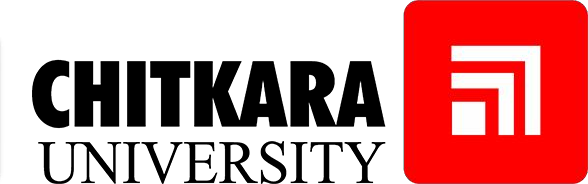
Artificial Intelligence and Machine Learning

Project Report Semester-IV (Batch-2022)

Case Study - Employee Salaries Dataset URL:-

<https://drive.google.com/file/d/1Qp8DnY1NR0OkMmXDpMZcBybJJgfAEjwV/view?usp=drive_link>



**Supervised By: Submitted By:**

Rajeev Bhardwaj Suyash Kumar Sharma

2210991000

**G-13**

**Department of Computer Science and Engineering**

Chitkara University Institute of Engineering & Technology,

**Chitkara University, Punjab**

* Display Top 10 Rows of the Dataset:

How can we view the top 10 rows of our dataset?

* Check Last 10 Rows of the Dataset:

How can we check the last 10 rows of our dataset?

* Find Shape of Our Dataset (Number of Rows and Number of Columns):

Where can we find the shape of our dataset, i.e., the number of rows and columns?

* Getting Information About Our Dataset:

How do we obtain information about our dataset, including total number of rows, total number of columns, datatypes of each column, and memory requirements?

* Check Null Values in the Dataset:

How do we identify and handle null values in our dataset?

* Drop ID, Notes, Agency, and Status Columns:

How can we remove the 'ID', 'Notes', 'Agency', and 'Status' columns from our dataset?

* Find Occurrence of the Employee Name (Top 5):

How do we find the top 5 occurrences of employee names in our dataset?

* Find the Number of Unique Job Titles:

Where can we find the count of unique job titles in our dataset?

* Total Number of Job Titles Containing 'Captain':

How many job titles in our dataset contain the term 'Captain'?

* Display All Employee Names from Fire Department:

How can we display all the employee names associated with the Fire Department?

* Find Minimum, Maximum, and Average BasePay:

How do we determine the minimum, maximum, and average BasePay in our dataset?

* Replace 'Not Provided' in EmployeeName Column to NaN:

How can we replace occurrences of 'Not Provided' in the 'EmployeeName' column with NaN?

* Drop Rows with More Than 5 Missing Values:

How do we drop rows in our dataset that have more than 5 missing values?

* Find the Job Title of Albert Pardini:

Where can we find the job title of the employee named Albert Pardini?

* How Much Does Albert Pardini Make (Include Benefits)?:

How can we determine the total compensation, including benefits, for Albert Pardini?

* Display Name of the Person Having the Highest BasePay:

How can we identify the employee with the highest BasePay and display their name?

* Find Average BasePay of All Employees Per Year:

How do we calculate the average BasePay for all employees per year?

* Find Average BasePay of All Employees Per Job Title:

How can we find the average BasePay for all employees grouped by job title?

* Find Average BasePay of Employees Having Job Title 'Accountant':

How do we determine the average BasePay for employees with the job title 'Accountant'?

* Find Top 5 Most Common Jobs:

Where can we find information on the top 5 most common job titles in our dataset?

# Library:-.

* Pandas

# Methods:-

1. Display Top 10 Rows of the Dataset:

Method: Use the head() function in Python (Pandas) or a similar function in other data analysis tools.

1. Check Last 10 Rows of the Dataset:

Method: Use the tail() function in Python (Pandas) or a similar function in other data analysis tools.

1. Find Shape of Our Dataset (Number of Rows and Number of Columns):

Method: Use the shape attribute in Python (Pandas) or similar functions in other tools.

1. Getting Information About Our Dataset:

Method: Use the info() function in Python (Pandas) or similar functions in other tools.

1. Check Null Values in the Dataset:

Method: Use isnull() or info() function in Python (Pandas) or similar functions in other tools.

1. Drop ID, Notes, Agency, and Status Columns:

Method: Use the drop() function in Python (Pandas) to remove specified columns.

1. Find Occurrence of the Employee Name (Top 5):

Method: Use value\_counts() on the 'EmployeeName' column and select the top 5 occurrences.

1. Find the Number of Unique Job Titles:

Method: Use nunique() or unique() on the 'Job Title' column to count unique job titles.

1. Total Number of Job Titles Containing 'Captain':

Method: Use str.contains() on the 'Job Title' column to filter for entries containing 'Captain', then count.

1. Display All Employee Names from Fire Department:

Method: Filter the dataset for entries where the 'Department' is 'Fire' and display the 'EmployeeName' column.

1. Find Minimum, Maximum, and Average BasePay:

Method: Use min(), max(), and mean() functions on the 'BasePay' column.

1. Replace 'Not Provided' in EmployeeName Column to NaN:

Method: Use replace() to replace occurrences of 'Not Provided' with NaN in the 'EmployeeName' column.

1. Drop Rows with More Than 5 Missing Values:

Method: Use dropna() with a threshold parameter to drop rows with more than 5 missing values.

1. Find the Job Title of Albert Pardini:

Method: Filter the dataset for entries where 'EmployeeName' is 'Albert Pardini' and retrieve the 'Job Title'.

1. How Much Does Albert Pardini Make (Include Benefits)?:

Method: Find the row for 'Albert Pardini' and sum the 'BasePay' and other relevant columns.

1. Display Name of the Person Having the Highest BasePay:

Method: Use idxmax() to find the index of the maximum 'BasePay' and retrieve the corresponding 'EmployeeName'.

1. Find Average BasePay of All Employees Per Year:

Method: Group by the 'Year' column and calculate the mean of 'BasePay' for each group.

1. Find Average BasePay of All Employees Per Job Title:

Method: Group by the 'Job Title' column and calculate the mean of 'BasePay' for each group.

1. Find Average BasePay of Employees Having Job Title 'Accountant':

Method: Filter for entries with 'Job Title' as 'Accountant' and calculate the mean of 'BasePay'.

1. Find Top 5 Most Common Jobs:

Method: Use value\_counts() on the 'Job Title' column and select the top 5 most common jobs.