Clinical data warehousing for evidence based decision making

Narra, Lekha; Sahama, Tony; Stapleton, Peta

Published in:
Digital Healthcare Empowering Europeans - Proceedings of MIE 2015

DOI:
10.3233/978-1-61499-512-8-329

Digital Healthcare Empowering Europeans
R. Cornet et al. (Eds.)

doi:10.3233/978-1-61499-512-8-329

Clinical Data Warehousing for Evidence Based Decision Making

Lekha Narra[a,1], Tony Sahama[b] and Peta Stapleton[b]
[a] *School of Electrical Engineering and Computer Science, Science and Engineering Faculty, Queensland University of Technology (QUT), Brisbane, Queensland 4000, Australia*
[b] *School of Psychology, Faculty of Society & Design, Bond University Gold Coast, Queensland 4229, Australia*

Abstract. Large volumes of heterogeneous health data silos pose a big challenge when exploring for information to allow for evidence based decision making and ensuring quality outcomes. In this paper, we present a proof of concept for adopting data warehousing technology to aggregate and analyse disparate health data in order to understand the impact various lifestyle factors on obesity. We present a practical model for data warehousing with detailed explanation which can be adopted similarly for studying various other health issues.

Keywords: Data warehousing, Obesity, Microsoft SQL Server, Health data, lifestyle

Introduction

Obesity has been a major issue of concern all over the world in recent decades due to its increasing prevalence at an alarming rate as well as its association with chronic and life threatening conditions [1]. In Australia, the prevalence of obesity has increased from 56.3% in 1995 to 61.2% in 2007-08 according to Australian Bureau of Statistics (ABS) 2013 [2]. Generally the rise in sedentary jobs and unhealthy food habits were considered major contributing factors for increasing obesity rates [3]. Hence in paper [4], we proposed and experimented with data warehousing approach for building a platform to analyse obesity in relation to nutrition, physical activity, age and gender. In this paper, we provide a proof of concept for the data warehousing approach proposed in [4] by generating reports to address the following questions.

1. What is the percentage of people in each weight category (underweight, normal weight, overweight, obese) for each type of milk consumed?
2. Examine the level to which the number of days exercised in a week is related to obesity?
3. Examine the trend of how the level of exercise undertaken is related to obesity?

[1] Lekha Narra, lekha.kolluri@connect.qut.edu.au

The above questions were selected in order to estimate how weight is affected by the type of milk consumed, number of days exercised and the level of exercise performed.

1 Methods

Australian population data related to obesity published by the ABS at [5] and [6] was used for the current case study. The source files at [5] and [6] provided summary data in Microsoft Excel spreadsheets where [5] presented weight of Australians with respect to various categories of their nutrition, physical activity and gender and [6] presented weight of Australians with respect to age. The various steps involved in building the data warehouse and generating reports are explained below.

1.1 Analysis of source data

As the data obtained from [5] and [6] was in summary form, raw data was synthesised based on this summary data as access to the original copy of the raw data was restricted. This aggregation was performed preserving the nature of source data and such that the averages match those present in the source files to the extent possible. This aggregated data was stored in Microsoft Excel spreadsheets.

1.2 Designing the data warehouse

This step involved choosing the architecture for the intended data warehouse as well as the choice of data model. Centralised data warehouse architecture was chosen after studying the suitability of various data warehouse architectures proposed by Ponniah, 2010 [7] with the current scenario. This architecture provides a consistent, integrated and flexible source of data where data is retrieved directly from the centralised data warehouse [8]. This choice was made because of the absence of multiple subject areas since the entire study was focused on weight factor in this case. The data warehouse was modelled using Dimensional Modelling and star schema that comprise of a fact table containing measurable attributes and attributes that link to multiple dimension tables around it. Hence each dimension has equal chance to be utilised in a query [7].

1.3 Building the data warehouse

This step started with the Extraction, Transform and Loading (ETL) process which involved extracting the data from source files (aggregated files in this case), transforming this data to match with the target system (i.e. data warehouse) and loading this transformed data in to the data warehouse. The ETL process was performed using SQL Server Integration Services (SSIS).

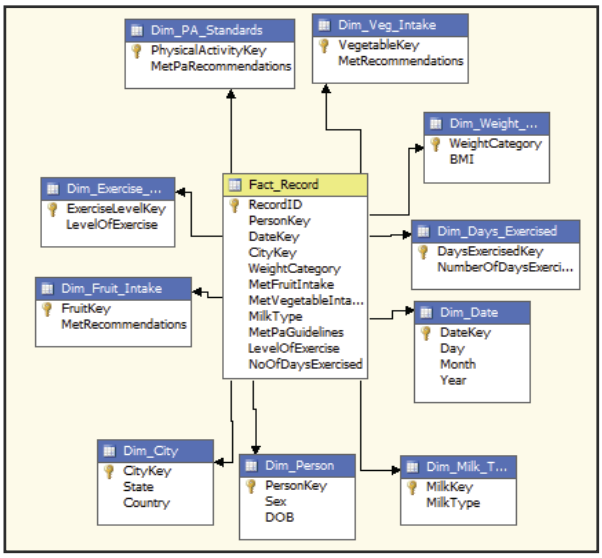

Figure 1: Star schema diagram for Obesity data warehouse

The dimensional model of the data warehouse is shown in Figure 1. Each dimension is linked to the fact table through the foreign key relationships. This enabled a person or a group of people's weight to be analysed with respect to their nutrition intake or physical activity performed or a combination of both.

1.4 Building OLAP Cubes

Cubes reduce the processing time by providing flexible access to summarised data by using pre-computed data [9].A single cube was built using all the dimensions in the data warehouse using count of records as measure [4].

1.5 Generating Reports

This step involved generating reports using data from the cube built in Step 1.4 to answer the questions mentioned in the Introduction section. SQL Server Reporting Services (SSRS) was used for generating reports in this case. The reports were laid out in a suitable form (i.e. matrix, line chart, column chart in this case) to clearly visualise and understand the relationship of weight with type of milk consumed and amount of exercise performed in the Australian population.

2 Results

Using the reports, the weight categories were analysed with respect to type of milk consumed, number of days exercised in a week and the level of exercise performed which were presented in Fig.2, Fig.3 and Fig.4 respectively.

Milk Type	NormalWeight	Obese	OverWeight	UnderWeight
Does not drink milk	5%	5%	5%	11%
Reduced fat	32%	35%	35%	14%
Skim	15%	18%	17%	6%
Whole milk/Regular/Full cream	47%	41%	42%	65%
Don't know	1%	1%	1%	4%
Total	100	100	100	100

Figure 2. Matrix style Report in SSRS presenting the percentage of people in each weight category consuming each type of milk

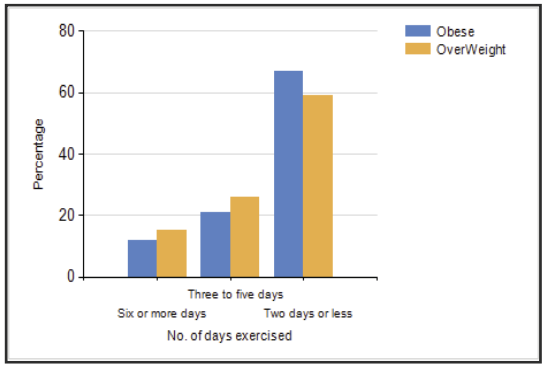

Figure 3. Column chart style Report in SSRS for analysis of overweight and obese weight categories in relation to number of days exercised in a week

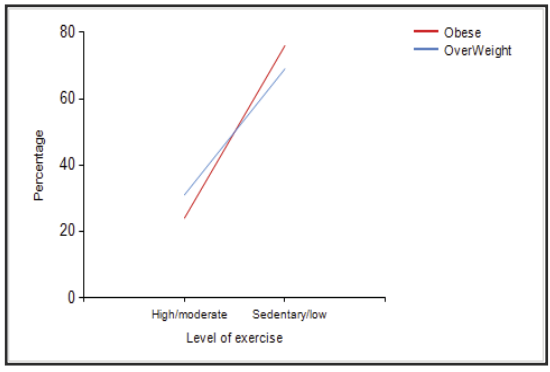

Figure 4. Line chart style report in SSRS for analysis of overweight and obese weight categories in relation to Level of exercise performed

It can be observed from the above figures that the type of milk consumed has no noticeable effect on the person's weight whereas there is significant increase in the overweight/obese category when the number of days exercised/ week reduced from six or more days to two days or less and when level of exercise moved from high/moderate to sedentary/low. These results match closely with the summary data published by the

ABS 2011 [5] and [6]. Minor inaccuracies were observed that resulted due to the rounding up of values while aggregating the raw data that was used to populate the obesity data warehouse. These inaccuracies may be considered negligible because original source data will be used for building a data warehouse in practical scenarios instead of aggregated data.

By applying a data warehousing approach to study various factors affecting obesity, data from disparate sources was successfully integrated and also enabled multi-dimensional analysis.

3 Discussion

On the whole, in the current case, data warehousing proved to be a promising technology for the study of obesity related data to effectively understand the impact of various lifestyle factors on obesity. Also the non-volatile nature of data warehouses enables the study of various health issues over a long period of time. This feature allows for study of other lifestyle diseases like cancer, which is only possible with detailed health data accumulated over a long period of time. In addition, analysis of health data can also be performed to identify unknown associations between various health factors by applying data mining techniques. Similarly this approach can be used for studying any health issue, not only to understand the influencing factors but also to study the effectiveness of various medical practices.

References

[1] Gill, T. P. (1997): Key issues in the prevention of obesity. BrMed Bull 53(2):359:388.

[2] Australian Bureau of Statistics (2013): Profiles of Health, Australia, 2011-13, cat. no. 4338.0, Canberra.

[3] Cameron, A.J., Welborn, T.A., Zimmet, P.Z., Dunstan, D.W., Owen, N., Salmon, J., Shaw, J.E. (2003): Overweight and obesity in Australia: the 1999-2000 Australian diabetes, obesity and lifestyle study (AusDiab). Medical Journal of Australia 178(9): 427-432.

[4] Narra, L., Sahama, T., & Stapleton, P.B. (accepted 2014). Clinical data warehousing: A business analytics approach for managing health data. Eighth Australasian Workshop on Health Informatics and Knowledge Management (HIKM, 2015), Sydney, Australia. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 164 - A. Anthony Maeder and Janes Warren, Eds.

[5] Australian Bureau of Statistics 2011, Overweight and Obesity in Adults in Australia: A Snapshot, 2007-08, Selected nutrition and physical activity characteristics by measured Body Mass Index(a), data cube: Excel spreadsheet, cat. no. 4842.0.55.001, Canberra.

[6] Australian Bureau of Statistics 2011, Overweight and Obesity in Adults in Australia: A Snapshot, 2007-08, Measured Body Mass Index by demographic and socio-economic characteristics, data cube: Excel spreadsheet, cat. no. 4842.0.55.001, Canberra.

[7] Ponniah, P. (2010). From Data Warehouse: The Building Blocks. In Data warehousing fundamentals for IT professionals. 32-34. Ponniah, P.(eds). 2nd ed.. A John Wiley & Sons, Inc., publications.

[8] Moody, D.L., Kortink, M.AR. (2000): From enterprise models to dimensional models: a methodology for data warehouse and data mart design. DMDW, p. 5.

[9] Han, J., Kamber, M. (2011). From Data Cube Technology. In Data Mining : Concepts and Techniques. 187. Han, J., Kamber, M. (eds). 3rd ed. n.p. Morgan Kaufmann Publishers Inc.