## Artificial Intelligence and Machine Learning

Project Report

Semester-IV (Batch-2022)

**Case Study**: - Train Data Assignment

[**Url:-**](about:blank) <https://drive.google.com/file/d/1E8AYysczbFa06IG6WfzbjtcE_tkACFOz/view?usp=sharing>

A red and white sign

Description automatically generated with low confidence

**Supervised By: Submitted By:**

Rajeev Thakur Kartik Taneja

Roll Number - 2210931007

Group - 14

**Department of Computer Science and Engineering**

Chitkara University Institute of Engineering & Technology,

Chitkara University, Punjab

**Description about Case Study: -**

● Read the Salaries and Train Dataset

● Display top 5 rows of the dataset

● Display last 3 rows of the dataset

● Find shape of our dataset ( number of rows and columns )

● Get information about our dataset

● Get overall statistics about the dataframe

● Filter the data

● Check the null values in the dataset

● Drop the column

● Handle missing values

● Categorical data encoding

● Describe univariate analysis

● Check How many people survived and died and plot it on graph

● Check how many passengers were in first, second, third class. Plot those figures on graph

● Display the number of male and female passengers.

● Describe bivariate analysis

● Who has better chance of survival male or female

● Which passenger has better chance of survival (First , second or third class)

● Describe feature engineering

**Library: -**

● Pandas , MatplotLib , mysql.connector

**Methods: -**

**1.** **head():** Description: Displays the first few rows of the data frame.

**2.** **tail():** Description: Displays the last few rows of the data frame.

**3.** **shape():** Description: Returns the shape (number of rows, number of columns) of the data frame.

**4.** **info():** Description: Provides basic information about the data frame, such as column types and missing values.

**5.** **Describe():** Description: It generates descriptive statistics of the numerical columns in a dataframe.

**6.** **Filter():** Description: It is used to select or filter specific columns from a Dataframe based on their labels or column names.

**7.** **isnull():** Description: Returns True/False for each value in the data frame, indicating whether the value is missing (NaN) or not.

**8.** **drop():** Description: Removes specific rows or columns from the data frame.

**9.** **Handle Missing Values:** It is used to ensure the quality and reliability of your analysis or machine learning model.

**10.** **value\_counts():** Description: Counts the unique values in a specific column of the data frame.

**11.** **Plt.figure():** Description: It initializes a new figure with a specific size.

**12.** **Plt.bar():** Description: It creates a bar plot.

**13.** **Plt.xticks():** Description: This method sets the x-axis tick labels.

**14.** **Plt.xlabel():** Description: It sets the label for x-axis.

**15.** **Plt.ylabel():** Description: It sets the label for y-axis.

**16.** **Plt.show():** Description: It displays the plot.