Eng: Huda Hemdan

Data Analysis and Visualization Project

Overview: This project involves extracting data from a website, preprocessing the data, and then analyzing and visualizing it using various tools such as Excel, Power BI, and Tableau.

Goal: The goal of this project is to demonstrate the ability to extract, preprocess, and analyze data from a website, and to create interactive and informative visualizations using various tools. The project aims to provide a comprehensive understanding of the data and to identify patterns and trends that can inform business decisions or answer research questions.

Step 1: Web Scraping

- Choose a website that has data you're interested in analyzing (e.g., a news website, an e-commerce website, etc.).
- Use a web scraping tool or library (e.g., BeautifulSoup, Scrapy, Selenium) to extract the data from the website.
- Store the extracted data in a CSV or JSON file.

Step 2: Data Preprocessing

- Import the extracted data into a Python script or a data preprocessing tool (e.g., Pandas, NumPy).
- Clean and preprocess the data by:
    - Handling missing values
    - Removing duplicates
    - Converting data types
    - Performing data normalization and feature scaling
- Store the preprocessed data in a new CSV or JSON file.

Step 3: Data Storage in MySQL

- Design a database schema to store the preprocessed data.
- Create a MySQL database and tables to store the data.
- Use SQL queries to insert the preprocessed data into the database.

Step 4: Data Analysis in Excel

- Import the preprocessed data into Excel.
- Use Excel's built-in functions and formulas to:
    - Calculate summary statistics (e.g., mean, median, mode)
    - Create charts and graphs to visualize the data
    - Perform data filtering and sorting
    - Identify insights and trends in the data

Step 5: Data Visualization in Power BI

- Import the preprocessed data into Power BI.
- Use Power BI's visualization tools to:
    - Create interactive dashboards and reports
    - Visualize the data using charts, graphs, and maps
    - Perform data filtering and sorting
    - Identify insights and trends in the data

Step 6: Data Visualization in Tableau

- Import the preprocessed data into Tableau.
- Use Tableau's visualization tools to:
    - Create interactive dashboards and reports
    - Visualize the data using charts, graphs, and maps

- Perform data filtering and sorting
- Identify insights and trends in the data

Example Project: Analyzing Movie Data

- Website: IMDB (Internet Movie Database)
- Data to extract: Movie titles, release years, genres, ratings, and cast lists
- Preprocessing: Handle missing values, remove duplicates, convert data types
- Excel analysis: Calculate summary statistics, create charts and graphs to visualize the data
- Power BI visualization: Create an interactive dashboard to visualize the data by genre, release year, and rating
- Tableau visualization: Create an interactive dashboard to visualize the data by cast member, movie title, and genre

Deliverables:

- Written Report: A written report summarizing the insights and trends found in the data.
- PowerPoint Presentation: A PowerPoint presentation showcasing the visualizations and insights.
- Excel File: The Excel file used to create the visualizations and perform data analysis.
- Power BI File: The Power BI file used to create the interactive dashboards and reports.
- Tableau File: The Tableau file used to create the interactive dashboards and reports.
- Python Script: The Python script used to preprocess the data and extract insights.
- MySQL Database Schema: The MySQL database schema used to store the preprocessed data.