**Exercise: Root cause analysis**

**Introduction**

Proficiency in data analysis is a versatile skill that extends beyond corporate requirements. Having honed your analytical skills with Adventure Works data, you decide to put them to the test in a practical scenario. While exploring the endless array of available online datasets, you came upon one containing information on a critical matter: the environmentally pressing issue of CO2 vehicle emissions.

In this exercise, you will embark on a practical scenario of building an insightful report from an actual dataset. You will use the capabilities of specialized visualizations like key influencers and decomposition trees to add impact to your report and swiftly recognize the hidden patterns behind raw information.

**Case study**

Environmental issues are on everyone's mind, and vehicle CO2 emissions stand out as a major contributor. A dataset filled with raw data about CO2 emissions per vehicle holds a wealth of valuable insights waiting to be uncovered through data analytics. In this case study, you’ll focus on a targeted investigation: using root cause analysis to identify the vehicle attributes that have the greatest impact on air pollution.

To do this, you will have to:

* Identify the key vehicle attributes that have the most impact on CO2 emissions in the environment.
* Identify the major contributors to pollution through a perceptive root cause analysis, leveraging the AI-driven visualizations of Power BI.
* Create relevant visualizations to enrich your report, including a decomposition tree, allowing users to freely explore and navigate the dataset's values.

**Instructions**

**Step 1: Download the report**

* Download the Power BI report file *CO2 emissions by vehicle* and open it in Power BI Desktop.

[CO2 emissions by vehicle](https://d3c33hcgiwev3.cloudfront.net/L9lgyxAARGiKmI2O6_gMRw_262d33ddd385463fa75a588c710878a1_CO2-emissions-by-vehicle.pbix?Expires=1712102400&Signature=PUiqhFm3ExXC6W5x2Nze3YD73KTbvadxFZTZB-9lkVH~WkRqp0rTNDIZMOYhBfP~2xxNp0~sZWv6PfSWHMTscTS0F-vYLpKDyBP8BSd6OK6cjuIEBns5y5TJ5mv70PdCeTY7xwUJFN~rMChAEl43hLUd7DorNn0RthJ8cCGv4aY_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[PBIX File](https://d3c33hcgiwev3.cloudfront.net/L9lgyxAARGiKmI2O6_gMRw_262d33ddd385463fa75a588c710878a1_CO2-emissions-by-vehicle.pbix?Expires=1712102400&Signature=PUiqhFm3ExXC6W5x2Nze3YD73KTbvadxFZTZB-9lkVH~WkRqp0rTNDIZMOYhBfP~2xxNp0~sZWv6PfSWHMTscTS0F-vYLpKDyBP8BSd6OK6cjuIEBns5y5TJ5mv70PdCeTY7xwUJFN~rMChAEl43hLUd7DorNn0RthJ8cCGv4aY_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

**Step 2: Identify the key influencers of CO2 emissions**

To commence your analysis, you'll take an initial look at the dataset you'll be working with, acquainting yourself with its values and attributes.

1. Navigate to the **Data** view of Power BI. Select the **Vehicles** table to familiarize yourself with the dataset.
2. Identify the **CO2 emission** column for analysis. Consider all attributes associated with it that could contribute to an increase or decrease in emissions.
3. Identify the column that won't contribute to specialized visualizations and thus can be excluded.
4. Navigate back to the **Report** View to initiate the creation of your report.

**Step 3: Report Creation**

To create the report, add a simple card visual to highlight the count of vehicles in your dataset. Introducing some general information about the dataset is a common practice in reporting. Then, with the help of the Key influencers' visual, start creating insightful charts in your report.

1. Create a basic card visual with the number of vehicles on the dataset, renaming it to *Vehicles analyzed*, to showcase the volume of the dataset.
2. Create the **Key influencers** visual to assist you in deriving insights from the dataset.
3. Add the appropriate column in the **Analysis** field, and all potential influencing factors as its attributes.
4. Explore each **Key influencer** calculated by the specialized visualization.
5. Notice the insightful scatter plot in the generated visualizations and recreate it within your report.

**Step 4: Use the Top segments tool to detect groups of influencing factors**

To understand if there is a potential group of influencing factors behind the CO2 emissions, use the **Top segments** tool of the visualization.

1. Navigate to the **Top segments** tool of the visualization.
2. Explore the top segments, identifying the main groups of attributes behind air pollution.
3. Observe an instance where a single factor holds such importance that it creates its own segment. This observation should validate the significance of the previously created scatter plot.
4. In the remaining segments, notice that apart from the field already visualized, there are two additional attributes creating segments and affecting the emissions.
5. Create a visualization highlighting the relationship between those two factors with emissions. Use the higher cardinality (number of distinct values) column as an axis, and the lower cardinality column as the legend.
6. Recognize the third attribute influencing the emissions that is not currently highlighted in the report. Select a visualization of your choice and add it to the report.

**Step 5: Build a decomposition tree with AI capabilities**

Having built a number of visualizations for your report, it is time to give your users the ability to explore the dataset freely with a specialized visualization. On a new page, add a decomposition tree and work with its data.

1. Create a new page on your report and select the decomposition tree visualization. Adjust it to fit the whole screen.
2. Add the average of the emission field and all its relevant attributes on their respective visualization fields.
3. Expand all attributes in a hierarchical function. You should notice that a specific field cannot be easily used in the decomposition tree due to its high cardinality.
4. Create 5 bins to group the column mentioned above into equal-sized groups. Remove the ungrouped column from the chart and add the newly created bins.
5. Using the power of AI analysis of the visualization, identify what is the lowest average CO2 emission on a vehicle with powertrain: Hybrid Electric Vehicle (HEV), taking into account all lowest emission scoring attributes.

**Conclusion**

The aim of this exercise was to underscore the invaluable role of data analysis in addressing pressing environmental issues. Throughout this activity, you used the capabilities of specialized visuals within Power BI to construct an enlightening report, employing authentic case study data. With the remarkable support of AI-powered visualizations, you crafted a captivating report that would captivate anyone concerned with the realm of CO2 emissions.

**Exemplar: Root cause analysis**

**Introduction**

In the exercise, *Root cause analysis,* you were assigned a real case scenario of actual environmental data to hone your analytical skills on Power BI reporting. With the immense help of Power BI specialized visuals on your side, your goal was to craft an insightful report, leveraging the significance of all available information in the dataset.

Your specific tasks were to:

* Create and explore the **Key influencers** visualization to identify the driving forces behind the CO2 emissions field.
* Use the information provided by the **Key influencers** visual to add insightful visualizations to your report.
* Create a decomposition tree to allow users to navigate themselves through the dataset.

This reading provides you with a detailed guide that you can use to compare your solution.

