

Name: Vinayak Soni
Roll No: 67

Practical No. 6

Aim: To perform ETL process for building a data warehouse using power query.

Theory: The ETL process (Extract, Transform, Load) is critical in building a data warehouse. Its main aim is to extract data from various sources, transform it into a suitable format for analysis, and load it into the data warehouse. Power Query is a data transformation and cleansing tool that can perform the ETL process in Microsoft Excel and Power BI. It allows users to connect to various data sources, filter, and clean data, and transform it into a format that can be easily loaded into a data warehouse. The ETL process using Power Query typically involves the following steps:

1. **Extraction:** Data is extracted from various sources such as databases, spreadsheets, or flat files.
2. **Transformation:** The extracted data is cleaned, filtered, and transformed into a format that is suitable for analysis. This includes tasks such as removing duplicates, renaming columns, converting data types, and creating calculated fields.
3. **Loading:** The transformed data is loaded into the data warehouse and stored for future analysis and reporting.

Screenshots:

Here we use a URL: [https://en.wikipedia.org/wiki/List_of_countries_by_GDP_\(nominal\)](https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)) and load it to power query

We perform transform and do preprocessing- Putting correct headers, changing data types, removing errors, etc.

Power Query Editor window showing the query 'GDP USD million by country'. The data is loaded from a table named 'Table1' and is displayed in a grid view. The columns are: Country/Territory, Life Region, Estimate, Year, Estimate, Year, Estimate, Year, Estimate, Year. The data is sorted by Year, showing values for 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, and 2021.

Country/Territory	Life Region	Estimate	Year	Estimate	Year	Estimate	Year	Estimate	Year
World		105168736	2013	96553077	2013	89328313	2013		
United States	Americas	26835099	2013	27986300	2013	23335081	2013		
China	Asia	59372586	2013	37794060	2013	31779411	2013		
Japan	Asia	4405738	2013	4813422	2013	4948878	2013		
Germany	Europe	4308854	2013	4223115	2013	4258835	2013		
India	Asia	3734882	2013	3271358	2013	3293471	2013		
United Kingdom	Europe	3338838	2013	3288880	2013	3133378	2013		
France	Europe	2919489	2013	2831473	2013	2857880	2013		
Italy	Europe	2348745	2013	2387750	2013	2337708	2013		
Canada	Americas	3088673	2013	3088136	2013	3088136	2013		
Brazil	Americas	3081235	2013	3088981	2013	3088981	2013		
Russia	Europe	3062649	2013	3775800	2013	3775802	2013		
South Korea	Asia	3721909	2013	3788134	2013	3838866	2013		
Australia	Oceania	3107568	2013	3163660	2013	3176932	2013		
Mexico	Americas	3660584	2013	3293038	2013	3272839	2013		
Spain	Europe	3470433	2013	3475177	2013	3477881	2013		
Indonesia	Asia	3395738	2013	3380500	2013	3386550	2013		
Netherlands	Europe	3060880	2013	3080007	2013	3062847	2013		
South Africa	Asia	3061861	2013	303341	2013	303341	2013		
Turkey	Asia	3249383	2013	3031779	2013	303833	2013		
Norway	Europe	308881	2013	303887	2013	303887	2013		
Poland	Europe	318881	2013	304048	2013	308618	2013		
Argentina	Americas	645283	2013	493490	2013	388867	2013		
Belgium	Europe	624240	2013	599479	2013	523864	2013		
Sweden	Europe	598851	2013	623438	2013	540864	2013		
Ireland	Europe	598881	2013	598360	2013	625880	2013		
Thailand	Asia	376231	2013	389582	2013	385796	2013		
Norway	Europe	324520	2013	482417	2013	382532	2013		

Finally loading the data in excel sheet:

Excel spreadsheet showing the data loaded from the Power Query query. The data is displayed in a table with columns: Country/Territory, Life Region, Estimate, Year, Estimate, Year, Estimate, Year, Estimate, Year. The data is sorted by Year, showing values for 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, and 2021.

Country/Territory	Life Region	Estimate	Year	Estimate	Year	Estimate	Year	Estimate	Year
World		105168736	2013	96553077	2013	89328313	2013		
United States	Americas	26835099	2013	27986300	2013	23335081	2013		
China	Asia	59372586	2013	37794060	2013	31779411	2013		
Japan	Asia	4405738	2013	4813422	2013	4948878	2013		
Germany	Europe	4308854	2013	4223115	2013	4258835	2013		
India	Asia	3734882	2013	3271358	2013	3293471	2013		
United Kingdom	Europe	3338838	2013	3288880	2013	3133378	2013		
France	Europe	2919489	2013	2831473	2013	2857880	2013		
Italy	Europe	2348745	2013	2387750	2013	2337708	2013		
Canada	Americas	3088673	2013	3088136	2013	3088136	2013		
Brazil	Americas	3081235	2013	3088981	2013	3088981	2013		
Russia	Europe	3062649	2013	3775800	2013	3775802	2013		
South Korea	Asia	3721909	2013	3788134	2013	3838866	2013		
Australia	Oceania	3107568	2013	3163660	2013	3176932	2013		
Mexico	Americas	3660584	2013	3293038	2013	3272839	2013		
Spain	Europe	3470433	2013	3475177	2013	3477881	2013		
Indonesia	Asia	3395738	2013	3380500	2013	3386550	2013		
Netherlands	Europe	3060880	2013	3080007	2013	3062847	2013		
South Africa	Asia	3061861	2013	303341	2013	303341	2013		
Turkey	Asia	3249383	2013	3031779	2013	303833	2013		
Norway	Europe	308881	2013	303887	2013	303887	2013		
Poland	Europe	318881	2013	304048	2013	308618	2013		
Argentina	Americas	645283	2013	493490	2013	388867	2013		
Belgium	Europe	624240	2013	599479	2013	523864	2013		
Sweden	Europe	598851	2013	623438	2013	540864	2013		
Ireland	Europe	598881	2013	598360	2013	625880	2013		
Thailand	Asia	376231	2013	389582	2013	385796	2013		
Norway	Europe	324520	2013	482417	2013	382532	2013		