Creating a Data Narrative

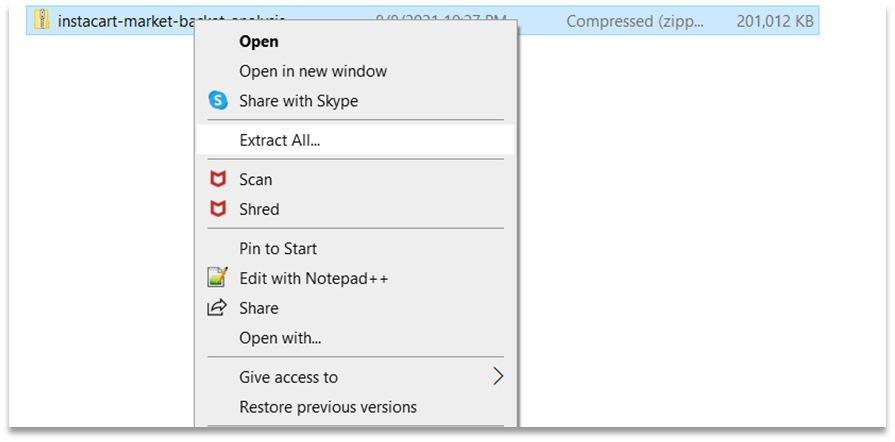
Please make sure you submit all screenshots with date / timestamps at the bottom right (windows users) or the top right (Apple users). Failure to do so will qualify your assignment for a discount.

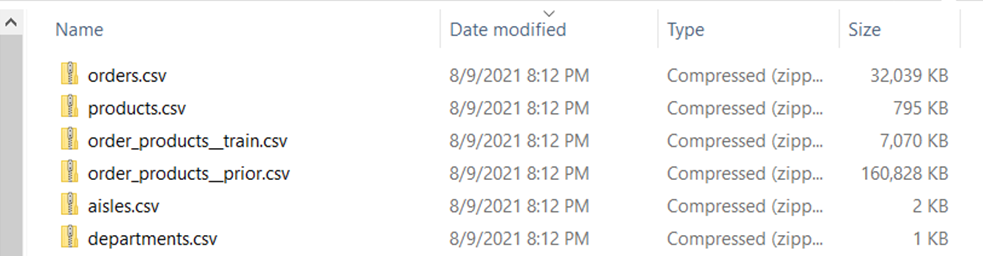
***Business Scenario***

The dataset for this assignment is obtained from Instacart, an electronic commerce grocery ordering and delivery organization. The dataset contains a sample of over 3 million grocery orders from more than 200,000 Instacart users. The dataset contains details about the orders placed by various users, the sequence of products purchased in each order, the week and hour of day the order was placed, and a relative measure of time between orders. It also has information about the aisle and department for the sold products.

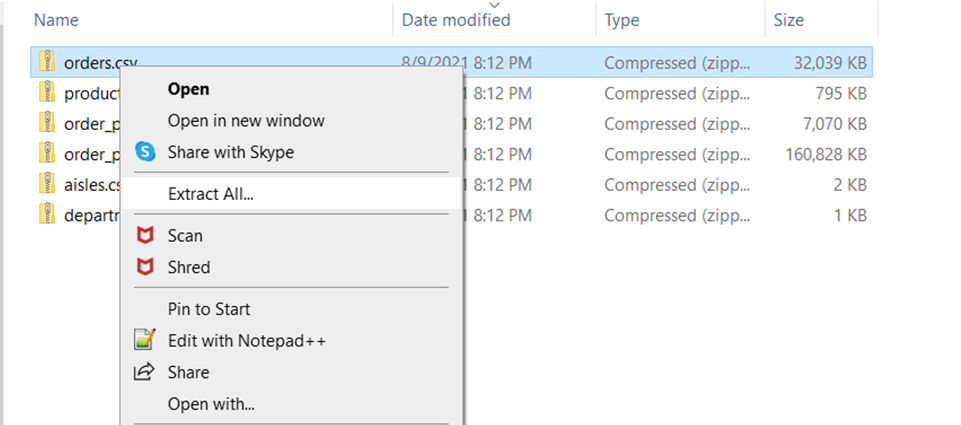
**Section 1: Acquiring Data**

1. Download the “instacart-market-basket-analysis” folder from e-learning.
2. *Unzip* the folder. (There are 6 compressed folders within this folder)

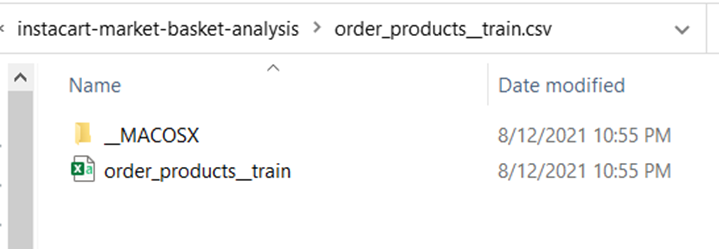




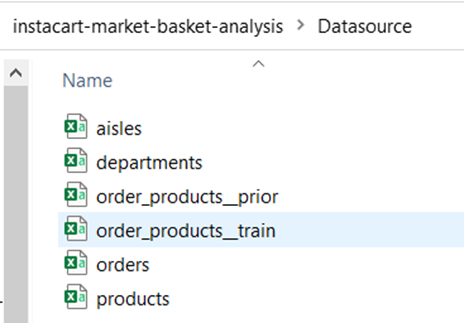
1. *Unzip* each of the 6 folders.



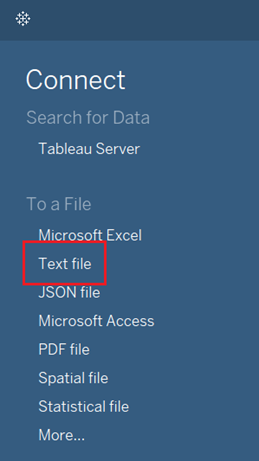
1. The csv files we will use in our analysis are located within each extracted folder.



