Perform Analytics in Power BI

Explore Statistical Summary

Data is often intertwined with statistics because statistics are one way in which you can explore your data

Statistics show you the **distribution of your data and help you identify key takeaways and trends and determine whether outliers exist**

Identify Outliers in Power BI

An **outlier** is a **type of anomaly in your data**, something **you didn't expect or that surprised you**, based on historical averages or results

You will want to **identify outliers to isolate data points** that **significantly differ from other data points**

But first you need to constitute what is an outlier

Process of **identifying outliers involves segmenting your data into two groups:**

- one group – outlier data
- other group – non-outlier data

The best way to **identify outlier groups**:

- **Use a visualization**
- **DAX formula**

These methods will ensure that **results are dynamic**

Use a Visual to Identify Outliers

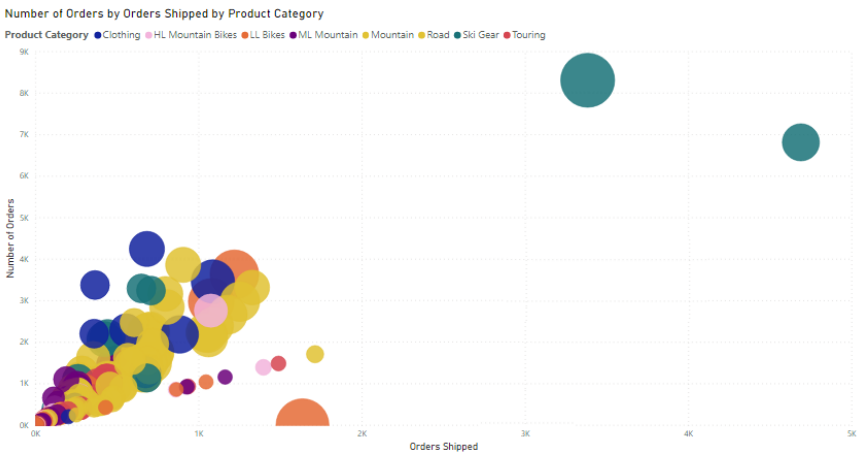
The **best visual** for **identifying outliers** is a **scatter chart**

It **shows the relationship between two numerical values**

Scatter charts display **patterns in large sets of data** and are, therefore, **ideal for displaying outliers**

The visual will update and display the data according to the selected fields, and you'll be able to **clearly identify the outliers in the data** – they are **isolated items** that are **away from the bulk of the data**

An example of a visual depiction of an outlier can be seen below (**the outlier is the data that is isolated from the bulk/cluster of data**)



Use DAX to Identify Outliers

You can also **use DAX** to **create a measure** that will **identify the outliers in your data**

An example **DAX** formula:

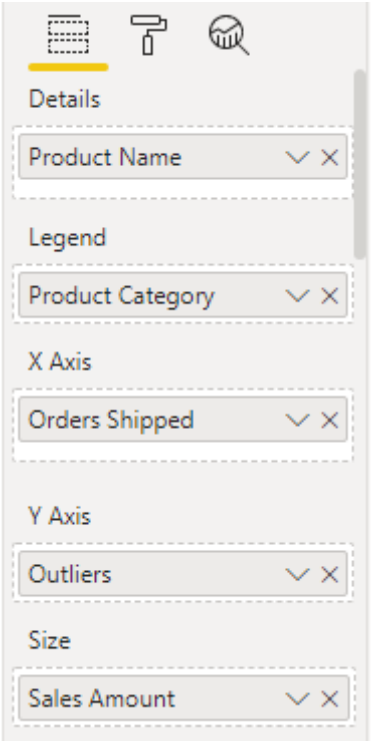
```
Outliers =
CALCULATE (
        [Order Qty],
        FILTER (
                VALUES (‘Product’[Product Name]) ,
                COUNTROWS (FILTER (‘Sales’, [Order Qty] >= [Min Qty])) > 0
        )
)
```

In the **DAX expression** above, **Order Qty** is a **measure** in the Sales table and **Min Qty** refers to the **lowest order quantity in the Sales table**

Once you have **created the outlier measure in DAX**.

You can **add it into a scatter chart visual** as it is **the best way of depicting an outlier in your data**.

Populate the **scatter chart** with the **fields associated with your DAX and the outlier measure**, as seen in the image below:



Group and Bin Data for Analysis

When creating visuals,

Power BI **aggregates your data into groups**, based on the values that it finds in the underlying data

You can **refine how those default groups are presented**

You can also **create new groups:**

by grouping two or more data points - **grouping**
putting values into equal-sized groups - binning

Grouping and Binning ensures that the **visuals** in the reports **display the data according to your preference**

Using these features will help you to **clearly view, analyse, and explore the data and trends in your visuals**

Additionally, you'll be able to **clearly view, analyse, and explore the data and trends in your visuals**

This will then provide your users with **more specific insights on their data**, which can **help drive business decisions**

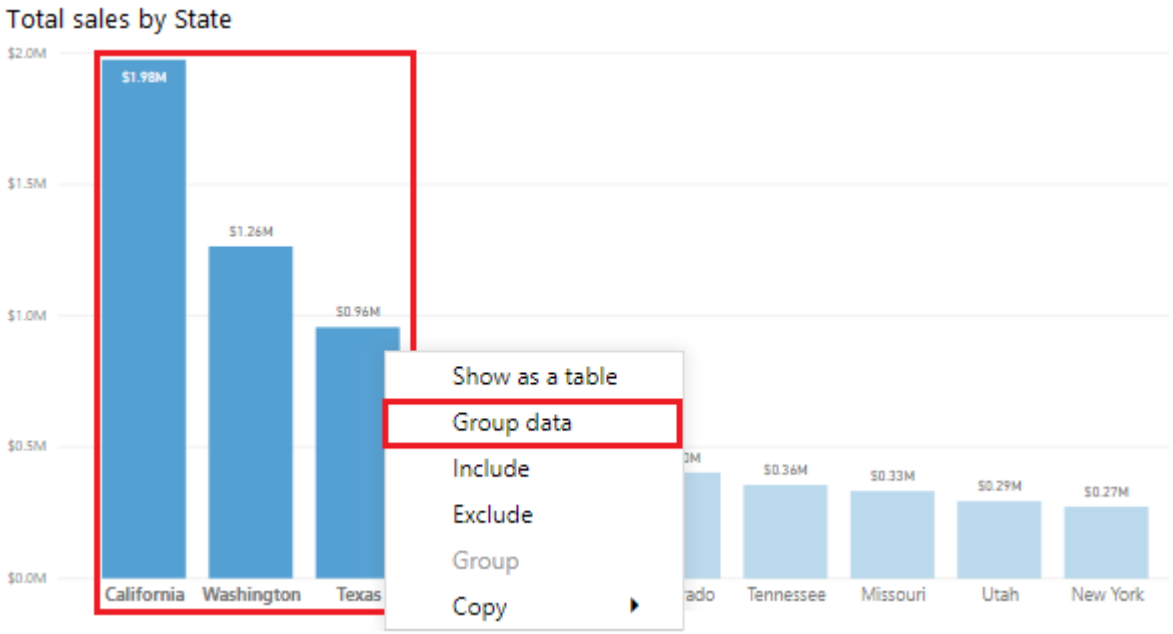
Grouping

Grouping is mainly used for **categories of data**

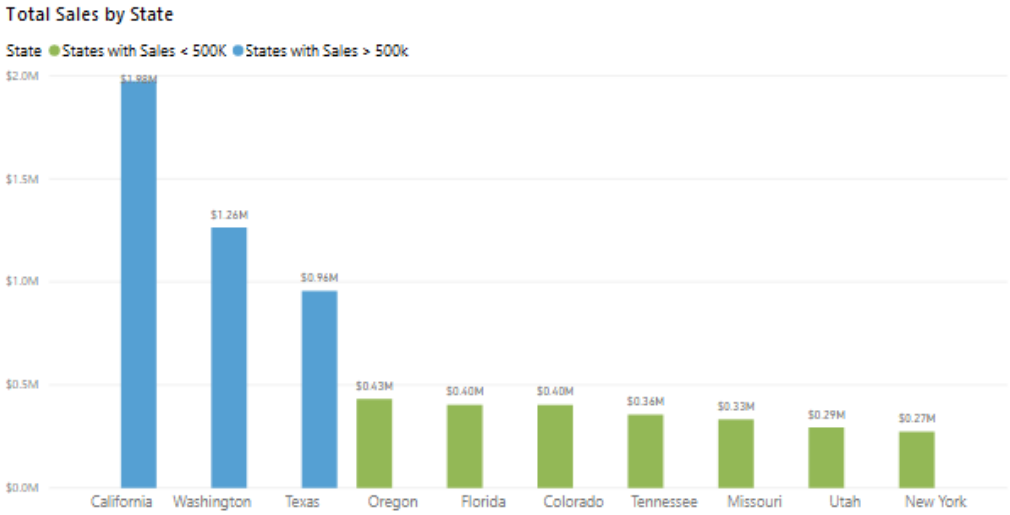
Create a Group

This section is best depicted by an example, e.g., assume you that **you want to group states with sales greater than 500,000 dollars**.

You the **Ctrl + click** the **data points that you want to group**, **Right-click** one of those selected data points and then **select the Group data option**



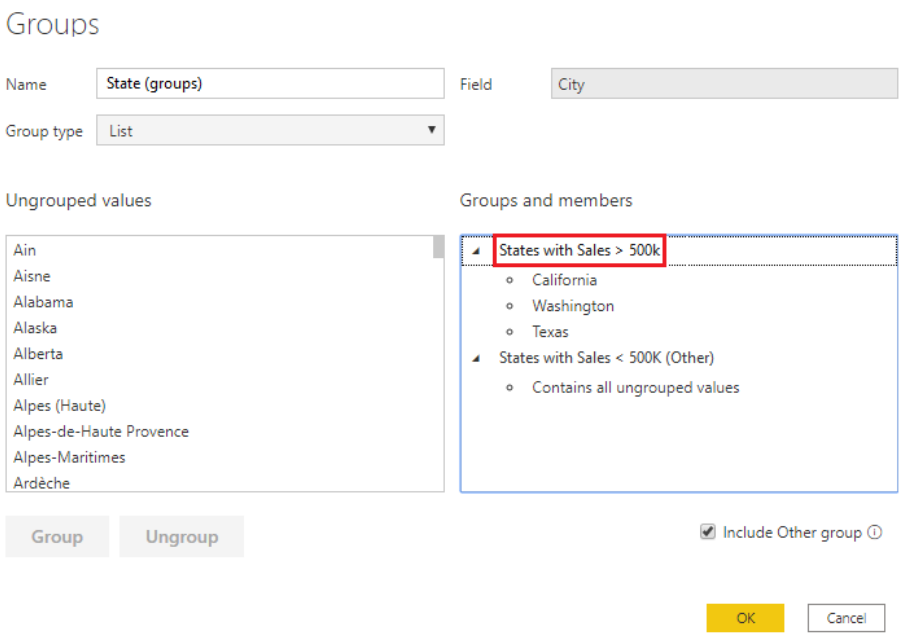
We then get the following visual after **grouping the above visual**



Edit a Group

You can also edit a previously-made group (e..g, the states with sales greater than $500k

This can be done by **right-clicking the group field** in either the **Legend bucket** or the **Fields pane** and then select the **Edit Groups**



Values in groups need to be **manually configured**

This means that, **any new items/values that appears** will be **automatically placed in the (Other) ungrouped section** of the graph

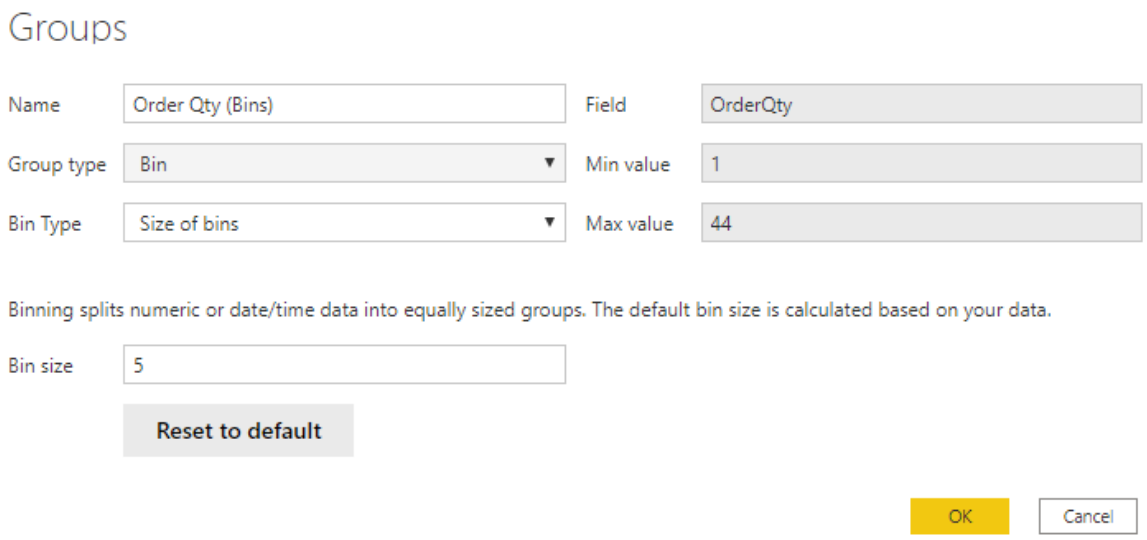
Hence, **you will need to manually place values/items into the created group**

Binning

Binning is similar to grouping, but is mainly used for **grouping continuous fields**, such as **numbers and dates**

Allows you to group numerical and time field data into **"bins" of equal size**

Allows you to **visualize and identify trends in your data in more meaningful ways**



This is again best explained with a visual

In the visual above, we want to **create bins (groups)** for the **Order Qty field.**

First start in the **Fields pane**, by **right-clicking the Order Qty field** that you want to create the bins for, and then **select New Group**
On the **Groups window** that appears, **set the Bin size** to the **size that you want**, adjust other settings are required and **select OK**

The values are then **separated/split into the equally sized groups** whose **total number of bins is USER-DECIDED** by the **value in the Bin size field**

Apply Clustering Techniques

Allows you to **identify a segment (cluster) of data** that is **similar to each other** but **dissimilar to the rest of the data**

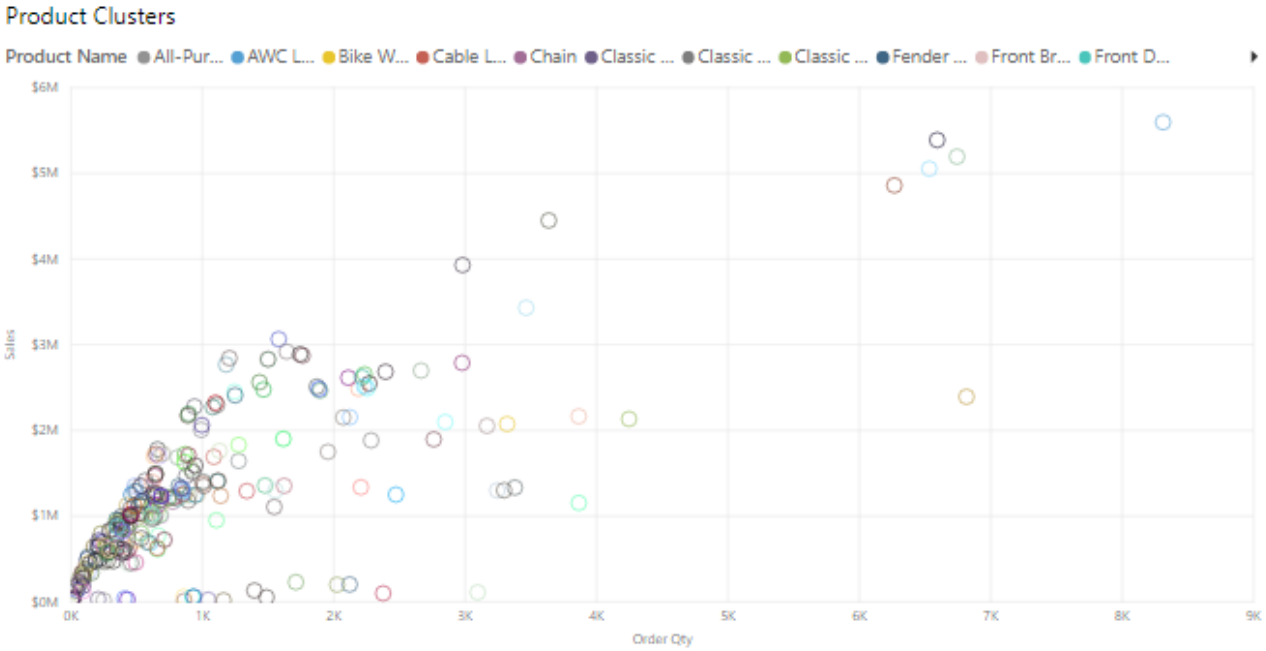
Clustering is different to the grouping technique accomplished previously

Clustering allows you to **quickly find groups of similar data points in a subset of your data**

It **analyses your dataset to identify similarities and dissimilarities** in the attribute values

and then it **separates the data that has similarities** into a **subset of the data**. These **subsets are referred to as clusters**

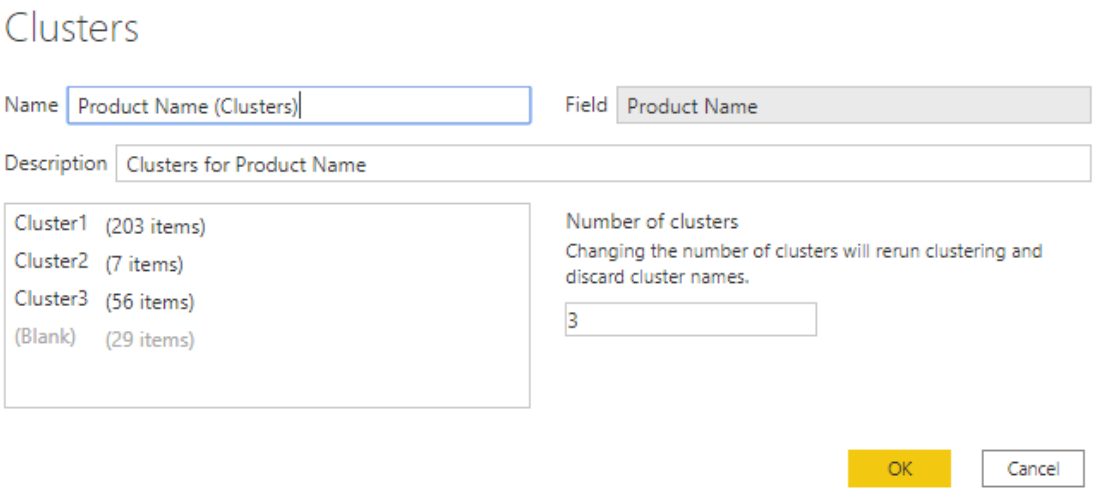
Take an initial unclustered dataset



Simply from looking, **it is difficult to discern any natural groups** – we can then apply **clustering to find groups of similar data points**

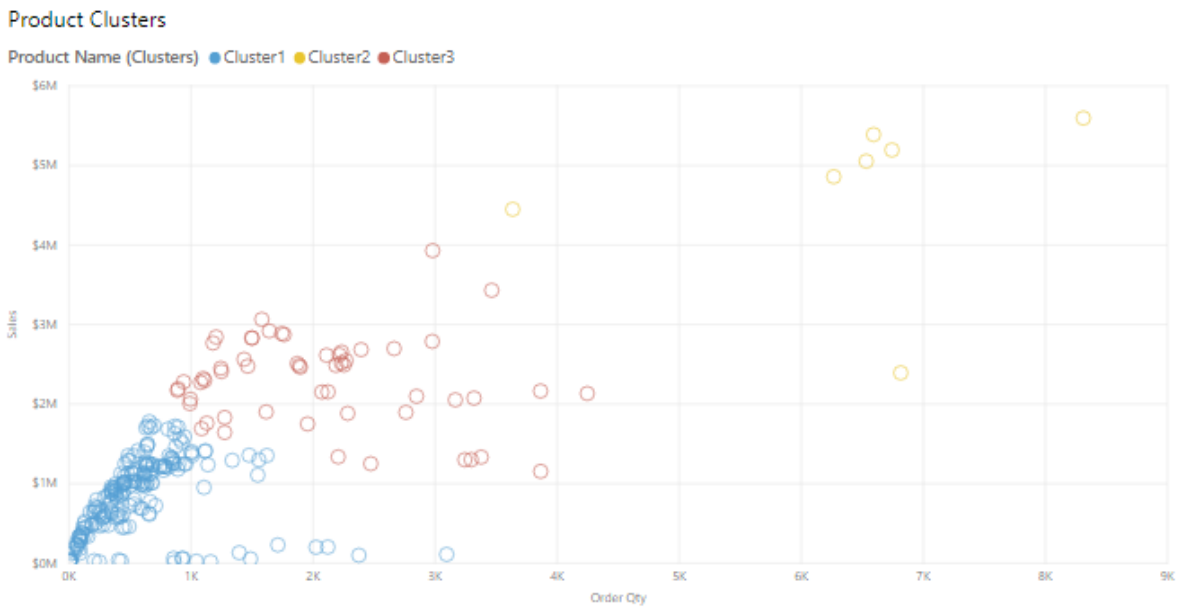
To **apply clustering to the scatter chart**, select **More options (…)** in the **upper-right corner of the visual** and then **select Automatically find clusters**

In the **Cluster window** that displays, you can then **edit the default name, field, and description** as seen in the image below:



Enter the **number of clusters that you want** (e.g., 3) into the box and **select OK**. Power BI will then **run the clustering algorithm** and **create a new categorical field with different cluster groups in it**

The image below **depicts the previous unclustered dataset with a 3 total clusters**



You can now **clearly see the clusters that are in the dataset** and **perform analysis on them**

NOTE that:
You can **recluster the data with different Number of clusters** but this will **discard the current clustering and cluster names**

We can also **use the clusters**, which is stored in the **scatter chart's Legend field well bucket**, as a source of **cross-highlighting** like **any other Legend field**

Use the Analyze Feature

The **Analyze feature** provides you with **additional analysis that is generated by Power BI for a selected data point**.

You might want to use this feature **if Power BI has found something that you haven't seen before**
or
if you want **Power BI to give you a different insight into your data**

This feature is **useful for analyzing why your data distribution looks the way that it does**.

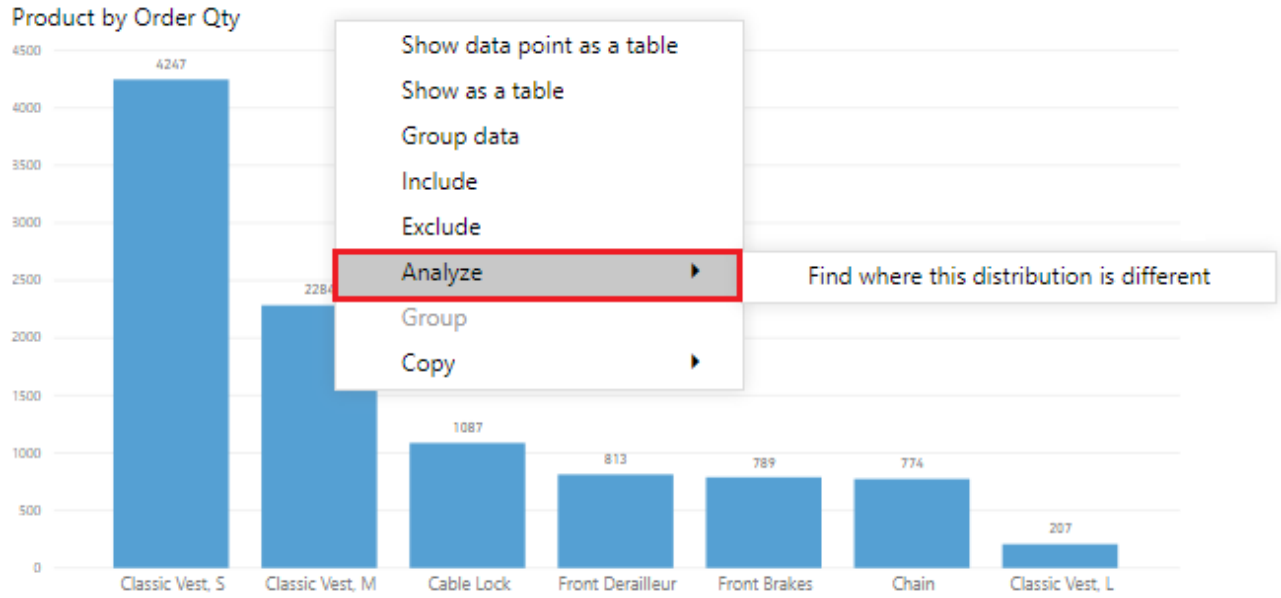
Instead of exploring the data manually, you can **use the Analyze feature** to get **fast, automated, insightful analysis of your data**

To use this feature, right-click a data point on the visual and then **hover** over the **Analyze option** to display **two further options:**

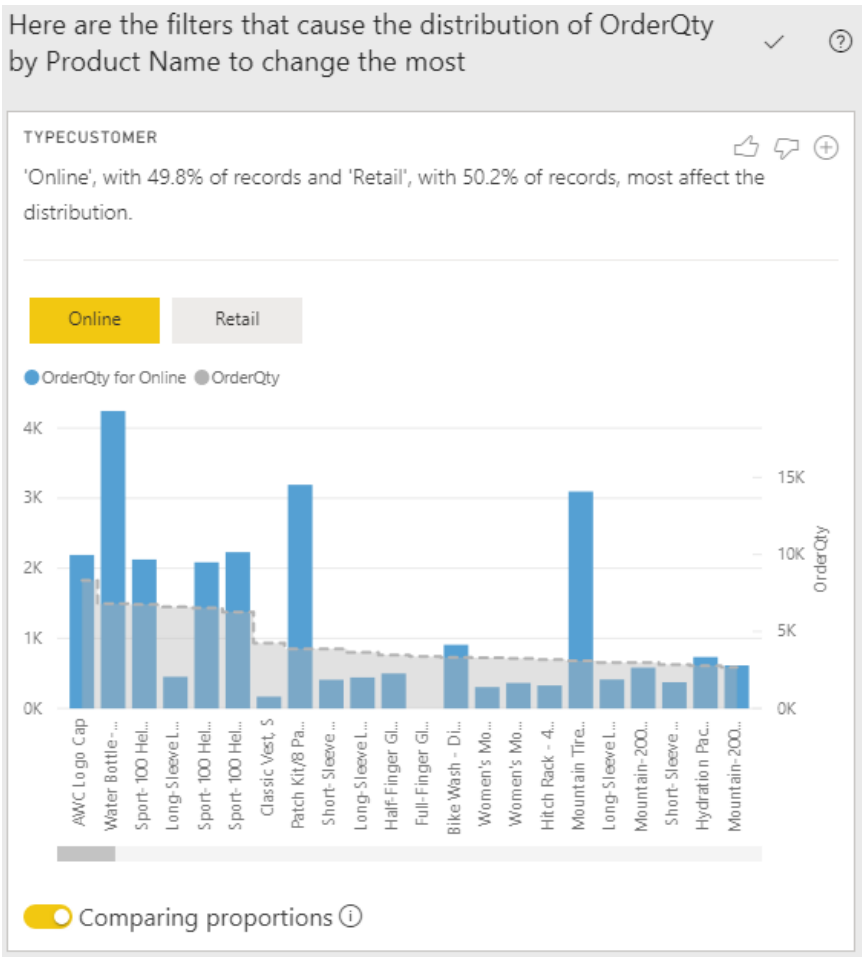
- **Explain the increase**
- **Find where the distribution is different**

The options that are available **will depend on the data point that you selected**

The **Analyze feature is demonstrated in the following images**:



In the example below, **you can select the Explain the increase option**, and a window displays with a new visual as illustrated in the image below:



If you **found these analysis useful,** you can **add the new visual to the report** so that **other users can view it**. Select the **plus (+)** icon in the **upper-right corner of the visual** to **add it to your report**

Review Quick Insights

this feature **uses machine learning algorithms** to **go over the entire dataset and produce insights (results)** for you quickly

GIANT NOTE
This feature only **AVAILABLE in POWER BI SERVICE ONLY**
This feature **DOES NOT WORK with DirectQuery**
This feature **ONLY WORKS with data that is IMPORTED** into Power BI

This feature is a **great way to build dashboards** when **you don't know where to start**
It can also
help you find insights you might have missed when building your reports

Get Quick Insights on your Dataset

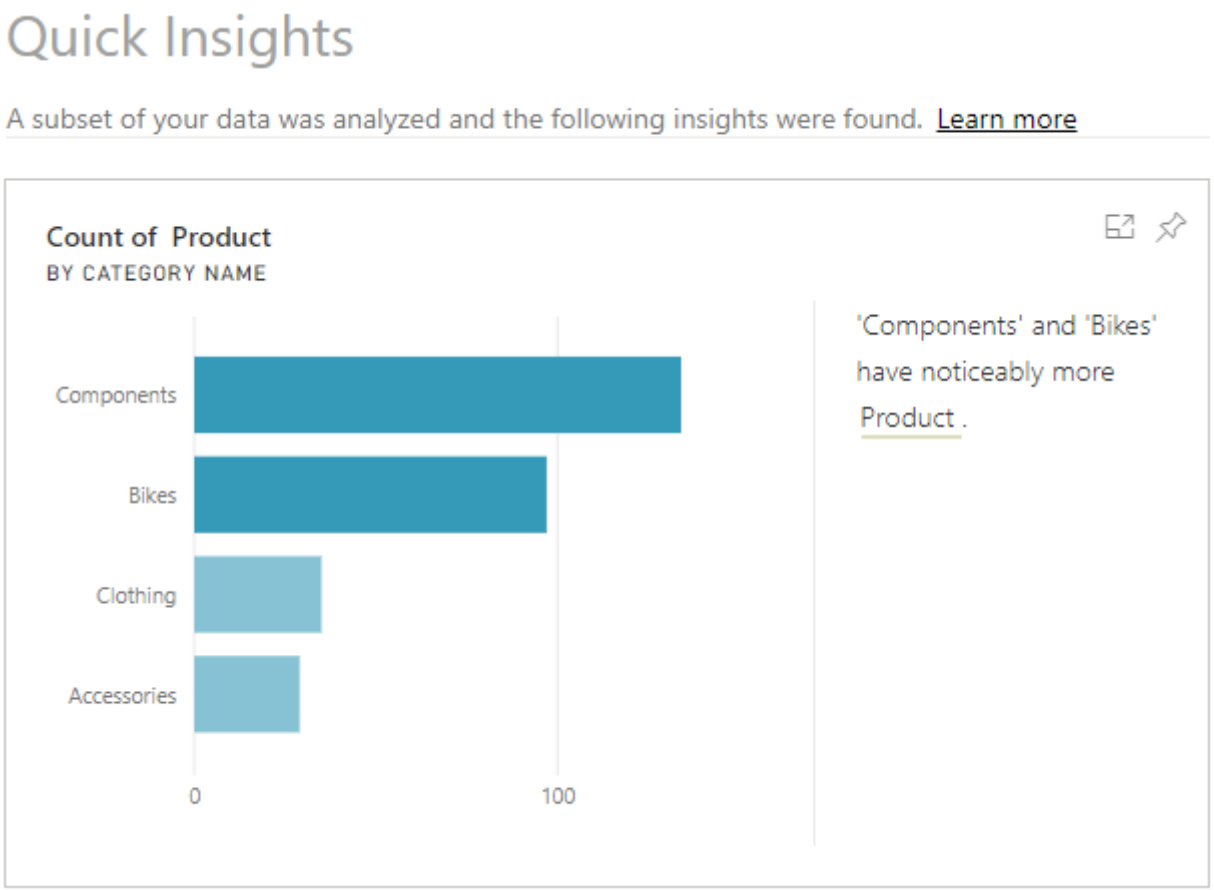
To get quick insights on your dataset, open the **Power BI Service** and then select the **Content tab**
Locate the **report** for which you want to **get quick insights** then select **More options (…)** => **Quick insights**

Once the algorithms has finished running.

Select the **View insights** to **open the Quick Insights page** for the selected dataset, and then **view the insights that Power BI has found for you**

The **Quick Insights page** contains **up to 32 separate insight cards**
each card has a chart or graph + **a short description**

The example below illustrates an insight return by the **Quick Insight service**



Add a Quick Insights Result Card to a Dashboard

If you see an insight card that is **particularly compelling,** you can **add it to your dashboard**

On the **Quick Insights page**, **hover over the card** and then **select the pin icon**. The **visual is then added to the dashboard** where you can **reposition it as required**

Interact with the Quick Insight Results

To take a closer look at the particular insight card, you can open it and perform the following actions:

- **Filter the visualization** by using the **available options in the Filters panel**
- **Pin the insight card** to a dashboard by **selecting the Pin visual**
- **Run insights on the card** (scoped insights) by selecting **Get insights** in the **upper-right corner**
- **Return to the original insight canvas** by **selecting the Exit Focus mode** in the upper-left corner

Apply AI Insights

This feature allows you to:

connect to a collection of pre-trained machine learning models that you can **apply to your data to enhance your data preparation efforts**

To apply **AI Insights to the data**:

go to **Power Query Editor** and select the **Column tab**, **three AI insight options will be available for you** to choose from:

- **Text Analytics**
- **Vision**
- **Azure Machine Learning**

NOTE:

A **Premium capacity/subscription** is required to **use the Text Analytics and Vision options**

Working with AI Visuals in Power BI

Power BI also has advanced AI capabilities that grant you:

the ability to **get an answer by asking a question using your own words** and **answering those questions back to you**

Use the Q & A Visual

This feature **lets you explore your data in your own words** by **allowing you to ask natural language questions** and then **providing you with answers to those questions**

The **ability to ask questions is VALUABLE to BOTH REPORT AUTHOR and REPORT USER**

For the Report Author

It gives you **ideas for the type of visuals that you can display in the report** and **lets you quickly add those visuals**

For the Report User

It gives an **effective tool that can be used to get quick answers to their questions about the data**

Add the Q & A Visual to your Report Canvas

To access the Q & A feature, you need to **add the Q & A visual to the report**. It can also be accessed from the **Visualizations pane**

The **Q & A feature is also available as a button**, useful for saving space on the report canvas

You can **start asking questions immediately** by **selecting one of the suggested questions** or by **entering a question into the question box**

Setting Up the Q & A Feature

You can set-up the underlying Q & A feature so that **it improves in answering questions about your data**

Basically, **teach the Q & A feature to better understand people**

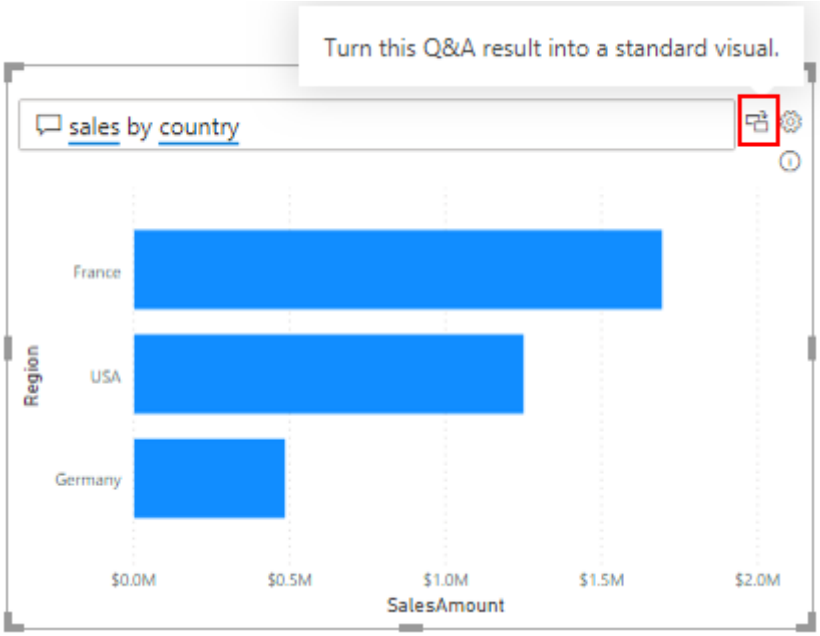
You can also **proactively monitor** and **review the questions** that are **coming through from users** and then **address misunderstandings or common typos**

Hence, you can **constantly calibrate the Q & A feature** so that **it provides better answers to your organization's questions**

Use the Q & A Feature to Build Visuals

Furthermore, when Power BI answers a question and you **find the visual result to be engaging or helpful** then you can **add it as a standard visual to your report**

In order to turn a **Q & A result into a standard visual,** select the icon next to the question box as shown in the image below:



The **Q & A feature is unique** in that **it DOES NOT require users to have knowledge of Power BI to use the visuals**

Find Important Factors w/ Key Influencers Visual

The **key influencer visual** helps you **understand the factors that are affecting a specific metric**.

