8) Data Visualization I

1. Use the inbuilt dataset 'titanic'. The dataset contains 891 rows and contains information about the passengers who boarded the unfortunate Titanic ship. Use the Seaborn library to see if we can find any patterns in the data.
2. Write a code to check how the price of the ticket (column name: 'fare') for each passenger is distributed by plotting a histogram.

In [1]:

**import** numpy **as** np

**import** pandas **as** pd

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

In [2]:

data **=** pd**.**read\_csv("titanic.csv")

In [3]:

data**.**head()

Out[3]:

**PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin Embarked**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **0** | 1 | 0 | 3 | Braund, Mr. Owen | male | 22.0 | 1 | 0 A/5 7.2500 NaN S  21171 |
|  |  |  |  | Harris |  |  |  |  |
|  |  |  |  | Cumings, Mrs. John |  |  |  |  |
|  | **1** 2 1 1 Bradley female 38.0 1 0 PC 17599 71.2833 C85 C  (Florence | | | | | | | | |

**2** 3 1 3

Heikkinen,

Miss.

Briggs Th...

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | Laina |  | | | | | | | | | |
|  | **3** | 4 | 1 | 1 | Futrelle,  Mrs. Jacques | female | 35.0 | 1 | 0 | 113803 | | 53.1000 | C123 | S |  |
|  |  |  |  |  | Heath (Lily May Peel) |  |  |  |  |  | |  |  |  |  |
|  | **4** | 5 | 0 | 3 | Allen, Mr. William | male | 35.0 | 0 | 0 | 373450 | | 8.0500 | NaN | S |  |
|  |  |  |  |  | Henry |  |  |  |  |  | |  |  |  |  |
| In [4]: | data**.**shape |  |  |  |  |  |  |  |  |  | |  |  |  |  |
| Out[4]: | (891, 12) | |  |  | |  | |  |  | |  | |  | | |
| In [5]: | data**.**describe() | |  |  | |  | |  |  | |  | |  | | |
| Out[5]: | **PassengerId** | | **Survived** | **Pclass** | | **Age** | | **SibSp** | **Parch** | | **Fare** | |  | | |
|  | **count** 891.000000 | | 891.000000 | 891.000000 | | 714.000000 | | 891.000000 | 891.000000 | | 891.000000 | |  | | |
|  | **mean** 446.000000 | | 0.383838 | 2.308642 | | 29.699118 | | 0.523008 | 0.381594 | | 32.204208 | |  | | |
|  | **std** 257.353842 | | 0.486592 | 0.836071 | | 14.526497 | | 1.102743 | 0.806057 | | 49.693429 | |  | | |

female 26.0 0 0

STON/O2. 7.9250 NaN S 3101282

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **min** | | 1.000000 | 0.000000 | 1.000000 | | 0.420000 | 0.000000 | | 0.000000 | 0.000000 |
| **25%** | | 223.500000 | 0.000000 | 2.000000 | | 20.125000 | 0.000000 | | 0.000000 | 7.910400 |
| **50%** | | 446.000000 | 0.000000 | 3.000000 | | 28.000000 | 0.000000 | | 0.000000 | 14.454200 |
| **75%** | | 668.500000 | 1.000000 | 3.000000 | | 38.000000 | 1.000000 | | 0.000000 | 31.000000 |
| **max** | | 891.000000 | 1.000000 | 3.000000 | | 80.000000 | 8.000000 | | 6.000000 | 512.329200 |
| In [6]: data**.**describe(include **=** 'object') | | | | | | | | | | |
| Out[6]: |  | **Name** | | **Sex** | **Ticket** | **Cabin** | **Embarked** |  | | |
|  | **count** | 891 | | 891 | 891 | 204 | 889 |  | | |
|  | **unique** | 891 | | 2 | 681 | 147 | 3 |  | | |
|  | **top** | Braund, Mr. Owen Harris | | male | 347082 | B96 B98 | S |  | | |
|  | **freq** | 1 | | 577 | 7 | 4 | 644 |  | | |

In [7]:

Out[7]:

In [8]:

data['Age'] **=** data['Age']**.**fillna(np**.**mean(data['Age']))

PassengerId 0

data**.**isnull()**.**sum()

|  |  |
| --- | --- |
| Survived | 0 |
| Pclass | 0 |
| Name | 0 |
| Sex | 0 |
| Age | 177 |
| SibSp | 0 |
| Parch | 0 |
| Ticket | 0 |
| Fare | 0 |
| Cabin | 687 |
| Embarked  dtype: int64 | 2 |

In [9]:

data['Cabin'] **=** data['Cabin']**.**fillna(data['Cabin']**.**mode()[0])

In [10]:

data['Embarked'] **=** data['Embarked']**.**fillna(data['Embarked']**.**mode()[0])

In [11]:

Out[11]:

In [12]:

Out[12]:

PassengerId 0

data**.**isnull()**.**sum()

Survived 0

Pclass 0

Name 0

Sex 0

Age 0

SibSp 0

Parch 0

Ticket 0

Fare 0

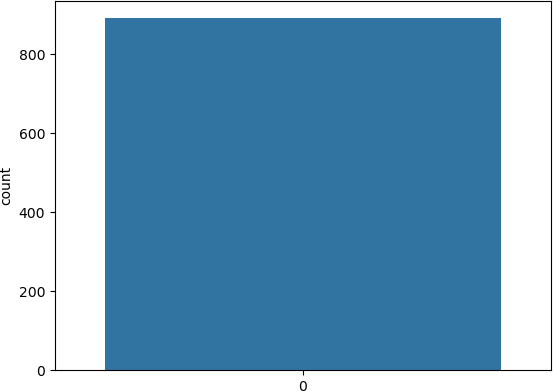
Cabin 0

Embarked 0

dtype: int64

sns**.**countplot(data['Survived'])

<Axes: ylabel='count'>



In [13]:

Out[13]:

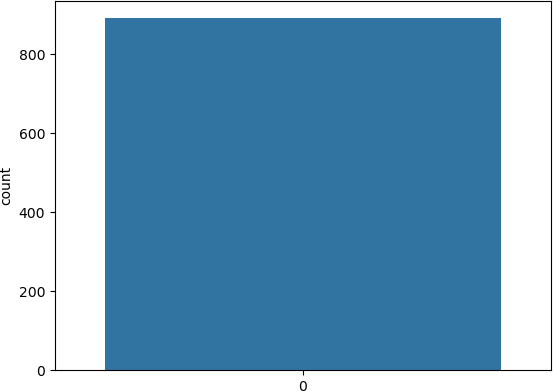
In [14]:

C:\Users\adesh\AppData\Local\Temp\ipykernel\_17200\901528773.py:1: FutureWarning: The def ault value of numeric\_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric\_only to sile

tc **=** data**.**corr() sns**.**heatmap(tc, cmap**=**"YlGnBu") plt**.**title('Correlation')

<Axes: ylabel='count'>

sns**.**countplot(data['Pclass'])



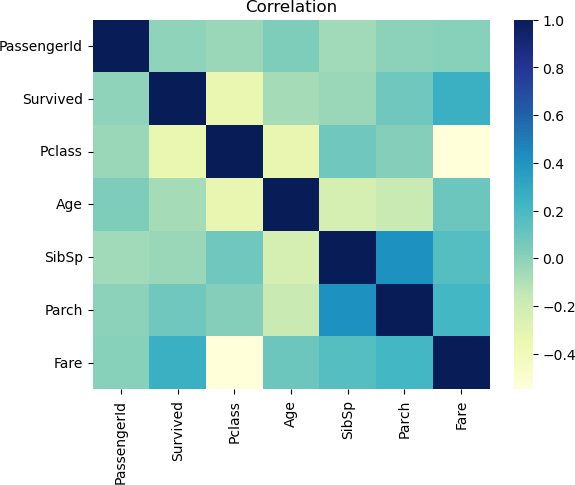
Out[14]:

In [15]:

Out[15]:

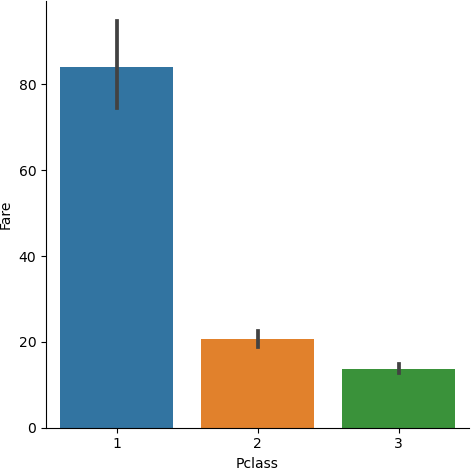
nce this warning. tc = data.corr()

Text(0.5, 1.0, 'Correlation')



sns**.**catplot(x**=**'Pclass', y**=**'Fare', data**=**data, kind**=**'bar')

<seaborn.axisgrid.FacetGrid at 0x1151fc377d0>



In [ ]: