PROJECT 06

HR MANAGEMENT

1. Data Preprocessing
2. Data Visualization

By:
Mehrdad Mansourdehghan

GitHUb

YouTube Channel

LinkedIn Profile

Send me E-mail

1. Project Goals

In this project, I'm going to clean and preprocess data in Microsoft Power Query, and then I use Microsoft Power BI to visualize data and create an interactive. dashboard. During this project, I tried to make dynamic solutions to build an automatic way for the next data.

2. Language, libraries, tools:

Language: Power Query, DAX
Libraries: -
IDE: -
Application: Microsoft Excel, Microsoft PowerBI

3. Data

This data is about working hours and shifts of employees in a company in three months. Each label of working has its own meaning, and you can see the details in the Excel sheet. In this data, we have all working days in three month "April", "May" and "June" for all employees and it is written that in each day what was the status of working in terms of HR management. Then based on the rules in the company, they calculate some measures for each of the employees and they must stick to the regulations. First understand the requirements of HR team:

1. The first requirement is that the HR manager wants to know what the preference of working for employee. They are willing to work from home or in the office and what reason is behind that. The reason is that, when people prefer to work from home, they want to work on Friday or Monday. So, it makes sense for the company to know on which day, they can have for example team building to other activity and when they can have the majority of people all together. Another reason is about capacity planning and company can share its capacity and resource (e.g., laptop or other resource) to everybody during week.
2. Second thing is HR team want to know the level of sickness to know whether the illness is dangerous for other people (like Covid-19) to not. I mean they want to have a sickness level.

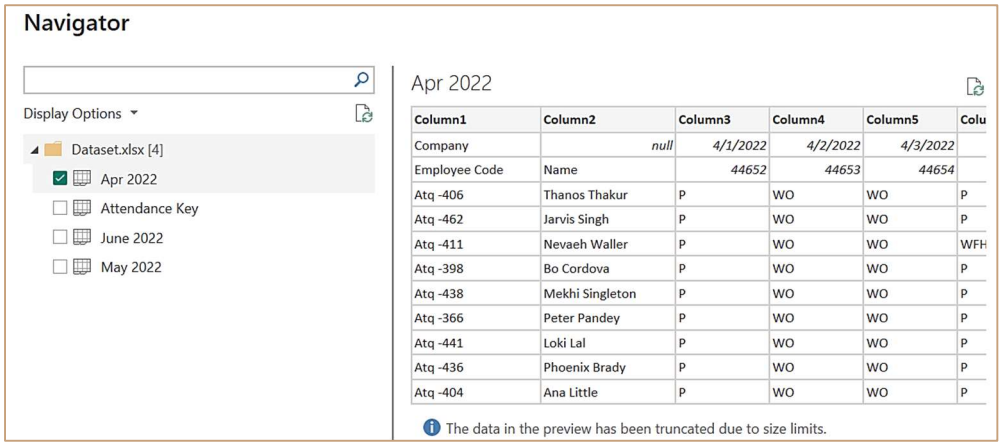
For example, if on a specific day, 15% or 10% of employees are leaving or absent, there might be a bad thing about prevalence of sickness in company, because it is not a coincidence that this number of people are leaving or absent. Thus, the company wants to have understand about that (even seasonal, flu or other illness) to take some precaution measures in advance.

Data Preprocessing

Here you can follow all steps that were taken in this part.

