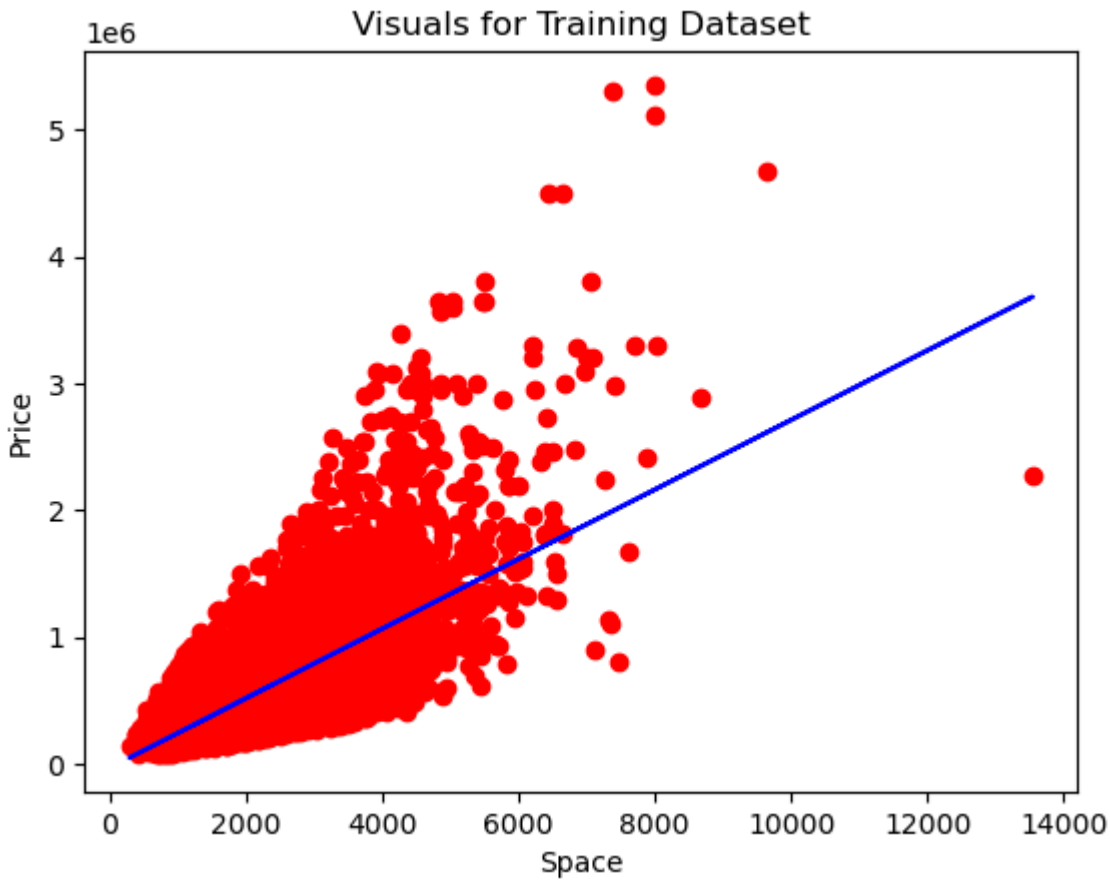
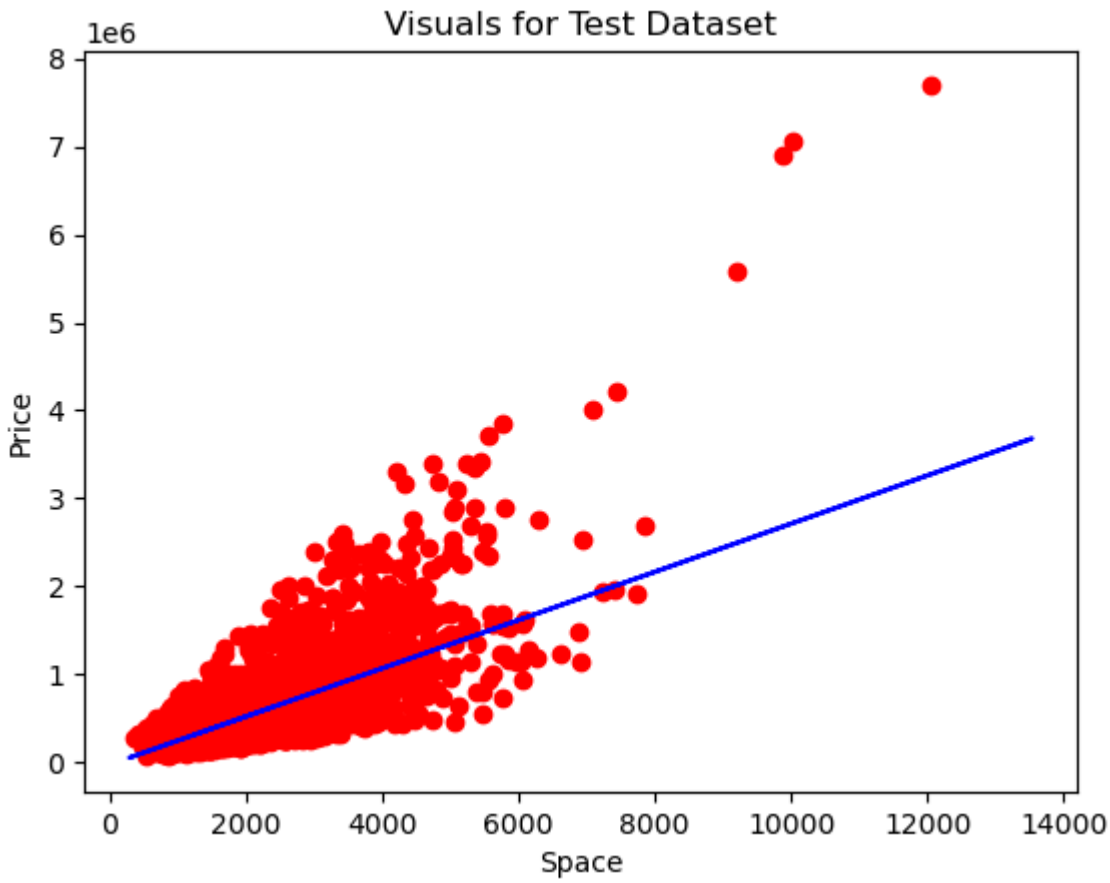
This is my first attempt on a simple linear regression model to predict housing prices. Still working on improving the accuracy of the model though. This is very simple code with not much functionality.

In [3]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# Optional: show full arrays
np.set_printoptions(threshold=np.inf)

# Load dataset (choose correct path depending on Kaggle vs local)
# On Kaggle:
# dataset = pd.read_csv("../input/house_data.csv")
# On Local PC:
dataset = pd.read_csv(r"../input/house-price-prediction\House price prediction\H

space = dataset['sqft_living']
price = dataset['price']

x = np.array(space).reshape(-1, 1)
y = np.array(price)

# Train-test split
from sklearn.model_selection import train_test_split
xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size=1/3, random_stat

# Fit Linear Regression
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(xtrain, ytrain)

# Predict
pred = regressor.predict(xtest)

# Visualize Training results
plt.scatter(xtrain, ytrain, color='red')
plt.plot(xtrain, regressor.predict(xtrain), color='blue')
plt.title("Visuals for Training Dataset")
plt.xlabel("Space")
plt.ylabel("Price")
plt.show()

# Visualize Test results
plt.scatter(xtest, ytest, color='red')
plt.plot(xtrain, regressor.predict(xtrain), color='blue')
plt.title("Visuals for Test Dataset")
plt.xlabel("Space")
plt.ylabel("Price")
plt.show()
```


Visuals for Training Dataset
1e6
Price
Space
0
1
2
3
4
5
0
2000
4000
6000
8000
10000
12000
14000


Visuals for Test Dataset
1e6
Price
Space
0
1
2
3
4
5
6
7
8
0
2000
4000
6000
8000
10000
12000
14000

In [ ]: