Lab 2: Data Preparation Exercise

This exercise is to help you practice the following objectives
- Prepare data for analysis and reporting
- Extract, clean, and transform data

Examine the attached data files. The first file contains data on movies and TV shows scrapped from IMDB website. This dataset has issues around inconsistent variable names for example in the year columns, null values, non-alpha numeric variables etc. The second dataset is a live data set on average salaries of managers in different domains. The data input for this dataset is free text and therefore has issues with consistency and outliers.

Apply what you have learned to load the data correctly to tableau prep and generate and execute a cleaning workflow.

Once completed, take a screenshot of the workflow, and do a two-page lab report of the cleaning steps and rationale behind it.

Submission

Submit the lab report in PDF format. This report should include required parts with screenshots, labels, and descriptions. All screens must be clear, original, and show the complete screen.

Grading

The lab is assessed based on:

- Completion of required tasks.
- Satisfaction of submission requirements.
- The quality of lab report, including completeness and explanation of screenshots.

Rubric:

Score	Summary	Rating Description
10	Outstanding work; beyond expectation.	Correctly developed all workflows and loaded data required. Satisfied all submission requirements. Excellent report with clear original screenshots and explanation.
8-9	Good work; meet expectations.	Correctly loaded data required with minor issues. Satisfied most submission requirements. Good report with clear original screenshots; explanation may need improvement.
6-7	Adequate work; need improvement.	Loaded data but with major issues or different from the tutorial. Satisfied most submission requirements. Too simple report without labels or explanations. screenshots, or unclear/unqualified screenshots.
<6	Lack of effort.	Unfinished steps in the workflow. Did not follow submission requirements. Poor report with unclear, partial or edited screenshots. No explanations at all.

Other Resources -Power BI

1. Getting started with Power Query

    a) https://docs.microsoft.com/en-us/power-query/power-query-what-is-power-query
    b) Using the query editor: https://docs.microsoft.com/en-us/power-bi/desktop-query-overview
    c) Common query tasks in Power BI Desktop: https://docs.microsoft.com/en-us/power-bi/desktop-common-query-tasks
    d) https://docs.microsoft.com/en-us/learn/modules/get-data-power-bi/

2. Getting data from various sources: https://docs.microsoft.com/en-us/power-bi/desktop-data-sources

    a) CSV: https://docs.microsoft.com/en-us/power-bi/desktop-connect-csv
    b) Excel: https://docs.microsoft.com/en-us/power-bi/desktop-connect-excel
    c) Web page:
        o https://docs.microsoft.com/en-us/power-bi/desktop-connect-to-web
        o https://docs.microsoft.com/en-us/power-bi/desktop-connect-to-data
        o https://docs.microsoft.com/en-us/power-bi/desktop-connect-to-web-by-example

live connection? Tutorial getting data from AW data mart
https://www.biinsight.com/connect-live-power-bi-to-ssas-multidimensional/

3. Data cleaning and transformation

    a) Data cleaning: https://www.digitalvidya.com/blog/data-cleaning-features-in-power-bi/
    b) Shape and combine data: https://docs.microsoft.com/en-us/power-bi/desktop-shape-and-combine-data
    c) Calculated column: https://docs.microsoft.com/en-us/power-bi/desktop-tutorial-create-calculated-columns
    d) Data types: https://docs.microsoft.com/en-us/power-bi/desktop-data-types
    e) String functions: https://www.tutorialgateway.org/power-bi-dax-string-functions/
    f) https://www.tutorialgateway.org/how-to-split-columns-in-power-bi/

4. Some more extensive exercises

    a) https://docs.microsoft.com/en-us/power-bi/desktop-tutorial-analyzing-sales-data-from-excel-and-an-odata-feed
    b) https://docs.microsoft.com/en-us/power-bi/desktop-tutorial-importing-and-analyzing-data-from-a-web-page
    c) WiseOwlTutorials: https://www.youtube.com/watch?v=cKEAf78qLFM&list=PLNIs-AWhQzcnSmNbeUpQrKuiqGplNB2hh&index=6 – complete four parts 3.1 to 3.4 (see all tutorials in the playlist). For part 3.2 and 3.4, once completed, take a screenshot of the relationship view (clearly showing all tables and relationships), and a screenshot of data view for each table (clearly showing data values). Include this part in the lab report.

Get data from sql server …

Power Query - Overview and Learning

- https://support.office.com/en-us/article/Power-Query-Overview-and-Learning-ed614c81-4b00-4291-bd3a-55d80767f81d
- https://support.office.com/en-us/article/Power-Query-101-008b3f46-5b14-4f8b-9a07-d3da689091b5

Power query with M language in Excel

https://support.office.com/en-us/article/Introduction-to-Microsoft-Power-Query-for-Excel-6e92e2f4-2079-4e1f-bad5-89f6269cd605