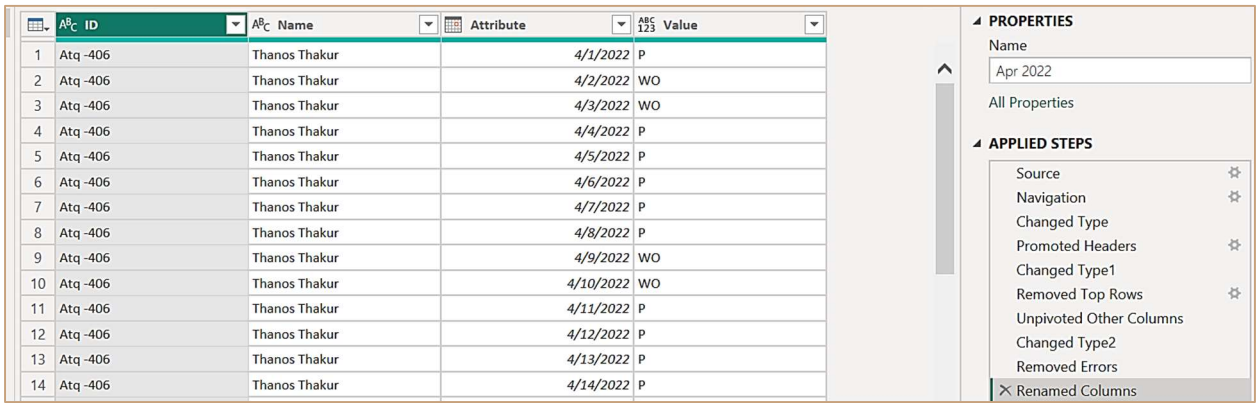
1. Transform Data and Build Parameter

Since the sheets have the diffrenet name and number of columns, we can Power Query feature to load the data and combin them. I run the process in the following, to create a dynamic model instead of doing manually for all sheets. First of all, I select just the first sheet and go to transform in Power Query.



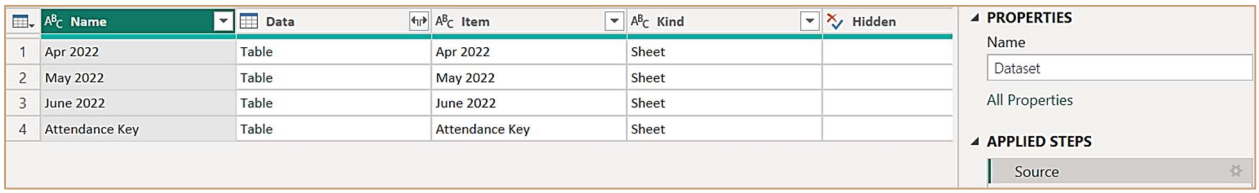
Column1	Column2	Column3	Column4	Column5	Colu
Company	null	4/1/2022	4/2/2022	4/3/2022	
Employee Code	Name	44652	44653	44654	
Atq -406	Thanos Thakur	P	WO	WO	P
Atq -462	Jarvis Singh	P	WO	WO	P
Atq -411	Nevaeh Waller	P	WO	WO	WFH
Atq -398	Bo Cordova	P	WO	WO	P
Atq -438	Mekhi Singleton	P	WO	WO	P
Atq -366	Peter Pandey	P	WO	WO	P
Atq -441	Loki Lal	P	WO	WO	P
Atq -436	Phoenix Brady	P	WO	WO	P
Atq -404	Ana Little	P	WO	WO	P

Now, in power query, do all needed data wrangling process. All these processes will be applied for all sheets later, as well.



	ID	Name	Attribute	Value
1	Atq -406	Thanos Thakur	4/1/2022	P
2	Atq -406	Thanos Thakur	4/2/2022	WO
3	Atq -406	Thanos Thakur	4/3/2022	WO
4	Atq -406	Thanos Thakur	4/4/2022	P
5	Atq -406	Thanos Thakur	4/5/2022	P
6	Atq -406	Thanos Thakur	4/6/2022	P
7	Atq -406	Thanos Thakur	4/7/2022	P
8	Atq -406	Thanos Thakur	4/8/2022	P
9	Atq -406	Thanos Thakur	4/9/2022	WO
10	Atq -406	Thanos Thakur	4/10/2022	WO
11	Atq -406	Thanos Thakur	4/11/2022	P
12	Atq -406	Thanos Thakur	4/12/2022	P
13	Atq -406	Thanos Thakur	4/13/2022	P
14	Atq -406	Thanos Thakur	4/14/2022	P

Then I create a function to do all above steps, automatically. For doing this, I duplicate the data and remove all steps except the source, and I changed the name to dataset.



	Name	Data	Item	Kind	Hidden
1	Apr 2022	Table	Apr 2022	Sheet	
2	May 2022	Table	May 2022	Sheet	
3	June 2022	Table	June 2022	Sheet	
4	Attendance Key	Table	Attendance Key	Sheet	

PROPERTIES
Name
Dataset
All Properties
APPLIED STEPS
Source

Now go back to Apr 2022 (the first data) and convert it to function. For doing this, I change something in advanced editor and I call it Transformer

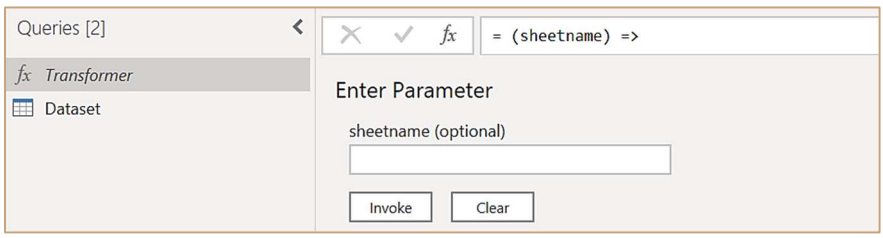
Advanced Editor

Apr 2022

Display Options

```
(sheetname) =>

let
    #"Changed Type" = Table.TransformColumnTypes(sheetname,{{"Column1", type text}, {"Column2", type text}, {"Column3", type any}, {"Column4",
    #"Promoted Headers" = Table.PromoteHeaders(#"Changed Type", [PromoteAllScalars=true]),
    #"Changed Type1" = Table.TransformColumnTypes(#"Promoted Headers",{{"Company", type text}, {"Column2", type text}, {"4/1/2022", type any}
    #"Removed Top Rows" = Table.Skip(#"Changed Type1",1),
    #"Unpivoted Other Columns" = Table.UnpivotOtherColumns(#"Removed Top Rows", {"Column2", "Company"}, "Attribute", "Value"),
    #"Changed Type2" = Table.TransformColumnTypes(#"Unpivoted Other Columns",{{"Attribute", type date}}),
    #"Removed Errors" = Table.RemoveRowsWithErrors(#"Changed Type2", {"Attribute"}),
    #"Renamed Columns" = Table.RenameColumns(#"Removed Errors",{{"Column2", "Name"}, {"Company", "ID"}})
in
    #"Renamed Columns"
```



Queries [2]

fx Transformer

Dataset

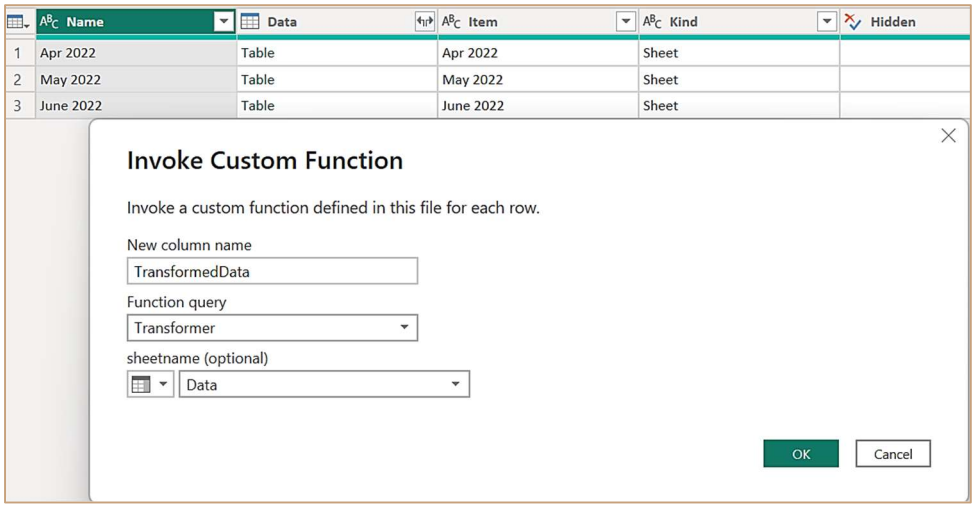
= (sheetname) =>

Enter Parameter

sheetname (optional)

Invoke Clear

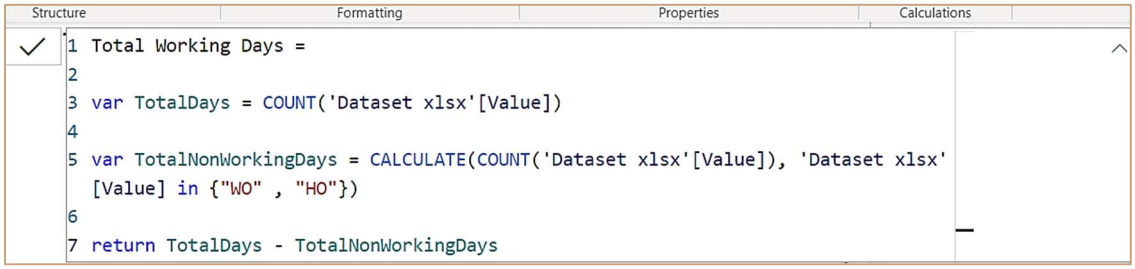
Now, in Dataset remove all un-needed rows and invoke function to it. It means I am appling the function to all these sheets.



Now dataset is ready!

2. Build Metric and Measures

To answer all the HR's requirements, we should create some metrics and measures. I use DAX to do this. First, I create a table for measures and in this table, I find the total days and total non-working days. In this dataset, "HO" and "WO" are non-working days (Week Off, Holiday Off). Then if is non-working days from all days, I will have just working days.




```
Total Working Days =

var TotalDays = COUNT('Dataset xlsx'[Value])

var TotalNonWorkingDays = CALCULATE(COUNT('Dataset xlsx'[Value]), 'Dataset xlsx'[Value] in {"WO" , "HO"})

return TotalDays - TotalNonWorkingDays
```

Then I must calculate the present days. This measure means whether the employee works at office ("P"), totally work at home ("WFH"), or half work from home ("HWFH"), I must calculate the present days. First of all, I want to calculate the number of days for working from home ("WFH") and ("HWFH"). So, for doing this, I consider if an employee has "WFH", it gets 1, if she/he has "HWFH",

gets 0.5 and else gets 0. Then, I can calculate all present days with aggregation of that new column and all rows that have "P".

```
#Presence_Score = SWITCH(
    TRUE(),
    'Dataset xlsx'[Value] = "WFH", 1,
    'Dataset xlsx'[Value] = "HWFH" , 0.5,
    'Dataset xlsx'[Value] = "P", 1,
    0
)
```

heet	Employment Code	Name	Date	Value	#Presence_Score
pr 2022	Atq -406	Thanos Thakur	4/11/2022	P	1
pr 2022	Atq -462	Jarvis Singh	4/11/2022	P	1
pr 2022	Atq -411	Nevaeh Waller	4/11/2022	P	1
pr 2022	Atq -438	Mekhi Singleton	4/11/2022	P	1
pr 2022	Atq -366	Peter Pandey	4/11/2022	P	1
pr 2022	Atq -441	Loki Lal	4/11/2022	P	1

Now, I can find total number of present days:

```
Total Present = SUM('Dataset xlsx'[#Presence_Score])
```

Also, we can find the percentage of presence:

```
% Present = DIVIDE([Total Present], [Total Working Days])
```

The next measure that I made is percentage of working from home ("WFH"). For doing this, I should divide all records that contain this type by all present days.

```
% WFH =

var TotalWFh = CALCULATE(COUNT('Dataset'[Value]), 'Dataset'[Value] in {"WFH"})

return TotalWFh / [Total Present]
```

Now, I do the same process for sick, and for doing this, I consider SL and HSL that mean sick leave and half sick leave.

