

SCM 651 Fall 2017 Group Assignment 4

Due: Midnight, Tuesday, November 14, 2017, Total Points = 45

This assignment is based on the data set SCM651 F17 data group assignment 4.xls

This data set has six worksheets from three data sources.

World Bank Data: The first four worksheets are prepared from the World Bank data base:

Country: Fields are country name and three letter country code. (You may note that there is some duplication this list includes all countries individually as well as the names of some groups of countries. For the assignment, ignore this fact and use the list as given.)

GDP: Fields are country code, and per capita GDP in 2014.

Life Expectancy: Fields are country code, and life expectancy at birth in 2014.

Internet: Fields are country code, and percentage of population with internet in 2014.

Smartphone Survey Data: The fifth worksheet is from survey data on smartphone usage collected by my students in Spring 2017. The fields are ID, gender, age, sources from which respondent gets news, brands of smartphone ever owned by respondent, and brand of smartphone used by respondent.

Orange Juice Data: The sixth worksheet has 135 cases drawn from the Dominicks data base and includes seven fields, ID, Store, Week, Brand, Move, Price and Feat.

Please import the six worksheets into Access and do the following in Access. Each task will give you a table. Export the tables in Excel. Submit the answer as an Excel file where each worksheet is a table of answers.

1(30 points) To answer question 1, please use the first four worksheets (from the World Bank data base).

1(a).(5 pt) Join the tables country and income to create a new table that has country name, country code, and GDP. The new table should have all country names. (For some countries, GDP will be missing in the new table.) Create a new field called income that has three categories: low (GDP is low), medium (GDP is medium), and high (GDP is high). Use your judgment to set the boundaries. For example, you can use conditional formatting in Excel to find the boundaries for highest 20% and lowest 20% of GDP. The resulting table is the answer to question 1.

Note:

- When importing the worksheet **country**, please mark that the first row has labels.
- Use switch or IIF to create the new field **income**. The handout on Access describes how to use Switch and IIF (see page 17 of the handout).
- If GDP is missing for a country, leave the cell for income blank. For example, you can use the following format:

Switch([GDP] between value1 and value2,"Low",[GDP] between value 3 and value 4,"Medium",[GDP] between value 5 and value6,"High",true,"")

The last "" creates a blank entry if GDP does not fall in any of the three ranges specified.

- Use country code to join the tables in Database Tools → relationships. In Join type, mark all cases from the table **country** are included.

1(b).(5 pt) Join the table that includes country name, country code, GDP and income (that is, the answer to question 1(a)) with the other two tables (life expectancy and internet) from the World Bank data base. Create a new table that has all country names and all the fields (country name, country code, GDP, income, life expectancy and internet). The resulting table is the answer to question 1(b). (There will be cases where one or more fields will be missing.)

Note:

- You have to join three tables here: the table which is the answer to 1(a), life expectancy, and internet. In join type, mark all cases from the answer to 1(a) as included.
- Sometimes we get an error message if we try to join multiple tables at the same time. If that happens, proceed sequentially: Table from 1(a) + life expectancy = Step 1, Step 1 + Internet = Step 2.

1(c).(5 pt) From the table in question 1(b), create a table for which there is no missing data for any of the three variables GDP, life expectancy and internet. Include all fields in this table. This table is the answer to question 1(c).

1(d).(5 pt) Use the table from question 1(c) to find the average of GDP, life expectancy, and internet for the three categories of income (low, medium and high). The table of results is the answer to question 1(d).

1(e).(5 pt) From the table in question 1(b), create a table for which GDP is not missing, life expectancy is missing, and internet is missing. Include all fields in the table. This table is the answer to question 1(e).

1(f).(5 pt) From the table in question 1(b), create a table for which internet is not missing, GDP is missing, and life expectancy is missing. This is the answer to question 1(f).

2.(10 points) Use the worksheet Smartphone Survey Data to answer this question.

Using the Switch or IIF function, create the following five 1/0 dummy variables:

- **Newspaper** (1 if News Source includes newspaper, 0 if not)
- **Radio** (1 if New Source includes radio, 0 if not)
- **Apple** (1 if cellphone brand owned includes Apple, 0 if not)
- **Samsung** (1 if cellphone brand owned includes Samsung, 0 if not)

Then, using the Switch or IIF function create a variable called **Cell** that is 1 if cellphone brand used is Apple, 2 if LG, 3 if Samsung, 4 if other, and blank if missing).

The answer to question 2 is the worksheet smartphone survey with the five new columns added. Please also paste the Switch or IIF commands used to prepare the answers.

3.(5 points) Use the orange juice worksheet to answer question 3. This worksheet includes one or more duplicates of the same rows. Each combination of the three fields STORE, WEEK and BRAND should be unique. Import the worksheet to Access and copy the table to a new table with structure only. Following the method described in the handout on Access, open the copy in design view, click Indexes under Table Tools, and create a primary key that is the combination of Store, Week and Brand. Save the change. Append the original table to the copy to remove duplicates. Your answer to question 3 is the table you get after removing duplicates.

Note: In problem 3, please use the method described in Case 3, pages 33-34 of the primer on Access.