1. Create a new folder called “Datasource” within the extracted “instacart-market-basket-analysis” folder.
2. Move all six of the csv files to this new folder.

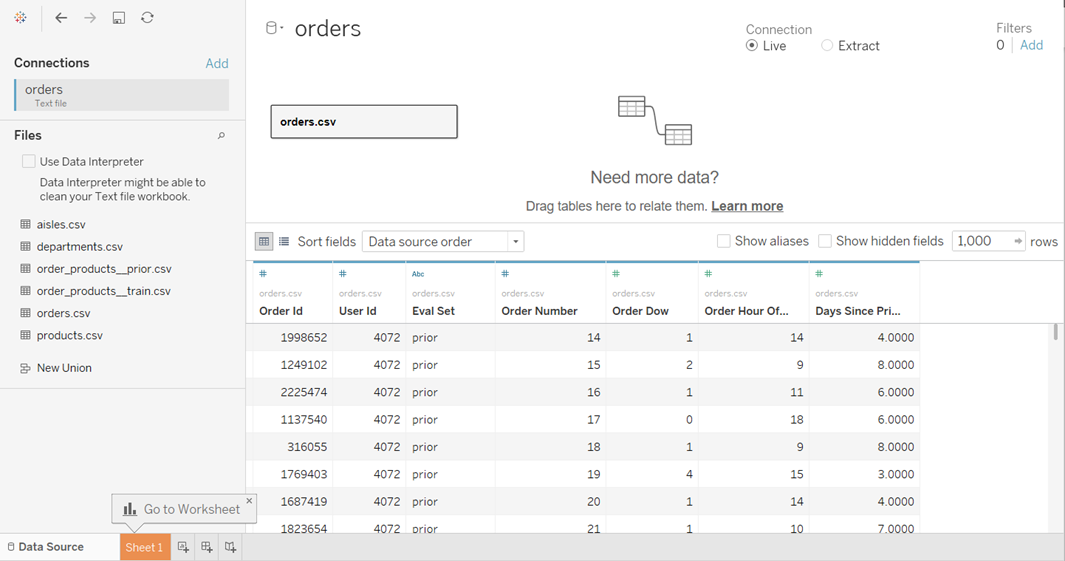


1. Download and launch the latest version of *Tableau Desktop*.
2. Under *Connect*, click *Text File* to connect to the downloaded data files.

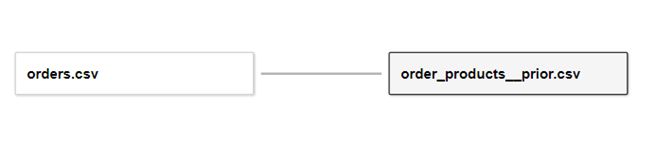


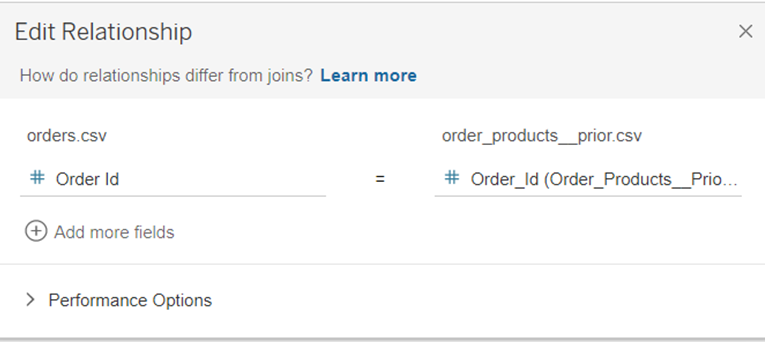
1. Navigate to the “Datasource” folder and select **orders.csv**.

Note: All the files in our folder are added using the *Files* section.



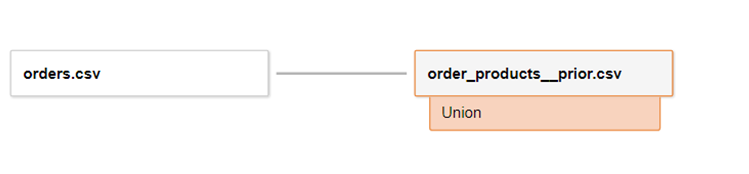
1. Drag **orders\_products\_\_prior.csv** from the *Files* panel next to the **orders.csv** box.





We are creating a *Relationship* between the **orders.csv** file and the **orders\_products\_train.csv** file. Relationships are a dynamic and flexible way to combine data from multiple data sources. They are like joins but have some differences. While using Relationships, data sources are not combined up front. A custom data source is created for every visualization you create based on the fields you use in your worksheet. Tableau automatically selects join types based on the fields being used in the visualization.

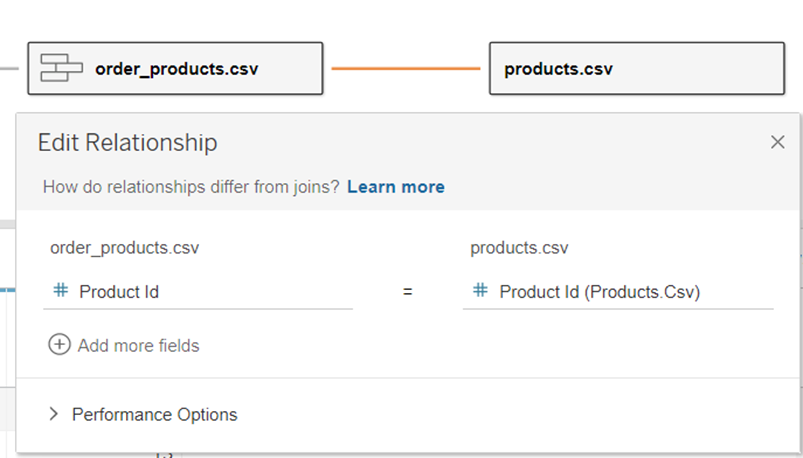
1. Make sure the correct fields of *Order Id* are used while creating the Relationship.
2. Next, drag **order\_products\_\_train.csv** from the *Files* panel right below the **order\_products\_\_prior.csv** box until you see an orange box that says union.



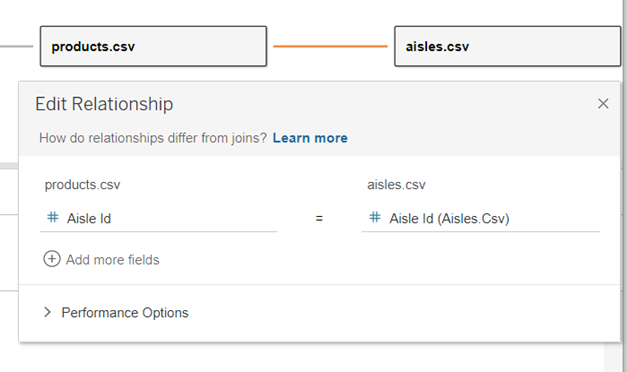


We have created a union of **order\_products\_\_prior.csv** and **order\_products\_\_train.csv**.

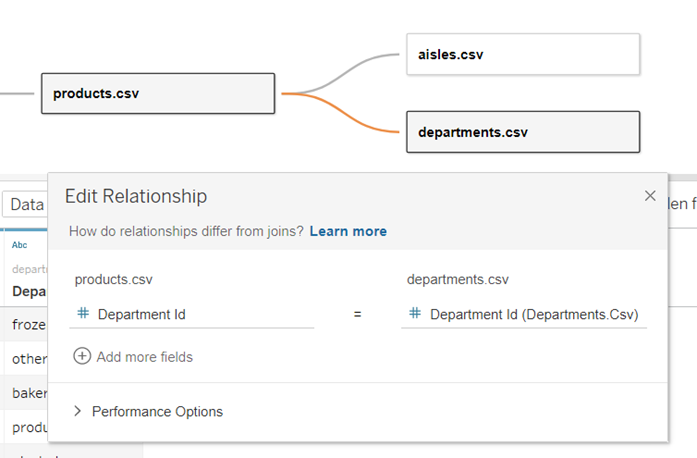
1. Right click on the newly created union box and rename it to **order\_products.csv.**
2. Drag **products.csv** next to the **order\_products.csv** box.



1. Drag **aisle.csv** next to the **products.csv** box.



1. Finally, drag **departments.csv** next to **products.csv** as well.



**Paste screenshots showing all the relationships…………………………………………………………………………………………………**

1. Click *Go to Worksheet*.

**Section 2: Defining the Audience**

(Refer to the reading assignment from Storytelling with Data: Let’s Practice)

What are the primary groups and individuals that are the focus of your communication?

List the executive role or position that is the key decision maker in your audience?

How much experience does your audience have working with this data? (shallow or deep; include at least one defining sentence of justification)

What is the scope of your audience? (wide or narrow; include at least one defining sentence of justification)

What does your audience care about? (you can circle back to this question after your analysis)

What action does your audience need to take? (you can circle back to this question after your analysis)

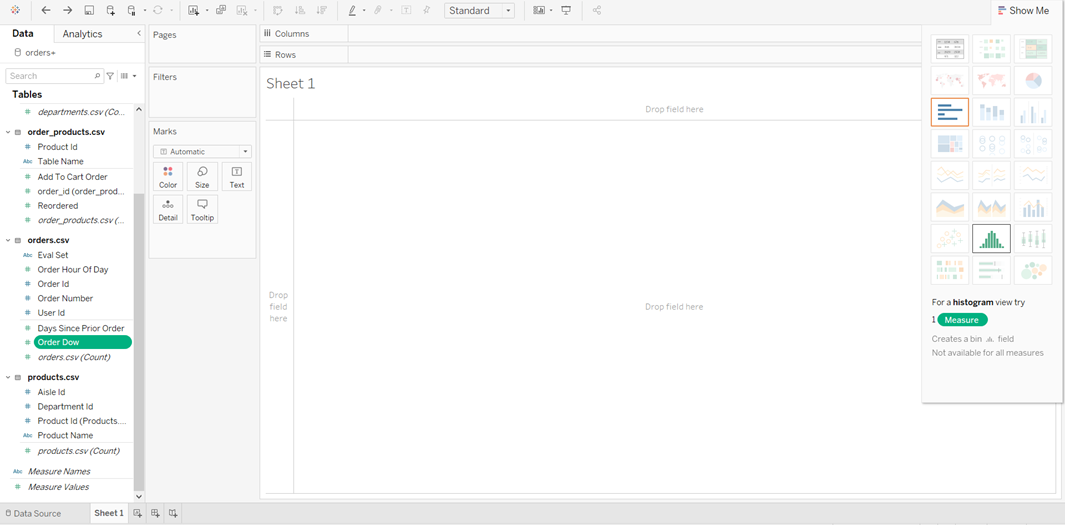
What are the stakes? (what matters, what is the point)

What is the big idea? (this will form the main storyline)

**Section 3: Data Analysis**

1. **Service Analysis**
2. Analyzing the Busiest Days of the Week (dow)
3. On the Data pane, under the **orders.csv** table, select **order dow**.
4. Click *Show me* on the toolbar and then click on the *Histogram* option.

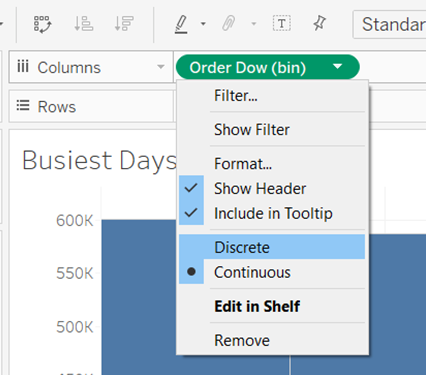
Note: *Show Me* can be accessed in the upper right corner of the screen. With field(s) selected, *Show Me* offers one-click options for chart types.



1. A *Histogram* visualization is automatically created for us. A new field called **Order Dow(bin)** is also added to the Data pane.
2. Select the newly created **Order Dow(bin)** field on the *Data* pane and click on the drop down.



1. Then select *Edit*.
2. Change the *bin size* to 1.
3. Click on the dropdown on the **Order Dow(bin)** pill on the *Columns* shelf and select *Discrete*. Notice that the pill changes in color from green to blue.



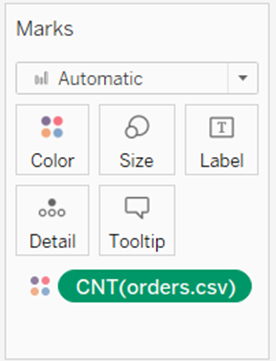
1. Rename the worksheet to “<Student Name> - Busiest Days of the Week” by double clicking the sheet tab at the bottom.

**Paste a screenshot of the worksheet…………………………………………………………………………………………..**

**Question 1: What are the two busiest days of the week?**

1. Analyzing the Busiest Hours
2. Click on the *New Worksheet* icon at the bottom of the page.
3. Drag **Order Hour of Day** (orders.csv table) from the *Data* pane to the *Columns* shelf. Then click on it using the *dropdown arrow* and select *Discrete*.
4. Next, drag **orders.csv (Count)** from the *Data* pane to the *Rows* shelf.
5. Drag **orders.csv (Count)** to *Color* on the *Marks* card.

Note: Tableau applies different colors to marks based on the type of field. For example, if you drag and drop a dimension (categorical data), each category is assigned a different color. If you drop a measure (continuous data), each mark in the view is colored based on its value.

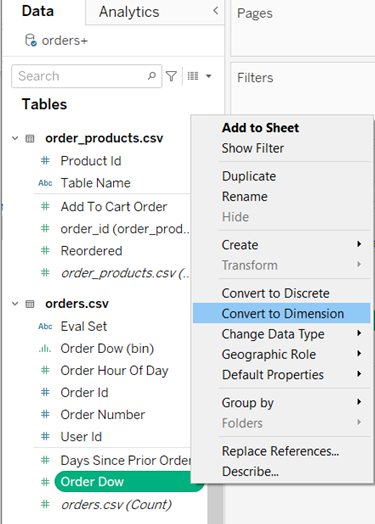


1. On the *Marks* card, click *Color > Edit Colors***.** From the drop down, choose *Red-Green Diverging*.

**Paste a screenshot of the worksheet…………………………………………………………………………………………..**

**Question 2: What are the busiest hours? List the top 11 busiest hours.**

1. **Analyzing the Busiest hours for each day.** Click on **Order Dow** on the *Data* pane and select *Convert to Dimension*. **Order Dow** is now converted to a dimension. Notice the change in color.

****

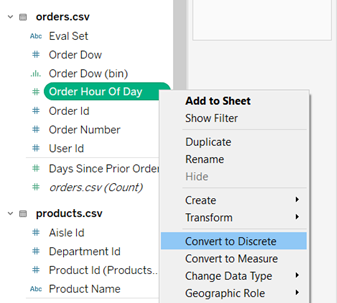
1. Drag **Order Dow** to the *Columns* shelf, to the left of **Order Hour of Day.** We can now analyze the busiest hours for each day of the week.
2. Rename the sheet “<Student Name> - Busiest Hours”.

**Paste a screenshot …………………………………………………………………………………………………………………………**

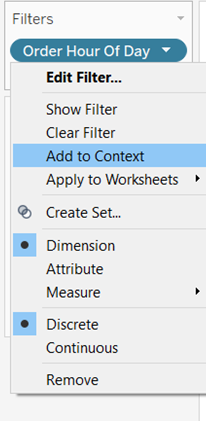
1. Top 10 Products in the Busiest Hours
2. Click on *New Worksheet*.
3. Drag **Product Name** from the *Data* Pane (under products.csv table) to the *Columns* shelf.

Note: For this worksheet, data is obtained from three tables – products, orders and order\_products.

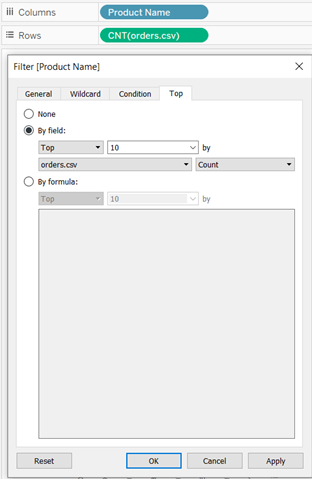
1. Drag **orders.csv (Count)** to the *Rows* shelf.
2. Change the **Order Hour of Day** field on the *Data* pane to *Convert to Discrete*.



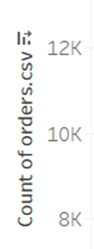
1. Drag **Order Hour of Day** to the *Filters* card. Click on the filter it and choose *Show Filter* to view the filter on the worksheet.
2. Select all the busiest hours we derived above in Question 2 and click *OK*.
3. Click on the drop down on the **Order Hour of Day** pill under the *Filters* card and select *Add to Context*.



1. Next drag **Product Name** to the *Filters* card.
2. On the window that pops up, navigate to the *Top* tab, and ensure the *Top 10* from **orders.csv** is selected.



1. Click *OK*.
2. Sort the visualization, by clicking on the *Sort* icon next to the *Row* axis label.



1. Rename the sheet to “<Student Name> - Top 10 Products in the Busiest Hours”.

**Paste a screenshot of the worksheet…………………………………………………………………………………………..**

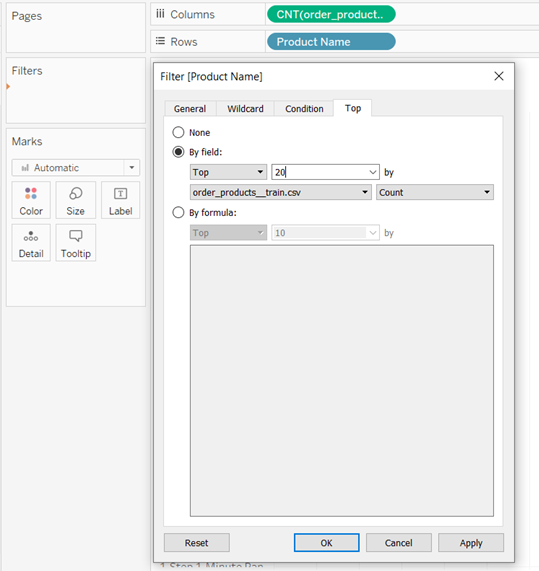
**Question 3: What product is sold most often during the busiest hours?**

1. Days Since Prior Order. How long after their previous order, do customers order again?
2. Click on *New Worksheet*.
3. Select the **Days Since Prior Order** field on the *Data* pane.
4. Click on *Show Me* and select the *Histogram* chart.
5. Click on *Show Me* again to close it.
6. Edit the newly created **Days Since Prior Order (bin)** field on the *Data* pane to change the *Bin size* to 1.
7. Rename the sheet to “<Student Name> - Days Since Prior Order”.

**Paste a screenshot of the worksheet…………………………………………………………………………………………..**

**Question 4: When are customers most likely to order again?**

1. **Product Analysis:**
2. Best Selling Products
3. Click on *New Worksheet*.
4. Drag **Product Name** from the products.csv table to the *Rows* shelf.
5. Drag **orders\_products.csv (Count)** from the *Data* pane (under orders\_products.csv table) to the *Columns* shelf.
6. Drag **Product Name** to the *Filters* card and choose the following options on the *Top* tab and click *OK*. We now have a list of the *Top 20* bestsellers.

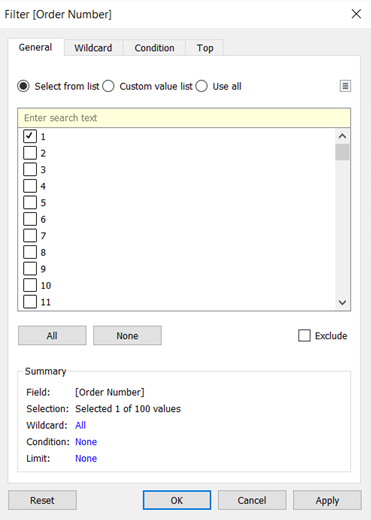


1. Using the menu bar icons for the view, *Sort* the visualization in *Descending* order.
2. Rename the sheet to “<Student name>-Best Selling Products”.

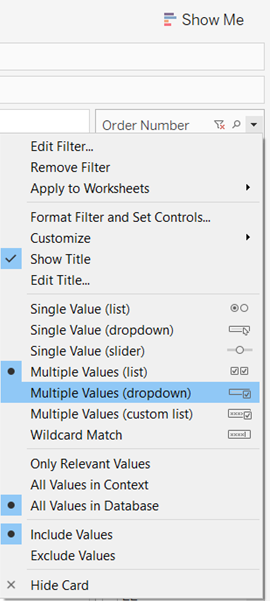
**Paste a screenshot of the worksheet…………………………………………………………………………………………..**

**Question 5: What are the Top 3 bestselling products?**

1. Top 10 Products on First Order. What are the top products that customers will most likely buy while ordering for the first time?
2. Drag **Order Number** to the *Filters* card and on the pop-up window select *1*.



1. Add the **Order Number** filter to context by selecting the *Add to Context* option.
2. Click on the **Order Number** in the *Filter* section and select *Show Filter* to display the filter on the worksheet.
3. Click on the dropdown at the top right corner of this filter and select *Multiple Values (dropdown)*.

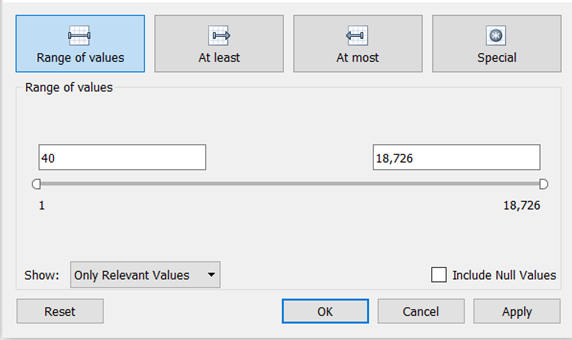


**Paste a screenshot………………………………………………………………………………………………………………………………..**

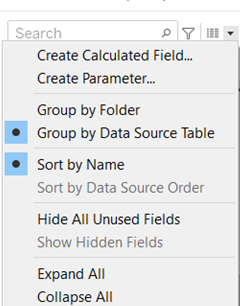
1. Top 10 Reordered Products in terms of their probability of being reordered
2. Click *New Worksheet*.
3. Drag the **Reordered** field from the orders\_products.csv table to the *Columns* shelf.
4. The default aggregation is *Sum*. Change it to *Average*.
5. Drag **Product Name** to *Rows*.

Note: The **Reordered** field has a value of either 1 or 0. 1 indicates that the product has been reordered. The order\_products table contains a list of orders placed by customers and all the products that constituted that order. So, by calculating the average of the **Reordered** field for each product, we are calculating the probability of the product being reordered

1. Drag **orders\_products.csv (Count)** to the *Filters* card and make the following changes. We want to focus our analysis on products that have been included in at least 40 orders. Set the filter to *Range of values* and set the minimum to *40*. Click *OK*.



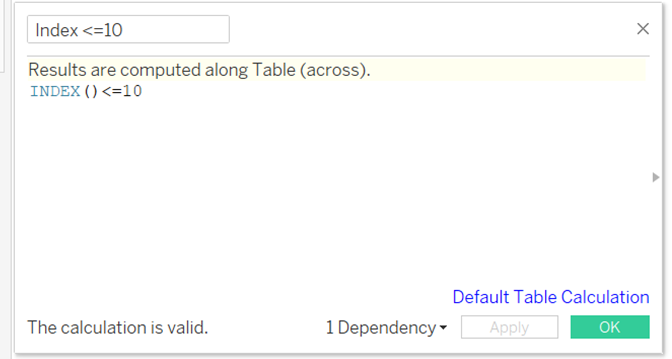
1. Using the menu bar icons, *Sort* the visualization view in *Descending* order.
2. On the *Data* pane, click on the *drop down arrow* next to the *Search* window and select the *Create Calculated Field* menu.



