**Introduction**

Data Mart is Danny’s latest venture and after running international operations for his online supermarket that specialises in fresh produce - Danny is asking for your support to analyse his sales performance.

In June 2020 - large scale supply changes were made at Data Mart. All Data Mart products now use sustainable packaging methods in every single step from the farm all the way to the customer.

Danny needs your help to quantify the impact of this change on the sales performance for Data Mart and it’s separate business areas.

The key business question he wants you to help him answer are the following:

* What was the quantifiable impact of the changes introduced in June 2020?
* Which platform, region, segment and customer types were the most impacted by this change?
* What can we do about future introduction of similar sustainability updates to the business to minimise impact on sales?

## Case Study Questions

The following case study questions require some data cleaning steps before we start to unpack Danny’s key business questions in more depth.

### **1. Data Cleansing Steps**

In a single query, perform the following operations and generate a new table in the data\_mart schema named clean\_weekly\_sales:

* Convert the week\_date to a DATE format
* Add a week\_number as the second column for each week\_date value, for example any value from the 1st of January to 7th of January will be 1, 8th to 14th will be 2 etc
* Add a month\_number with the calendar month for each week\_date value as the 3rd column
* Add a calendar\_year column as the 4th column containing either 2018, 2019 or 2020 values
* Add a new column called age\_band after the original segment column using the following mapping on the number inside the segment value
* Add a new demographic column using the following mapping for the first letter in the segment values:
* Ensure all null string values with an "unknown" string value in the original segment column as well as the new age\_band and demographic columns
* Generate a new avg\_transaction column as the sales value divided by transactions rounded to 2 decimal places for each record

### **2. Data Exploration**

1. What day of the week is used for each week\_date value?
2. What range of week numbers are missing from the dataset?
3. How many total transactions were there for each year in the dataset?
4. What is the total sales for each region for each month?
5. What is the total count of transactions for each platform
6. What is the percentage of sales for Retail vs Shopify for each month?
7. What is the percentage of sales by demographic for each year in the dataset?
8. Which age\_band and demographic values contribute the most to Retail sales?
9. Can we use the avg\_transaction column to find the average transaction size for each year for Retail vs Shopify? If not - how would you calculate it instead?

### **3. Before & After Analysis**

This technique is usually used when we inspect an important event and want to inspect the impact before and after a certain point in time.

Taking the week\_date value of 2020-06-15 as the baseline week where the Data Mart sustainable packaging changes came into effect.

We would include all week\_date values for 2020-06-15 as the start of the period **after** the change and the previous week\_date values would be **before**

Using this analysis approach - answer the following questions:

1. What is the total sales for the 4 weeks before and after 2020-06-15? What is the growth or reduction rate in actual values and percentage of sales?
2. What about the entire 12 weeks before and after?
3. How do the sale metrics for these 2 periods before and after compare with the previous years in 2018 and 2019?

### **4. Bonus Question**

Which areas of the business have the highest negative impact in sales metrics performance in 2020 for the 12 week before and after period?

* region
* platform
* age\_band
* demographic
* customer\_type

Do you have any further recommendations for Danny’s team at Data Mart or any interesting insights based off this analysis?