In [1]:

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

Matplotlib is building the font cache; this may take a moment.

In [7]:

```python
data=pd.read_csv("Q1_Dataset.txt")
data.shape
```
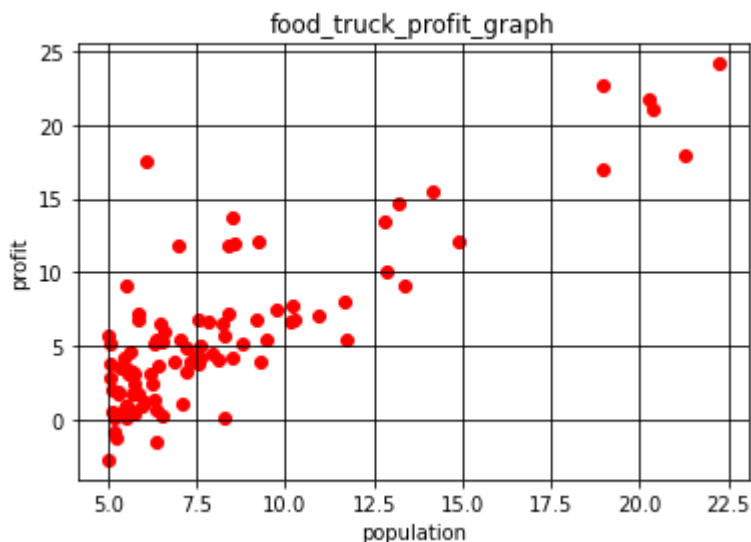
Out[7]:

(97, 2)

In [8]:

```python
x=data[['population']].values
y=data[['profit']].values
```

In [9]:

```python
%matplotlib inline
plt.scatter(x,y,c='r',label='scatter_data')
plt.xlabel("population")
plt.ylabel("profit")
plt.title('food_truck_profit_graph')
plt.grid(True,color='k')
plt.show()
```


In [10]:

```python
k = LinearRegression()
```

In [11]:

```python
k.fit(x,y)
```

Out[11]:

```
LinearRegression()
```

In [25]:

```python
print('c value:',k.intercept_)
```
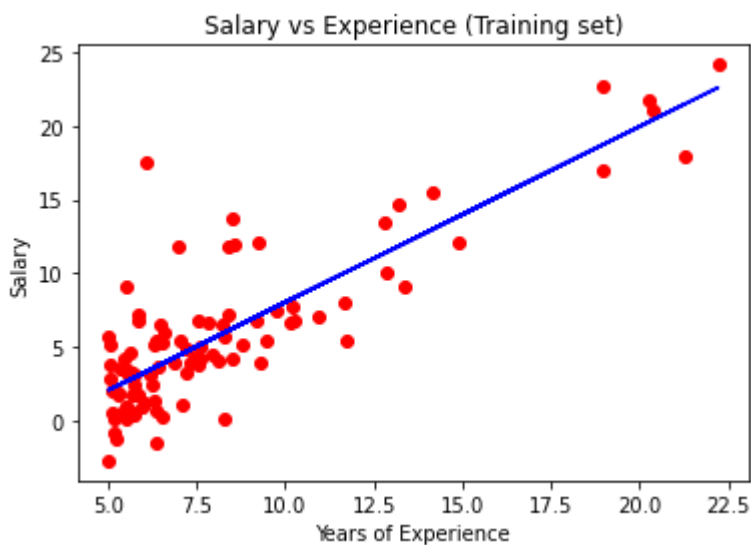
```
c value: [-3.89578088]
```

In [16]:

```python
print("m value:",k.coef_,)
```

```
m value: [[1.19303364]]
```

In [17]:

```python
y_pred = k.predict(x)
plt.scatter(x, y, color = 'red')
plt.plot(x,y_pred, color = 'blue')
plt.title('Salary vs Experience (Training set)')
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.show()
```


In [18]:

```python
from sklearn.metrics import r2_score
r_sq=r2_score(y,y_pred)
r_sq
```

Out[18]:

```
0.7020315537841397
```

In [20]:

```python
from sklearn.metrics import mean_squared_error
rmse=mean_squared_error(y,y_pred)
rmse
```

Out[20]:

8.953942751950358

In [23]:

```python
n1=4.5
n2=6.5
print("profit from 45000 people city is",k.predict([[n1]])*10000,'$')
print('profit from 65000 people city is ',k.predict([[n2]])*10000,'$')
```

```
profit from 45000 people city is [[14728.70520541]] $
profit from 65000 people city is  [[38589.37808921]] $
```

In [ ]: