## Artificial Intelligence and Machine Learning

Project Report

Semester-IV (Batch-2022)

**Case Study**: - Salaries Dataset

[Url:-](about:blank) A red and white sign

Description automatically generated with low confidence

**Supervised By: Submitted By:**

Rajeev Bharadwaj Tanya Batra

Roll Number: -2210990907

Group - 13

**Department of Computer Science and Engineering**

## Chitkara University Institute of Engineering & Technology,

## Chitkara University, Punjab

**Description about Case Study: -**

* Read dataset Salaries
* Display Top 10 rows
* Display the Last 10 rows
* Check the shape of dataset
* Check null values in the Dataset
* How many rows and columns are in our dataset, the datatypes of each column, and the memory requirement
* Drop ID, Notes, Agency, and Status Columns
* Find the occurrence of Employee name (Top 5)
* Find the number of unique job titles
* Total number of job titles containing the captain
* Display all Employee names from the fire department
* Find the minimum, maximum, and average base pay
* Replace ‘Not Provided’ in the Employee Column with NaN
* Drop the rows that have more than 5 missing values
* Find the job title of Albert Pardini
* How much does Albert Pardini make?
* Display the name of the person having the highest base pay
* Display the average base pay of all employees per year
* Find the average base pay of all employees per job title
* Find the average base pay of all employees having job title Accountant
* Find Top 5 most common jobs

**Library: -**

* Pandas

**Methods: -**

1. **read\_csv():**

Description: Reads a CSV file and converts it into a data frame.

1. **tail():**

Description: Displays the last few rows of the data frame.

1. **head():**

Description: Displays the first few rows of the data frame.

1. **shape():**

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

1. **info():**

Description: Provides basic information about the data frame, such as column types and missing values.

1. **isnull():**

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

1. **sum():**

Description: Calculates the sum of values in each column of the data frame.

1. **dropna():**

Description: used to remove missing (NaN) values from a Data Frame or Series

1. **contains():**

Description: Checks if a specified substring or value is present in a column of the data frame.

1. **max():**

Description: Returns the maximum value in a column of the data frame.

1. **min():**

Description: Returns the minimum value in a column of the data frame.

1. **mean():**

Description: Calculates the mean (average) value of a column in the data frame.

1. **len():**

Description: Returns the number of rows in the data frame

1. **sort\_values():**

Description: Helps arrange the data in either ascending or descending order based on the values in the specified columns

1. **to\_numeric:**

Description: Used to convert the values of a Series to numeric format.

1. **value\_counts():**

Description: Counts the unique values in a specific column of the data frame.

1. **replace():**

Description: Used to replace values in a Data Frame or Series.