Faculty of Computing and Information Technology/Department of Computer Science and Mathematics

BACS3013 Data Science

Tutorial 2 (Introduction to Data Science and Big Data Analytics - cont)

Q1. Figure 1.1 gives a high-level overview of the analytics process model.

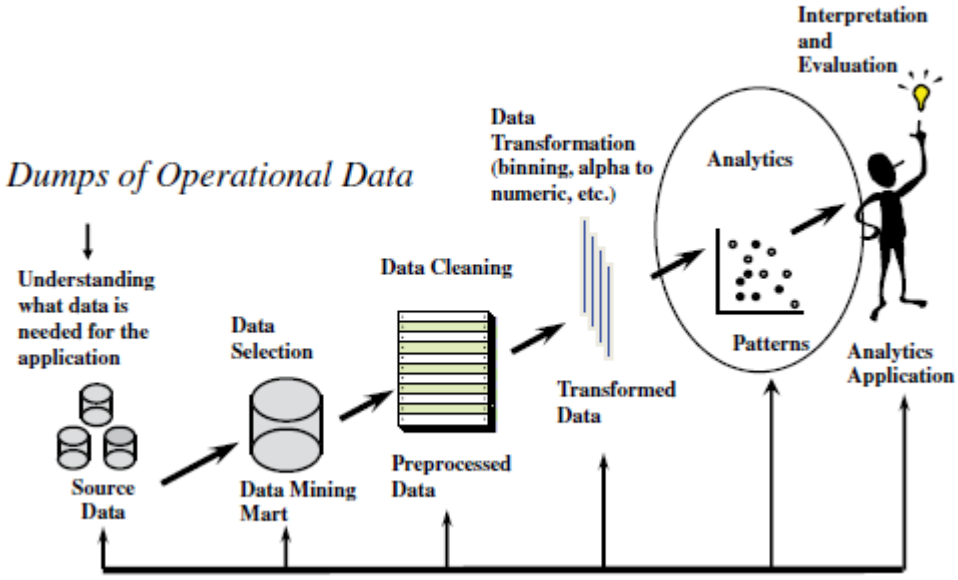

Figure 1.1 The Analytics Process Model

By referring to a real-life application as an example, describe how data scientist uses this model to design an analytics application. (13 marks)

A data scientist is helping a University to analyze the level of student satisfaction for various programmes offered in that University. (1m)

At the source data (data collection) stage, he decides to collect the comments from students (through online questionnaire survey and feedback form) as the input data. (2m)

At the data selection stage, he gathers all forms and store in a (cloud) data warehouse. (2m)

At the data cleaning stage, he removes all inconsistencies, such as missing values, outliers, by using a mathematical model. (2m)

At the data transforming stage, he transforms all questionnaire scores into different numerical coding and aggregation, such as mean, maximum number, frequency etc. (2m)

At the data analytics stage, he clusters all data inputs into few distinct factors, by using clustering methods. (2m)

At the interpretation and evaluation stage, he interprets the results and concludes the findings. For instance, it is found that the top three distinct factors are tuition fee, quality of teaching, and facility. He may even show the mathematical relationship or equation for these three factors and satisfaction score. (2m)

Another example:

A data analyst is requested to retrieve the individual voices from audio recordings recorded by customer service officer. (1m)

At the source data (data collection) stage, he decides to collect several audio recordings. (2m)

At the data selection stage, he gathers all audio recordings and store in a (cloud) data warehouse. (2m)

At the data cleaning stage, he removes all inconsistencies, such as background noises and empty voices by using some filtering methods. (2m)

At the data transforming stage, he transforms all conversation recordings into digital signals/structured data. (2m)

At the data analytics stage, he recovers the respective individual voices (especially customers) using some blind source separation or dimensional reduction methods. (2m)

At the interpretation and evaluation stage, he evaluates the results and validate the quality of recovered individual voices. (2m)

(Note, at this stage, the manager might ask him to further analyze customer's preference and sentiments. He may then use the analytics process model (Figure 1.1) again to do another analysis, ie: sentiment analysis from speech.)

Q2. List out **two (2)** analytics applications that may be useful for the following sectors:

(a) Government (2 marks)

Tax avoidance, social security fraud (2m)

(b) Risk Management (2 marks)

Credit Risk Management, Market Risk Management (2m)

(c) Web (2 marks)

Social media analytics, web analytics (2m)

(d) Marketing (2 marks)

Customer Segmentation, response modelling (2m)

(e) Logistics (2 marks)

Demand Forecasting, Supply chain analytics(2m)

BACS3013 Data Science

Q3. List out **five (5)** software tools used in data analytics. (5 marks)

R, Python, SQL, Matlab, Excel, Tableau, Hadoop, SPSS, SAS, etc (5m)

Q4. List out **two (2)** analytics applications that can be integrated as part of Smart Campus project. For each application, state the suitable software tool and briefly describe how it works. (10 marks)

Application: Analysis on Students' feedback using text mining method (1m)

Software tool: R, Matlab, SAS, Tableau etc (1m)

Description: Generally, by end of semester, students will give their comments or feedback about the quality of lecturers, facilities, etc. A data scientist may collect all these comments and store as .csv files. He may then analyze the common expressions stated in the feedback by visualizing a word cloud using some text mining methods. (3m)

Application: Car park analysis and reporting (1m)

Software tool: Tableau, Excel, R, Matlab, SAS, etc – for analysis and reporting (1m)
or any complete car park management solutions like ParkingEye etc

Description: A data scientist may analyze the use of car park(s) in the campus and then give a comprehensive view of staffs/students/visitors' activities. By capturing the movements of all vehicles visiting the site, a huge amount of insight may be derived, such as:

- **Repeat visit analysis**
- **Heap maps of popular time and dates**
- **Capacity and utilization**
- **Stay duration analysis**
- **Etc**

(3m)

Note: It will help the University management in making a good car park management, such as number of car park slots reserved for staff and/or students, time and duration for opening non-main gates, placement of security guards etc.