ASSIGNMENT – 1 REPORT

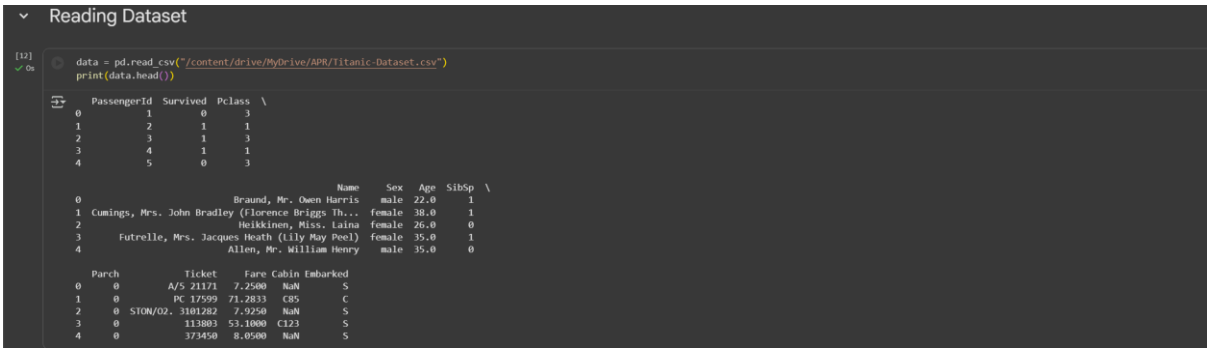
-Vaishika Agrawal (2201AI42)

1. Introduction

The Titanic dataset is a classic dataset for binary classification, aiming to predict whether a passenger survived or not based on features such as age, sex, ticket class, and fare.
In this assignment, I applied **Logistic Regression** to perform survival prediction.

2. Dataset Description

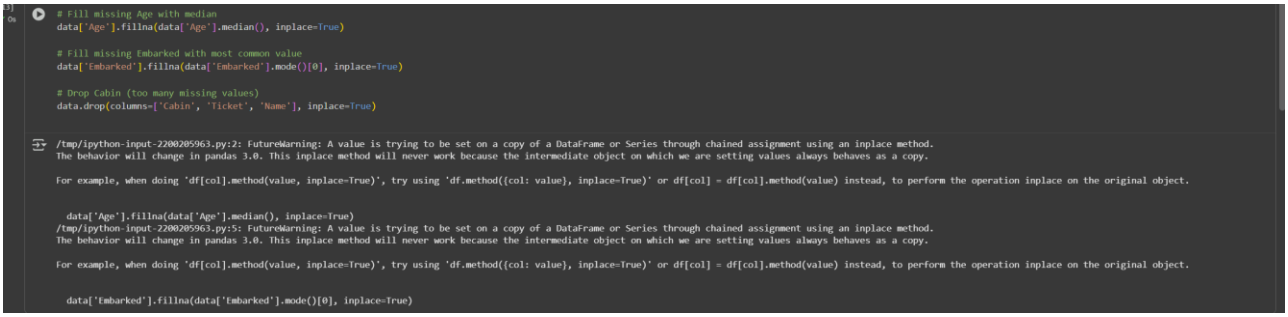
- **Rows:** 891 passengers
- **Columns:** PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked
- **Target variable:** Survived (0 = Did not survive, 1 = Survived)



3. Steps Performed

3.1 Data Preprocessing

- Handled missing values (‘Age’ filled with median, ‘embarked’ with mode).
- Converted categorical variables (‘Sex’ mapped to 0/1, ‘Embarked’ mapped to 0/1/2).
- Removed irrelevant columns (PassengerId, Name).



3.2 Feature Selection

```
X = data.drop('Survived', axis=1)
y = data['Survived']
```

Final features used for model training:
Pclass, Sex, Age, SibSp, Parch, Fare, Embarked_Q, Embarked_S

3.3 Train-Test Split

- Split data into training (80%) and testing (20%).

```
[16]
✓ 0s
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

3.4 Model Training

- Applied **Logistic Regression** with max_iter=500.

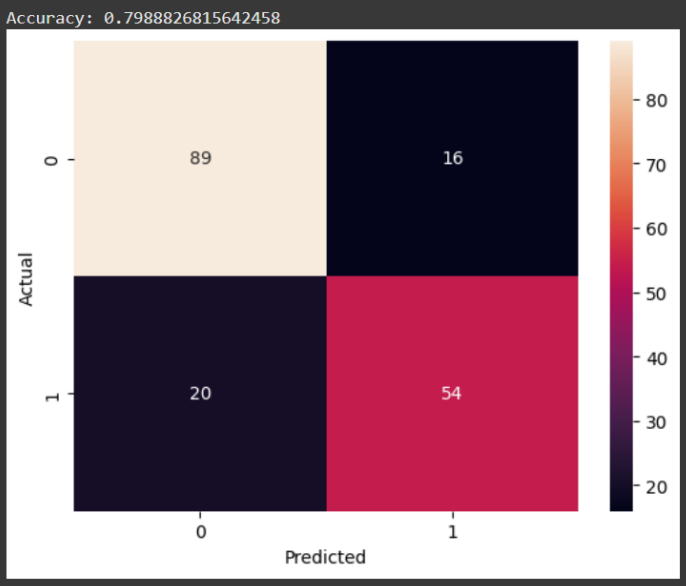
```
[17]
✓ 0s
model = LogisticRegression(max_iter=500)
model.fit(X_train, y_train)

LogisticRegression
LogisticRegression(max_iter=500)
```

3.5 Model Evaluation

- Evaluated with **Accuracy, Confusion Matrix, Precision, Recall, and F1-score**.

Final results

```
Accuracy: 0.7988826815642458
```




```
Classification Report:
              precision    recall  f1-score   support

           0       0.82      0.85      0.83       105
           1       0.77      0.73      0.75        74

    accuracy                           0.80       179
   macro avg       0.79      0.79      0.79       179
weighted avg       0.80      0.80      0.80       179
```

Accuracies

Training Accuracy: 0.8061797752808989

Training Error: 0.1938202247191011

Testing Accuracy: 0.7988826815642458

Testing Error: 0.2011173184357542

Colab notebook link: Assignment-1