In [1]:
```
import pandas
from pandas import DataFrame
import matplotlib.pyplot as plt

from sklearn.linear_model import LinearRegression
```

In [2]:
```
df = pandas.read_excel('run_report_3.xlsx')
df = df.drop(columns=['Unnamed: 2'])
df
```

Out[2]:

	distance	pace
0	26.30	8.41
1	26.20	9.43
2	13.10	8.16
3	13.10	9.09
4	9.67	6.48
5	9.32	7.38
6	6.40	6.55
7	6.20	7.00
8	3.11	6.11
9	3.11	6.34
10	3.08	5.59
11	1.00	5.11

In [3]:
```
df.describe()
#df['mins_per_mile'] = ['']
```

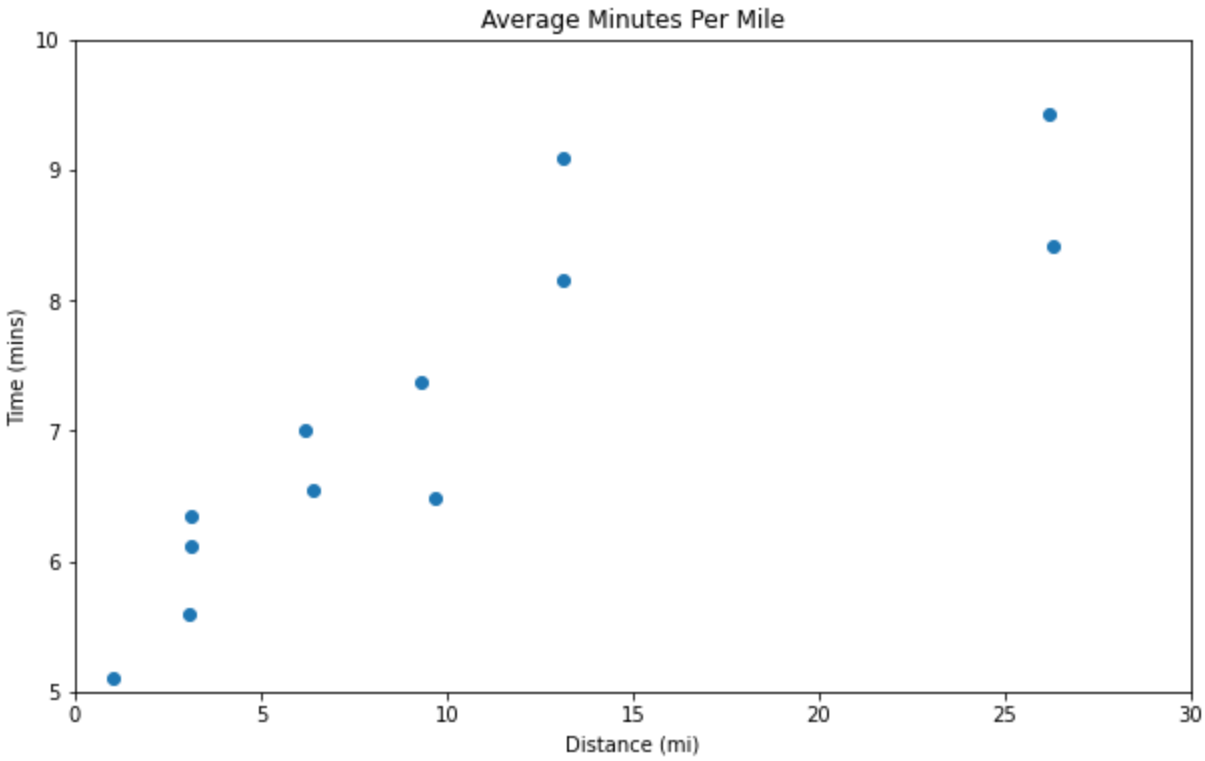
Out[3]:

	distance	pace
count	12.000000	12.000000
mean	10.049167	7.137500
std	8.523352	1.374608
min	1.000000	5.110000
25%	3.110000	6.282500
50%	7.860000	6.775000
75%	13.100000	8.222500
max	26.300000	9.430000

In [4]:
```
X = DataFrame(df, columns=['distance'])
y = DataFrame(df, columns=['pace'])
```

In [5]:
```
plt.figure(figsize=(10,6))
plt.scatter(X, y, alpha=1)
plt.title('Average Minutes Per Mile')
```

```
plt.xlabel('Distance (mi)')
plt.ylabel('Time (mins)')
plt.ylim(5, 10)
plt.xlim(0, 30)
plt.show()
```



In [6]:
```
regression = LinearRegression()
regression.fit(X, y)
```

Out[6]:
```
LinearRegression()
```

In [7]:
```
regression.coef_ # theta_1
```

Out[7]:
```
array([[0.1393062]])
```

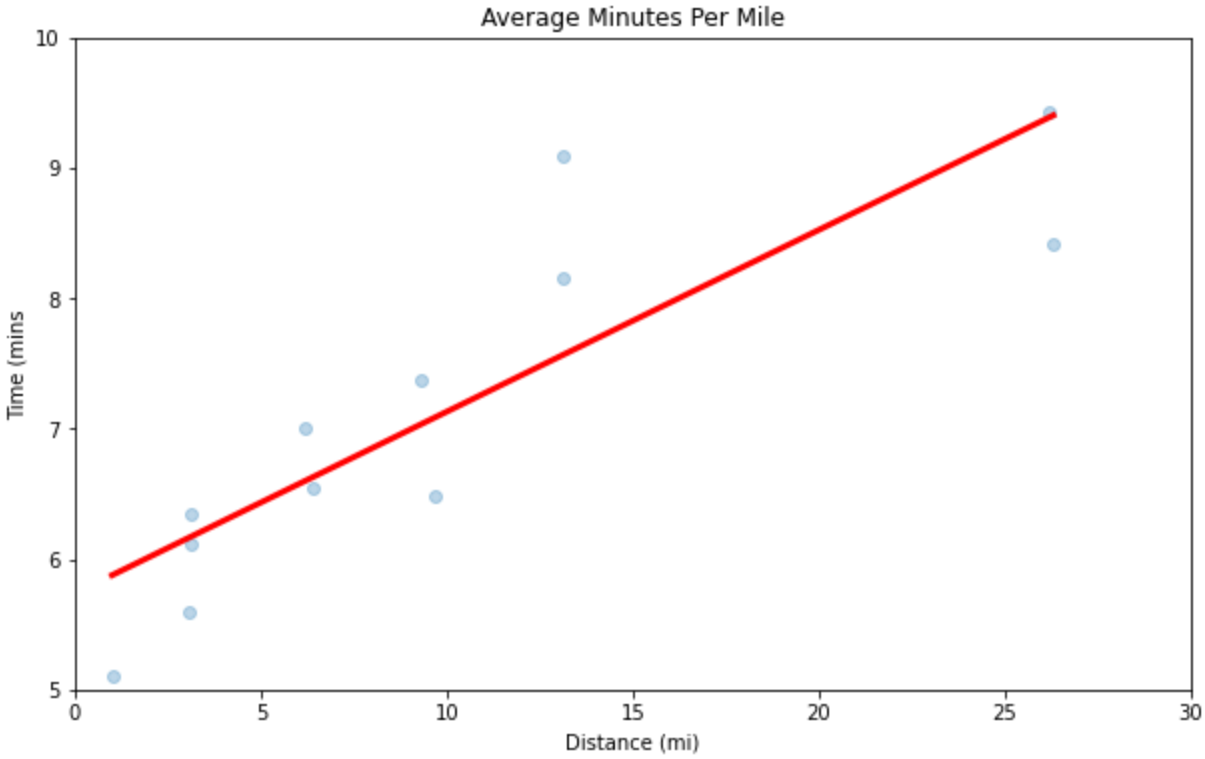
In [8]:
```
regression.intercept_
```

Out[8]:
```
array([5.73758876])
```

In [9]:
```
plt.figure(figsize=(10,6))
plt.scatter(X, y, alpha=0.3)

# Adding the regression line here:
plt.plot(X, regression.predict(X), color='red', linewidth=3)

plt.title('Average Minutes Per Mile')
plt.xlabel('Distance (mi)')
plt.ylabel('Time (mins')
plt.ylim(5, 10)
plt.xlim(0, 30)
plt.show()
```



In [10]:
```
regression.score(X, y)
```

Out[10]:
```
0.7461118274152458
```

In [ ]:

In [ ]:

In [ ]: