

HW1

Instructions

- To complete this homework assignment, download the related files, create the necessary visualizations, and answer the questions. Follow the directions below to complete this assignment.

Requirements

- One member of your group should submit a .twbx file for each question. Be sure to complete all steps of all questions before submitting your final file.

Q1

Connecting to Data and Combining Data

1. Connect to the Data_to_Clean.xls file.
2. Use the data interpreter to clean the data.
3. Split the City/State column and rename the resulting columns.
4. Change the Sales and Profit fields to both be currencies, Discount to be a %, and Row ID to a dimension.
5. Save the data source so it can be reused.
6. Combine it with the SuperStoreManagers.xlsx file.
7. Create a bar chart to find out who was the best manager by sales and profit.

Q2

You are trying to find which of your customers has the most purchases (transactions) for a new store you opened called Wholesale Carbonated Beverages. You also want to know the quantity sold for each of your products. You have two Excel spreadsheets as your data sources. Download the Wholesale Carbonated Beverages Products and Customers.xlsx file and the Wholesale Carbonated Beverages Transactions.xlsx file. One contains sheets or tables for Products and Customers, and the other contains Transactions. You need to connect to these two data sources, create relationships between the various tables, and build visualizations to answer the following questions:

1. Use View Data for each table to identify the common fields between the tables. Which two tables contain the Transaction ID field?
2. Create the relationship using all the tables. Create a visualization showing the transaction count by customer name in descending sort order. Which customer had the most purchases?

3. On a new worksheet, create a visualization showing the quantity sold by product name in descending sort order. Which product had the highest quantity sold?

Q3

You have information about orders and returns for an office supply store in a data source called **Orders and Returns.xlsx**. This data source contains two tables: **Orders** and **Returns**. You'll join them and then build two views, and answer the following questions.

In the first part of the scenario, you want to see only orders by product category which have been returned.

In the second part of the scenario, you want to see how the number of orders returned compares to the number of orders not returned by product category.

Scenario Part One

1. Which join type should you use to see and compare only orders which have been returned?
The **Orders** table is on the left, and the **Returns** table will be on the right.
2. Using the appropriate join type, create a join between the **Orders** table and the **Returns** table using the **Order ID** fields. Create a view that shows the distinct count of returned orders, and name the view **Orders Returned**. How many orders were returned for the Tables product category?
3. Which product categories had the most returned orders?

Scenario Part Two

1. Which join type should you use to see all orders regardless if they have been returned?
The **Orders** table is on the left and the **Returns** table on the right.
2. Change the join type to the appropriate join type for your intended analysis. In the visualization you created in scenario part one, use the Status field to differentiate between orders returned and not returned using color. Of the three product categories that had the most returns, which category had the most orders made overall?
3. Of the product categories with no returns, which had the most orders overall?

Q4

Imagine that you have sales orders in one data source and sales targets in another. You would like to create two visualizations to identify information that still needs to be obtained so you can compare all your sales against their targets. You decide to connect to these two data sources, blend them together using a common dimension, and then build two visualizations to answer these questions:

1. On the first worksheet named "Orders First", build a bar chart with **Orders** as the primary data source, showing **Sales** broken down by **Category**. Create a combined axis chart that

includes **Quota**. Swap the rows and columns axes, edit the bar colors to show **Sales** in green and **Quota** in gray. Use the view you have built to answer the following question. Which product categories do not have sales target information?

2. On the second worksheet named “Targets First”, build a bar chart with **Targets** as the primary data source, showing **Quota** broken down by **Category**. Create a combined axis chart that

includes **Sales**. Swap the rows and columns axes, edit the bar colors to show **Sales** in green and **Quota** in gray. Use the view you have built to answer the following question. Which product categories do not have any sales yet?

Q5

You are researching air travel trends, and you want to know the total distance specific airlines traveled in July 2021. You have an Excel workbook with several similarly-named tables, each of which contains data for a different week in July 2021, and one of which contains data for July 2020. You would like to use a wildcard search to combine all the July 2021 tables in a union in Tableau Desktop so you can analyze data for the entire month at once in Tableau.

1. Create a wildcard union to include the four 2021 tables but not the 2020 table. Then rename the union “Unioned Airline Data July 2021”. Explore the fields and field values in the data grid and metadata grid for reference fields and null values. Note the **Path** and **Sheet** reference fields, automatically created by Tableau. Note the similarly named fields **Carrier** and **Name (Carrier)**, as well as **Carrier Code** and **Code (Carrier)**. Some of these fields contain many null values. Recall what you learned about fields that are added after a union. What next step should you take?
2. After merging the mismatched fields, rename the newly merged columns to “Airline” and “Airline Code”. Build a view that shows total distance traveled by airline across all the weeks of data in July 2021. Looking at the total distance each airline traveled across the four weeks of data, which airline traveled the largest total distance?