Experiment No. 2

Aim: Data Visualization from Extraction Transformation and Loading (ETL) Process

Objective: In this experiment, we will be able to transform raw data into valuable insights, thereby empowering businesses with the tools needed for data-driven decision-making.

Theory:

A data representation tool is the user interface of the whole business intelligence system. Before it can be used for creating visuals, the data goes through a long process. This is basically a description of how BI works.

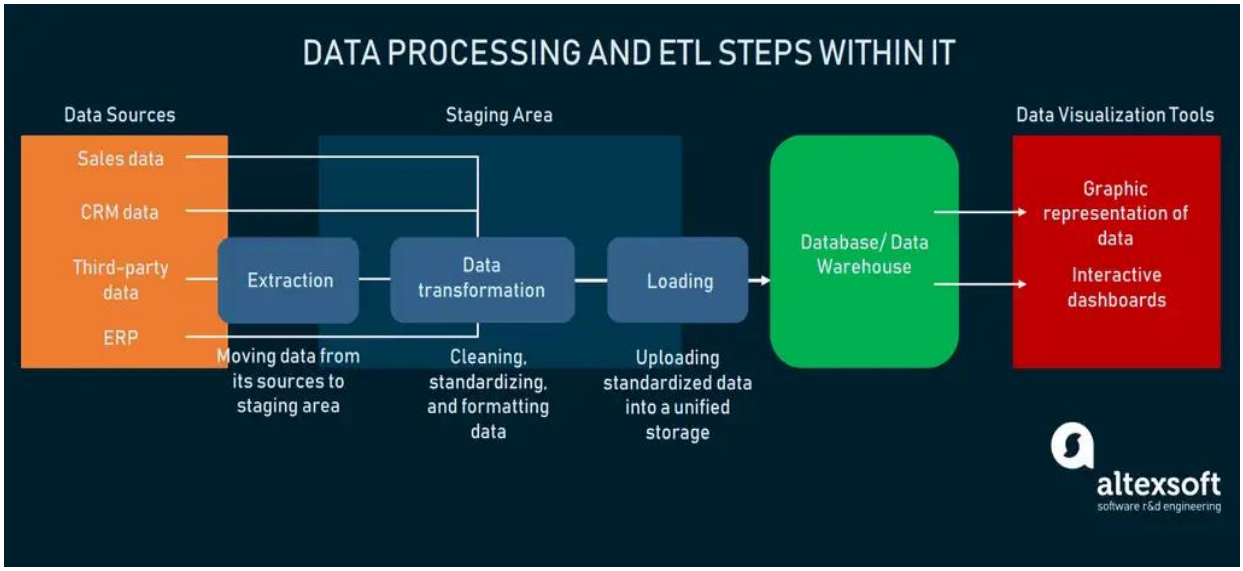
First things first, we should define data sources and data types that will be used.

Then transformation methods and database qualities are determined.

Following that, the data is sourced from its initial storages, for example, Google Analytics, ERP, CRM, or SCM system.

Using API channels, the data is moved to a staging area where it is transformed. Transformation assumes data cleaning, mapping, and standardizing to a unified format.

Further, cleaned data can be moved into a storage: a usual database or data warehouse. To make it possible for the tools to read data, the original base language of datasets can also be rewritten.



Data visualization actually takes place in the whole process. Most modern BI interfaces have a wide number of options concerning the choice of how to use data for visuals. In most cases, there is a command dashboard with a drag-and-drop interface that allows to:

Connect the data source to the system via API (or custom integration)
Choose the dataset to work with

Choose the type of visualization
Place multiple visuals on the dashboard
Create interactive elements to manipulate the data
Modify visuals as the data updates
Type information manually
Save reports
Share reports

