

Group Assignment: Business Analytics Case Study

In groups of 5, you are required to develop a Business Analytics (BA) system for E-Zone. You are expected to design a high-level BA framework with a detailed dimensional model design. Additionally, you will need to implement components of your BA framework, generate insights from data and suggest ways for managerial action.

Case Description

Please see: E-Zone Consumer Goods Case Study

Assessment

This assignment is worth 30% of your assessment for the subject.

Case Study – Report Requirements

The submission should be about 3000 words in length and take the form of a report describing and justifying the design of the star schema and predictive models. **The length of 3000 words does not include appendices.** A suggested outline for the sections of the report is:

1. **Executive Summary:** An executive summary is a short summary of the key information in the document and is written for a business person. It should contain a clear recommendation together with evidence to help a business person to decide whether to go ahead with the BA project.
2. **Business Analytics Framework:** You should present a discussion of business analytics specific to E-Zone, why it is necessary and the benefits they can bring to this organisation.
3. **Design of a Dimensional Model:** Your group is required to design a star schema for the case study. Include a diagram of your star schema¹ as well as justifications for the design choices that you have made. The star schema must be documented with a diagram and a dictionary describing each of the attributes within the fact and dimension tables. Within your data dictionary you should include details about sourcing of the data in the star schema from data files in the operational systems. The data dictionary needs to contain enough detail for your dimensional model design so that the model can be clearly understood.
4. **Implementation of the Dimensional Model:** Your group is required to implement your dimensional model. This involves two steps: (1) data integration and (2) data visualization.
 - 4.1. data integration:
 - Use the Pentaho Kettle Data Integration tool to create your fact and dimension tables.
 - Add a screenshot of your Kettle transformations to the text and briefly explain your transformations and steps.
 - Attach and submit your Kettle transformations along with your report.
 - 4.2. data visualization.
 - Use Qlik Sense to build dashboard(s) to communicate key information to managers.
 - The dashboard(s) should help managers with the **key decision areas highlighted in case description.**
 - Add a screenshot of your dashboard(s) to the text
 - Attach and submit your Qlik Sense app along with your report.

¹ You can use **Data model viewer on Qlik Sense** for this. Other options are using ERDPlus (<https://erdplus.com/>) online or drawing your design using MS Word.

5. **Development of a predictive model:** Your group needs to build an analytical model that predicts overall profits per month for each region
 - You need to use RapidMiner to build your predictive model
 - Add a screenshot of your RapidMiner process to the text
 - Attach and submit your RapidMiner file along with your report.
6. **Generate Business Insights and Strategise:** Your group is required to do analysis of data. You need to use the digital dashboards you have built to generate business insights for six problem areas that the managers are facing. In addition to generating business insights, you need to propose action that managers can take to grow their company.
7. **Appendix 1 – Data Dictionary:** Include a data dictionary for each of your fact and dimensions tables. Include an entry for each data item including its name, a brief description of its meaning, and describe how it will be sourced from the existing databases and include a description of any necessary transformations.
8. **Appendix 2 – Kettle transformations, Qlik visualisations and RapidMiner Process:** You also need to zip your Kettle, Qlik and RapidMiner files and submit along with your assignment. **Submissions without Kettle, Qlik, and RapidMiner attachments will not be marked.** Late submissions of these files will incur penalties as discussed in subject eCP.
9. **Appendix 3 – SQL (optional):** If you prefer to use SQL for implementing your dimensional model, include all of the SQL statements that you used including Create statements, Key creation and any indexing you deem necessary.
10. **Appendix 4 – Work Breakdown:** – Detail the breakdown of work of the team members for this assignment. This should be a detailed account of what each team member accomplished as part of the assignment. **Submissions without work breakdown will not be marked.**

Assessment Criteria

Aspect of review	Comprehensive	Proficient	Basic	Below Basic	Poor
Executive Summary 5%					
Business Analytics Discussion 10%					
Dimensional Model Design 20%					
Dimensional Model Implementation 25%					
Building Predictive Model 10%					
Generating business insights and strategising 25%					
Data Dictionary 5%					
Kettle transformations, Qlik visualizations and RapidMiner Process (HURDLE)					
SQL (optional)					
Work Breakdown (HURDLE)					

Submission

This assignment is due at **14:00 on 21 October 2019.**

For late submissions, please refer to guidelines outlined on eCP:

http://www.courses.uq.edu.au/student_section_loader.php?section=5&profileId=93817

You are required to submit an electronic version of your work through the LMS (See the Assessment Link on the LMS and **look for the link labelled Assignment Submission**. You will have two opportunities to submit, in case you make a mistake on your first submission or wish to revise it. Please ensure that you put your report through Turnitin BEFORE submitting it to Blackboard, to make sure that your assignment is plagiarism free. By submitting your work in this manner plagiarism is checked for you and you will be able to see if you have any issues with this in your work. Use Turnitin as many times as you need to fix any similarity issues before the due date. Also, by submitting the Assignment via the LMS you are agreeing that the Assignment is your own work.

Please compress your Kettle, Qlik and RapidMiner files as well as your Data Dictionary in a zip file. Submit the Zip file along with your report on both Blackboard assignment and Turnitin.

Academic Honesty

Academic misconduct occurs when students portray someone else's work as their own. There are many ways in which Academic misconduct can occur. Some of these are:

- **Sham Paraphrasing:** Material copied verbatim from text and source acknowledged in-line but represented as paraphrased.
- **Illicit Paraphrasing:** Material paraphrased from text without in-line acknowledgement of source.
- **Other Plagiarism:** Material copied from another student's assignment with the knowledge of the other student.
- **Verbatim Copying:** Material copied verbatim from text without in-line acknowledgement of the source.
- **Recycling:** Same assignment submitted more than once for different subjects.
- **Ghost Writing:** Assignment written by third party and represented as own work.
- **Purloining:** Assignment copied from another student's assignment or other person's papers without that person's knowledge.

The university is committed to graduating students with "a profound respect for truth, and for the ethics of scholarship... we want our graduates to be capable of independent thought, to be able to do their own work, and to know how to acknowledge the work of others" (Prof. Peter McPhee). As such, the university takes a dim view of students who are not able to correctly acknowledge the work of others, or who try to pass this work off as their own. All students should be aware of the following web site: <https://my.uq.edu.au/information-and-services/manage-my-program/student-integrity-and-conduct/academic-integrity-and-student-conduct> which provides practical advice to students about how not to be involved in academic misconduct. In particular look at the Quick Checklist section (paraphrased below)

Quick checklist

To be certain to acknowledge sources fairly and avoid plagiarising, review this checklist before beginning to write your essay and again after you have completed your first draft².

1. What type of source are you using: your own independent material, common knowledge, or someone else's independent material?
2. If you are quoting someone else's material, is the quotation exact? Have you used quotation marks for quotations embedded in your text? Have you set off block quotes with an extra space before and after the quote, single spacing within the quote, and left indenting of all lines of the block quote? Are omissions shown with ellipses and additions with square brackets?

² Reproduced with permission of Dr Stephen Morgan, Faculty of Economics and Commerce, UniMelb

3. If you are paraphrasing someone else's material, have you rewritten it in your own words and sentence structures? Does your paraphrase employ quotation marks when you resort to the author's exact language? Have you represented the author's meaning without distortion?
4. Have you acknowledged each use of someone else's material?
5. Do all references contain complete and accurate information on the sources you have cited?

E-Zone Consumer Goods Case Study

E-Zone Consumer Goods (ECG) is one of the largest online retailers in Australia that sell consumer goods locally, as well as internationally in Canada and USA. Their product range includes household, personal health, beauty, children items, and other products. E-Zone essentially connects small businesses from all over Australia to potential customers in a seamless way. Those merchants are able to sell their products through E-Zone and ship them directly to the customers using ECG logistics partners.

After a customer purchases a product from E-Zone the seller gets notified to fulfil that order. Once the customer receives the product, or the estimated delivery date is due, the customer gets a satisfaction survey by email where he/she can give a note for the purchase experience and write down some comments.

ECG has been very successful in recent years and is a highly profitable business. The manager would like to expand the business further. However, in order to make these decisions he needs a better understanding of his sales, products, customers, suppliers, and marketing position. He has heard from vendors that business analytics will provide him with the capability to make more informed decisions. You have been hired as a consultant to propose a business analytics solution to the management team.

The Insights Needed for Decision-Making

Decisions about expanding the business are made by the CEO, with input from general managers of various functions. Although all managers have a wealth of experience and a strong “gut feel” about the business, the CEO wants evidence based on data to inform his decision-making. In particular, he needs insights about delivered orders, including product, customer, sales territory, time and vendor information, and lead performance. Therewith, the CEO wants to know expected regional sales performance.

1. *Who are the key customers?*

You should identify ECG’s key customers. Your BA system should provide information about unit sales, dollar sales and margin for customers for various time periods (including year, month and quarter) and about customer segments (teenage - under 20 years old, adult - under 65 years old, senior - 65 years old and over, 1st Jan 2019 is a cut-off date for age calculation). It should also provide information about the locations in which the most profitable customers live.

2. *Which products are the most profitable?*

You should identify ECG’s most profitable product categories and subcategories. Moreover, your BA system should provide information about unit sales, dollar sales and margin for products for various time periods (including year, season and month).

3. *Which sales territories are the most profitable?*

You should identify ECG’s most profitable *sales territories (states)* and salespersons who are underperforming. Moreover, your BA system should provide information about unit sales, dollar sales and margin for each sales territory for various time periods (including year, month and quarter) along with the assigned salesperson.

4. *Which time periods are the most profitable?*

You should identify ECG’s most profitable times. Moreover, your BA system should provide information about the unit sales, dollar sales and margin for various time periods including Monthly, Quarterly, Yearly and Sundays.

5. *Review score per product category*

You should identify product categories with the lowest review scores and analyse if delivery times affect product review scores.

6. *Lead performance*

You should analyse lead types, lead profile behaviours, and the type of media where the lead was acquired to assess the effectiveness of marketing channels and performance of the closed deals. Your BA system should provide information about the number of deals closed, percentage of the deals closed and the number of days it took to close the deal for sellers for various time periods (including year, month and quarter).

7. *(Expected) Regional Performance*

In the next step, the CEO wants you to predict the profits in your domestic market Australia. To that end, you select the monthly aggregated data sample (`regional_performance_sample_data.csv` in `Sample Data.zip` file) and use the sales area (0 = Other Countries, 1 = Australia), the local seasons as well as the profit of the previous month (`profit_lag1`) to predict future profits. Other countries and Spring serve as reference categories. Make sure to account for missing values by setting them equal to zero. Based on the obtained regression results, briefly explain the effects of seasons and sales areas on profits. Please also use this regression to predict profits you would expect in December in Australia if you made a profit of 120,000 in November.

The Current Information Systems

The company uses two separate OLTP databases for marketing and selling items: the marketing system and the sales system.

The marketing system includes information about the sellers that filled-in requests of contact to sell their products on ECG. After a lead fills in a form, a filter is made to select the ones that are qualified to sell their products at ECG. They are called Marketing Qualified Leads (MQLs). After a qualified lead fills in a form he/she is contacted by a Sales Development Representative. At this step some information is checked and more information about the lead is gathered.

The sales system includes information about orders made by customers.

Data for the BA system will be sourced from two of the existing operational systems, the marketing system and the sales system.

1. The Sales System

The sales system is a package based on a SQL Server database running on a windows-based computer. Examples of data in relevant SQL Server tables are shown in the zipped files on BlackBoard assignment folder Called Sample Data. The ER model for the sales system is shown in Figure 1 below.

Please note:

In general, a customer order might have multiple items and each item might be fulfilled by a distinct seller.

“Order_Items_Dataset” table includes data about the items purchased within each order. Example:

The order_id = 00143d0f86d6fbd9f9b38ab440ac16f5 has 3 items (same product). Each item has the freight calculated accordingly to its measures and weight. To get the total freight value for each order you just have to sum. The total order_item value is: $21.33 * 3 = 63.99$ The total freight value is: $15.10 * 3 = 45.30$. Therewith, the freight value is to be paid by the sellers, that’s why it reduces the overall profit.

“Order_Reviews_Dataset” table includes data about reviews specific to different orders. And the customer can update the review for the same order multiple times. You will need to use the latest updated version of the reviews for each order in your analysis.

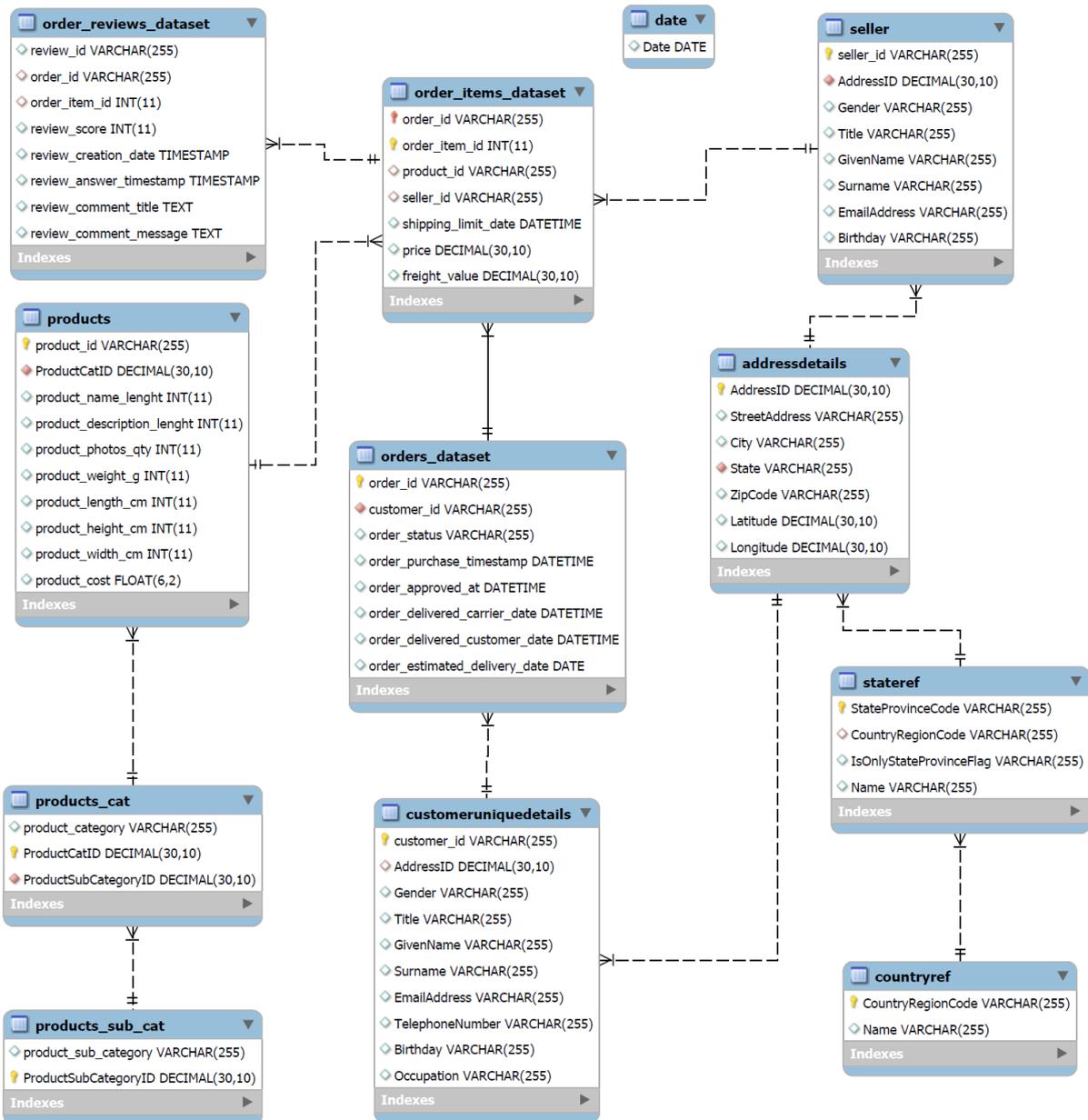


Figure 1: Sales ER Model (Definition of attributes are provided in the data dictionary in Assignment folder on LMS)

2. The Marketing System

The marketing system is also package based on SQL Server and runs of a windows-based computer. Examples of data in relevant tables are also shown below on BlackBoard assignment folder. Definition of attributes are provided in the data dictionary. The ER model for the marketing system is shown in Figure 2 below.

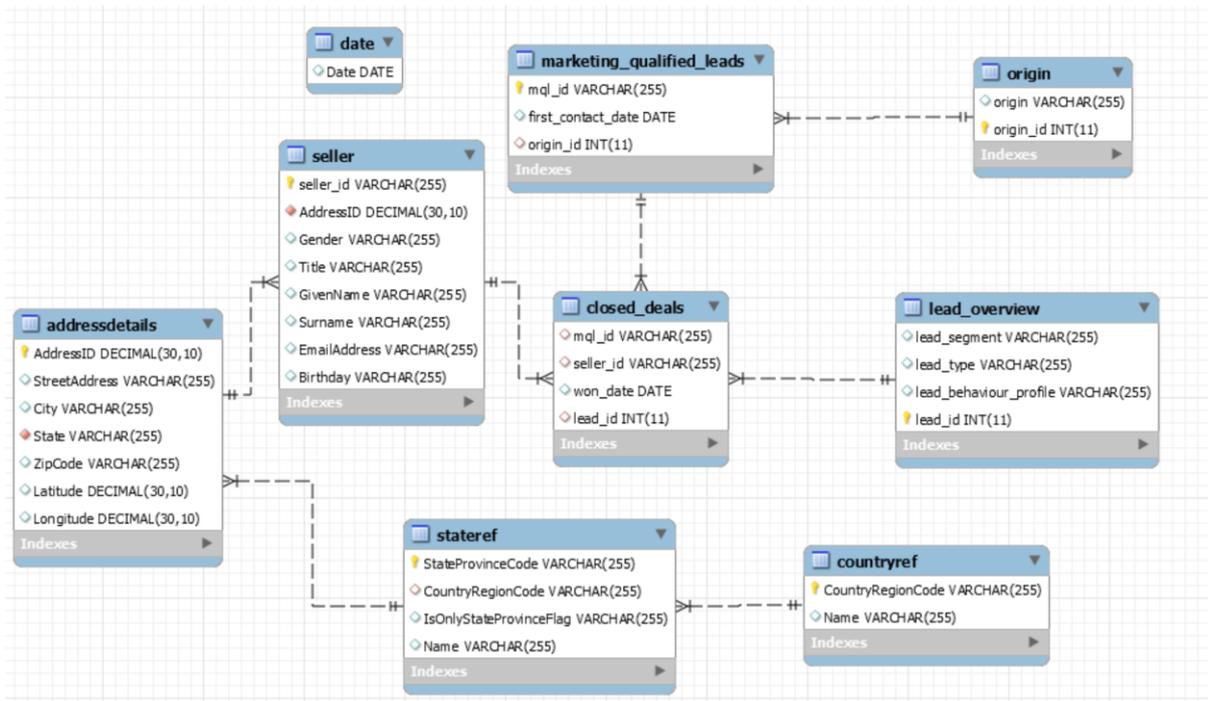


Figure 2: Marketing ER Model (Definition of attributes are provided in the data dictionary in Assignment folder on LMS).

Connecting to ECG OLTP databases

Use the following information to connect to ECG OLTP database and source the data:

server: ezone.database.windows.net
 port: 1433
 username: database_user@ezone
 password: dbh@\$p@\$Wrd
 database: bism

Good luck with your assignment.