upGrad

#LifeKoKaroLift

Data Science Program

Module Name: SQL & Visualization

Course : Visualization

Lecture On : Advanced Visualisations Using Tableau

Instructor :

upGrad

Today's Agenda

- Revision
- Donut Charts
- Pareto Charts
- Packed Bubble Charts
- Highlight Tables
- Control Charts
- LOD Expressions: Include and Fixed
- LOD Expressions: Exclude
- Motion Charts
- Bullet Charts
- Gantt Charts
- Likert Scale Charts
- Hexbin Charts
- Best Practices
- Key Takeaways

Revision

upGrad

In previous session, we learnt:

- Treemaps and Grouping
- Dashboards - I
- Joins and Splits
- Numeric and String Functions
- Logical and Date Functions
- Histograms and Parameters
- Scatter Plots
- Dual Axis Charts
- Top N parameters and Calculated Fields
- Stacked Bar Charts
- Dashboards - II and Filter Actions
- Storytelling

Donut Charts

upGrad

- The first advanced visualisation that you learn about is a donut chart, which is a variation of a pie chart that looks like a doughnut.
- The hole in the centre can be used to display additional information.
- Let’s learn how to create a donut chart by first creating a pie chart and then manipulating it.
- A donut chart can be used for understanding the share in sales for different categories, such as furniture, office supplies and technology. Note that a donut chart is also a dual-axis chart.
- Donut charts can be used to analyse the amount of time spent by multiple users on different social media platforms.
- The hole in each donut chart can be used to store the user’s name, age, income, etc. This may be difficult to achieve in a pie chart.

Pareto Charts

upGrad

- A Pareto chart is a combination of a bar graph and a line chart.

- The bars in this visualisation are arranged in descending order, while the line chart is a cumulative total of the values in the bar graph. Let's learn about a Pareto chart in detail and create one in Tableau.

- Pareto charts are generally used to identify the most important factors out of a large set of factors.

- For example,
  - In the visualisation that you created, you can observe that the top four categories make up about 50% of the total sales.

  - Also, adding a reference line helps in identifying the top categories that are required to reach the reference percentage.

Packed Bubble Charts

upGrad

- You can use a packed bubble chart instead of a bar chart in certain cases to make your dashboard more attractive.
- Its main advantage is that you can easily notice the values that stand out due to their larger bubble size as compared to the other values.
- Now, do you know how easy or difficult it is to create a packed bubble chart in Tableau? Let’s find out.
- You can add another variable to the colour component to further enrich your chart.

Highlight Tables

upGrad

- You may have observed that data is generally stored in a tabular format. This is because a table helps in organising the information.
- However, as a table gets bigger and the number of values increases, it becomes increasingly difficult to extract insights quickly.
- Highlight tables differ from normal tables in the following two ways:
    - They generally contain numeric values.
    - The background of each cell is coloured with an intensity corresponding to its value.
- Let’s learn how to create a highlight table in Tableau.

Highlight Tables

upGrad

- Highlight tables are mostly used to compare numeric values among categorical variables.
- Generally, the darker the shade of a cell, the higher its value.
- Observe that there are 17 rows and 12 columns in the second highlight table, which amounts to 204 values.
- Watching these values without colour coding can be quite cumbersome, especially if you need to quickly identify which subcategories performed really well in which months.

Control Charts

upGrad

- A control chart is a subtle variation of a line chart that you read about in the previous module.
- Control chart is also used to study how data changes over time.
- Additionally, it contains fixed horizontal lines (limits) to determine if a certain process has values within the limits.
- In other words, a control chart is used to check whether the said process is under ‘control’ or not.
- Let’s learn about control charts in detail, and also understand how to create them in Tableau.

Control Charts

upGrad

- You saw how the upper and lower limits were calculated based on the mean and standard deviation of the profits for each month.

- Now, let's take a look at the functions that are used to calculate the limits.

    - **WINDOW_AVG()**: This function is similar to the window functions that you learnt about in the 'Advanced SQL and Best Practices' module. As Amit mentioned, this function is used to calculate the average of the values of the variable in question. However, it only considers the values present in the visualisation, out of the total values present in the data set.

    - **WINDOW_STDEV()**: This function is similar to the WINDOW_AVG() function, except that it calculates the standard deviation of the values present in the visualisation.

Control Charts

upGrad

- Now, let's look at the following formulae that are used to calculate the lower and upper limits for the control chart.
- **Lower Bound:**

  WINDOW_AVG(SUM([Profit])) - WINDOW_STDEV(SUM([Profit]))

- **Upper Bound:**

  WINDOW_AVG(SUM([Profit])) + WINDOW_STDEV(SUM([Profit]))

- The lower and upper bounds in the formulae above have been considered to be one standard deviation away from the mean.

- The number of standard deviations may vary in different cases, but it is generally considered to be one.

- Finally, you also learnt how to identify outliers in the data set by using a control chart.

Control Charts

upGrad

- The formula given below will return outliers, i.e., all values that lie below the lower bound or above the upper bound.

  SUM([Profit]) < [Lower Bound] OR SUM([Profit]) > [Upper Bound]

- Let's learn to add parameters to a control chart to make the upper and lower bounds dynamic.

- The formulae used to calculate the lower and upper bounds dynamics, similar to the ones explained in the previous video, are given below.

- Lower Bound:

WINDOW_AVG(SUM([Profit])) - WINDOW_STDEV(SUM([Profit])) * [Contrl Std Dev]

- **Upper Bound:**

WINDOW_AVG(SUM([Profit])) + WINDOW_STDEV(SUM([Profit])) * [Contrl Std Dev]

LOD Expressions: Include and Fixed

upGrad

- The concept of Level of Detail (LOD) expressions is a feature in Tableau used to add multiple levels of granularity to your visualisations.
- It helps you answer questions that cannot generally be answered with visualisations created by using only data fields, as they are originally present in the dataset.
- There are three types of LOD expressions: Include, Fixed and Exclude.
- An Include LOD expression can be used to analyse data for cities, provided that you have data for states.
- It ‘includes’ the dimension of a city in its calculation even if it is not present in the visualisation.

LOD Expressions: Include and Fixed

upGrad

- Now, let’s look at another type of LOD expression: Fixed LOD and it’s importance in business intelligence.
- As the name suggests, a Fixed LOD expression does not make any computations using additional dimensions that are not already present in the view or the visualisation. In the video above,
- Calculating a Fixed LOD expression simplifies identifying the percentage of sales made for each subcategory.

LOD Expressions: Exclude

upGrad

- An Exclude LOD expression functions exactly opposite to the way an Include LOD expression works.

- An Include LOD expression was used to calculate city-wise data on the basis of the state, while an Exclude LOD expression was used to calculate the total sales made in each state and display that data along with the individual sales information for each city.

Motion Charts

upGrad

"Animation can explain whatever the mind of man can conceive. This facility makes it the most versatile and explicit means of communication yet devised for quick mass appreciation."

–- Walt Disney

- Most of the time, moving pictures are definitely better than still ones. The charts and graphs that you have learnt about so far,
- Motion charts are dynamic; they move on the screen. Due to this, they are also called animation charts.
- They are mostly used to display changes in data over time.
- Let's learn more about motion charts and understand how you can create them in Tableau.

Motion Charts

upGrad

- You can use a motion chart to view snapshots of sales and profits for each month one after the other, like a slideshow or an animation.
- In Tableau, you can change the speed of the animation as well as the direction.
- Finally, you saw how adding the month and year in the title of the chart makes it much easier to understand the time duration that is currently being displayed in the animation.
- ‘Show history’ is a highly useful option to show how the data values of a particular category and subcategory changed over a period of time.
- We would recommend you go ahead and try some motion charts in Tableau on your own and have a little fun!

Bullet Charts

upGrad

- A bullet chart is a variation of a bar chart, where you can find out if a certain measure of a categorical variable is less than, equal to or greater than the recommended value.
- Let’s learn how to create bullet charts in Tableau.
- Among many graphs under the ‘Show Me’ option, a bullet chart is readily available for use.
- A good example of using a bullet chart can be to compare the current sales performance of a team with the target sales to be achieved.
- By doing so, the team can easily determine the metrics in which it is struggling to achieve the required revenue.

