In this blog we are going to learn about Power BI .pbit files, Power BI desktop file Meta data, Extraction of Power BI Meta data and saving it as an excel file using .pbit file and a simple python code using libraries like Pandas, OS, Regex, JSON and dax\_extract.

The blog will go through the following topics:

* What is Power BI and .pbix files
* What is Power BI .pbit files
* What is the Meta data in a Power BI Desktop file
* Extraction of Meta data using Python.
  1. Exporting .pbix file as .pbit file
  2. Unzipping .pbit file to get DataModelSchema file
  3. Reading .pbit file and DataModelSchema file in python as Dictionary
  4. Extracting Measures from the Dictionary
  5. Extracting calculated columns from the Dictionary
  6. Extracting Relationships from the Dictionary
  7. Saving extracted data as an Excel file.
* Summary

**What is Power BI and .pbix files**

Power BI is a market leading business intelligence tool by Microsoft for Cleaning, Modifying and Visualizing raw data to come up with actionable insights. Power BI comes with its own data transformation engine called power query and a formula expression language called DAX (Data Analysis Expressions).

DAX gives power BI the ability to calculate new columns, dynamic measures and tables inside power Bi desktop.

By default, Power BI report files are saved with .pbix extension which is a renamed version of a ZIP file which contains multiple components, such as the visuals, report canvas, model metadata, and data.

**What is Power BI .pbit file**

.pbit is a template file created by power Bi desktop which is also a renamed version of a ZIP file that contains all the Meta data for the power BI report but doesn’t contain the data itself. Once we extract .pbit file we get a DataModelSchema file along with other files which contain all the Meta data of a power BI desktop file.

Later in this blog we will be using these .pbit and DataModelSchema files to extract Power BI desktop Meta data.

**What is the Meta data in a Power BI Desktop file**

Meta data in a Power BI desktop is everything with regards to what you see when in the Report view. All the details like name, source, expression, data type of all the tables, Calculated tables, Calculated Columns, Calculated Measures, Relationships and Lineage between different tables of the model, Hierarchies, Parameters etc. can be considered as Meta data.

Here in this Blog, we will mainly focus on extracting Calculated Measures, Calculated Columns and Relationships.

**Extraction of Meta data using Python**

We have used Python to process and extract the JSON from DataModelSchema and .pbit file. Once we had JSON, we converted it to python dictionary and extracted required Meta data.

Below are the steps we will need to achieve the requirement:

1. **Exporting .pbix file as .pbit file:** There are two ways to save our power BI desktop file as .pbit file.
   1. Once we are in Power BI desktop, we have an option to save our file as power BI template(.pbit) file
   2. We can go to File-->Export-->Power BI Template and save the .pbit file at the desired directory.
2. **Unzipping .pbit file to get DataModelSchema file:** We can directly unzip the .pbit file using the 7z-Zip file manager or any other file manager. Once we Unzip the file, we will get a folder with the same name as that of the .pbit file. Inside the folder we will get the DataModelSchema file, we will have to change its extension to .txt for reading in python.
3. **Reading .pbit and Data model schema file in python:** We have an option to directly read the .pbit file in python using the dax\_extract library. Second option to read the text file in python and using the JSON module convert it into a python dictionary. Code can be found in the github repository link given at the end of this file.
4. **Extracting Measures from the dictionary:** The dictionary that we get consists details of all the tables as separate lists, Individuals tables have details related to the columns and measures belonging to that table, we can loop on each table one by one and get details of Columns, Measures etc. Below is sample output of the python code can be found in the github repository link given at the end of this file.

|  | **table Number** | **table Name** | **Measure Name** | **Measure Expression** |
| --- | --- | --- | --- | --- |
| **0** | 5 | Query Data | % Query Resolved | CALCULATE(COUNT('Query Data'[Client ID]),'Quer... |
| **1** | 5 | Query Data | Special Query Percentage | CALCULATE(COUNT('Query Data'[Client ID]),'Quer... |
| **2** | 6 | Asset Data | Client Retention Rate | CALCULATE(COUNT('Asset Data'[Client ID]),'Asse... |

1. **Extracting calculated columns from the Dictionary:** Like how we extracted the measures we can loop on each table and get details of all the calculated columns. Below is the sample output of the python code can be found in the github repository link given at the end of this file

|  | **table no** | **Table Name** | **name** | **expression** |
| --- | --- | --- | --- | --- |
| **6** | 2 | Calendar | Day | DAY('Calendar'[Date]) |
| **7** | 2 | Calendar | Month | MONTH('Calendar'[Date]) |
| **8** | 2 | Calendar | Quarter | CONCATENATE("Q",QUARTER('Calendar'[Date]) ) |
| **9** | 2 | Calendar | Year | YEAR('Calendar'[Date]) |

1. **Extracting relationships from the dictionary:** Data for relationships is available in the model key of the data dictionary and can be easily extracted. Below is the sample output of the python code can be found in the github repository link given at the end of this file .

|  | **From Table** | **From Column** | **To Table** | **To Column** | **State** |
| --- | --- | --- | --- | --- | --- |
| **0** | Operational Data | Refresh Date | LocalDateTable\_50948e70-816c-4122-bb48-2a2e442... | Date | ready |
| **1** | Query Data | Client ID | Client Data | Client ID | ready |
| **2** | Query Data | Query Date | Calendar | Date | ready |
| **3** | Asset Data | Client ID | Client Data | Client ID | ready |
| **4** | Asset Data | Contract Maturity Date | LocalDateTable\_d625a62f-98f2-4794-80e3-4d14736... | Date | ready |
| **5** | Asset Data | Enrol Date | Calendar | Date | ready |

1. **Saving Extracted data as an Excel file:** All the extracted data can be saved in empty lists and these lists can be used to derive a pandas data frame. This pandas dataframe can be exported as Excel and easily used for reference and validation purposes in a complex model. Below snapshot gives an idea of how this can be done.

**Summary**

In this blog we learnt about extracting metadata from .pbit and DataModelSchema file. We have created a Python script that allows users to enter the file location of .pbit and DataModelSchema file and then metadata extraction along with excel generation can be automated. The code can be found on the below github link Also sample excel files can be downloaded from below github link.

[YashShrivastava07/PowerBI-Metadat-extractor (github.com)](https://github.com/YashShrivastava07/PowerBI-Metadat-extractor)