Assignment: Olist Marketing Funnel

Unlike previous assignments, the workbook and datasets must be downloaded separately. A key skill being tested by the assignment is the ability to establish connections and relationships between multiple data tables. Thus, they will not be pre-connected to the workbook, and you'll have to add them to the data source as part of the task list.

As a short introduction to the data, Olist is a Brazilian e-commerce company that facilitates the delivery of products from sellers to customers all across the country. Olist provides a central platform for customers to find products they want to purchase, while they also work with sellers to provide the logistics to deliver their products. (In many ways, this is similar to eBay.) In this assignment, we'll focus on the seller side of things. In particular, we'll be looking at aspects of the marketing funnel that takes a prospective seller from being a lead to closing, and becoming an active seller on the Olist platform.

Below, **text in bold** represents the name of a table (csv), while *text in italics* represents the name of a field (column).

1. Let's start by loading in the marketing leads table (**olist_marketing_qualified_leads_dataset**). A lead is generated by a prospective seller arriving at a landing page via some means, such as search engine results (paid or organic), social media posts, or even direct traffic onto the website.
    a. After connecting to the data, on Sheet 1, create a bar chart of the number of leads generated by each form of *origin*.
    b. Which method of generating leads was the most popular overall?
    c. How many leads were generated from an unknown or Null origin?

2. Next, we can combine the marketing leads data with the closed leads table (**olist_closed_leads_dataset**). After a lead is generated, they are contacted by Sales Development Representatives (SDRs) and Sales Representatives (SRs), who consult with the lead on becoming a seller on the Olist platform. If the lead is successfully converted into a seller, the lead is said to have been closed; only those closed leads are represented in this new table.
    a. Add the closed leads table to the data source through use of a relationship. Make sure that the relationship is defined through the *mql_id* feature that is common to both tables.

b. To easily analyze whether or not a lead became a seller, engineer a boolean feature "*Is Seller*" that takes the value True if the lead was closed or False if the lead failed to close.
    - *Hint*: The *seller_id* field will have a value if a lead is present in both data tables, but will be Null if a lead did not close.
c. Modify the plot you created on Sheet 1 using the new "*Is Seller*" feature into a stacked bar chart, so that each bar is divided into segments based on if leads were closed or not.
d. How many leads that came in through paid search results were closed?

3. Let's also look at how long it takes for different types of leads to close.
    a. Engineer a feature "Days to Close" that reports the number of days it took for a lead to be closed, i.e. the number of days between the *won_date* from the **closed leads** table and *first_contact_date* from the **qualified leads** table.
    b. On Sheet 2, create a bar chart of the average number of days to close each lead, grouped by the seller's *business segment*.
    c. Among all segments with at least 10 closed leads (*Hint*: use a filter using the appropriate dataset!), which business segment has the lowest average number of days to close?

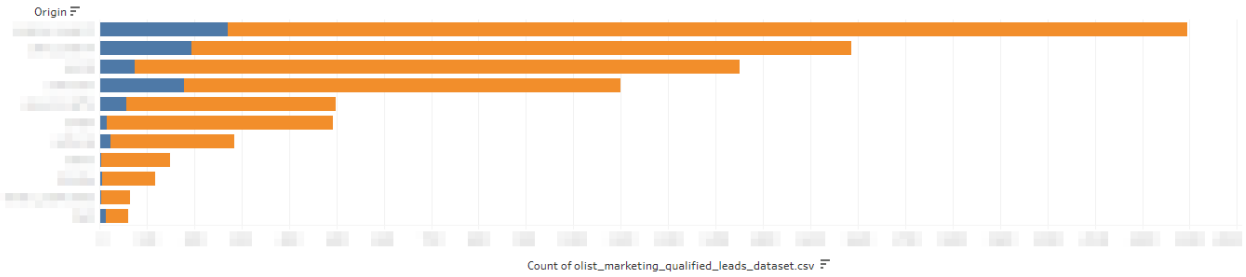
4. Finally, let's add in the final table, the seller information (**olist_sellers_dataset**). This table lists information about where the seller originates from.
    a. Add the sellers table to the data source through the use of a relationship. Make sure that the relationship is defined through the *seller_id* feature that is common to the **closed leads** and **sellers** tables.
    b. On Sheet 3, create a bar chart of the number of closed leads by the state that they are based out of.
    c. The seller information and closed leads information come from different sources, so they don't overlap completely. For how many closed leads do we not have their state information (i.e. their *state* is Null)?
    d. On Sheet 4, create a table with the seller's state on one axis, the lead origin on the other axis, and the number of closed leads generated from each combination as text values. (*Hint*: Use the Label facet of the Marks card)
    e. Are there any states where there were more leads generated through paid search than organic search? If so, which states were these?

Visualization Reference

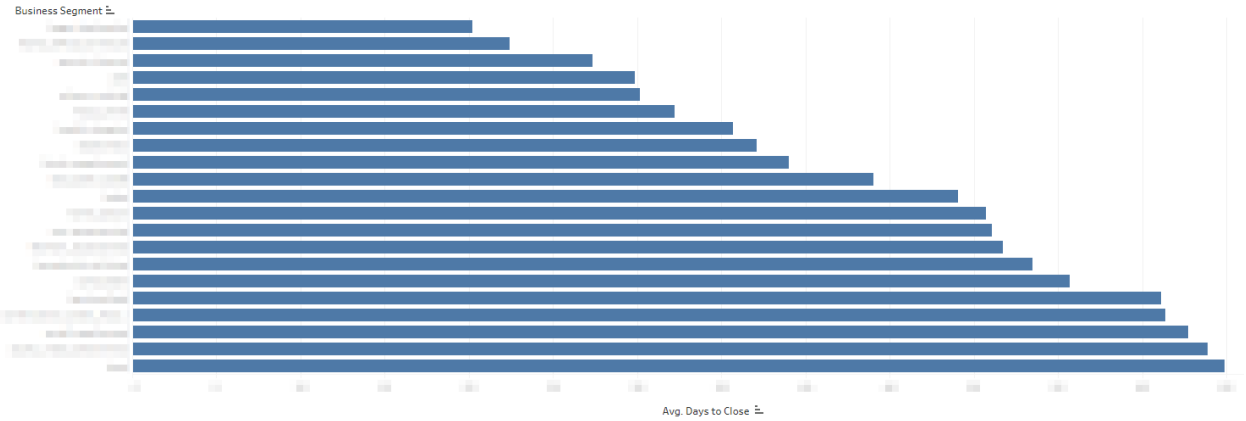
You can check your work on the SkillBuilder assignment by referencing the images below. Each image is a general depiction of what you should see on each Sheet of the workbook, with various details hidden or omitted. **Your visualizations do not need to look exactly like these reference images!** These should just be a general guide to check if you're on the right track.

Sheet 1 (Question 2c)

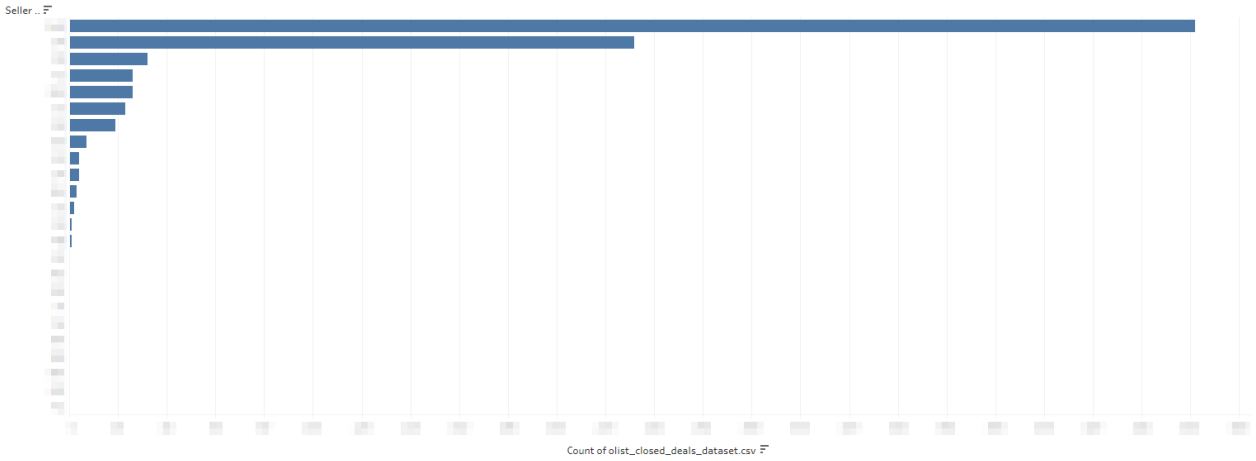
The initial bar chart created in Question 1a will not include the colored sections.



Sheet 2 (Question 3b)



Sheet 3 (Question 4b)



Sheet 4 (Question 4d)

	Seller State
Origin	Null
Null	10