JMAN – Lenskart Data Pipeline and Dashboard

**Problem Definition :-**

***Web Scraping, Data Transformation, and Visualization of Lenskart Product Information***

1. **Extraction**

In this step we automate the process of scraping product information from the Lenskart website's eyeglasses page using Selenium for web automation and BeautifulSoup for HTML parsing. The data collected includes product ids, product names, ratings, number of reviews, sizes, and prices, which is then saved into a CSV file.

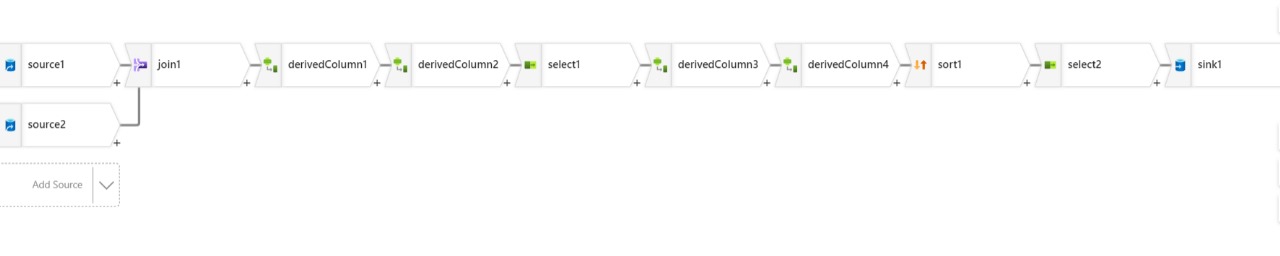
The script performs the following main steps:

1. **Initialize WebDriver**: Set up the Selenium WebDriver for Microsoft Edge.
2. **Navigate to URL**: Open the target webpage.
3. **Scroll to Bottom**: Automatically scroll to the bottom of the page to load all products.
4. **Extract HTML Content**: Retrieve the HTML content of the fully loaded page.
5. **Parse HTML**: Use BeautifulSoup to parse the HTML and extract product details.
6. **Store Data**: Store the extracted data in a pandas DataFrame.
7. **Save to CSV**: Save the DataFrame to a CSV file.
8. **Close WebDriver**: Close the Selenium WebDriver.
9. **Transformation**

After scraping and storing product data in a CSV file, the next step involved transforming the data to ensure consistency and proper formatting. This included changing the data format of certain fields and rearranging the fields for better organization. The transformations were performed using dataflows.

**Objectives**

1. **Format Standardization**: Ensure all fields have a consistent format suitable for analysis.
2. **Field Arrangement**: Reorganize the fields for better readability and usability.



1. **Power BI**

After transforming the data, the next step was to load it into Power BI for visualization and analysis. Power BI automatically created a data model that relates the tables, enabling efficient data exploration and reporting.

After loading and modeling the data in Power BI, the next phase involved learning about different visualization techniques and creating measures to enhance data analysis. This phase also included creating specific visualizations, such as tracking the quantity sold per year.

**Objectives**

1. **Load Data into Power BI**: Import the transformed CSV data into Power BI
2. **Automatic Data Modeling**: Leverage Power BI's automatic data modeling to establish relationships between tables.
3. **Data Visualization**: Create interactive visualizations to analyze and present the data.

