

Artificial Intelligence and Machine Learning

Project Report

Semester-IV (Batch-2022)

Case Study: -Employee Salary

Url:- https://drive.google.com/file/d/1wD80WiLgzFvIQnMENWxydYId2lpHIzRi/view?usp=drive_link

CHITKARA
UNIVERSITY



view?usp=drive_link

Supervised By:

Rajeev Thakur

Submitted By:

Vrinda

Roll Number: -2210

Group - 14

**Department of Computer Science and Engineering
Chitkara University Institute of Engineering & Technology,
Chitkara University, Punjab**

Description about Case Study: -

- Read dataset of Employee Salary
- Display Top 10 rows
- Display the Last 10 rows
- Find shape of our dataset
- Getting information about our dataset like Total number of rows, columns, datatype of each column and memory requirement.
- Check null values in the Dataset
- Drop ID, NOTES , AGENCY A=ND STATUS COLUMN
- Total number of jobs titles contain captain
- Display all the employee name from fire department
- Find minimum, maximum and average basepay
- Replace 'Not Provided' in employee name column to NaN
- Drop the rows having more than 5 missing values
- Find the job title of Albert Pardini
- How much Albert Pardini make
- 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 all employee having job title accountant
- Find top 5 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. sum():

Description: Calculates the sum of values in each column of the data frame.

8. drop():

Description: Removes specific rows or columns from the data frame.

9. value_counts():

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

10. nunique():

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

11. contains():

Description: Checks if a specified substring or value is present in a column of the data frame.

12. max():

Description: Returns the maximum value in a column of the data frame.

13. min():

Description: Returns the minimum value in a column of the data frame.

14. mean():

Description: Calculates the mean (average) value of a column in the data frame.

15. len():

Description: Returns the number of rows in the data frame

16. apply():

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

17. groupby():

Description : applying a function, and combining the results

18. unique():

Description: used to find the unique values from a series.

19. iloc():

Description: selecting and manipulating data in DataFrames and Series

20. idxmax():

Description: Return the row label of the maximum value