT
8	1009	Male	Finance	65000	UK	2020	1009	Male	Fina
9	1010	Female	HR	null	Australia	2018	1010	Female	HR
10	1011	Male	Marketing	58000	Germany	2019	1011	Male	Mar
11	1012	Female	Operations	54000	Canada	2022	1012	Female	Ope
12	1013	Male	IT	63000	UK	2018	1013	Male	IT
13	1014	Female	Finance	67000	USA	2019	1014	Female	Fina
14	1016	Female	Sales	49000	Germany	2020	1016	Female	Sale
15	1017	Male	Operations	60000	Canada	2017	1017	Male	Ope
16	1018	Female	Marketing	51000	UK	2018	1018	Female	Mar
17	1019	Male	Finance	68000	USA	2022	1019	Male	Fina
18	1020	Female	IT	64000	Australia	2016	1020	Female	IT

Filtered Rows.EmpID	Filtered Rows.Gender	Filtered Rows.Department	Filtered Rows.Salary	Filtered Rows.Country	Filtered Rows.Year of Joining
1001	Male	Sales	50000	USA	2018
1002	Female	HR	55000	Canada	2019
1003	Male	IT	60000	UK	2017
1004	Female	Marketing	48000	USA	2020
1006	Female	Operations	52000	Germany	2021
1007	Male	Sales	48000	Canada	2019
1008	Female	IT	62000	USA	2017
1009	Male	Finance	65000	UK	2020
1010	Female	HR	null	Australia	2018
1011	Male	Marketing	58000	Germany	2019
1012	Female	Operations	54000	Canada	2022
1013	Male	IT	63000	UK	2018
1014	Female	Finance	67000	USA	2019
1016	Female	Sales	49000	Germany	2020
1017	Male	Operations	60000	Canada	2017
1018	Female	Marketing	51000	UK	2018
1019	Male	Finance	68000	USA	2022



Changed Type.EmpID	Changed Type.Gender	Changed Type.Department	Changed Type.Salary	Changed Type.Country	Changed Type.Year of Joining
1001	Male	Sales	50000	USA	
1002	Female	HR	55000	Canada	
1003	Male	IT	60000	UK	
1004	Female	Marketing	48000	USA	
1005	Male	Finance	70000	Australia	
1006	Female	Operations	52000	Germany	
1007	Male	Sales	48000	Canada	
1008	Female	IT	62000	USA	
1009	Male	Finance	65000	UK	
1010	Female	HR	53000	Australia	
1011	Male	Marketing	58000	Germany	
1012	Female	Operations	54000	Canada	
1013	Male	IT	63000	UK	
1014	Female	Finance	67000	USA	
1015	Male	HR	52000	Australia	
1016	Female	Sales	49000	Germany	
1017	Male	Operations	60000	Canada	
1018	Female	Marketing	51000	UK	
1019	Male	Finance	68000	USA	
1020	Female	IT	64000	Australia	

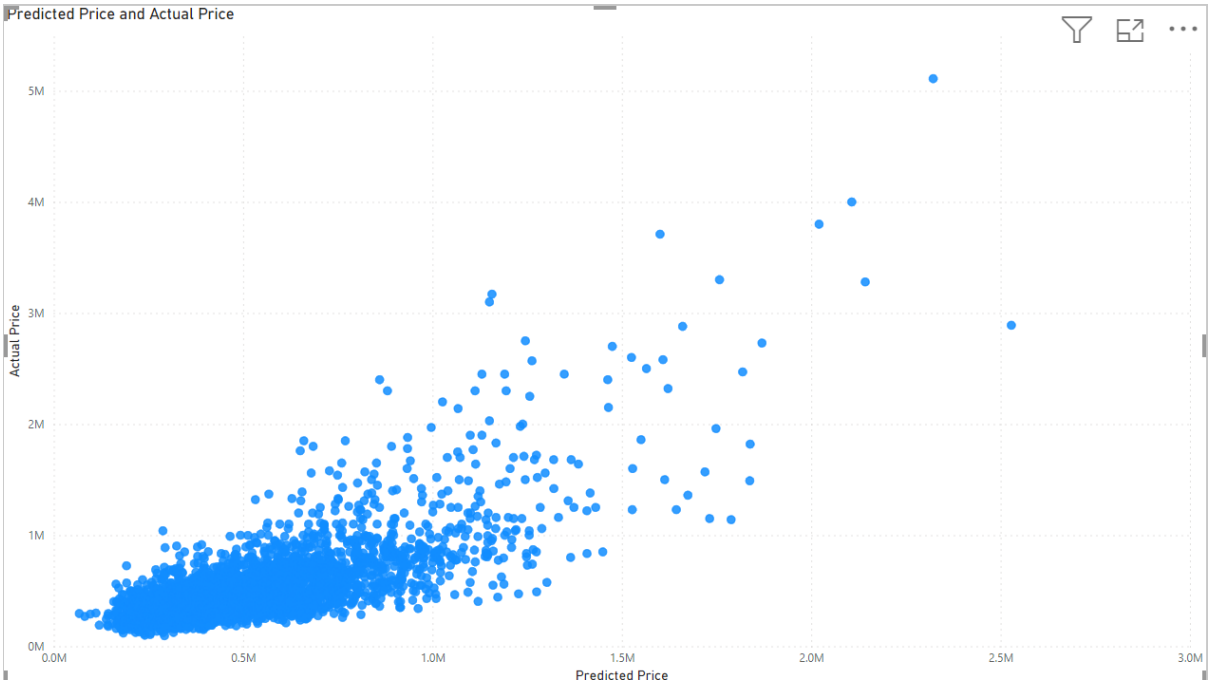
2. Visualize the result of any Machine Learning algorithm on any dataset of your choice in PowerBI.

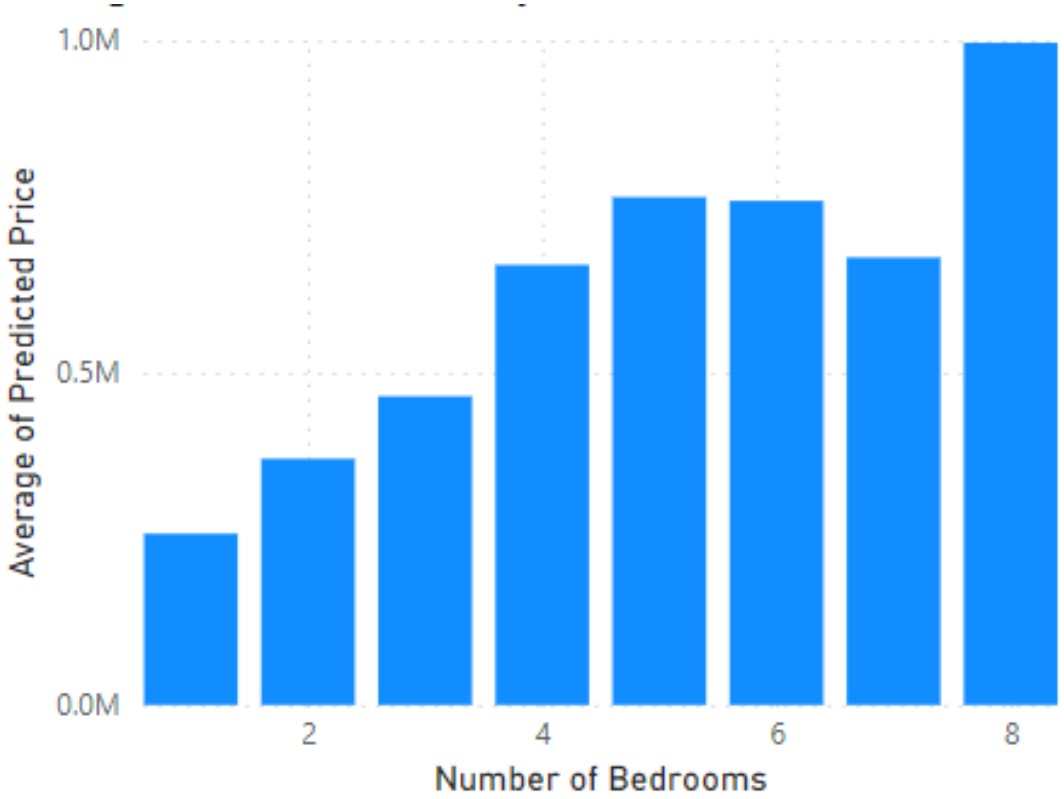
ML code:

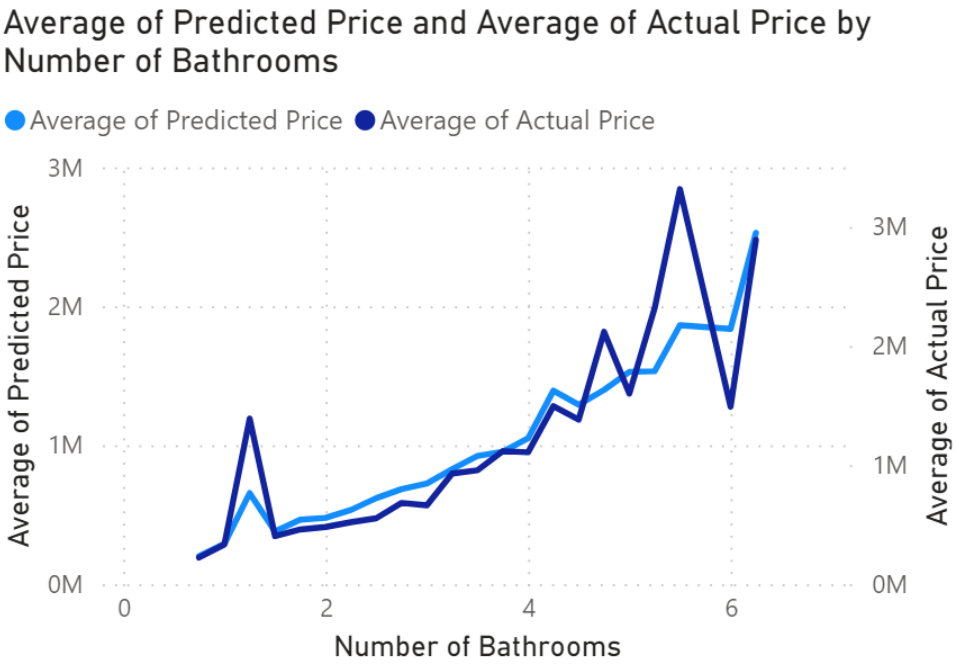
```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# Load the dataset
dataset_path = r"D:\6th sem\CHS\powerbi\dataset_houseprediction\House Price India.csv"
house_df = pd.read_csv(dataset_path)

# Select the specified columns as features
X = house_df[['number of bedrooms', 'number of bathrooms', 'living area', 'lot area', 'number of floors']]

# Target variable
y = house_df['Price']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Train the model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
print("Mean Squared Error:", mse)
```

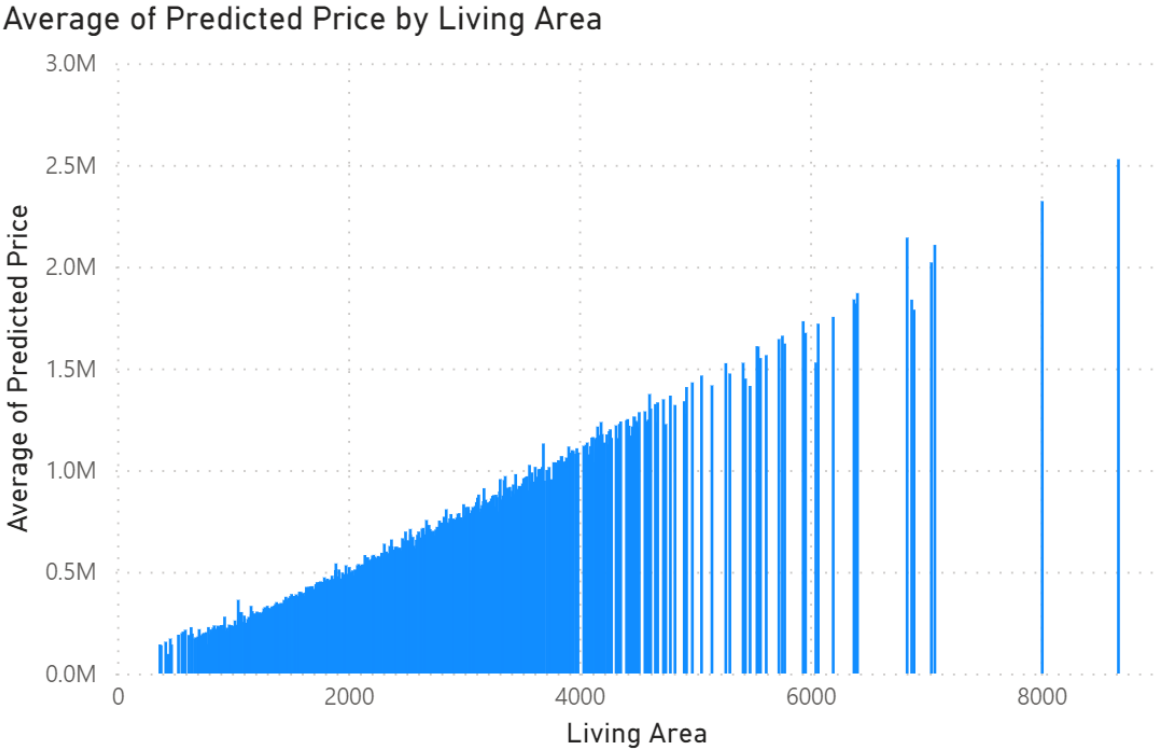
```python
# Create a DataFrame with X_test, y_test, and y_pred
results_df = pd.DataFrame({
    'Number of Bedrooms': X_test['number of bedrooms'],
    'Number of Bathrooms': X_test['number of bathrooms'],
    'Living Area': X_test['living area'],
    'Lot Area': X_test['lot area'],
    'Number of Floors': X_test['number of floors'],
    'Actual Price': y_test,
    'Predicted Price': y_pred
})

# Export the DataFrame to a CSV file
results_df.to_csv('house_price_predictions.csv', index=False)
```

VISUALIZATION:






Average of Predicted Price and Average of Actual Price by Number of Bathrooms
Average of Predicted Price
Average of Actual Price
3M
2M
1M
0M
Average of Predicted Price
Average of Actual Price
0
2
4
6
Number of Bathrooms


Average of Predicted Price by Living Area
3.0M
2.5M
2.0M
1.5M
1.0M
0.5M
0.0M
Average of Predicted Price
0
2000
4000
6000
8000
Living Area