Problem Statement : Mr Anil is planning to go on a budget vacation trip. He is confused with multiple choices. Develop an ML algorithm to help find the best budget trip based on the distance (kilometer) using DataScience methodology.

```python
import pandas as pd
print(pd.__version__)
```

```
1.3.5
```

```python
df = pd.read_csv("/content/Budget_Trip .csv")
df
```

	distance(Km)	Budget
0	100	1800
1	75	1350
2	50	900
3	150	2850
4	175	3325
5	200	3700
6	350	6300
7	400	7200
8	500	9500
9	520	9360
10	700	12950

```python
print(type(df))
```

```
<class 'pandas.core.frame.DataFrame'>
```

```python
X=df.drop('Budget',axis=1)
X
```

	distance(Km)
0	100
1	75
2	50
3	150
4	175
5	200
6	350

```python
y = df['Budget']
y
```

```
0      1800
1      1350
2       900
3      2850
4      3325
5      3700
6      6300
7      7200
8      9500
9      9360
10    12950
Name: Budget, dtype: int64
```

DATA VISUALIZATION

```python
from matplotlib import pyplot as plt
```

```python
plt.scatter(X,y,s=100,marker='*',color="red")
plt.plot(X,y,color='pink')
```

```
[<matplotlib.lines.Line2D at 0x7f73735a7a50>]
```

ML MODEL

```
from sklearn import linear_model
```

```
reg=linear_model.LinearRegression()
```

```
reg.fit(X,y)
```

```
LinearRegression()
```

```
reg.predict([[300]])
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/base.py:451: UserWarning: X does not
  "X does not have valid feature names, but"
array([5518.98000886])
```

```
reg.score(X,y)
```

```
0.9986895622712859
```

```
m=reg.coef_
m
```

```
array([18.42225122])
```

```
b=reg.intercept_
b
```

```
-7.695356738392547
```

```
inputX=300
output_Budget=m*inputX+b
output_Budget
```

```
array([5518.98000886])
```

```
import pickle
```
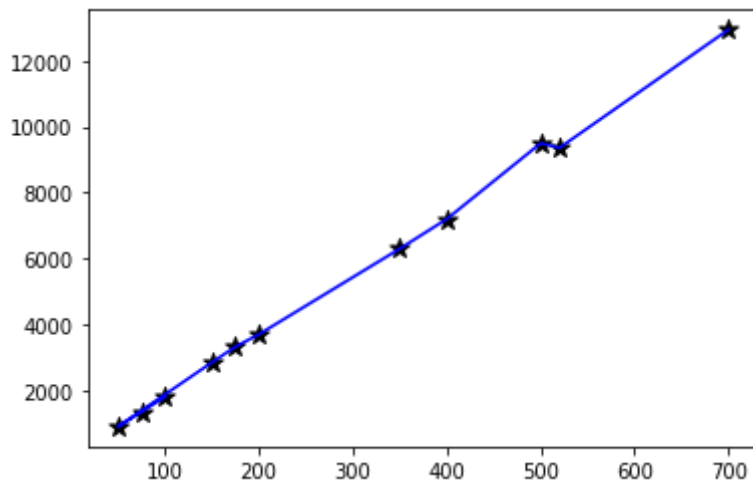
```
with open('BudgetTripmodel','wb') as f:
    pickle.dump(reg,f)
```

```
ytest=reg.predict(X)
ytest
```

```
array([ 1834.52976513,  1373.97348466,   913.4172042 ,  2755.64232606,
        3216.19860653,  3676.754887  ,  6440.0925698 ,  7361.20513073,
        9203.4302526 ,  9571.87527697, 12887.88049633])
```
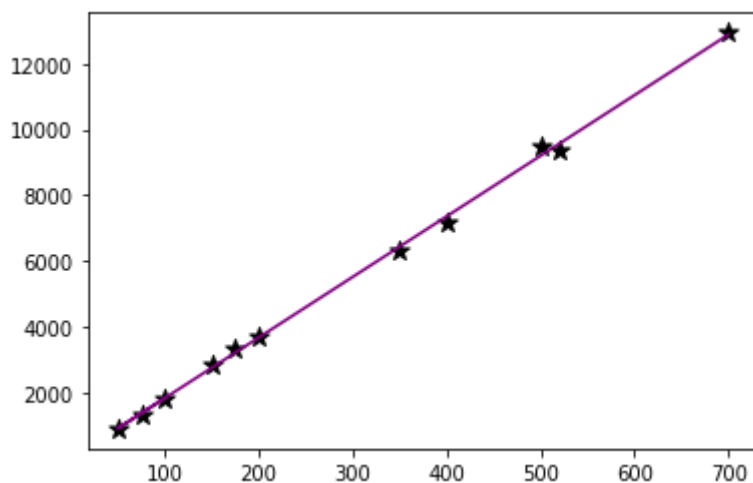
```
plt.scatter(X,y,s=100,marker='*',color="black")
plt.plot(X,y,color='blue')
```

```
[<matplotlib.lines.Line2D at 0x7f73669a9dd0>]
```



```
plt.scatter(X,y,s=100,marker='*',color="black")
plt.plot(X,ytest,color='purple')
```

```
[<matplotlib.lines.Line2D at 0x7f736695af90>]
```



```
with open('/content/BudgetTripmodel','rb') as f:
    reg=pickle.load(f)
```

```
reg.predict([[300]])
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/base.py:451: UserWarning: X does not
  "X does not have valid feature names, but"
array([5518.98000886])
```

```python
df=pd.read_csv('/content/ClientFile.csv')
df
```

	distance(Km)
0	500
1	780
2	900
3	1100
4	1400
5	1700
6	2100
7	3300
8	4500
9	4600

```python
predictedprice=reg.predict(df)
predictedprice
```

```
array([ 9203.4302526 , 14361.66059383, 16572.33074007, 20256.7809838 ,
        25783.4563494 , 31310.131715  , 38679.03220247, 60785.73366488,
        82892.43512728, 84734.66024915])
```

```python
print(type(predictedprice))
```

```
<class 'numpy.ndarray'>
```

```python
df['Budget']=predictedprice
df
```

	distance(Km)	Budget	
0	500	9203.430253	
1	780	14361.660594	
2	900	16572.330740	
3	1100	20256.780984	
4	1400	25783.456349	

```
df.to_csv('ClientOutputFile.csv',index=False)
```

6	2100	38679.032202	
7	3300	60785.733665	
8	4500	82892.435127	
9	4600	84734.660249	

✓   0s   completed at 7:16 PM