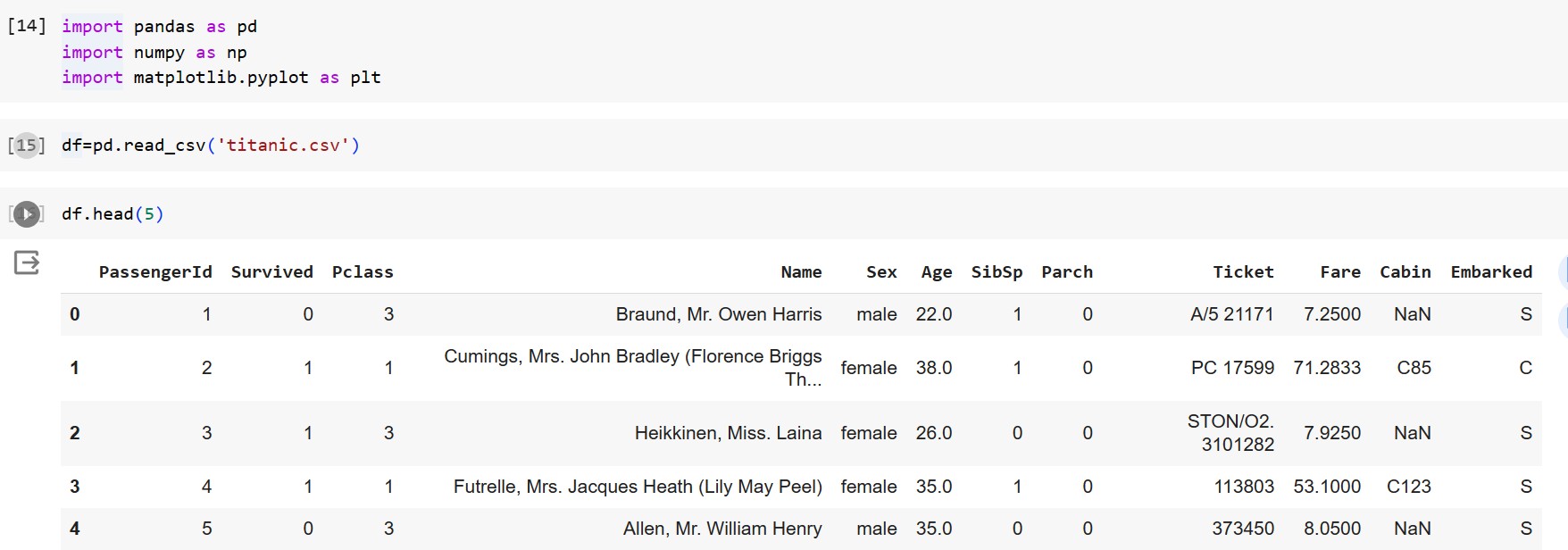
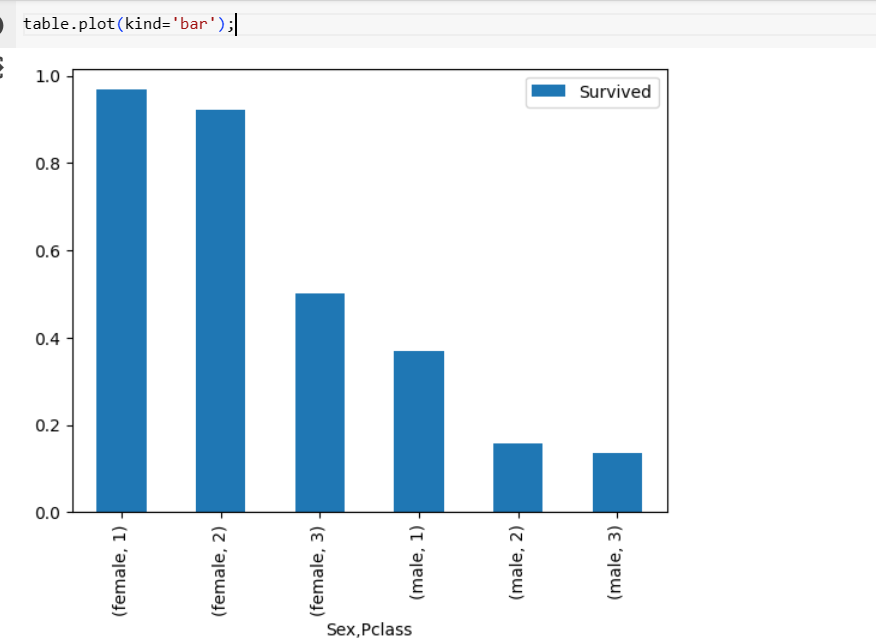
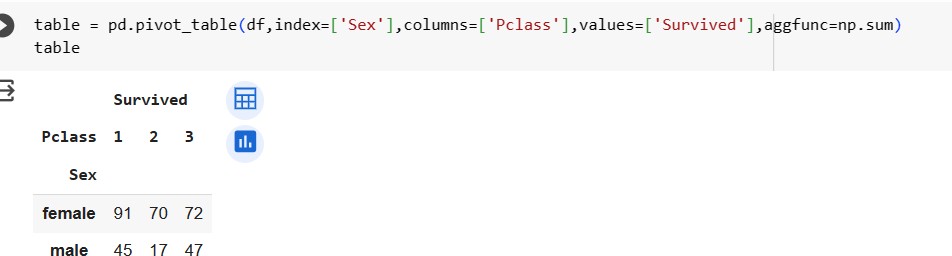
1. Create flexible data aggregations using pivot tables and Represent data visually using pivot charts

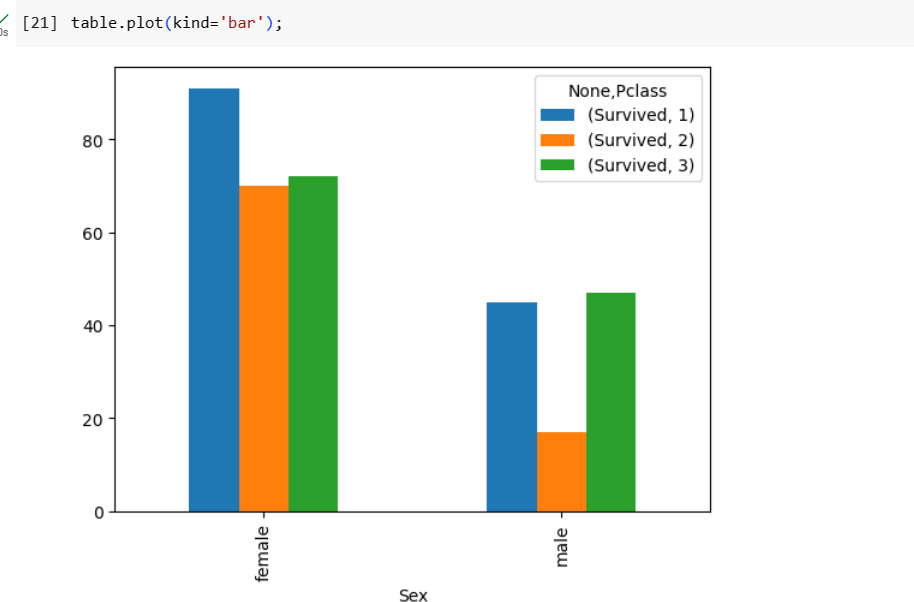


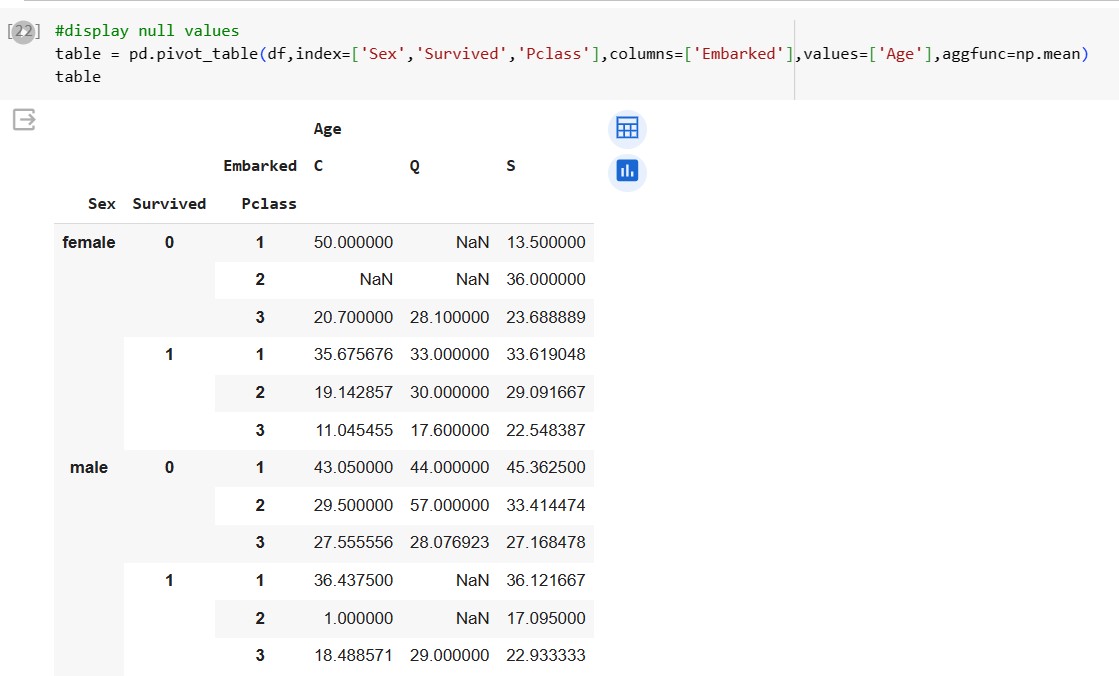


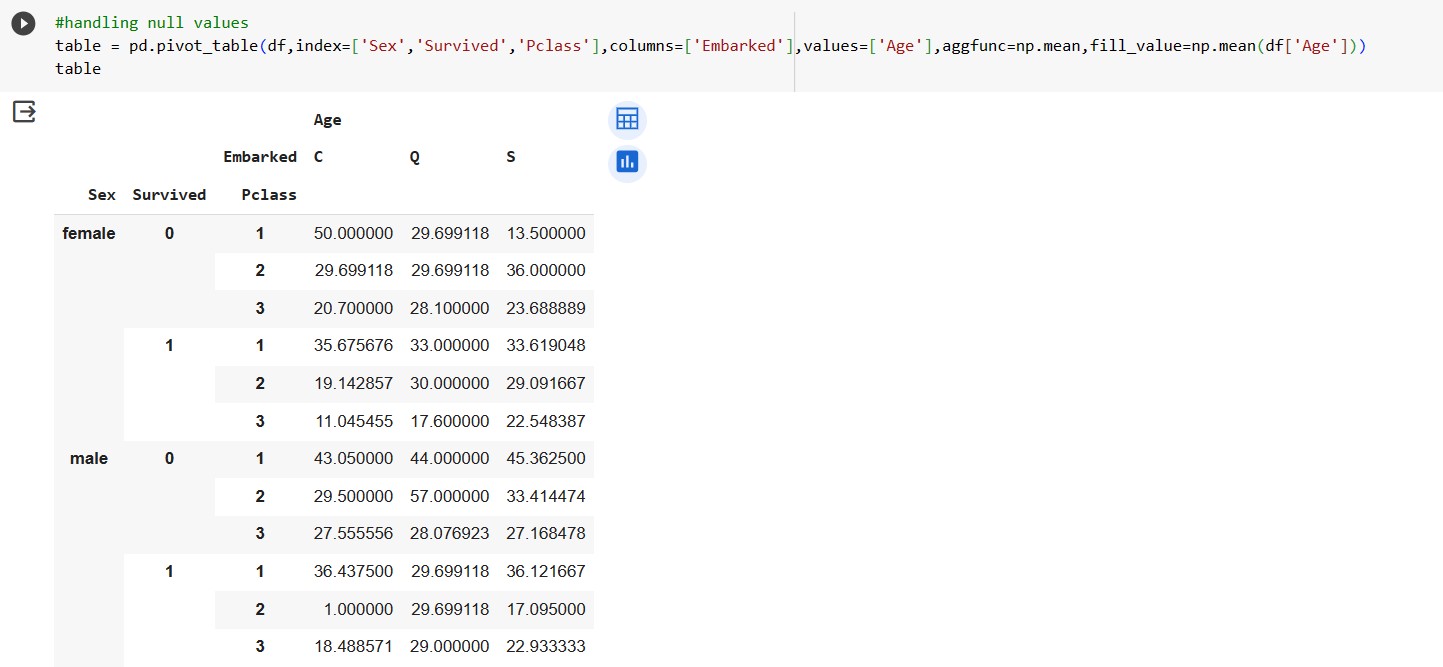












import pandas as pd import numpy as np

df = pd.read\_csv("tem.csv") df

|  |  |  |
| --- | --- | --- |
|  | **city** | **temperature** |
| **0** | Mumbai | 34 |
| **1** | Chennai | 38 |
| **2** | Hyderabad | 43 |
| **3** | Banagalore | 30 |
| **4** | Pune | -4 |
| **5** | Kochi | 33 |
| **6** | Goa | 50 |
| df.shape |  |  |

(7, 2)

df.dtypes

city object

temperature int64 dtype: object

df.head()

### city temperature

* 1. Mumbai 34
  2. Chennai 38
  3. Hyderabad 43
  4. Banagalore 30
  5. Pune -4

df.tail(3)

### city temperature

1. Pune -4
2. Kochi 33
3. Goa 50

df.isnull()

### city temperature

1. False False
2. False False
3. False False
4. False False
5. False False
6. False False
7. False False

df.isnull().sum()

city 0

temperature 0

dtype: int64

df.count()

city 7

temperature 7

dtype: int64

df.info()

<class 'pandas.core.frame.DataFrame'> RangeIndex: 7 entries, 0 to 6

Data columns (total 2 columns):

# Column Non-Null Count Dtype

1. city 7 non-null object
2. temperature 7 non-null int64 dtypes: int64(1), object(1)

memory usage: 240.0+ bytes

gk = df.groupby('city') gk=gk.get\_group('Mumbai') gk

### city temperature

**0** Mumbai 34

## Downlod Following CSV files and do all operations iris.csv

1. titanic.csv

## car.csv

1. Iris.csv Solve the following 1)Download csv from google 2)upload in juypter notebook 3)load/read csv file 4)display count of rows and

## Columns 5) Display data type of each column 6)Displat first 3 record 7)Display last 3 record 8) Display count of null values 9)Display info of file 10 Diplay the data of one category

import pandas as pd

df = pd.read\_csv("titanic.csv") df

### PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket Fa

**0** 1 0 3

**1** 2 1 1

Braund, Mr. Owen

Harris

Cumings, Mrs. John Bradley (Florence Briggs Th...

Heikkinen,

male 22.0 1 0 A/5 21171 7.25

female 38.0 1 0 PC 17599 71.28

STON/O2.

Miss. Laina

|  |  |  |  |
| --- | --- | --- | --- |
| **2** | 3 | 1 | 3 |
| **3** | 4 | 1 | 1 |

female 26.0 0 0

3101282 7.92

Futrelle,

Mrs. Jacques Heath (Lily May

Peel)

female 35.0 1 0 113803 53.10

df.shape

(891, 12)

df.dtypes

PassengerId int64

Survived int64

Pclass int64

Name object

Sex object

Age float64

SibSp int64

Parch int64

Ticket object

Fare float64

Cabin object

Embarked object dtype: object

df.head(3)

### PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket Fare

**0** 1 0 3

Braund, Mr. Owen

Harris

male 22.0 1 0 A/5 21171 7.2500

df.tail(3)

### PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket Fare C

**888** 889 0 3

Johnston,

Miss. Catherine

Helen

female NaN 1 2

W./C. 23.45

6607

df.isnull().sum()

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

df.count()

|  |  |
| --- | --- |
| PassengerId | 891 |
| Survived | 891 |
| Pclass | 891 |
| Name | 891 |
| Sex | 891 |
| Age | 714 |
| SibSp | 891 |
| Parch | 891 |
| Ticket | 891 |
| Fare | 891 |
| Cabin | 204 |
| Embarked | 889 |
| dtype: int64 |  |

df.info()

<class 'pandas.core.frame.DataFrame'> RangeIndex: 891 entries, 0 to 890

Data columns (total 12 columns):

# Column Non-Null Count Dtype

1. PassengerId 891 non-null int64
2. Survived 891 non-null int64
3. Pclass 891 non-null int64

|  |  |  |  |
| --- | --- | --- | --- |
| 3 | Name | 891 non-null | object |
| 4 | Sex | 891 non-null | object |
| 5 | Age | 714 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(5), object(5) memory usage: 83.7+ KB

gk = df.groupby('Pclass') gk=gk.get\_group(3)

gk

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **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 |
| **2** | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.9250 | NaN | 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 |
| **7** | 8 | 0 | 3 | Palsson, Master. Gosta Leonard | male | 2.0 | 3 | 1 | 349909 | 21.0750 | NaN | S |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **882** | 883 | 0 | 3 | Dahlberg, Miss. Gerda Ulrika | female | 22.0 | 0 | 0 | 7552 | 10.5167 | NaN | S |

Sutehall, Mr. Henry Jr male 25.0 0 0 SOTON/OQ

|  |  |  |  |
| --- | --- | --- | --- |
| **884** | 885 | 0 | 3 |
| **885** | 886 | 0 | 3 |
| **888** | 889 | 0 | 3 |

392076

7.0500 NaN S

Rice, Mrs. William (Margaret female 39.0 0 5 382652 29.1250 NaN Q Norton)

Johnston, Miss. Catherine Helen female NaN 1 2 W./C. 6607 23.4500 NaN S "Carrie"

df = pd.read\_csv("Iris.csv") df

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Id** | **SepalLengthCm** | **SepalWidthCm** | **PetalLengthCm** | **PetalWidthCm** | **Species** |
|  | **0** 1 | 5.1 | 3.5 | 1.4 | 0.2 | Iris-setosa |
|  | **1** 2 | 4.9 | 3.0 | 1.4 | 0.2 | Iris-setosa |
|  | **2** 3 | 4.7 | 3.2 | 1.3 | 0.2 | Iris-setosa |
|  | **3** 4 | 4.6 | 3.1 | 1.5 | 0.2 | Iris-setosa |
|  | **4** 5 | 5.0 | 3.6 | 1.4 | 0.2 | Iris-setosa |
|  | **...** ... | ... | ... | ... | ... | ... |
|  | **145** 146 | 6.7 | 3.0 | 5.2 | 2.3 | Iris-virginica |
|  | **146** 147 | 6.3 | 2.5 | 5.0 | 1.9 | Iris-virginica |
|  | **147** 148 | 6.5 | 3.0 | 5.2 | 2.0 | Iris-virginica |
|  | **148** 149 | 6.2 | 3.4 | 5.4 | 2.3 | Iris-virginica |
|  | **149** 150 | 5.9 | 3.0 | 5.1 | 1.8 | Iris-virginica |

150 rows × 6 columns

df.shape

(150, 6)

df.dtypes

Id int64

SepalLengthCm float64 SepalWidthCm float64 PetalLengthCm float64 PetalWidthCm float64

Species object dtype: object

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| df.head(3)  **Id** | **SepalLengthCm** | **SepalWidthCm** | **PetalLengthCm** | **PetalWidthCm** | **Species** |
| **0** 1 | 5.1 | 3.5 | 1.4 | 0.2 | Iris-setosa |
| **1** 2 | 4.9 | 3.0 | 1.4 | 0.2 | Iris-setosa |
| **2** 3 | 4.7 | 3.2 | 1.3 | 0.2 | Iris-setosa |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| df.tail(3) | **Id** | **SepalLengthCm** | **SepalWidthCm** | **PetalLengthCm** | **PetalWidthCm** | **Species** |
| **147** | 148 | 6.5 | 3.0 | 5.2 | 2.0 | Iris-virginica |
| **148** | 149 | 6.2 | 3.4 | 5.4 | 2.3 | Iris-virginica |
| **149** | 150 | 5.9 | 3.0 | 5.1 | 1.8 | Iris-virginica |

df.isnull().sum()

Id 0

SepalLengthCm 0

SepalWidthCm 0

PetalLengthCm 0

PetalWidthCm 0

Species 0

dtype: int64

df.count()

|  |  |
| --- | --- |
| Id | 150 |
| SepalLengthCm | 150 |
| SepalWidthCm | 150 |
| PetalLengthCm | 150 |
| PetalWidthCm | 150 |
| Species  dtype: int64 | 150 |

df.info()

<class 'pandas.core.frame.DataFrame'> RangeIndex: 150 entries, 0 to 149

Data columns (total 6 columns):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # |  | Column | Non-Null Count |  | Dtype |
| 0 |  | Id | 150 non-null |  | int64 |
| 1 |  | SepalLengthCm | 150 non-null |  | float64 |
| 2 |  | SepalWidthCm | 150 non-null |  | float64 |
| 3 |  | PetalLengthCm | 150 non-null |  | float64 |
| 4 |  | PetalWidthCm | 150 non-null |  | float64 |
| 5 |  | Species | 150 non-null |  | object |

dtypes: float64(4), int64(1), object(1) memory usage: 7.2+ KB

gk = df.groupby('Species')

gk=gk.get\_group('Iris-setosa') gk

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Id** | **SepalLengthCm** | **SepalWidthCm** | **PetalLengthCm** | **PetalWidthCm** | **Species** |
| **0** 1 | 5.1 | 3.5 | 1.4 | 0.2 | Iris-setosa |
| **1** 2 | 4.9 | 3.0 | 1.4 | 0.2 | Iris-setosa |
| **2** 3 | 4.7 | 3.2 | 1.3 | 0.2 | Iris-setosa |
| **3** 4 | 4.6 | 3.1 | 1.5 | 0.2 | Iris-setosa |
| **4** 5 | 5.0 | 3.6 | 1.4 | 0.2 | Iris-setosa |
| **5** 6 | 5.4 | 3.9 | 1.7 | 0.4 | Iris-setosa |
| **6** 7 | 4.6 | 3.4 | 1.4 | 0.3 | Iris-setosa |
| **7** 8 | 5.0 | 3.4 | 1.5 | 0.2 | Iris-setosa |
| **8** 9 | 4.4 | 2.9 | 1.4 | 0.2 | Iris-setosa |
| **9** 10 | 4.9 | 3.1 | 1.5 | 0.1 | Iris-setosa |
| **10** 11 | 5.4 | 3.7 | 1.5 | 0.2 | Iris-setosa |
| **11** 12 | 4.8 | 3.4 | 1.6 | 0.2 | Iris-setosa |
| **12** 13 | 4.8 | 3.0 | 1.4 | 0.1 | Iris-setosa |
| **13** 14 | 4.3 | 3.0 | 1.1 | 0.1 | Iris-setosa |
| **14** 15 | 5.8 | 4.0 | 1.2 | 0.2 | Iris-setosa |
| **15** 16 | 5.7 | 4.4 | 1.5 | 0.4 | Iris-setosa |
| **16** 17 | 5.4 | 3.9 | 1.3 | 0.4 | Iris-setosa |
| **17** 18 | 5.1 | 3.5 | 1.4 | 0.3 | Iris-setosa |
| **18** 19 | 5.7 | 3.8 | 1.7 | 0.3 | Iris-setosa |
| **19** 20 | 5.1 | 3.8 | 1.5 | 0.3 | Iris-setosa |
| **20** 21 | 5.4 | 3.4 | 1.7 | 0.2 | Iris-setosa |
| **21** 22 | 5.1 | 3.7 | 1.5 | 0.4 | Iris-setosa |
| **22** 23 | 4.6 | 3.6 | 1.0 | 0.2 | Iris-setosa |
| **23** 24 | 5.1 | 3.3 | 1.7 | 0.5 | Iris-setosa |
| **24** 25 | 4.8 | 3.4 | 1.9 | 0.2 | Iris-setosa |
| **25** 26 | 5.0 | 3.0 | 1.6 | 0.2 | Iris-setosa |
| **26** 27 | 5.0 | 3.4 | 1.6 | 0.4 | Iris-setosa |
| **27** 28 | 5.2 | 3.5 | 1.5 | 0.2 | Iris-setosa |
| **28** 29 | 5.2 | 3.4 | 1.4 | 0.2 | Iris-setosa |
| **29** 30 | 4.7 | 3.2 | 1.6 | 0.2 | Iris-setosa |
| **30** 31 | 4.8 | 3.1 | 1.6 | 0.2 | Iris-setosa |
| **31** 32 | 5.4 | 3.4 | 1.5 | 0.4 | Iris-setosa |
| **32** 33 | 5.2 | 4.1 | 1.5 | 0.1 | Iris-setosa |
| **33** 34 | 5.5 | 4.2 | 1.4 | 0.2 | Iris-setosa |
| **34** 35 | 4.9 | 3.1 | 1.5 | 0.1 | Iris-setosa |
| **35** 36 | 5.0 | 3.2 | 1.2 | 0.2 | Iris-setosa |
| **36** 37 | 5.5 | 3.5 | 1.3 | 0.2 | Iris-setosa |
| **37** 38 | 4.9 | 3.1 | 1.5 | 0.1 | Iris-setosa |
| **38** 39 | 4.4 | 3.0 | 1.3 | 0.2 | Iris-setosa |
| **39** 40 | 5.1 | 3.4 | 1.5 | 0.2 | Iris-setosa |
| **40** 41 | 5.0 | 3.5 | 1.3 | 0.3 | Iris-setosa |
| **41** 42 | 4.5 | 2.3 | 1.3 | 0.3 | Iris-setosa |
| **42** 43 | 4.4 | 3.2 | 1.3 | 0.2 | Iris-setosa |
| **43** 44 | 5.0 | 3.5 | 1.6 | 0.6 | Iris-setosa |
| **44** 45 | 5.1 | 3.8 | 1.9 | 0.4 | Iris-setosa |
| **45** 46 | 4.8 | 3.0 | 1.4 | 0.3 | Iris-setosa |

import pandas as pd import numpy as np

df = pd.read\_csv("Book1.csv") df

|  |  |  |
| --- | --- | --- |
| **city** | **temperature** | **humidity** |
| **0** new york | 65 | 56 |
| **1** new york | 65 | 66 |
| **2** new york | 66 | 60 |
| **3** mumbai | 75 | 80 |
| **4** mumbai | 68 | 80 |

import statistics

statistics.stdev(df['humidity'])

11.171392035015153

## Finding Frequency

count = df['city'].value\_counts() print(count)

new york 3

mumbai 2

Name: city, dtype: int64

count = df.groupby(['city']).count() print(count)

temperature humidity

city

mumbai 2 2

new york 3 3

## Double-click (or enter) to edit

df.mean()

<ipython-input-32-c61f0c8f89b5>:1: FutureWarning: The default value of numeric\_only in DataFrame.mean is deprecated. In a future version df.mean()

temperature 67.8

humidity 68.4

dtype: float64

df.median()

<ipython-input-33-6d467abf240d>:1: FutureWarning: The default value of numeric\_only in DataFrame.median is deprecated. In a future versi df.median()

temperature 66.0

humidity 66.0

dtype: float64

df.mode(numeric\_only=True)

### temperature humidity

**0** 65 80

df.describe()

|  |  |  |
| --- | --- | --- |
|  | **temperature** | **humidity** |
| **count** | 5.000000 | 5.000000 |
| **mean** | 67.800000 | 68.400000 |
| **std** | 4.207137 | 11.171392 |
| **min** | 65.000000 | 56.000000 |
| **25%** | 65.000000 | 60.000000 |
| **50%** | 66.000000 | 66.000000 |
| **75%** | 68.000000 | 80.000000 |
| **max** | 75.000000 | 80.000000 |

temperature\_variance = df['temperature'].var() print(temperature\_variance)

17.7

## Double-click (or enter) to edit

humidity\_variance = df['humidity'].var() print(humidity\_variance)

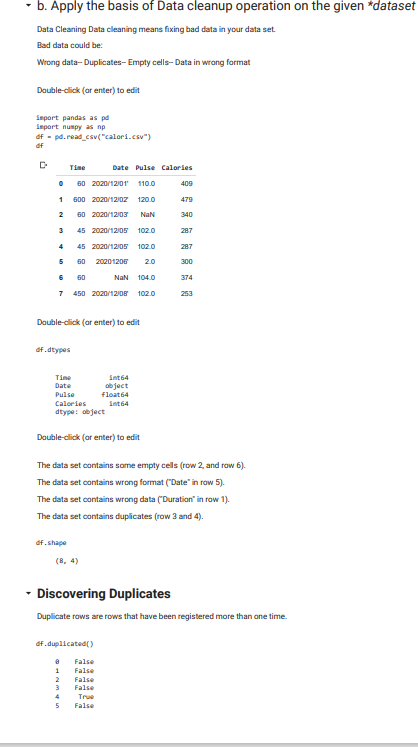
124.79999999999998

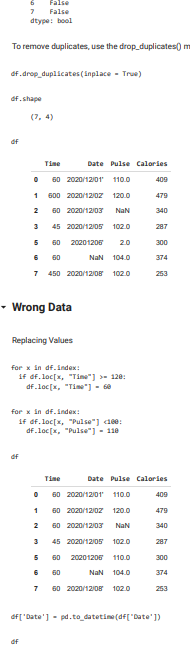
temperature\_stddev = df['temperature'].std() print(temperature\_stddev)

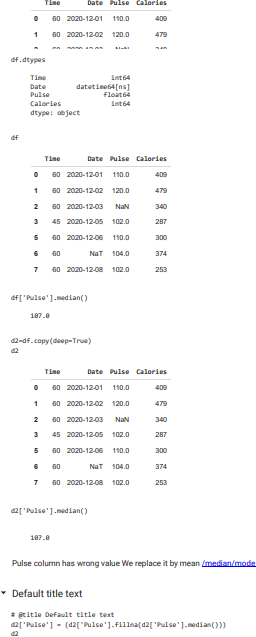
4.207136793592526

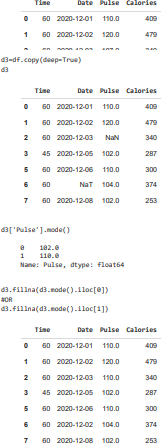
humidity\_stddev = df['humidity'].std() print(humidity\_stddev)

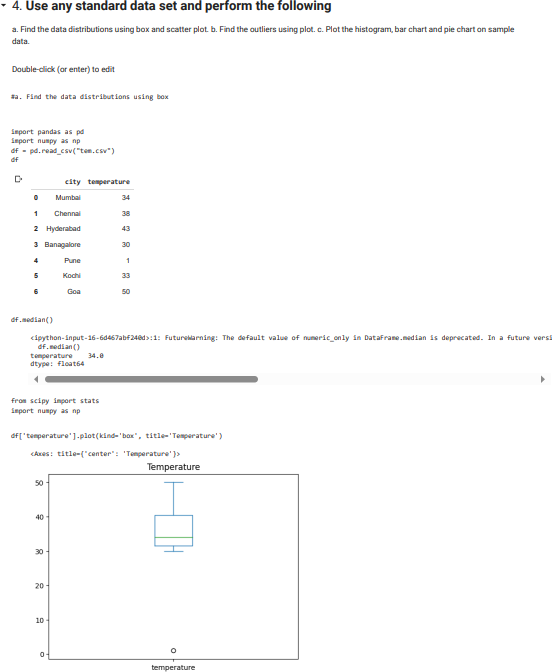
11.171392035015153

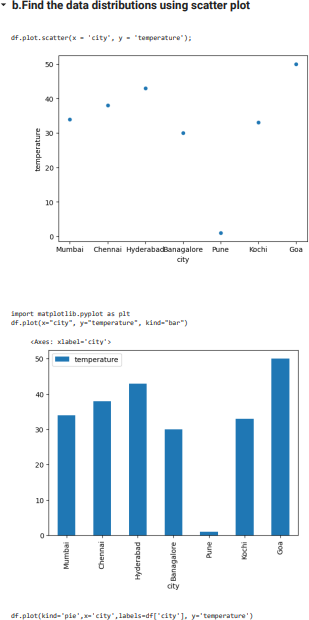


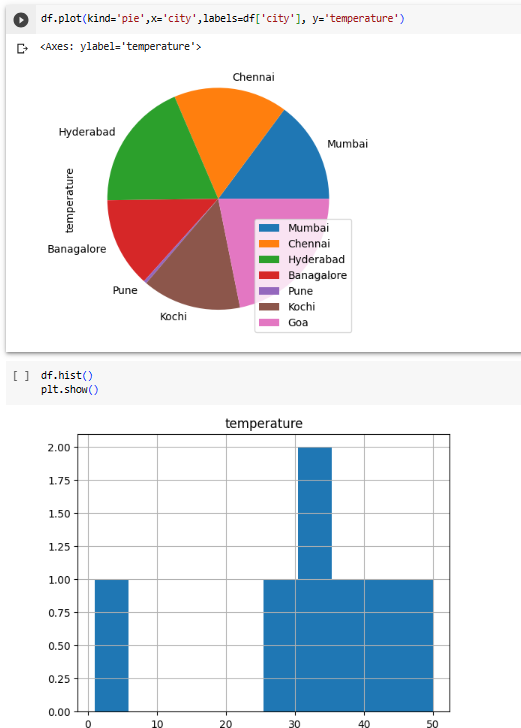


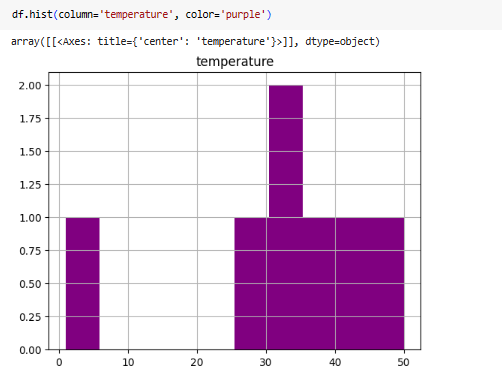






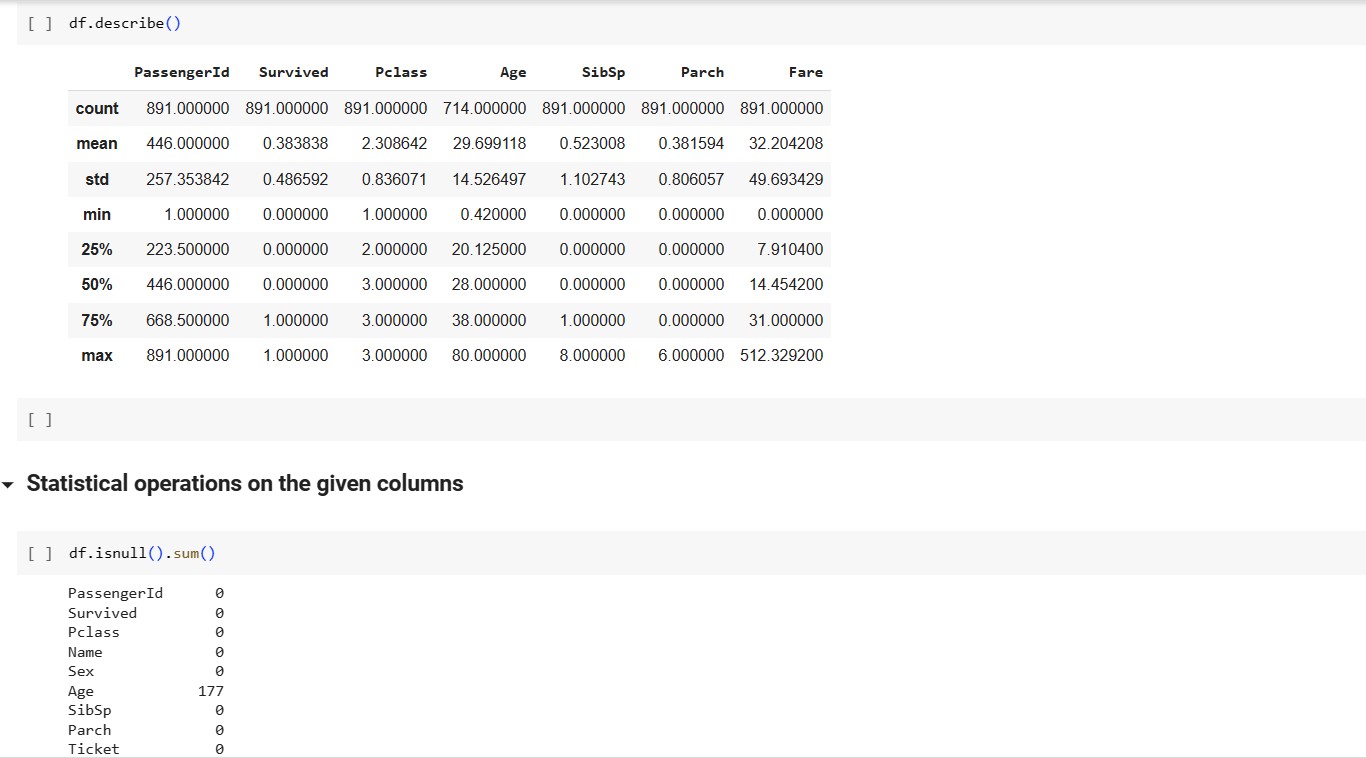


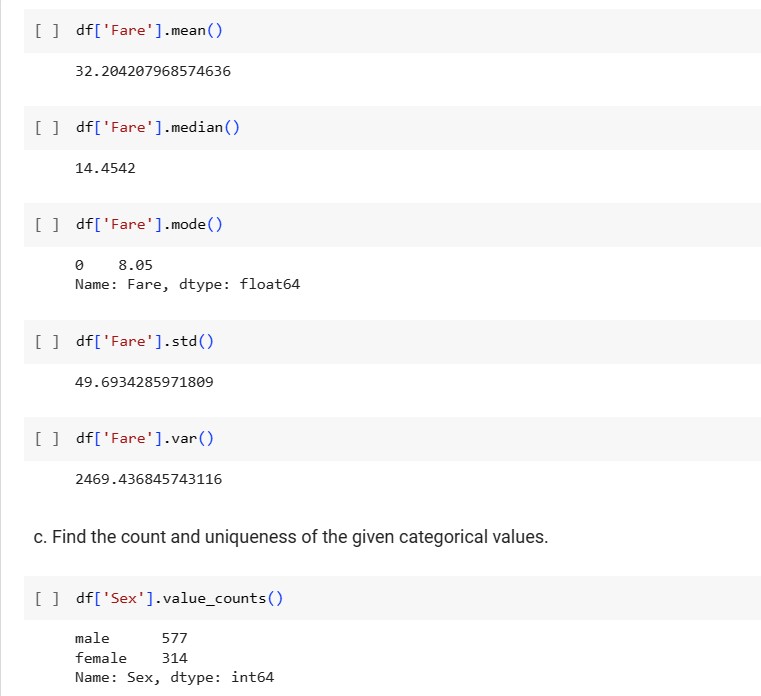


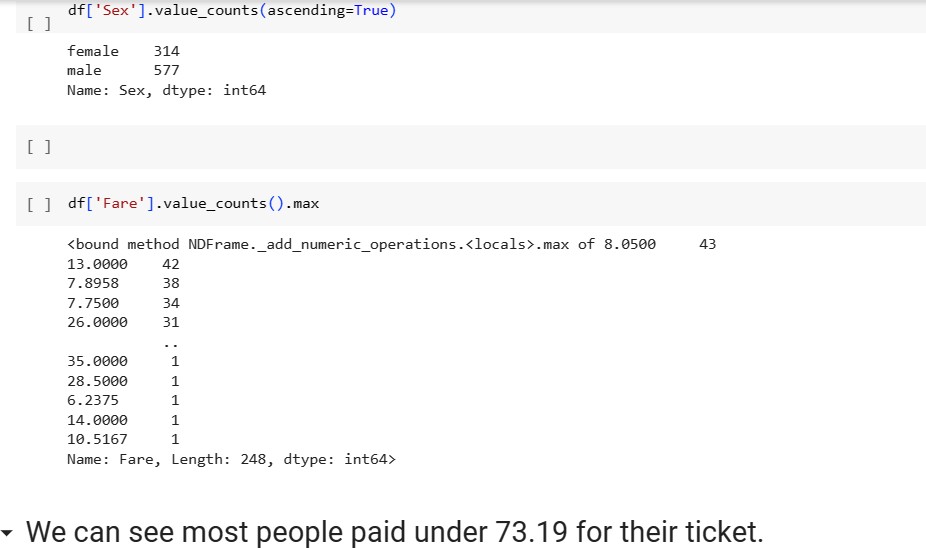


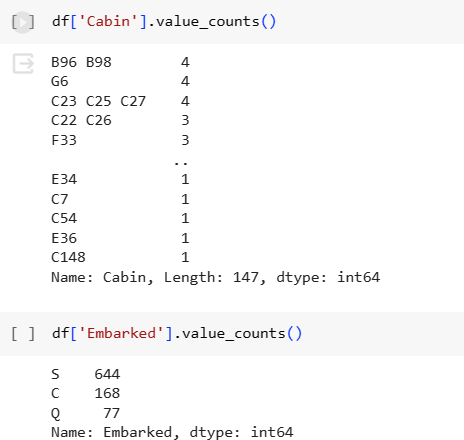


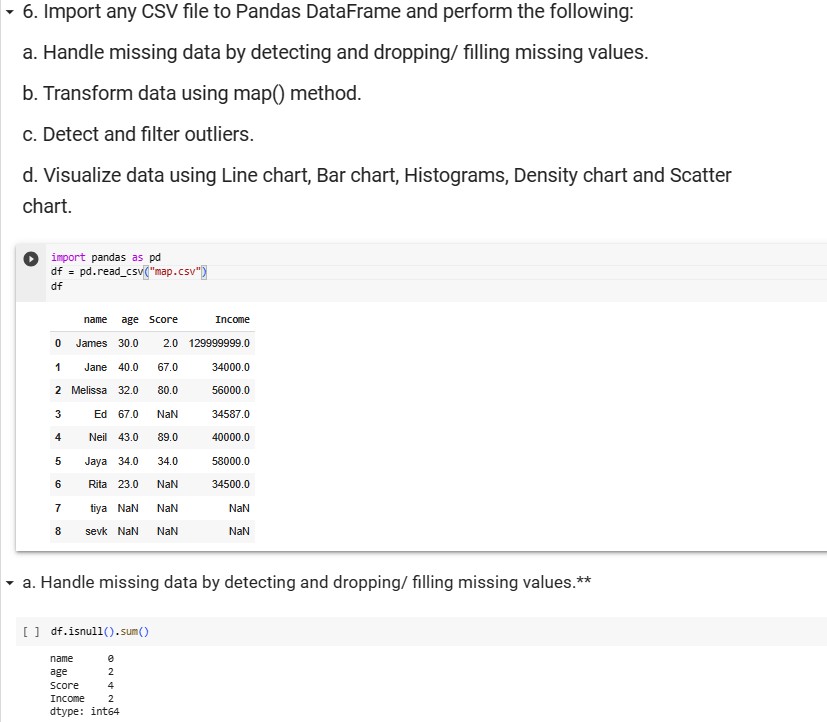


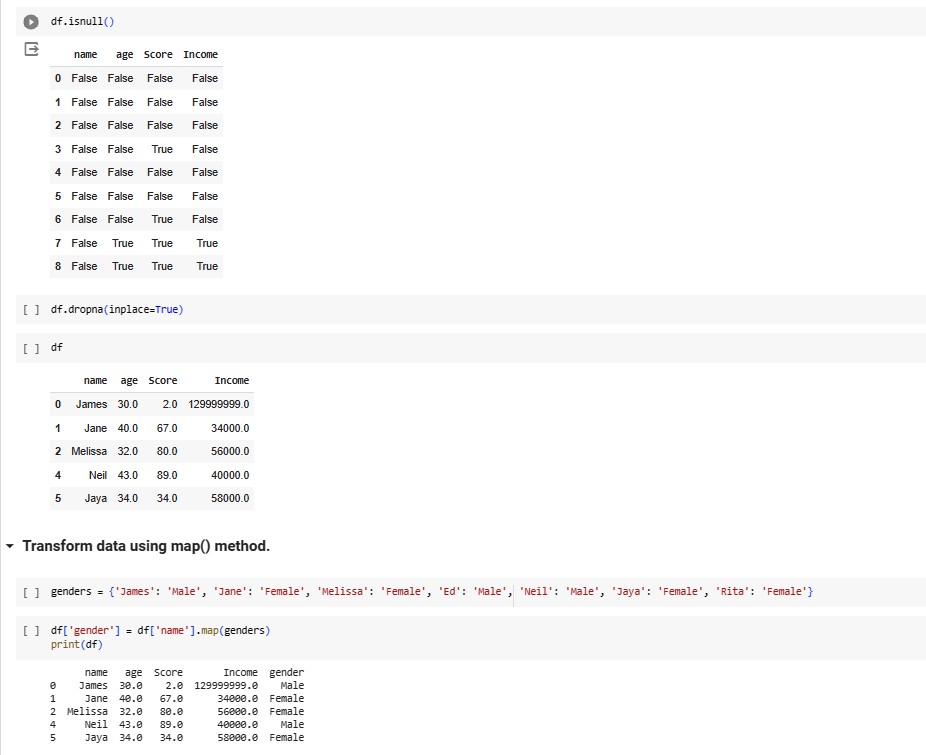


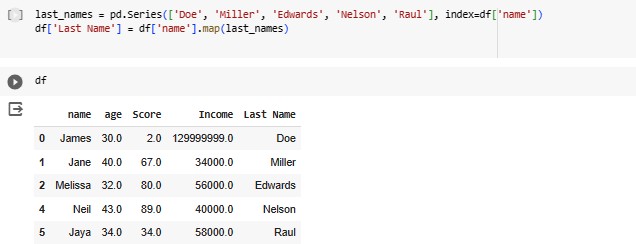


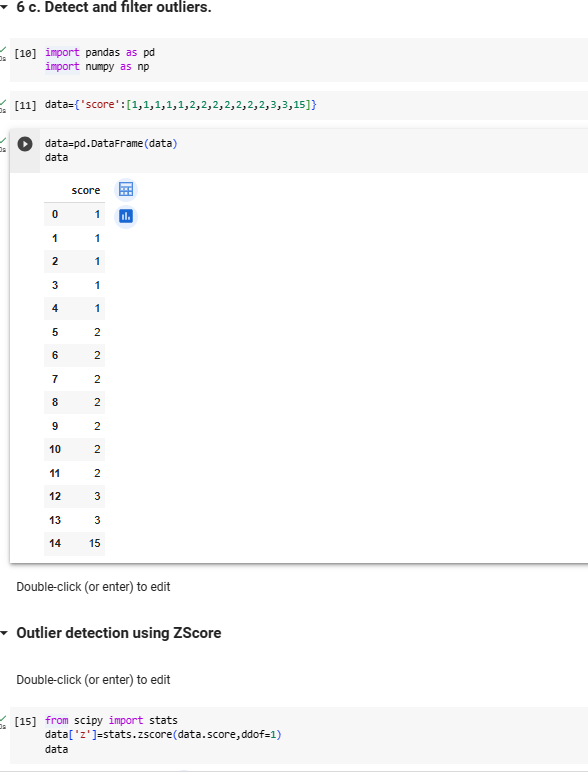


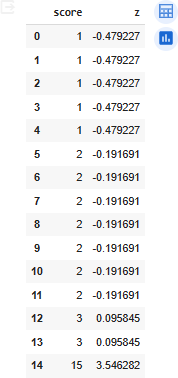


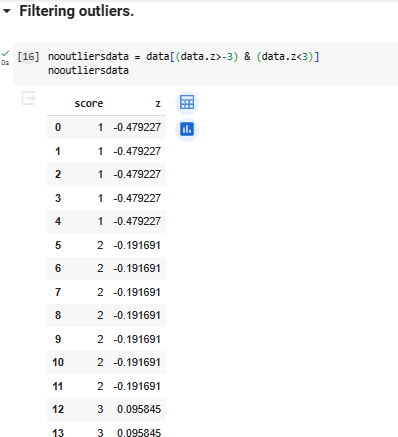


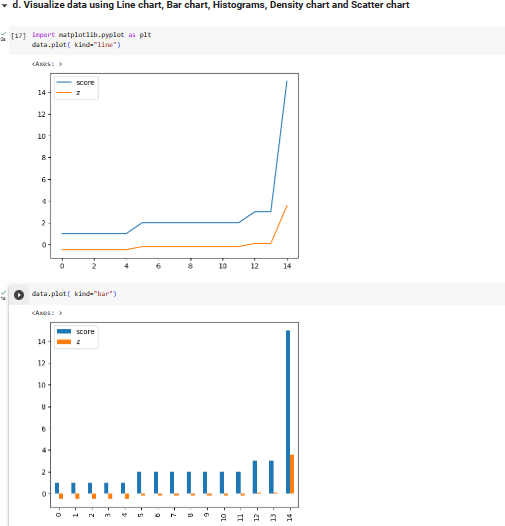


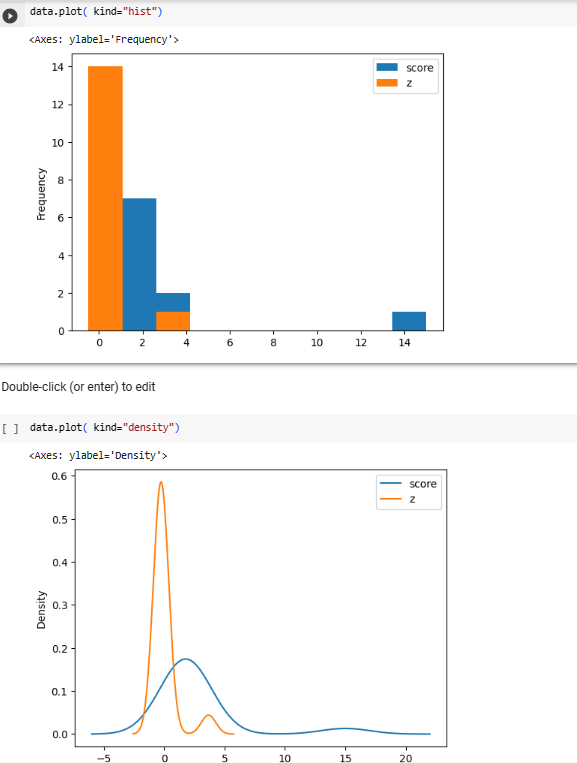


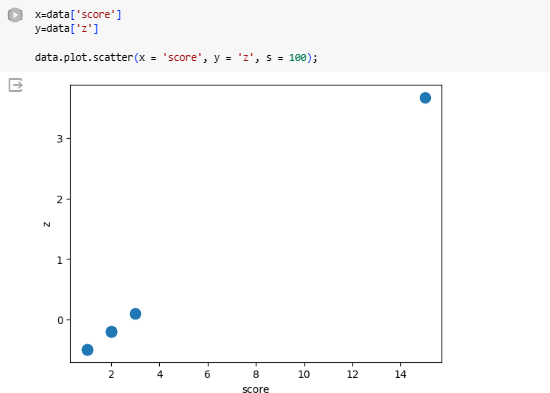


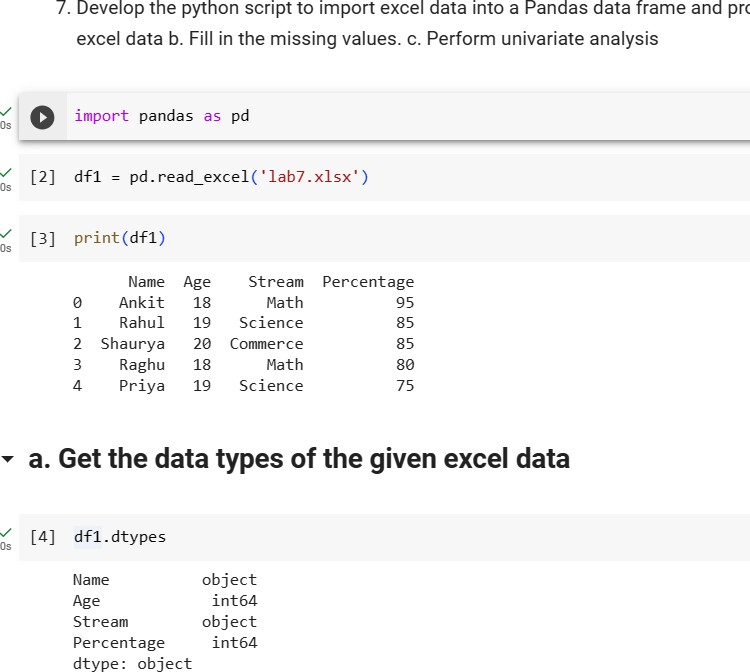


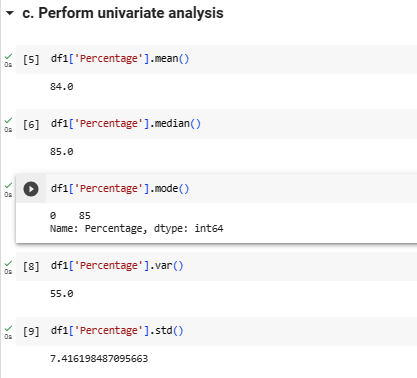


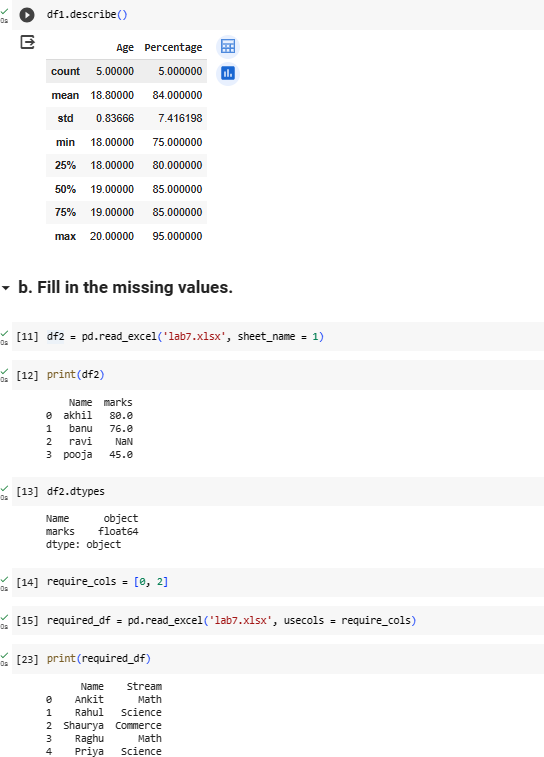


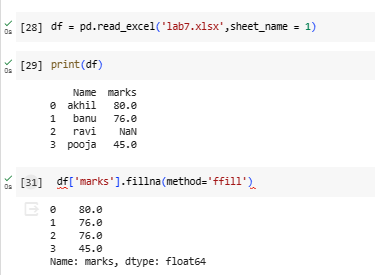


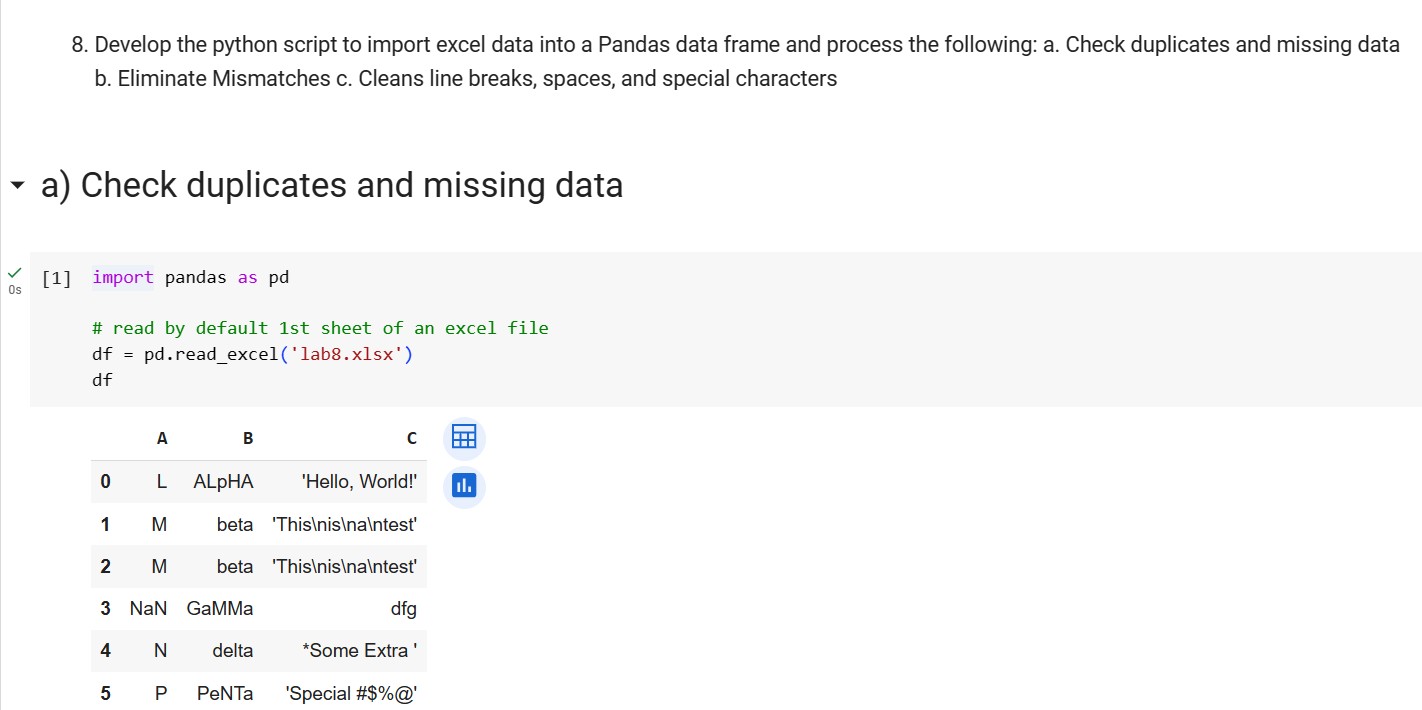


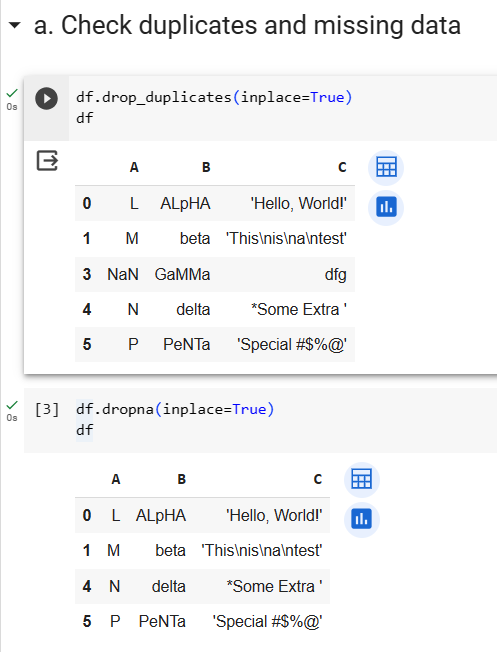


