

## Importar módulos que necesitaremos para este portátil

```
In [ ]: import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import train_test_split
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

## Cargar el conjunto de datos de entrenamiento

```
In [ ]: bike_data = pd.read_csv('daily-bike-share.csv')
bike_data['day'] = pd.DatetimeIndex(bike_data['dteday']).day
numeric_features = ['temp', 'atemp', 'hum', 'windspeed']
categorical_features = ['season', 'mnth', 'holiday', 'weekday', 'workingday', 'weathersit']
bike_data[numeric_features + ['rentals']].describe()
print(bike_data.head())
```

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	\
0	1	1/1/2011	1	0	1	0	6	0	
1	2	1/2/2011	1	0	1	0	0	0	
2	3	1/3/2011	1	0	1	0	1	1	
3	4	1/4/2011	1	0	1	0	2	1	
4	5	1/5/2011	1	0	1	0	3	1	

	weathersit	temp	atemp	hum	windspeed	rentals	day
0	2	0.344167	0.363625	0.805833	0.160446	331	1
1	2	0.363478	0.353739	0.696087	0.248539	131	2
2	1	0.196364	0.189405	0.437273	0.248309	120	3
3	1	0.200000	0.212122	0.590435	0.160296	108	4
4	1	0.226957	0.229270	0.436957	0.186900	82	5

## Características y etiquetas separadas

Después de separar el conjunto de datos, ahora tenemos matrices entumecidas llamadas **X** que contienen las características y **y** que contienen las etiquetas.

```
In [ ]: X, y = bike_data[['season', 'mnth', 'holiday', 'weekday', 'workingday', 'weathersit', 'temp', 'atemp', 'hum', 'windspeed', 'rentals']]
```

## Dividir los datos 70%-30% en conjunto de entrenamiento y conjunto de prueba

```
In [ ]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30, random_state=42)
print ('Training Set: %d rows\nTest Set: %d rows' % (X_train.shape[0], X_test.shape[0]))
Training Set: 511 rows
Test Set: 220 rows
```

## Encaja un modelo de lazo en el set de entrenamiento

```
In [ ]: from sklearn.linear_model import Lasso

model = Lasso().fit(X_train, y_train)
print (model, "\n")

Lasso()
```

## Evaluar el modelo utilizando los datos de prueba

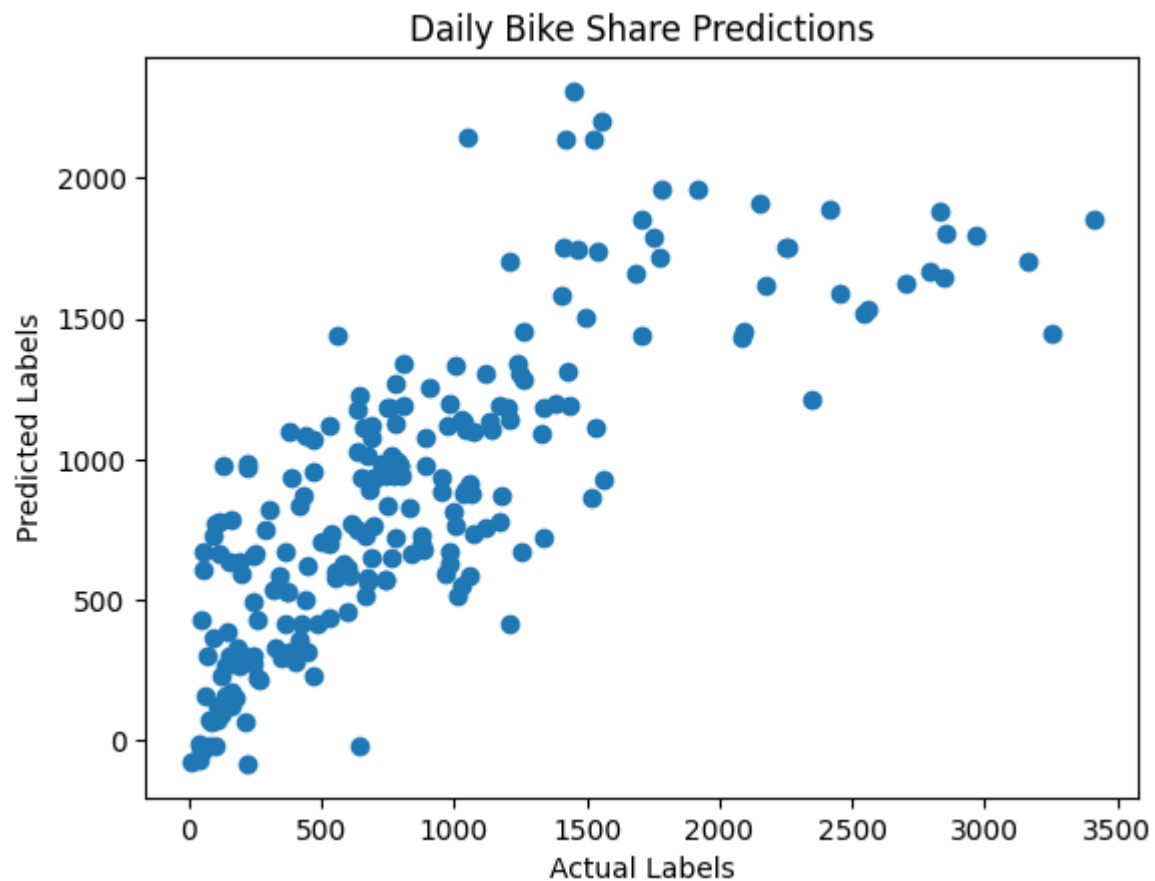
```
In [ ]: predictions = model.predict(X_test)
mse = mean_squared_error(y_test, predictions)
print("MSE:", mse)
rmse = np.sqrt(mse)
print("RMSE:", rmse)
r2 = r2_score(y_test, predictions)
print("R2:", r2)
```

```
MSE: 201155.70593338404
RMSE: 448.5038527519959
R2: 0.6056468637824488
```

## Gráfico predicho vs real

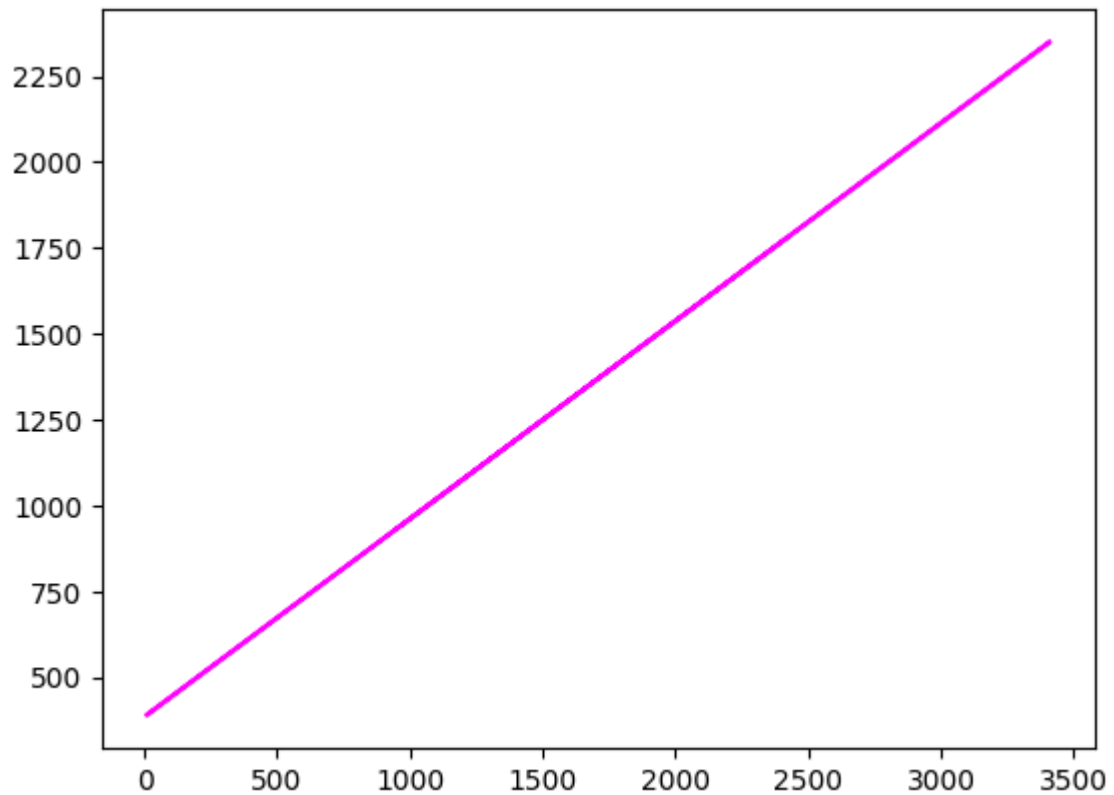
```
In [ ]: plt.scatter(y_test, predictions)
plt.xlabel('Actual Labels')
plt.ylabel('Predicted Labels')
plt.title('Daily Bike Share Predictions')
```

```
Out[ ]: Text(0.5, 1.0, 'Daily Bike Share Predictions')
```



## superponer la línea de regresión

```
In [ ]: z = np.polyfit(y_test, predictions, 1)
p = np.poly1d(z)
plt.plot(y_test, p(y_test), color='magenta')
plt.show()
```



## Entrenar el modelo