1. Enter the following calculation.

*Index <=10*

Click *OK* when finished. Index is a function that returns the first N rows of the view.



1. Drag this new **Index Calculated Field** to the *Filters* card and select *True* on the pop-up window.
2. Rename the analysis worksheet “<Student name>-Top 10 Reordered Products”.

**Paste a screenshot of the worksheet…………………………………………………………………………………………..**

**Question 6: What is the most reordered product? List its probability of being reordered as well**

1. Number of Items Purchased in an Order
2. Click *New Worksheet*.
3. Using the formula below, create a new *Calculated Field* called **Number of Items per Order**.

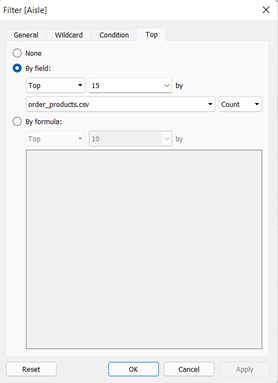


1. Using the *Show Me* option, create a *Histogram* chart for the calculated field **Number of Items per Order** and change the bin size to *1*.
2. Rename the sheet “<Student name>-Number of Items Purchased per Order”.

**Paste a screenshot of the worksheet…………………………………………………………………………………………..**

**Question 7: How many items do customers most often order?**

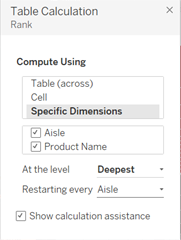
1. **Category Analysis:**
2. Treemap of Aisles and Products – Most Popular Aisles
3. Click *New Worksheet*.
4. On the *Data* pane, select **Aisle** from the aisle.csv table and holding down the *ctrl key* for *multiple selection*, also select the **order\_products.csv (Count)** field. Next, select the *Treemap* chart from the options in the *Show Me* option.
5. Drag **Aisle** to the *Filters* card and make the following selections. Using the *Top* tab, select *By field*, insert *Top 15* by **order\_products.csv** and *Count*. Click *OK*.



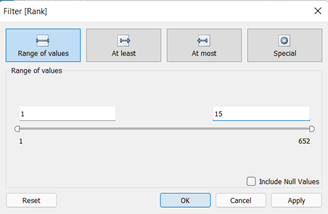
1. Drag the **Aisle** field to the *Marks Color* encoding. Drag the **Product Name** field to the *Marks Label* encoding.
2. Create a new *Calculated Field* called **Rank** using the formula found below.



1. Drag the new *Calculated Field* called **Rank** to the *Filters* card, click on it and select *Edit Table Calculation*. Select *Specific Dimensions*, *Aisle*, *Product Name*, *at the level Deepest*, *Restarting every Aisle*, and *Show calculation assistance*.



1. Next, edit the **Rank** Filter and change the *Range of values* from *1* to *15*. Click *OK*.



1. Rename the analysis worksheet “<Student name>-Treemap of Aisles and Products”.

**Paste a screenshot of the worksheet…………………………………………………………………………………………..**

**Question 8: List the top 3 aisles that customers mostly order from**

1. How often (frequency) are products from each of the departments sold
2. Click *New Worksheet*.
3. On the *Data* pane, select **Department** from the departments.csv table and holding the *ctrl key* multiple select, also choose the **order\_products.csv (Count)** field.
4. Create a *Bubble* chart using the *Show Me* function.
5. Drag the **Department** field to the *Marks Color* encoding.
6. On the toolbar, change the fit of the worksheet from *Standard* to *Entire View*.
7. Rename the worksheet “<Student name>-Number of Products Sold per Department”

**Paste a screenshot of the worksheet…………………………………………………………………………………………..**

**Question 9: List the top 2 departments with the highest number of products sold?**

1. **Top Products**
2. Ranking each hour of each day of the week based on number of orders placed
3. We need to Recreate the “<Student Name> - Busiest Hours” worksheet. We will perform this with instructions to copy the analysis worksheet.
4. Create a new blank analysis worksheet.
5. At the very bottom right corner of the workbook, click on the *Filmstrip* icon in the menu bar.

Icon

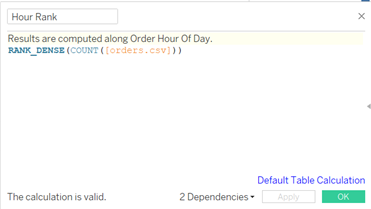
Description automatically generated with medium confidence

1. Select the Busiest Hours worksheet from the thumbnails presented. Right click and select the menu *Copy*.

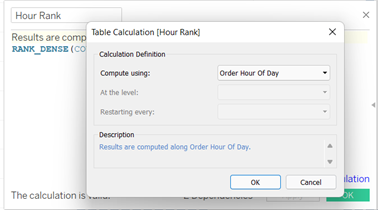
Chart, bar chart

Description automatically generated

1. Create a new worksheet. Right click on the new worksheet and select the menu *Paste*. At the bottom of the workbook select the *Tabs* icon to the right of the *Filmstrip* icon.
2. Remove the **CNT(orders.csv)** encoding from the *Marks* card.
3. Create a new calculated field called **Hour Rank** using the formula in the screenshot below.



1. Click on the blue *Default Table Calculation* function make the selection **Order Hour of the Day**. Click *OK*.



1. Drag the newly created **Hour Rank** calculated field to the *Color* and *Label*encodings on the *Marks* card.
2. Rename the sheet “<Student name>-Busiest Hours – Rank”.

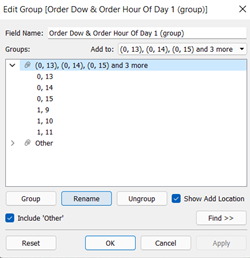
**Paste a screenshot of the worksheet…………………………………………………………………………………………..**

1. Grouping Order Hour of Day and Order DOW
2. Recreate the “<Student Name> - Busiest Hours-Rank” worksheet using the *Copy* and *Paste* approach.
3. Select the top three hours for Order Hour of Day for Order DOW 0(Saturday) and 1(Sunday). You can select multiple bars us the *ctrl key*.

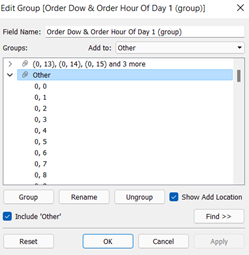
Chart, bar chart, histogram

Description automatically generated

1. On the pop-up that appears, click on the *Group Members*icon and select *All Dimensions*.
2. A new group fieldcalled **Order Dow & Order Hour Of Day (group)** is created in the *Data* pane.
3. Click on the newly created **Order Dow & Order Hour Of Day (group)** and select *Edit Group*. Rename the group that you created to **Weekend Busiest Hours**.



1. Click on the *Other* group and holding the *ctrl key* multiple select hours 16, 17, and 18 for each day.

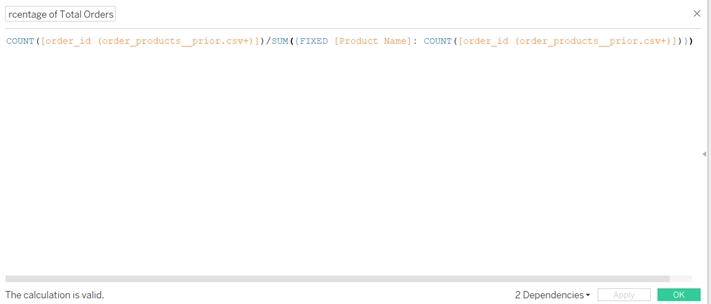
 Graphical user interface

Description automatically generated

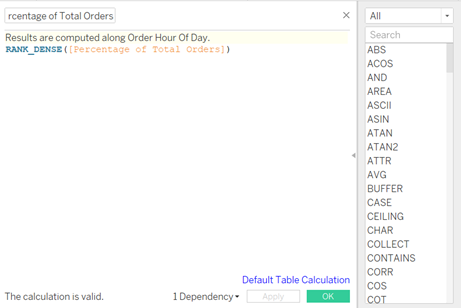
1. Next, click on the *Group* icon and rename the group **Evening Hours**.
2. Click on the *Other* group icon again and holding the *ctrl key* select the top 3 busiest hours for Friday. Rename the group **Friday Rush Hours**.
3. *Group* the 6, 7, and 8th hour of each day of the week and rename it **Early Birds**.
4. *Group* the 21, 22, and 23rd hour of each day and rename it **Late Night Hours**.
5. Similarly, *group* the top 3 busiest hours for each day Monday through Thursday and rename it **Monday-Thursday Rush Hours**.
6. Remove **Hour Rank** from the *Marks* card and drag the **Order Dow and Order Hour of Day (group)** to *Color* encoding on the *Marks* card.
7. Rename the sheet “<Student name>-Busiest Hours – Groups”.

**Paste a screenshot of the worksheet…………………………………………………………………………………………..**

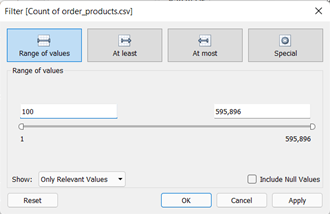
1. Analyzing the top products ordered per group
2. Click *New Worksheet*.
3. Create a calculated field called **Percentage of Total Orders** using the formula shown below.



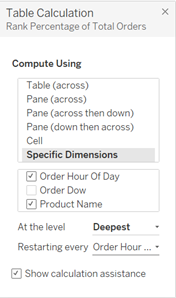
1. Create another calculated field **Rank - Percentage of Total Orders** as shown below. Alter the *Default Table Calculation* setting to compute using the field **Order Hour Of Day**.



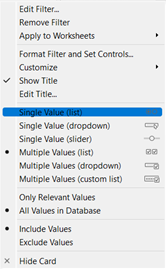
1. Drag **Order Dow** and **Order Hour of Day** to the *Columns* shelf.
2. Drag **Product Name** to the *Columns* shelf.
3. Drag **order\_products.csv (Count)** to the *Filters* card.
4. Select the **CNT(order\_products.csv)** filter and choose *Edit Filter*. Limit the *Range of values* to a minimum of *100*. Click *OK*.

****

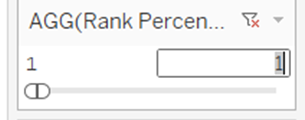
1. Next add **Rank - Percentage of Total Orders** to the *Filter* card. Click on the drop down arrow and select *Edit Table Calculation*. Choose *Specific Dimensions*, *Order Hour of Day*, *Product Name*, *At the level Deepest*, *Restarting every Order Hour*, and *Show calculation assistance*.



1. Click on the dropdown arrow again and select the menu path *Show Filter*.
2. Drag **Order Dow & Order Hour of Day (group)** to the *Filters* card and select **Friday Rush Hours** from the list. Next, click on it and select *Show Filter*. Change the filter to a *Single Value List*.



1. Drag the **Percentage of Total Orders** field to *Rows*.
2. Drag the **Product Name** field to the *Color* and *Label* encodings on the *Marks* card.
3. Change the **AGG(Rank Percentage of Total)** filter to display the top product for each hour.



1. On the toolbar, change the fit of the worksheet from *Standard* to *Entire View*.
2. Rename the sheet “<Student name>-Top Products per Group-Friday Rush Hours”.
3. Duplicate this worksheet using the copy worksheet approach and select **Early Birds** on the **Order Dow & Order Hour of Day (group)**.
4. Rename the sheet “<Student name>-Top Products per Group-Early Birds”.
5. We need a separate worksheet for each customer group for our data narrative. Repeat the copy worksheet approach for each customer group in **Order Dow & Order Hour of Day (group)**. You should have a total of 6 worksheets, one for each of the Order Dow & Order Hour of Day groups. **(Early Birds, Evening Hours, Friday Rush Hours, Late Night Hours, Monday -Thursday Rush Hours and Weekend Busiest Hours)** You do not have to create a worksheet for the *Other* group.

**Paste a screenshot of each of the 6 customer group worksheets…………………………………………………………………………………………..**

**Section 4: Curate your Views**

(Refer to the Lecture on Narrative Visualization)

Curate your visualization views by selecting the best evidence displaying insights. To expedite the process, we recommend using this list of visualization worksheets that you created. Make sure to remove all extra worksheets and dashboards from your workbook leaving only the tabs for worksheets you will use.

* Busiest Hours
* Busiest Days of the Week
* Days Since Prior Order
* Number of Items Purchased
* Best Selling Products
* Treemap of Aisles and Products
* All six of the Top Products per Group worksheets

**Section 5: Organizing the Storyline with Storyboarding**

(Refer to the reading assignment from Storytelling with Data: Let’s Practice)

(Refer to the Lecture on Narrative Visualization)

A good way to review the evidence you have is to write the title of each visualization on a sticky note and place them on a flip chart or white board. Storyboarding is the process of trying out different views in a sequence to tell a good story.

**Insert a smartphone picture of your storyboarding process including the view titles written on sticky notes in sequence…………………………………………………………………………………………..**

Choose your preferred storyline approach using Freytag’s Pyramid or Neil Cohn’s Visual Grammar. Organize your storypoints into a meaningful storyline. List your storypoints in a storyline sequence. Make sure to list the storyline approach you selected and each phase of the storyline beside the view. An example is provided below.

Freytag’s Pyramid

1. Busiest Hours - Beginning
2. Busiest Days of the Week – Rising Action
3. …

**Provide your sequenced list of views in a numbered list. Make sure to include the storyline approach as the heading.…………………………………………………………………………………………..**

**Section 5: Creating a Story in Tableau**

(Refer to the Lecture on Narrative Visualization)

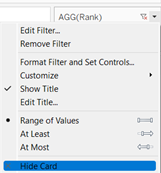
1. To create a story, click on the *New Story* icon on the bottom right of the workbook. You will find all the worksheets you have created in the worksheet tabs to the left.
2. To create a story point, select the worksheet you want as your first story point story from the list in the *Story* panel. Drag the worksheet you selected to the *Drag a sheet here* canvas on the right.

Graphical user interface, application, Word

Description automatically generated

1. Before you drag a worksheet to the story, best practice suggests hiding all the legends and filters from each worksheet. We are building a story, not an online application. Each frame/view in your story should only contain the storypoint and the visualization. We are sequentially narrating a story that grew from our analysis. It is not appropriate for us to leave a bunch of selection options in a view and expect the audience to perform analysis. Let’s hide those selectable options. In each worksheet, choose the worksheet tab and then select the dropdown arrow on the *Filter* or *Legend* and select *Hide Card*.

Note: In some scenarios, it may be acceptable to have one view or dashboard with a couple of selection options to handle questions about alternative scenarios. A data narrative should not contain more than one view with selections.



1. To create each new story point, click on *Blank* in the *Story* panel.



(Refer to the reading assignment from Storytelling with Data: Let’s Practice)

**Section 6: Adding Rhetoric to Your Story**

(Refer to the Lecture on Narrative Visualization)

From your analysis you are going to shape a story about a new promotional program for reducing the time between grocery orders. Review the explanations of each type of rhetoric and what scenario is best for this purpose.

**List the type of Rhetoric that is best for this story .…………………………………………………………………………………………..**

1. Compose rhetorical captions for each story point by clicking on the *Add a caption* box. These captions will be captured above each view. Each caption should capture the most interesting insight or fact presented in that view. The phrase chosen should be consistent with the type of rhetoric being used.
2. Make sure that all of your captions are a single phrase or sentence.

Note: Captions are not titles. Titles that frame the analysis in each view are created in the analysis worksheets.

Graphical user interface, text, application, email, website

Description automatically generated

**Provide a Screenshot of all of the Captions in your Story similar to the one above.…………………………………………………………………………………………..**

**Section 7: Formatting the Story**

Data narratives should be polished and formatted for high grade publishing.

1. Select the menu path *Story>Format*.
2. For the overall story *Title*, set the *Font* to *Tableau Regular, 16pt*.
3. For the *Navigator* captions, set the *Font* to *Tableau Book, 11pt*. Set the *Shading* to the teal blue on the right side of the color panel.
4. For Text Objects, set the *Font* to *Tableau Book, 10pt*.

Graphical user interface, application, Word

Description automatically generated

**Provide a Screenshot of your story format settings similar to the one above.…………………………………………………………………………………………..**

1. In the Story pane, click on *A Drag to add text* and place the text box at the bottom left of the first view. Change the font to *8*. Type “Data source: Instacart.” in the text box. Select *left justification* for the text.

**Provide a Screenshot of the first view in your story with the data source text box.…………………………………………………………………………………………..**

1. Format the Titles in each worksheet used in your story. Click on the tab for the worksheet used in your first view. Right click on the *Title* and select the menu *Edit>Title*. Change the title from <Sheet Name> to the title of the analysis. Change the font size to *12*. You must highlight the text and then select the font size to change the text. Click *OK*. Repeat this process for each worksheet in your story.

Graphical user interface, text, email

Description automatically generated

1. Format the Axis Titles in each worksheet used in your story. Right click on the *Legend Label* used on the left(rows, y axis) of the visualization. Select the *Edit Axis* menu. Edit the *Axis Title* with a more accurate phrase for your analysis. Edit the Axis Title on the top(columns, x axis) in the same manner. An example of a better *Axis Title* for “Dow” would be “Day of the Week” or “Weekday”. Repeat this process for each worksheet used in your story.

Graphical user interface, application

Description automatically generated

**Provide a Screenshot of the first view in your story with edited Titles and Axis Titles.…………………………………………………………………………………………..**

**Provide a Screenshot of the last view in your story with edited Titles and Axis Titles.…………………………………………………………………………………………..**

**Section 8: Managing Deliverables**

1. Attach the assignment document in Microsoft Word in eLearning. Include only screenshots and answers with proper question numbers.
2. Attach your Tableau workbook file saved with the extension TWBX in eLearning
3. Submit your Answer Document and your TWBX file in eLearning.