Name :Vedant Arvind Gole Task : Week 2

In [4]:
```
import pandas as pd
```

In [5]:
```
df = pd.read_csv("titanic.csv")
```

In [6]:
```
df.head(10)
```

Out[6]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C

In [7]:
```
df.shape
```

Out[7]:
```
(891, 12)
```

In [8]:
```
df.isnull()
```

Out[8]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	False	False	False	False	False	False	False	False	False	False	True	False
1	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	True	False
3	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	True	False
...	...	...	...	...	...	...	...	...	...	...	...	...
886	False	False	False	False	False	False	False	False	False	False	True	False
887	False	False	False	False	False	False	False	False	False	False	False	False
888	False	False	False	False	False	True	False	False	False	False	True	False
889	False	False	False	False	False	False	False	False	False	False	False	False
890	False	False	False	False	False	False	False	False	False	False	True	False

891 rows × 12 columns

In [9]:
```
df.isnull().sum()
```

Out[9]:
```
PassengerId       0
Survived          0
Pclass            0
Name              0
Sex               0
Age             177
SibSp             0
Parch             0
Ticket            0
Fare              0
Cabin           687
Embarked          2
dtype: int64
```

In [10]:
```
df.rename(columns={'Pclass' : 'PassengerClass','SibSp': 'SiblingsSpouses'},inplace=True)
```

In [11]:
```
df
```

Out[11]:

	PassengerId	Survived	PassengerClass	Name	Sex	Age	SiblingsSpouses	Parch	Ticket	Fare	Cabin	Embark
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	
887	888	1	1	Graham, Miss. Margret Edith	female	19.0	0	0	112053	30.0000	B42	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	

891 rows × 12 columns

In [12]:
```
df.drop(['Cabin','Ticket'],axis=1,inplace=True)
```

In [13]:
```
df
```

Out[13]:

	PassengerId	Survived	PassengerClass	Name	Sex	Age	SiblingsSpouses	Parch	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	7.2500	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	53.1000	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	8.0500	S
...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	13.0000	S
887	888	1	1	Graham, Miss. Margaret	female	19.0	0	0	30.0000	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	23.4500	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	30.0000	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	7.7500	Q

891 rows × 10 columns

In [14]:
```
df.fillna({'Age':df['Age'].median()},inplace=True)
```

In [15]:
```
df
```

Out[15]:

	PassengerId	Survived	PassengerClass	Name	Sex	Age	SiblingsSpouses	Parch	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	7.2500	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	53.1000	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	8.0500	S
...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	13.0000	S
887	888	1	1	Graham, Miss. Margaret [illegible]	female	19.0	0	0	30.0000	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	28.0	1	2	23.4500	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	30.0000	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	7.7500	Q

891 rows × 10 columns

In [16]:
```
df.dropna(subset=['Embarked'],inplace=True)
```

In [17]:
```
df
```

Out[17]:

	PassengerId	Survived	PassengerClass	Name	Sex	Age	SiblingsSpouses	Parch	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	7.2500	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	53.1000	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	8.0500	S
...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	13.0000	S
887	888	1	1	Graham, Miss. Margaret [illegible]	female	19.0	0	0	30.0000	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	28.0	1	2	23.4500	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	30.0000	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	7.7500	Q

889 rows × 10 columns

In [18]:
```
df[["Name","Age","Fare"]].head(10)
```

Out[18]:

	Name	Age	Fare
0	Braund, Mr. Owen Harris	22.0	7.2500
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	38.0	71.2833
2	Heikkinen, Miss. Laina	26.0	7.9250
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	35.0	53.1000
4	Allen, Mr. William Henry	35.0	8.0500
5	Moran, Mr. James	28.0	8.4583
6	McCarthy, Mr. Timothy J	54.0	51.8625
7	Palsson, Master. Gosta Leonard	2.0	21.0750
8	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	27.0	11.1333
9	Nasser, Mrs. Nicholas (Adele Achem)	14.0	30.0708

In [19]:
```
df[df['Age'] > 30]
```

Out[19]:

	PassengerId	Survived	PassengerClass	Name	Sex	Age	SiblingsSpouses	Parch	Fare	Embarked
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	71.2833	C
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	53.1000	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	8.0500	S
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	51.8625	S
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	26.5500	S
...	...	...	...	...	...	...	...	...	...	...
873	874	0	3	Vander Cruyssen, Mr. Victor	male	47.0	0	0	9.0000	S
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	83.1583	C
881	882	0	3	Markun, Mr. Johann	male	33.0	0	0	7.8958	S
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	29.1250	Q
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	7.7500	Q

303 rows × 10 columns

In [20]:
```
df[df['Fare'] > 50]
```

Out[20]:

	PassengerId	Survived	PassengerClass	Name	Sex	Age	SiblingsSpouses	Parch	Fare	Embarked
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	71.2833	C
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	53.1000	S
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	51.8625	S
27	28	0	1	Fortune, Mr. Charles Alexander	male	19.0	3	2	263.0000	S
31	32	1	1	Spencer, Mrs. William Augustus (Marie Eugenie)	female	28.0	1	0	146.5208	C
...	...	...	...	...	...	...	...	...	...	...
856	857	1	1	Wick, Mrs. George Dennick (Mary Hitchcock)	female	45.0	1	1	164.8667	S
863	864	0	3	Sage, Miss. Dorothy Edith "Dolly"	female	28.0	8	2	69.5500	S
867	868	0	1	Roebling, Mr. Washington Augustus II	male	31.0	0	0	50.4958	S
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	52.5542	S
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	83.1583	C

158 rows × 10 columns

In [21]:
```
df.sort_values(by='Age', ascending=False,inplace=True)
```

In [22]:
```
df.head(10)
```

Out[22]:

	PassengerId	Survived	PassengerClass	Name	Sex	Age	SiblingsSpouses	Parch	Fare	Embarked
630	631	1	1	Barkworth, Mr. Algernon Henry Wilson	male	80.0	0	0	30.0000	S
851	852	0	3	Svensson, Mr. Johan	male	74.0	0	0	7.7750	S
493	494	0	1	Artagaveytia, Mr. Ramon	male	71.0	0	0	49.5042	C
96	97	0	1	Goldschmidt, Mr. George B	male	71.0	0	0	34.6542	C
116	117	0	3	Connors, Mr. Patrick	male	70.5	0	0	7.7500	Q
745	746	0	1	Crosby, Capt. Edward Gifford	male	70.0	1	1	71.0000	S
672	673	0	2	Mitchell, Mr. Henry Michael	male	70.0	0	0	10.5000	S
33	34	0	2	Wheadon, Mr. Edward H	male	66.0	0	0	10.5000	S
456	457	0	1	Millet, Mr. Francis Davis	male	65.0	0	0	26.5500	S
280	281	0	3	Duane, Mr. Frank	male	65.0	0	0	7.7500	Q

In [23]:
```
avg = df.groupby('PassengerClass')['Age'].mean
```

In [24]:
```
print(avg )
```

```
<bound method GroupBy.mean of <pandas.core.groupby.generic.SeriesGroupBy object at 0x000001AFD21124E0>>
```

In [25]:
```
avg_fare= df.groupby('PassengerClass')['Fare'].sum
```

In [26]:
```
avg_fare
```

Out[26]:
```
<bound method GroupBy.sum of <pandas.core.groupby.generic.SeriesGroupBy object at 0x000001AFD2113440>>
```

In [28]:
```
df.drop_duplicates(inplace=True)
```

In [29]:
```
df = df.drop_duplicates()
```

In [31]:
```
df.shape
```

Out[31]: (889, 10)

In [32]:
```
df.isnull().sum()
```

Out[32]:
```
PassengerId        0
Survived           0
PassengerClass     0
Name               0
Sex                0
Age                0
SiblingsSpouses    0
Parch              0
Fare               0
Embarked           0
dtype: int64
```

In [33]:
```
df['FamilySize'] = df['SiblingsSpouses'] + df['Parch']
```

In [35]:
```
df.head(5)
```

Out[35]:

	PassengerId	Survived	PassengerClass	Name	Sex	Age	SiblingsSpouses	Parch	Fare	Embarked	FamilySize
630	631	1	1	Barkworth, Mr. Algernon Henry Wilson	male	80.0	0	0	30.0000	S	0
851	852	0	3	Svensson, Mr. Johan	male	74.0	0	0	7.7750	S	0
493	494	0	1	Artagaveytia, Mr. Ramon	male	71.0	0	0	49.5042	C	0
96	97	0	1	Goldschmidt, Mr. George B	male	71.0	0	0	34.6542	C	0
116	117	0	3	Connors, Mr. Patrick	male	70.5	0	0	7.7500	Q	0

In [40]:
```
df.to_csv("cleaned_titanic.csv",index = False)
```

In [42]:
```
data = pd.read_csv("cleaned_titanic.csv")
```

In [43]:
```
data.head(5)
```

Out[43]:

	PassengerId	Survived	PassengerClass	Name	Sex	Age	SiblingsSpouses	Parch	Fare	Embarked	FamilySize
0	631	1	1	Barkworth, Mr. Algernon Henry Wilson	male	80.0	0	0	30.0000	S	0
1	852	0	3	Svensson, Mr. Johan	male	74.0	0	0	7.7750	S	0
2	494	0	1	Artagaveytia, Mr. Ramon	male	71.0	0	0	49.5042	C	0
3	97	0	1	Goldschmidt, Mr. George B	male	71.0	0	0	34.6542	C	0
4	117	0	3	Connors, Mr. Patrick	male	70.5	0	0	7.7500	Q	0

In [ ]: