# *Assignment 5 – Data Management*

Date assigned: Thursday, April 7, 2016

Date assignment due: **Friday, April 22, 2016, 11:50pm**

**Learning Objectives**

Upon successful completion of this assignment, the student will be able to:

* Use simple formulas and functions in Excel
* Create charts in Excel
* Describe data mining and how it is used
* Analyze a case study from a business intelligence perspective

To do:

**Part A – Using Excel**

**Create an Excel document called YourUserId\_E01\_A05\_Budget.xls, and save it in the Assignment folder in the 420-E01 folder in your H drive.**

1. Create a personal budget in Excel.

|  |  |  |  |
| --- | --- | --- | --- |
| Assume that you are a student living on your own in an apartment. You own a car and you are paying your own bills. Use the information below to create a budget for your first year of financial independence:   * Rent is $800 a month * Utilities are $200 a month from January through May and October through December. From June through September, utilities are $100 a month. * You spend $4800 a year on food. You can divide this into 12 equal payments. * You spend $1200 a year on entertainment. * School tuition is due January 1 and August 1-each tuition payment is $500 per semester. * Your car is paid off * You pay $1900 a year for insurance that is divided into twelve monthly payments. * You spend $30 per week on gas * From Sept – April, you have a part time job at Tim Horton’s and your take home pay is $200 per week, after taxes have been deducted * From May – August, you have a full time co-op job and your take home pays is $400 per week * From May – August, you work one shift a week at Tim Horton’s and your take home pay is $50 * You receive a $3000 scholarship, which is paid to you in two equal installments in Sept. and January. | |  | | --- | |  | |  | |
| 1. Create an Excel worksheet that shows your expenses by month. Title the worksheet "Expenses". Type your name next to the title. 2. Use the months (January through December) as your column headings and rent, utilities, food, entertainment, tuition, gas and insurance as row headings. Bold each column and row heading. 3. Total each row and total each column. Add up all of your expenses. **Use the SUM function**. 4. Create a 3-D column chart as a new sheet that shows monthly expenses for: rent, utilities, food, gas, insurance, tuition and entertainment by month. Use "Monthly Expenses" as the chart title in 14-point font. Include your name in the title. Make sure that your axis have appropriate titles, and remove the legend. 5. Create a pie chart as a new sheet that shows total expenses for category. HINT: select only the data that you need by using the CTRL key and the mouse click. Title the pie graph "Overall Expenses" and include your name in the title. Make sure that the legend contains the expense categories. 6. Create a new sheet that shows your income by month. Title the worksheet “Income”. Type your name next to the title. 7. Use the months (January through December) as your column headings and Tim Horton’s, Co-op job and scholarship as row headings. Bold each column and row heading. 8. Total each row and total each column. Add up all of your income. **Use a formula.** 9. Create a new sheet that shows your summary financial information. Title the worksheet “Summary”, and make it the first sheet. 10. Create a row called Total Income. Use a formula and absolute addresses to extract the total income from the Income worksheet. 11. Create a row called Total Expenses. Use a formula and absolute addresses to extract the total expenses from the Expenses worksheet. 12. Create a row called Net Income which uses the formula Total Income – Total Expenses to calculate the net income. 13. Create a row called In Debt, which determines whether you are in debt. If the Net Income is positive, then display No. If the Net Income is negative, display Yes. Use the IF function to determine this. 14. Use a formula to calculate the current date, and display that at the top of your Summary sheet. 15. Ensure that all cells used have a border around it and that the row and column headings are bolded. 16. Add a fill colour to all column headings. |  |

**Part B – Data Management and Data Warehousing**

**Save this document as YourUserId\_E01\_A05, and save it in the Assignment folder in the 420-E01 folder in your H drive.**

1. Determine whether each of the following would call for an OLAP or an OLTP database. Explain your reasoning.
   1. An automobile marketer wants to improve business activity; therefore, he wants to view data from different perspectives in order to decide where to focus his efforts.

OLAP

* 1. The "Best Foot Forward" company wants to build a database to track the progress of its shoe sales, by style and by month.

OLAP

* 1. Wikipedia wants to allow users to update information online. Changes made by on line users are marked for review before being publicly released.

OLTP

* 1. A large college with declining enrolment wants to determine where the majority of students come from to allow them to perform targeted marketing to those locations.

OLAP

* 1. eBay needs to process information on its bidders quickly to allow “last-second” bids.

OLTP

1. Use the Internet to research the following questions:
   1. Explain the concept of “Big Data”, and why data analytics is such a growing field of employment.
   * Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying and information privacy.
   * People want to keep track of everything that goes on so that they can market things towards people and be very specific with things. Big data is used in research in science to find statistics in hundreds of thousands of runs through simulations, running social science experiments, etc.
   1. What technical skills are needed to obtain a data analytics job position?
   * You need to have database experience and experience in dealing with analytics or statistics. A lot of the skills needed are taught in social science in methods. Learning about statistical sciences and knowing what kind of data is relevant and knowing how to pick what data needs to be analyzed.
   1. What are the three biggest Business Intelligence software applications used in the industry for data analytics?
   * Oracle
   * IBM SPSS analytics
   * Microsoft Azure
2. Read the case study titled “The Case for Business Intelligence at NetFlix” provided to you in your handout. Answer the following questions:
   1. What is the impact to NetFlix if the information contained in its database is of low quality?
   * You end up having customer’s that don’t feel as satisfied with your service or you end up recommending people movies that aren’t along the lines of something your customers would like to watch and they end up thinking it’s a poor service.

* 1. How might NetFlix resolve issues of poor information in their customer movie reviews?
  + By filtering customer reviews of the best reviews to the worst ones so that people see all the good stuff and really need to have to dig down to find the bad stuff.
  1. Why is database technology so important to NetFlix and its business model?
  + Their business model is highly based on customer satisfaction and appealing to the customer, so they need to know a lot more about each individual customer to keep them satisfied.
  1. Why must NetFlix cleanse or scrub the information in its database?
  + Because people’s interests vary and multiple people use the same account with more or less frequency and a lot of bits of information can be outlyers, so some of it might not be relevant to the rest of the data. It needs to be kept up to date. If you’ve been recommending a movie to someone for 6 months and they still haven’t watched it, maybe you should stop recommending it and give them something else to watch.
  1. Choose one of the three common forms of data mining analysis (cluster analysis, association detection, statistical analysis) and explain how NetFlix is using it to gain Business Intelligence.
  + Statistical analysis can be used to find out how often a user will watch recommended content and then try and find out when in the day and what days of the week they’re more likely to watch it along with what genres of movies or TV shows they’re more likely to watch so that you can recommend things that they’re most likely to watch at the time they’re most likely to watch it. They also find out how many people have watched things similar to you and try to figure out what the odds are of you watching something they recommend and pick the one with the highest likelihood.

1. In your own words, explain the following data warehousing concepts:
   1. What are the similarities and differences between a data warehouse and a data mart?
   * Data warehouses are meant to store data for extensive periods of time. Once information goes in, it isn’t frequently updated or cleaned up. Data warehouses keep track of all information for all parts of an organization.
   * Data marts are used to keep track of small data in shorter term. Data marts are used to keep information for one part of an organization. Data marts are much cheaper, often faster, but much smaller.
   1. What problem are data warehouses trying to solve?
   * Storing massive amounts of information is incredibly difficult, so data warehouses are there to try and organize all the information from everywhere within an organization into one place.
   1. Find an example online of a company that uses data warehouses, and explain how they are used.
   * Walmart uses data warehouses to keep track of everything going on inside their stores at all times. They know who bought what when, they know what the most popular shelves are in the store, they know the order that people tend to walk through their stores, everything. This is so that they can optimize their advertisement and ultimately their sales. If suppliers know that their shelf currently isn’t full, then they can start manufacturing more product to fit that shelf space.

**Marking Scheme**

|  |  |
| --- | --- |
|  | Marks |
| Part A Question 1 – Budget | 45 |
| Part B Question 1 – OLAP vs OLTP | 10 |
| Part B Question 2 – Data Analytics | 10 |
| Part B Question 3 – NetFlix case study | 10 |
| Part B Question 4 - Data warehouses | 10 |
| Organization | 5 |
| Total | 90 |

**To submit**

The following file should be uploaded to Moodle:

* YourUserName\_E01\_A05\_Budget.xls
* YourUserName\_E01\_A05 with the answers to Part B