Data Visualization

Assignment – 8.1

edureka!

Assignment – 8.1

Scenario 1

Let's consider a hypothetical situation wherein you are employed by Netflix, an American media production and services provider company headquartered in Los Gatos, California. It was founded in 1997 by Reed Hastings and Marc Randolph in Scotts Valley, California.

Your responsibility is to bring relevant insights out of the rating dataset, which incorporates the user ratings, rating score, and release year for the shows available on Netflix. This analysis can help the firm make better production-related business decisions.

Note: Please take the Screen shots of the work book for each problem and send it in a Doc file along with the Packaged workbooks

Required Data Description:

- Title: Names of the shows produced
- Release Year: Release Year of the show
- User Rating Score: Ratings provided by the user to the shows
- Rating: Shows are rated and categorized into UR, R, PG, etc., as per the censor board

Problem 1

Find the trend in the number of shows produced (Title Field) over the years to analyze the spikes in the pattern.

- **Release Year** field is the year of release. (Suggestion: Use the calculated field to find out the unique number of shows released per year)
- Turn on the marking label only for the year having the maximum shows produced, in a dynamic manner such that the label is turned on and off accordingly, as the year of maximum shows releases changes dynamically during the next data update
- Besides, color the line chart based on the number of shows released that year

Problem 2

The production team is planning for the next seasons of the shows having the best User Ratings

- Plot a bar chart by the **Average User Rating Score**
- Provide the user the option to choose top N shows by the **User Rating Score** and identify the shows for the subsequent season production based on the user ratings. (Suggestion: Use parameter with filters to allow the user to enter the inputs of Top N shows and use the same in the view)

Problem 3

Find the sum of the average number of shows under each **Rating** field. The expected output is numeric.

- Calculation Assistance: We need to calculate the average number of shows for each **Rating** field (Present in dimensions). Subsequently, we need to sum up these averages as per the requirements (Suggestion: Use LOD expressions to calculate the number of shows at the **Rating** field level and average it out at the overall level)

Scenario 2

Let's consider a hypothetical situation wherein you are employed by a US survey agency that analyzes the electoral data.

Your responsibility is to bring relevant insights out of the sample dataset, which incorporates the party-wise expected vote share. You are required to analyze the same and prepare a report for the upcoming journal article.

Required Data Description:

- Party: Names of the contesting parties
- Candidate: Names of the contesting candidates from the parties
- State Code, State Name, and County Name (Geography-specific info)
- Vote count: Vote counts at the row-level
- County Vote Count: Vote counts at the county level

Problem 1

Find the percentage share of the vote count for each party and present the same in decreasing order to figure out who's leading the survey results (Suggestion: Use Quick table calculation)

- Turn on the mark labels
- Color code the bars by their respective vote share

Problem 2

Find the candidate from the state, Washington DC, having the largest vote share in percentage and create a bar chart using the same data. The data will be published in a newspaper to support the article about the popularity of candidates from each party in the President's State. (Suggestion: Use Quick table calculation)

Problem 3

Identify the top 5 states where Clinton is having the highest vote count (Suggestion: Avoid using calculated fields over here)

- Provide the option to use the same statistical piece of data for other candidates as well
- Turn on the mark labels for the plot
- Also, provide a color gradient as per the vote share
- Choose the graph type as per the requirements