**Artificial Intelligence and Machine Learning**

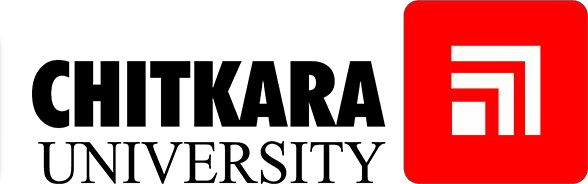
Project Report

Semester-IV (Batch-2022)

Case Study - Adult Dataset

URL:

<https://drive.google.com/file/d/1_T16APdj1seOcvqEu7kMSRj37X4LCPDu/view?usp=drive_link>



|  |  |
| --- | --- |
| **Supervised By:** | **Submitted By:** |
| Rajeev Bhardwaj | Shreya Tewari |

2210990839

# G-13

**Department of Computer Science and Engineering**

Chitkara University Institute of Engineering & Technology,

**Chitkara University, Punjab**

**Descrip on about Case Study:-**

1. Display Top 10 Rows:
2. How can we view the top 10 rows of our dataset?
3. Display Last 10 Rows:
4. How can we check the last 10 rows of our dataset?
5. Find Shape of Our Dataset (Number of Rows and Number of Columns):
6. Where can we find the shape of our dataset, i.e., the number of rows and columns?
7. Gettinng Informa on About Our Dataset:
8. How do we obtain information about our dataset, including the total number of rows, total number of columns, data type of each column, and memory requirements?
9. Fetch Random Samples from Dataset (50%):
10. How do we randomly sample 50% of the data from our dataset?
11. Check Null Values in Dataset:
12. How do we identify and handle null values in our dataset?
13. Perform Data Cleaning (Replace '?' with NaN) and Plot in Graph with Seaborn Library: How can we replace occurrences of '?' with NaN in our dataset as part of data cleaning, and then plot the results using the Seaborn library?
14. Drop All Rows Having Missing Values:
15. How do we remove all rows from our dataset that contain missing values?
16. Check for Duplicate Data and Drop Them:
17. How can we identify and drop duplicate rows in our dataset?
18. Get Overall Sta s cs About the DataFrame:
19. How do we obtain overall stastical informa on about the en re DataFrame?
20. Bivariate Analysis on Graph:
21. How can we perform bivariate analysis and visualize relationships between two variables using graphs?
22. Replace Salary Values ['<=50k', '>50k'] with 0 and 1:
23. How do we convert salary values '<=50k' and '>50k' to numerical values 0 and 1 in our dataset?
24. Which Workclass is Getting the Highest Salary?:
25. How can we determine which workclass has the highest average salary in our dataset?
26. Who has a Be er Chance to Get Salary >50k, Male or Female?:
27. How do we analyze and compare the chances of getting a salary >50k between male and female individuals?
28. Convert Workclass Column Datatype to Category Data:
29. How can we convert the data type of the 'Workclass' column to the category data type?

**Library:-.**

* Pandas
* Seaborn
* Matplotlib

**Methods:-**

1. Display Top 10 Rows:

Method: Use the head() func on.

1. Display Last 10 Rows:

Method: Use the tail() func on.

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

Method: Use the shape a ribute.

1. Getting Informa on About Our Dataset:

Method: Use the info() func on.

1. Fetch Random Samples from Dataset (50%):

Method: Use the sample() func on.

1. Check Null Values in Dataset:

Method: Use the isnull() func on.

1. Perform Data Cleaning (Replace ‘?’ with NaN) and Plot in Graph with Seaborn Library:

Method: Use the replace() func on for data cleaning and Seaborn library for plo ng.

1. Drop All Rows Having Missing Values:

Method: Use the dropna() func on.

1. Check for Duplicate Data and Drop Them:

Method: Use the duplicated() func on and drop\_duplicates() func on.

1. Get Overall Statistics About the DataFrame:

Method: Use the describe() func on.

1. Bivariate Analysis on Graph:

Method: Use the Seaborn or Matplotlib library for plo ng bivariate analysis.

1. Replace Salary Values ['<=50k', '>50k'] with 0 and 1:

Method: Use the replace() func on.

1. Which Workclass is Getting the Highest Salary?:

Method: Use the groupby() func on and calculate the mean.

1. Who has a Be er Chance to Get Salary >50k, Male or Female?:

Method: Use the groupby() func on and calculate the percentage.

1. Convert Workclass Column Datatype to Category Data:

Method: Use the astype() func on.