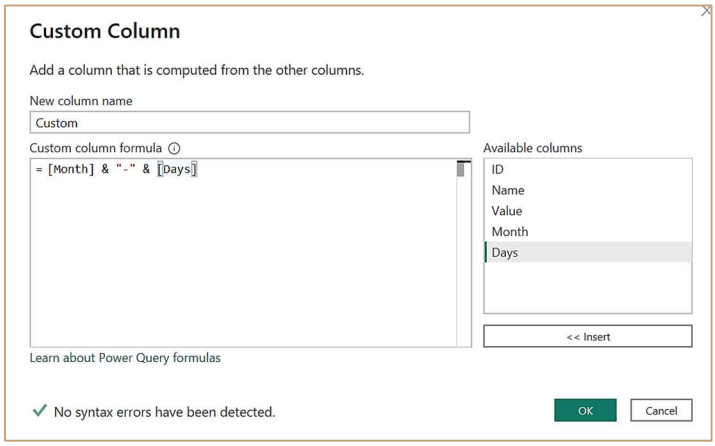
```
#Leave_Score = SWITCH(
    TRUE(),
    'Dataset'[Value] = "SL", 1,
    'Dataset'[Value] = "HSL" , 0.5,
    0
)
```

```
Total Leave = SUM('Dataset'[#Leave_Score])
```

```
% Leave = DIVIDE([Total Leave], [Total Working Days])
```

3. Normalize Date column.

To have a better view about the date, I normalize the date and I break it down to year, month, and weekdays. First of all, we should create a date column. Since so far, we have a column that contains the name of month, and another column containing number of days in month. So, we can combine them to create a date column.



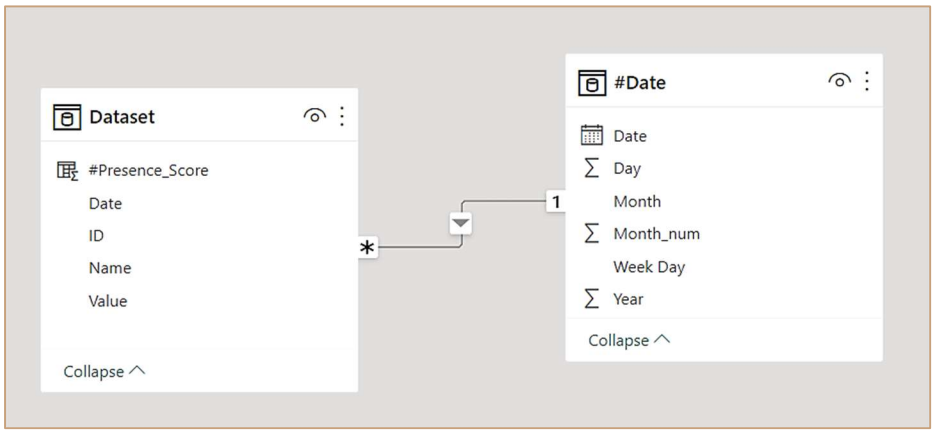
ID	Name	Value	Date
Atq -406	Thanos Thakur	P	4/1/2023
Atq -406	Thanos Thakur	WO	4/2/2023
Atq -406	Thanos Thakur	WO	4/3/2023
Atq -406	Thanos Thakur	P	4/4/2023
Atq -406	Thanos Thakur	P	4/5/2023
Atq -406	Thanos Thakur	P	4/6/2023
Atq -406	Thanos Thakur	P	4/7/2023
Atq -406	Thanos Thakur	P	4/8/2023

Then we create a separate table for the date.

```
1  #Date = ADDCOLUMNS(
2      CALENDAR(MIN('Dataset'[Date]), MAX('Dataset'[Date])),
3      "Year", YEAR([Date]),
4      "Month_num", MONTH([Date]),
5      "Month", FORMAT([Date], "MMMM"),
6      "Day", DAY([Date]),
7      "Week Day", FORMAT([Date], "DDDD"))
```

Date	Year	Month_num	Month	Week Day	Day
4/1/2023 12:00:00 AM	2023	4	April	Saturday	1
4/2/2023 12:00:00 AM	2023	4	April	Sunday	2
4/3/2023 12:00:00 AM	2023	4	April	Monday	3
4/4/2023 12:00:00 AM	2023	4	April	Tuesday	4
4/5/2023 12:00:00 AM	2023	4	April	Wednesday	5
4/6/2023 12:00:00 AM	2023	4	April	Thursday	6
4/7/2023 12:00:00 AM	2023	4	April	Friday	7

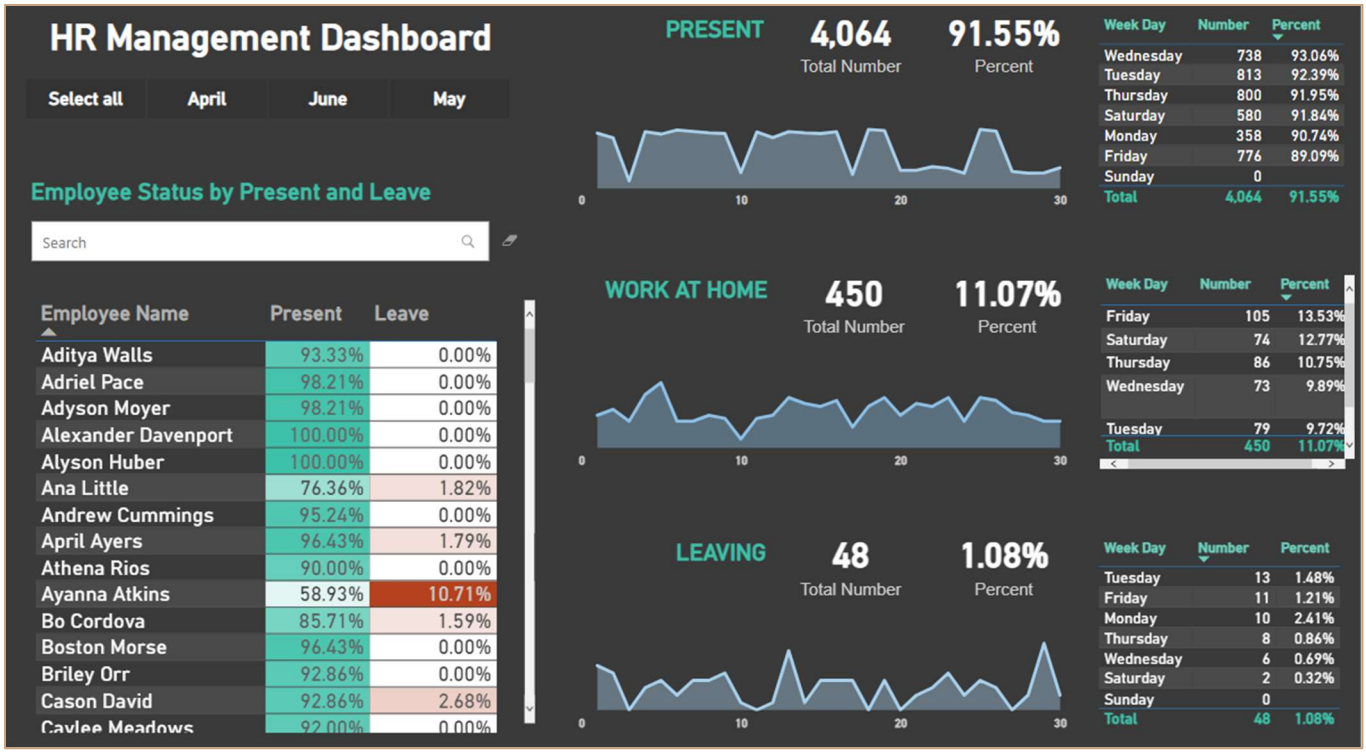
Then we make a relationship between the date and the old date.



So, from now on, we use the new date wherever we need date values.

Data Visualization

In this section we want to analyze data based on visualization, because I believe the best way to analyze the data is in visualized way. So, by doing this we can answer some ad-hoc questions that might be asked in daily-basis business. So far, we have a good image of the data, and we can help managers or users who are willing to have insight about the data. During this section, from time to time I had to come back to data preprocessing and creating new measures.



On the top left, we see a filter that you can use to search the name of an employee. In the bottom, we see the last status of all employees by their percentage of present and leave. This table has conditional formatting that can show the differences with color. In right side of dashboard and from top to down, we see the status of present, work from home and leaving with their cards and weekday table. Also, we see a line chart that shows the trend for each of them during a month.

- END -