NAME ⇒ Kamal deep

Roll no ⇒ 2501660006

Course name ⇒ BCA cyber security

Topic ⇒ Fundamentals of Data collection and cleaning

Signature ⇒

Assignment ⇒ 2

1) Learning objectives
2) Identify different types and sources of data.
2) Data can be categorized by its sources (Internal, External, Primary or secondary) and by its type (quantitative or qualitative). Internal sources are from with in an organization, while external sources are from outside.

Types of Data :-

(i) Quantitive Data : Numerical data can be measured and calculated, such as sales figures or employee headcount

(ii) Qualitative Data : Non-numerical data that describes qualities or characteristics, such as customer feed back, observation, or interview transcripts.

Sources of Data

(i) Internal sources : Data that originates and is collected from with in an organization, likes sales records, company reports or data from public database.

Ply data collection methods like surveys, interviews and observations.

service interviews and observations are qualitative data collection methods used to gather in depth insights Interviews involve direct conversation to collect verbal responses, while observations involves watching and recording behaviour in a natural setting. Both are used to understand experiences perceptions, and behaviours, often in the context of social science, research or service quality analysis.

service interviews

- What it is A conversation between a researcher and one or more participants to gather detailed information through oral questions and responses.
- How it works: Researchers ask open-ended question to understand a person's beliefs, attitude and experiences

ex ⇒ Asking customers about their experience with a new service by having them describe their feelings and the steps they took to use it.

3) Perform basic data cleaning including handling missing values and duplicates

3) Basic Data cleaning involves identifying and correcting errors, specifically by handling missing values (through deletion or imputation) and removing duplicates (by deleting redundant entries). A systematic process first assesses Data quality, then removes duplicates and irrelevant Data, Fixes structural errors, handles missing Data, and Finally validates the cleaned data.

sure data integrity and ethical handling during processing
Implement technical and organization measures such as
access controls, encryption, validation, and regular
audits, while also adhering to ~~co~~ ethical principles
like transparency, privacy, ~~min~~ minimizing data collection.

Assignments tasks

Task 1 ÷ Design a short survey (5-10 questions) to collect
Feed back about student satisfaction at your college

Your feedback is invaluable in helping us improve
our certification programs. this survey will take
approximately 3-5 minutes to complete. All responses
are confidential

Section 1: General Program Experience.

1) Which certification program did you complete or are you
currently enrolled in?

- [list Program A]
- [list Program B]
- [list Program C]

Other (Please specify)

2) How satisfied were you with the overall quality of the
~~co~~ certification program?

- very satisfied
- satisfied
- Neutral
- pissatisfied
- very dissatisfied

5 effectively do you feel ~ - - - - - - - field?

- Extremely effective
- very effective
- Moderately effective
- Slightly effective

4) How would you rate the instructor's knowledge of the subject matter?

- Excellent
- Good
- Average
- Poor
- very Poor

5) How satisfied are you with the accessibility and support provided by Faculty and staff?

- very satisfied
- Satisfied
- Neutral
- Dissatisfied
- very Dissatisfied

6) Where the online resources, library materials, ~ - - ~ ?

- Yes completely
- Mostly
- Neutral
- Not really
- No, not at all

Task 2: Collect 10-15 responses (real or simulated) and store them in a spreadsheet.

You can neither use an online form (like Google Forms or Microsoft Forms) to collect real responses automatically or manually enter simulated data into an existing spreadsheet file.

method 1: using online Forms
(Recommended for real responses)

Steps using Google Forms:

1. Create the Form
2. Link to a spreadsheet.
3. Share the Form
4. Collect Responses
5. Query the spreadsheet

Using Microsoft Forms:

1. Create the Forms
2. Link to Excel
3. Share the Form.

method 2: manually entering simulated Data

1) Open a spreadsheet
2) Setup Headers
3) enter simulated Responses
4) Save the File

Task 3: Identify and Fix issues such as missing data, duplicates entries, or inconsistent Formatting.

A systematic process of data cleaning (or data cleansing) is required. This involves profiling that data to detect errors and then applying various techniques to correct them, often using tools like spreadsheets, SQL, or programming libraries.

1: Identify the issues (Data Profiling)

missing Data

Duplicate Entries

- Inconsistent Formatting

Duplicate entries

- Remove duplicates
- Consolidate Records

Step2. Fix the Issues

missing Data

- Remove Records
- Impute values
- Predictive Modeling

Tools and Best Practices

- Spread sheets
- Programming language
- Specialized software

Task 4: write a short note on ethical conviderations while collecting personal data.

Ethical data collection requires prioritizing informed consent, privacy, and transparency. Key consideration include clearly explaining how data will be used, securing the data against breaches, only collecting necessary information and using it for the purpose it was originally intended, while also being accountable for any misuse.

Informed consent.

voluntary and clear

No coercion.

Right to withdraw

Transparency and accountability

- openness
- Accountability

Privacy and security

Protect data

minimize risk

Confidentiality

Responsible Use

- Purpose limitation
- Avoid bias.