Assignment 2

**Due**: September 16, 2016 (Submit through Canvas by 11:59pm)

**Type**: Individual (OK to ask others for simple clarification question)

**Grade**: 8%

**WARNING**: There are two parts to this assignment – Part A and Part B. The assignment is not something that be done in one sitting or the day before the deadline. You need to really step back and ask what problem you are trying to address and how to break the bigger problem into smaller questions to solve using SQL. Given the nature of questions, there will be subjectivity and variance in the answers.

Please submit SQL statement (well formatted) with results in a Word document. Whenever there are more than 10 records in the result, please copy and paste the first 10 records.

# PART A – Coffee Sales

The following questions are based on the Coffee store sales data. Answer the following queries and you must use SQL to extract data and not eyeball some data to answer the questions.

1. Just for starters - SQL questions:
   1. In each state, find the area codes with sales more than 10% the average sales of all area codes within that state for the year 2013.
   2. Find the products with profit margins as percentage of sales (profits/sales) of at least 15%. Display the results in descending order of total actual sales. Round the percentage to two digits using ROUND(….,2) function.
   3. Find AreaIDs where the total profits from leaves in 2012 are two times greater than that from beans.
2. DECLINING PROFITS:
   1. Which are the top 5 area codes with declining profits and how much did the profits decline for these 5 area codes?
   2. Among the five profit-declining area codes, are the profits consistently declining for all products? If not, identify the products for which they had significantly higher profit decline.
3. BUDGETED Numbers:
   1. All the budgeted numbers are expected targets for 2012 and 2013. Identify the top 5 states for the year 2012 that have substantially higher actual numbers relative to budgeted numbers for profits and sales.
   2. Identify area codes within these 5 states that beat budgeted sales and profits significantly (You need to define what significant means here).
4. PRODUCT related:
   1. In each market, which products have the greatest increase in profits?
   2. In each market, which **product types** have greatest increase in sales?
   3. Have all products within the product types show similar behavior, or some products within a product type have greatest increase in sales?
5. MARKETING EXPENSES (LOWEST):
   1. Which top 5 states have the lowest market expenses as a percentage of their sales?
   2. Do the above 5 states also have the highest profits as a percentage of sales?
   3. Are there any particular product(s) within these markets with the least marketing expenses?
6. MARKETING EXPENSES (highest):
   1. Which 5 states have the highest marketing expenses as a percentage of sales?

Are these marketing expenses justified? (Note: you need to think how you will justify high marketing expenses)?

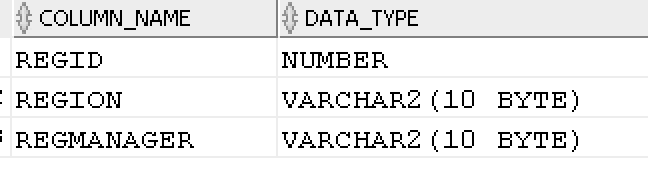
* 1. In each of these 5 states, do any area codes spend too much on marketing expenses relative to others?

1. STRATEGY:
   1. You are in a high-level strategy meeting to discuss how to improve performance. This may involve shutting down stores in losing area codes and/or expanding in very profitable/high growth area. Evaluate the data and recommend which stores to close and where?
   2. Where should the firm focus on expanding?

# PART B: Office Product

The data files are available on Canvas. Here are the tables you need to create; Note: PK is primary key and FK is the foreign key.

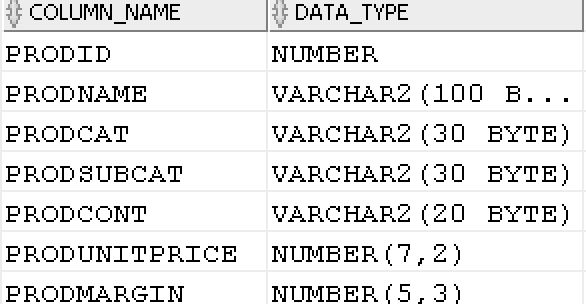
**TABLE: MANAGERS (REGID is the PK)**



CONSTRAINT:

REGION can be only ‘East’, ‘South’, ‘Central’, ‘West’.

