**Lab - Creating a Visualization**

**Objectives**

In this activity, you will use Tableau Desktop Public, a free version of the Tableau desktop product to create meaningful visualizations to answer questions.

**Part 1: Using a Public Dataset on Tableau**

**Part 2: Using Data to Create Meaningful Visualizations**

**Background/Scenario**

Visualizations are often used to spot trends or to uncover previously unknown correlations within a dataset. The right visualization can show areas of a business that need attention, or where marketing programs should be targeted.

In this lab, you will use a fictional company dataset that is publicly available in Tableau. Using different types of visualizations, you will create meaningful charts and graphs that can illustrate issues existing in the business.

**Required Resources**

* PC with internet access
* Tableau Desktop App

**Instructions**

**Part 1: Using a Public Dataset on Tableau**

**Step 1: Connect to the dataset.**

1. Open Tableau Desktop. In the start page, select **Sample Data Sets** in the **Discover** pane on the right.
2. Scroll down to the **Entertainment** category. Download the **Hollywood’s Most Profitable Stories** Dataset (csv) by clicking on **Dataset (csv)** on the right of the window.
3. When the download is completed, return to Tableau Public Desktop. In the **Connect** pane, select **Text File**.
4. Locate and open the file **HollywoodsMostProfitableStories.csv** that you just downloaded.

**Step 2: Explore the public dataset.**

Data analysis starts with a question or an issue. Using this movie dataset, an analyst can answer questions such as “What genre of movie gets the best audience scores?” or “Which movie studio produced the most revenue in a specific year?”

1. When you load the dataset, the **Data Source** summary page appears. Here you can verify the types of data contained in each field, and quickly make changes to any type that is not correct.
2. Scroll over to the right in the table on the right until you see the **Year** column. Click the **Type** icon (#) above the **Year**.

What is the type of data that is actually contained in the Year column?

Date

Why is the type shown for the Year field not correct?

Because it takes the year as a number

1. Change the type on the **Year** field to **the correct type**.

What happens to the data in the column?

It changes to 1 / january of the year

1. Change the type for the Year field back to **Number (whole)**.
2. Review the fields that are contained in this dataset.

What might be another question you could answer through analysis of this data?

Answer Area

**Part 2: Using Data to Create Meaningful Visualizations**

**Step 1: Question: What genre of movie gets the best scores?**

Let’s start with the question, “What genre of movie gets the highest audience scores?”

1. Click the **Sheet 1** tab at the bottom of the screen to open the workspace. The fields in your data table are in the left panel.

Note: If Tableau still shows the movies database from the previous lab, right-click on the title of the database at top left and select **Close**. If a popup appears, select **OK**.

1. Drag the field Genre to the Columns shelf at the top of the workspace.
2. Drag the Audience score % field to the Rows shelf at the top of the workspace. The audience score is automatically shown as a sum when you move it to the Rows shelf. For this analysis, however, the average score for each genre is needed. Use the down arrow next to the SUM(Audience score%) on the Rows shelf to select Measure(Sum) > Average.
3. Tableau should now present a column chart showing the average audience scores for each genre.

Based on the resulting column chart, which genre has the highest audience score?

Answer . FANTASY

1. To look a little deeper into this analysis, it is necessary to make sure that the highest rated genre’s audience score is based on a large number of high scores and not a very small number of very high scores, which would make the results unreliable. Add a count of the scores onto the chart by dragging the **Audience score %** field to the **Row** shelf. This will cause a second chart to open showing the number of ratings per genre.
2. Use the dropdown arrow on the new **Audience score %** entry to change the measure from SUM to COUNT by selecting **Measure(Sum) > Count**. Questions:

Which genre has the highest number of scores?

Answer COMEDY

How does the count of scores for the genre with the highest audience score impact the analysis?

The analysis could be misleading. The fantasy genre has a limited number of audience ratings, and all of them are quite high.

**Question: Which movie studio produced the most revenue in a specific year?**

To answer the second question, you will need to start a new worksheet for this analysis.

1. Open a second sheet using the leftmost **+** icon at the bottom of the screen.
2. Start by dragging the **Lead Studio** to the Columns shelf. Then add **Year** to the Rows shelf. A grid showing which studios had data in which years opens.
3. Find the bars associated with the year 2010 in the grid; there are 11. Select them by drawing a box over all of them with your mouse or by selecting them individually. Right-click or hover over the selected bars and choose **Keep Only**. The other parts of the grid disappear.
4. Now drag **Worldwide Gross** to the Rows shelf as well and remove Year (right click and select Remove or click on the bubble and hit the delete key). You should see a column graph. In this context, “Gross” means the total amount of money the studio made that year from the movies in the dataset.

Note: Make sure your window is wide enough to see all of the studios in the dataset.

Which studio took in the most money on the movies in the dataset during 2010?

Answer. WARNER BROS WITH 548

1. Now let’s adjust the measures to show the average gross per picture: in the Rows shelf, change the **SUM(Worldwide Gross)** to an AVG function by using the dropdown arrow to select **Measure(Sum) > Average**.

Now which studio took in the most average gross revenue per picture?

Answer DISNEY

1. You can rename the sheets by double clicking the sheet name ("Sheet 3,” for example) and entering new names.

**Reflection Question**

Think about how the different views of the same dataset give different insights and tell different stories about the data. Why is it important to have representative and relevant data when you begin your analysis?

Answer It is really important because if your data aint relevant it could lead to misleading conclusion. So that why its important to know the question we ask and also have wide relevant data.