## Artificial Intelligence and Machine Learning

Project Report

Semester-IV (Batch-2022)

**Case Study**: -Employee Salary

[Url:-](file:///C:\Users\spars\Desktop\AIML%20Project\-) https://drive.google.com/file/d/1n8D48Z461Qr6uE-JTvRNL9dj1-H9pxhN/view?usp=sharing

A red and white sign

Description automatically generated with low confidence

**Supervised By: Submitted By:**

Rajeev Thakur Yashika

Roll Number: -2210990976

Group - 14

**Department of Computer Science and Engineering**

## Chitkara University Institute of Engineering & Technology,

## Chitkara University, Punjab

**Description about Case Study: -**

* Read dataset of Employee Salary
* Display Top 10 rows
* Display the Last 10 rows
* Find shape of our dataset
* Getting information about our dataset like Total number of rows, columns, datatype of each column and memory requirement.
* Check null values in the Dataset
* Drop ID, NOTES , AGENCY A=ND STATUS COLUMN
* Total number of jobs titles contain captain
* Display all the employee name from fire department
* Find minimum, maximum and average basepay
* Replace ‘Not Provided’ in employee name column to NaN
* Drop the rows having more than 5 missing values
* Find the job title of Albert Pardini
* How much Albert Pardini make
* Display name of the person having the highest basepay
* Find average basepay of all employee per year
* Find average basepay of all employee per job title
* Find average basepay of all employee having job title accountant
* Find top 5 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. **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. **nunique():**

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

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. **apply():**

Description: Applies a function to transform the values in the data frame.

1. **groupby():**

Description : applying a function, and combining the results

1. **unique():**

Description: used to find the unique values from a series.

1. **iloc():**

Description:  selecting and manipulating data in DataFrames and Series

1. **idxmax():**

Description: Return the row label of the maximum value