Report on the Analysis of the Data Set:

Task 1: To get to know the Data Set, its Types,

Internship: AI & Machine Learning Internship.

Data Set: Titanic Data Set.

1. Introduction: The purpose of this task was to understand the Data Set before implementing any form of Machine Learning. Understanding your Data set provides you with insights on the types of Data, as well as any areas of Missed Data, your Target Variable, and important information about how your Data Set is positioned for the implementation of Machine Learning.

In order to analyse the Data Set used for this report, we have selected the Titanic Data Set.

2. An Overview of the Data Set:

- 891 total rows
- 12 total columns
- Structured tabular Data Set (e.g. a table)
- Each row is a passenger, each column has information pertaining to that passenger.

3. Types of Data Identified:

- Numerical Features:

PassengerId, Survived, Pclass, Age, SibSp, Parch and Fare;

- Categorical Features:

Name, Sex, Ticket, Cabin and Embarked.

- Binary Features:

Survived: Is it a 0/1 Value; Sex: Male/Female.

A detailed understanding of each Data Type is very important. Each Data Type is treated differently in the Data Preparation phase of Machine Learning.

4. Summary Statistics:

The average Age of passengers was about 29.7 years. Passengers ranged in Age from 0.42-80 years. The average Fare for passengers was approximately 32.20. All Data Types were Integer, Float and Object.

5. Missing Values:

Identified Columns with Missing Values:

Age: 177 Missing Values;

Cabin: 687 Missing Values;

Embarked: 2 Missing Values.

These Missing Values will need to be dealt with during the Data Preparation phase.

6. Target and Features:

Target Variable = Survived

Features = All other Columns. (Including Age, Sex, Pclass and Fare).