In [1]:
```
#importing the library
import sqlite3
import numpy as np
import pandas as pd
```

In [2]:
```
#connecting the jupyter notebook with DB
conn = sqlite3.connect('titanic.db')
cursor = conn.cursor()
```

In [3]:
```
# creating the table column name  in DB

create_table_query= """
CREATE TABLE IF NOT EXISTS titanic(
Passenger_Id integer primary key,
survived INTEGER,
pclass INTEGER,
name TEXT,
sex TEXT,
age REAL,
SibSp INTEGER,
parch INTEGER,
fare REAL,
Embarked TEXT
)
"""
```

In [4]:
```
#execute and commit the table
cursor.execute(create_table_query)
conn.commit()
```

In [5]:
```
#checking the table info
cursor.execute("PRAGMA table_info(titanic)")
print(cursor.fetchall())
```

```
[(0, 'Passenger_Id', 'INTEGER', 0, None, 1), (1, 'survived', 'INTEGER', 0, None,
0), (2, 'pclass', 'INTEGER', 0, None, 0), (3, 'name', 'TEXT', 0, None, 0), (4, 'se
x', 'TEXT', 0, None, 0), (5, 'age', 'REAL', 0, None, 0), (6, 'SibSp', 'INTEGER',
0, None, 0), (7, 'parch', 'INTEGER', 0, None, 0), (8, 'fare', 'REAL', 0, None, 0),
(9, 'Embarked', 'TEXT', 0, None, 0)]
```

In [6]:
```
# read the titanic data
data= pd.read_csv('titanic.csv')
data
```

Out[6]:

| | Passenger_Id | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Fare | Embark |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 0.0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | 7.2500 | Southampt |
| **1** | 2 | 1.0 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | 71.2833 | Cherbou |
| **2** | 3 | 1.0 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | 7.9250 | Southampt |
| **3** | 4 | 1.0 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 53.1000 | Southampt |
| **4** | 5 | 0.0 | 3 | Allen, Mr. William Henry | male | 35.0 | 0 | 0 | 8.0500 | Southampt |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **1304** | 1305 | NaN | 3 | Spector, Mr. Woolf | male | NaN | 0 | 0 | 8.0500 | Southampt |
| **1305** | 1306 | NaN | 1 | Oliva y Ocana, Dona. Fermina | female | 39.0 | 0 | 0 | 108.9000 | Cherbou |
| **1306** | 1307 | NaN | 3 | Saether, Mr. Simon Sivertsen | male | 38.5 | 0 | 0 | 7.2500 | Southampt |
| **1307** | 1308 | NaN | 3 | Ware, Mr. Frederick | male | NaN | 0 | 0 | 8.0500 | Southampt |
| **1308** | 1309 | NaN | 3 | Peter, Master. Michael J | male | NaN | 1 | 1 | 22.3583 | Cherbou |

1309 rows × 10 columns

# replacing the nan value as it giving error while inserting the data and there is no concept of null in sqlite so we replace it by -9999

In [7]:

```
data.replace(np.nan, -9999, inplace=True)
data
```

Out[7]:

| | Passenger_Id | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Fare | Emba |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 0.0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | 7.2500 | Southam |
| **1** | 2 | 1.0 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | 71.2833 | Cherb |
| **2** | 3 | 1.0 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | 7.9250 | Southam |
| **3** | 4 | 1.0 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 53.1000 | Southam |
| **4** | 5 | 0.0 | 3 | Allen, Mr. William Henry | male | 35.0 | 0 | 0 | 8.0500 | Southam |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **1304** | 1305 | -9999.0 | 3 | Spector, Mr. Woolf | male | -9999.0 | 0 | 0 | 8.0500 | Southam |
| **1305** | 1306 | -9999.0 | 1 | Oliva y Ocana, Dona. Fermina | female | 39.0 | 0 | 0 | 108.9000 | Cherb |
| **1306** | 1307 | -9999.0 | 3 | Saether, Mr. Simon Sivertsen | male | 38.5 | 0 | 0 | 7.2500 | Southam |
| **1307** | 1308 | -9999.0 | 3 | Ware, Mr. Frederick | male | -9999.0 | 0 | 0 | 8.0500 | Southam |
| **1308** | 1309 | -9999.0 | 3 | Peter, Master. Michael J | male | -9999.0 | 1 | 1 | 22.3583 | Cherb |

1309 rows × 10 columns

check the output of database table

In [8]:
```
query="SELECT * FROM titanic"
df = pd.read_sql_query(query, conn)
df
```

Out[8]:

| | Passenger_Id | survived | pclass | name | sex | age | SibSp | parch | fare | Embarked |
|---|---|---|---|---|---|---|---|---|---|---|

# inserting the data and commit the changes

In [9]:
```
for index, row in data.iterrows():
    insert_data_query = f"""
    INSERT INTO titanic (Passenger_Id, survived, pclass, name, sex, age, SibSp, par
    ({row['Passenger_Id']}, {row['Survived']}, {row['Pclass']}, '{row['Name']}', '
    """
    cursor.execute(insert_data_query)
```

In [10]:
```
#cursor.execute("delete from titanic")

# Commit the changes
conn.commit()
```

In [11]:
```
#checking the complete output
query="SELECT * FROM titanic"
df = pd.read_sql_query(query, conn)
df
```

Out[11]:

| | Passenger_Id | survived | pclass | name | sex | age | SibSp | parch | fare | Emba |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | 7.2500 | Southam |
| **1** | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | 71.2833 | Cherb |
| **2** | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | 7.9250 | Southam |
| **3** | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 53.1000 | Southam |
| **4** | 5 | 0 | 3 | Allen, Mr. William Henry | male | 35.0 | 0 | 0 | 8.0500 | Southam |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **1304** | 1305 | -9999 | 3 | Spector, Mr. Woolf | male | -9999.0 | 0 | 0 | 8.0500 | Southam |
| **1305** | 1306 | -9999 | 1 | Oliva y Ocana, Dona. Fermina | female | 39.0 | 0 | 0 | 108.9000 | Cherb |
| **1306** | 1307 | -9999 | 3 | Saether, Mr. Simon Sivertsen | male | 38.5 | 0 | 0 | 7.2500 | Southam |
| **1307** | 1308 | -9999 | 3 | Ware, Mr. Frederick | male | -9999.0 | 0 | 0 | 8.0500 | Southam |
| **1308** | 1309 | -9999 | 3 | Peter, Master. Michael J | male | -9999.0 | 1 | 1 | 22.3583 | Cherb |

1309 rows × 10 columns

In [12]:

```
# total no of people embarked from three different location. most of them was from
query="SELECT Embarked,count(*) FROM titanic group by Embarked"
df = pd.read_sql_query(query, conn)
df
```

Out[12]:

| | Embarked | count(*) |
|---|---|---|
| **0** | -9999 | 2 |
| **1** | Cherbourg | 270 |
| **2** | Queenstown | 123 |
| **3** | Southampton | 914 |

In [13]:
```
# maximum age of person travelling  in the voyage
query="SELECT * FROM titanic order by age desc limit 1"
df = pd.read_sql_query(query, conn)
df
```

Out[13]:

| | Passenger_Id | survived | pclass | name | sex | age | SibSp | parch | fare | Embarked |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 631 | 1 | 1 | Barkworth, Mr. Algernon Henry Wilson | male | 80.0 | 0 | 0 | 30.0 | Southampton |

In [14]:
```
#total people travelling in titanic was 1309 out which 843 are male and 466 are fe
query="SELECT sex,count(*) FROM titanic group by sex"
df = pd.read_sql_query(query, conn)
df
```

Out[14]:

| | sex | count(*) |
|---|---|---|
| **0** | female | 466 |
| **1** | male | 843 |

