

Project | Data Wrangling (Excel)

Project Overview

This project is focused on wrangling data and analyzing data using pivot tables and complex formulas.

This project is an excellent opportunity to showcase the skills you are learning in Excel. Therefore, you are encouraged to not only focus on getting the correct answer but also to focus on your submission appearance. Once you score 90% or higher on this project, you can submit your project to your [professional portfolio](#) in Wix, which you can later share throughout your professional journey as you apply for promotions or new jobs. Therefore, professional appearance should also be a priority in your submission.

Project Resources

In this assignment, we have three (3) mock data sets to use. These are not real data sets but randomized mock data sets of various dimensions and measures to help explore the concepts discussed in this module. The data sets were created using: <https://www.mockaroo.com/>

Download and save these 3 data sets:

- [Data_Set_1.csv](#)
- [Data_Set_2.csv](#)
- [Data_Set_3.csv](#)

Format & Submission

Your submission should be one Excel Workbook.

Pay careful attention to the instructions on how to label each Excel tab. Be sure to label each part of the project according to the instructions.

Submit your file and make sure to follow this naming convention:
Project1_LastName.xlsx (e.g., Project1_Smith.xlsx)

Project Instructions

Part A: Wrangling

1. Import Data_Set_2.csv and Data_Set_3.csv into Data_Set_1.csv. Each data set should be a separate tab labeled as the original file name.
2. Format the data in Data_Set_1 as a table with named columns.
3. Sort the data in Data_Set_1 in ascending ID order.
4. Data_Set_1 is missing three (3) columns of data. This data can be found in Data_Set_2. Use a formula learned in class to bring these additional columns (UserName, CompanyName, JobTitle) into columns N, O, and P in the Data_Set_1 tab. Validate that the data came over accurately.
5. Create a new column on the Data_Set_1 tab in column Q called "FullName" and use an equation learned in class to combine the first and last names into the same cell. Be sure to include a space between the first and last names.
6. Replace all of the hyphens in the credit_card_types with spaces.
7. Freeze the top row.
8. Filter data to only include customers using an American Express credit card.
9. Change the tab name from "Data_Set_1" to "Part A".

Part B: Analyzing Customer Demographics

(Use the best equation to solve these problems. Do not use pivot tables, and make sure "flexible" equations are used, meaning the data set can be updated without the need to update your equations. You may have the "right" answer, but if your answer does not change as data is updated, points will be deducted).

Create a new tab labeled "Part B". Answer Parts B1-B3 on this tab.

1. What is the **unique** (no duplicate counting) count of customer diversity (race)? In other words, determine the unique number of White, Black or African American, Japanese, etc. customers. (Hint: use the remove duplicates tool and create another tab using a unique list of FullNames).
2. What is the **unique** (no duplicate counting) count of customer gender? In other words, determine the unique number of male and female customers.
3. Create a table with the total sales amount by gender and race in the data set. Format the sales as currency. Which gender/race combination has the most / least total sales? Use conditional formatting to illustrate the magnitude.
4. Create a graph illustrating the total sales amount by gender and race.

Part C: Analyzing Customer Sales

(Use the best equation to solve these problems. Do not use pivot tables, and make sure "flexible" equations are used, meaning the data set can be updated without the need to update your equations. You may have the "right" answer, but if your answer does not change as data is updated,

points will be deducted).

Create a new tab labeled "Part C". Answer Parts C1-C3 on this tab.

1. Determine the top 5 "best" customers defined by the most number of purchases.
2. Determine the top 5 "best" customers defined by the total value of purchases.
3. Use an equation learned in class and pull in the companies of those customers that made the top 5 number of purchases from Part C1 and C2.

Part D: Wrangling

Change the tab's name from "Data_Set_3" to "Part D". Answer Part D1 on this tab.

1. Create three additional columns called Month, Month#, and Year. Use formulas learned in class to extract the month name (e.g., January), month number (e.g., 1), and year into columns Q, R, and S, respectively.

Part E: Analyzing Sales

(Do not use equations to solve these problems; use pivot tables.)

Create a new tab labeled "Part E". Answer Parts E1-E5 on this tab using the data in the tab labeled "Part D."

1. Create a table for the total sales amount by year. Which year had the highest / least amount of sales? Format the sales as currency. Use conditional formatting to illustrate the magnitude of the sales. Create a graph to illustrate total sales

by year.

2. Create a table of the average sales amount by month using the month name from D1 (e.g., January). Which month had the highest / lowest average sales amount? Format the sales as currency. Use conditional formatting to illustrate the magnitude of the sales. Create a graph to illustrate the average sales amount by month.
3. Show the total sales amount by the customer's company. Order the list by total sales in descending order (most to least). Format your pivot table to only show the top 5 companies. Format the sales as currency.
4. Show the total sales amount by credit card company. Order the list by total sales in descending order. Which credit card company has the most sales, and which has the least? Format the sales as currency. Use conditional formatting to illustrate the magnitude of the sales.
5. Show the average sales amount by state. Which state has the highest / lowest sales average? Format the sales as currency. Use conditional formatting to illustrate the magnitude of the sales.