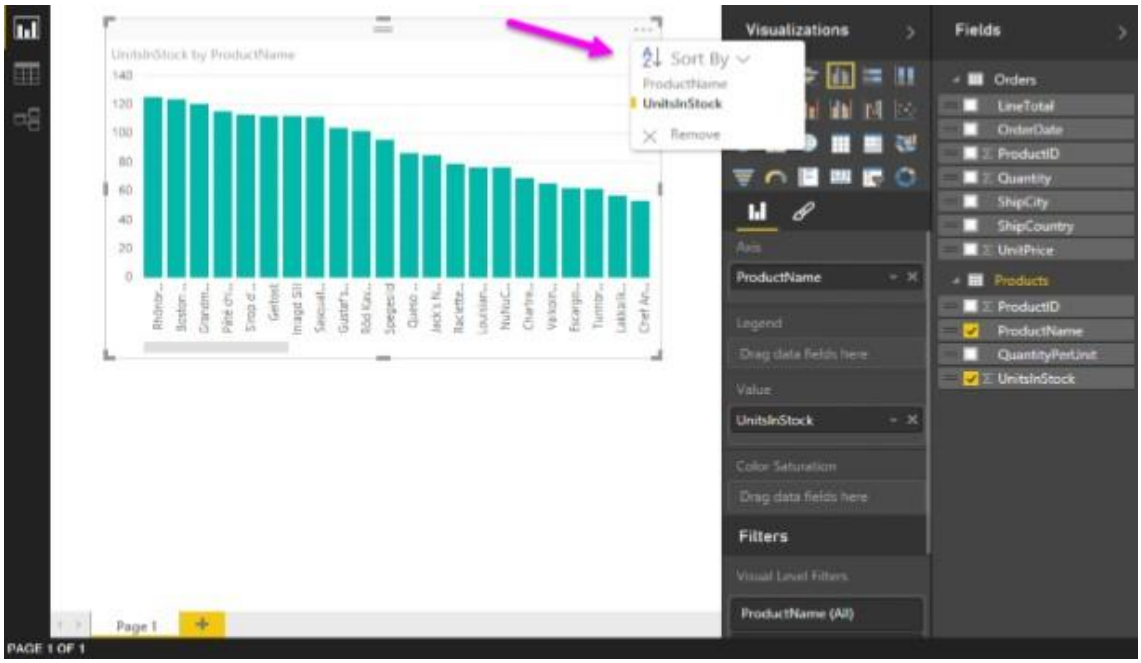
Data Visualization from ETL Process

Power BI Desktop lets you create a variety of visualizations to gain insights from your data. You can build reports with multiple pages and each page can have multiple visuals. You can interact with your visualizations to help analyze and understand your data

In this task, you create a report based on the data previously loaded. You use the Fields pane to select the columns from which you create the visualizations.

Step 1: Create charts showing Units in Stock by Product and Total Sales by Year

1. Drag UnitsInStock from the Field pane (the Fields pane is along the right of the screen) onto a blank space on the canvas. A Table visualization is created. Next, drag ProductName to the Axis box, found in the bottom half of the Visualizations pane. Then we then select Sort By > UnitsInStock using the skittles in the top right corner of the visualization.



2. Drag OrderDate to the canvas beneath the first chart, then drag LineTotal (again, from the Fields pane) onto the visual, then select Line Chart. The following visualization is created.

3. Next, drag ShipCountry to a space on the canvas in the top right. Because you selected a geographic field, a map was created automatically. Now drag LineTotal to the Values field; the circles on the map for each country are now relative in size to the LineTotal for orders shipped to that country.

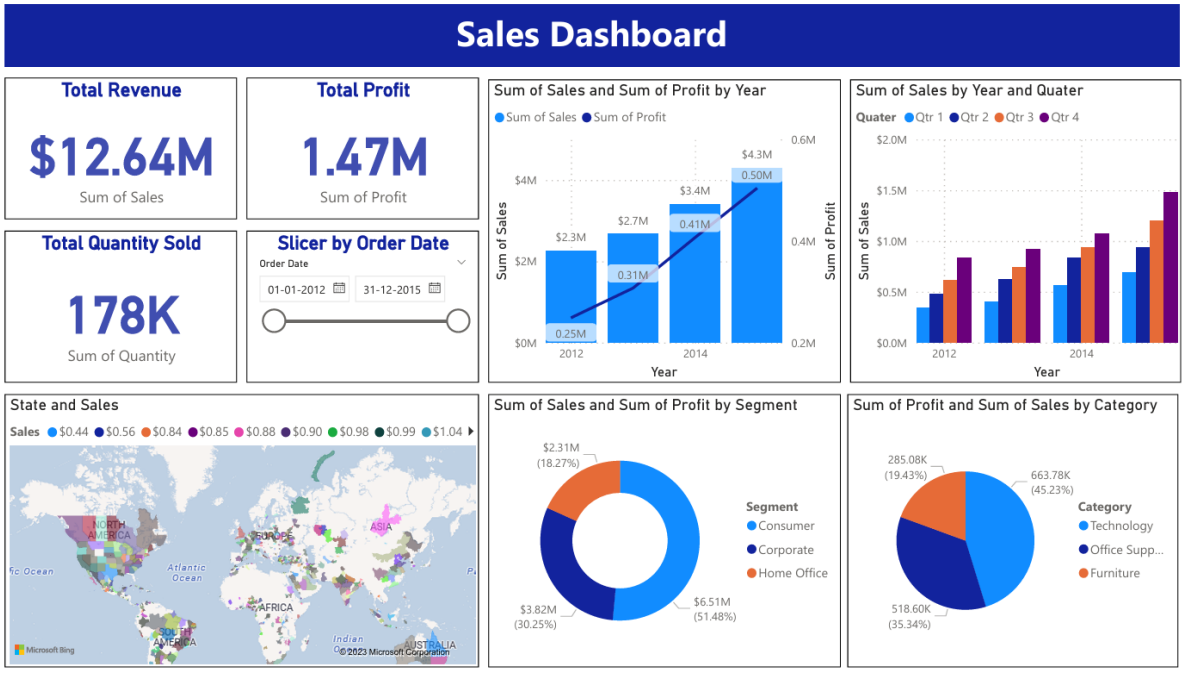
Step 2: Interact with your report visuals to analyze further

Power BI Desktop lets you interact with visuals that cross-highlight and filter each other to uncover further trends.

1. Click on the light blue circle centered in Canada. Note how the other visuals are filtered to show Stock (ShipCountry) and Total Orders (LineTotal) just for Canada.

Input: global_superstore_2016.csv

Output:



Applications:

Power BI Data Modeling and Visualization:

Scenario: Constructing meaningful data models in Power BI for insightful visualizations and reports.

Application: The ETL process shapes the data in a way that aligns with Power BI data models. Relationships between tables are established, and calculations may be pre-aggregated during the ETL to enhance Power BI performance. This ensures that the data is ready for effective reporting and analysis.

Conclusion:

Summarize the overall process, emphasizing the effectiveness of Power BI in turning raw data into compelling visualizations and actionable insights.

Questions:

1. How can ETL tools improve data visualization?

2. What are some best practices for using ETL tools?

3. ETL vs ELT

4.What is data visualization: how it works, types of data to visualize, visualization formats

Answers:

1. ETL tools enhance data visualization by streamlining the process of extracting data from various sources, transforming it into a suitable format for visualization, and loading it into visualization platforms. By automating these tasks, ETL tools ensure that the data presented in visualizations is accurate, consistent, and up-to-date, thereby improving the quality and reliability of insights derived from visualizations.

2. Best practices for using ETL tools encompass several key aspects. These include implementing robust data validation processes to ensure data quality, maintaining comprehensive documentation to facilitate understanding and troubleshooting, employing effective error handling mechanisms to handle exceptions gracefully, optimizing performance through techniques like parallel processing and efficient resource utilization, and designing systems with scalability in mind to accommodate growing data volumes and user demands.

3. ETL (Extract, Transform, Load) involves extracting data from source systems, transforming it to fit the target schema, and then loading it into the target database or data warehouse. ELT (Extract, Load, Transform), on the other hand, involves loading raw data into the target system first and then performing transformations within the target system itself.

4. Data visualization refers to the graphical representation of data to facilitate understanding, analysis, and decision-making. It works by converting raw data into visual formats such as charts, graphs, maps, and dashboards. Types of data commonly visualized include numerical data (e.g., sales figures), categorical data (e.g., product categories), temporal data (e.g., time series), and geospatial data (e.g., geographic distributions). Visualization formats vary depending on the nature of the data and the insights sought, with common formats including bar charts, pie charts, line graphs, scatter plots, heatmaps, and treemaps.