Case Study Project - Coffee Shops

We have included a [sample solution](#) at the end of this document. Check it out to see what a passing solution to this case study looks like.

You can use any tool that you want to do your analysis and create visualizations. You will only need to use DataCamp Workspace to write up your summary in Markdown.

We have included everything you need to complete the project in this document. Read it in detail before you get started.

Company Background

Caffeine Form is a company creating coffee cups from recycled material. Although they started selling the products on their website last year, the results were not as good as they expected. To better enter the market, they decided to collaborate with local coffee shops to advertise and sell their coffee cups.

The marketing team is trying to come up with the best criteria to choose possible collaborators by investigating the local market. They think focusing on the places with the most reviews might have the biggest impact as more good reviews will create more excitement for their cups.

Customer Questions

They would like you to answer the following questions to help:

- What is the most common place type in this local market?
- How does the range in number of reviews differ across all shops?
- How does the number of reviews vary across each place type?

Dataset

The dataset contains the information about all coffee shops in this new market, where one row is one shop. The dataset can be downloaded from [here](#).

The dataset needs to be validated based on the description below:

Column Name	Criteria
Region	Character, one of 10 possible regions (A to J) where coffee shop is located.
Place name	Character, name of the shop.
Place type	Character, the type of coffee shop, one of "Coffee shop", "Cafe",

	"Espresso bar", and "Others".
Rating	Numeric, coffee shop rating (on a 5 point scale).
Reviews	Numeric, number of reviews provided for the shop. Remove the rows if the number of reviews is missing.
Price	Character, price category, one of "$", "$$" or "$$$".
Delivery option	Binary, describing whether there is a delivery option, either True or False.
Dine in option	Binary, describing whether there is a dine-in option, either True or False. Replace missing values with False.
Takeout option	Binary, describing whether there is a takeout option, either True or False. Replace missing values with False.

Submission Requirements

1. You are going to create a written report summarizing your findings. Use the project task list provided below for guidance in the tasks you should complete and information to include in the report.

2. You can use any tools you want to do your analysis and create visualizations.

3. You will need to use DataCamp Workspace to write up your findings and share visualizations. Use the Markdown Guide for your reference to write your report in the DataCamp Workspace.

4. You must use the data we provide for the analysis.

5. Use the grading rubric provided below to check your work before submitting the report.

Project Task List

Data Validation

1. Check the data matches the criteria in the data dictionary.
2. Describe the validation tasks you completed and what you found. Have you made any changes to the data to enable further analysis?

Data Discovery and Visualization

1. Use exploratory analysis methods to answer the business questions in the project brief.
2. Create at least two different data visualizations to demonstrate the characteristics of variables.
3. Create at least one data visualization to demonstrate the relationship between two or more variables.
4. Describe what you found in the analysis and how the visualizaitons answer the business questions in the project brief.

Grading Rubric

You will be graded against the following criteria. You must pass all criteria to pass this part of the certification.

Domain	*Description*	**Sufficient**	**Insufficient**
Data Validation	Assess data quality and perform validation tasks	Has validated all variables against provided criteria and where necessary has performed cleaning tasks to result in analysis-ready data.	Has not conducted all the required checks and/or has not cleaned the data. May have removed data rather than performed cleaning tasks.
Data Visualization	Create data visualizations to demonstrate the characteristics of data and represent relationships between features.	Has created at least two different types of data visualization that highlight characteristics of individual variables after validation. Has created at least one visualization that shows the relationship between two variables. Has used visualizations that support the findings being presented.	Has used the same visualization throughout. Has not included graphics to represent single variables and relationships. Has not used visualizations that support the findings being presented.
Communication	Presents data concepts to small, diverse audiences	For each analysis step, has explained their findings and/or the reasoning for selecting approaches.	Has not provided a summary for each step (data validation, exploratory analysis).

Sample Solution

You can find a sample solution from a published workspace link [here](#). The sample solution demonstrates the required format for the final submission (i.e. a published workspace), and sufficient content needs to be included against the grading rubric. However, the sample solution is not the only solution.