1. **DATA**

The seven (7) different datasets that were utilized in this project were retrieved from seven (7) distinct sources and converted into a comparable format before being placed into the data warehouse for business analytics.

Below are the list of datasets used in this project:

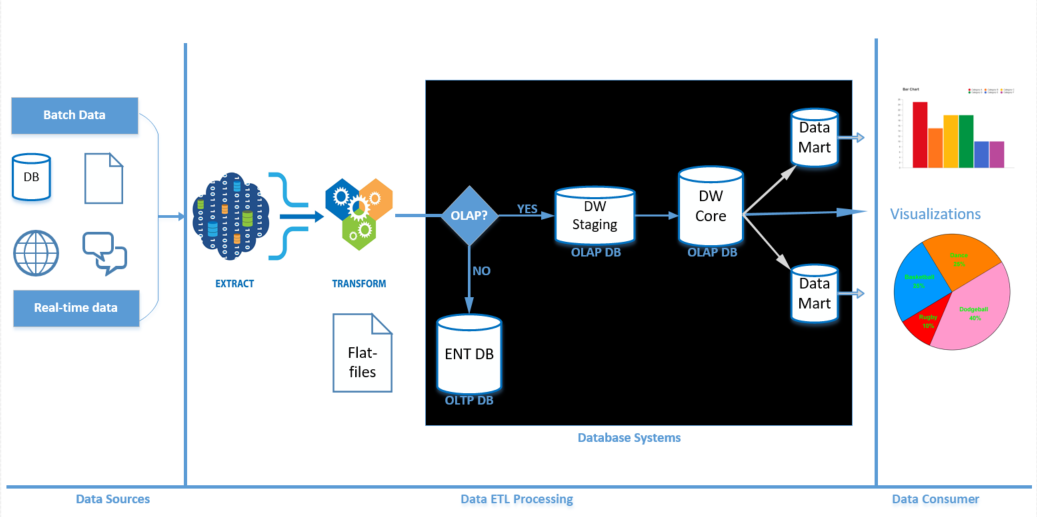
* API Links
  + data API 1
  + data API 2
  + data API 3
* Countries in the world by population (2023) <https://www.worldometers.info/world-population/population-by-country/>: This dataset is real time world population according to Worldometer that was scrapped from their public website
* CSV file: The major goal of the CSV file was to examine how it might be merged with other distinct file types. The CSV file is a local file that was saved on the local disk C. It functioned as a data source for the CSV file.
* JOSN file: The major goal of the JSON file was to examine how it might be merged with other distinct file types. The JSON file is a local file that was saved on the local disk C. It functioned as a data source for the JSON file.
* XML file: The major goal of the XML file was to examine how it might be merged with other distinct file types. The XML file is a local file that was saved on the local disk C. It functioned as a data source for the XML file.

The tools and process used for selection of datasets were based on my experience, speaking of which includes the followings:

1. Python: The language was used because it’s currently one of the best for data project and I also have good knowledge of it as well.
2. SQL server: In the current database environment, SSMS is a superb business database system that can carry out a variety of data management duties. I was forced to choose it as the main database system after carefully examining my experience with SSMS and its data warehouse deployment capabilities.
3. **METHODS**

The methods adopted in this project involves the process of determining the architecture, interfaces, and data for the system in order to comply with certain requirements. An organized method for creating and engineering systems is necessary for the success of the project. When creating a system, it's important to consider every aspect of the infrastructure, from the hardware and software to the data and the manner it's kept.

The overall system design image:



The list below clearly defines the methods involved in the project development:

1. **Data collection:** Python is used to acquire data from both internal and external sources, including third-party suppliers and external sources like databases, APIs, flat-files, and web services, in any format that is appropriate.
2. **Data Processing:** The next step in the data aggregation pipeline is data processing. To be used by the business intelligence system, the data must be cleansed and processed. Data cleansing, data validation, and data translation into a standard format are often required for this.
3. **Data Loading:** In a corporate situation, most data may not be saved in script or Excel files. Many SQL-based relational databases, such as SQL Server, PostgreSQL, and MySQL, are in use, and a number of substitute databases have seen significant growth in popularity. The database to utilize is often determined by the performance, data integrity, and scalability needs of an application. In this project, the Postgres database will be used to load and store the data.
4. **Data Storage:** The next step in the pipeline of data aggregation is the storage of the processed data. A database or flat-file system might be used to carry do this. The data has to be kept in a secure and organized way so that the business intelligence system may easily access and use it.
5. **Design of Business Intelligence System:** Designing a business intelligence system should take into consideration the needs for reporting, data storage, data quality, security, and other factors. ETL, or extract, transform, and load, should be implemented across the system to move data from data sources into data storage. Additionally, the system should be able to analyze data, visualize data, and generate reports.