## **Artificial Intelligence and**

## **Machine Learning**

Project Report

Semester-IV (Batch-2022)

**Case Study**: - Amazon Purchases Dataset

**Url:-**

[**https://drive.google.com/file/d/1UceW0qGLkMMFLoq60xes3Q5ag65PrxGK/view?usp=sharing**](https://drive.google.com/file/d/1UceW0qGLkMMFLoq60xes3Q5ag65PrxGK/view?usp=sharing)

A red and white sign

Description automatically generated with low confidence

**Supervised By: Submitted By:**

Rajeev Thakur Name - Sneha Miglani

Roll Number -2210990858

Group - G-13

**Department of Computer Science and Engineering**

## **Chitkara University Institute of Engineering & Technology,**

## **Chitkara University, Punjab**

**Description about Case Study: -**

* Read dataset Amazon Purchases
* Display Top 10 rows
* Display the Last 10 rows
* Check the datatype of Each column
* Check null values in the Dataset
* How many rows and columns are in our dataset
* Highest and lowest purchase price in the dataset
* Average purchase price
* How many people have French 'fr' as their Language
* The job title contains engineer
* Find the email of the person with the following IP address: 132.207.160.22
* How many people have Mastercard as their Credit Card Provider and Purchased above 50?
* Find the Email of the person with the following Credit Card number:4105595335494659
* How Many People Purchase During the AM and How Many People Purchase During PM?
* How Many People Have a Credit Card That Expires In 2020
* Top 5 Most Popular Email Providers (Gmail.com, Yahoo.com, etc.)

**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.