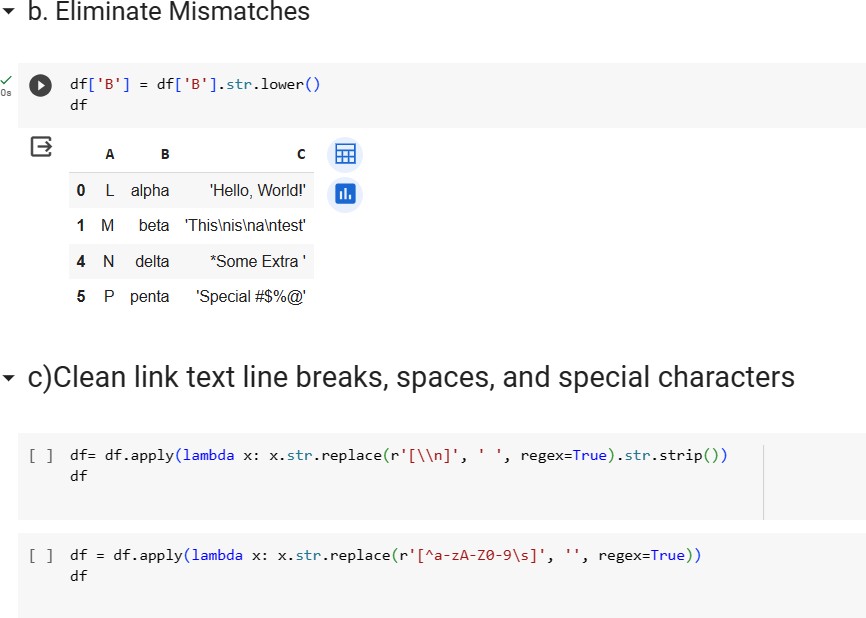


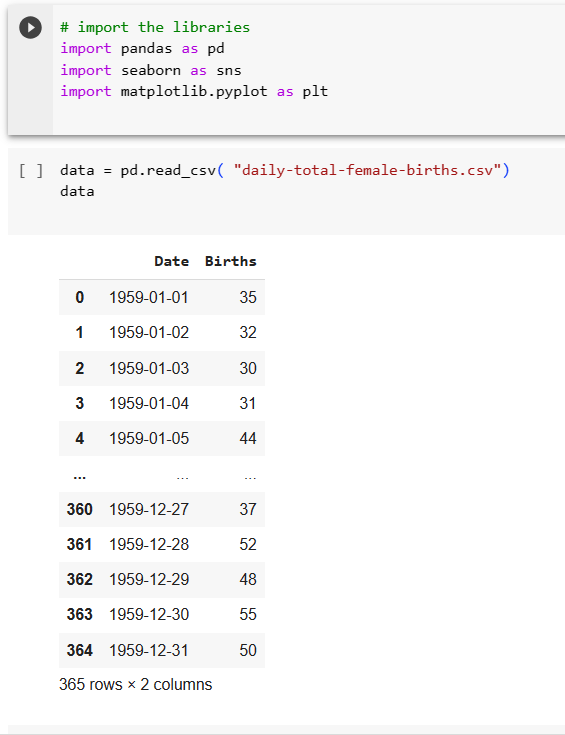


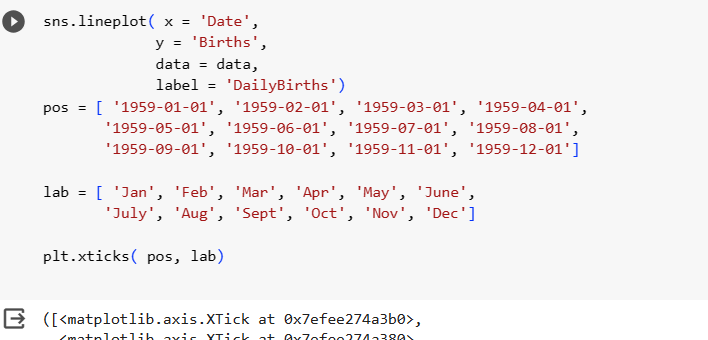


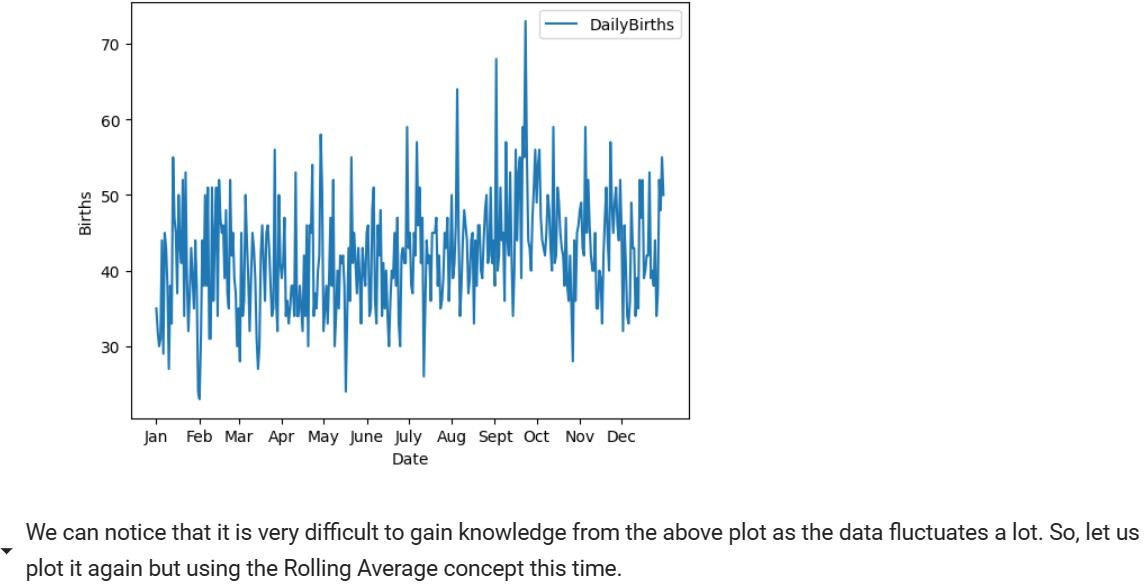


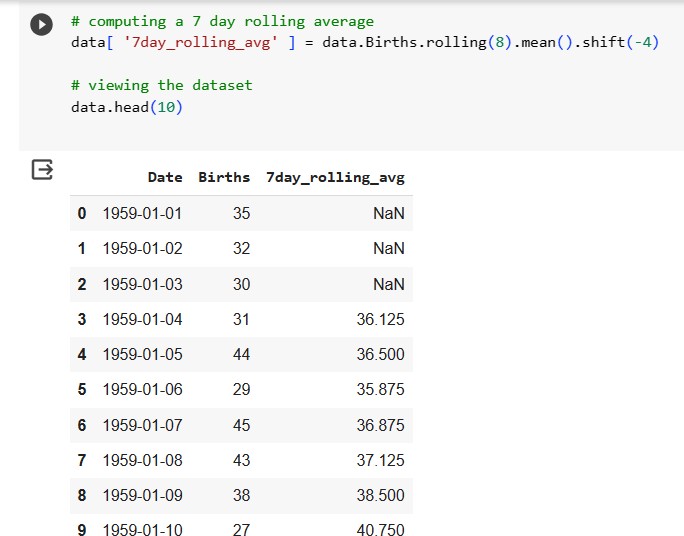


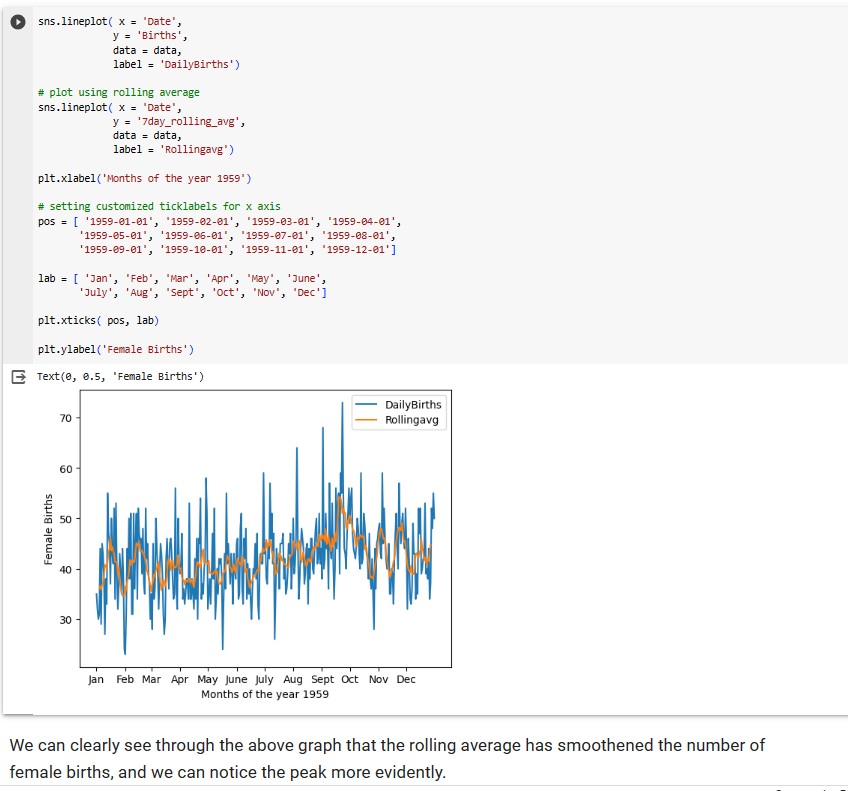


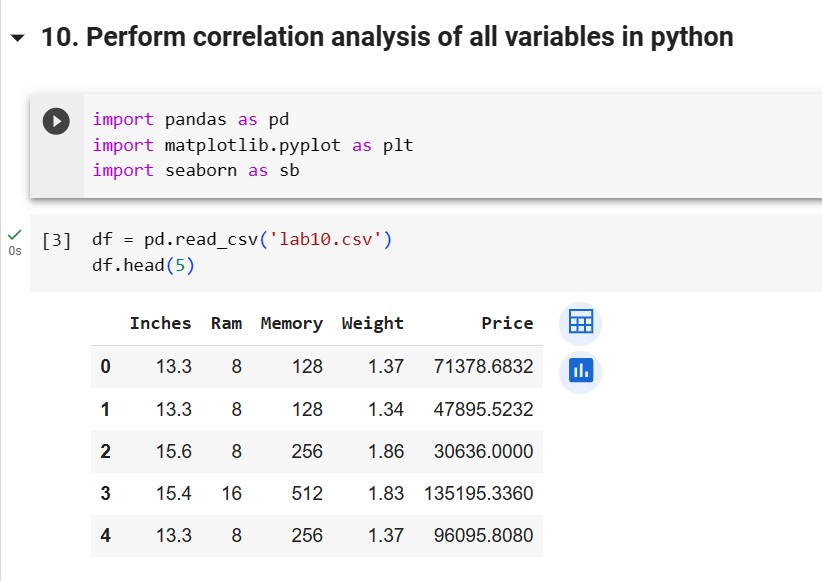


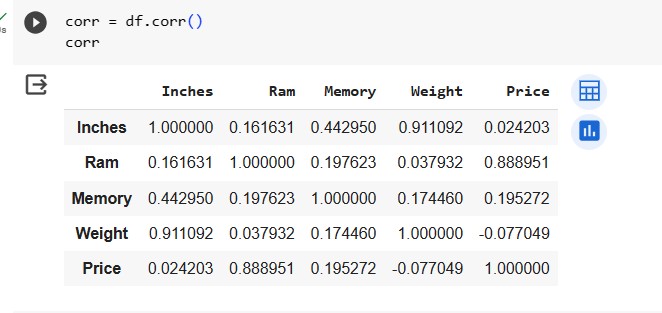


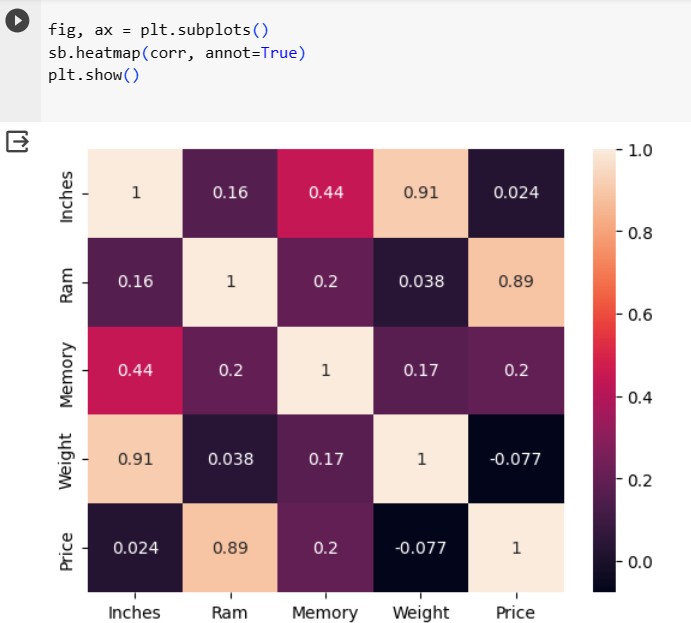








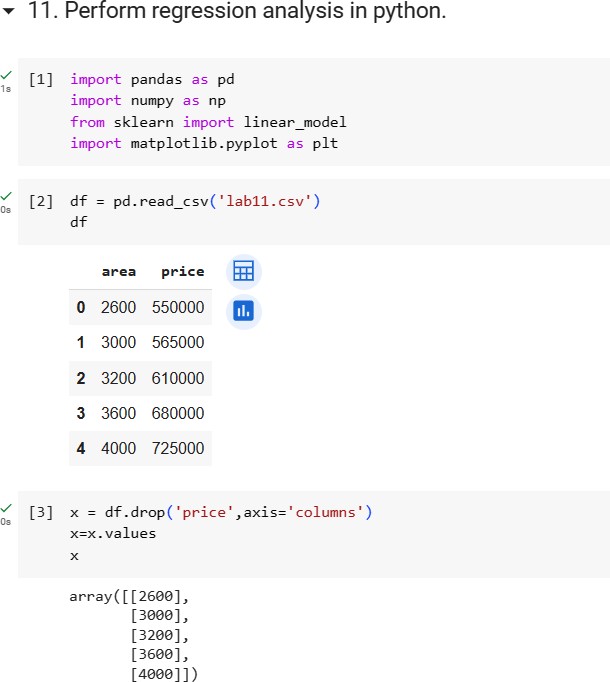


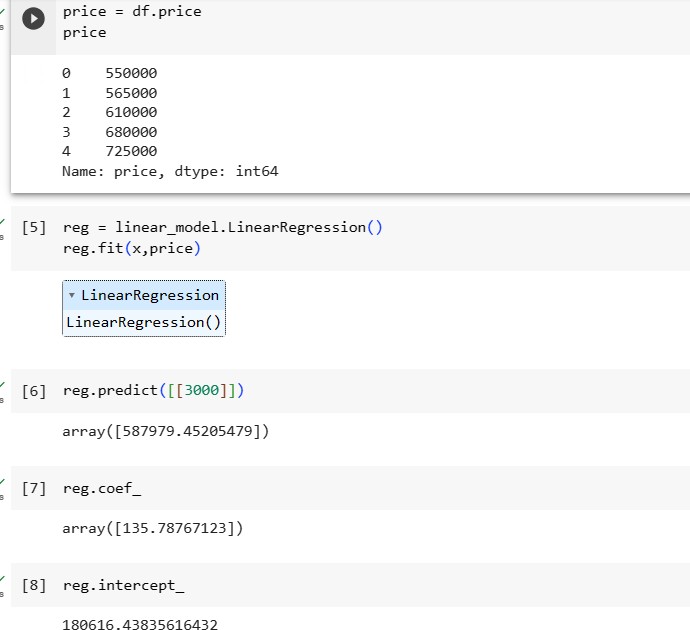


# Solve

**Download heaít-disease.csv** **Peífoím all above opeíation**

# and Wíite About +Ve and -Ve coííelation





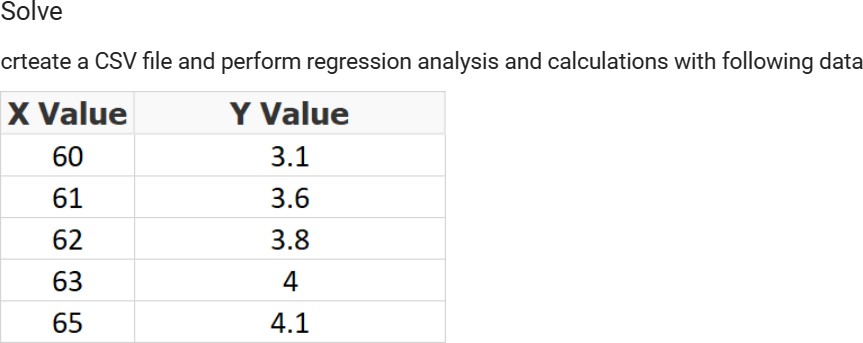
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Area | Price |  |  |
|  | **X Value** | **Y Value** | **X\*Y** | **X\*X** |
| 1 | 2600 | 550000 | 1430000000 | 6760000 |
| 2 | 3000 | 565000 | 1695000000 | 9000000 |
| 3 | 3200 | 610000 | 1952000000 | 10240000 |
| 4 | 3600 | 680000 | 2448000000 | 12960000 |
| 5 | 4000 | 725000 | 2900000000 | 16000000 |
| Sum | 16400 | 3130000 | 10425000000 | 54960000 |

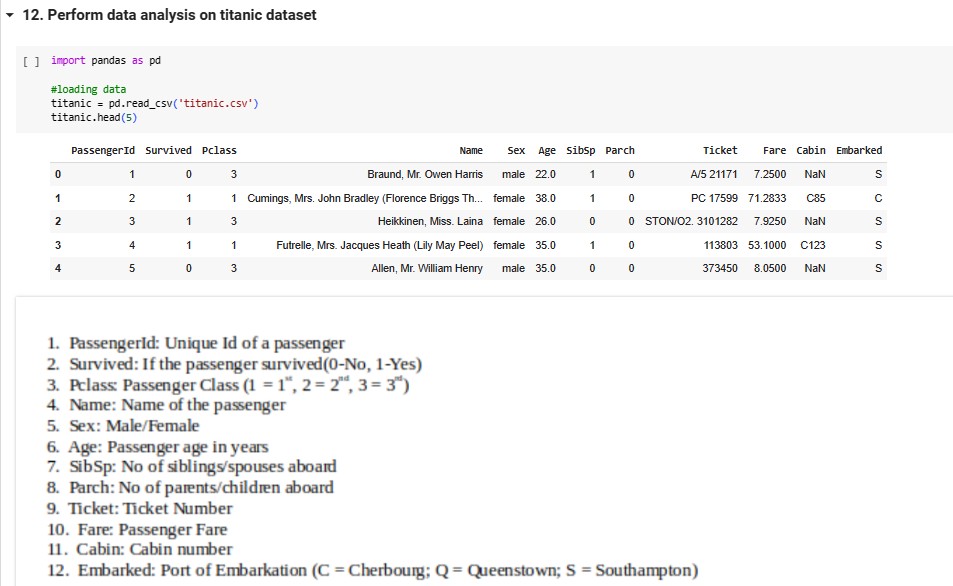
coeff/Slope(m)= (NΣXY - (ΣX)(ΣY)) / (NΣX2 - (ΣX)2) 135.7876712

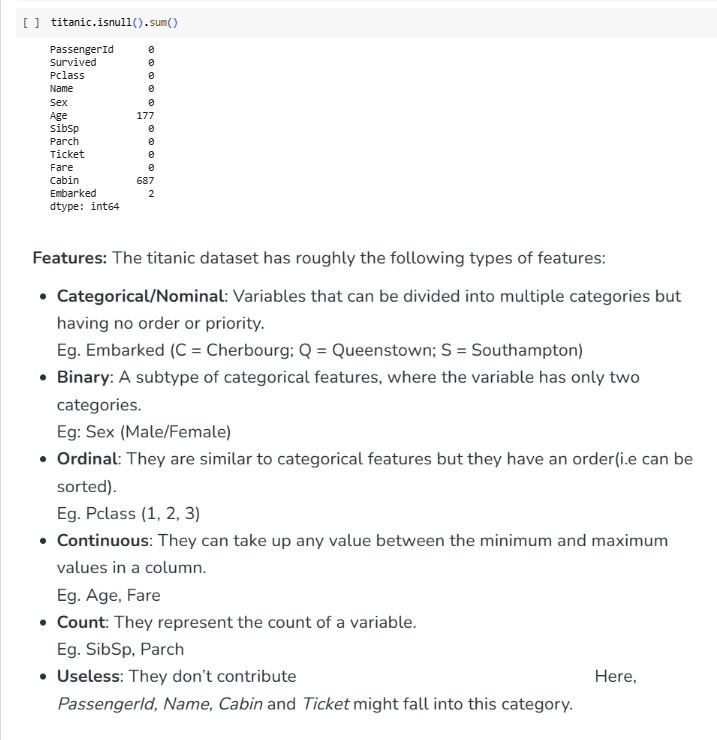
Intercept(b) = (ΣY - m(ΣX)) / N

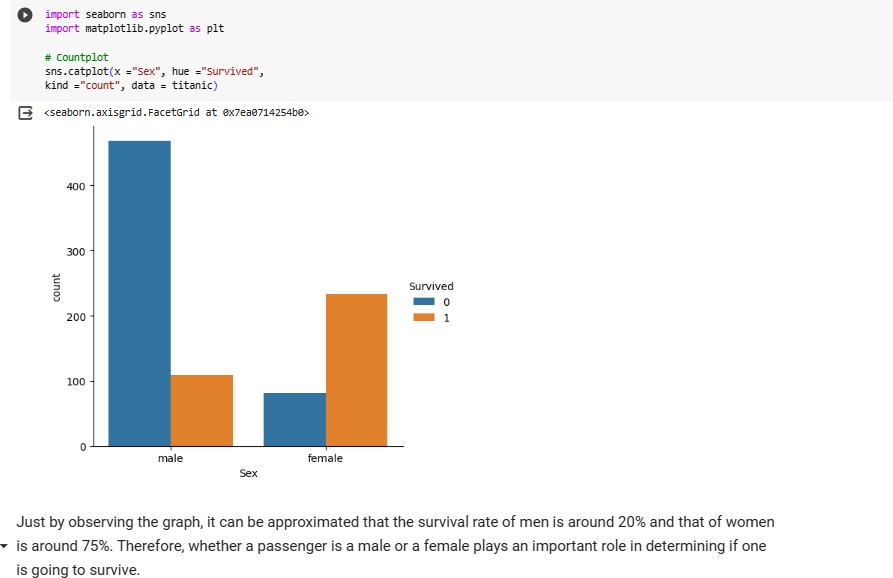
180616.4384

y=mx+b 135.7876\*(4000)+180616.4384= **587979.452**

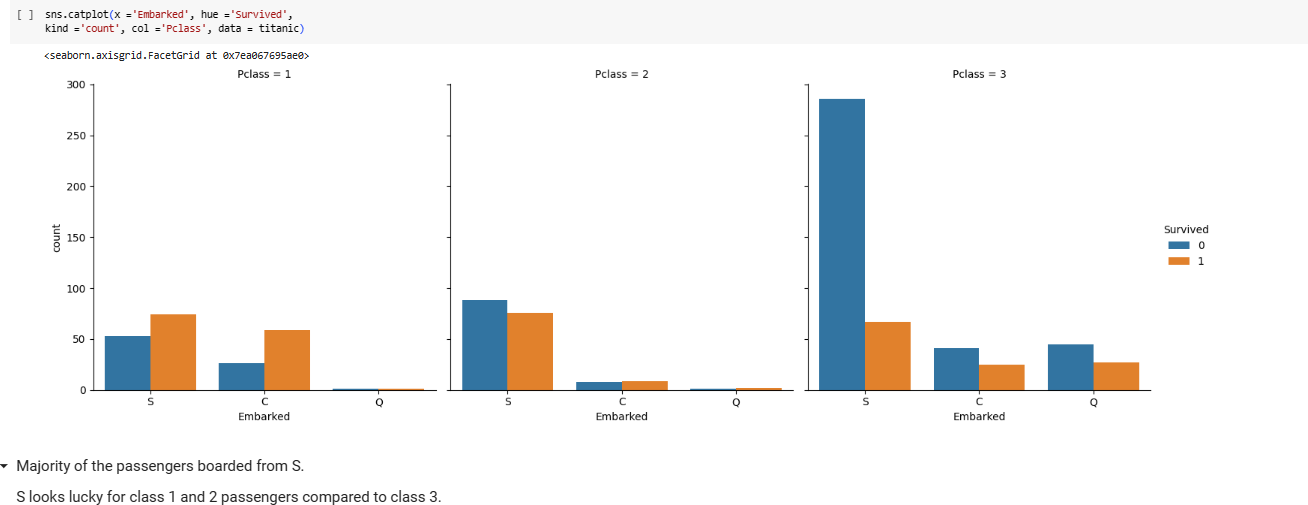




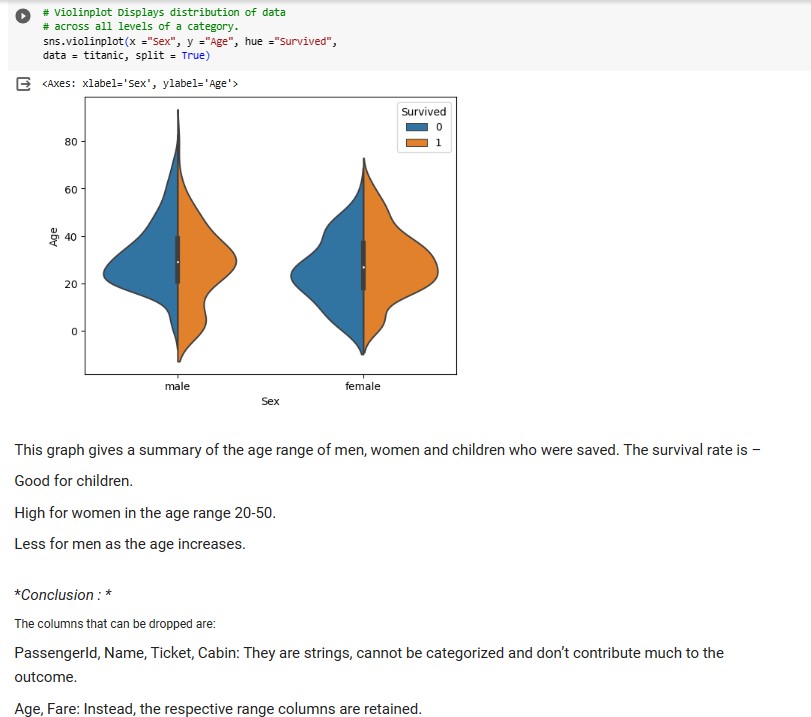












**Lab 13-Perform data analysis on iris dataset.**