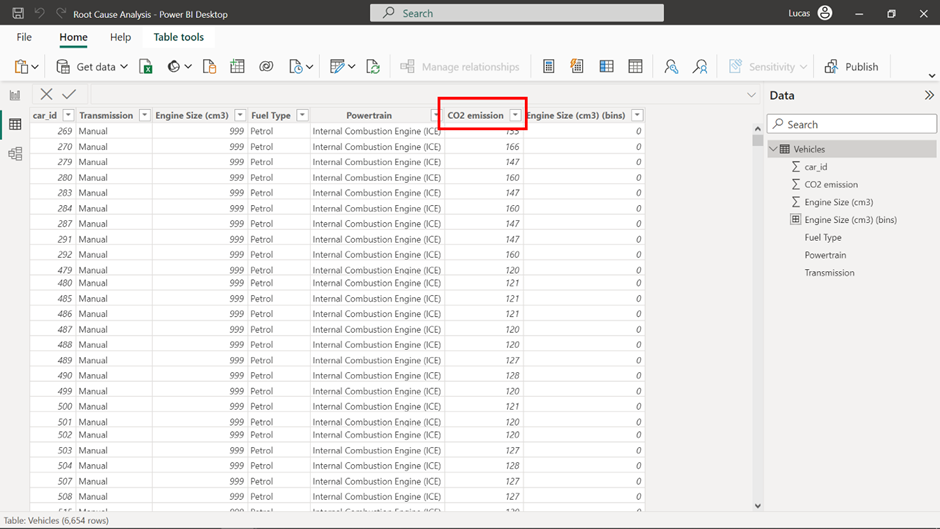
**Instructions**

**Step 1: Download the report**

* Download the Power BI report file *CO2 emissions by vehicle* and open it in Power BI Desktop.

**Step 2: Identify the key influencers of CO2 emissions**

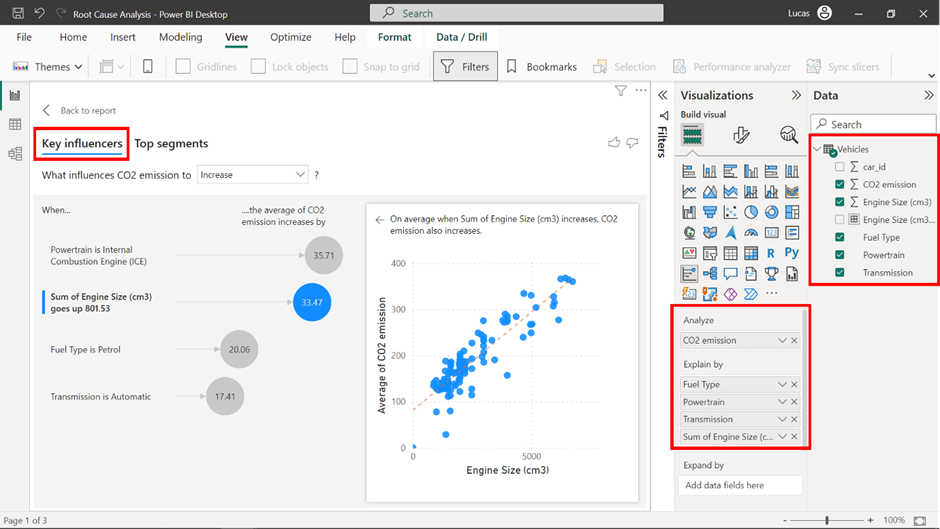
1. Navigate to the **Data** view of Power BI. Select the **Vehicles** table to identify all necessary columns to be used on the specialized visuals.
2. The **CO2 emission** field is the main column to be analyzed. **Transmission**, **Engine Size (cm3)**, **Fuel Type**, and **Powertrain** are the attributes that might explain the increase or decrease in a vehicle’s CO2 emissions.



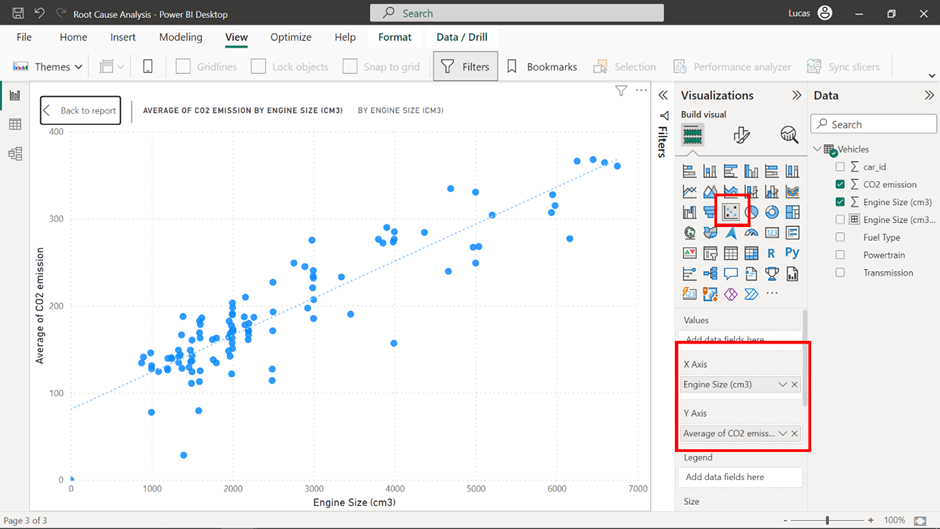
1. The **car\_id** column serves as an index column for the table and holds no specific analytical value in specialized visuals.
2. Navigate back to the **Report** view to initiate the creation of your report.