**TABLE: PRODUCTS (ProdID is the PK)**

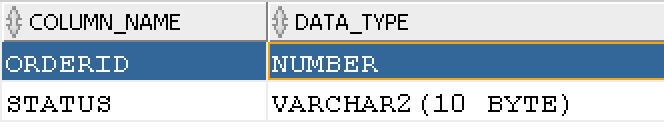


CONSTRAINTS:

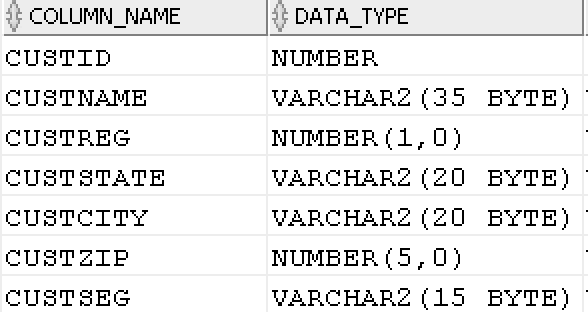
PRODCAT can only be ‘Technology’ ‘Furniture’ or ‘Office Supplies’

PRODCONT take on only ‘Jumbo Drum’, ‘Medium Box’, ‘Jumbo Box’, ‘Wrap Bag’, ‘Large Box’, ‘Small Box’, ‘Small Pack’

**TABLE: ORDERS (OrderID is the PK)**



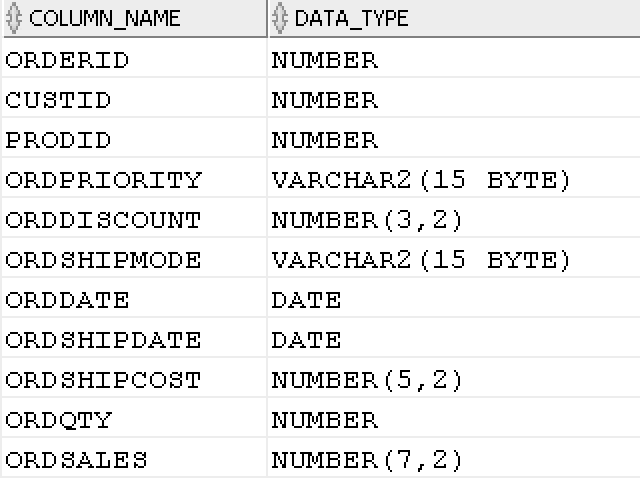
**TABLE: CUSTOMERS (CustID is the PK; CustReg is the FK on delete cascade)**



CONSTRAINT:

CUSTSEG can be only Home Office ‘Corporate’, ‘Small Business’, ‘Consumer’.

**TABLE: ORDERDET (OrderID (FK), CustID (FK), ProdID (FK) are together a PK; All FK are on delete restrict)**



CONSTRAINTS

ORDPRIORITY can be ‘Low’, ‘Medium’, ‘High’, ‘Critical’, ‘Not Specified’

ORDSHIPMODE can be ‘Regular Air’, ‘Delivery Truck’, ‘Express Air’

**TASKS:**

Do the following and copy into Word document the DDL, DML, results, and any errors. Like in Part A, please copy and paste the first 10 rows if there are more than 10 rows in the answer.

QUESTION 1: Create the 5 tables given above. You should define primary keys, foreign keys, and other CHECK constraints. And, load the data from Excel spreadsheet.

QUESTION 2: ORDER Cancellations

1. What fraction of the orders was cancelled?
2. What were the sales from cancelled orders?
3. Who are the top five customers in terms of cancelled orders?

QUESTION 3: CUSTOMER related:

1. Who are the top 10 customers in terms of revenues generated?
2. Are there customers who buy mostly some categories of products and there is a potential for them to buy other product categories?

QUESTION 4: There are differences in the actual (theoretical) price ((unit price \* number of units\*(1-discount) + shipping cost) and the actual sales for all products. There are some discounts and shipping costs. Yet, there are discrepancies in the theoretical sales and actual sales.

1. How much more or less are the actual sales value compared to the theoretical sales value?
2. Are certain managers generally pricing more or less than theoretical sales? Analyze the differences based on the regions/managers.

QUESTION 5: these are product related questions:

1. Products have numbers within its name. Identify the product names with digits in their name. (hint: use REGEXP\_LIKE)
2. Which are the top 5 selling products during the year 2011?
3. Which are the top 10 products with greatest total profit margin? (i.e., sales\*margin).
4. Identify the worst five products in terms of sales?