```
In [ ]: from sklearn.tree import DecisionTreeRegressor
        from sklearn.tree import export_text

        model = DecisionTreeRegressor().fit(X_train, y_train)
        print (model, "\n")
```

DecisionTreeRegressor()

## Visualizar el árbol del modelo

```
In [ ]: tree = export_text(model)
        print(tree)
```

```

|--- feature_6 <= 0.45
|   |--- feature_4 <= 0.50
|       |--- feature_7 <= 0.32
|           |--- feature_8 <= 0.41
|               |--- feature_1 <= 2.50
|                   |--- feature_9 <= 0.33
|                       |--- feature_8 <= 0.36
|                           |--- value: [558.00]
|                               |--- feature_8 > 0.36
|                                   |--- value: [515.00]
|                                       |--- feature_9 > 0.33
|                                           |--- value: [317.00]
|                                               |--- feature_1 > 2.50
|                                                   |--- feature_8 <= 0.40
|                                                       |--- feature_7 <= 0.27
|                                                           |--- value: [981.00]
|                                                               |--- feature_7 > 0.27
|                                                                   |--- value: [968.00]
|                                                                       |--- feature_8 > 0.40
|                                                                           |--- feature_7 <= 0.28
|                                                                               |--- value: [532.00]
|                                                                                   |--- feature_7 > 0.28
|                                                                                       |--- value: [710.00]
|                                                                                           |--- feature_8 > 0.41
|                                                                                               |--- feature_7 <= 0.25
|                                                                                                   |--- feature_6 <= 0.18
|                                                                                                       |--- feature_8 <= 0.43
|                                                                                                           |--- value: [284.00]
|                                                                                                               |--- feature_8 > 0.43
|                                                                                                                   |--- feature_7 <= 0.10
|                                                                                                                       |--- value: [150.00]
|                                                                                                                           |--- feature_7 > 0.10
|                                                                                                                               |--- feature_2 <= 0.50
|                                                                                                                                   |--- feature_8 <= 0.50
|                                                                                                                                       |--- value: [73.00]
|                                                                                                                                           |--- feature_8 > 0.50
|                                                                                                                                               |--- feature_8 <= 0.68
|                                                                                                       |--- value: [68.00]
|                                                                                                           |--- feature_8 > 0.68
|                                                                                                               |--- value: [67.00]
|                                                                                               |--- feature_2 > 0.50
|                                                                                                   |--- value: [117.00]
|                                                                                                       |--- feature_6 > 0.18
|                                                                                                           |--- feature_9 <= 0.17
|                                                                                                               |--- feature_6 <= 0.21
|                                                                                                                   |--- value: [123.00]
|                                                                                                                       |--- feature_6 > 0.21
|                                                                                                                           |--- value: [140.00]
|                                                                                                       |--- feature_9 > 0.17
|                                                                                                           |--- feature_6 <= 0.19
|                                                                                                               |--- value: [333.00]
|                                                                                               |--- feature_6 > 0.19
|                                                                                                   |--- feature_8 <= 0.53
|                                                                                                       |--- feature_9 <= 0.21
|                                                                                                           |--- value: [251.00]
|                                                                                                               |--- feature_9 > 0.21
|                                                                                                                   |--- feature_2 <= 0.50
|                                                                                                                       |--- value: [205.00]
|                                                                                                                           |--- feature_2 > 0.50
|                                                                                                                               |--- value: [217.00]

```

```

|--- feature_8 > 0.53
|   |--- feature_1 <= 7.00
|   |   |--- value: [288.00]
|   |   |--- feature_1 > 7.00
|   |   |   |--- value: [275.00]
|--- feature_7 > 0.25
|   |--- feature_9 <= 0.11
|   |   |--- value: [706.00]
|   |--- feature_9 > 0.11
|   |   |--- feature_8 <= 0.54
|   |   |   |--- feature_5 <= 1.50
|   |   |   |   |--- feature_7 <= 0.26
|   |   |   |   |   |--- value: [309.00]
|   |   |   |   |   |--- feature_7 > 0.26
|   |   |   |   |   |   |--- feature_0 <= 2.50
|   |   |   |   |   |   |   |--- feature_9 <= 0.16
|   |   |   |   |   |   |   |   |--- value: [408.00]
|   |   |   |   |   |   |   |   |--- feature_9 > 0.16
|   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |--- feature_0 > 2.50
|   |   |   |   |   |   |   |   |   |--- feature_8 <= 0.48
|   |   |   |   |   |   |   |   |   |   |--- value: [440.00]
|   |   |   |   |   |   |   |   |   |   |--- feature_8 > 0.48
|   |   |   |   |   |   |   |   |   |   |   |--- value: [502.00]
|   |   |   |   |   |   |   |--- feature_5 > 1.50
|   |   |   |   |   |   |   |   |--- value: [618.00]
|   |--- feature_8 > 0.54
|   |   |--- feature_6 <= 0.29
|   |   |   |--- feature_8 <= 0.63
|   |   |   |   |--- value: [354.00]
|   |   |   |   |--- feature_8 > 0.63
|   |   |   |   |   |--- value: [318.00]
|   |   |--- feature_6 > 0.29
|   |   |   |--- feature_1 <= 7.00
|   |   |   |   |--- value: [195.00]
|   |   |   |   |--- feature_1 > 7.00
|   |   |   |   |   |--- value: [155.00]
|--- feature_7 > 0.32
|   |--- feature_9 <= 0.25
|   |   |--- feature_6 <= 0.37
|   |   |   |--- feature_7 <= 0.36
|   |   |   |   |--- feature_6 <= 0.36
|   |   |   |   |   |--- feature_1 <= 10.50
|   |   |   |   |   |   |--- feature_7 <= 0.33
|   |   |   |   |   |   |   |--- feature_9 <= 0.21
|   |   |   |   |   |   |   |   |--- feature_9 <= 0.20
|   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |--- feature_9 > 0.20
|   |   |   |   |   |   |   |   |   |   |--- value: [1047.00]
|   |   |   |   |   |   |   |--- feature_9 > 0.21
|   |   |   |   |   |   |   |   |--- value: [724.00]
|   |   |--- feature_7 > 0.33
|   |   |   |--- feature_8 <= 0.83
|   |   |   |   |--- feature_9 <= 0.16
|   |   |   |   |   |--- value: [694.00]
|   |   |   |   |   |--- feature_9 > 0.16
|   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |--- feature_8 > 0.83
|   |   |   |   |   |   |--- value: [879.00]
|   |--- feature_1 > 10.50

```

```

|--- feature_8 <= 0.57
|   |--- feature_8 <= 0.51
|   |   |--- value: [943.00]
|   |--- feature_8 > 0.51
|   |   |--- feature_7 <= 0.33
|   |   |   |--- value: [1156.00]
|   |   |--- feature_7 > 0.33
|   |       |--- truncated branch of depth 2
|--- feature_8 > 0.57
|   |--- feature_9 <= 0.10
|   |   |--- feature_8 <= 0.66
|   |   |   |--- value: [955.00]
|   |   |--- feature_8 > 0.66
|   |       |--- truncated branch of depth 2
|   |--- feature_9 > 0.10
|   |   |--- feature_1 <= 11.50
|   |   |   |--- value: [922.00]
|   |   |--- feature_1 > 11.50
|   |       |--- value: [767.00]
|--- feature_6 > 0.36
|   |--- value: [1658.00]
|--- feature_7 > 0.36
|   |--- feature_1 <= 6.00
|   |   |--- value: [331.00]
|   |--- feature_1 > 6.00
|   |   |--- feature_1 <= 11.50
|   |   |   |--- value: [560.00]
|   |   |--- feature_1 > 11.50
|   |       |--- value: [538.00]
|--- feature_6 > 0.37
|   |--- feature_9 <= 0.24
|   |   |--- feature_9 <= 0.15
|   |   |   |--- feature_8 <= 0.67
|   |   |   |   |--- feature_1 <= 10.50
|   |   |   |   |   |--- value: [2252.00]
|   |   |   |   |--- feature_1 > 10.50
|   |   |   |       |--- value: [2290.00]
|   |   |   |--- feature_8 > 0.67
|   |   |   |   |--- feature_1 <= 10.50
|   |   |   |   |   |--- value: [1619.00]
|   |   |   |--- feature_1 > 10.50
|   |   |   |   |--- feature_1 <= 11.50
|   |   |   |   |   |--- value: [1249.00]
|   |   |   |   |--- feature_1 > 11.50
|   |   |   |       |--- value: [1153.00]
|   |   |--- feature_9 > 0.15
|   |   |   |--- feature_8 <= 0.51
|   |   |   |   |--- value: [1651.00]
|   |   |   |--- feature_8 > 0.51
|   |   |   |   |--- feature_9 <= 0.20
|   |   |   |   |   |--- feature_6 <= 0.42
|   |   |   |   |   |   |--- feature_1 <= 2.00
|   |   |   |   |   |   |   |--- value: [1070.00]
|   |   |   |   |   |   |--- feature_1 > 2.00
|   |   |   |   |       |--- truncated branch of depth 2
|   |   |   |--- feature_6 > 0.42
|   |   |   |   |--- value: [1188.00]
|   |   |--- feature_9 > 0.20
|   |   |   |--- feature_6 <= 0.43
|   |   |       |--- value: [665.00]

```

```

|--- feature_6 > 0.43
|--- value: [642.00]
|--- feature_9 > 0.24
|--- value: [2301.00]
--- feature_9 > 0.25
|--- feature_9 <= 0.32
|--- feature_1 <= 2.50
|--- value: [397.00]
|--- feature_1 > 2.50
|--- feature_8 <= 0.64
|--- value: [982.00]
|--- feature_8 > 0.64
|--- feature_7 <= 0.40
|--- value: [640.00]
|--- feature_7 > 0.40
|--- value: [480.00]
|--- feature_9 > 0.32
|--- feature_7 <= 0.41
|--- value: [120.00]
|--- feature_7 > 0.41
|--- value: [121.00]
--- feature_4 > 0.50
|--- feature_6 <= 0.34
|--- feature_1 <= 2.50
|--- feature_7 <= 0.29
|--- feature_7 <= 0.19
|--- feature_7 <= 0.14
|--- feature_8 <= 0.43
|--- value: [95.00]
|--- feature_8 > 0.43
|--- feature_3 <= 1.50
|--- value: [86.00]
|--- feature_3 > 1.50
|--- value: [89.00]
|--- feature_7 > 0.14
|--- feature_8 <= 0.46
|--- feature_9 <= 0.32
|--- value: [61.00]
|--- feature_9 > 0.32
|--- value: [75.00]
|--- feature_8 > 0.46
|--- feature_9 <= 0.30
|--- feature_9 <= 0.21
|--- feature_3 <= 2.00
|--- value: [42.00]
|--- feature_3 > 2.00
|--- truncated branch of depth 2
|--- feature_9 > 0.21
|--- feature_6 <= 0.16
|--- value: [41.00]
|--- feature_6 > 0.16
|--- value: [38.00]
|--- feature_9 > 0.30
|--- value: [25.00]
|--- feature_7 > 0.19
|--- feature_3 <= 4.50
|--- feature_9 <= 0.26
|--- feature_3 <= 2.50
|--- feature_8 <= 0.60
|--- feature_9 <= 0.19

```



```

|--- truncated branch of depth 2
|--- feature_9 > 0.19
|--- truncated branch of depth 2
|--- feature_8 > 0.60
|--- feature_8 <= 0.66
|--- value: [186.00]
|--- feature_8 > 0.66
|--- truncated branch of depth 2
--- feature_3 > 2.50
|--- feature_8 <= 0.51
|--- feature_6 <= 0.22
|--- truncated branch of depth 2
|--- feature_6 > 0.22
|--- value: [82.00]
|--- feature_8 > 0.51
|--- feature_7 <= 0.23
|--- value: [15.00]
|--- feature_7 > 0.23
|--- truncated branch of depth 4
--- feature_9 > 0.26
|--- feature_5 <= 2.50
|--- feature_9 <= 0.31
|--- value: [72.00]
|--- feature_9 > 0.31
|--- value: [64.00]
|--- feature_5 > 2.50
|--- value: [34.00]
--- feature_3 > 4.50
|--- feature_8 <= 0.47
|--- value: [115.00]
|--- feature_8 > 0.47
|--- feature_9 <= 0.27
|--- feature_9 <= 0.14
|--- value: [149.00]
|--- feature_9 > 0.14
|--- value: [148.00]
|--- feature_9 > 0.27
|--- value: [174.00]
--- feature_7 > 0.29
|--- feature_3 <= 4.50
|--- feature_5 <= 1.50
|--- feature_6 <= 0.31
|--- feature_6 <= 0.29
|--- value: [206.00]
|--- feature_6 > 0.29
|--- feature_9 <= 0.20
|--- value: [173.00]
|--- feature_9 > 0.20
|--- value: [163.00]
|--- feature_6 > 0.31
|--- value: [218.00]
--- feature_5 > 1.50
|--- feature_7 <= 0.33
|--- value: [74.00]
|--- feature_7 > 0.33
|--- value: [135.00]
--- feature_3 > 4.50
|--- feature_5 <= 1.50
|--- feature_6 <= 0.32
|--- value: [310.00]

```

```

|--- feature_6 > 0.32
|   |--- value: [307.00]
|--- feature_5 > 1.50
|   |--- value: [227.00]
--- feature_1 > 2.50
    |--- feature_9 <= 0.20
        |--- feature_9 <= 0.12
            |--- feature_9 <= 0.06
                |--- feature_7 <= 0.33
                    |--- value: [362.00]
                |--- feature_7 > 0.33
                    |--- value: [337.00]
            --- feature_9 > 0.06
                |--- feature_3 <= 3.50
                    |--- feature_5 <= 1.50
                        |--- value: [143.00]
                    --- feature_5 > 1.50
                        |--- feature_3 <= 2.00
                            |--- value: [174.00]
                        |--- feature_3 > 2.00
                            |--- value: [178.00]
                    --- feature_3 > 3.50
                        |--- feature_3 <= 4.50
                            |--- feature_7 <= 0.28
                                |--- value: [254.00]
                            |--- feature_7 > 0.28
                                |--- value: [243.00]
                        --- feature_3 > 4.50
                            |--- feature_6 <= 0.30
                                |--- value: [261.00]
                            |--- feature_6 > 0.30
                                |--- value: [268.00]
                    --- feature_9 > 0.12
                        |--- feature_8 <= 0.64
                            |--- feature_8 <= 0.45
                                |--- feature_9 <= 0.14
                                    |--- value: [316.00]
                                |--- feature_9 > 0.14
                                    |--- value: [245.00]
                            --- feature_8 > 0.45
                                |--- feature_9 <= 0.14
                                    |--- feature_8 <= 0.64
                                        |--- value: [491.00]
                                    |--- feature_8 > 0.64
                                        |--- value: [429.00]
                                --- feature_9 > 0.14
                                    |--- feature_6 <= 0.30
                                        |--- feature_9 <= 0.17
                                            |--- truncated branch of depth 3
                                        |--- feature_9 > 0.17
                                            |--- truncated branch of depth 2
                                    |--- feature_6 > 0.30
                                        |--- feature_0 <= 2.50
                                            |--- value: [359.00]
                                        |--- feature_0 > 2.50
                                            |--- truncated branch of depth 3
                                --- feature_8 > 0.64
                                    |--- feature_7 <= 0.31
                                        |--- value: [168.00]
                                    |--- feature_7 > 0.31

```

```

|--- feature_9 <= 0.13
|   |--- value: [349.00]
|--- feature_9 > 0.13
|   |--- feature_1 <= 7.50
|   |   |--- value: [289.00]
|   |--- feature_1 > 7.50
|   |   |--- value: [314.00]
|--- feature_9 > 0.20
|   |--- feature_8 <= 0.78
|   |   |--- feature_9 <= 0.23
|   |   |   |--- feature_9 <= 0.21
|   |   |   |   |--- feature_0 <= 3.00
|   |   |   |   |   |--- feature_9 <= 0.21
|   |   |   |   |   |   |--- value: [221.00]
|   |   |   |   |   |--- feature_9 > 0.21
|   |   |   |   |   |   |--- value: [222.00]
|   |   |   |--- feature_0 > 3.00
|   |   |   |   |--- value: [198.00]
|   |   |--- feature_9 > 0.21
|   |   |   |--- feature_5 <= 1.50
|   |   |   |   |--- feature_6 <= 0.23
|   |   |   |   |   |--- value: [123.00]
|   |   |   |   |--- feature_6 > 0.23
|   |   |   |   |   |--- value: [137.00]
|   |   |   |--- feature_5 > 1.50
|   |   |   |   |--- feature_1 <= 3.50
|   |   |   |   |   |--- value: [191.00]
|   |   |   |   |--- feature_1 > 3.50
|   |   |   |   |   |--- value: [177.00]
|   |--- feature_9 > 0.23
|   |   |--- feature_3 <= 4.50
|   |   |   |--- feature_8 <= 0.56
|   |   |   |   |--- feature_0 <= 1.50
|   |   |   |   |   |--- feature_9 <= 0.26
|   |   |   |   |   |   |--- value: [203.00]
|   |   |   |   |   |--- feature_9 > 0.26
|   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |--- feature_0 > 1.50
|   |   |   |   |--- feature_8 <= 0.43
|   |   |   |   |   |--- value: [317.00]
|   |   |   |   |--- feature_8 > 0.43
|   |   |   |   |   |--- value: [326.00]
|   |   |--- feature_8 > 0.56
|   |   |   |--- feature_8 <= 0.62
|   |   |   |   |--- feature_1 <= 11.50
|   |   |   |   |   |--- value: [139.00]
|   |   |   |   |--- feature_1 > 11.50
|   |   |   |   |   |--- value: [150.00]
|   |   |   |--- feature_8 > 0.62
|   |   |   |   |--- feature_8 <= 0.71
|   |   |   |   |   |--- value: [247.00]
|   |   |   |   |--- feature_8 > 0.71
|   |   |   |   |   |--- value: [195.00]
|   |--- feature_3 > 4.50
|   |   |--- feature_9 <= 0.23
|   |   |   |--- value: [456.00]
|   |   |--- feature_9 > 0.23
|   |   |   |--- feature_0 <= 1.50
|   |   |   |   |--- value: [247.00]
|   |   |--- feature_0 > 1.50

```

file:///F:/9no/extraccion de datos/microsoft/experimentación modelos eficaces.html

[illegible]

[illegible]

```

|--- feature_7 > 0.37
|--- feature_7 <= 0.39
|--- value: [796.00]
|--- feature_7 > 0.39
|--- feature_8 <= 0.73
|--- feature_1 <= 7.00
|--- value: [447.00]
|--- feature_1 > 7.00
|--- value: [470.00]
|--- feature_8 > 0.73
|--- value: [548.00]
--- feature_6 > 0.45
--- feature_4 <= 0.50
|--- feature_8 <= 0.83
|--- feature_1 <= 10.50
|--- feature_9 <= 0.28
|--- feature_6 <= 0.81
|--- feature_8 <= 0.67
|--- feature_9 <= 0.09
|--- value: [3065.00]
|--- feature_9 > 0.09
|--- feature_8 <= 0.66
|--- feature_9 <= 0.15
|--- feature_9 <= 0.14
|--- truncated branch of depth 8
|--- feature_9 > 0.14
|--- truncated branch of depth 3
|--- feature_9 > 0.15
|--- feature_9 <= 0.16
|--- truncated branch of depth 4
|--- feature_9 > 0.16
|--- truncated branch of depth 10
|--- feature_8 > 0.66
|--- value: [3031.00]
--- feature_8 > 0.67
--- feature_7 <= 0.67
--- feature_7 <= 0.50
--- feature_8 <= 0.79
--- feature_6 <= 0.47
--- value: [1138.00]
--- feature_6 > 0.47
--- truncated branch of depth 2
--- feature_8 > 0.79
--- value: [2207.00]
--- feature_7 > 0.50
--- feature_7 <= 0.51
--- value: [3155.00]
--- feature_7 > 0.51
--- feature_6 <= 0.69
--- truncated branch of depth 7
--- feature_6 > 0.69
--- truncated branch of depth 6
--- feature_7 > 0.67
--- feature_6 <= 0.74
--- feature_3 <= 0.50
--- feature_6 <= 0.73
--- value: [1249.00]
--- feature_6 > 0.73
--- value: [1298.00]
--- feature 3 > 0.50

```

```
--- feature 9 > 0.10
```



```

--- feature_9 <= 0.23
|--- feature_3 <= 2.50
|   |--- feature_8 <= 0.64
|   |   |--- feature_7 <= 0.46
|   |   |   |--- feature_1 <= 6.50
|   |   |   |   |--- value: [838.00]
|   |   |   |   |--- feature_1 > 6.50
|   |   |   |   |   |--- value: [922.00]
|   |   |   |--- feature_7 > 0.46
|   |   |   |   |--- feature_0 <= 3.00
|   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |--- feature_0 > 3.00
|   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |--- feature_8 > 0.64
|   |   |   |--- feature_9 <= 0.14
|   |   |   |   |--- feature_7 <= 0.49
|   |   |   |   |   |--- value: [699.00]
|   |   |   |   |--- feature_7 > 0.49
|   |   |   |   |   |--- value: [637.00]
|   |   |--- feature_9 > 0.14
|   |   |   |--- feature_0 <= 3.00
|   |   |   |   |--- value: [409.00]
|   |   |   |--- feature_0 > 3.00
|   |   |   |   |--- value: [486.00]
|--- feature_3 > 2.50
|   |--- feature_7 <= 0.50
|   |   |--- feature_9 <= 0.13
|   |   |   |--- value: [655.00]
|   |   |--- feature_9 > 0.13
|   |   |   |--- feature_6 <= 0.46
|   |   |   |   |--- value: [516.00]
|   |   |   |--- feature_6 > 0.46
|   |   |   |   |--- truncated branch of depth 4
|   |--- feature_7 > 0.50
|   |   |--- feature_6 <= 0.54
|   |   |   |--- feature_6 <= 0.53
|   |   |   |   |--- value: [735.00]
|   |   |   |--- feature_6 > 0.53
|   |   |   |   |--- value: [695.00]
|   |   |--- feature_6 > 0.54
|   |   |   |--- feature_0 <= 3.00
|   |   |   |   |--- value: [550.00]
|   |   |   |--- feature_0 > 3.00
|   |   |   |   |--- value: [559.00]
--- feature_9 > 0.23
|--- feature_7 <= 0.47
|   |--- feature_6 <= 0.47
|   |   |--- feature_8 <= 0.43
|   |   |   |--- value: [745.00]
|   |   |--- feature_8 > 0.43
|   |   |   |--- value: [614.00]
|   |--- feature_6 > 0.47
|   |   |--- value: [471.00]
|--- feature_7 > 0.47
|   |--- feature_5 <= 1.50
|   |   |--- feature_3 <= 2.50
|   |   |   |--- value: [905.00]
|   |   |--- feature_3 > 2.50
|   |   |   |--- value: [834.00]
|   |--- feature_5 > 1.50

```

```

|--- value: [1008.00]
|--- feature_7 > 0.53
|--- feature_9 <= 0.13
|--- feature_7 <= 0.72
|--- feature_6 <= 0.68
|--- feature_8 <= 0.82
|--- feature_9 <= 0.09
|--- feature_7 <= 0.59
|--- truncated branch of depth 2
|--- feature_7 > 0.59
|--- truncated branch of depth 2
|--- feature_9 > 0.09
|--- feature_6 <= 0.64
|--- truncated branch of depth 6
|--- feature_6 > 0.64
|--- truncated branch of depth 3
|--- feature_8 > 0.82
|--- value: [1334.00]
|--- feature_6 > 0.68
|--- feature_9 <= 0.12
|--- feature_8 <= 0.59
|--- value: [1177.00]
|--- feature_8 > 0.59
|--- feature_5 <= 1.50
|--- value: [1363.00]
|--- feature_5 > 1.50
|--- truncated branch of depth 2
|--- feature_9 > 0.12
|--- feature_9 <= 0.13
|--- feature_3 <= 1.50
|--- value: [989.00]
|--- feature_3 > 1.50
|--- truncated branch of depth 2
|--- feature_9 > 0.13
|--- value: [1233.00]
|--- feature_7 > 0.72
|--- feature_8 <= 0.61
|--- feature_9 <= 0.12
|--- feature_9 <= 0.11
|--- value: [872.00]
|--- feature_9 > 0.11
|--- value: [921.00]
|--- feature_9 > 0.12
|--- value: [778.00]
|--- feature_8 > 0.61
|--- feature_8 <= 0.70
|--- value: [673.00]
|--- feature_8 > 0.70
|--- value: [568.00]
|--- feature_9 > 0.13
|--- feature_8 <= 0.53
|--- feature_7 <= 0.73
|--- feature_9 <= 0.15
|--- value: [1281.00]
|--- feature_9 > 0.15
|--- feature_8 <= 0.51
|--- feature_1 <= 6.50
|--- truncated branch of depth 6
|--- feature_1 > 6.50
|--- truncated branch of depth 4

```

```

|--- feature_8 > 0.51
|--- feature_6 <= 0.65
|--- value: [1242.00]
|--- feature_6 > 0.65
|--- value: [1032.00]
|--- feature_7 > 0.73
|--- value: [1405.00]
--- feature_8 > 0.53
|--- feature_5 <= 1.50
|--- feature_9 <= 0.17
|--- feature_9 <= 0.17
|--- feature_3 <= 3.50
|--- truncated branch of depth 7
|--- feature_3 > 3.50
|--- truncated branch of depth 2
|--- feature_9 > 0.17
|--- feature_3 <= 2.50
|--- value: [1128.00]
|--- feature_3 > 2.50
|--- value: [1198.00]
--- feature_9 > 0.17
|--- feature_6 <= 0.78
|--- feature_6 <= 0.72
|--- truncated branch of depth 7
|--- feature_6 > 0.72
|--- truncated branch of depth 5
|--- feature_6 > 0.78
|--- feature_7 <= 0.74
|--- value: [662.00]
|--- feature_7 > 0.74
|--- value: [606.00]
--- feature_5 > 1.50
|--- feature_3 <= 1.50
|--- feature_8 <= 0.76
|--- feature_7 <= 0.55
|--- truncated branch of depth 2
|--- feature_7 > 0.55
|--- truncated branch of depth 4
|--- feature_8 > 0.76
|--- feature_9 <= 0.20
|--- value: [653.00]
|--- feature_9 > 0.20
|--- value: [630.00]
--- feature_3 > 1.50
|--- feature_9 <= 0.19
|--- feature_9 <= 0.15
|--- truncated branch of depth 5
|--- feature_9 > 0.15
|--- truncated branch of depth 3
|--- feature_9 > 0.19
|--- feature_1 <= 8.00
|--- truncated branch of depth 5
|--- feature_1 > 8.00
|--- value: [428.00]
--- feature_3 > 4.50
|--- feature_1 <= 4.00
|--- value: [2469.00]
|--- feature_1 > 4.00
|--- feature_8 <= 0.72
|--- feature_9 <= 0.12

```

```

|--- feature_9 <= 0.08
|--- value: [1325.00]
|--- feature_9 > 0.08
|--- feature_8 <= 0.65
|--- feature_7 <= 0.62
|--- value: [1516.00]
|--- feature_7 > 0.62
|--- value: [1511.00]
|--- feature_8 > 0.65
|--- value: [1379.00]
|--- feature_9 > 0.12
|--- feature_7 <= 0.77
|--- feature_1 <= 6.50
|--- feature_8 <= 0.52
|--- feature_8 <= 0.36
|--- value: [898.00]
|--- feature_8 > 0.36
|--- truncated branch of depth 2
|--- feature_8 > 0.52
|--- feature_8 <= 0.58
|--- truncated branch of depth 2
|--- feature_8 > 0.58
|--- truncated branch of depth 3
|--- feature_1 > 6.50
|--- feature_6 <= 0.72
|--- feature_6 <= 0.69
|--- truncated branch of depth 2
|--- feature_6 > 0.69
|--- truncated branch of depth 3
|--- feature_6 > 0.72
|--- feature_8 <= 0.51
|--- value: [1366.00]
|--- feature_8 > 0.51
|--- truncated branch of depth 2
|--- feature_7 > 0.77
|--- feature_9 <= 0.17
|--- value: [829.00]
|--- feature_9 > 0.17
|--- value: [670.00]
|--- feature_8 > 0.72
|--- feature_9 <= 0.20
|--- feature_1 <= 5.50
|--- feature_6 <= 0.61
|--- value: [909.00]
|--- feature_6 > 0.61
|--- value: [1417.00]
|--- feature_1 > 5.50
|--- feature_7 <= 0.53
|--- value: [1182.00]
|--- feature_7 > 0.53
|--- feature_9 <= 0.16
|--- feature_9 <= 0.14
|--- truncated branch of depth 2
|--- feature_9 > 0.14
|--- truncated branch of depth 2
|--- feature_9 > 0.16
|--- value: [1045.00]
|--- feature_9 > 0.20
|--- feature_7 <= 0.57
|--- value: [529.00]

```

[illegible]

## Evaluar el modelo utilizando los datos de prueba

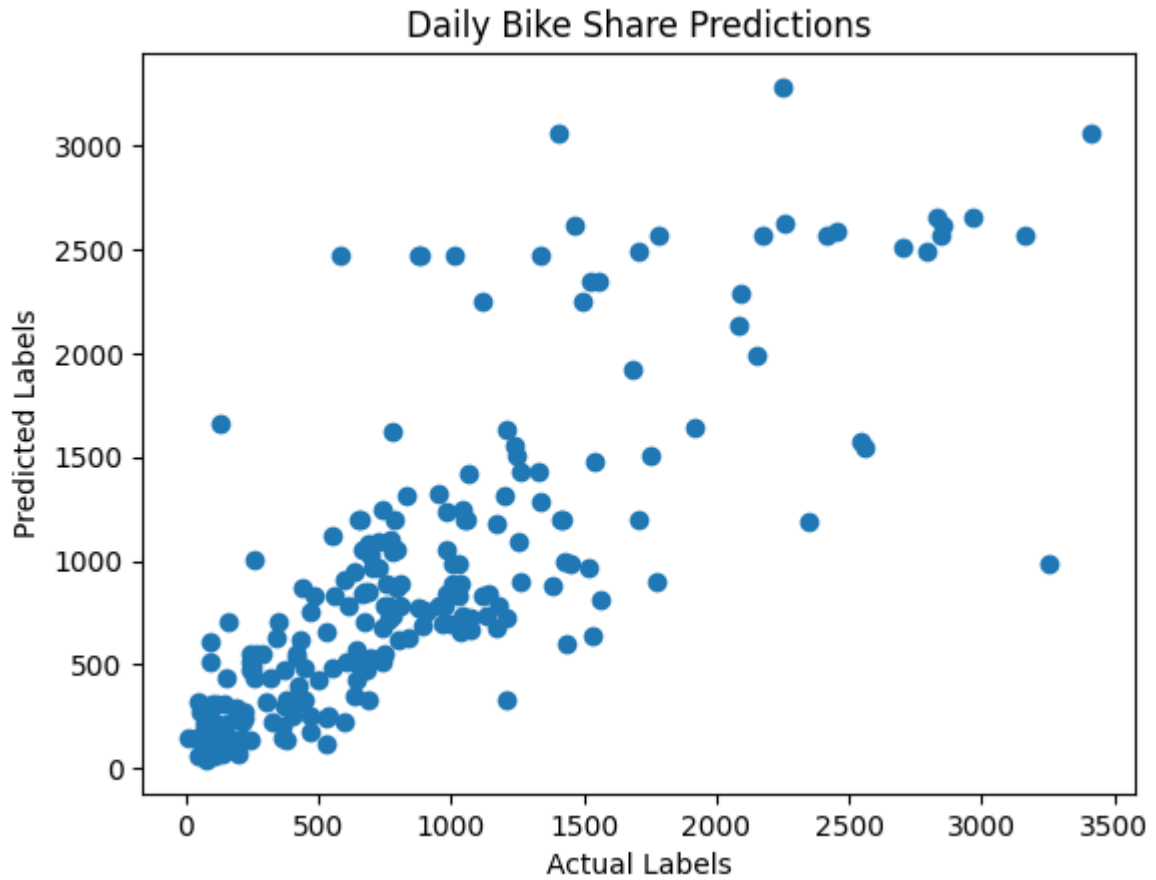
```
In [ ]: predictions = model.predict(X_test)
mse = mean_squared_error(y_test, predictions)
print("MSE:", mse)
rmse = np.sqrt(mse)
print("RMSE:", rmse)
r2 = r2_score(y_test, predictions)
print("R2:", r2)
```

MSE: 225757.46818181817  
RMSE: 475.139419730481  
R2: 0.5574166529906086

## Gráfico predicho vs real

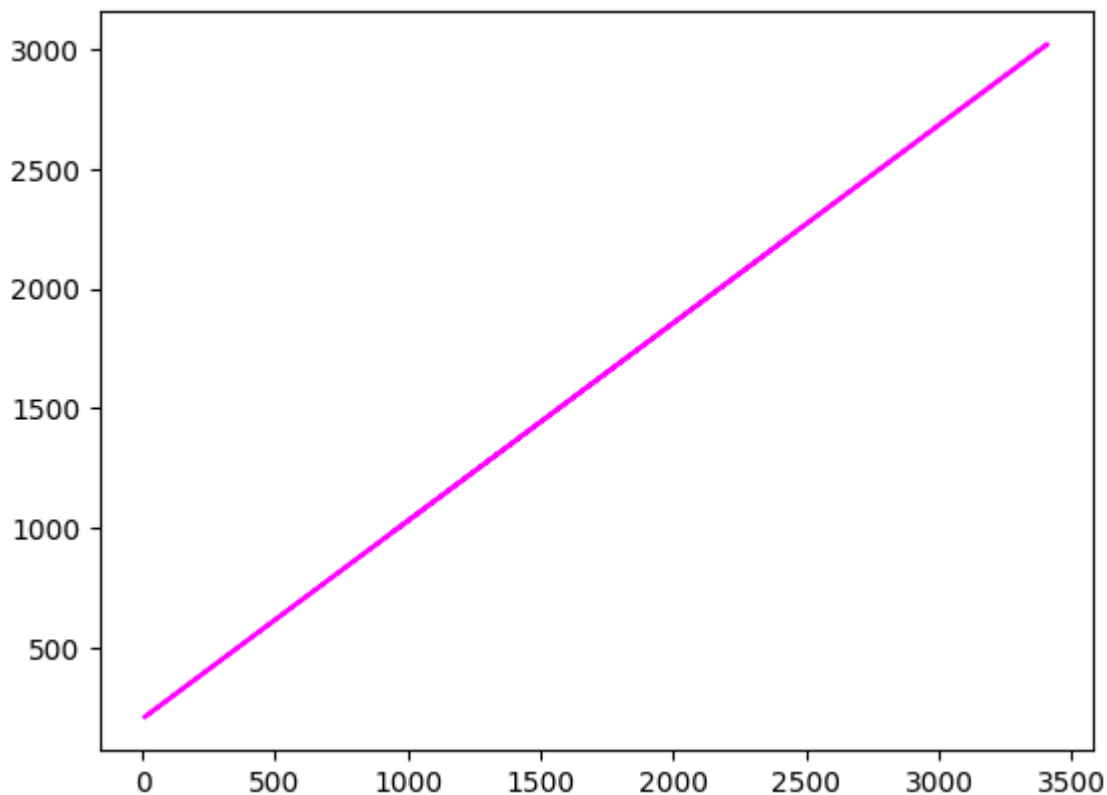
```
In [ ]: plt.scatter(y_test, predictions)
plt.xlabel('Actual Labels')
plt.ylabel('Predicted Labels')
plt.title('Daily Bike Share Predictions')
```

```
Out[ ]: Text(0.5, 1.0, 'Daily Bike Share Predictions')
```



## superponer la línea de regresión

```
In [ ]: z = np.polyfit(y_test, predictions, 1)
p = np.poly1d(z)
plt.plot(y_test, p(y_test), color='magenta')
plt.show()
```



## Entrenar el modelo

```
In [ ]: from sklearn.ensemble import RandomForestRegressor

model = RandomForestRegressor().fit(X_train, y_train)
print (model, "\n")

RandomForestRegressor()
```

## Evaluar el modelo utilizando los datos de prueba

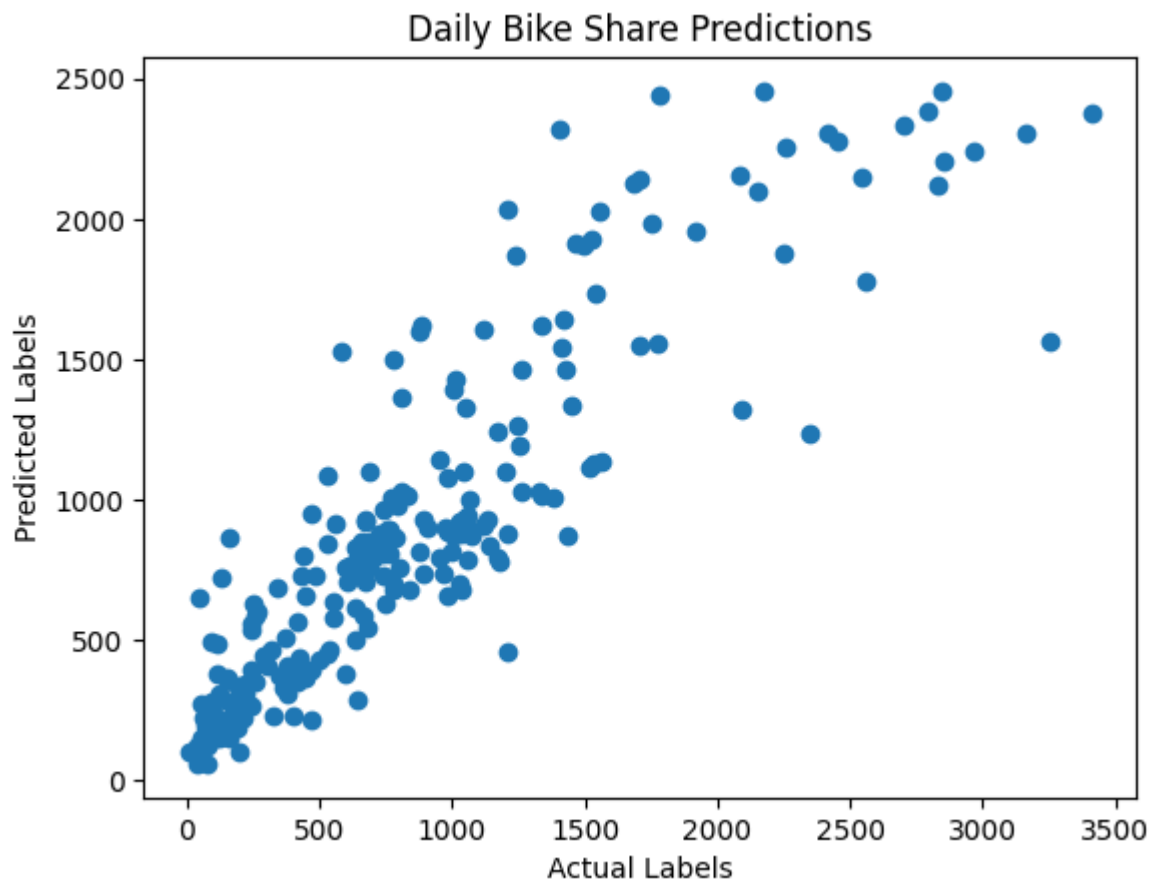
```
In [ ]: predictions = model.predict(X_test)
mse = mean_squared_error(y_test, predictions)
print("MSE:", mse)
rmse = np.sqrt(mse)
print("RMSE:", rmse)
r2 = r2_score(y_test, predictions)
print("R2:", r2)

MSE: 114270.9676290909
RMSE: 338.0398905885086
R2: 0.7759789400253472
```

## Gráfico predicho vs real

```
In [ ]: plt.scatter(y_test, predictions)
plt.xlabel('Actual Labels')
plt.ylabel('Predicted Labels')
plt.title('Daily Bike Share Predictions')
```

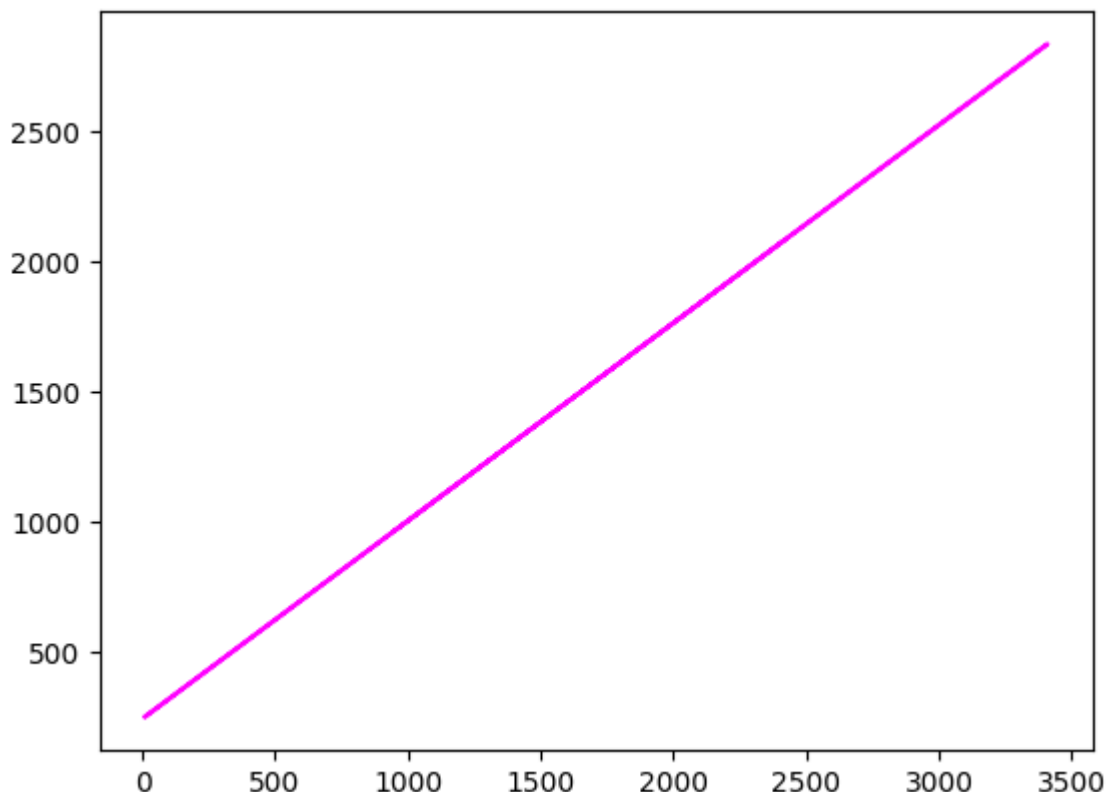
```
Out[ ]: Text(0.5, 1.0, 'Daily Bike Share Predictions')
```



## superponer la línea de regresión

```
In [ ]: z = np.polyfit(y_test, predictions, 1)
p = np.poly1d(z)
plt.plot(y_test, p(y_test), color='magenta')
plt.show()
```





## Entrenar el modelo

```
In [ ]: from sklearn.ensemble import GradientBoostingRegressor
```

## Encajar un modelo de lazo en el set de entrenamiento

```
In [ ]: model = GradientBoostingRegressor().fit(X_train, y_train)
print (model, "\n")
```

GradientBoostingRegressor()

## Evaluar el modelo utilizando los datos de prueba

```
In [ ]: predictions = model.predict(X_test)
mse = mean_squared_error(y_test, predictions)
print("MSE:", mse)
rmse = np.sqrt(mse)
print("RMSE:", rmse)
r2 = r2_score(y_test, predictions)
print("R2:", r2)
```

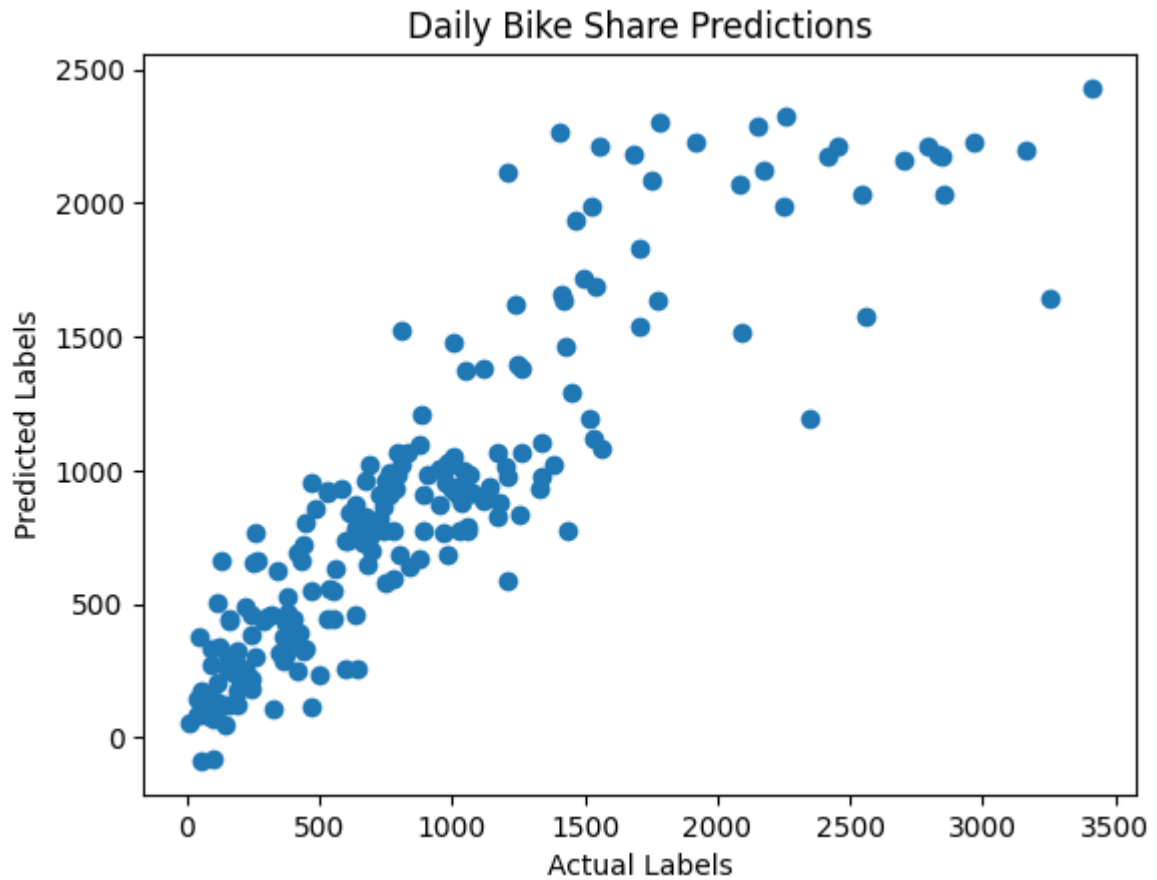
MSE: 103982.62941888033  
RMSE: 322.4633768645369  
R2: 0.7961485813528819

## Gráfico predicho vs real

```
In [ ]: plt.scatter(y_test, predictions)
```

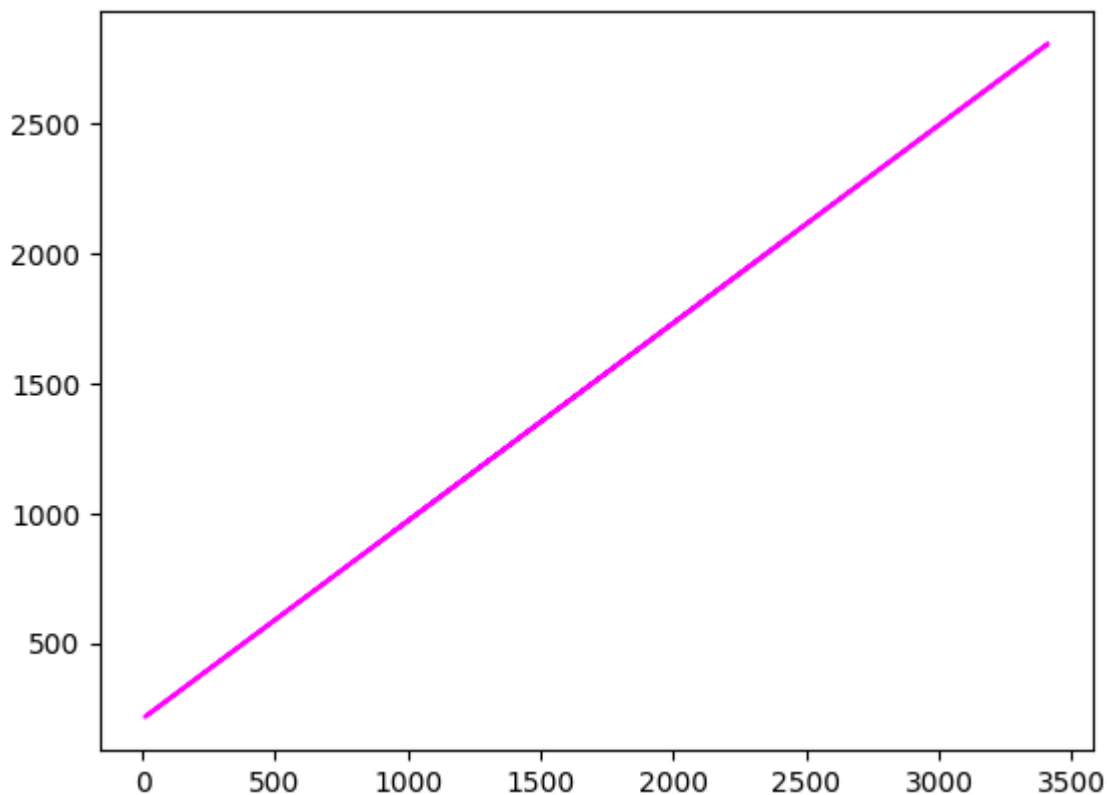
```
plt.xlabel('Actual Labels')
plt.ylabel('Predicted Labels')
plt.title('Daily Bike Share Predictions')
```

Out[ ]: Text(0.5, 1.0, 'Daily Bike Share Predictions')



## superponer la línea de regresión

```
In [ ]: z = np.polyfit(y_test, predictions, 1)
p = np.poly1d(z)
plt.plot(y_test, p(y_test), color='magenta')
plt.show()
```



## Optimización y guardado de modelos

### Importar módulos que necesitaremos para este portátil

```
In [ ]: import pandas as pd
        from sklearn.linear_model import LinearRegression
        from sklearn.metrics import mean_squared_error, r2_score
        from sklearn.model_selection import train_test_split
        import numpy as np
        import matplotlib.pyplot as plt
        %matplotlib inline
```

### Cargar el conjunto de datos de entrenamiento

```
In [ ]: bike_data = pd.read_csv('daily-bike-share.csv')
        bike_data['day'] = pd.DatetimeIndex(bike_data['dteday']).day
        numeric_features = ['temp', 'atemp', 'hum', 'windspeed']
        categorical_features = ['season', 'mnth', 'holiday', 'weekday', 'workingday', 'weathersit']
        bike_data[numeric_features + ['rentals']].describe()
        print(bike_data.head())
```

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	\
0	1	1/1/2011	1	0	1	0	6	0	
1	2	1/2/2011	1	0	1	0	0	0	
2	3	1/3/2011	1	0	1	0	1	1	
3	4	1/4/2011	1	0	1	0	2	1	
4	5	1/5/2011	1	0	1	0	3	1	

	weathersit	temp	atemp	hum	windspeed	rentals	day
0	2	0.344167	0.363625	0.805833	0.160446	331	1
1	2	0.363478	0.353739	0.696087	0.248539	131	2
2	1	0.196364	0.189405	0.437273	0.248309	120	3
3	1	0.200000	0.212122	0.590435	0.160296	108	4
4	1	0.226957	0.229270	0.436957	0.186900	82	5

## Características y etiquetas separadas

Después de separar el conjunto de datos, ahora tenemos matrices numpy llamadas **X** que contienen las características y **y** que contienen las etiquetas.

```
In [ ]: X, y = bike_data[['season', 'mnth', 'holiday', 'weekday', 'workingday', 'weathersit', 'temp
```

## Dividir los datos 70%-30% en conjunto de entrenamiento y conjunto de prueba

```
In [ ]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30, random_state=42)

print ('Training Set: %d rows\nTest Set: %d rows' % (X_train.shape[0], X_test.shape[0]))

Training Set: 511 rows
Test Set: 220 rows
```

## Entrenar el modelo

```
In [ ]: from sklearn.ensemble import GradientBoostingRegressor, RandomForestRegressor
```

## Encajar un modelo de lazo en el set de entrenamiento

```
In [ ]: model = GradientBoostingRegressor().fit(X_train, y_train)
print (model, "\n")

