You are part of Quantium'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 Zilinka.

Hi,

Thanks for your feedback earlier, I'm glad you find my follow up emails helpful in ensuring your 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 an 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 in 3 weeks so if you could submit your analysis by mid next week that will give us great amount of time to discuss findings and pull together the report.

Keep up the good work!

Thanks,

Zilinka'

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.

We have chosen to complete this task in R, however, you will also find Python to be a useful tool in this piece of analytics. We have also provided an R solution template if you want some assistance in getting through this Task.

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 customers etc.

As we are in the early stages of this analysis Zilinka has asked us to submit our initial findings, so please save your code as a .pdf file and upload it to unlock the example answer.

Pro analytics Tip: While the data set would not normally be considered large some operations may still take some time to run.

This is normal and to be expected when dealing with data sets at Quantium. Whilst your analysis is based on the whole data set you may want to try some strategies to try running sample solutions on a smaller subset of data I.e. create a sample solution using some of the data while it uploads.

We think you've got everything you need for this task but if you're stuck and want a hand, drop us an email at admin@insidesherpa.com mentioning this task.