Performance Assessment D211 – Advanced Data Acquisition

Part 3

Doug Haunsperger

College of IT, Western Governors University

February 19, 2024

## C1. Dashboard Alignment

According to the scenario provided by WGU, our hospital chain data analyst team has been “asked to investigate the extent to which readmission is a problem” for our hospital system (*D210,* n.d.). The scenario also states that the data gives us “the opportunity to predict readmission based on other conditions and factors of the patient” (*D210,* n.d.). Given this, I chose to imagine that there was an ongoing project by the hospital to control patients’ Vitamin D levels to minimize readmissions.

I chose to compare our patients’ Vitamin D blood serum levels to those of the total US population. To do so, I combined the WGU medical data with the 2017-2018 National Health and Nutrition Examination Survey (NHANES) performed by the Centers for Disease Control. (CDC, 2022).

## C2. BI Tool

I chose to use Tableau as a business intelligence tool. It is a “BI and data visualization tool that leverages visual analytics to empower people and organizations in making the most of their data” (Leung, 2021). It allows for easy connection to our PostgreSQL database, and quickly creating visual dashboards from that data. As Leung says, “These dashboards go a long way in aligning an organization’s strategic efforts, uncovering critical insights, and speeding up enterprise-wide decision-making.” (2021).

## C3. Data Cleaning & Preparation

To prepare the data for analysis, I first had to import and clean the desired data from the NHANES study. The CDC releases this data in the SAS Transport format. I had to load these files into the freely available SAS Universal Viewer, and then export them as CSV. Once in CSV, I used Excel to delete columns I was not interested in. I next imported the CSV files into the PostgreSQL database by creating tables and using the pgAdmin import interface as described in section A4 of the Part 1 README document. The NHANES data is largely numerically coded (i.e. for the Gender column, a value of ‘1’ means ‘Male’, ‘2’ means ‘Female’, etc.) - see data dictionaries at (CDC, 2022). When connecting our WGU patient data with the NHANES data in Tableau, I used the custom SQL code shown in section A4 of the README document to do the conversion of these numeric codes to meaningful text strings.

Finally, I used Tableau’s rename function to give the NHANES data columns meaningful names.

## C4. Dashboard Creation

To create the submitted dashboard, I first created four worksheets – Vitamin D Serum Levels, Marital Status, Gender, and Income Level.

### Vitamin D Serum Levels

For this worksheet, I performed the following steps:

* + Create calculated field “VitD (nmol/L)” for WGU patient Vitamin D levels in nmol/L to match the NHANES data (conversion factor of 2.5 from ng/mL)
  + Create bins for the WGU “VitD (nmol/L)” and NHANES “Tot Vitd” measures, each 5 nmol/L in width.
  + Place these created bins as columns, and the NHANES and patient CNTs as rows
  + Set both rows and columns to dual axis to create a dual histogram
    - Measure Name is set to color the marks, and given meaningful names “NHANES Respondents” and “Our patients”
  + Capped the x-axis display at 200 nmol/L due to a few outliers in the NHANES data

### Marital Status

For this worksheet, since the WGU patient data and NHANES data are joined from different tables with marital status in different fields, I used a calculated field “Marital Status” to unify the marital status from each. The code for this is as follows:

IF [NHANES Marital Status] == 'Married' OR [Marital] == 'Married' THEN 'Married'

ELSEIF [NHANES Marital Status] == 'Divorced' OR [Marital] == 'Divorced' THEN 'Divorced'

ELSEIF [NHANES Marital Status] == 'Separated' OR [Marital] == 'Separated' THEN 'Separated'

ELSEIF [NHANES Marital Status] == 'Never Married' OR [Marital] == 'Never Married' THEN 'Never Married'

ELSEIF [NHANES Marital Status] == 'Widowed' OR [Marital] == 'Widowed' THEN 'Widowed'

ELSEIF [NHANES Marital Status] == 'Partnership' THEN 'Partnership'

ELSE 'NULL' END

I then performed the following steps:

* + Set the new “Marital Status” measure as the row, and the NHANES and patient CNTs as columns.
  + Set dual axis on the columns
  + Change the CNTs to a quick table calculation of “Percent of Total”
  + Filter out NULL responses
  + Measure Name is set to color the marks, and given meaningful names “NHANES Respondents” and “WGU Patients”

### Gender, Income Level

These worksheets were set up identically to Marital Status, except using different calculated field code as follows:

Gender -

IF [NHANES Gender] == 'Male' OR [Gender] == 'Male' THEN 'Male'

ELSEIF [NHANES Gender] == 'Female' OR [Gender] == 'Female' THEN 'Female'

ELSE 'Other' END

Income Level -

IF [NHANES Income Range] == '< $5k' OR [Income] < 5000 THEN '< $5k'

ELSEIF [NHANES Income Range] == '$5k - $10k' OR [Income] < 10000 THEN '$5k - $10k'

ELSEIF [NHANES Income Range] == '$10k - $15k' OR [Income] < 15000 THEN '$10k - $15k'

ELSEIF [NHANES Income Range] == '$15k - $20k' OR [Income] < 20000 THEN '$15k - $20k'

ELSEIF [NHANES Income Range] == '$20k - $25k' OR [Income] < 25000 THEN '$20k - $25k'

ELSEIF [NHANES Income Range] == '$25k - $35k' OR [Income] < 35000 THEN '$25k - $35k'

ELSEIF [NHANES Income Range] == '$35k - $45k' OR [Income] < 45000 THEN '$35k - $45k'

ELSEIF [NHANES Income Range] == '$45k - $55k' OR [Income] < 55000 THEN '$45k - $55k'

ELSEIF [NHANES Income Range] == '$55k - $65k' OR [Income] < 65000 THEN '$55k - $65k'

ELSEIF [NHANES Income Range] == '$65k - $75k' OR [Income] < 75000 THEN '$65k - $75k'

ELSEIF [NHANES Income Range] == '$75k - $100k' OR [Income] < 100000 THEN '$75k - $100k'

ELSEIF [NHANES Income Range] == '> $100k' OR [Income] > 100000 THEN '> $100k'

ELSE 'NULL' END

Once all four worksheets were created, I created the dashboard, renamed it “Demographic Comparison in Vitamin D Levels”, and added the 4 worksheets to it. I made the ‘Measure Names’ legend floating and renamed it ‘Legend’. Finally, I set each of the 3 demographic worksheets to ‘Use as Filter’ so that a selection of any demographic component (i.e., income of ‘> $100k’) will filter the Vitamin D histogram to only those patients who fit those criteria.

## Additional Data Set Insights

The HRRP data includes predicted and actual readmission rates for all reporting hospitals in the country (CMS, 2023). I aggregated these scores at the state and regional level to provide a comparison baseline for our hospital system.

The NHANES data includes demographic information and Vitamin D blood serum laboratory data for a representative cross-section of the US population (CDC, 2022). I combined this data with our hospital dataset to be able to compare Vitamin D levels with the national population, allowing filtering by income level, gender, marital status, or age.

## C5. Executive Decision Support

The scenario identifies three types of executives in the audience: Senior VP, VP of Research, and Regional VPs (WGU, n.d.).

I imagine that the VP of Research is behind my notional ‘Vitamin D Control Project’ and can see how well the hospital is controlling patient Vitamin D levels and can relate this to success (or lack thereof) in reducing readmissions.

The VPs can also examine patient demographics compared to the nationally representative sample in the NHANES data to decide about facility needs, outreach programs, etc. In this case, we can see that our patients’ income levels are generally lower than average, and we have fewer married patients and relatively more separated/divorce/widowed patients.

## C6. Analysis Limitations

We only have a single snapshot of patient data. It would be useful to collect additional data series to see how things trend over time.

We have data for number of children in household and patient age, which could easily be added to the demographic dashboard using the same techniques as above. However, there is much more demographic data available in the NHANES dataset which we do not collect for our patients. Likewise, there is data we collect that is not present in the NHANES set – additional public data sets would need to be researched and acquired to make useful comparisons of that data.

## D. References

Centers for Disease Control and Prevention (CDC). April 2022. National Center for Health Statistics. *National Health and Nutrition Examination Survey 2017-2018.* <https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/releasenotes.aspx?BeginYear=2017>

Leung, K. September 2021. Datacamp Blog. *How Tableau Helps Your Organization Achieve Greater Data Insights.* <https://www.datacamp.com/blog/how-tableau-helps-your-organization-achieve-greater-data-insights>

Western Governors University (WGU). n.d. *D210 Medical Data Considerations and Dictionary.* <https://access.wgu.edu/ASP3/aap/content/ds0fh43lkfd9tf85kvmd.zip>