**Practice with CASE statements – Analyzing Health Data**

Text

Description automatically generatedThis assignment provides a chance to work with SQL CASE statements to organize data by categories (for subsequent analysis) and to create new columns of textual analytics. SQL CASE statements are used to convert continuous data (i.e., age) into categorical groups. This categorization reduces the cardinality and granularity of the data (please refer to the glossary as necessary to understand these truths) enabling the use of different chart types such as pivot/matrix tables, pivot charts, and slicers.

The general process of this assignment is to first create the SQL query that creates some new columns of analytics in your resultset, then use PowerBI to perform analysis of the dataset using the new textual dimension columns created by the CASE statements. The dataset is a compilation of over 70,000 patient health records of indicators of heart health and potential Coronary Artery Disease. The dimensions and metrics are shown in the above table (i.e. blood pressure, physical dimensions, and heart rate when exercising using a [treadmill stress test](https://www.health.harvard.edu/heart-disease/cardiac-exercise-stress-testing-what-it-can-and-cannot-tell-you)). In a treadmill stress test, the doctor keeps raising the speed of the treadmill until the patient is overwhelmed, exhausted and begging to stop the test (or their heart rate reaches 200 beats per minute). The patient’s heartrate is monitored and measured at the point of stopping the test. If the patient’s recorded treadmill test heart rate is within 85% of their max heart rate (in the image below) then they hit their peak and are deemed ‘at risk’). Conversely a patient whose heart rate does not spike to maximum allowable (85% of the numbers below for their age) heartrate is healthy, and a person whose treadmill heartrate reaches only 50% or 60% of their max are considered extremely healthy (hard for them to spike their heart rate, as athletes are accustomed to the exercise, and their body’s cardiovascular system has adjusted to exercise.

In this research project you will be able to use PowerBI to generate interesting insights. Motivating the future use of arrays, you may also discover that there are some analytics that are exceedingly difficult or impossible to generate, without usage of arrays.

Procedure

1. Write a SQL SELECT query that pulls all the columns from the table above (currently in the featherman\_analytics database on cb-ot-devst05.ad.wsu.edu…..also check devst06 as needed).

2. Notice the AvgMaxHeartRate is empty in the [Health\_heart\_experimental] table. When you query this table use a CASE statement to fill this column with values. The maximum heart rate for each of age group is shown in the image in the appendix. Th AvgMaxHeartRate column should display 85% of the published number for their age group.

3. USE SQL CASE statements to ***create three new columns*** each that categorizes the patients  
  
a) Make a new column that has age groups that match the groups defined in the appendix  
b) Make a new column that that has BMI groups that match the groups defined [here](https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html) or similar BMI chart.  
c) Make a new column that categorizes the results of each patients treadmill test. As discussed above a person whose treadmill test hit 199 these people are much less healthy than others whose treadmill test remained below the max for their age group (85% of the number in the image below). You create the categories for this column  
  
d) Add other columns to your SQL Select statement to perform some analysis (the analysis can either be performed in SQL or in PowerBI).   
  
e) Perfect this query in SSMS and then copy it into the appendix of your turn-in.

3) Next open PowerBI and copy the query to pull the data into PowerBI (watch provided PowerBI videos or content from a different source as necessary). Build a report that includes a series of charts and tables that analyze the data. Add your interpretation and commentary next to each chart or table that describes the findings, perhaps connecting the findings of the current chart/table with the prior. Build a brief report of your charts/tables and analysis.   
  
  
***Some suggestions for charts/tables include the following****:*  
  
a) crosstab/matrix reports of age group and BMI group (both of these groups created by the CASE statements)  
for the values place **counts** of the ID. Next click the Show value as option for the values and make a matrix report of interesting measures using the % of Row Total and % of column total summarization.   
  
b) You can also make some interesting matrix reports using counts and percentages, such as percent people in an age category or BMI group cross tabulated (broken into columns) by weight, or categorical results of the treadmill test. For example this analysis can enable a comparison of BMI by age group, or treadmill results by age group, or treadmill results by BMI group (cross-tabulated by age). Report your insights. A stacked column chart might help visualization of results.  
  
c) Column charts of BMI group sliced by Age group or other slicer might be interesting.  
  
d) Line charts of age and average weight, age and height/weight ratio, average blood pressure, etc.4. Reflecting on the data and analysis: Include a paragraph that provides a few suggestions for additional data needed, such as more demographic data that might help to tease out more insights from the dataset. Also provide a few suggestions as to what other grouping of data (or of new data) would be useful.   
  
5. Finish with a statement of any concerns with the dataset or approach that might jeopardize the validity of your findings. Add a statement of any roadblocks you ran into, such as wanting to create a new metric but that was impossible with your current SQL knowledge. A future assignment will give you a chance to remove these roadblocks, by usage of arrays (table variables).

**Turn-in**

Analyze the data and write a brief 3- or 4-page report (including cropped, screenshot images of your charts and matrix reports). Higher grades for professional, well-formatted and thoughtful analysis.

**Appendix**

Here is the chart needed to calculate the average maximum heart rate by age group

**Table

Description automatically generated**