GradientBoostingRegressor()
```

## Evaluar el modelo utilizando los datos de prueba

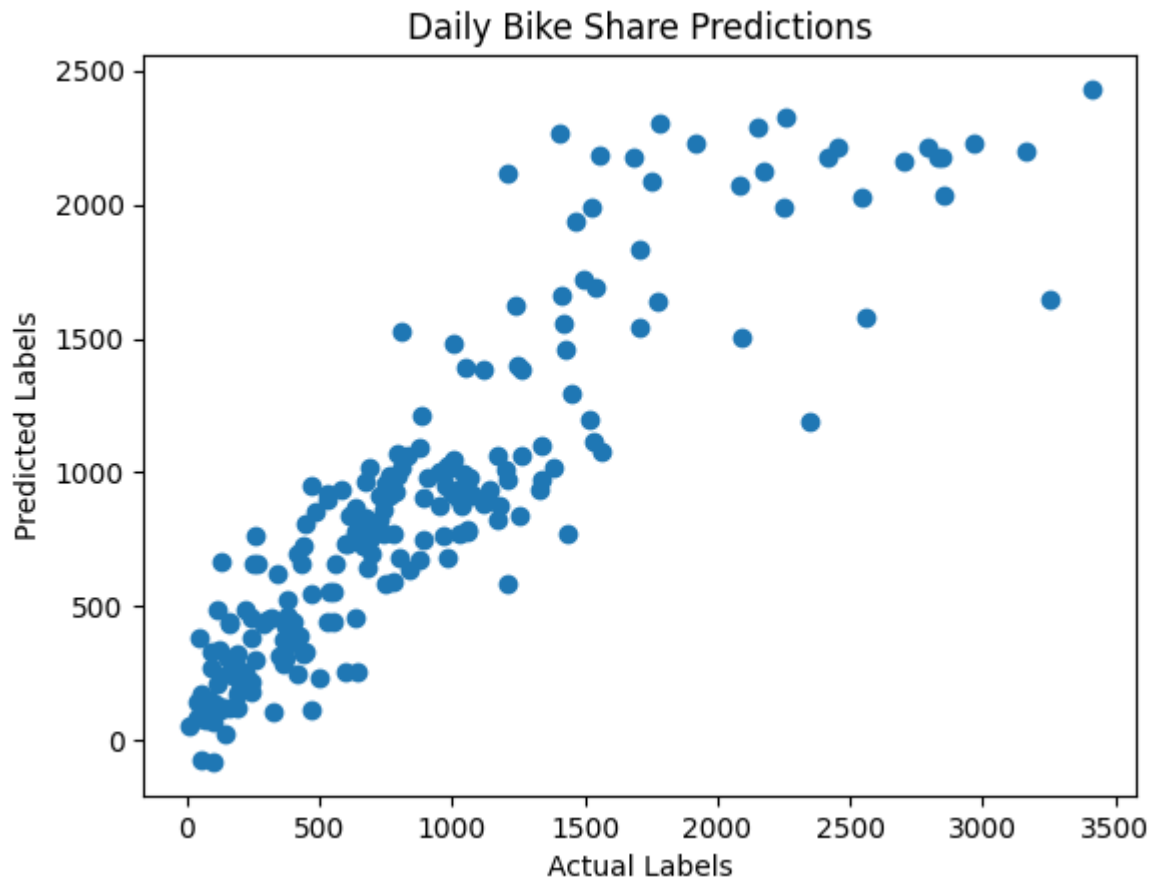
```
In [ ]: predictions = model.predict(X_test)
mse = mean_squared_error(y_test, predictions)
print("MSE:", mse)
rmse = np.sqrt(mse)
print("RMSE:", rmse)
r2 = r2_score(y_test, predictions)
print("R2:", r2)
```

MSE: 103749.79480282562  
RMSE: 322.1021496401811  
R2: 0.7966050390041085

## Gráfico predicho vs real

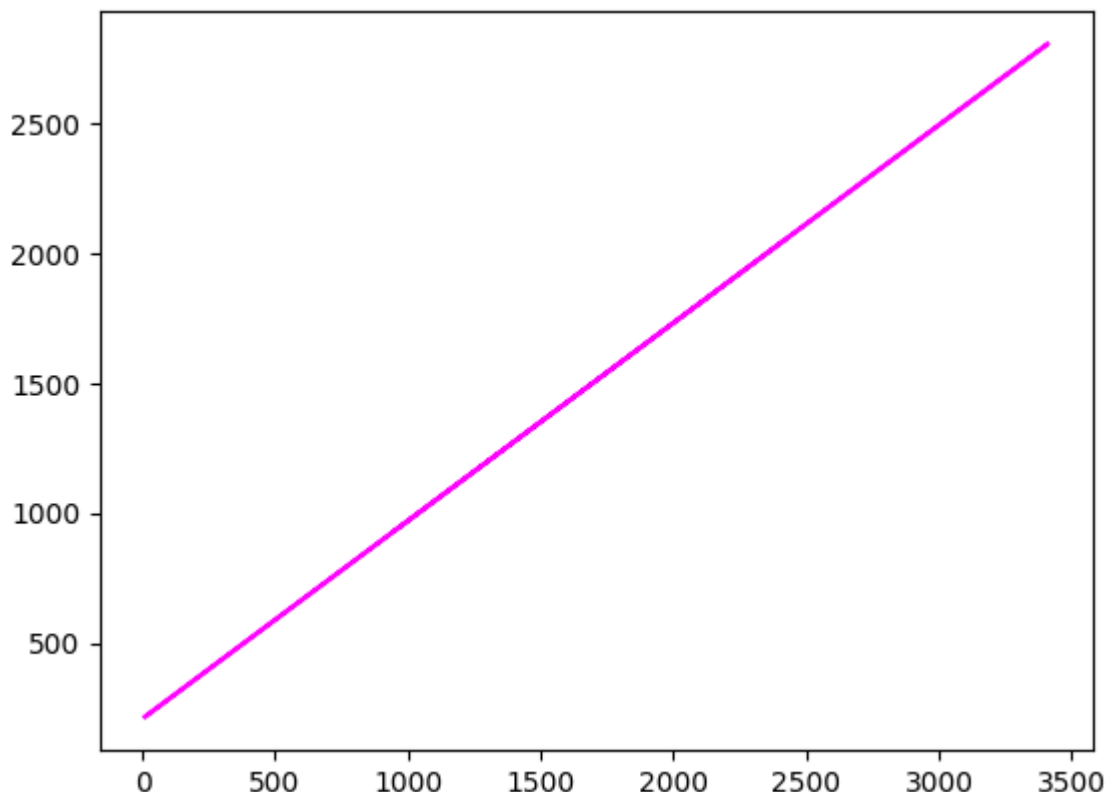
```
In [ ]: plt.scatter(y_test, predictions)
plt.xlabel('Actual Labels')
plt.ylabel('Predicted Labels')
plt.title('Daily Bike Share Predictions')
```

```
Out[ ]: Text(0.5, 1.0, 'Daily Bike Share Predictions')
```



## superponer la línea de regresión

```
In [ ]: z = np.polyfit(y_test, predictions, 1)
p = np.poly1d(z)
plt.plot(y_test, p(y_test), color='magenta')
plt.show()
```



## Usar un algoritmo de aumento de gradiente

```
In [ ]: from sklearn.model_selection import GridSearchCV
        from sklearn.metrics import make_scorer, r2_score
        alg = GradientBoostingRegressor()
```

## Pruebe estos valores de hiperparámetros

```
In [ ]: params = {
        'learning_rate': [0.1, 0.5, 1.0],
        'n_estimators' : [50, 100, 150]
        }
```

## Encuentra la mejor combinación de hiperparámetros para optimizar la métrica R2

```
In [ ]: score = make_scorer(r2_score)
        gridsearch = GridSearchCV(alg, params, scoring=score, cv=3, return_train_score=True)
        gridsearch.fit(X_train, y_train)
        print("Best parameter combination:", gridsearch.best_params_, "\n")
```

Best parameter combination: {'learning\_rate': 0.1, 'n\_estimators': 100}

## Consigue el mejor modelo

```
In [ ]: model=gridsearch.best_estimator_
        print(model, "\n")
```

GradientBoostingRegressor()

## Evaluar el modelo utilizando los datos de prueba

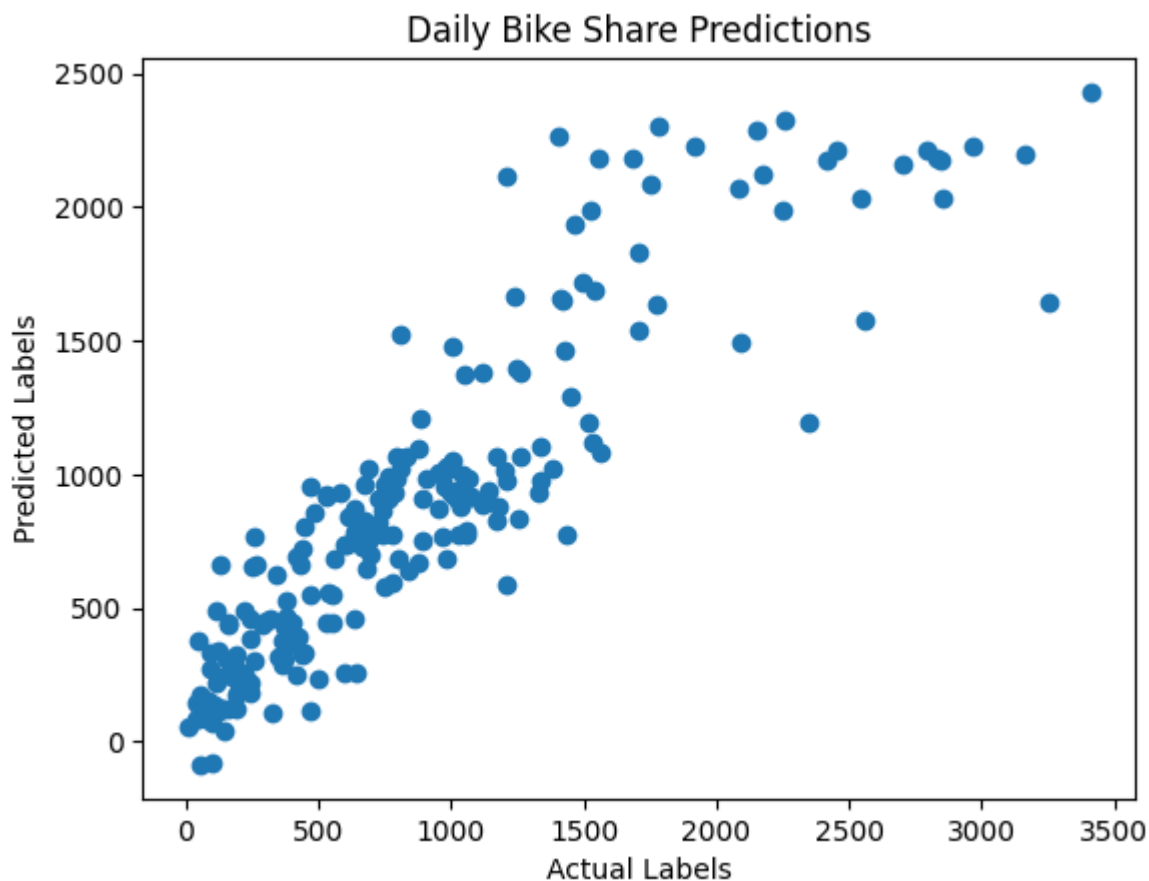
```
In [ ]: predictions = model.predict(X_test)
mse = mean_squared_error(y_test, predictions)
print("MSE:", mse)
rmse = np.sqrt(mse)
print("RMSE:", rmse)
r2 = r2_score(y_test, predictions)
print("R2:", r2)
```

MSE: 104215.62469388853  
RMSE: 322.82444872389783  
R2: 0.7956918087398606

## Gráfico predicho vs real

```
In [ ]: plt.scatter(y_test, predictions)
plt.xlabel('Actual Labels')
plt.ylabel('Predicted Labels')
plt.title('Daily Bike Share Predictions')
```

Out[ ]: Text(0.5, 1.0, 'Daily Bike Share Predictions')



## superponer la línea de regresión

```
In [ ]: z = np.polyfit(y_test, predictions, 1)
p = np.poly1d(z)
plt.plot(y_test, p(y_test), color='magenta')
plt.show()
```

