# P1

```
import numpy as np, pandas as pd, matplotlib.pyplot as plt, seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

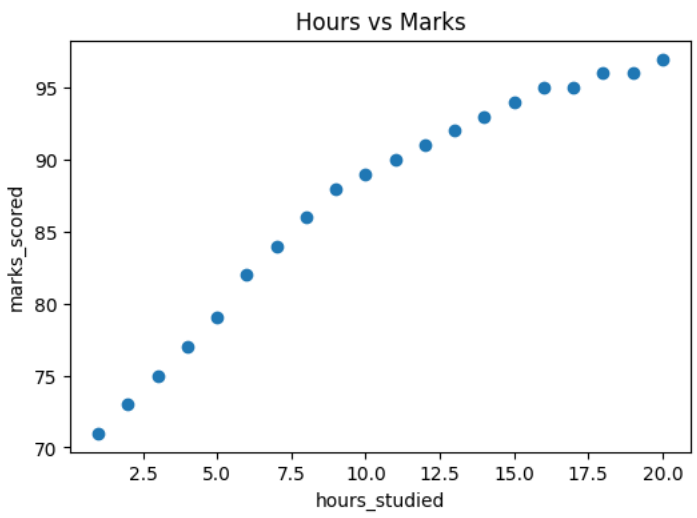
```
RANGE = 20

df = pd.DataFrame({
    'hours_studied': np.arange(1, RANGE+1),
    'marks_scored': [71, 73, 75, 77, 79, 82, 84, 86, 88, 89, 90, 91, 92, 93, 94, 95, 95, 96, 96, 97]
})
df.to_csv('data.csv', index=False)
print(df)
```

```
    hours_studied  marks_scored
0               1            71
1               2            73
2               3            75
3               4            77
4               5            79
5               6            82
6               7            84
7               8            86
8               9            88
9              10            89
10             11            90
11             12            91
12             13            92
13             14            93
14             15            94
15             16            95
16             17            95
17             18            96
18             19            96
19             20            97
```

```
df_copy = df.copy()
```

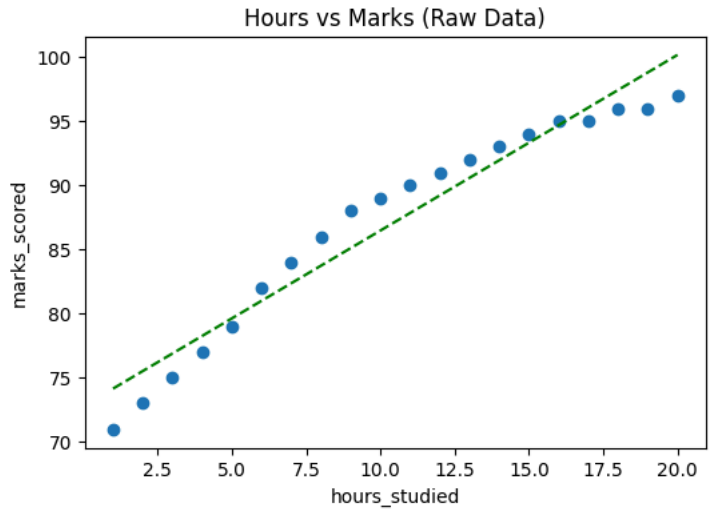
```
plt.figure(figsize=(6,4))
plt.scatter(df_copy['hours_studied'],df_copy['marks_scored'])
plt.xlabel('hours_studied')
plt.ylabel('marks_scored')
plt.title('Hours vs Marks')
plt.show()
```




```
x,y = df_copy['hours_studied'],df_copy['marks_scored']
m,c = np.polyfit(x,y,1)
print(f"Slope: {m}\nIntercept: {c}")
```

```
Slope: 1.3691729323308284
Intercept: 72.77368421052633
```

```
plt.figure(figsize=(6,4))
plt.scatter(df_copy['hours_studied'],df_copy['marks_scored'])
plt.plot(x,m*x+c,color='green',linestyle='--')
plt.xlabel('hours_studied')
plt.ylabel('marks_scored')
plt.title('Hours vs Marks (Raw Data)')
plt.show()
```



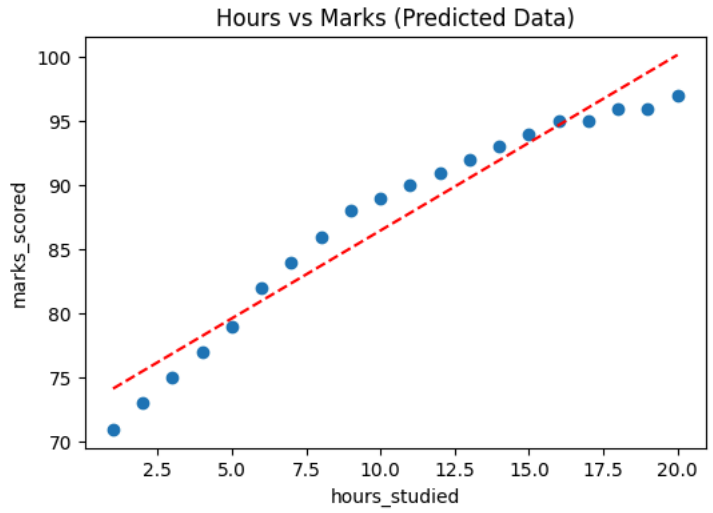

```
from sklearn.linear_model import LinearRegression

model = LinearRegression()
model.fit(df_copy[['hours_studied']],df_copy['marks_scored'])

m1, c1 = model.coef_ , model.intercept_
print("Slope:",m1[0],"\nIntercept:",c1)
```

```
Slope: 1.3691729323308273
Intercept: 72.77368421052631
```

```
  plt.figure(figsize=(6,4))
  plt.scatter(df_copy['hours_studied'],df_copy['marks_scored'])
  plt.plot(x,m1*x+c1,color='red',linestyle='--')
  plt.xlabel('hours_studied')
  plt.ylabel('marks_scored')
  plt.title('Hours vs Marks (Predicted Data)')
  plt.show()
```




```
test= pd.DataFrame({
    'hours_studied':[RANGE+1, RANGE+2]
})
print(test)
```

```
   hours_studied
0             21
1             22
```

```
predict = model.predict(test)
test['marks_scored']=predict
print(test)
```

```
   hours_studied  marks_scored
0             21    101.526316
1             22    102.895489
```

```
print("MSE --> ",np.mean((model.predict(df_copy[['hours_studied']])-df_copy['marks_scored'])**2))
```

```
MSE -->  3.8959022556391063
```

```
new_df= pd.concat([df,test],ignore_index=True)
print(new_df)
```

```
    hours_studied  marks_scored
0               1     71.000000
1               2     73.000000
2               3     75.000000
3               4     77.000000
4               5     79.000000
5               6     82.000000
6               7     84.000000
7               8     86.000000
8               9     88.000000
9              10     89.000000
10             11     90.000000
11             12     91.000000
12             13     92.000000
13             14     93.000000
14             15     94.000000
15             16     95.000000
16             17     95.000000
17             18     96.000000
18             19     96.000000
19             20     97.000000
20             21    101.526316
21             22    102.895489
```

```
model1 = LinearRegression()
model1.fit(new_df[['hours_studied']],new_df['marks_scored'])

m2, c2 = model1.coef_ , model1.intercept_
print("Slope:",m2[0],"\nIntercept:",c2)

x,y = new_df['hours_studied'],new_df['marks_scored']

plt.figure(figsize=(6,4))
plt.scatter(new_df['hours_studied'],new_df['marks_scored'])
plt.plot(x,m2*x+c2,color='red',linestyle='--')
plt.xlabel('hours_studied')
plt.ylabel('marks_scored')
plt.title('Hours vs Marks (New Data)')
plt.show()
```

```
Slope: 1.3691729323308273
Intercept: 72.77368421052631
```