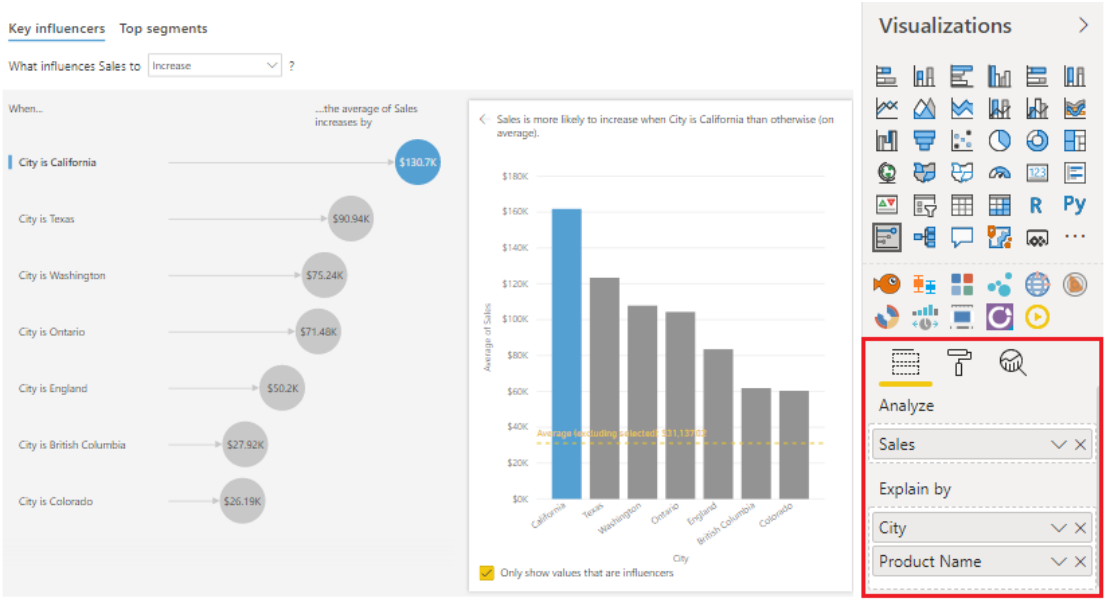
The visual **analyses your data for you, ranks the factors that matter** and then **displays those factors as key influencers**

The visual also **helps you to contrast the relative importance of these factors**,

meaning that **you can build your visuals while understanding what factors impact those visuals** and **why the visuals appear the way they do**

We can then **add a Key influencers visual** to the report by selecting the **Key influencers icon** on the **Visualization pane**

The visual can be configured by:



You can also use the **What influences… drop-down list** to **see what caused the data to decrease or increase**.

In the image above, the insight is that **sales for California is likely to be 130,700.00 higher than sales in other cities**

Use the Decomposition Tree Visual to Break Down a Measure

The **decomposition tree visual** will **automatically aggregate your data** and lets you **drill down into your dimensions** so that **you can view your data across multiple dimensions**

In order to describe the best way for using decomposition tree, it is better to describe a scenario

You have **built visuals for the Supply Chain team**, but the team wants to be able to **analyse the percentage of products that the organization has on back order**

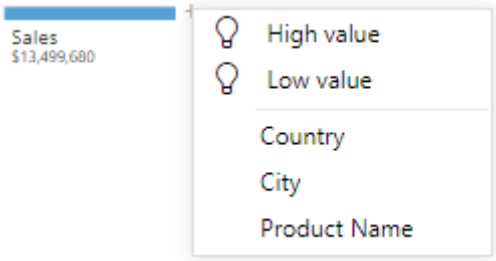
in other words, **the percentage of products that are out of stock**

Hence, the **decomposition tree** does enable you to **conduct root cause analysis in order to understand a measure better**

Add the **Decomposition tree visualization** through the **Visualization pane**

Then in the **Analyze field**, also **add the measure or aggregate** that **you want to analyse**
In the **Explain by field** well, **add the dimension(s) that you want to drill into**. In this case,

The **decomposition tree** allows you to **drill into your stated categories** as seen by the **list dropdown**



In this case, you want to **analyse the Sales field** by **drilling down into a number of dimensions**, such as **Country, City, and Product**

At the **top of the list of dimensions**, are **two additional options** marked as lightbulb icons.

These options are referred to as **AI splits**

They'll **automatically find high** and **low values in the data for you**

AI Splits

This function works by **considering all available fields** and **determining which one to drill into** in order to **get the highest / lowest value of the measure** that is being analysed

The following image **illustrates the results of selecting the High value AI split**