## **Artificial Intelligence and Machine Learning**

Project Report

Semester-IV (Batch-2022)

**Case Study**: - Employee Salaries Dataset

[Url:-](about:blank) https://drive.google.com/file/d/1Vo3zGqi2BN8JoGQqU9sZHK7GYxORfe-s/view?usp=drive\_link

A red and white sign

Description automatically generated with low confidence

**Supervised By: Submitted By:**

Rajeev Thakur Yazvi Balooni

Roll Number: -2210990985

Group - 14

**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
* Find the shape of database
* Getting information about our dataset like total no. of rows, columns, datatype of each column and memory requirement
* Check null values in the dataset
* Drop ID, notes, agency and status columns
* Find Occurence of the employee Name(top5)
* Find the number of unique job titles
* Total number of job titles contains captain
* Display all the employee names from fire department
* Find min,max and average basepay
* Replace 'Not provided' in emplyeename column to NaN
* Drop the rows having more than 5 missing values
* Find the job title of Albert Pardini
* How much Albert Pardini make(include benefits)
* Display name of the person having the highest basepay
* Find average basepay of all employee per year
* Find avg basepay of employee 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. **head():**

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

1. **tail():**

Description: Displays the last 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. **drop():**

Description: Removes specific rows or columns from the data frame.

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

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

1. **unique():**

Description: Used to finf unique values from a series

1. **len():**

Description: Returns the number of rows in the data frame

1. **pd.to\_numeric( ):**

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

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. **replace():**

Description: Used to replace specific values in a DataFrame or Series with other values.

1. **dropna():**

Description: Used to remove rows or columns from a DataFrame that contain missing values (NaN, null values)

1. **loc[]:**

Description: Indexer is used to access rows and columns by label(s) in a DataFrame**.**

1. **idxmax():**

Description: Used to return the index label of the first occurrence of the maximum value in a Series.

1. **groupby():**

Description: Used to group data in a DataFrame based on one or more columns, and then perform some operation (such as aggregation, transformation, or filtering) on each group separately.