## Artificial Intelligence and Machine Learning

Project Report

Semester-IV (Batch-2022)

**Case Study**: - Employee Salaries Dataset

A red and white sign

Description automatically generated with low confidence

[Url:-](about:blank) <https://drive.google.com/file/d/1XRy_iGit_4RkbbazMTXGjQH28LQNqrYB/view?usp=drive_link>

**Supervised By: Submitted By:**

Rajeev Thakur Uday Malhotra

Roll Number: -2210990919

Group - 14

**Department of Computer Science and Engineering**

## Chitkara University Institute of Engineering & Technology,

## Chitkara University, Punjab

**Description about Case Study: -**

# 1. Display TOP 10 ROWS OF THE DATASET

# 2. CHECK LAST 10 ROWS OF THE DATASET

# 3. FIND SHAPE OF OUR DATASET(NUMBER OF ROWS AND NUMBER OF COLUMNS)

# 4. FIND SHAPE OF OUR DATASET(NUMBER OF ROWS AND NUMBER OF COLUMNS)

5. CHECK NULL VALUES IN THE DATASET

6. DROP ID,NOTES,AGENCY AND STATUS COLUMNS

7. FIND OCCURRENCE OF THE EMPLOYEE NAME(TOP5)

8. FIND THE NUMBER OF UNIQUE JOB TITLES

9. TOTAL NUMBER OF JOBS TITLES CONTAINS CAPTAIN

10. DISPLAY ALL THE EMPLOYEE NAMES FROM FIRE DEPARTMENT

11. FIND MINIMUM,MAXIMUM AND AVERAGE BASE PAY​

# 12. REPLACE 'NOT PROVIDED' IN EMPLOYEENAME COLUMN TO NaN

13.drop the rows having more than 5 missing values​

# 14. find the job title of ALBERT PARDINI

​

# 15. HOW MUCH ALBERT PARDINI MAKE (INCLUDE BENEFITS)?

# 16. DISPLAY NAME OF THE PERSON HAVING THE HIGHEST BASE PAY

1

# 17. FIND AVERAGE BASE PAY OF ALL EMPLOYEE PER YEAR

# 18. FIND AVERAGE BASE PAY OF ALL EMPLOYEE PER JOB TITLE

1

# 19. FIND AVERAGE BASE PAY OF EMPLOYEE HAVING JOB TITLE ACCOUNTANT

# 20. FIND TOP5 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. **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.

1. **apply():**

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

**18. Replace()** : Replace values given in to replace with value