## Artificial Intelligence and Machine Learning

Project Report

Semester-IV (Batch-2022)

**Case Study**: - Employee Salaries Dataset

[Url:-](file:///C:\Users\spars\Desktop\AIML%20Project\-) [Salaries Assignment.ipynb](https://drive.google.com/file/d/1TvLx89Se-y-43orhQ6xPpcYe5qM-AMd8/view?usp=sharing)

A red and white sign

Description automatically generated with low confidence

**Supervised By: Submitted By:**

Rajeev Thakur Vanshika

Roll Number: -2210990944

Group - 14

**Department of Computer Science and Engineering**

## Chitkara University Institute of Engineering & Technology,

## Chitkara University, Punjab

**Description about Case Study: -**

* Display Top 10 rows
* Display the Last 10 rows
* Find the shape of the database
* Getting information about our dataset like total no. of rows, total np. of 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(top 5)
* Find the number of unique Job Titles
* Total number of jobs titles contains captain
* Display all the employee Names from fire department
* Find min,max and Avg Basepay
* Replace 'not provided' in employeename 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 avg basepay of all employee per year
* Find avg basepay of employee having job title accountant
* Find top 5 most common jobs

**Library: -**

* Pandas : Resource will be the same in case of pandas (csv, excel and sql database) but in pandas we are processing the data and retrieving statistics output

**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.
2. **groupby() :** Description: It is used to group data based on some condition .
3. **unique() :**

Description: The unique functionin pandas is used to find the unique values from

a series.

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. **value\_counts():**

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