**Step 3: Report creation**

1. Create a basic card visual and add the **car\_id** column from the table with the **Count** function. Keep in mind that because each table row indicates a distinct vehicle and holds no blank values, any column with the **Count** function can be used, delivering the same result. Double-click on the field name to rename it to **Vehicles analyzed**.
2. Locate and select the **Key influencers** visual from the **Visualizations** pane.
3. Add the **CO2 emission** column to the **Analyze** field and all other attribute columns (excluding **car\_id**) in the **Explain by** field.
4. Explore each key influencer calculated by the specialized visualization.

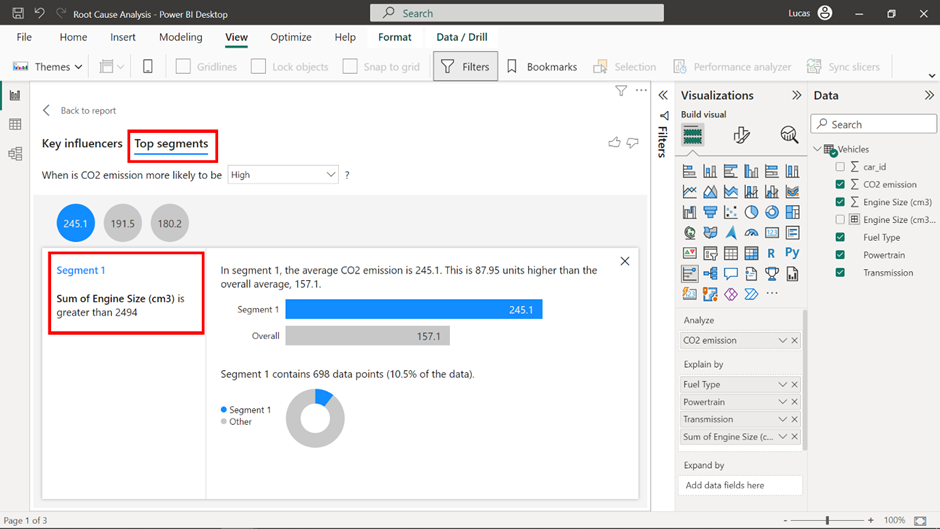


5. Note the insightful scatter plot about the relationship of **Engine size (cm3)** with **Average of CO2 emission** in the generated visualizations. Recreate it in the report by selecting a scatter plot and adding the columns mentioned above in the **X-axis** and **Y-axis** respectively. The **Y-axis** will default to **Sum of CO2 emission**, so be sure to change it to **Average** using the down arrow symbol at the right side of the field.

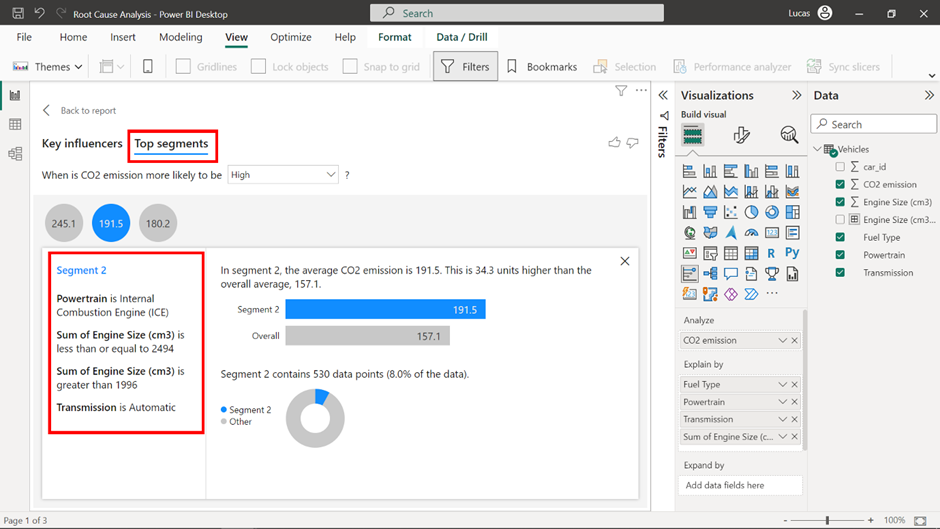


**Step 4: Utilize the Top segments tool to detect groups of influencing factors**

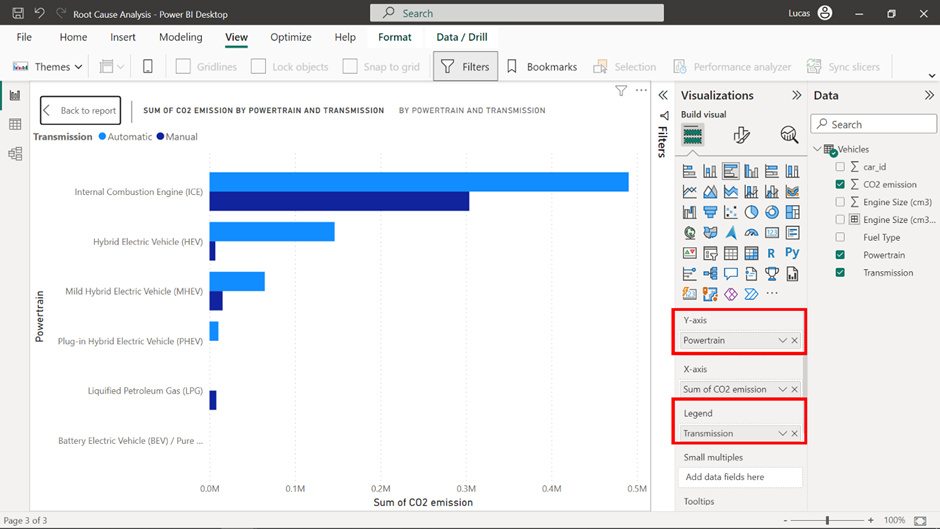
1. Navigate to the **Top segments** tool of the visualization. It is the second tab on the upper side of the visualization.
2. Explore the top segments, identifying the main groups of attributes behind air pollution.
3. Note that **Sum of Engine Size** is a segment by itself as it heavily influences the CO2 emissions of the vehicle. This observation confirms the significance of the scatter plot that was added in the previous step.



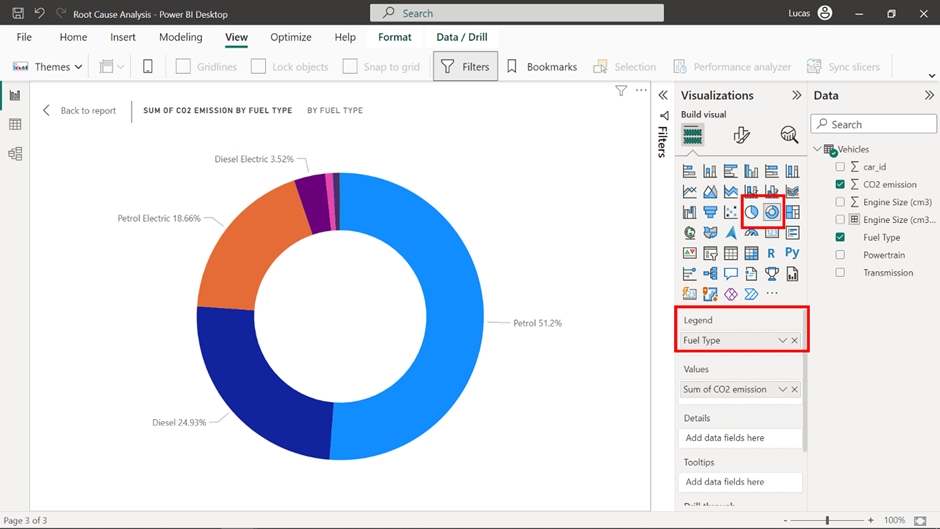
4. In the remaining segments, notice that apart from the field already visualized, **Powertrain** and **Transmission combined** are important factors when it comes to CO2 emissions.



5. Create a visualization highlighting the relationship between these two factors with emissions. Use **Powertrain** in the **Y-axis**, as it has a higher cardinality than **Transmission**, which can be added to the **Legend**.

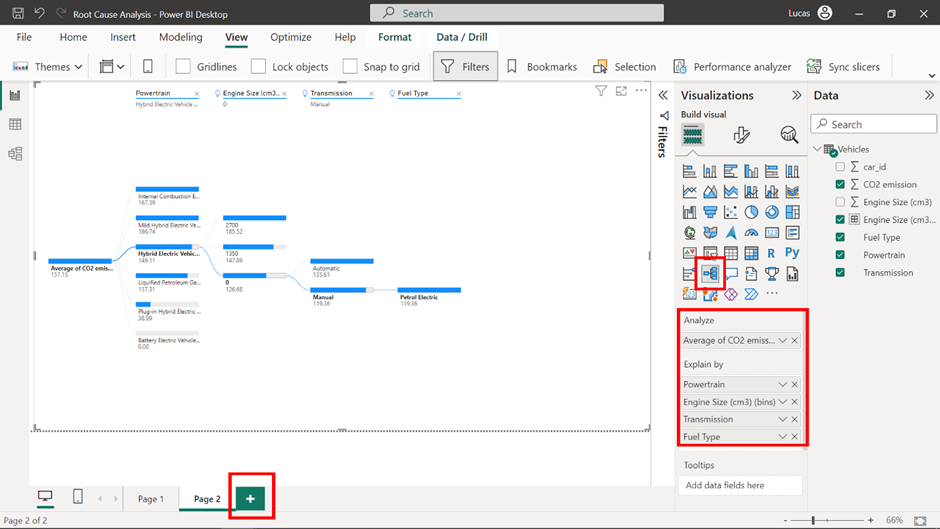


6. The **Fuel type** field has not been included in the report. Identifying the three categories of fuel types that consist of more than 90% of CO2 emissions. A pie or donut chart would be a fitting visualization for this field.



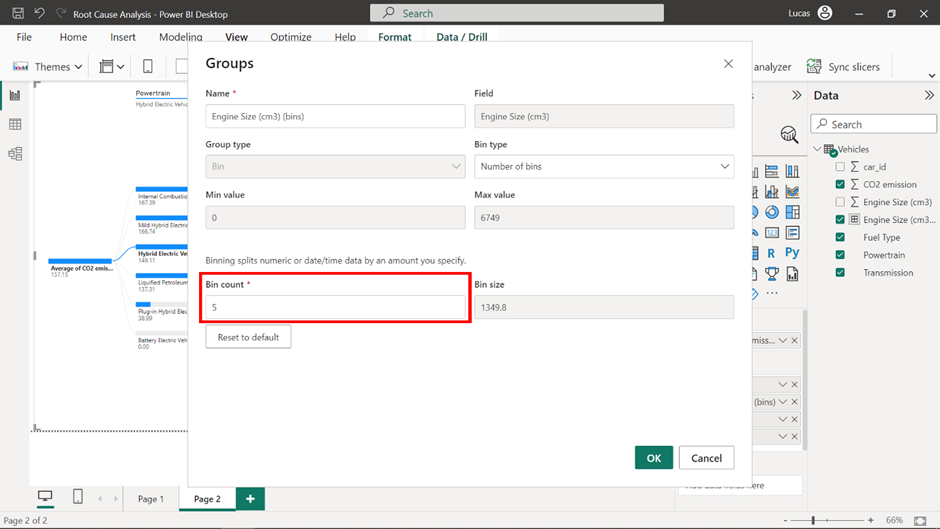
**Step 5: Build a decomposition tree with AI capabilities**

1. Select the **+** symbol on the bottom of the canvas to create a new page on your report, and then select the decomposition tree visualization in the **Visualizations** pane. Adjust the visual to fit the whole screen.
2. Add the average of the **CO2 emission** in the **Analyze** field and all its relevant attributes in the **Explain by** field.

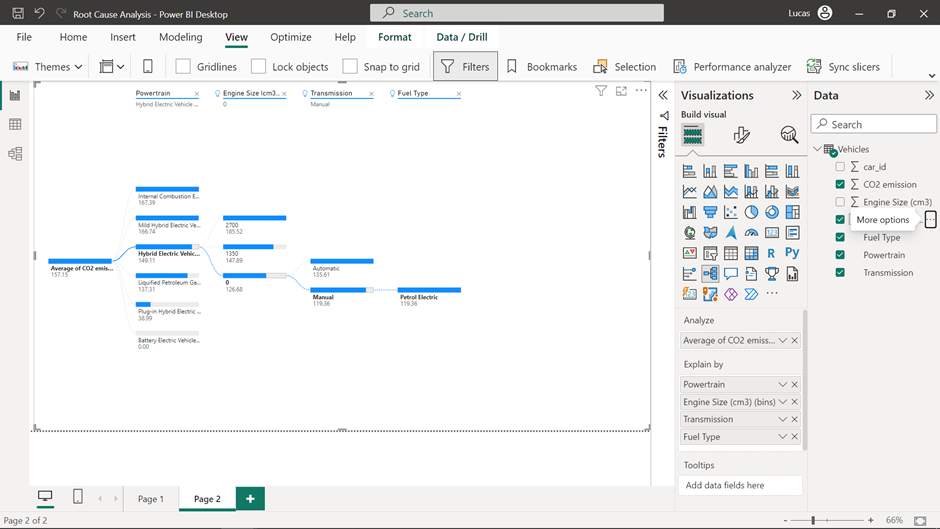


3. Expand all attributes in a hierarchical function using the plus sign at the right side of each bar. **Engine size (cm3)**, having a high cardinality, cannot be successfully visualized in the decomposition tree.

4. Right-click on the column **Engine size (cm3).** Select **New group** to group the data points into equal-sized bins by selecting five as the number of bins. Remove the ungrouped column from the chart and add the newly created bins.



5. Using the power of AI analysis on the visualization, identify what is the lowest average CO2 emission on a vehicle with **Powertrain: Hybrid Electric Vehicle (HEV)**. To do this, close all previously opened tabs of the visual, and navigate using **Power train** as the first hierarchy selection, and then **Low value** on the rest of the attributes to leverage the AI functionality of the visualization.



**Conclusion**

By completing this exercise, you explored how to create an accessible report using the capabilities of specialized visualizations in Power BI, like key influencers and decomposition trees to add impact to your report and swiftly recognize the hidden patterns behind raw information. With the AI visuals of Power BI as your tools, you build a report from scratch, issuing a definitive presentation of vehicle attributes’ association with air pollution from CO2.