1

Data preparation and customer analytics

Conduct analysis on your client's transaction dataset and identify customer purchasing behaviours to generate insights and provide commercial recommendations.

Practical skills you will gain from working on this module:

Data validation

Data visualisation

Data wrangling

Programming

Compare your work with real model solutions created by the Quantium team.

Source : https://www.theforage.com/virtual-internships/prototype/NkaC7knWtjSbi6aYv/Data-Analytics

1.

Data preparation and Customer analytics

Conduct analysis on your client's transaction dataset and identify customer purchasing behaviours to generate insights and provide commercial recommendations.

Tasks:

- 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.

Data Provided:

1. QVI_transaction_data

   Potato chips Transaction of 1 year

	A	B	C	D	E	F	G	H
1	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR	PROD_NAME	PROD_QTY	TOT_SALES
2	43390	1	1000	1	5	Natural Chip Compny SeaSalt175g	2	6
3	43599	1	1307	348	66	CCs Nacho Cheese 175g	3	6.3
4	43605	1	1343	383	61	Smiths Crinkle Cut Chips Chicken 170g	2	2.9
5	43329	2	2373	974	69	Smiths Chip Thinly S/Cream&Onion 175g	5	15
6	43330	2	2426	1038	108	Kettle Tortilla ChpsHny&Jlpno Chili 150g	3	13.8
7	43604	4	4074	2982	57	Old El Paso Salsa Dip Tomato Mild 300g	1	5.1
8	43601	4	4149	3333	16	Smiths Crinkle Chips Salt & Vinegar 330g	1	5.7
9	43601	4	4196	3539	24	Grain Waves Sweet Chilli 210g	1	3.6
10	43332	5	5026	4525	42	Doritos Corn Chip Mexican Jalapeno 150g	1	3.9
11	43330	7	7150	6900	52	Grain Waves Sour Cream&Chives 210G	2	7.2
12	43602	7	7215	7176	16	Smiths Crinkle Chips Salt & Vinegar 330g	1	5.7
13	43332	8	8294	8221	114	Kettle Sensations Siracha Lime 150g	5	23
14	43603	9	9208	8634	15	Twisties Cheese 270g	2	9.2
15	43329	13	13213	12447	92	WW Crinkle Cut Chicken 175g	1	1.7
16	43600	19	19272	16686	44	Thins Chips Light& Tangy 175g	1	3.3
17	43604	20	20164	17136	54	CCs Original 175g	1	2.1

2. QVI_purchase_behaviour

   Customer Data for a particular region. This is based on overall behaviour of the customer.

A	B	C
LYLTY_CARD_NBR	LIFESTAGE	PREMIUM_CUSTOMER
1000	YOUNG SINGLES/COUPLES	Premium
1002	YOUNG SINGLES/COUPLES	Mainstream
1003	YOUNG FAMILIES	Budget
1004	OLDER SINGLES/COUPLES	Mainstream
1005	MIDAGE SINGLES/COUPLES	Mainstream
1007	YOUNG SINGLES/COUPLES	Budget
1009	NEW FAMILIES	Premium
1010	YOUNG SINGLES/COUPLES	Mainstream
1011	OLDER SINGLES/COUPLES	Mainstream

Problem Statement provided by Quantium :

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.

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. If you aren't familiar with R or Python we would recommend searching a few online courses to help get you started. We have also provided an R solution template if you want some assistance in getting through this Task. Whilst its 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.

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 model answer.

Note: that this is an open-ended case study that can be approached in many ways. Model answer is in R.

Note : I have done the analysis in Python3 in a Jupyter Notebook. I have provided them in the folder.