Add Your Name Here

Big Data and Data Visualisation - Assessment

Portfolio of tasks

# Instructions

For this module, you are required to provide evidence of the completed tasks listed below.

Add the evidence/screenshots required below and submit this document and the Power BI files as part of your submission.

|  |  |  |
| --- | --- | --- |
| **Power BI - Assessment Sheet** | | |
| **Student Name** |  |  |
| **Student ID** |  |

Once you have provided a screenshot for the tasks below, save this file as a PDF file and submit it together with the Power BI files that you have used for the tasks.

Please note that for each of the tasks below you need to show that

1. You can work with data to produce a dashboard
2. The dashboard must demonstrate an understanding of the theory of human visual perception, appropriate chart and design choices for the scenario.

# Data Analytics Task 1: Data Pre-processing and Report Design (30 marks)

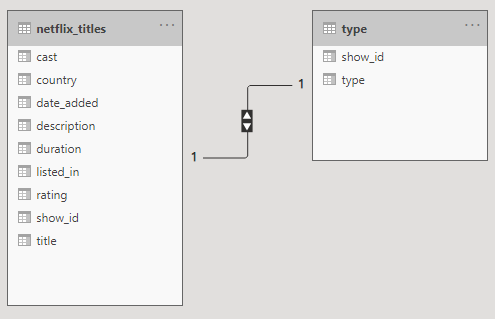
You are asked to take screenshots of your work as you performs the steps below. Make sure you check the table on Page 4 before starting the tasks.

This task uses data in tv shows and movies available on Netflix.

Download the file “Netflix\_titles.csv” from Blackboard.

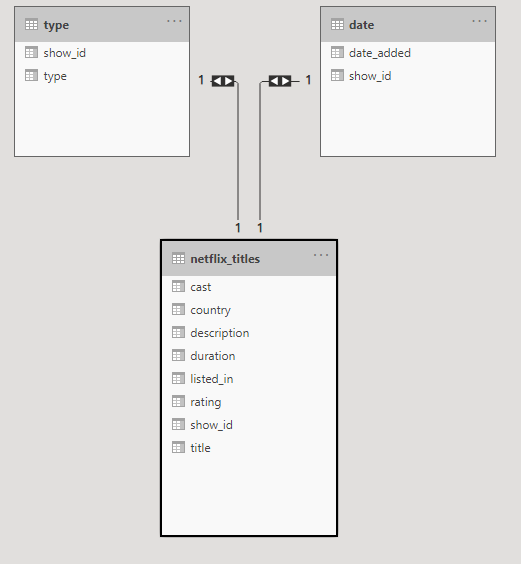
Below there is the set of tasks that you need to complete. Provide the screenshots required in the table at the end of the tasks.

1. Import the data. This is a .csv file, not an Excel file. Do you remember how to load it?
2. Delete the ”director” column
3. Remove all the empty rows in the “cast” column
4. Create a new table “Type” – this table must contain only the “show\_id” column and the “Type” column
5. Remove the “type” column from the Netflix\_titles table
6. Close and apply your changes, go to report view and check that you have two tables with one relationship as shown below. If not, you have to create the relationship manually.



If the double arrow is not showing in your model, don’t worry. You are probably using a different version of Power BI. What matters is that you have a 1-to-1 relationship (1s on both sides).

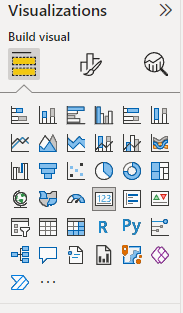
1. Then create another table called “Date”, containing only the show\_id and date\_added columns.
2. Remove the date\_added column from the Netflix\_titles table.
3. Close&Apply your changes, you should now have the three tables below:



If the double arrows are not showing in your model, don’t worry. You are probably using a different version of Power BI. What matters is that you have a 1-to-1 relationship (1s on both sides).

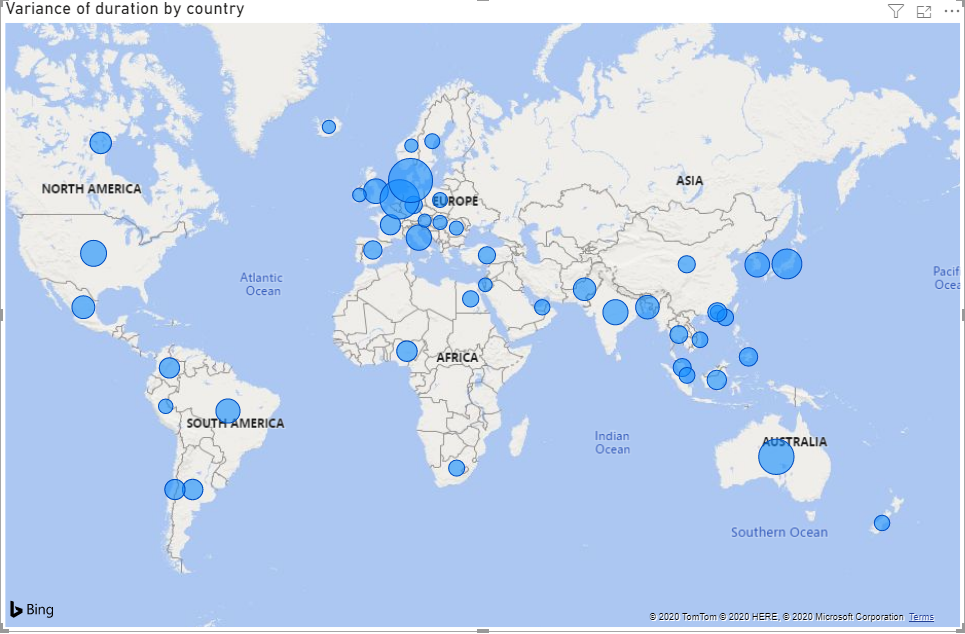
1. Then, remove all the “TV Show” types from the dataset by deleting all the rows in the “type” table containing “TV Shows”.
2. In the “duration” column remove all the rows that contain the word “Season”, e.g. “1-Season”, “2-Seasons”, “1 s”, “2 s”, etc...
3. Remove the word “min” from each row of the “duration” column. You should leave only the numbers in each row.
4. Change the data type of “duration” into whole number

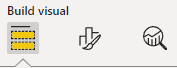
The next part of this task will focus on your reporting skills. You will have to **explore** the different Power BI reporting tools to be able to complete these tasks.

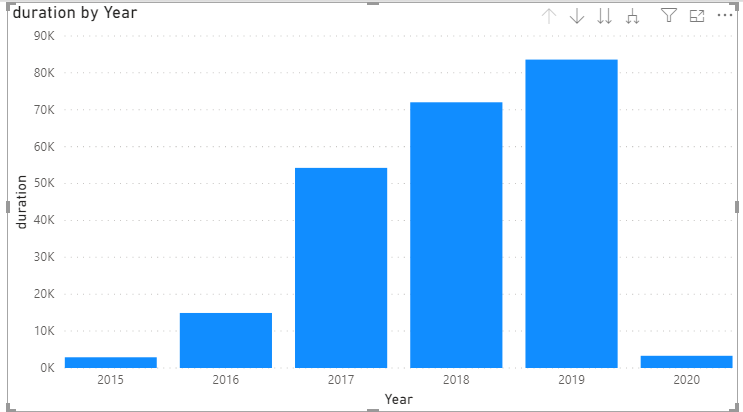


Remember that you can select the chart and use the Format and Analytics tab to change any formatting options.

1. Create a map similar to the one below that shows the duration by country (please note that your map will look different than this example).



Explore the format tab to complete the tasks below 

1. Change the title text size to 16.
2. Centre the title
3. Change the colours of the blue bubbles into red
4. Filter all values of “duration” to show items greater than 60 minutes. *Hint: use the Filters panel*
5. On a new page of your repot, create the chart on the right that shows the duration by year.
6. Change the colour of the bars into red
7. Remove the title of your chart.
8. Change the name of the y-axis into “Duration in minutes”.
9. Explore and use the Analytics tab to add the “Average Line”.
10. Label the average line “Average duration in minutes”.
11. Show both the label of the line and the value
12. Change the colour of the labels into black
13. Change the colour of the average line into black.

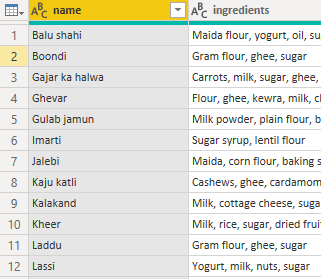
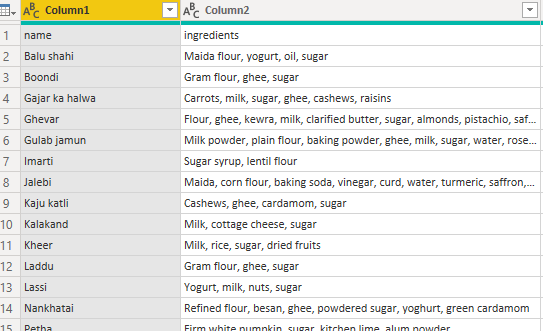
|  |  |  |
| --- | --- | --- |
| **Questions** | | **Provide a screenshot of the required questions** |
| 4 | Create a new table “Type” – this table must contain only the “show\_id” column and the “Type” column |  |
| 9 | Close&Apply your changes, you should now have the three tables |  |
| 13 | Change the data type of “duration” into whole number |  |
| 14 | Create a map that shows the duration by country |  |
| 18 | Filter all values of “duration” to show items greater than 60. |  |
| 22 | Change the name of the y-axis into “Duration in minutes”. |  |
| 27 | Change the colour of the average line into black. |  |
| *Independent Tasks* | | |
|  | Add two more charts to further investigate the dataset | The treemap visually represents which countries contribute the most to total movie duration in the dataset. Each rectangle's size corresponds to the sum of movie minutes per country, while the color intensity indicates the average movie length—darker shades represent longer average durations. This chart helps quickly identify both the most productive countries (by total content length) and those that produce lengthier films on average, offering a clear, space-efficient overview of geographic content distribution. |
|  |  | This chart shows the distribution of movies by country based on the count of unique *show\_id* values in the dataset. It highlights which countries contribute the highest number of movies, with the United States and India leading by a wide margin, followed by the United Kingdom and several European and Asian countries. The horizontal bar layout makes it easy to compare countries and quickly identify dominant contributors to the movie catalog. This chart was selected because it effectively illustrates geographic trends in movie production and availability, helping stakeholders understand regional dominance, content diversity, and potential gaps in international representation. |
| Arrange your tiles/charts to create a well-designed dashboard (use the knowledge covered in the lessons). Add a title, change the colours/formatting options of the charts.  Feel free to add more charts if you think you need more.  Once you are happy with the dashboard design, take a screenshot and paste it below. | | |
|  | | |
| Use the space below to explain how you did apply any of the theoretical aspects of data visualization covered in this module (for example, you could justify the selection of a specific chart or design choices for the scenario). Please provide three different points.  **1. Choosing Effective Chart Types** For analyzing movie data by country, I selected two complementary charts: a bar chart showing the count of movies by country and a treemap representing total duration by country. The bar chart was chosen because it clearly and simply compares discrete quantities—making it easy to see which countries have the most movies. The treemap, on the other hand, was chosen to display part-to-whole relationships, where each country’s rectangle size corresponds to total movie minutes. This allows viewers to quickly grasp both quantity and contribution to total runtime in one visual.  **2. Using Pre-attentive Visual Cues** I applied color and size as pre-attentive attributes to enable quick interpretation. In the treemap, darker shades indicate higher average movie durations, while larger rectangles represent greater total minutes. These visual cues help the viewer instantly identify patterns—such as which countries produce longer films or contribute the most content—without needing to read labels or numbers closely. This leverages how the human eye naturally perceives differences in color intensity and area.  **3. Supporting Comparative Analysis** By placing the bar chart and treemap side by side, the dashboard supports direct comparison between the number of movies per country and their cumulative duration. This dual-view approach reveals insights such as whether a country with many films also has longer runtimes, or if a smaller film industry produces lengthier content. Clear titles, labeled axes, and consistent color schemes ensure the information is organized and accessible, helping users draw meaningful conclusions efficiently. | | |

# Data Analytics Task 2: Data Modelling (25 marks)

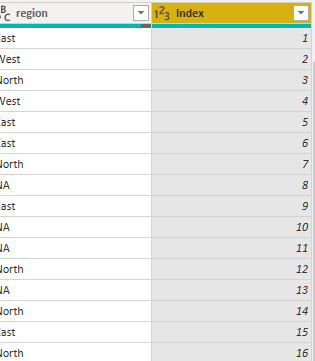
The tasks below use the Restaurant Orders dataset available on Blackboard. Download the file before starting the tasks.

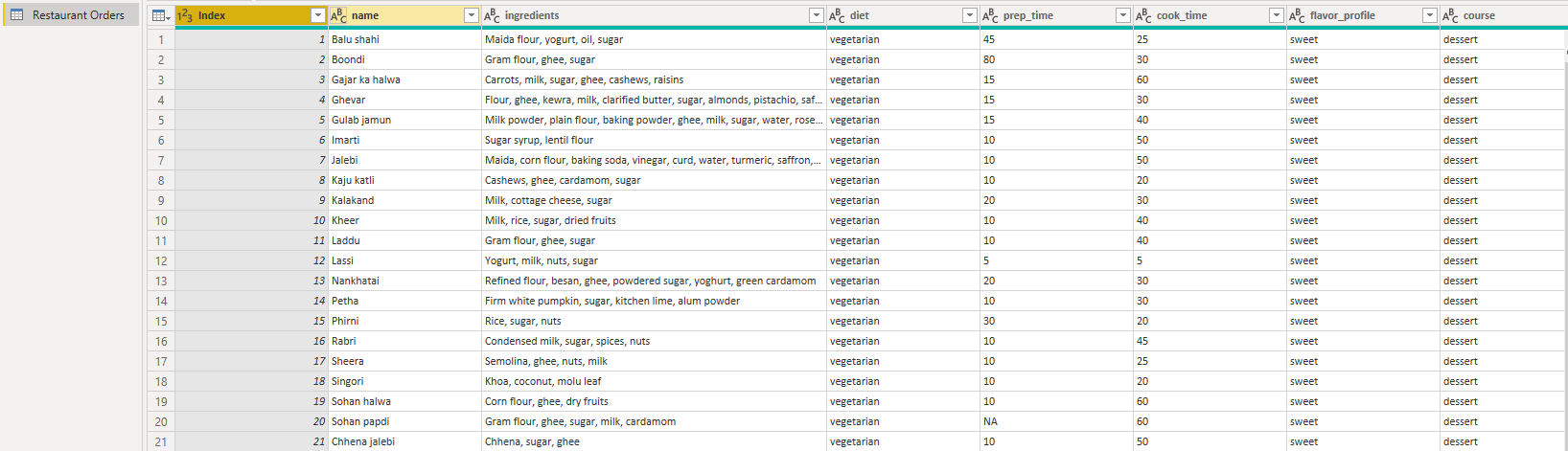
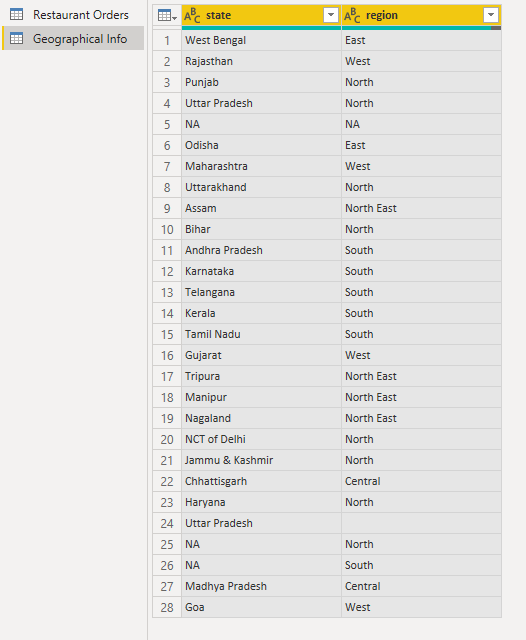
Below there is the set of tasks that you need to complete. Add the required screenshots in the table after the tasks.

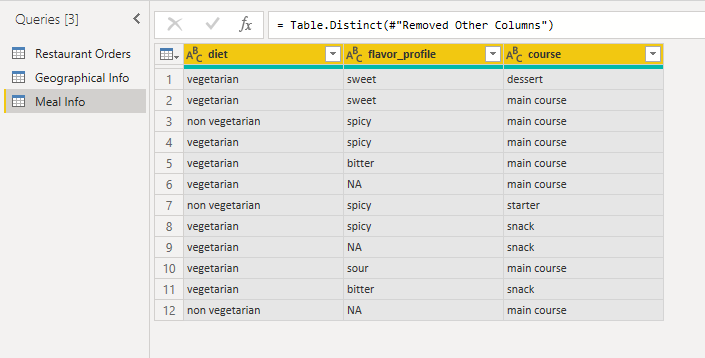
1. Load the data “restaurant\_orders.csv” file
2. Open the Power Query Editor (click on “Transform Data/Edit Queries”)
3. Use the right tool to promote\use first row as headers (see example below). If your Power BI version has already done this for you, then move to step 4



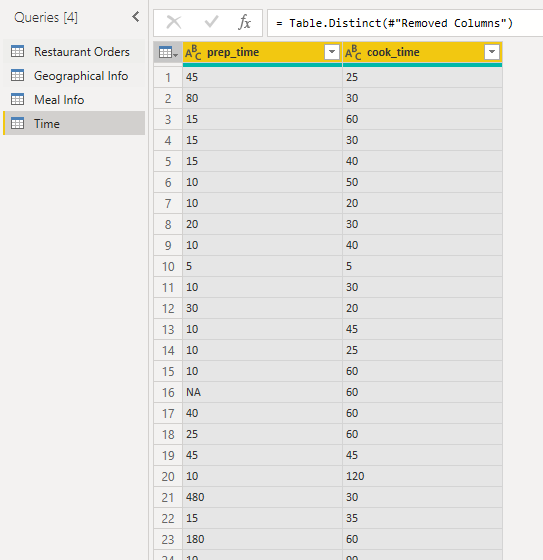
1. Add an index column as below (starting at 1). If you are not sure how to do this, you should Google “add an index column in Power BI”, select one of the results and follow the steps shown. Here is the new “Index” column.



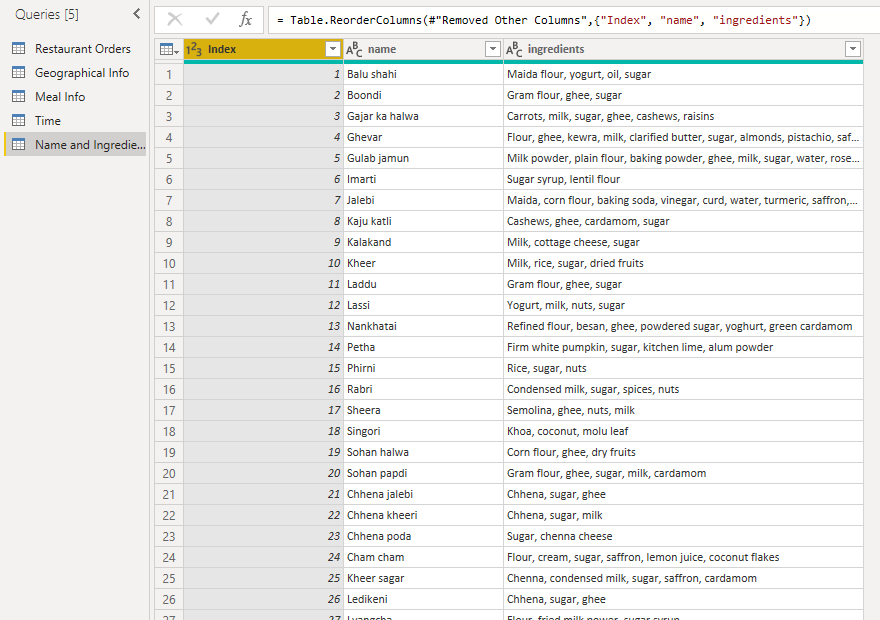
1. Move the column to the left of the table:
2. Now, use the duplicate function to split the big table into the smaller tables below. Review the steps covered in the lesson if you are not sure how to do this.
3. Geographical Info Table (note the length of the table…how would you remove duplicates from both columns simultaneously? Do you remember what key we use to select multiple columns?)
4. In the Geographical Info table, remove the two rows containing NA in the “State” column and the “blank” rows from the “region” column. The final table will have 24 rows.
5. Meal Info Diet (note the length of the table, no duplicates).

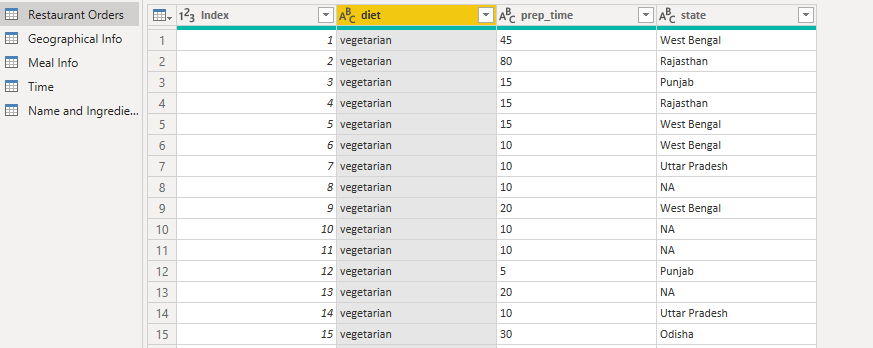


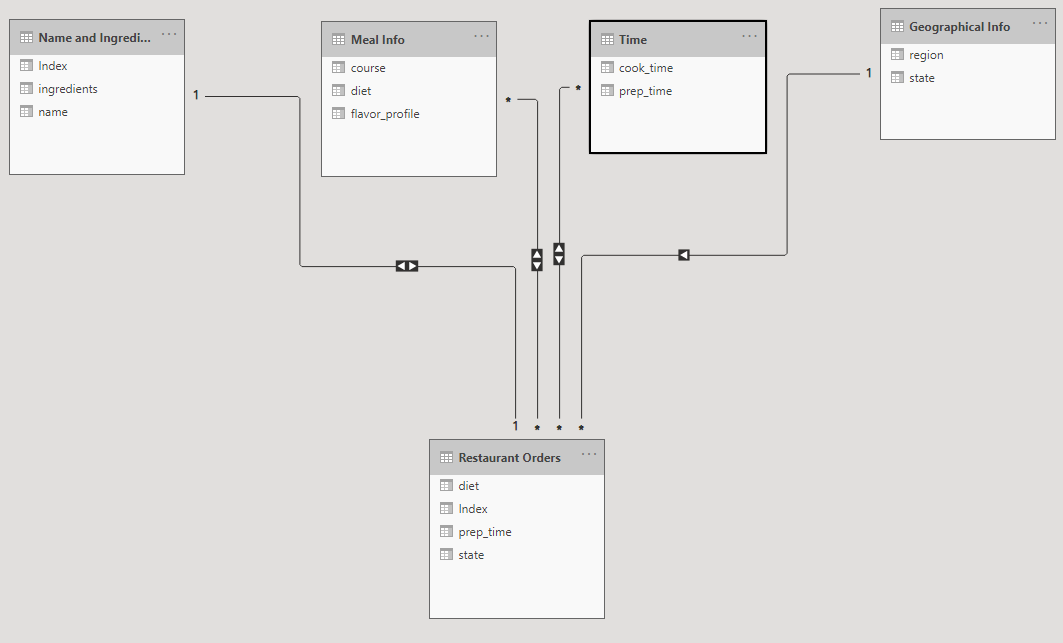
1. Time table (77 Rows)



1. Name and Ingredients Table (255 Rows)

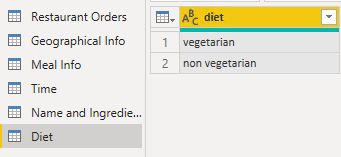


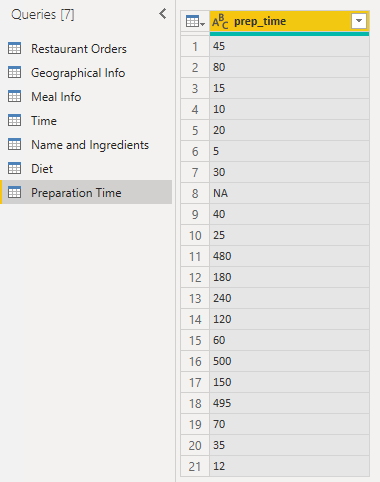
1. Finally, modify the Restaurant Orders table to show only the columns below (255 Rows) 
2. Remove the “NA” from the state column. The final table should have 231 rows.
3. Use the 5 tables to create the data model below. Review the steps covered in the lessons if you are not sure how to do this.



There are 2 many-to-many relationships that are not good…let’s create bridge tables to transform them into 1-to-many relationships. If you don’t remember how to do this, review the lessons or watch the Youtube video [here](https://www.youtube.com/watch?v=4zmSzDeiYt4&t=722s) (the whole Youtube video is relevant to data modelling, but from minute 13:00 is the part that you’ll need for this task).

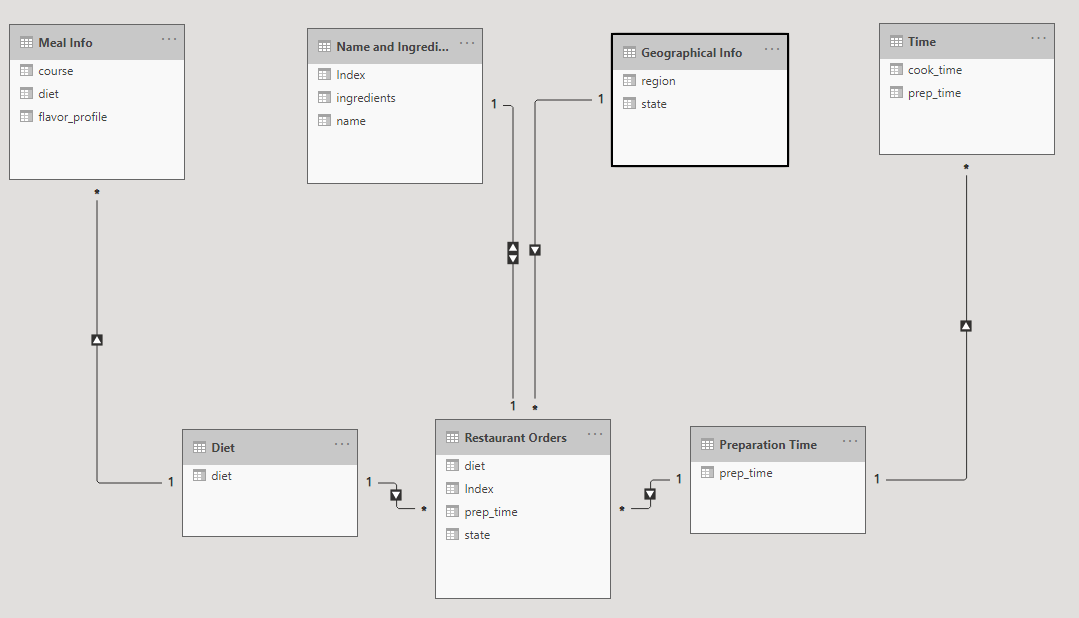
You should create the bridge tables below.

1.  Diet table(2 rows)



1. Preparation time table (21 rows).
2. Remove NA and duplicates from this table (20 rows)
3. Use the bridge tables and the correct relationships to obtain the model below.

All the many-to-many relationships have been replaced by many-to-1 relationships.

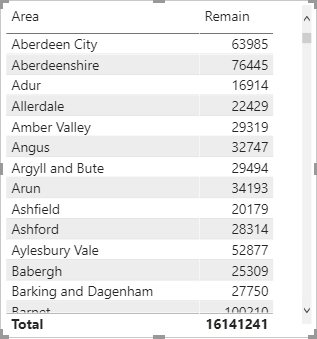


|  |  |  |
| --- | --- | --- |
| **Questions** | | **Provide a screenshot for the required questions** |
| 5 | Move the column to the left of the table |  |
| 6g | Remove the “NA” from the state column. The final table should have 231 rows. |  |
| 8 | Use the bridge tables and the correct relationships to obtain the model below. |  |
| *Independent Tasks* | | |
| A | Use the report view to create 3 charts to answer the questions below. |  |
|  | Use only the Restaurant Orders table to find which was the average preparation time by diet (vegetarian or non-vegetarian). You might need to change the data type of preparation time and remove NAs from that column, do you remember how to do this? | *The clustered column chart was used to analyse the average preparation time by diet using the Restaurant Orders table. This chart clearly compares vegetarian and non-vegetarian meals by showing the average preparation time for each category. The visualisation indicates that vegetarian meals have a higher average preparation time compared to non-vegetarian meals. Using a column chart makes it easy to observe differences between the two diet types and supports effective comparison.* |
|  | Use only the Geographical Info and Restaurant Orders tables to show the number of orders by state (the index column might be useful in this case). Which state had the highest number of orders? | *A bar chart was created using the Geographical Info and Restaurant Orders tables to show the number of orders by state. The index column was used as a count of orders. The chart shows that Punjab has the highest number of orders, followed by Gujarat. Bar charts are effective for comparing quantities across categories, and this visualisation clearly highlights which states have the most customer demand.* |
|  | Use only the Restaurant Orders table to find how many orders were vegetarian and how many were not vegetarian. | *A pie chart was used to display the proportion of vegetarian and non-vegetarian orders using the Restaurant Orders table. The chart shows that* ***non-vegetarian orders account for 90.55% of total orders****, while vegetarian orders represent a much smaller percentage. The pie chart effectively communicates the dominance of non-vegetarian meals by visually emphasising the share of each category.* |
| Arrange your tiles/charts to create a well-designed dashboard (use the knowledge covered in the lessons). Add a title, change the colours/formatting options of the charts.  Feel free to add more charts if you think you need more. If you are getting errors when creating new charts, this might be due to the relationships.  Once you are happy with the dashboard design, take a screenshot and paste it below. | | |
|  | | |
| Use the space below to explain how you did apply any of the theoretical aspects of data visualization covered in this module (for example, you could justify the selection of a specific chart or design choices for the scenario). Please provide three different points.  *1.*  *Column and bar charts were selected because they are suitable for comparing values across categories. The clustered column chart allowed an easy comparison of average preparation time between vegetarian and non-vegetarian meals, while the bar chart clearly showed differences in the number of orders across states. These chart types support quick understanding and accurate comparison, which aligns with best practices in data visualization.*  *2.*  *A pie chart was used to represent the proportion of vegetarian and non-vegetarian orders because it effectively shows part-to-whole relationships. This made it easy to identify that non-vegetarian orders form the majority of total orders. Using a pie chart in this context helps viewers quickly grasp the distribution without needing to interpret detailed numbers.*  *3.*  *The dashboard layout and formatting were designed to improve clarity and readability. Consistent colours, clear titles, and proper spacing between visuals were applied to reduce visual clutter and guide the viewer’s attention. These design choices follow data visualization principles that emphasise simplicity, consistency, and ease of interpretation.* | | |

# Data Analytics Task 3: DAX and M Language (20 marks)

In this task, you will be required to apply DAX and M Language.

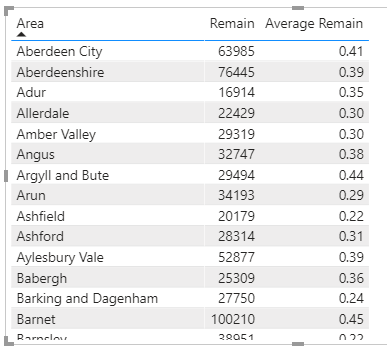
1. Download the Excel workbook (Voting Figures) from Blackboard and import it on Power BI. This file contains information about the Brexit referendum votes. You can find more info about the Electoral Commission source data from a link [here](https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/elections-and-referendums/past-elections-and-referendums/eu-referendum/results-and-turnout-eu-referendum).
2. Under report view, use the table visual to create the table below:



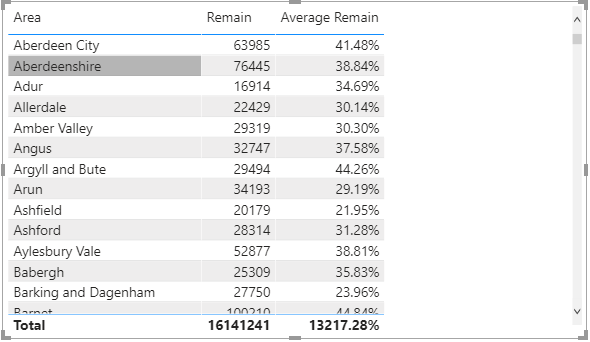
1. Create a new Measure “Average Remain” to work the proportion of Remain votes out of the total electorate. Proportion means division…so your formula should include a DIVIDE operator (“/”) to show the total remain vote (SUM) divided by the total (SUM) electorate for each area.

*The measure should be a simple calculation of the form***SUM(X) / SUM(Y)***.*

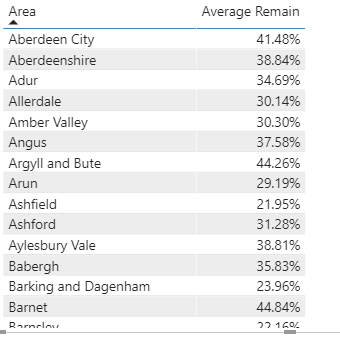
1. Add the measure to the table. Here's what the final table should look like.



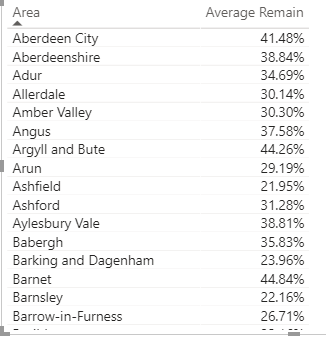
1. Show the average Remain column as a percentage (2 decimal places) as shown below.



1. Remove the Remain column.
2. Remove the Total values at the bottom of the table, since it’s not a meaningful value in this context. The new table is shown below.

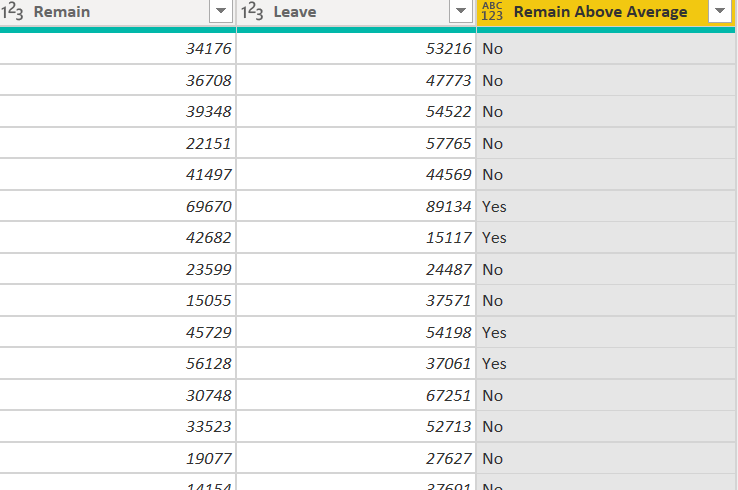


1. Change the “Style” of the table into “None”. The new table is shown below



1. Go on Transform Data and use M Language to add a new Custom Column “Remain Above Average”.

If the value in the Remain Column is higher than or equal to 42254 (this is the average of that column and I have already calculated for you), then the column should display “Yes”, else display “No”. The new column should look like the one below.



1. Look for the advanced Editor to find the M Language associated with the step you have just performed in Step 9. It should look similar to the one below.

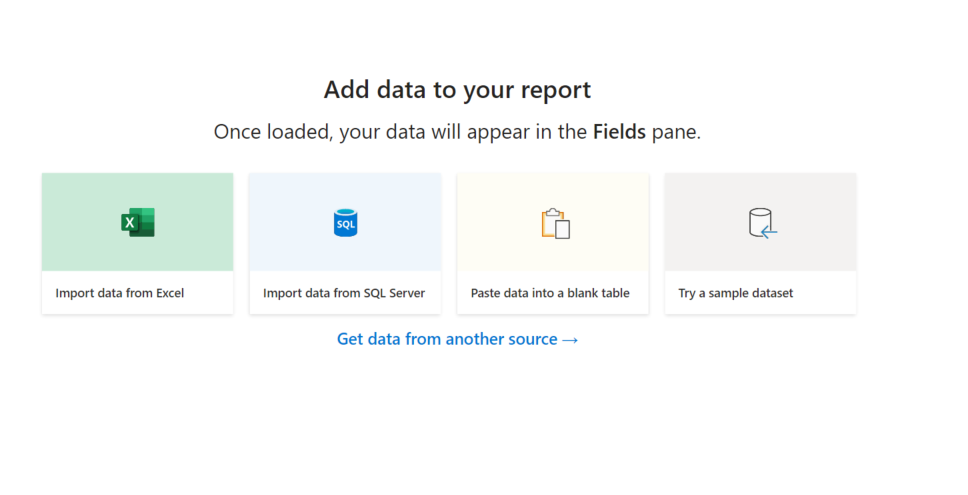


|  |  |  |
| --- | --- | --- |
| **Questions** | | **Provide a screenshot for the required questions** |
| 2 | Under report view, use the table visual to create the table below | (add here a screenshot of the table including the first 20 rows) |
| 4 | Add the measure to the table. | (add here a screenshot of the table including the first 20 rows) |
| 7 | Remove the Total values at the bottom of the table, since it’s not a meaningful value in this context. | (add here a screenshot of the table including the first 20 rows) |
| 8 | Change the “Style” of the table into “None”. | (add here a screenshot of the table including the first 20 rows) |
| 9 | Go on Transform Data and use M Language to add a new Custom Column “Remain Above Average”. | (add here a screenshot of the instruction used to create the column) |
|  |  | (add here a screenshot of the new column including the first 20 rows) |
| 10 | Look for the advanced Editor to find the M Language associated with the step you have just performed in Step 9 | (add here a screenshot of the Advanced Editor, Highlight/Explain which line corresponds to the step performed in question 9) |

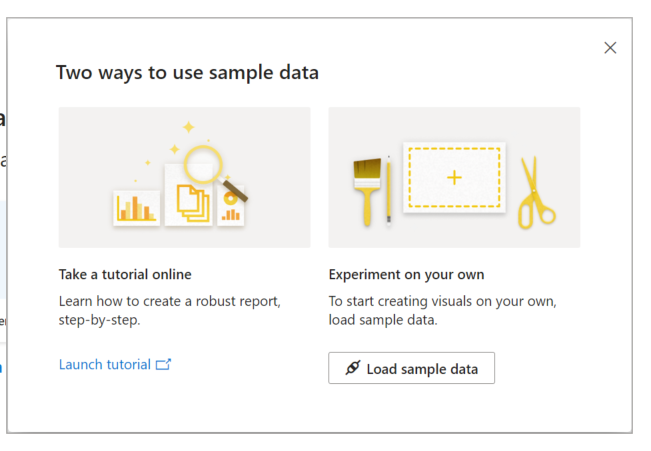
# Data Analytics Task 4: Dashboard Design (25 marks)

In this task, you will be required to perform some data analytics steps to analyse a financial dataset.

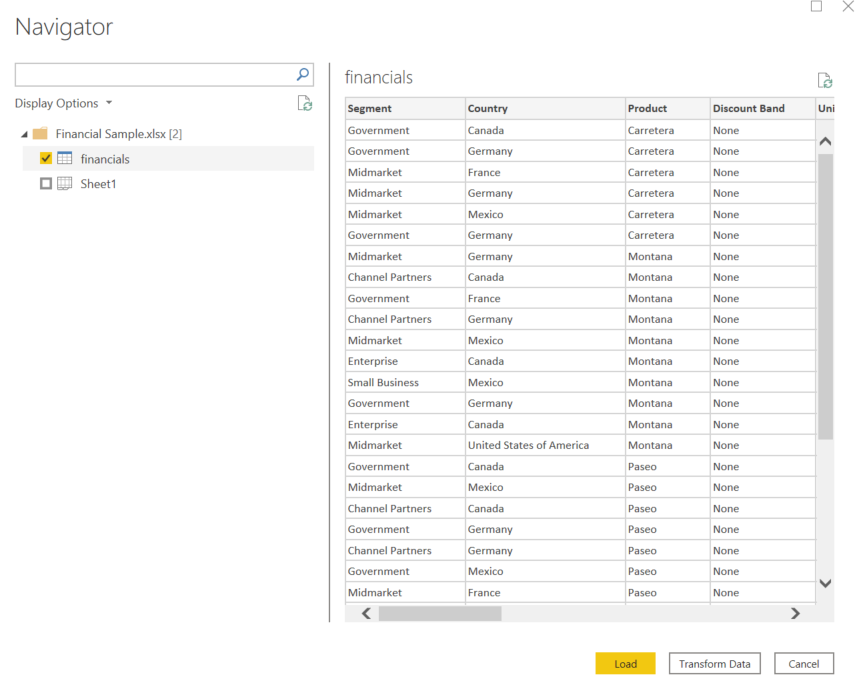
To complete the tasks below, open Power BI and select “Try a sample dataset”.



Then select “Load sample data”.



Now, load the financials table into Power BI



Create a Power BI Dashboard to analyse the data and complete the tasks below.

You can use any charts/tools to answer the questions. For this task, you are not required to work on the data (using Transform Data) or change the data model.

For each task, add the answer and a screenshot of the chart that you used to answer the question (use the Snipping Tool to select only the chart that contains the answer). Feel free to use Google if you need help and or you want to explore new tools/charts.

|  |  |  |
| --- | --- | --- |
| **Business Questions** | | **Answer and Screenshot to support your answer**  **Why do you think this was a suitable chart?** |
| 1 | Which product had the highest profit? |  |
| 2 | How many different products are in the dataset? |  |
| 3 | What was the gross sale associated with the “Paseo” product? |  |
| 4 | How many units were sold in 2014? |  |
| 5 | How many units were sold in May (over the two years)? |  |
| 6 | Which month had the highest number of units sold? |  |
| 7 | How many different countries are in the dataset? |  |
| 8 | Which two months had the lowest number of units sold? |  |
| 9 | Create a chart to show the profit of the different countries. |  |
| 10 | When was the first product sold? |  |
| 11 | On which product would you invest in the next year? Why? |  |
| 12 | What was the total “gross sales” over the two years? |  |
| 13 | Arrange your tiles/charts to create a well-designed dashboard (use the knowledge covered in the lessons). Add a title, change the colours/formatting options of the charts. Once you are happy with the dashboard design, take a screenshot and paste it below. | |
| *(Add here the screenshot of the final dashboard)* | | |
| Use the space below to explain how you did apply any of the theoretical aspects of data visualization covered in this module (for example, you could justify the selection of a specific chart or design choices for the scenario). Please provide three different points.  1.  2.  3. | | |

Once you have provided a screenshot for the tasks above, save this file as a PDF file and submit it together with the Power BI files that you have used for the tasks.