## **Artificial Intelligence and Machine Learning**

Project Report

Semester-IV (Batch-2022)

**Case Study**: - Employees Salaries Datasets

[**Url:-**](about:blank) <https://drive.google.com/file/d/1VhqTPqbIHX5jeXUDsaW05qNGLsO8kwcw/view?usp=sharing>

A red and white sign

Description automatically generated with low confidence

**Supervised By: Submitted By:**

Rajeev Thakur Umarpreet singh Grover

Roll Number - 2210990926

Group - 14

**Department of Computer Science and Engineering**

## **Chitkara University Institute of Engineering & Technology,**

## **Chitkara University, Punjab**

**Description about Case Study: -**

* Read employees salary dataset
* Display Top 10 rows
* Display Last 10 rows
* How many rows and columns are in the dataset
* Display info about dataset
* Check null values in the Dataset
* Drop Id, notes, agency and status columns
* Find occurrence of the employee’s 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
* 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 average BasePay of all Employee per year
* Find average basepay of all employee per job title
* Find average 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.
2. **tail():** Description: Displays the last few rows of the data frame.
3. **head():** Description: Displays the first few rows of the data frame.
4. **shape():** Description: Returns the shape (number of rows, number of columns) of the data frame.
5. **info():** Description: Provides basic information about the data frame, such as column types and missing values.
6. **isnull():** Description: Returns True/False for each value in the data frame, indicating whether the value is missing (NaN) or not.
7. **drop():** Description: Removes specific rows or columns from the data frame.
8. **value\_counts():** Description: Counts the unique values in a specific column of the data frame.
9. **nunique():** Description: Returns the count of unique values in a specific column of the data frame.
10. **contains():** Description: Checks if a specified substring or value is present in a column of the data frame.
11. **Replace():** Replace values given in to replace with value
12. **Dropna():** Returns a new DataFrame object unless the inplace parameter is set to True
13. **max():** Description: Returns the maximum value in a column of the data frame.
14. **min():** Description: Returns the minimum value in a column of the data frame.
15. **mean():** Description: Calculates the mean (average) value of a column in the data frame.