Gantt Charts

upGrad

- The module development process carried out at upGrad involves several steps.
- Each step can have one of the following values: **not started**, **in progress** and **completed**.
- This planning is essential to understand which modules are on track and which are not.
- Various steps or phases are followed by Content Strategists in the Data Science program.
- Multiple people from different teams are involved in each phase. Tracking the progress of each step for multiple modules can become a nightmare for a program manager. This is where Gantt charts come in handy.

Gantt Charts

upGrad

- The module development process carried out at upGrad involves several steps.
- Each step can have one of the following values: **not started**, **in progress** and **completed**.
- This planning is essential to understand which modules are on track and which are not.
- Various steps or phases are followed by Content Strategists in the Data Science program.
- Multiple people from different teams are involved in each phase. Tracking the progress of each step for multiple modules can become a nightmare for a program manager. This is where Gantt charts come in handy.
- Let’s see the usage of Gantt charts to understand whether a project is on track or not.

Likert Scale Charts

upGrad

- As a consumer, you keep coming across surveys on various aspects of different products.

- These poll questions are designed to understand and analyse consumer sentiment and behaviour. Consider the sample poll question given below.

- Do you think this module will help you progress in your career as a data analyst? Select one of the following.
  - Strongly disagree
  - Disagree
  - Neither agree nor disagree
  - Agree
  - Strongly agree

- What you just saw is a classic example of a Likert scale. It is a rating scale that is commonly used in surveys.

Likert Scale Charts

upGrad

- A Likert scale chart is nothing but a modified bar chart used to visualise a Likert scale.
- We will use the data set given on the platform. This data set measures the feedback given by 400 customers for a particular hotel. Its metadata has also been given.
- Let’s learn about Likert scale charts.
- It is easy to determine the primary feature(s) of a product that customers particularly disliked.
- Notice how you need to pivot the data to make it more readable. You also need to join the survey file to the metadata file in order to understand what aspect of customer service is mapped to what question.

Likert Scale Charts

upGrad

- The calculation used to create a Likert scale based on the ratings given by customers is provided below:

```
IF [Rating] = 1 THEN 'Very Bad'
ELSEIF [Rating] = 2 THEN 'Bad'
ELSEIF [Rating] = 3 THEN 'Neutral'
ELSEIF [Rating] = 4 THEN 'Satisfied'
ELSE 'Very Satisfied' END
```

- You also learnt that in this case, you can further refine this scale in the following manner:

```
IF [Rating] < 3 THEN 1
ELSEIF [Rating] = 3 THEN 0.5
ELSE 0 END
```

Likert Scale Charts

upGrad

- Now, we will create a Likert scale chart for the given data set. Let's learn about another new term: the 'Gantt Start'.

- It is easy to calculate the total responses given. The formula used is given below: **TOTAL(COUNT([Survey Response]))**

- You also learnt that a Gantt Start is used to determine the offset point from where every bar in a Likert scale chart starts. The calculation used to find the Gantt Start is given below: **- [Total Negative Score] / [Total Responses]**

- Along with the point at which negative ratings start coming in, it is also important to take into account the number of ratings for each rating description. This can be determined by looking at the size of each bar.

- The calculation used for determining 'Sizing' is given below:

  COUNT([Survey Response]) / [Total Responses]

Likert Scale Charts

upGrad

- Along with determining the Gantt Start for each bar, it is also important to find the point at which the next rating will start.
- Let’s learn about the calculation required to determine this.
- This value is also known as the **Gantt Percent Start**.
- The calculation used is given below:

PREVIOUS_VALUE([Gantt Start]) + ZN(LOOKUP([Sizing], -1))

Likert Scale Charts

upGrad

- Let’s take a closer look at the functions used to calculate the Gantt Percent Start.
    - **PREVIOUS_VALUE()**: It returns the value of the calculation carried out in the previous row. If the current row is the first row of the partition, it returns the given expression itself.
    - **LOOKUP(expression, [offset])**: It returns the value of the expression in a target row, specified as a relative offset from the current row. For example, LOOKUP([Sizing], -2) will return the value of ‘sizing’ present in two rows above the current row.
    - **ZN(expression)**: It returns the expression if it is not null; otherwise, it returns zero.

Hexbin Charts

upGrad

- In this segment, you will learn the importance of hexbin charts for visualising data and how to create them in Tableau. Before proceeding, download the ZIP code dataset given on the platform.

- **Hexbin charts**, or hexbin maps, are called so because they consist of a large number of hexagonal ‘bubbles’. Let’s know more about this interesting visualisation.

- Hexbin charts are derived from density maps. They are generally used as a replacement for treemaps when the number of data points is very large.

- Their distinctive shape makes it easier to group similar values together and find data points that stand out in the visualisation.

Hexbin Charts

upGrad

- To create a hexbin chart, you need to use the HEXBINX() and HEXBINY() functions that are provided in Tableau.
- Note that the hexagon shape is not available by default. You will need to download an image in PNG format and store it with the other shapes in the folder where Tableau is installed in your system.
- You have now learnt how to create and use a large number of advanced visualisations in Tableau.
- Before summarising all the visualisations, it is important to understand the best practices that you need to follow while creating charts and dashboards.

Best Practices

upGrad

- In any programming language, you need to know the best practices to be followed.

- This is to ensure that you write clean and efficient code that others can easily understand and work upon.

- For example, you learnt about the following best practices in the module on ‘Advanced SQL and Best Practices’:
    - Comment your code by using a hyphen (-) for a single line and (/* … */) for multiple lines of code.
    - Always use table aliases when your query involves more than one source table.
    - Assign simple and descriptive names to columns and tables.
    - Write SQL keywords in upper case and the names of columns, tables and variables in lower case.

Best Practices

upGrad

- Similarly, there is a set of directives that you should ideally follow when creating a visualisation in Tableau. Let's start by learning some of these.

- The best practices to keep in mind while creating charts and dashboards, as explained in the video above, can be summarised as follows:

- **Define the purpose:** Define the purpose of your visualisation and focus on achieving that objective.

- **Consider your audience**: Consider your audience and their degree of familiarity with the subject and the underlying data.

- **Provide context**: Ensure that your visualisation contains supporting information that provides the source of the data and the time of its collection, and explains how it relates to your audience.

Best Practices

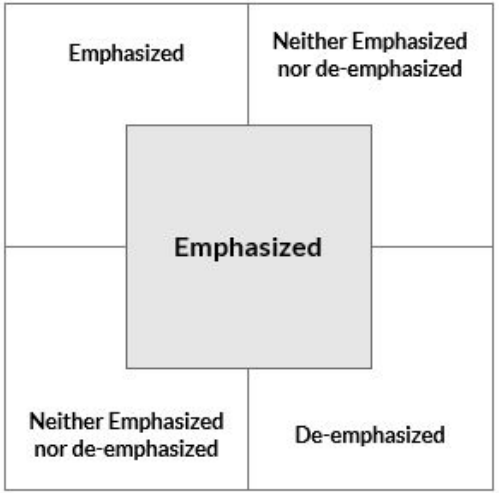
upGrad

- **Show numbers**: Show the values for all the critical components of your visualisation.
- **Present the most important information first**: This is especially important while sorting data and arranging multiple visualisations in a dashboard.
- **Pay attention to aesthetics**: Make effective use of colour, formatting, font and size.
- Remove the third dimension of depth in charts.
- Eliminate the grid lines from a visualisation.
- Eliminate variations in colour that do not encode any meaning.

Best Practices

upGrad

- Remove data that is less relevant.
- Condense data using summaries.
- Emphasise important data using visual attributes such as color, size and shape.
- Emphasise important data by its position on the dashboard. Refer to the diagram given below.



Key Takeaway

- Revision
- Donut Charts
- Pareto Charts
- Packed Bubble Charts
- Highlight Tables
- Control Charts
- LOD Expressions: Include and Fixed
- LOD Expressions: Exclude
- Motion Charts
- Bullet Charts
- Gantt Charts
- Likert Scale Charts
- Hexbin Charts
- Best Practices
- Key Takeaways

upGrad

#LifeKoKaroLift

Thank You!