**Department of Computer Science and Engineering (Data Science) AY -2021-22**

**Lab Manual**

**Subject: Foundations of Data Analysis Laboratory (DJ19DSL303)**

**Semester: III**

**Experiment 1**

**(Data Visualization)**

**NAME:RISHABH PATIL SAPID:60009200056**

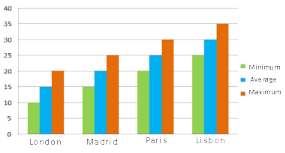
**DIV/BATCH:K/K2 DATE:16/11/2021**

**Aim**:

Create new measures on a given dataset and visualize them using a bar graph.

**Theory**:

A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called a column chart.



A Grouped Bar Graph denoting the temperatures in different cities on a specific day

A bar graph shows comparisons among discrete categories. One axis of the chart shows the specific categories being compared, and the other axis represents a measured value. Some bar graphs present bars clustered in groups of more than one, showing the values of more than one measured variable.

Bar graphs/charts provide a visual presentation of categorical data. Categorical data is a grouping of data into discrete groups, such as months of the year, age group, shoe sizes, and animals. These categories are usually qualitative. In a column (vertical) bar chart, categories appear along the horizontal axis and the height of the bar corresponds to the value of each category.

Bar charts have a discrete domain of categories and are usually scaled so that all the data can fit on the chart. When there is no natural ordering of the categories being compared, bars on the chart may be arranged in any order. Bar charts arranged from highest to lowest incidence are called Pareto charts.

Grouped (clustered) and stacked:

In grouped (clustered) bar charts, for each categorical group there are two or more bars color-coded to represent a particular grouping. For example, a business owner with two stores might make a grouped bar chart with different colored bars to represent each store: the horizontal axis would show the months of the year and the vertical axis would show revenue.

Alternatively, a stacked bar chart stacks bars on top of each other so that the height of the resulting stack shows the combined result. Stacked bar charts are not suited to data sets having both positive and negative values.

**Lab Assignments to complete in this session**

Use the given dataset and perform the following tasks:

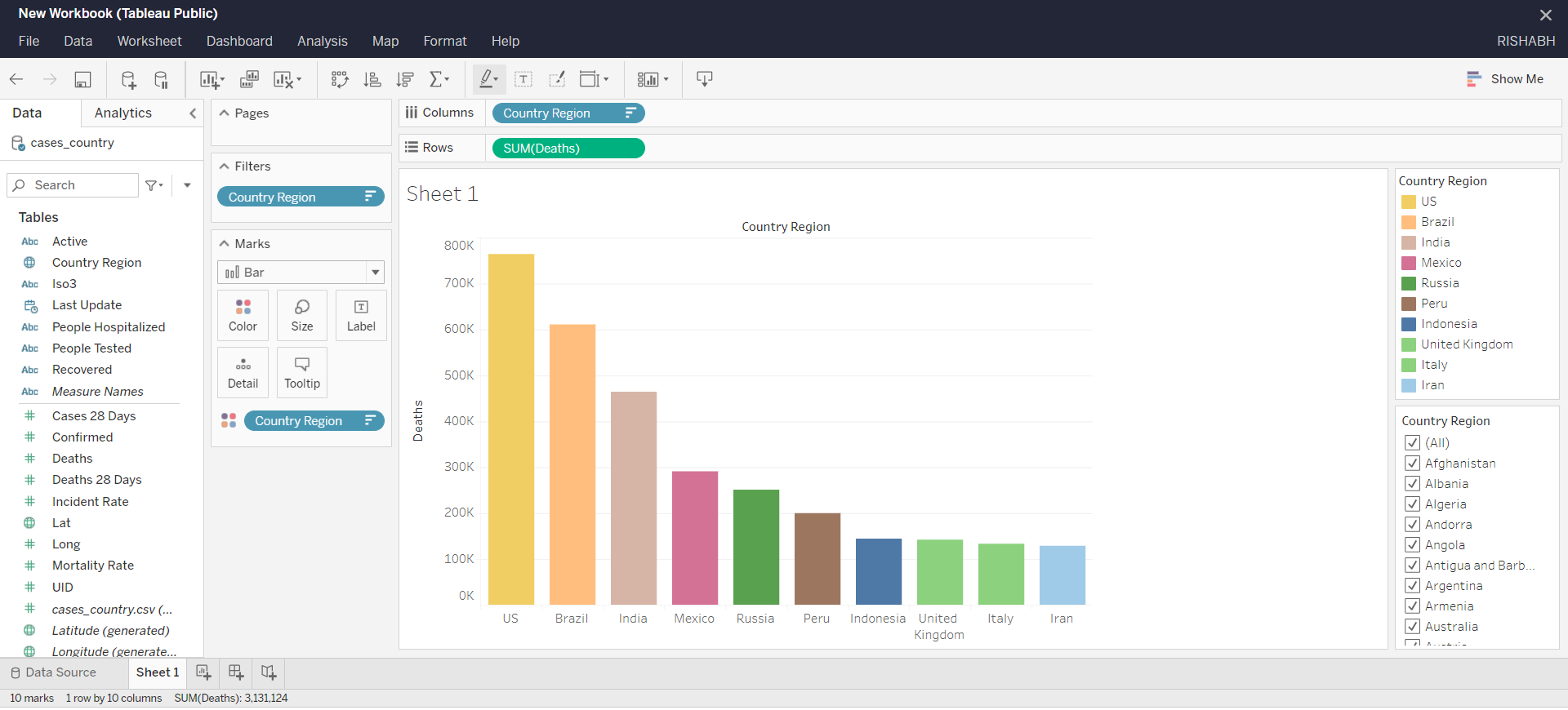
**Dataset:**

**1.COVID CASES**

https://raw.githubusercontent.com/CSSEGISandData/COVID-19/web-data/data/cases\_country.csv Compatible Web Data Connector for easy importing:

https://basic-csv-wdc.herokuapp.com/

a) The United Nations (UN) wants to know what Countries are suffering more deaths due to Covid-19, in order to donate additional vaccines and prevent the spread of Covid, from the additional funds and vaccine budget of this month.

