## Artificial Intelligence and Machine Learning

Project Report

Semester-IV (Batch-2022)

**Case Study**: - Employee Salary Dataset

[Url:-](about:blank) <https://drive.google.com/file/d/1nrJRb49Oqj5N-lNgFrEvg2irGKvAVqW4/view?usp=sharing>

A red and white sign

Description automatically generated with low confidence

|  |  |
| --- | --- |
| **Supervised By:**  **Rajeev Thakur** | **Submitted By:**  **Umang Shakya**  **Roll Number: 2210990925**  **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 last 10 rows
* Find shape of dataset
* Getting information about dataset
* Check null values in dataset
* Drop ID, Notes, Agency and Status Columns
* Find occurrence of employee name
* Find number of unique title jobs
* Find number of job titles containing Captain
* Display all the employee names from Fire Department
* Find minimum, maximum and average Basepay
* Replace “Not Provided” in EmployeeName column to NaN
* Drop the rows having more than 5 missing values
* Find Job title of “Albert Pardini”
* How much Albert Pardini makes(including Benefits)
* Display name of person having highest Basepay
* Find average Basepay of all employees per year
* Find average Basepay of all employees 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.

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.