



King County, Washington Housing Prediction



Dataset

`kc_house_data.csv`

<https://www.kaggle.com/shivachandel/kc-house-data>

Preparing

```
price           0
bedrooms        0
bathrooms       0
sqft_living     0
sqft_lot        0
floors          0
waterfront      0
view            0
condition       0
grade           0
sqft_basement   0
yr_built        0
yr_renovated    0
zipcode         0
lat             0
long            0
sqft_living15   0
sqft_lot15      0
dtype: int64
```

```
In [18]: from sklearn.model_selection import train_test_split
```

```
In [19]: x = df.loc[:,['bedrooms','bathrooms','floors','sqft_living','zipcode','sqft_basement']]
```

```
In [20]: x.head()
```

```
Out[20]:
```

	bedrooms	bathrooms	floors	sqft_living	zipcode	sqft_basement
0	3	1.0	1.0	1180	98178	0
1	3	2.0	2.0	2570	98125	400
2	2	1.0	1.0	770	98028	0
3	4	3.0	1.0	1960	98136	910
4	3	2.0	1.0	1680	98074	0

```
In [21]: x_train2, x_test2, y_train2, y_test2 = train_test_split(
x, y, test_size=0.25)
```

```
In [22]: x_train2.head()
```

Algorithm

Linear Regression

LINEAR

```
In [24]: from sklearn.linear_model import LinearRegression
```

```
In [25]: linear = LinearRegression()
```

```
In [26]: linear.fit(x_train2,y_train2)
```

```
Out[26]: LinearRegression()
```

```
In [27]: y_pred2=linear.predict(x_test2)
```

```
In [28]: y_pred2
```

```
Out[28]: array([[460486.0940296 ],  
                [228169.44832641],  
                [787214.7442747 ],  
                ...,  
                [863498.41632012],  
                [197255.35817373],  
                [504038.72450384]])
```

API

API

```
In [49]: # GET /calculatePrice/:bedrooms/:bathrooms/:floors/:sqft_living/:zipcode/:sqft_basement
request = json.loads(REQUEST)
bedrooms_var = request['path'].get('bedrooms')
bathrooms_var = request['path'].get('bathrooms')
floors_var = request['path'].get('floors')
sqft_living_var = request['path'].get('sqft_living')
zipcode_var = request['path'].get('zipcode')
sqft_basement_var = request['path'].get('sqft_basement')

price = linear.predict(np.array([[bedrooms_var,bathrooms_var,floors_var,sqft_living_var,zipcode_var,sqft_basement_var]]))[0][0]
print({'price' : "%d"} % price)
```