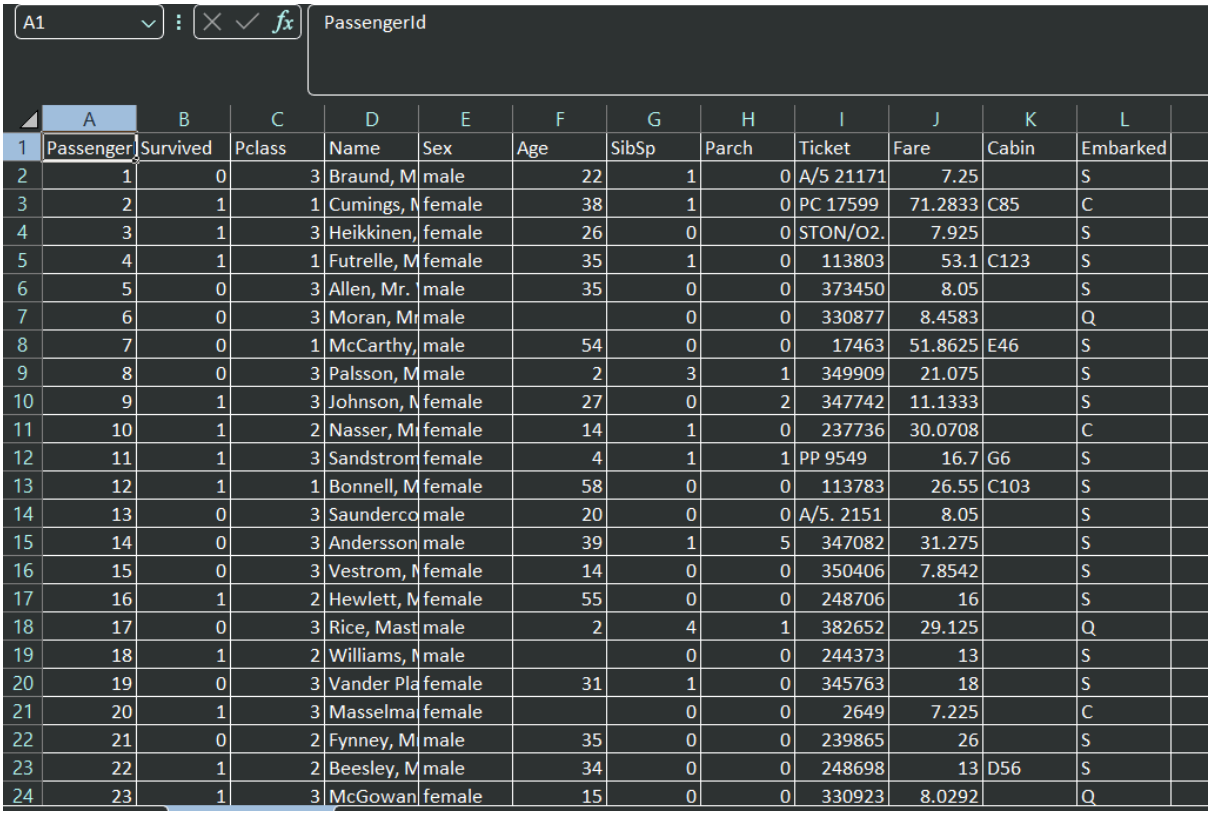
TASK 1

TITANIC SURVIVAL PREDICTION

- Use the Titanic dataset to build a model that predicts whether a passenger on the Titanic survived or not. This is a classic beginner project with readily available data.
- The dataset typically used for this project contains information about individual passengers, such as their age, gender, ticket class, fare, cabin, and whether or not they survived.

DATASET CLICK HERE

TITANIC DATASET



A1 | PassengerId

	A	B	C	D	E	F	G	H	I	J	K	L
1	Passenger	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
2	1	0	3	Braund, M	male	22	1	0	A/5 21171	7.25		S
3	2	1	1	Cumings, M	female	38	1	0	PC 17599	71.2833	C85	C
4	3	1	3	Heikkinen,	female	26	0	0	STON/O2.	7.925		S
5	4	1	1	Futrelle, M	female	35	1	0	113803	53.1	C123	S
6	5	0	3	Allen, Mr.	male	35	0	0	373450	8.05		S
7	6	0	3	Moran, Mr	male		0	0	330877	8.4583		Q
8	7	0	1	McCarthy,	male	54	0	0	17463	51.8625	E46	S
9	8	0	3	Palsson, M	male	2	3	1	349909	21.075		S
10	9	1	3	Johnson, M	female	27	0	2	347742	11.1333		S
11	10	1	2	Nasser, Mr	female	14	1	0	237736	30.0708		C
12	11	1	3	Sandstrom	female	4	1	1	PP 9549	16.7	G6	S
13	12	1	1	Bonnell, M	female	58	0	0	113783	26.55	C103	S
14	13	0	3	Saunderco	male	20	0	0	A/5. 2151	8.05		S
15	14	0	3	Andersson	male	39	1	5	347082	31.275		S
16	15	0	3	Vestrom, M	female	14	0	0	350406	7.8542		S
17	16	1	2	Hewlett, M	female	55	0	0	248706	16		S
18	17	0	3	Rice, Mast	male	2	4	1	382652	29.125		Q
19	18	1	2	Williams, M	male		0	0	244373	13		S
20	19	0	3	Vander Pla	female	31	1	0	345763	18		S
21	20	1	3	Masselma	female		0	0	2649	7.225		C
22	21	0	2	Fynney, M	male	35	0	0	239865	26		S
23	22	1	2	Beesley, M	male	34	0	0	248698	13	D56	S
24	23	1	3	McGowan	female	15	0	0	330923	8.0292		Q

CODE

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
import tkinter as tk
from tkinter import filedialog

# Step 1: File Selection
def select_file():
    """Open a dialog for user to select the Titanic dataset file."""
    root = tk.Tk()
    root.withdraw()  # Hide the root window
    file_path = filedialog.askopenfilename(title="Select Titanic Dataset CSV File",
                                           filetypes=[("CSV Files", "*.csv")])
    return file_path

# Select the Titanic dataset
file_path = select_file()

if not file_path:
    print("No file selected. Exiting...")
else:
    # Step 2: Load the Dataset
    print(f"Loading dataset from: {file_path}")
    titanic_data = pd.read_csv(file_path)

    # Step 3: Handle Missing Values
    titanic_data['Age'].fillna(titanic_data['Age'].median(), inplace=True)
    titanic_data.drop(columns=['Cabin'], inplace=True)  # Drop 'Cabin' column
    titanic_data['Embarked'].fillna(titanic_data['Embarked'].mode()[0], inplace=True)

    # Step 4: Encode Categorical Variables
    label_encoder = LabelEncoder()
    titanic_data['Sex'] = label_encoder.fit_transform(titanic_data['Sex'])  # male = 1, female = 0
    titanic_data['Embarked'] = label_encoder.fit_transform(titanic_data['Embarked'])  # C, Q, S to 0, 1, 2

    # Step 5: Drop Irrelevant Columns
    titanic_data.drop(columns=['PassengerId', 'Name', 'Ticket'], inplace=True)

    # Step 6: Define Features (X) and Target (y)
    X = titanic_data.drop(columns=['Survived'])
    y = titanic_data['Survived']
```

```
# Step 7: Split Data into Train and Test Sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Step 8: Train the Model
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)

# Step 9: Evaluate the Model
y_pred = model.predict(X_test)

print("\nModel Evaluation Results:")
print(f"Accuracy: {accuracy_score(y_test, y_pred):.2f}")
print("\nConfusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))

# Step 10: Predict Survival for a New Input
print("\n--- Survival Prediction ---")
print("Enter the following details for survival prediction:")
try:
    Pclass = int(input("Enter Passenger Class (1 = First, 2 = Second, 3 = Third): "))
    Sex = input("Enter Sex (male or female): ").strip().lower()
    Age = float(input("Enter Age: "))
    SibSp = int(input("Enter Number of Siblings/Spouses Aboard: "))
    Parch = int(input("Enter Number of Parents/Children Aboard: "))
    Fare = float(input("Enter Fare Amount: "))
    Embarked = input("Enter Port of Embarkation (C, Q, S): ").strip().upper()

    # Encode inputs
    Sex = 1 if Sex == "male" else 0
    Embarked = {"C": 0, "Q": 1, "S": 2}.get(Embarked, 0)
```

```
    # Encode inputs
    Sex = 1 if Sex == "male" else 0
    Embarked = {"C": 0, "Q": 1, "S": 2}.get(Embarked, 0)

    # Create input array
    new_data = pd.DataFrame({
        'Pclass': [Pclass],
        'Sex': [Sex],
        'Age': [Age],
        'SibSp': [SibSp],
        'Parch': [Parch],
        'Fare': [Fare],
        'Embarked': [Embarked]
    })

    # Predict survival
    prediction = model.predict(new_data)[0]
    survival_status = "Survived" if prediction == 1 else "Did Not Survive"
    print(f"\nPrediction: The passenger {survival_status}.")
except ValueError as e:
    print(f"Invalid input! Please try again. Error: {e}")
```

OUTPUT

```
>>>
= RESTART: C:/Users/HP/OneDrive/Desktop/CA/TASK-1.py
Loading dataset from: C:/Users/HP/Downloads/archive/Titanic-Dataset.csv

Model Evaluation Results:
Accuracy: 0.82

Confusion Matrix:
 [[92 13]
 [19 55]]

Classification Report:
              precision    recall  f1-score   support

           0       0.83      0.88      0.85       105
           1       0.81      0.74      0.77        74

    accuracy                           0.82       179
   macro avg       0.82      0.81      0.81       179
weighted avg       0.82      0.82      0.82       179


--- Survival Prediction ---
Enter the following details for survival prediction:
Enter Passenger Class (1 = First, 2 = Second, 3 = Third): 2
Enter Sex (male or female): MALE
Enter Age: 20
Enter Number of Siblings/Spouses Aboard: 1
Enter Number of Parents/Children Aboard: 2
Enter Fare Amount: 100
Enter Port of Embarkation (C, Q, S): C

Prediction: The passenger Did Not Survive.
>>>
```

BY KALPANA RAWAT