```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
```

```
df = pd.read_csv('CO2 Emissions_Canada.csv')
df.head()
```

	Make	Model	Vehicle Class	Engine Size(L)	Cylinders	Transmission	Fuel Type	Fuel Consumption City (L/100 km)	Fuel Consumption Hwy (L/100 km)	Fuel Consumption Comb (L/100 km)	Fuel Consumption Comb (mpg)	Em
0	ACURA	ILX	COMPACT	2.0	4	AS5	Z	9.9	6.7	8.5	33	
1	ACURA	ILX	COMPACT	2.4	4	M6	Z	11.2	7.7	9.6	29	
2	ACURA	ILX HYBRID	COMPACT	1.5	4	AV7	Z	6.0	5.8	5.9	48	
3	ACURA	MDX 4WD	SUV - SMALL	3.5	6	AS6	Z	12.7	9.1	11.1	25	
4	ACURA	RDX AWD	SUV - SMALL	3.5	6	AS6	Z	12.1	8.7	10.6	27	

Next steps: Generate code with df | New interactive sheet

```
X = df[['Engine Size(L)']]
y = df['CO2 Emissions(g/km)']
```

```
model = LinearRegression()
model.fit(X, y)

beta_1 = model.coef_[0]
print("Original Regression Coefficient (β₁):", beta_1)
```

```
Original Regression Coefficient (β₁): 36.77731518641943
```

```
B = 2000
n = len(df)
bootstrap_coefs = []

for i in range(B):
    sample = df.sample(n, replace=True)

    X_sample = sample[['Engine Size(L)']]
    y_sample = sample['CO2 Emissions(g/km)']

    model = LinearRegression()
    model.fit(X_sample, y_sample)

    bootstrap_coefs.append(model.coef_[0])

bootstrap_coefs = np.array(bootstrap_coefs)
```

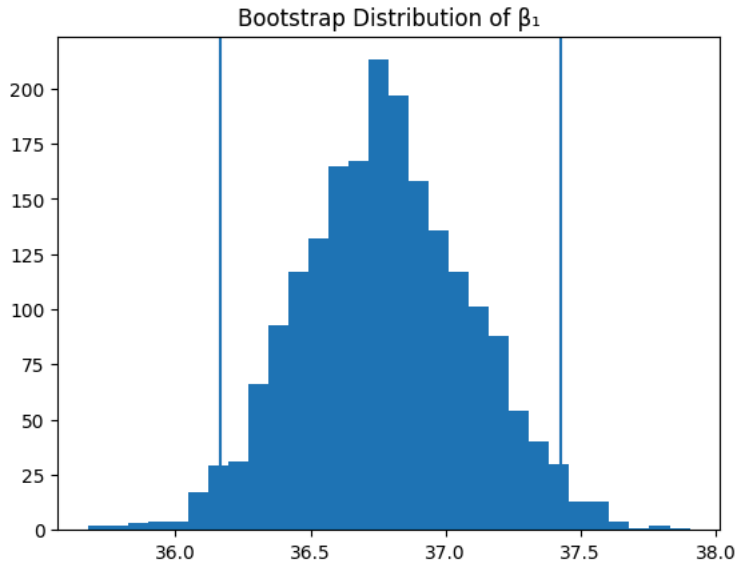
```
ci_90 = np.percentile(bootstrap_coefs, [5, 95])
print("90% Confidence Interval:", ci_90)
```

```
90% Confidence Interval: [36.27254783 37.31779341]
```

```
ci_95 = np.percentile(bootstrap_coefs, [2.5, 97.5])
print("95% Confidence Interval:", ci_95)
```

```
95% Confidence Interval: [36.16824772 37.42470936]
```

```
plt.hist(bootstrap_coefs, bins=30)
plt.axvline(ci_95[0])
plt.axvline(ci_95[1])
plt.title("Bootstrap Distribution of β₁")
plt.show()
```


Bootstrap Distribution of β₁
200
175
150
125
100
75
50
25
0
36.0
36.5
37.0
37.5
38.0