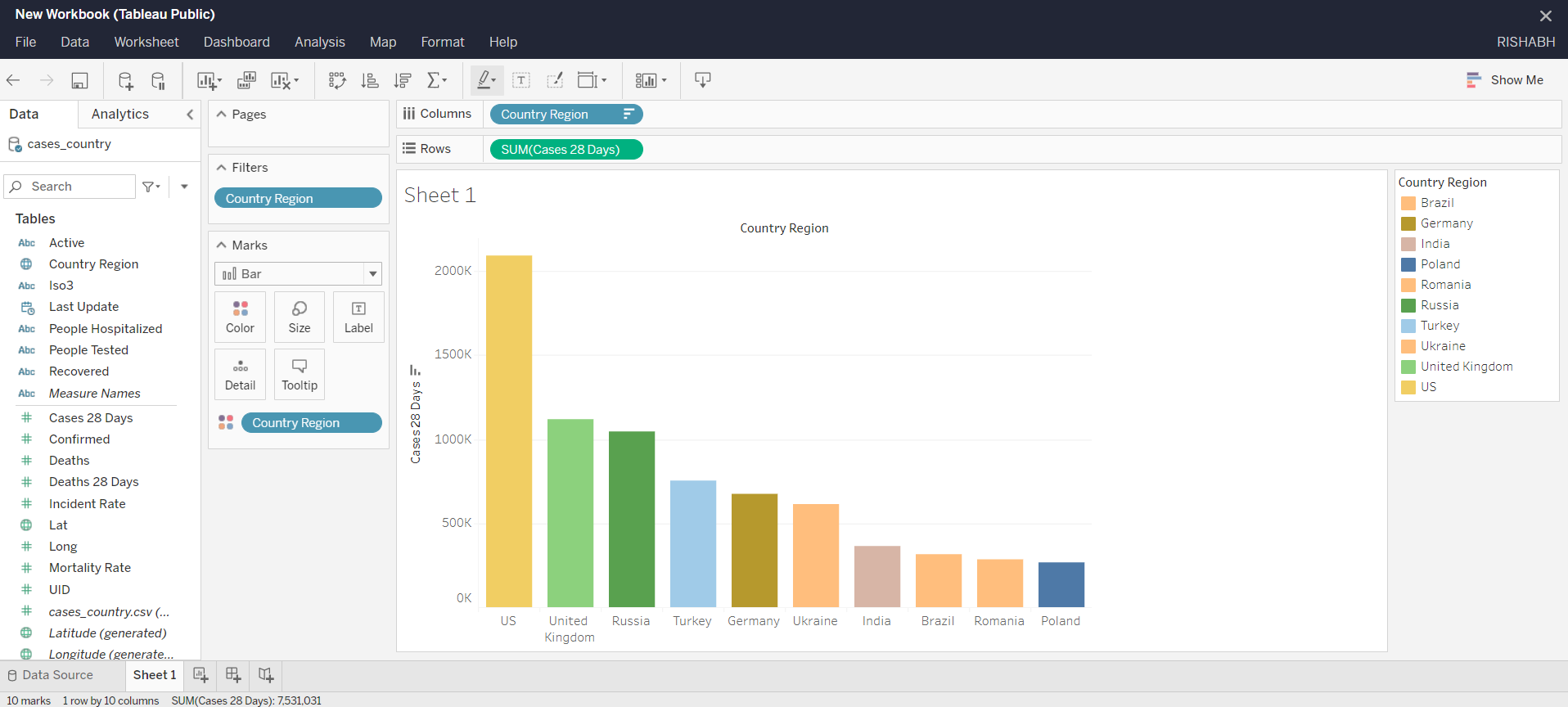


Here, according to the question, we need to plot a bar graph to see which countries are suffering more deaths due to Covid-19. So, in this above graph, we set

Y-axis: Deaths X-axis: Country Region

Hence, we got the required Bar Graph. Thus we get the top countries suffering more deaths due to Covid-19.

b) Find out what are the top 10 Countries suffering severely due to covid in the past month eligible for additional vaccine donations.

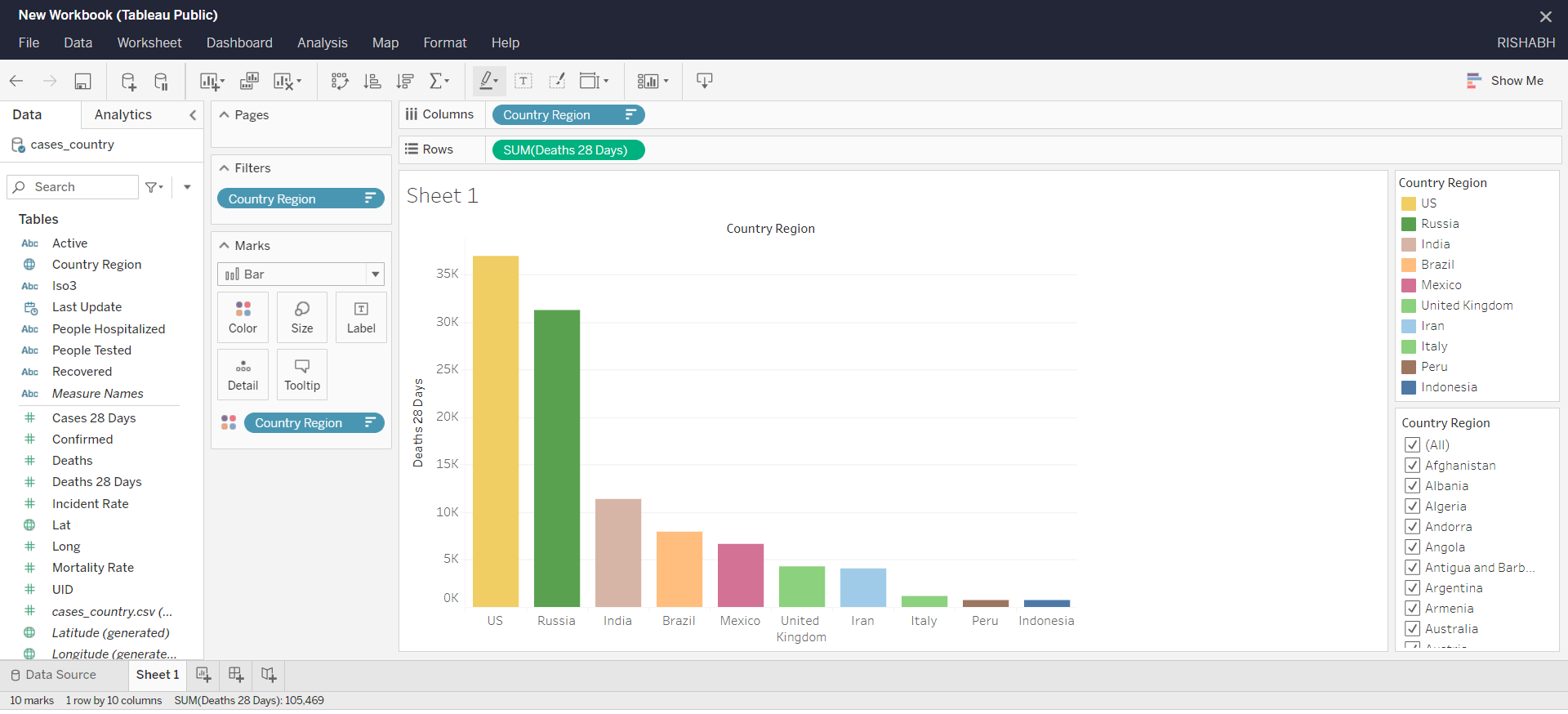


Here, according to the question, we need to plot a bar graph to see which countries are severely suffering due to Covid-19 in this month. So, in this above graph, we set

Y-axis: Cases (28 days) X-axis: Country Region

Also, since we needed only the top 10 countries which are severely suffering from Covid-19, we filtered the data for the top 10 countries. Hence, we got the required Bar Graph. Thus, the top 10 countries are the US, UK, Russia, Turkey, Germany, Ukraine, India, Brazil, Romania, and Poland.

c) Need for vaccines and funding are estimated from the number of deaths in the past month.



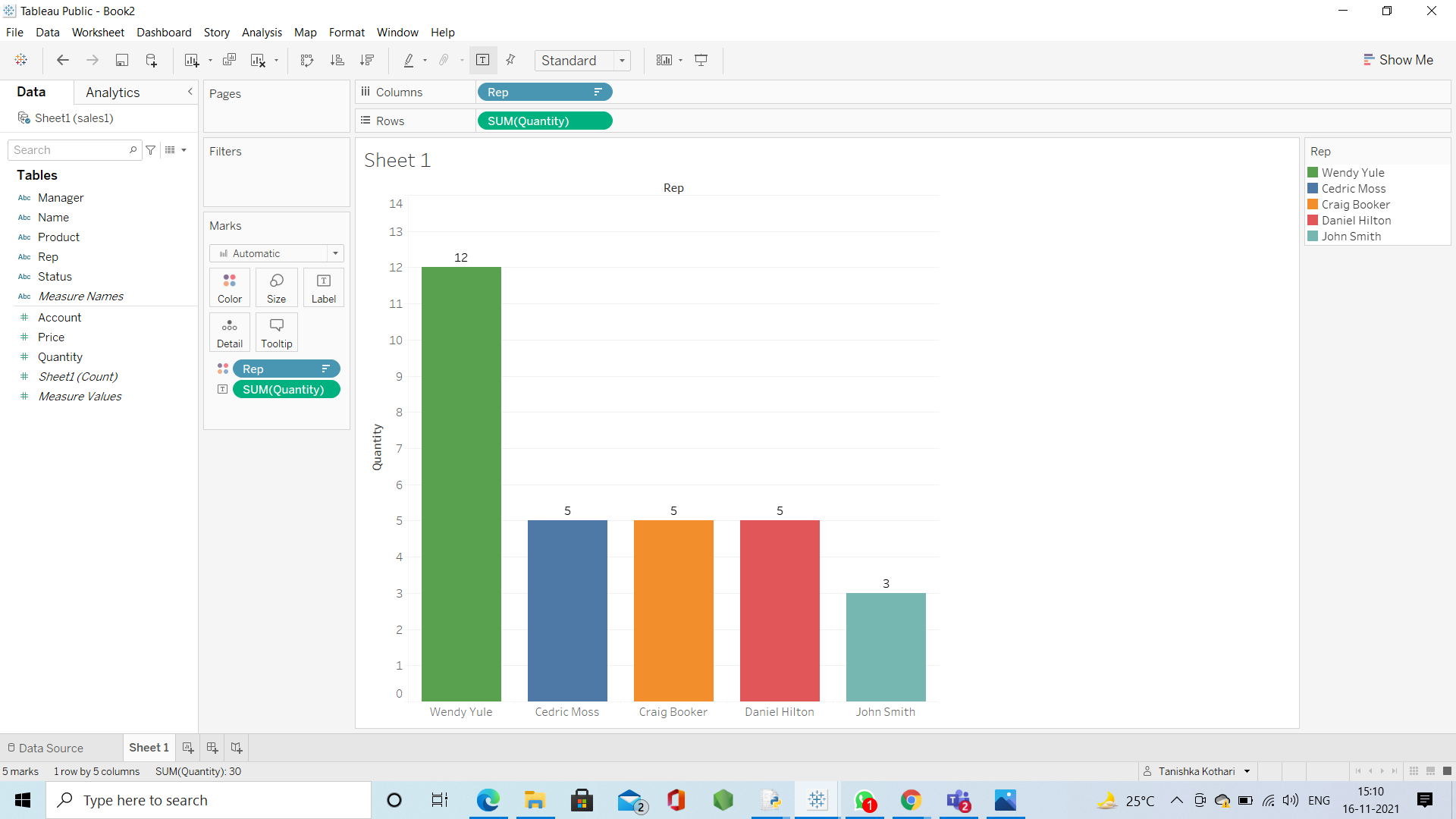
Here, according to the question, we need to plot a bar graph to see which countries are suffering more deaths due to Covid-19 in this month. So, in this above graph, we set

Y-axis: Deaths (28 days) X-axis: Country Region

Hence, we got the required Bar Graph. Thus we get the top countries suffering more deaths due to Covid-19

**2.SALES**

a)which representative has sold highest quantity

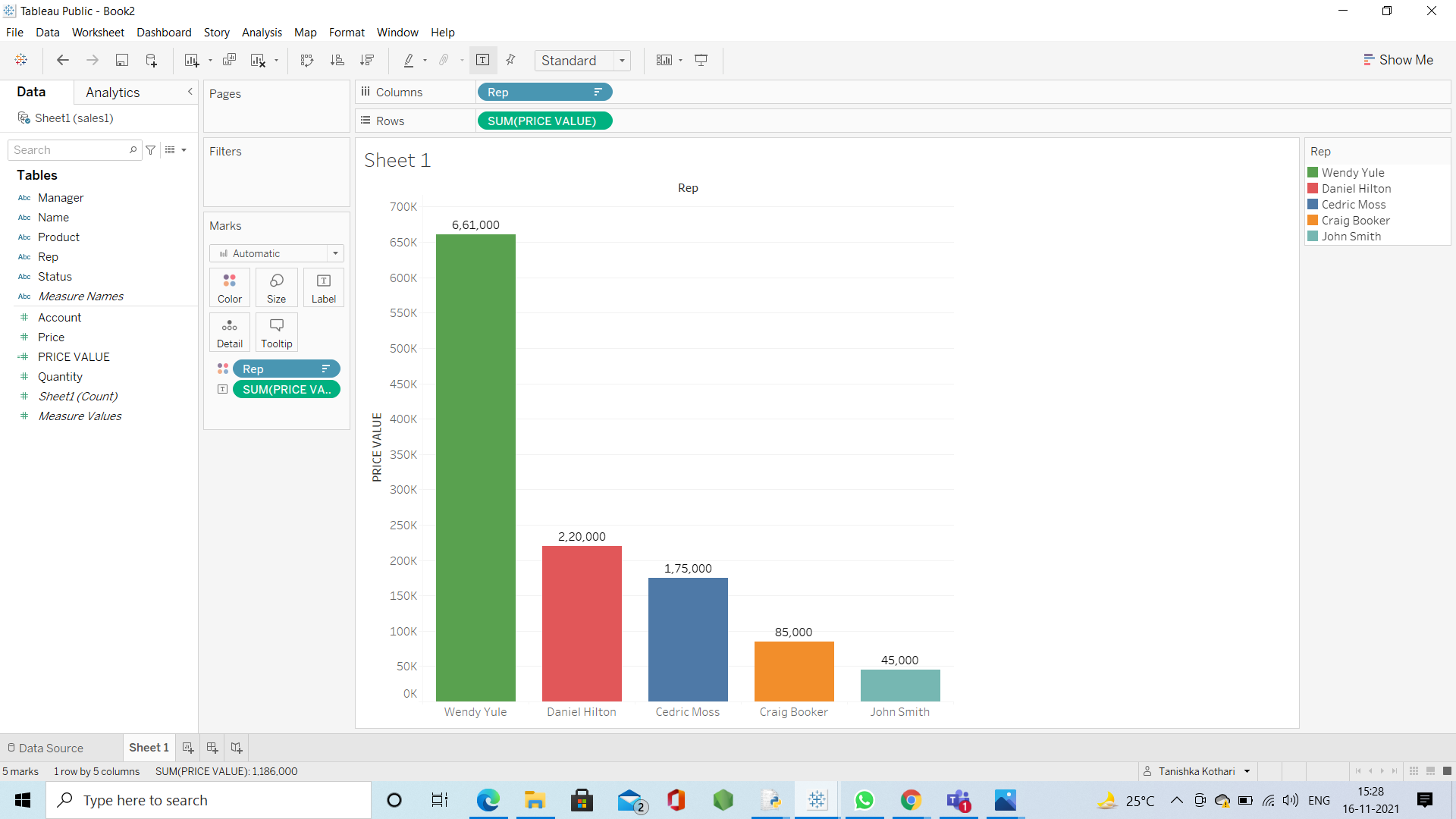


Here, according to the question, we need to plot a bar graph to get which representative had sold the highest quantity. So, in this above graph, we set

Y-axis: Quantity X-axis: Rep

Hence, we got the required Bar Graph. Thus, from the bar graph we know that Wendy Yule has sold the highest quantity

b)which representative has sold highest price value



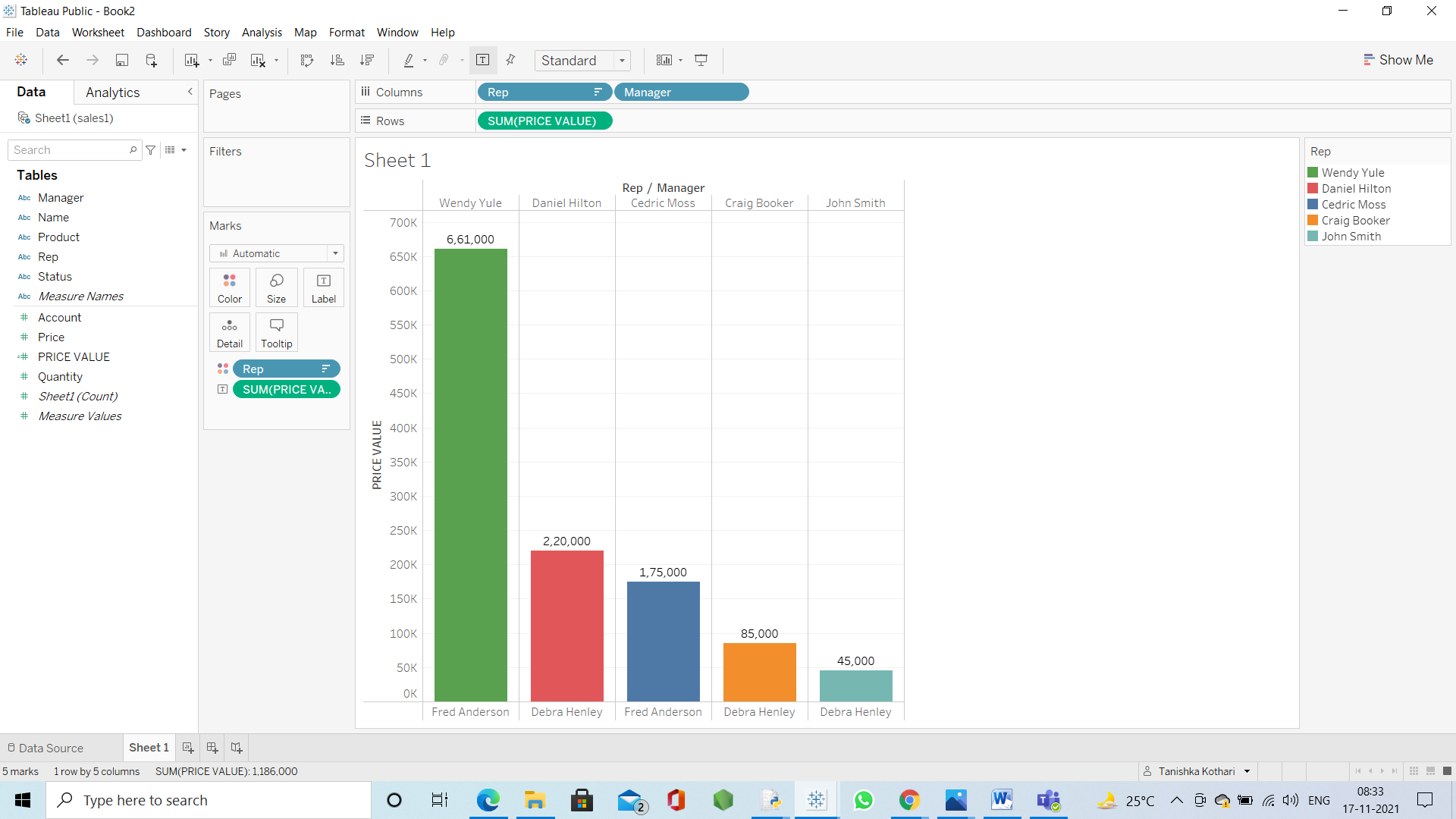
Here, according to the question, we need to plot a bar graph to get which representative had sold the highest price value. Here, since we need to get the price value, we have to make a calculated field named Total Sale.

Total Sale = [Price]\*[Quantity]

So, in this above graph, we set Y-axis: Total Sale X-axis: Rep

Hence, we got the required Bar Graph. Thus, from the bar graph we know that Wendy Yule has sold the highest price value.

c)who is the manager of the representative who sold highest price

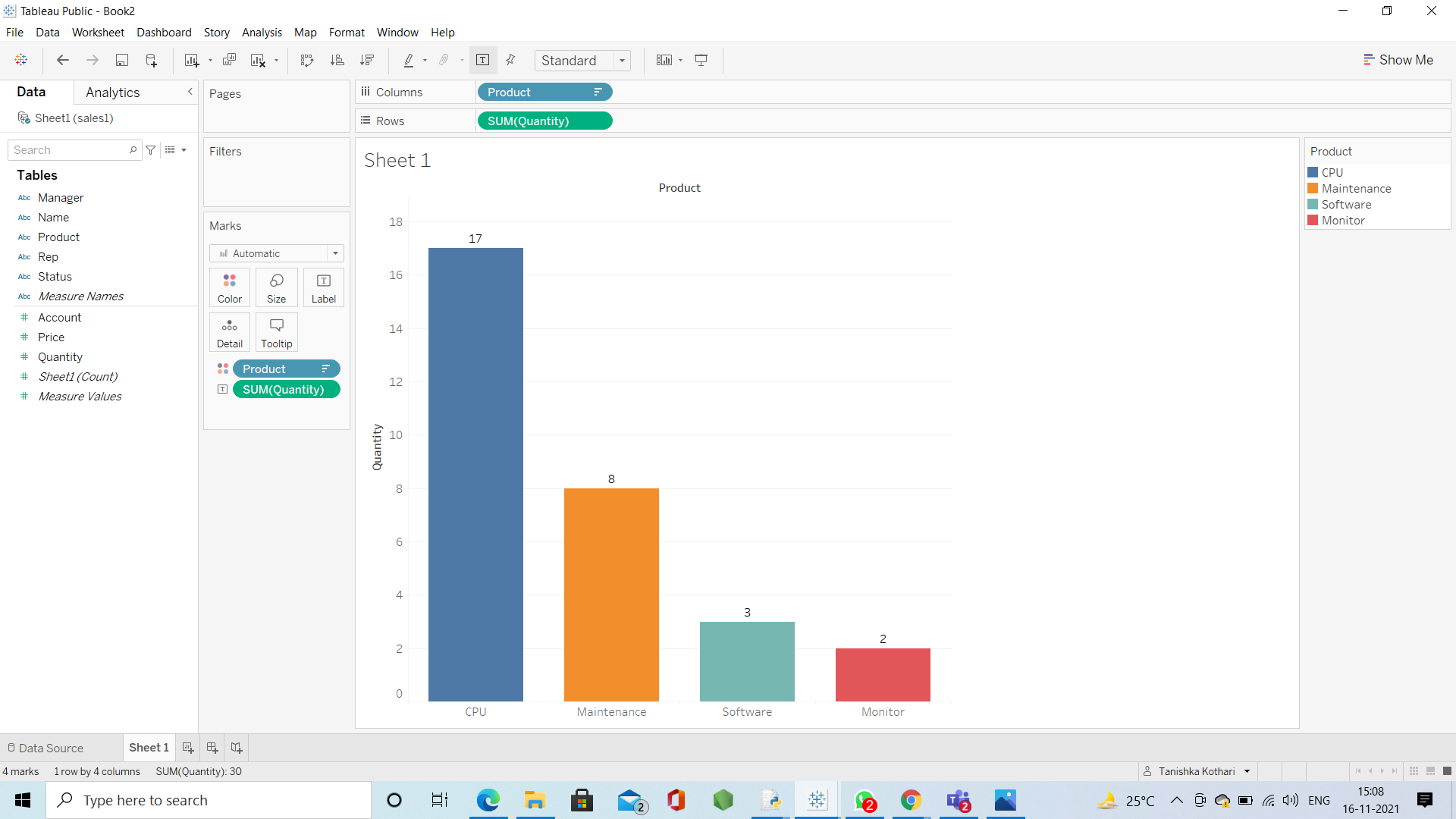


Here, according to the question, we need to plot a bar graph to get which representative had sold the highest price value. Here, since we need to get price value, we have to use Total Sale, which is a calculated field. Total Sale = [Price]\*[Quantity]

So, in this above graph, we set Y-axis: Total Sale X-axis: Rep

Hence, we got the required Bar Graph. Thus, from the bar graph we know that Wendy Yule has sold the highest price. And since Fred Anderson is Manager of Wendy Yule, the answer to the question will be Fred Anderson.

d)Which product is sold most



Here, according to the question, we need to plot a bar graph to get which product is sold the most. So, in this above graph, we set

Y-axis: Quantity X-axis: Product

Hence, we got the required Bar Graph. Thus, from the bar graph we know that CPUs are sold the most.