Semester Group Consulting Project Title:
Design and Implementation of a Data Warehouse
for a Retail Store with Store-level Data

**Consulting Report-1: Requirements Gathering to Create Business Questions
And Domain Understanding**

Absolutely no discussions between project groups about the entire project.

Overall Objective of this Project

The objective of this project is to design and develop a data warehouse for a retail store chain called Dominick's Fine Food (DFF). DFF has many branches around Chicago area. The data for this project includes store-level data covering shelf management and pricing from all the branches of DFF. According to the website, "this is historical data. … From 1989 to 1994, Chicago Booth and Dominick's Finer Foods entered into a partnership for store-level research into shelf management and pricing. Randomized experiments were conducted in more than 25 different categories throughout all stores in this 100-store chain. (A brief history of the store is in https://www.company-histories.com/Dominicks-Finer-Foods-Inc-Company-History.html). As a by-product of this research cooperation, approximately nine years of store-level data on the sales of more than 3,500 UPCs is available in this dataset. This data is unique for the breadth of its coverage and for the information available on retail margins." The data in SAS V6 or V7 format can be accessed from the following web site: https://www.chicagobooth.edu/research/kilts/datasets/dominicks

However, I have downloaded the data for your convenience, and you need to access CSV data files from Canvas link. The data is stored in my Google Drive in zipped form.   The link will be provided in the Portal.

Data Types

The data is stored in a number of files in folder. Please unzip it to see the files. The data set is about 5 GB and is dirty. There are several data sets in this folder. One can divide them into two categories (look for the codebook and manual in the portal and also in the research.chicagobooth.edu web site) –

    (A) General Files:
- *The Customer Count file*: Ccount (The Customer Count file contains information on store traffic and coupon usage, by store.)
- *Store-level Demographics*: Demographics (The demographic file contains detailed information based on the U.S. Government census.)

    (B) Category Specific Files

There are files for each category studied in the course of the project. Each category is referenced by a three-letter acronym.

- *UPC files*: upcxxx (where xxx stands for the category acronym) The UPC files contain one record for each UPC in a category. They contain information about product name, size, commodity code, etc. The files are sorted by UPC.
- *Movement files*: wxxx (where xxx stands for the category acronym) The movement files contain weekly sales data for each UPC in each store for over 5 years. The variables included in these files comprise: price, unit sold, profit margin, deal code, etc. The files are sorted by UPC, store, week.

Activities for Consulting Report-1
This report will concentrate on your understanding of the project and will focus on the data and development of the <u>business questions</u> for the data warehouse. The activities involved are as follows:

1. As some data are in CSV format and others are in text and HTML format (see the website and the manual), you need to be careful how to download the data. Look at the metadata of all these OLTP source files in the manual.

2. For this report, selectively download **some** of the data in the Lab or your laptop and study them. <u>Do not put them in SQL Server right now</u>. Import the sample data to <u>text</u> files or to Excel. <u>Understand the data by using Excel, Excel Charts, and Pivot tables and charts (if needed)</u>. You might want to download JMP from http://www.jmp.com/ for 30-day use. This is a very important step.

3. Do domain research FOR marketing and sales using the above web site to see <u>what possible questions are useful</u>. We will discuss what they are in Activity 1. Most of the questions <u>should focus on marketing aspects</u> of a retail store. The URL is once again https://www.chicagobooth.edu/research/kilts/datasets/dominicks . You can also do online search to understand the issues with this kind of retail data. CITE EVERYTHING THAT YOU ARE GOING TO USE IN YOUR REPORT.

4. Based on your analysis of the sample data (#2), your research (#3), and class slides develop a set of <u>10</u> business questions (also called OLAP questions) that are useful for DFF and implementable for the project. (Hint: Use Excel tables and charts to show if the questions are implementable or not. Write the reasons why these business questions (BQs) are important to DFF. These BQs are OLAP questions. <u>Make use of Pivot tables (if needed) as much as possible.</u>

Submission of the Report
A report in a report format must be submitted that shows all the above activities have done. Make sure that your report follows EXACTLY the structure given below. All activities refer to the activities listed earlier:

Section 1. **Introduction**
- Talk about DFF and the problems for it that you are dealing with in the group project

Section 2. **Details about the Data**
- Understanding of the data (covers Activity #2) ... (25%)
- Describe metadata for all the OLTP source files (covers Activity#1)…(0%)
- Draw ERDs (use Visio) to show the relationships among the data (Activity #1).… (10%)

Section 3. **Subject Area Understanding**
- Read some of the papers (you don't have to read all. 3-5 good papers would suffice) from the web site URL. Also research online on retail domain. List out some of the problems that you from your reading that are relevant for this case. Also create reference list and attach it at the end of the paper. Cite them in the text as (Sen, 1993. "the time of the paper," vol# etc., year) (Activity#3) … (25%)

Section 4. **Business Questions**
- Based on your understanding of the data, metadata and papers, list out <u>10</u> business questions (Activity#4).
- Give support from the data to each question listed above to show that they are indeed useful for DFF. Use Excel, its statistics tools, Pivot tables and charts (Pivot or otherwise).
- <u>Prioritize the business questions</u>. Rationalize your prioritizations, that is, why are these priorities are important for DFF to have? … (40%)

Grading

- Presentation of the report   20% - 10
- Quality of the content   80% - 40
    1. Describe the metadata for all the OLTP source files
    2. ERDs   (10% - 4)
    3. Understanding of data (data analysis using Excel/Charts/Pivot tables) (25%- 10)
    4. Subject Area understanding (research etc.)   (25% - 10)
    5. List out 10 business questions and prioritize them. <u>Rationalize your prioritizations</u>. (40% - 16)