Buenafe, Lorenz Angelo N.
1915058
CPE 019 - CPE32S9

PART 1

Part 1: Import the Libraries and Data

```
from google.colab import drive
drive.mount("/content/drive")
```

```
Mounted at /content/drive
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
path = "/content/drive/MyDrive/CPE 019 Emerging Technologies 3/HOA 4.1/titanic_train.csv"
df = pd.read_csv(path)
```

```
df["Sex"] = df["Sex"].apply(lambda toLabel: 0 if toLabel == 'male' else 1)
df["Age"].fillna(df["Age"].mean(), inplace=True)
```

```
df.info()
```

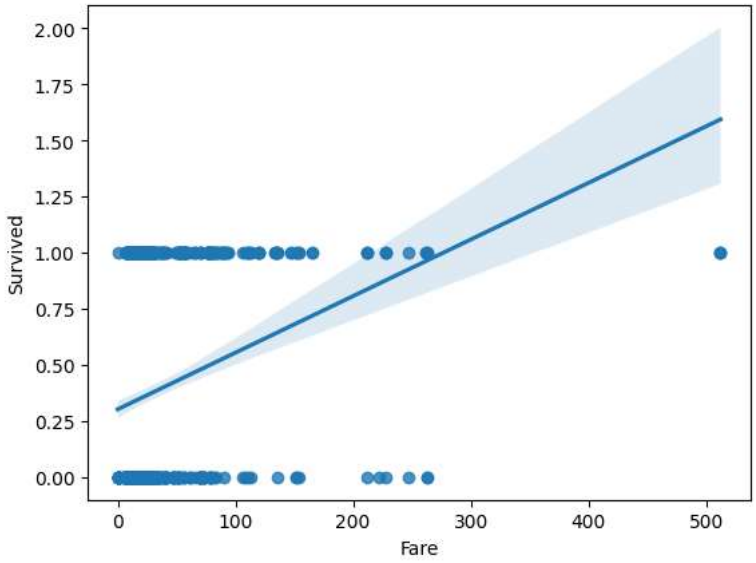
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  891 non-null    int64
 1   Survived     891 non-null    int64
 2   Pclass       891 non-null    int64
 3   Name         891 non-null    object
 4   Sex          891 non-null    int64
 5   Age          891 non-null    float64
 6   SibSp        891 non-null    int64
 7   Parch        891 non-null    int64
 8   Ticket       891 non-null    object
 9   Fare         891 non-null    float64
 10  Cabin        204 non-null    object
 11  Embarked     889 non-null    object
dtypes: float64(2), int64(6), object(4)
memory usage: 83.7+ KB
```

```
df.describe()
```

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch
count	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	0.352413	29.699118	0.523008	0.381594
std	257.353842	0.486592	0.836071	0.477990	13.002015	1.102743	0.806057
min	1.000000	0.000000	1.000000	0.000000	0.420000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	0.000000	22.000000	0.000000	0.000000
50%	446.000000	0.000000	3.000000	0.000000	29.699118	0.000000	0.000000
75%	668.500000	1.000000	3.000000	1.000000	35.000000	1.000000	0.000000
max	891.000000	1.000000	3.000000	1.000000	80.000000	8.000000	6.000000

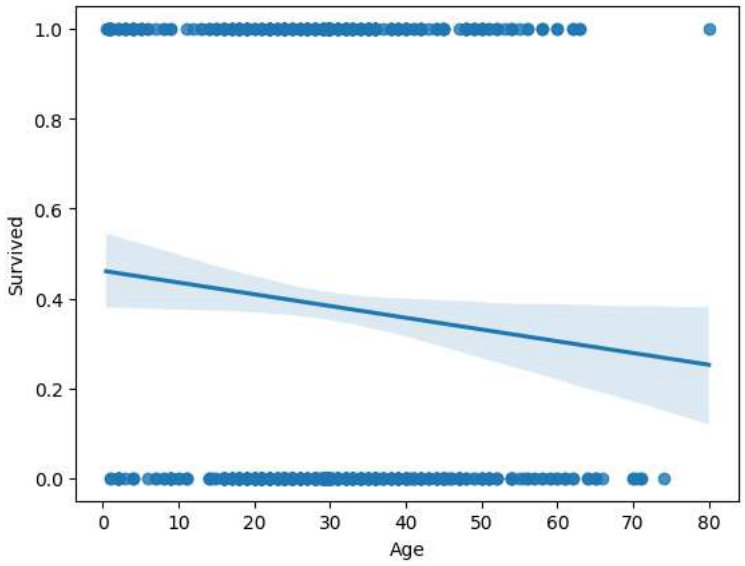
Part 2: Plot the Data

```
sns.regplot(x="Fare", y="Survived", data=df);
```



Part 3: Perform Simple Linear Regression on the SURVIVAL feature column

```
sns.regplot(x="Age", y="Survived", data=df);
```



› PART 2

[ ] ↳ *15 cells hidden*