Predicting Housing Prices

In [1]:
```
%matplotlib inline
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
```

1. Loading Home Data

In [3]:
```
home_data = pd.read_csv('./home_data.csv')
df = home_data[['price', 'zipcode']]
```

Quiz1

We found the zip code with the highest average house price.
What is the average house price of that zip code?

In [9]:
```
zip_avg_price = pd.DataFrame(data = df.groupby('zipcode', as_index=False).mea
zip_avg_price.sort_values(['price'], ascending=False).head(5)
```

Out[9]:

	zipcode	price
24	98039	2.160607e+06
3	98004	1.355927e+06
25	98040	1.194230e+06
48	98112	1.095499e+06
41	98102	9.012582e+05

Quiz2

What fraction of the houses have living space between 2000 sq.ft. and 4000 sq.ft.?

In [15]:
```
square_range_df = home_data[(home_data['sqft_living'] > 2000) & (home_data['s
fraction_target = (len(square_range_df.index)/len(home_data.index))
print(fraction_target)
```

```
0.42187572294452413
```

Quiz3

What is the difference in RMSE between the model trained with my_features and the one trained with advanced_features?

In [23]:
```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
```

In [33]:
```
total_size = len(home_data)
test_ratio = 0.2
```

```
train_data = home_data.iloc[0: int(total_size*(1-test_ratio))]
test_data = home_data.iloc[int(total_size*(1-test_ratio)):]

my_features = ['bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot', 'floors',
advanced_features = ['bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot', 'flo

my_feat_x = train_data[my_features]
advanced_x = train_data[advanced_features]
y = train_data['price']

model = LinearRegression()
model.fit(my_feat_x, y)

my_feat_RMSE = np.sqrt(mean_squared_error(test_data['price'], model.predict(t

model = LinearRegression()
model.fit(advanced_x, y)

advanced_RMSE = np.sqrt(mean_squared_error(test_data['price'], model.predict(

print(f'Diff btw. the regression model for two type of features : {np.abs(my_
```

```
Diff btw. the regression model for two type of features : 48280.64826140503
```

In [ ]: