EBUS3030 Assignment 2

Stavros Karmaniolos c3160280@uon.edu.au
Jay Rovacsek c3146220@uon.edu.au
Jacob Litherland c3263482@uon.edu.au
Edward Lonsdale c3252144@uon.edu.au

October 24, 2018

Contents

1 Assignment Overview & Requirements

Business Intelligence - EBUS3030
Assignment 2

Assignment Outcomes

This assignment requires multiple outputs to be created to exhibit your understanding of business intelligence/data analysis through an example 'real world' question that is comparable to what you may be asked of you as you become an IT professional.

Key outcomes to be delivered are: Data Modelling/ETL to get the data in a usable format, Output of your analysis, Report summarising your findings and a presentation to the class of your work. The presentation is expected to concentrate more on your findings/recommendations as if it were a situation where you are presenting the response to the CEO.

Assignment Question

The CEO of 'BIA Inc' had been speaking to the Sales Executive and had heard about some recent work you had completed and thought you might be able to assist them with a problem they have.

"I've heard that you helped the Sales Exec recently with understanding more about our Newcastle site. I would now like your help to get a handle on the whole business. As you are aware, the Newcastle office is just 1 of 10 sites we have across the country. Unfortunately, sales in some items have dropped across the country in recent years and we are currently running at a loss.

We need to consider consolidating our company offices. We need to reduce costs for the longevity of the company as a whole. I need you to get some numbers together around the performance of our 10 offices, so that I can factor this information into any decision regarding which office (or offices) we might consider closing.

I would like a summary of recent numbers and some trend analysis as well please. It would be great if you could also project sales for the next 12 months for each office as well. It would be helpful if you could indicate the 3 most popular and 3 least popular items in each of our stores, as well as the worst performing items for the company as a whole.

As you can imagine, this is a very sensitive topic so, as part of your response I want you to provide the justification as to which office we may close. Our decision will upset some people and I want to make sure we have all the background information on hand. If you can provide a Ranking of offices based on your analysis that would be wonderful.

…. I believe you started to bring together a data store of this information from the Newcastle Office, can you expand that and load all of the sales information for all offices and complete your analysis.
"

Assignment Deliverables

Using the data file provided in Excel and notes about the data (*Assignment 2 – data.xlsx*), you are required to complete the following elements as part of the assignment.

- Data Model/Data Load Process
  - Provide an overview of the data model & ETL process completed to get the data ready for analysis
  - Ensure you record any assumptions you have made as part of this component and your reasoning behind the assumption.
- Analysis including any predictive work undertaken
  - Provide the SQL and raw output of your base analysis
  - Provide workings of the predictive work you completed for the trending & prediction on future sales.
  - Ensure you record any assumptions you have made as part of this component and your reasoning behind the assumption (this includes answers to any relevant questions put to management (your lecturer) during workshops.

- Executive Summary & Presentation in response to business question.
    - Provide an Executive report and Dashboard
        - Your Executive report should include a short Executive brief/summary that presents a clear concise response back to the CEO question about possible downsizing of operations including evidence/justification.
        - A dashboard should allow quick comparisons between the sites to be undertaken as well as contain at least one element of 'predictive' analysis
        - Present the material as if it is to be consumed in a formal boardroom meeting
            - All members of the team need to participate in a (15 – 20 minute) presentation to be given as part of the lab in Week 11.
            - Please be aware that any of the company Executive may ask questions as you present your findings.

NB: As part of your response, you should also specifically include any assumptions/external information you have made/used throughout the process as well as any quality processes/checking you have completed or limitations you discovered.

Breakup of assignment Marks (total course mark for assignment = Assignment Part B submission (28% + Presentation Two (7%) = 35%.

Assignment Component	Percentage Allocation
Data Model/ETL	10%
Core Analysis	50%
Executive Summary/Evidence	25%
Dashboard	15%
	100%

Key Documents Required & Format

You are required to upload all files in a single zip file (including any presentation items for the team delivery within the lab) via blackboard to the Assignment Two drop folder by 12 noon, Thursday 25th October. You will also be required to submit a paper copy of your report at the beginning of the presentation workshop (make sure this is printed well before the workshop and has a group Assessment Cover sheet signed by ALL team members).

NB: Only 1 load per team only but it should contain all of the deliverable items.

Your data model should include a printout of an ER diagram using the notation described in lectures. It should also include a printout of your SQL schema showing Primary and foreign keys, as well as all attributes.

Your presentation is worth 7% of the course mark. It should simulate presenting the report to management. You should time your presentation to be between 12-15 minutes with 5-10 minutes for questions. You presentation should include a demonstration of your dashboard, results and recommendations from your analysis. Your presentation marks will contain components for organisation, comprehension of presented results, and timing.

2 Executive Summary

2.1 Datamart Business Rules

The following business rules were provided to be used in the context of this assignment:

- At BIA all customers interacts are in an online environment.We only support electronic orders.
- Returning Customers can provide POI information via the web interface and look up their record and that will flow with the sale
- The sales associate can complete the order form/sale for the client.
- Each sale will have a receipt number/id.
- A receipt can have many line items
- Each line item can only be for a single item, but the customer can purchase multiples of the same item.
- After consultation with your team, we have made the following change to discount applied to sales: Where a customer has multiple line items, any sale with 5 or more row items (containing at least 5 different items) is provided a 5% discount.
- The system automatically handles the total for the sale by looking up the item, then multiplying the costs per item by number purchased, and then should store this final field total as a record in the system (but should also be able to see clearly sales that were provided a discount.
- Store Item prices can change at any point, however the price the customer pays is the amount listed for the store item that is sold on the sale date. We need to keep a record of all store item prices historically so that we can determine what the store item price was at any particular past date.
- Only 1 BIA sales assistant can be attributed to any receipt
- Customers may visit multiple stores for purchases (ie they are not locked to a particular store). As a result, all customer records are replicated across all stores, so they do not need to be re-recorded at a store by store level.

With these considerations in mind, the following report was created to outline the discovery, creation and polish to satisfy the assignment requirements.

3 Data Model

The following section outlines the models used in the design and creation of the database. It includes the EER Model with relations, attributes and relationships as well as the database schema from Microsoft SQL Server Management Studio.

3.0.1 Database Schema

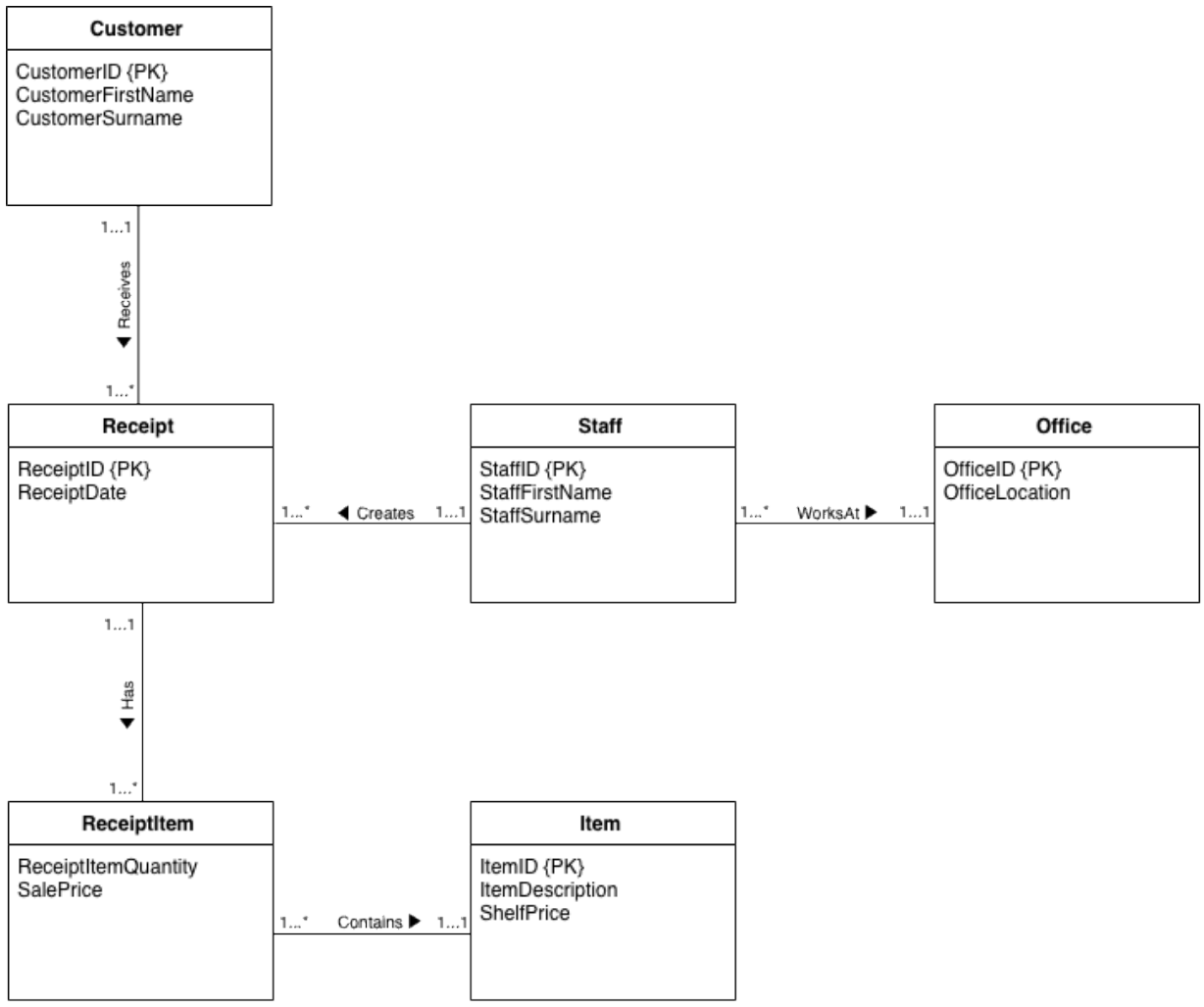
The below data model is only a suggestion and is still subject to change into the future. A full create script can be found in the appendix

It must be noted that the structure of this data model is less than efficient, and it would be expected in a datamart situation that only at lower levels of data would this schema remain responsive in the manner it is now, as the outline suggests the datamart is not necessarily the most suitable design for future use, however suits very well currently.

It would be expected that only at extremely large data sets would this model prove a bad design. In such cases a model more representative of the snowflake or star schema would be heavily advised.

3.0.2 EER Diagram

An EER diagram of the suggested data model is provided below.



4 Data Load Process (ETL/ELT)

Initial import of the data supplied in the xlsx file generated a very basic table that allowed us to analyze the data for potential outliers, confirm the business requirements of the data and then create tables from which the data model was derived.
The Imported table structure was as follows:

Assignment2Data

Column Name	Data Type	Allow Nulls
Sale_Date	datetime2(7)	☐
Reciept_Id	int	☐
Customer_ID	nvarchar(50)	☐
Customer_First_Name	nvarchar(50)	☐
Customer_Surname	nvarchar(50)	☐
Staff_ID	nvarchar(50)	☐
Staff_First_Name	nvarchar(50)	☐
Staff_Surname	nvarchar(50)	☐
Staff_office	int	☐
Office_Location	nvarchar(50)	☐
Reciept_Transaction_Row_ID	int	☐
Item_ID	int	☐
Item_Description	nvarchar(50)	☐
Item_Quantity	int	☐
Item_Price	float	☐
Row_Total	float	☐
		☐

A decision to leave this initial import table as default was made to allow easy reference to the initially supplied excel data file.

4.1 Quality Assurance Processes

Maybe include some C# code references or whatnot.

4.2 Assumptions and Reasoning

4.2.1 Item

4.2.2 ReceiptItem

4.2.3 Receipt

4.2.4 Staff

4.2.5 Customer

4.2.6 Office

5 Base Analysis

5.1 Notes on Analysis

5.2 Raw Results

5.2.1 Total Number of Sales

5.2.2 Total Sales Value

5.2.3 Total Sales Per Store

5.2.4 Total Items Sold

5.2.5 Discounted Sales Ratio

5.2.6 Total Sales Value per Staff Member

5.2.7 Average Sale Value Overall

5.2.8 Average Sale Per Store

5.3 Total Number of Customers

5.4 Top Team-member(s) Analysis

Analysis over all stores, correlation to store we want to nuke?

5.5 Customer Analysis

Analysis per store - top 3?

5.5.1 Customer Frequency

Can we predict future trends in customers?

5.6 Items Per Sale

5.7 Item Popularity

Top 3 best and worst items overall, correlation to any stores?

5.8 Worst Performing Item

Correlation to store?

6 Conclusion and Recommendations

References

[1] Reasons against TSQL Money type: Stackoverflow User; *SQLMenace* https://stackoverflow.com/questions/582797/should-you-choose-the-money-or-decimalx-y-datatypes-in-sql-server

[2] Microsoft TSQL documentation of Decimal/Numeric types https://docs.microsoft.com/en-us/sql/t-sql/data-types/decimal-and-numeric-transact-sql?view=sql-server-2017

[3] Microsoft documentation: WITH common_table_expression (Transact-SQL) https://docs.microsoft.com/en-us/sql/t-sql/queries/with-common-table-expression-transact-sql?view=sql-server-2017

[4] Upselling - Business Dictionary http://www.businessdictionary.com/definition/upselling.html

7 Appendix