Batch: C2   Roll No.:   16010122267
Experiment 01

Title: Data Collection and finalizing dataset from problem domain

Objective:

1. To learn how to collect the dataset
2. To learn sources of dataset
3. To assess the dataset based on Metrics to Measure Data Quality
4. To finalize the features of dataset

Course Outcome:

CO1: Learn how to locate and download datasets, extract insights from that data and present their findings in a variety of different formats.

Books/ Journals/ Websites referred:

www.kaggle.com
www.geeksforgeeks.org
https://numpy.org/doc/

Resources used:

(Dataset link)

Theory:

First, determine the specific data you require. Next, select suitable methods or tools to gather this information effectively. Thoroughly validate the collected data to ensure its accuracy. Address any missing or incomplete data points appropriately.

Following points should be written by students

- Problem domain (Healthcare, Ecommerce, Education, Finance, agriculture etc.)
- Motivation for the selected Domain
- Brain stormed features of Dataset (Based on Domain Selected)
- Search for dataset
- Justification for choosing above dataset
- Source of dataset (Link Needs to be given)
- Sample of Finalized dataset (First 5 Records)
- Data Dictionary
- Column wise summary

1. **Domain:** Ecommerce
2. **Motivation:** To analyze the profitability of online sales.
3. **Dataset features:** Sales channels, SKU codes, product details, MRPs customer payment methods, dates, fulfilment, quantities, currency.
4. **Dataset search:** Explored ecommerce databases and platforms like Kaggle.
5. **Justification:** Relevant for analyzing ecommerce sales strategies and profitability.
6. **Dataset Source: https://www.kaggle.com/datasets/thedevastator/unlock-profits-with-e-commerce-sales-data?datasetId=2699295&sortBy=voteCount**
7. **Data Dictionary:** Clarification of ecommerce terminology and variables.
8. **Column-wise summary**: Sales channels, product attributes, pricing, transaction specifics for comprehensive ecommerce insights.

Conclusion (Students should write in their own words):

In conclusion, we learnt how to choose appropriate datasets from the appropriate sources and brainstorms ideas to obtain valuable insights from the raw data.

Post Lab Question:

1. **Explain Role of Data in the Application Design.**
2. **Write different types of Data with Example.**

Ans1:
Data plays a big role in how an app works. It helps decide how things are organized and how users interact with the app. Data also helps make the app smart, like understanding what users want. It's like the app's brain, making it run well, keeping info safe, and helping people use it easily.

Ans2:

1. **Qualitative Data:** Represents non-numeric information like audio, images, symbols, or text. It's useful for understanding perceptions and customer preferences in market research and strategy design. Example: Customer reviews, social media posts.
2. **Nominal Data:** Consists of categories without inherent order or meaning, such as color, gender, or product names. It's used for labeling and classification. Example: Eye colors (blue, brown, green).
3. **Ordinal Data:** Exhibits natural order but lacks consistent intervals, suitable for ranking and sequences. It can't be used for arithmetic operations. Example: Education levels (elementary, middle school, high school, college).
4. **Quantitative Data:** Can be expressed in numbers and used for statistical analysis. It includes measurable attributes like price, weight, height, or temperature. Example: Ages of students in a class.
5. **Discrete Data:** Comprises whole numbers that can't be divided into fractions. It's countable and finite, suitable for items like counts or scores. Example: Number of books on a shelf.
6. **Continuous Data:** Can take any value within a range, including fractional numbers. It's also countable and finite, and often used for measurements. Example: Temperature readings, height of individuals.

Understanding these data types is essential for appropriate data analysis, visualization, and decision-making in various fields.