

# **Data Warehouse**

## **Experiment No.: 6**

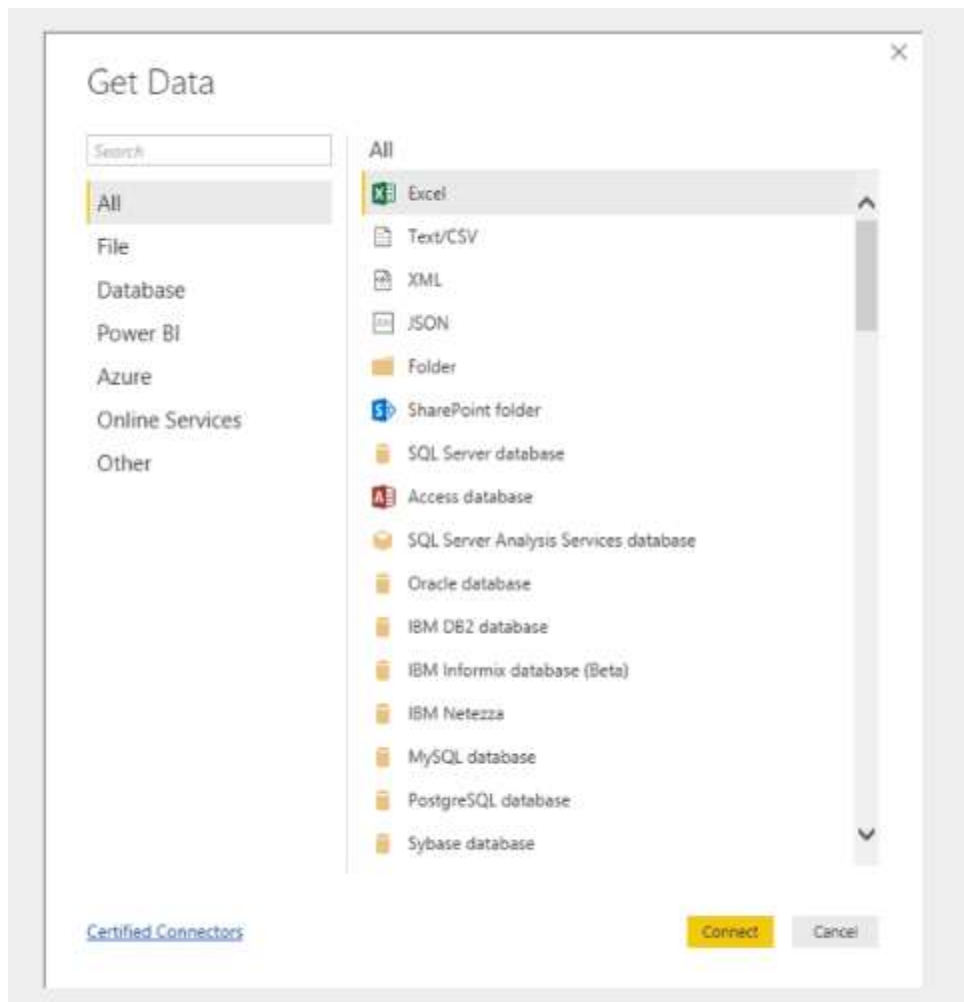
**Perform the Extraction and  
Transformation process to load the data  
using Sqlserver / Power BI.**

# Experiment No. 6

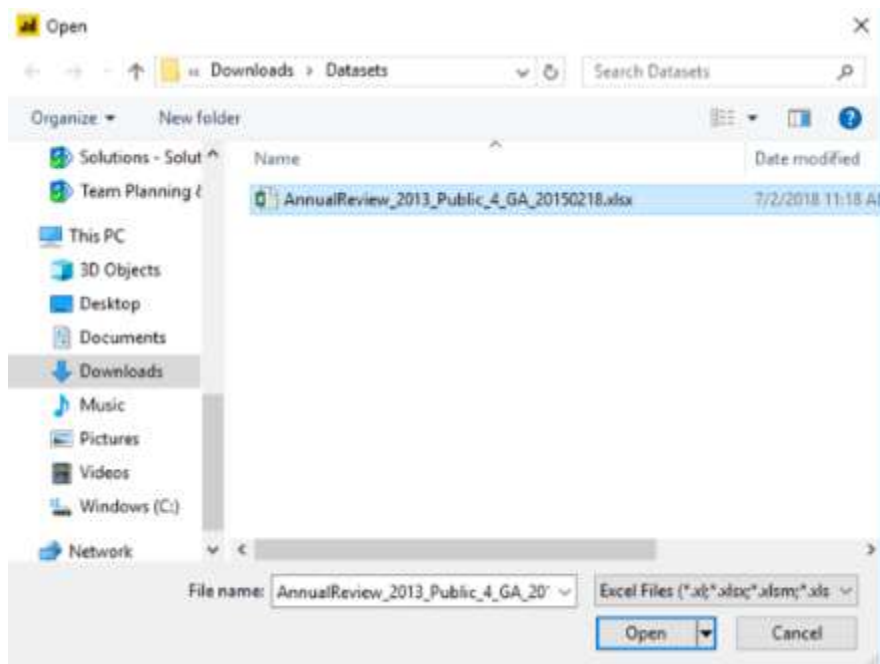
1. **Aim:** Perform the Extraction and Transformation process to load the data using Sqlserver / Power BI.
2. **Objectives:** To extract the data for transformation process.
3. **Course Outcomes:** Initiate the ETL process by connecting the data source and extract the data, which encompasses a wide range of options, including databases, Excel files, web services, and more.
4. **Hardware / Software Required:** Power BI tool to extract and load data into the staging area.
5. **Theory:** ETL is Extract, Transform, Load. It's a form of a data pipeline to integrate various data sources. Without it, your analytical reports and dashboards look old because of outdated data. ETL helps update them so your reports are current. Meanwhile, Microsoft defines Power BI: *Power BI is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights.*

**The first step is to launch Power BI Desktop, then follow these steps:**

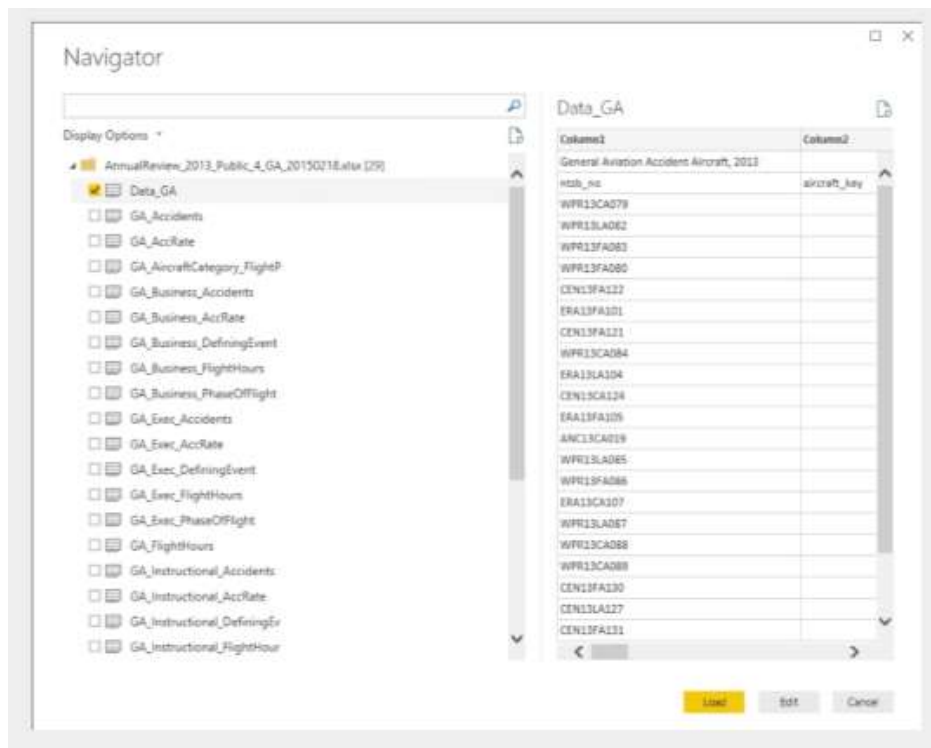
1 From the Power BI splash screen or toolbar, click on “Get Data,” select the Excel connector and click “Connect.”



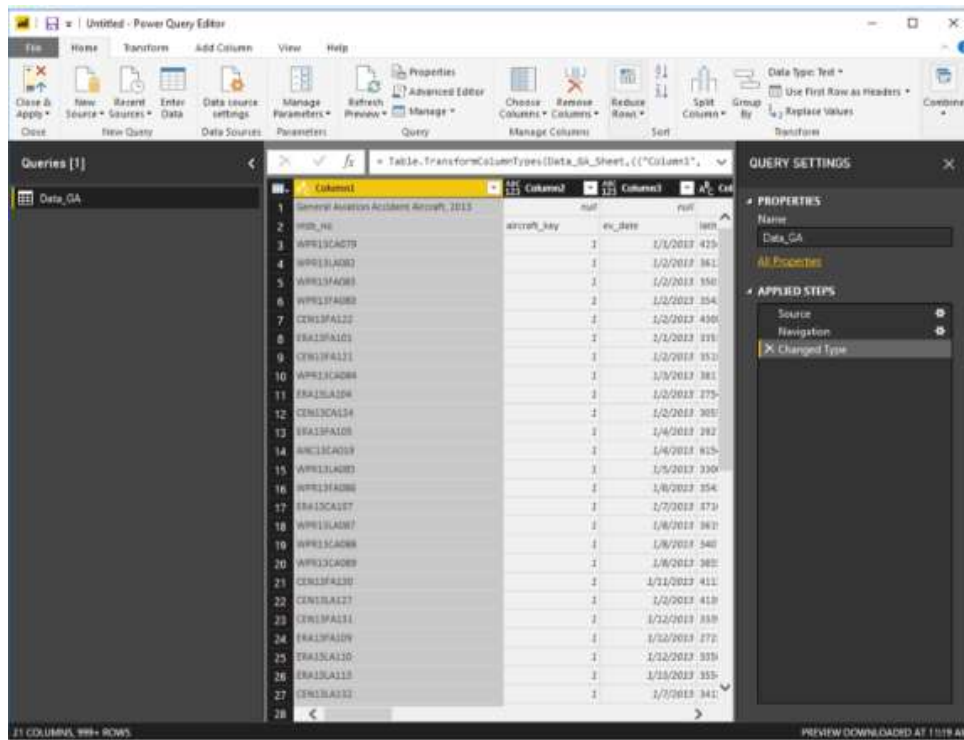
2. Browse for the Excel file, select it from the list and click “Open.”



3 On the Navigator dialog, select “Data\_GA” and click “Edit.”



The Power Query Editor opens to shape and transform our data.



You can see that, on the right, the Query Settings pane lists all of the Applied Steps taken so far. These steps were applied automatically to indicate the path of the source file (Source), the columns that were automatically discovered (Navigation) and the data types automatically detected (Changed Type).

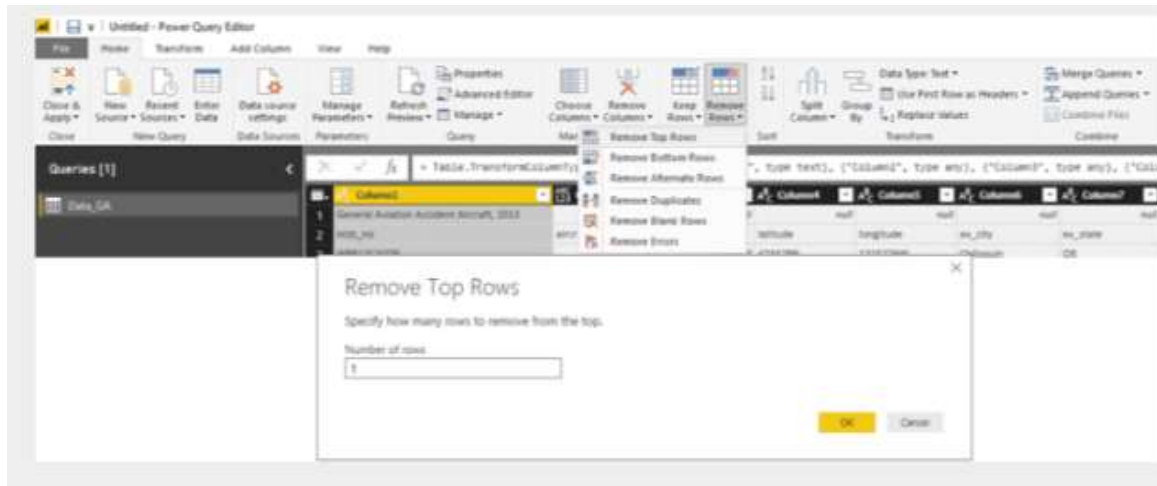
## 6. Results/Output Analysis:

The sample screenshots of extracting and loading data is to be demonstrated.

In cases where data is presented as a table with all of the appropriate headers and no empty rows, these steps will do the job for us accurately. In our case, the source excel file needs a little bit of clean-up—so we’ll shape and transform the data as described below:

### *i. Set Appropriate Column Headers.*

In this step, we’ll need to remove the first row because our column headers are actually in the second row. To remove the first row, click on “Remove Rows” from the Power Query Editor toolbar, click on “Remove Top Rows” and type “1” in the Number of rows dialog box.

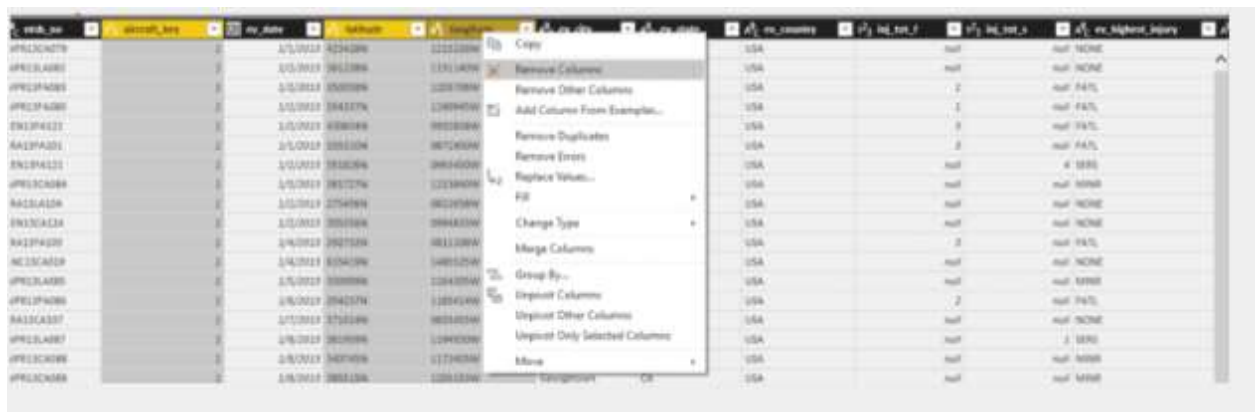


Now, we can indicate that the first row should be used as our column headers. To do this, click the “Use First Row as Headers” option on Power Query toolbar.



## ii Remove Unnecessary Columns

In this step, we need to remove all of the columns that we don’t need to analyze. To remove columns, select the columns while pressing the CTRL key, then right-click and select “Remove Columns.”



## iii Replace Null Values

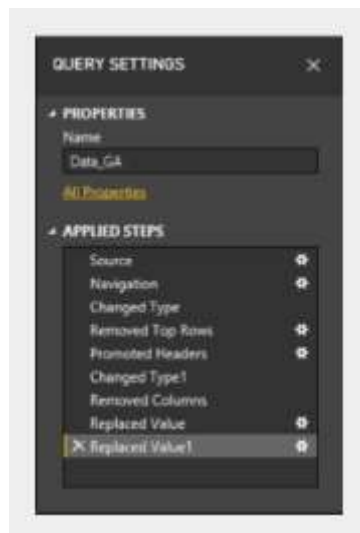
In this step, we need to replace null values with the number zero to ensure accurate analysis and to standardize values. In this example, we can replace all the null values for the “inj\_tot\_f” (Fatal Injuries)

and “inj\_tot\_s” (Serious Injuries) columns. To do this, right-click on any row with a null value and select “Replace Values...” then type the number “0” in the Replace With dialog box and click “OK.”



Repeat for any remaining column with null values. All null values will be now replaced with 0.

Each of these data shaping and transformation steps have been recorded in the Applied Steps pane as seen below. So, whenever data in the data source changes, all you need to do is refresh your Power BI file to reflect the changes. Power Query will then apply the same steps we applied, which will refresh the data in the internal tables as well as all visualizations that reference them.



- Conclusions:** ETL process of extraction by connecting the data source to extract the data, was performed to continue with the transformation stage of ETL process.

**8. Viva Questions:** A list of potential questions related to the ETL operations can be expected.

**9. References:**

1. Kimball Group: Kimball Group's Website offers articles and resources on dimensional modeling and data warehousing.
2. "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling" by Ralph Kimball and Margy Ross
3. Building the Data Warehouse" by William H. Inmon
4. <https://agilethought.com/blogs/extract-transform-load-data-power-bi/>