In [15]:
```
#out of 466 female, 233 were survived, 81 not survived  and there no data about 15
#out of 843 male, 109 only survived, 468 not survived  while there no data about 2
query="SELECT sex,survived, count(*) as Is_survived FROM titanic group by sex,survi
df = pd.read_sql_query(query, conn)
df
```

Out[15]:

| | sex | survived | Is_survived |
|---|---|---|---|
| **0** | female | -9999 | 152 |
| **1** | female | 0 | 81 |
| **2** | female | 1 | 233 |
| **3** | male | -9999 | 266 |
| **4** | male | 0 | 468 |
| **5** | male | 1 | 109 |

In [16]:
```
# Anna Ward share the highest cost of fare.
query="SELECT  NAME,SEX,AGE,EMBARKED,MAX(fare) as maxium_fair_paid FROM titanic"
df = pd.read_sql_query(query, conn)
df
```

Out[16]:

| | name | sex | age | Embarked | maxium_fair_paid |
|---|---|---|---|---|---|
| **0** | Ward, Miss. Anna | female | 35.0 | Cherbourg | 512.3292 |

In [17]:
```
# maximum number people travelling in the trip having the age between 20 to 60 yea
query="SELECT COUNT(*) AS NO_OF_PEOPLE, round(AGE/10,0)*10 AS AGE_GROUP FROM titan
df = pd.read_sql_query(query, conn)
df
```

Out[17]:

| | NO_OF_PEOPLE | AGE_GROUP |
|---|---|---|
| **0** | 2 | 80.0 |
| **1** | 11 | 70.0 |
| **2** | 54 | 60.0 |
| **3** | 109 | 50.0 |
| **4** | 169 | 40.0 |
| **5** | 292 | 30.0 |
| **6** | 300 | 20.0 |
| **7** | 58 | 10.0 |
| **8** | 51 | 0.0 |
| **9** | 263 | -10000.0 |

In [18]:
```
#details of all traveller between age of 20 and 60
query="SELECT * FROM titanic where age between 20 and 60"
df = pd.read_sql_query(query, conn)
df
```

Out[18]:

| | Passenger_Id | survived | pclass | name | sex | age | SibSp | parch | fare | Embarke |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | 7.2500 | Southampto |
| **1** | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | 71.2833 | Cherbour |
| **2** | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | 7.9250 | Southampto |
| **3** | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 53.1000 | Southampto |
| **4** | 5 | 0 | 3 | Allen, Mr. William Henry | male | 35.0 | 0 | 0 | 8.0500 | Southampto |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **783** | 1299 | -9999 | 1 | Widener, Mr. George Dunton | male | 50.0 | 1 | 1 | 211.5000 | Cherbour |
| **784** | 1303 | -9999 | 1 | Minahan, Mrs. William Edward (Lillian E Thorpe) | female | 37.0 | 1 | 0 | 90.0000 | Queenstow |
| **785** | 1304 | -9999 | 3 | Henriksson, Miss. Jenny Lovisa | female | 28.0 | 0 | 0 | 7.7750 | Southampto |
| **786** | 1306 | -9999 | 1 | Oliva y Ocana, Dona. Fermina | female | 39.0 | 0 | 0 | 108.9000 | Cherbour |
| **787** | 1307 | -9999 | 3 | Saether, Mr. Simon Sivertsen | male | 38.5 | 0 | 0 | 7.2500 | Southampto |

788 rows × 10 columns

In [19]:

```
query="SELECT min(age) as min_age,max(age) as max_age, avg(age) as avg_age, count(*
df = pd.read_sql_query(query, conn)
df
```

Out[19]:

| | min_age | max_age | avg_age | total_count | Embarked |
|---|---|---|---|---|---|
| **0** | 38.0 | 62.0 | 50.000000 | 2 | -9999 |
| **1** | -9999.0 | 71.0 | -2122.546593 | 270 | Cherbourg |
| **2** | -9999.0 | 70.5 | -5922.727642 | 123 | Queenstown |
| **3** | -9999.0 | 80.0 | -1419.035284 | 914 | Southampton |

In [20]:
```
#showing no of traveller as per passenger class
query="SELECT pclass,count(*) as people FROM titanic group by pclass"
df = pd.read_sql_query(query, conn)
df
```

Out[20]:

| | pclass | people |
|---|---|---|
| **0** | 1 | 323 |
| **1** | 2 | 277 |
| **2** | 3 | 709 |

In [21]:
```
#maximum no of people survive from passenger class 1
query="SELECT pclass, survived,count(*) as people_survived FROM titanic group by su
df = pd.read_sql_query(query, conn)
df
```

Out[21]:

| | pclass | survived | people_survived |
|---|---|---|---|
| **0** | 1 | -9999 | 107 |
| **1** | 2 | -9999 | 93 |
| **2** | 3 | -9999 | 218 |
| **3** | 1 | 0 | 80 |
| **4** | 2 | 0 | 97 |
| **5** | 3 | 0 | 372 |
| **6** | 1 | 1 | 136 |
| **7** | 2 | 1 | 87 |
| **8** | 3 | 1 | 119 |

In [22]:
```
# total fare paid category wise, for class 3 showing nr=egative due to nan values r
query="Select pclass,sum(FARE) as total_amount_paid_class FROM titanic group by pcl
df = pd.read_sql_query(query, conn)
df
```

Out[22]:

| | pclass | total_amount_paid_class |
|---|---|---|
| **0** | 1 | 28265.4043 |
| **1** | 2 | 5866.6374 |
| **2** | 3 | -580.5548 |

In [23]:
```
#list of people who paid fare more than 50
query="Select *  FROM titanic where fare>50"
```

```
df = pd.read_sql_query(query, conn)
df
```

Out[23]:

| | Passenger_Id | survived | pclass | name | sex | age | SibSp | parch | fare | Embarl |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | 71.2833 | Cherbo |
| **1** | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 53.1000 | Southamp |
| **2** | 7 | 0 | 1 | McCarthy, Mr. Timothy J | male | 54.0 | 0 | 0 | 51.8625 | Southamp |
| **3** | 28 | 0 | 1 | Fortune, Mr. Charles Alexander | male | 19.0 | 3 | 2 | 263.0000 | Southamp |
| **4** | 32 | 1 | 1 | Spencer, Mrs. William Augustus (Marie Eugenie) | female | -9999.0 | 1 | 0 | 146.5208 | Cherbo |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **235** | 1292 | -9999 | 1 | Bonnell, Miss. Caroline | female | 30.0 | 0 | 0 | 164.8667 | Southamp |
| **236** | 1294 | -9999 | 1 | Gibson, Miss. Dorothy Winifred | female | 22.0 | 0 | 1 | 59.4000 | Cherbo |
| **237** | 1299 | -9999 | 1 | Widener, Mr. George Dunton | male | 50.0 | 1 | 1 | 211.5000 | Cherbo |
| **238** | 1303 | -9999 | 1 | Minahan, Mrs. William Edward (Lillian E Thorpe) | female | 37.0 | 1 | 0 | 90.0000 | Queensto |
| **239** | 1306 | -9999 | 1 | Oliva y Ocana, Dona. Fermina | female | 39.0 | 0 | 0 | 108.9000 | Cherbo |

240 rows × 10 columns

In [24]:
```
#count of female and male pay fare more than 50
query="Select sex,count(sex)  FROM titanic where fare >50.0 group by sex"
df = pd.read_sql_query(query, conn)
df
```

Out[24]:

| | sex | count(sex) |
|---|---|---|
| **0** | female | 130 |
| **1** | male | 110 |

In [25]:
```
#count of male and female travelling and there age greater than 20
query="Select sex,count(sex)  FROM titanic where age >20 group by sex"
df = pd.read_sql_query(query, conn)
df
```

Out[25]:

| | sex | count(sex) |
|---|---|---|
| **0** | female | 279 |
| **1** | male | 519 |

In [ ]:

In [ ]: