Data at Scale: **Coursework**

In this coursework, you will be given a dataset on shopping retails of a company from its retailers in different locations. The dataset was the original copy downloaded from their website and it needs to be cleared up before you use it if necessary. In this coursework, each student will have an individual dataset.

Task:

The company are looking to increase their marketing spend in one of their four locations. In order to select which, management is interested in the regional performance of their stores across the country, as they know regions have different spending behavior patterns, perceptions and interests. They are particularly interested to understand what kind of customer behavior they can expect from their customers in different cities, and will use this information to plan their marketing and distributional tactics accordingly.

To do this they are interested in understanding their customer base, and importantly customer metrics/trends, in all four locations in addition to general performance statistics.

In this coursework, you are required to provide a report comparing relevant customer metrics and their performances across their shops' customers and sales performance. Your final recommendation will be interesting to the company, but importantly they also want to be presented with the information so they can review the information alongside other information they have gathered from different departments. Besides, the company are also attempting to discover particular items that worth more marketing efforts. To achieve so, they want you to provide a cohort analysis on the given item on a monthly basis and also conduct some comparative analysis across popular items' customers and sales performance.

Your task is to prepare a presentation, accompanying set of tables/graphs and short written comparative analysis and executive summary. The presentation will be presented by yourself at an upcoming meeting at a date to be confirmed. The tables/graphs will then be distributed (via the media they were presented, still images, interactive tableau graphs, webpages) along with the written report and executive summary for further consideration by management.

Some additional information:

The original dataset was uploaded to the directory '/home/lab test 1/coursework' in the Msclab server along with the 'columns.txt' which contains the column names for each feature. Each student should use the dataset named by his/her student ID.

DO NOT DELETE ANY TABLE IN THE DATABASE.

Your report and presentation should include at a minimum:

- A cohort analysis on the given item on a monthly basis.
- A comparative analysis of the store performance in terms of sales and profit relative to the size of the store.
- A comparative analysis of popular items. There is no one formula for many of these and therefore you must justify your choice. Do not try and do all the KPIs, rather pick ones so that you can make a coherent and justified recommendation. Feel free to implement things not listed in the lecture and add additional information from external data sources.

Suggested steps:

1) Determine a set of key performance indicators to use to compare the four regions. Consider not only the performance at the final period, but the trends at a weekly or monthly level. Justify your choice of KPIs.
2) Implement these KPIs in SQL. Review the results and come up with a recommendation based on the results.
3) Create a set of graphs/tables and/or dashboard for the managers that compares the four regions performance.
4) Using the graphs/tables/dashboard as a basis, create a presentation to summarize your findings.
5) Create an accompanying report. The format of the report is included at the end of this document.

NOTE: This is likely to be an iterative process. You may use any tools/programs you like.

Submission:

You will be required to submit:

1) Your presentation (you will have 6min + 2min Q&A when presenting)
2) The code / documentation / tableau files so that I can re-create your graphs/tables **from the originally provided data.** i.e. you must include steps (code) you have run to clean the data (if any).
    - Tableau files need to be exported as a packaged workbook. This can be done via the File menu in Tableau: File → Export as Packaged Workbook...
3) A short report (in the format below) including:
    a) the cohort analysis based on the item given to you
    b) the definitions of your KPIs as implemented
    c) a short (< 1 page) executive summary of the recommendation as to which site the company should expand to base on your analysis. Include key arguments in short form
    d) a discussion of the comparative analysis of the three sites based on the KPI's you have implemented. This should cover all the points you mention in your presentation. Be succinct. You may use bullet points. Focus on quality and clarity of the insights. Shorter, clearer work will get higher marks than unnecessarily long text. Maximum of 4 pages with no more than 2,000 words.

Marking:

Presentation: 30%

The presentation and how clear your argument is based on your presented facts. The quality of the graphs/tables (and how they are organized) in quickly providing a comparison and allowing a user to verify and explore the statements you have made in your presentation.

Report (MUST follow the format proved in this document): 70%, broken down as follows.

The quality of your written cohort analysis, comparative analysis, and executive summary. Justification and relevance of the selected KPIs.	40%
Correctness of implementation and data cleaning. Your implementation vs. what you said you measuring will be checked.	20%
The documentation of your work. Could someone else in the class easily repeat your analysis?	10%

Note: You all have individual datasets. Your results will not be comparable.

HINTS:

Your analysis should demonstrate critical thinking and reflection. The University provides this short guide which you may find helpful**.** This is one of many study resources for writing critically, although this list is broad covering many different subjects and styles. More general advice on writing is provided online by Student Services.

Dr. Gavin Smith, the module convenor of UK campus, has written a guide on technical writing specifically for this MSc, focusing on writing your dissertation in Semester 3. While this is not strictly relevant yet, almost all points hold as reports, like your dissertations, are a form of technical writing. Note, however, that an executive summary differs slightly from an abstract. A discussion on how to write a good executive summaries can be found here (however the example is not great/that relevant). A much better set of examples can be found here.

Due Date:

The deadline for the coursework submission is 24th December 2019 by 4.00 pm. An electronic copy should be submitted to the Assignment link at the module page on Moodle. Five marks will be deducted for each day (or part thereof) if coursework is submitted after the official deadline date without an extension having been obtained. Except in exceptional circumstances, late submission penalties will apply automatically unless a claim for extenuating circumstances is made within five calendar days following an assessment deadline. Coursework must be submitted before the presentation which will be held between 27th to 31st December 2019. The exact date will be confirmed via Moodle (announcement and update on the main page).

Report format.
The report MUST comprise of the four sections below. Deviating from this format will result in lost marks.

Section 1: Cohort Analysis

You should finish the cohort analysis based on given item.

Section 2: The KPIs

This section should comprise only of blocks in the format below. An example is provided, followed by an empty block to use as a template.

Example

KPI Description (in words): Active customers per month
KPI formula: count(unique customers), per month, per store
Steps to realize KPI: 1) Create view via SQL query. CREATE VIEW kpi_active_customers AS SELECT COUNT(DISTINCT customer_id) FROM cw.baskets GROUP BY store_code, DATE_TRUNC('month', purchased_at) ORDER BY DATE_TRUNC('month', purchased_at); 2) Visualized via Tableau as graph titled "Active customers per month". See the Tableau file.
Additional Notes: None.
KPI Description (in words):
KPI formula:
Steps to realize KPI:
Additional Notes:

Section 3: Executive Summary

1 page max.

Section 4: Comparative Analysis

4 pages with 2,000 words max.