## Artificial Intelligence

## and

## Machine Learning

Report

Semester-IV (Batch-2022)

Case Study: Ecommerce dataset

<https://drive.google.com/file/d/1y92-a7vvKeyEhxanv3-t4Z_ZPOFdxpvw/view?usp=drive_link>

## A red and white sign Description automatically generated with low confidence

**Supervised By: Submitted By:**

Faculty Name Name : Simranpreet Kaur

Roll Number :2210990854

Group : G13(A)

**Department of Computer Science and Engineering**

Chitkara University Institute of Engineering & Technology,

Chitkara University, Punjab

**DESCRIPTION**

**Libraries:**

**Pandas:** Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data.

The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008.

Pandas allows us to analyze big data and make conclusions based on statistical theories.

Pandas can clean messy data sets, and make them readable and relevant.

**Questions:**

1. display top 10 rows
2. Check last 10 rows of the datset
3. **Check Datatype of each column**
4. **Check Null values in the dataset**
5. **How many Rows and Columns are there in our dataset?**
6. **Highest and lowest purchase prices**
7. **Average Purchase Price**
8. **How many people have french 'fr' as their Language**
9. **Job Title contains Engineer**
10. **find email of the person with the following IP address:132.207.160.22**
11. **How many People have mastercard as their credit card provider and made a purchase above 50**
12. **find email of the person with the following credit card number 4664825258997302**
13. **How many People Purchase during the AM and How many people purchase during PM?**
14. **How many people have a credit card that expires in 2020?**
15. **top 5 most popular email providers(e.g gmail.com,yahoo.com,etc)**

**Methods:**

1. pandas.read\_csv(): **p**andas.read\_csv() is a function in the pandas library that reads a comma-separated values (csv) file into a Dataframe. It has many parameters that allow you to customize how the file is read, such as sep, header, index\_col, usecols, dtype, skiprows, nrows, na\_values, and more.
2. Data.head(): data.head(10) is a method that returns the first 10 rows of the DataFrame data. It is useful for getting a quick overview of the data or checking the format and column names. You can change the number of rows by passing a different value to the method, such as data.head(5) or data.head(20)
3. **data.tail(10) :** data.tail(10) is a method that returns the last 10 rows of the DataFrame data. It is useful for getting a quick overview of the data or checking the format and column names. You can change the number of rows by passing a different value to the method, such as data.tail(5) or data.tail(20)
4. **data.isnull().sum():** data.isnull().sum() is a method that returns the number of missing values in each column of the DataFrame data. It is useful for detecting and handling missing data in your analysis. It returns a Series with the column names and the count of null values.
5. Len(): The len() function is a built-in function in Python that returns the number of items in an object. It can be used with sequences, such as strings, lists, tuples, and ranges, or with collections, such as dictionaries, sets, and frozen sets.
6. **max():** The max() function is a built-in function in Python that returns the maximum value among the given arguments.
7. **min():** The min() function is a built-in function in Python that returns the smallest value among the given arguments or the smallest item in an iterable object.
8. **mean():**The mean() function is a built-in function in Python that returns the average of a list of numbers. It takes a list as an argument and returns a single number that is the sum of all the numbers in the list divided by the length of the list.
9. **count():** the count() function can be used with strings, lists, tuples, and other iterable objects. It takes one argument, which is the value to search for, and returns the number of times it appears in the object.
10. str.contains: The str.contains() function is used to test if a pattern or a regular expression is contained within a string of a Series or an Index. It returns a boolean Series or Index based on whether the given pattern or regex is found in each element of the Series or Index.
11. **value\_counts():** The value\_counts() function is used to get a Series containing counts of unique values in a pandas Series or DataFrame. It can be used to analyze the frequency distribution of categorical or numerical data.
12. Split(): The value\_counts() function is used to get a Series containing counts of unique values in a pandas Series or DataFrame. It can be used to analyze the frequency distribution of categorical or numerical data.
13. Apply(): The apply() function is a common function in many programming languages that allows you to call a function with a given context and arguments.
14. Append(): The apply() function is a common function in many programming languages that allows you to call a function with a given context and arguments.