

TITAINIC DATASET

SUB= INTRODUCTION TO PROBLEM SLOVING

NAME= VISHAL VERMA

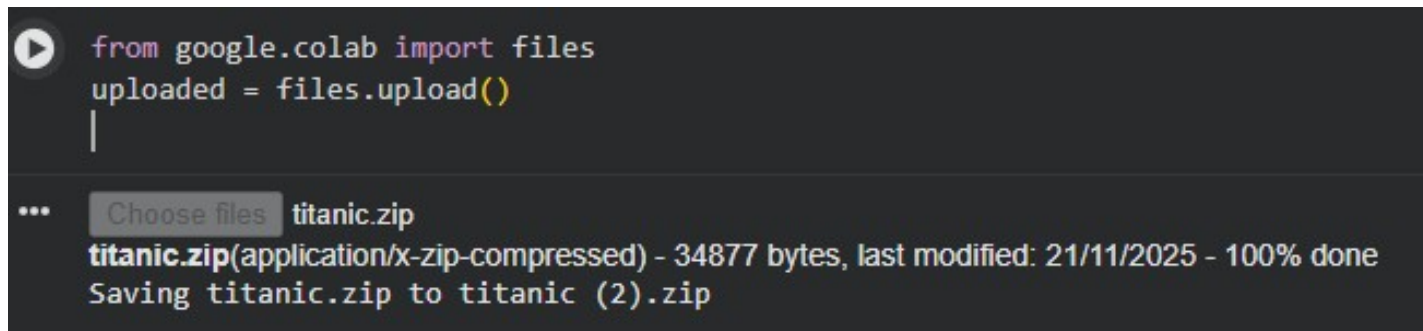
REG NO = 25MIP10001

INTROUCTION

- The Titanic survival dataset is one of the most popular dataset in data science. It contains information about passengers on the Titanic ship, such as their, age ,gender , ticket class, fare and most importantly. Whether they survived or not.
- The aim of the project is to --
- Upload the dataset(in csv format) to Google Colab
- Extract the file
- Load the CSV dataset
- Explore basic dataset information
- Perform basic analysis
- Calculate the survival count

DATA UPLOAD

- First , we have to download the dataset. After that, we will write a code to upload the file in Google colab.



The screenshot shows a Google Colab interface. At the top, there is a code cell with a play button icon on the left. The code inside the cell is: `from google.colab import files`, `uploaded = files.upload()`, and a cursor on the next line. Below the code cell, there is a file upload interface. It shows a button labeled "Choose files" followed by the filename "titanic.zip". Below this, a status message reads: "titanic.zip(application/x-zip-compressed) - 34877 bytes, last modified: 21/11/2025 - 100% done". At the bottom, it says "Saving titanic.zip to titanic (2).zip".

```
from google.colab import files
uploaded = files.upload()
|
```

... Choose files titanic.zip
titanic.zip(application/x-zip-compressed) - 34877 bytes, last modified: 21/11/2025 - 100% done
Saving titanic.zip to titanic (2).zip

LOAD & VIEW DATA

- This command is used in Python to load the data from the file that we have uploaded.

```
import pandas as pd

df = pd.read_csv("extracted_files/train.csv")
df.head()
```

	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

BASIC DATASET

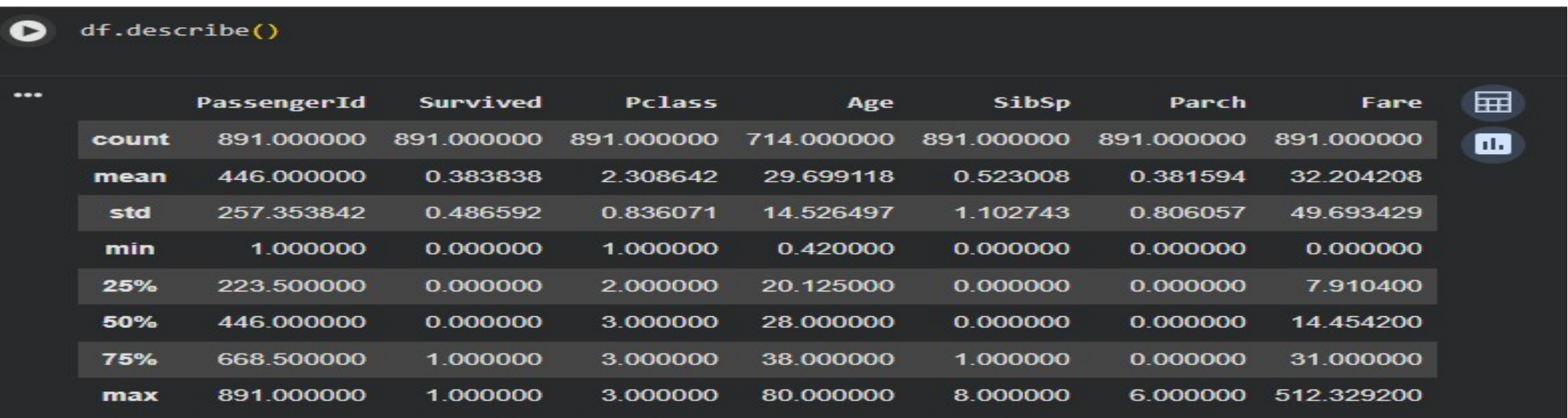
column types refer to the kind of data stored in each column are the empty or undefined entries in a dataset.

```
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    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
```

STATISTICAL SUMMARY

The `describe()` function is used to generate a statistical summary of the dataset. It provides important information such as the mean, minimum value, maximum value, standard deviation, and quartiles for each numerical column. This helps us understand the overall distribution and basic characteristics of data.



A screenshot of a Jupyter Notebook interface. At the top, a code cell contains the command `df.describe()`. Below it, the output is displayed as a table with 8 columns: `PassengerId`, `Survived`, `Pclass`, `Age`, `SibSp`, `Parch`, and `Fare`. The rows represent statistical measures: `count`, `mean`, `std`, `min`, `25%`, `50%`, `75%`, and `max`. The table is styled with alternating light and dark gray rows. On the right side of the table, there are two icons: a grid icon and a bar chart icon.

	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
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

Survival count

To check how many passengers survived:

```
df['Survived'].value_counts()
```

	count
Survived	
0	549
1	342

dtype: int64

Summary

- In this project , the Titanic Survival dataset was uploaded, extracted, loaded , and explored. We examined dataset information. Performed descriptive statistical analysis, and computed the survival count. This project provides a simple and effective introduction to data analysis using python.