Assignment-1

Subject: Web Analytics and Business Intelligence

Student Name: Haripriya Naga Neelima Yalamakayala

Student ID : 8939330

Date: 22nd October,2024

Task1- Data Cleaning

Step 1: Data is imported to the python program: Loaded the dataset into a Pandas Data Frame. Initial few rows of the dataset are displayed.

A screenshot of a computer

Description automatically generated

Step 2: Remove Irrelevant Columns: Removed irrelevant columns from the dataset. Displaying the remaining relevant fields.

A screenshot of a computer

Description automatically generated

Step3: Handle Missing Values: Checked for missing values and dropped rows with missing required fields. Updated dataset shape and missing values count.

A screenshot of a computer screen

Description automatically generated

Step4: Ensure Unique Identifiers: Identified and removed duplicate records based on 'Object ID'. Updated dataset shape after removing duplicates.

A screenshot of a computer screen

Description automatically generated

Step5: Ensure Consistent Data Formats: Converted 'Object Begin Date' and 'Object End Date' fields into consistent datetime format.

A screenshot of a computer

Description automatically generated

Step6: Validate Logical Consistency: Validated the logical consistency of date fields. Removed rows where 'Object End Date' was before 'Object Begin Date'.

A screenshot of a computer program

Description automatically generated

Step7: Export the Cleaned Dataset: Final dataset cleaned and exported to 'cleaned\_dataset.csv' for further analysis.

A screenshot of a computer program

Description automatically generated

Task 2 - BI and Data related Questions

1. What are the 3 Vs of Data and explain each one in detail?

Ans:

The three Vs of data stand for important characteristics that describe large data and are necessary to comprehend its complexity:

* Volume: Volume defines the quantity of data produced and kept. Due to constant generation from several sources, including social media, corporate transactions, IoT sensors, etc., the amount of data in today's digital world is huge. For instance, businesses like Facebook and Google gather terabytes (TB) or petabytes (PB) of data every day.
* Velocity: Velocity describes how quickly data is created, processed, and examined. Businesses may obtain insights and make choices more quickly when data is handled at a faster pace. Processing data in real-time can be necessary, particularly when it comes to stock market or Internet of Things data.
* Variety: Variety refers different kinds of data that are created. A variety of forms are available for data, including semi-structured (XML, JSON), structured (databases), and unstructured (video, audio, photos, social media posts). Managing different formats is essential for precise analysis.

1. List capabilities of Business Intelligence systems.

Ans:

* Data integration: BI systems are able to combine information from a variety of different sources, including spreadsheets, databases, ERP and CRM platforms, and external data streams (such as social media and market trends). BI solutions enable businesses to produce a single version of the truth for improved decision-making by combining data from several systems into a single view.
* Data Warehousing: To store large amounts of recent and historical data, BI systems frequently make use of data warehousing technologies. This makes analysis and querying more effective. In order to facilitate its retrieval for reporting and analytics, data is frequently cleansed, processed, and arranged.
* Data analytics: BI systems include advanced analytical features like machine learning, statistical analysis, data mining, and predictive analytics. In order to support strategic decision-making, these methods assist businesses in finding patterns, predicting future trends, and extracting insights from past data.
* Data Visualization: Data Visualization is basically converting complicated data into understandable visual representations is one of BI systems' most potent features. Users can quickly gain insights and understand massive datasets in an easily readable format with the use of data visualization. It helps uncover patterns, correlations, and outliers that could otherwise be obscured in unprocessed data.
* OLAP (Online Analytical Processing): OLAP tools, which enable multi-dimensional data processing, are frequently found in BI systems. With OLAP, users may change data dynamically for a more thorough analysis and slice and dice data along several dimensions (such as time, geography, and product).
* Data Governance and Security: In order to ensure data integrity, compliance, and security, BI solutions come with strong data governance features. They control access to various data types, guaranteeing that private data is protected and that only authorized individuals are able to see or modify particular datasets.

1. Different types of data with example for each type.

Ans:

Data can be divided into many categories according on its structure and format. Among the most widely used categories of data are:

* Structured Data:

Structured data is arranged according to a predetermined schema and is usually kept in tables with rows and columns.

Example: Employee records in an HR system; a customer database with fields such as Customer ID, Name, and Email.

* Unstructured Data:

Unstructured Data is hard to organize because it lacks a set format or organization.

Examples include emails, pictures, movies, and posts on social media.

* Semi-structured Data:

Semi-structured Data has certain organizational characteristics, such tags or markers, but doesn't strictly adhere to a schema.

Examples include log files, JSON files, and XML files.

1. Define data visualization.

Ans:

Data visualization is the process of representing information and data graphically using maps, graphs, charts, and other visual aids. By visually representing data, it facilitates the understanding of patterns, trends, and insights and makes it simpler for stakeholders to base choices on the information.

1. What is a KPI and provide an example (apart from one that has been already provided within the lecture)

Ans:

A measured figure that shows how well a company is accomplishing its main goals is called a KPI (Key Performance Indicator). KPIs are used to assess how well a department, project, or organization is doing at achieving particular goals.  
  
Example: Customer Retention Rate: This KPI shows how loyal and satisfied customers are by calculating the percentage of customers a business keeps over a specific time period.

1. What is a BI system?

Ans:

A technology-driven method for assessing data and providing managers, executives, and other end users with useful information that can help in decision-making is known as a business intelligence (BI) system. To provide insights, BI systems combine analytics, data mining, data visualization, and data warehousing.

1. What are the 5 C's of Data for data preparation and the purpose of each?

Ans:

The 5 C's of Data that are essential steps in preparing data for analysis:

* + Clean: Make sure there are no mistakes, inconsistencies, or unnecessary information in the data.
  + Complete: Verify that all important data points are present and that the information is thorough.
  + Correct: Confirm that the information is true and in line with actual facts.
  + Consistent: Verify that the data is input and formatted consistently throughout the dataset (e.g., date formats).
  + Current: Verify that the information is correct and relevant to the analysis. Inaccurate conclusions may result from using outdated data.

1. What are some Key Success Factors of a Successful BI Program and explain each factor?

Ans:

Key Success Factors of a Successful BI Program:

Strong Executive Support:

A effective business intelligence program requires the support of upper management. The relevance of data-driven decision-making is made clear when CEOs actively support BI efforts. Additionally, this aids in obtaining the required capital, assets, and dedication from all organizational levels. In addition to ensuring that the BI program is in line with overall business strategy, executive assistance aids in overcoming resistance to change.

Clearly Defined Business Goals and KPIs:

A BI program needs to be in line with particular corporate goals. Clearly defining the organization's goals is crucial before BI technologies are implemented. Having clearly defined Key Performance Indicators (KPIs) guarantees that the BI system is tracking the appropriate metrics to provide value, regardless of the objective—improving customer satisfaction, supply chain optimization, or sales performance.

Data Governance and Quality Management:

It is essential to guarantee the security, consistency, and quality of data. Decision-making can be delayed and incorrect insights can result from poor data quality. To guarantee that data is correct, dependable, and current, a successful BI program needs robust data governance procedures. This entails setting up procedures for data entry, cleansing, and testing in addition to making sure that data complies with security and privacy laws.

**User Adoption and Training**:

+If consumers lack the necessary skills, even the best BI tools are useless. A BI program's success depends on its investment in user assistance and training. To make wise judgments, end users must be at ease using the tools and know how to extract and evaluate the data. A user-friendly interface and ongoing training are essential for promoting adoption across departments.

**Cross-Department Collaboration**:

BI programs often cut across multiple departments (finance, marketing, HR, etc.), so collaboration is vital. All stakeholders must contribute and work together for a BI activity to be effective. Involving several departments guarantees that various teams can benefit from BI insights and that data needs are appropriately met. A culture of data sharing and collaborative intelligence is promoted by this partnership.

**Scalability and Flexibility of BI Tools**:

The BI tools that are used should be flexible enough to accommodate new technologies and evolving business requirements, as well as scalable enough to expand with the company. Additionally, because the corporate environment is always changing, flexibility is crucial. A good business intelligence program makes use of solutions that facilitate seamless integration with current systems and are able to manage growing data quantities and complexity over time.

**Agility and Continuous Improvement**:

A successful BI program is a continuous process rather than a one-time implementation. In order to handle shifts in market conditions, corporate needs, and technical advancements, the BI approach should be flexible and agile. The BI system, tools, and procedures should be reviewed and updated on a regular basis to guarantee that the program continues to offer useful and timely insights.