TASK 1:

You are part of a company's retail analytics team and have been approached by your client, the Category Manager for Chips, who wants to better understand the types of customers who purchase Chips and their purchasing behaviour within the region.
The insights from your analysis will feed into the supermarket's strategic plan for the chip category in the next half year.

You have received the following email from your manager, Gabby.

'Hi,

Welcome again to the team, we love having new graduates join us!

I just wanted to send a quick follow up from our conversation earlier with a few pointers around the key areas of this task to make sure we set you up for success.

Below I have outlined your main tasks along with what we should be looking for in the data for each.

Examine transaction data – look for inconsistencies, missing data across the data set, outliers, correctly identified category items, numeric data across all tables. If you determine any anomalies make the necessary changes in the dataset and save it. Having clean data will help when it comes to your analysis.

Examine customer data – check for similar issues in the customer data, look for nulls and when you are happy merge the transaction and customer data together so it's ready for the analysis ensuring you save your files along the way.

Data analysis and customer segments – in your analysis make sure you define the metrics – look at total sales, drivers of sales, where the highest sales are coming from etc. Explore the data, create charts and graphs as well as noting any interesting trends and/or insights you find. These will all form part of our report to Julia.

Deep dive into customer segments – define your recommendation from your insights, determine which segments we should be targeting, if packet sizes are relative and form an overall conclusion based on your analysis.

Make sure you save your analysis in the CSV files and your visualisations – we will need them for our report.

Looking forward to reviewing your work.

Thanks,

Gabby.

Here is your task
We need to present a strategic recommendation to Julia that is supported by data which she can then use for the upcoming category review; however to do so we need to analyse the data to understand the current purchasing trends and behaviours. The client is particularly interested in customer segments and their chip purchasing behaviour. Consider what metrics would help describe the customers' purchasing behaviour.

You can choose to complete this task in R or Python. Whilst it's possible to complete the task in Excel you may find the size of the data and the nature of the tasks is such that it is more difficult to complete in Excel.

To get started, download the resource csv data files below and begin performing high level data checks such as:

Creating and interpreting high level summaries of the data
Finding outliers and removing these (if applicable)
Checking data formats and correcting (if applicable)

You will also want to derive extra features such as pack size and brand name from the data and define metrics of interest to enable you to draw insights on who spends on chips and what drives spends for each customer segment. Remember our end goal is to form a strategy based on the findings to provide a clear recommendation to Julia the Category Manager so make sure your insights can have a commercial application.

Download the QVI Transaction Data & QVI Purchase Behaviour datasets.
As we are in the early stages of this analysis Gabby has asked us to submit our initial findings, so please save your code as a pdf or html file, also save your dashboard as a pbix or twb file.

NOTE:

LIFESTAGE: Customer attribute that identifies whether a customer has a family or not and what point in life they are at e.g. are their children in pre-school/primary/secondary school.

PREMIUM_CUSTOMER: Customer segmentation used to differentiate shoppers by the price point of products they buy and the types of products they buy. It is used to identify whether customers may spend more for quality or brand or whether they will purchase the cheapest options.

TASK 2:

Here is the background information on your task

You are part of a company's retail analytics team and have been approached by your client, the Category Manager for Chips, has asked us to test the impact of the new trial layouts with a data driven recommendation to whether or not the trial layout should be rolled out to all their stores.

You have received the following email from Gabby.

'Hi,

Thanks for your feedback earlier, I'm glad you find my follow up emails helpful in ensuring you're on the right track.

For this part of the project we will be examining the performance in trial vs control stores to provide a recommendation for each location based on our insight. Below are some of the areas I want you to focus on, of course if you discover any other interesting insights feel free to include them in your findings.

Select control stores – explore the data and define metrics for your control store selection – think about what would make them a control store. Look at the drivers and make sure you visualise these in a graph to better determine if they are suited. For this piece it may even be worth creating a function to help you.

Assessment of the trial – this one should give you some interesting insights into each of the stores, check each trial store individually in comparison with the control store to get a clear view of its overall performance. We want to know if the trial stores were successful or not.

Collate findings – summarise your findings for each store and provide any recommendation that we can share with Julia outlining the impact on sales during the trial period.

Remember when working with a client visualisations are key to helping them understand the data. Be sure to save all your visualisations so we can use them later in our report. We are presenting to our client.

Keep up the good work!

Thanks,

Gabby

Here is your task:

Julia has asked us to evaluate the performance of a store trial which was performed in stores 77, 86 and 88.

To get started use the QVI_data dataset below or your output from task 1 and consider the monthly sales experience of each store.

This can be broken down by:

total sales revenue
total number of customers
average number of transactions per customer

Create a measure to compare different control stores to each of the trial stores to do this, write a function to reduce having to re-do the analysis for each trial store. Consider using Pearson correlations or a metric such as a magnitude distance e.g. 1- (Observed distance – minimum distance)/(Maximum distance – minimum distance) as a measure.

Once you have selected your control stores, compare each trial and control pair during the trial period. You want to test if total sales are significantly different in the trial period and if so, check if the driver of change is more purchasing customers or more purchases per customer etc.

As we are in the early stages of this analysis Gabby has asked us to submit our initial findings.

TASK 3:

Here is the background information on your task:

Task 3 is targeted specifically at building your ability to recognise commercial, actionable insights from your analysis and displaying it in a clear and concise way for your client, with minimal jargon.

As both technical tasks 1 and 2 were open ended in terms of insights, this task answer will focus on the layout and the order of your inclusions, including where to include graphs, taglines, written insights and recommendations.

As part of a company's retail analytics team, you have been conducting a range of analysis on transaction and purchase behaviour data to provide key recommendations to your client, the Category Manager of chips, who is putting together their strategic plan.

Here is your task:

With our project coming to an end it's time for us to send a report to Julia, based on our analytics from the previous tasks. We want to provide her with insights and recommendations that she can use when developing the strategic plan for the next half year.

For this report, we need to include data visualisations, key callouts, insights as well as recommendations and/or next steps.

We recommend you use a tool like PowerPoint (or similar) to create your report.

One way to create your visualisations (graphs, charts etc) is to save your output files from task 1 & 2 and generate your charts in Power BI or Tableau. You may find it helpful to create chart templates however that is not essential. We are looking for consistency here – if a series is blue in one chart it needs to remain blue throughout the presentation. Once you have created the charts in your Power BI or Tableau file, paste them into your PowerPoint in the order that suits best.

When you have finished creating your report, remember to save it as a PDF to submit.

All in all the submissions required are:
- Your code files in pdf or html format
- Your Tableau or Power BI dashboards
- Your PowerPoint report

There are 3 datasets for you to use for your analysis. The first two for the first task and the final one for the second task.

Best,
Gabby.