Sample Practical Task

Instructions

* Utilize any tools you are comfortable with, such as Tableau, Power BI, or MS Excel, to accomplish the required tasks.
* Document your solutions in a Word file.
* Ensure that all visualizations created to fulfill the tasks are included.
* There is no need to provide any code.

Background

**Caffeine Form** is a company that produces coffee cups from recycled materials. They sell their cups to coffee shops through their website and aim to establish direct partnerships with these shops.

The company believes that collaborating with stores that have a higher number of reviews will enhance their product marketing. Their current objective is to focus on understanding the relationship between the types of shops and the quantity of reviews they receive. They now require a report detailing how the types of shops are correlated with the number of reviews.

Data

The dataset can be downloaded from [here](https://s3.amazonaws.com/talent-assets.datacamp.com/coffee.csv?).

**Column Name Criteria**

|  |  |
| --- | --- |
| Column Name | *Information* |
| Region | *Nominal. Where the store is located. One of 10 possible regions (A to J). Missing values should be replaced with “Unknown”.* |
| Place Name | Nominal. The name of the store. Missing values should be replaced with “Unknown”. |
| Place Type | Nominal. The type of coffee shop. One of “Coffee shop”, “Cafe”, “Espresso bar”, and “Others” Missing values should be replaced with “Unknown”. |
| Rating | Ordinal. Average rating of the store from reviews. On a 5 point scale. Missing values should be replaced with 0. |
| Review | Continuous. The number of reviews given to the store. Missing values should be replaced with the overall median number. |
| Price | Ordinal. The price range of products in the store. One of “$”, “$$” or “$$$” Missing values should be replaced with ”Unknown” |
| Delivery Option | Nominal. If delivery is available. Either True or False Missing values should be replaced with False. |
| Dine in Option | Nominal. If dine in is available. Either True or False Missing values should be replaced with False. |
| Take out Option | Nominal. If take away is available. Either True or False Missing values should be replaced with False. |

Tasks

Submit your answers directly in the word file.

**1. For every column in the data:**

a. State whether the values match the description given in the table above.

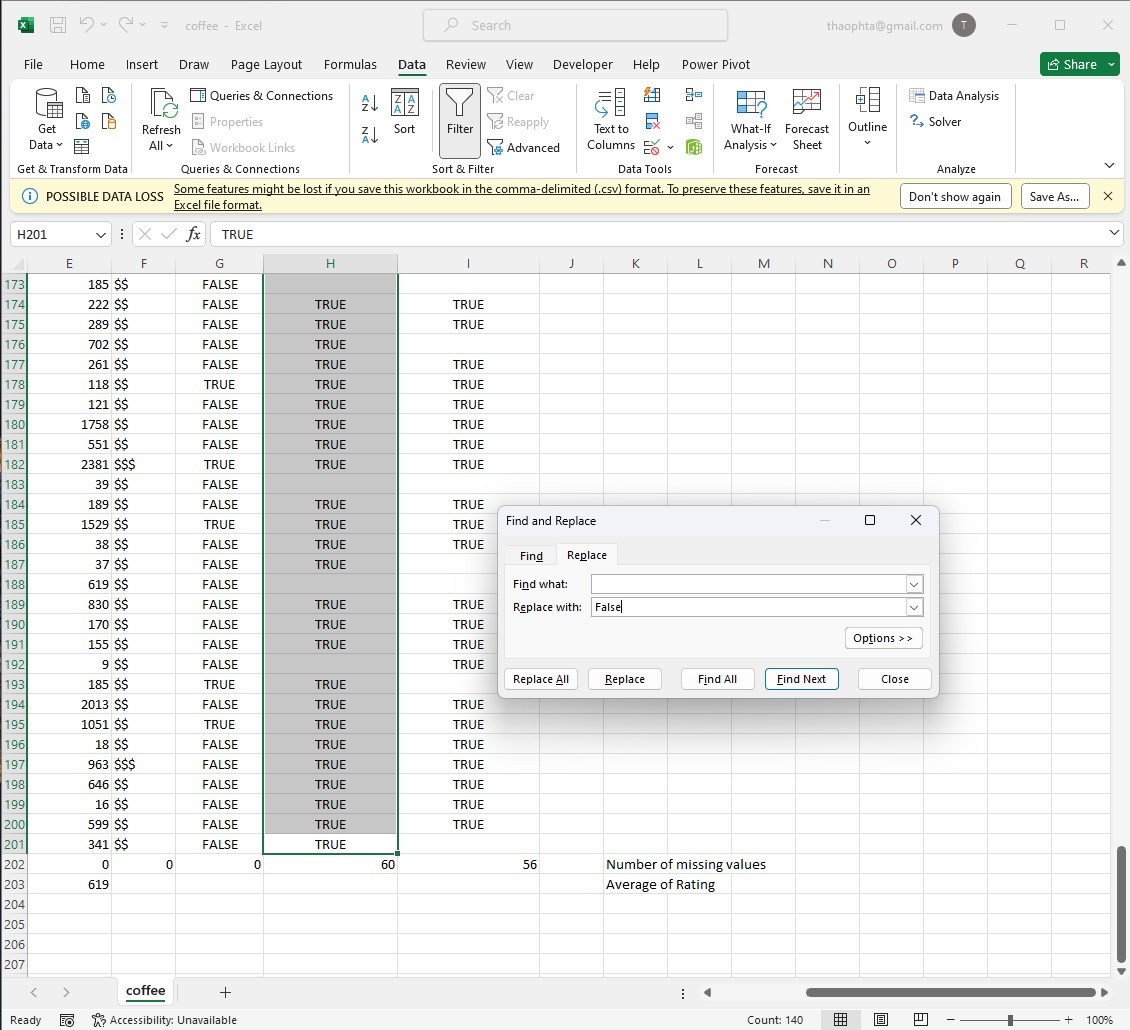
|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Column Name** | **Data Type Given** | **Data Type Recommended** |
| 1 | Region | Nominal | Agree (Nominal) |
| 2 | Place Name | Nominal | Agree (Nominal) |
| 3 | Place Type | Nominal | Agree (Nominal) |
| 4 | Rating | Ordinal | Agree (Ordinal) |
| 5 | Review | Continuous | Discrete |
| 6 | Price | Ordinal | Agree (Ordinal) |
| 7 | Delivery Option | Nominal | Binominal |
| 8 | Dine in Option | Nominal | Binominal |
| 9 | Take out Option | Nominal | Binominal |

b. State the number of missing values in each column.

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Column Name** | **Number of missing values** | **Value to be replaced (missing values)** |
| 1 | Region | 0 | Unknown |
| 2 | Place Name | 0 | Unknown |
| 3 | Place Type | 0 | Unknown |
| 4 | Rating | 2 | 0 |
| 5 | Review | 2 | Median number (619) |
| 6 | Price | 0 | Unknown |
| 7 | Delivery Option | 0 | False |
| 8 | Dine in Option | 60 | False |
| 9 | Take out Option | 56 | False |

c. Explain the steps taken to align values with the provided descriptions if any discrepancies were identified.

* Check all data and replace the missing values as shown in photo below.
* Under each column, use the following formula to count number of missing values in A column: “=COUNTBLANK(A2:A201)”. Then copy that formula to all columns, we will have the results as shown in the table above.
* Use formula “=ROUND(AVERAGE(E2:E202),0)” to calculate the median value in Review (Result is 619)
* Relace the blank cells in each column with values as shown in table above using Ctrl+F. First select whole column, Ctrl + F, select Replace then input the replaced value. Repeat process for each column.



***Figure 1: Replace missing values***

**2. Create a visualization that shows which is the most common type of coffee store. Use**

**the visualization to:**

1. State which category of the variable place type the most observations.

* Coffee shop is the category with the most observation (visualisation below).

1. Explain whether the observations are balanced across categories.

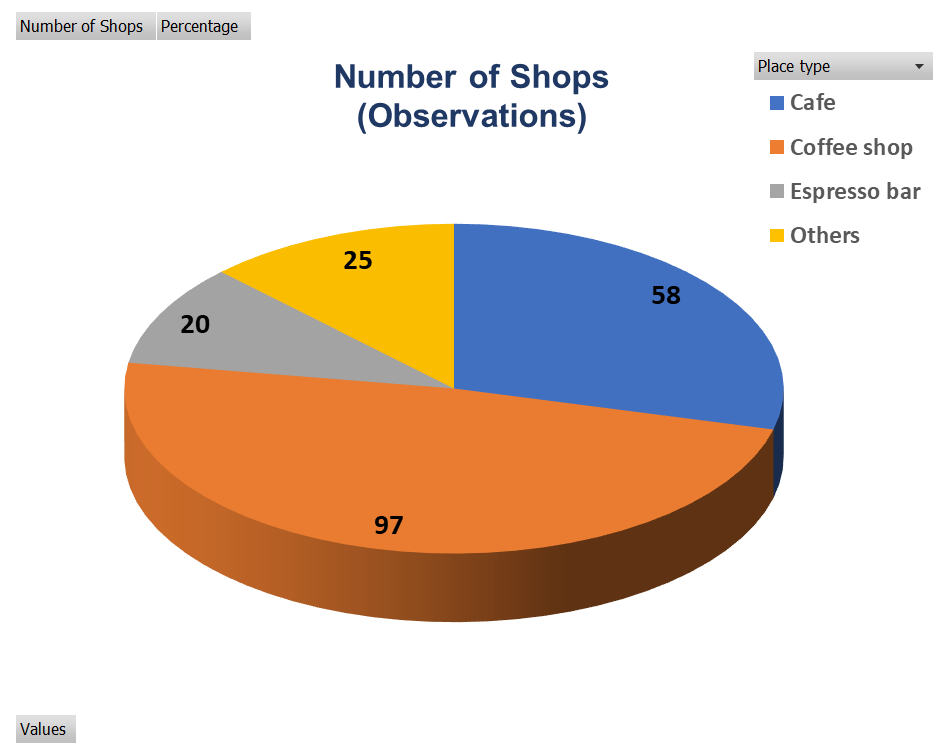
* Not balance, 4 categories with number of observations as: 97 (Coffee shop), 58 (Café), 20 (Expresso bar) and 25 (Others)

***Visualisation Process: using PivotTable and PivotChart***

* Insert the Pivot Table for dataset in which add “Place type” to Rows, “Place type” to Values twice (both using Count function). Change name for columns: Number of shops (Number of Observations) and other is Percentage over total observations as follows:

|  |  |  |
| --- | --- | --- |
| **Type of Store** | **Number of Shops** | **Percentage** |
| Cafe | 58 | 29.00% |
| Coffee shop | 97 | 48.50% |
| Espresso bar | 20 | 10.00% |
| Others | 25 | 12.50% |
| **Grand Total** | **200** | **100.00%** |

* Then Insert Pivot chart based on this PivotTable as in picture below.

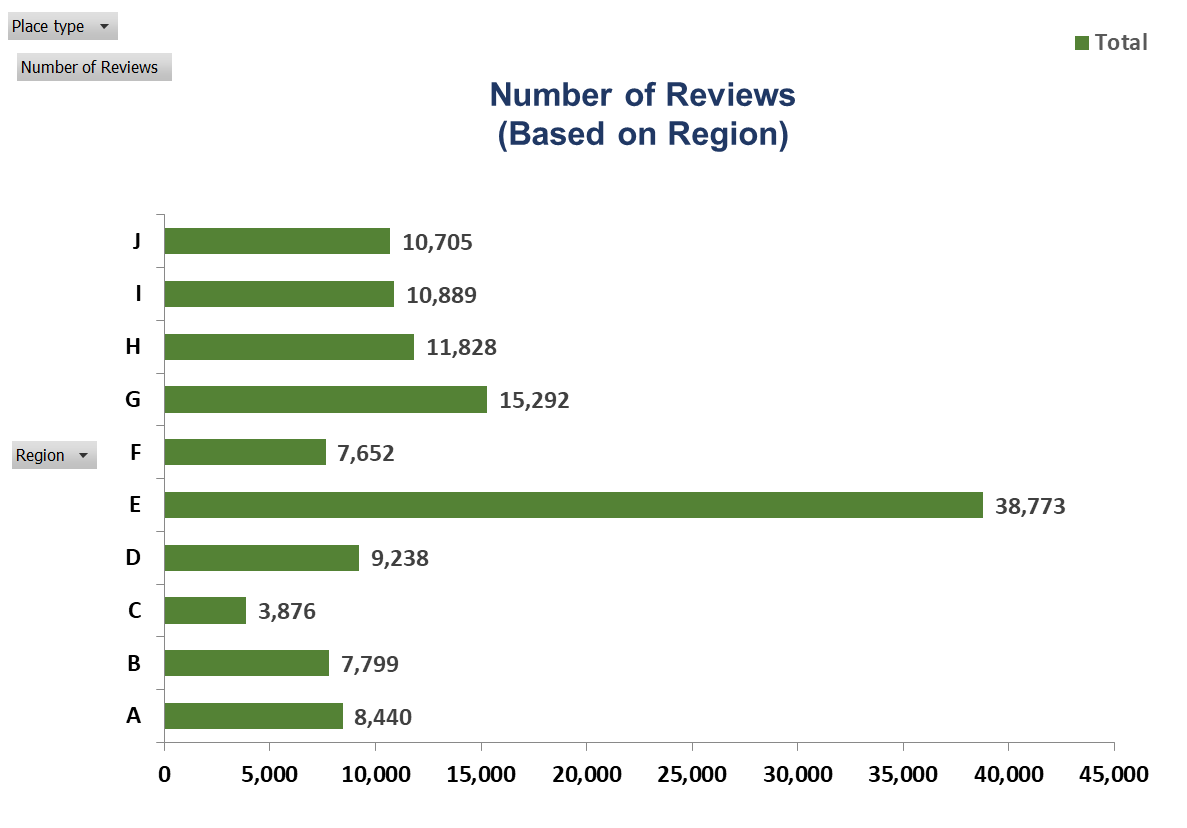
****

***Figure 2: Visualisation showing Number of Shops (Observations)***

**3. Describe the distribution of all of the number of reviews. Your answer must include a**

**visualization that shows the distribution.**

* Same in order to visualise that, I used PivotTable and PivotChart.
* Insert the Pivot Table for dataset in which add “Region” to Rows, “Review” to Values (using Sum function). Change name for columns: Region and Number of Reviews. Also add “Place type” into Filters to see number of reviews for each type of place in each region.
* Insert Pivot chart based on this PivotTable as in picture below.

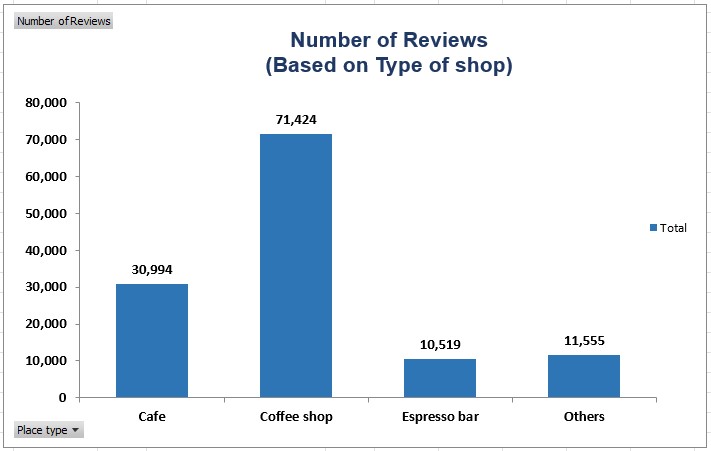
****

***Figure 3: Visualisation showing Number of Reviews based on Region (Distribution)***

4. **Describe the relationship between type of store and number of reviews. Your answer**

**must include a visualization to demonstrate the relationship.**

* Same in order to visualise that, I used PivotTable and PivotChart.
* Insert the Pivot Table for dataset in which add “Place type” to Rows, “Review” to Values (using Sum function). Change name for columns: Type of Store and Number of Reviews.
* Insert Pivot chart based on this PivotTable as in picture below.

******

***Figure 4: Visualisation showing Number of Reviews based on Type of Shops***