Advanced Data Acquisition

WGU Masters of Data Analytics course D211

Project by Natalie Toler, Student ID: 011148248

Contents

Part One - Data Dashboards

A1. The Datasets

As I am working with the same data as the previous course, D210, I have utilized the same external dataset. That data is the Healthcare Cost and Utilization Project (HCUP) data of insurance payments across the states for the year 2019. Since the HCUP data is more comprehensive than the WGU data I have modified it to include only one quarter of the 2019 costs. The HCUP data has also been cleaned and anonymized already as per the need to keep privacy for healthcare data. Since this is the case and there was no ID column I also added a sequential order column to maintain integrity of the data.

The WGU data is hosted in the virtual lab. The CSV for the HCUP data has been included in the submission of this PA.

A2. Dashboard installation

In order to access the dashboard the virtual space has to be set up with the SQL queries so that Tableau can correctly set up the dashboard. In order to facilitate the set up I have included the SQL queries and tableau dashboard file. Since the project is hosted on a virtual site, in order to install the dashboard the provided documents must be emailed or uploaded into the cloud so that the user can access the files.

Instructions for Dashboard Installation:

1. Upload the provided files into the virtual lab space.
2. Open pgAdmin and navigate to the "*medical_data*" database.
3. Right click on "*medical_data*" and select the Query tool.
4. Copy and paste the SQL Query from the provided "*sqlquery*" file into the Query tool.
5. Please ensure that the "*hcup_data.csv*" file is either in location "*C:\Users\Public\Downloads*" which can be done by dragging and dropping the file to the location in the file explorer application or change the line in the SQL query to accurately reflect where you have stored the "*hcup_data.csv*" file.
6. Click on the play button to execute the Query.
7. Once complete close pgAdmin.
8. Double click on the file "*WGU and HCUP*"
9. When prompted to connect to postgreSQL use the username "*postgres*" and the password "*Passw0rd!*"
10. The dashboards will appear and can be interacted with.

A3. Dashboard Navigation

The provided file includes the individual visualizations along with the two dashboards and a story presentation. In order to view the dashboards comprehensively the user should navigate to the story presentation.

The first dashboard consists of two maps of the United States of America with each state being shaded darker or lighter based on the average cost of hospitalization within the state in either the WGU dataset or the HCUP dataset. Both of these maps can be filtered by the state, for instance if a user would like to see only Wyoming, West Virginia, and Vermont they may go to the state drop down filter and select those three states to better compare individual states to each other and between the datasets.

The WGU map also has a dropdown menu filter for the services that a patient had while in hospital which will allow the user to further filter the cost by what services cost on average by state. The HCUP data has a dropdown menu filter for Hospitalization Type so that the cost can be filtered that way as well. Each state on the maps also has a tooltip that will display the average cost when hovered over.

The second dashboard contains two sets of comparative data. The WGU hospital data broken down by the average cost of each type of initial admission recorded in the data for each state. This visual can be filtered in several ways. First of all if the user clicks on "Elective", "Emergency", or "Observation" they can reorder the columns to show the max or min average. Alternatively by clicking on the state abbreviations the rows will reorder to be the max or min admission type for that state. By clicking the top most left area of the visual the user can return the data to alphabetic order. To the upper right the user can also filter the data by age group, either 18-64 or 65+. I have also included a second filter for initial admission so that a user can filter by admission type in a cleaner manner for better analysis.

Likewise the corresponding HCUP visual has all of the same ways to reorder the data. There are also two filters to the bottom right, one for hospitalization type and one for the expected payer type.

Below these two visuals are two fairly simple charts that show the cost of hospitalization by the age groups that are recorded in the datatypes. For WGU I have sorted the ages into two categories, 18-64 and 65+. I have included filters for initial admission type and state to further breakdown the data. For the HCUP data the ages are split by the expected payer, with many more options than the WGU data. This chart also has filters for Hospitalization type and state.

A4. SQL Query

The SQL query used to create the table used for the dashboards is included below and as a separate "SQL Query" file.

```
ALTER TABLE public.patient ADD COLUMN source text;
UPDATE public.patient SET source = 'WGU';
```

```
ALTER TABLE public.servicesaddon ADD FOREIGN KEY (patient_id)
REFERENCES public.patient(patient_id);

CREATE TABLE wgu_compare AS (
        SELECT pat.age, loc.state, pat.initial_days, pat.totalcharge,
admis.initial_admission, services.services, complication.complication_risk,
pat.source
        FROM patient AS pat
        INNER JOIN admission AS admis ON pat.admis_id=admis.admins_id
        INNER JOIN location AS loc ON pat.location_id=loc.location_id
        INNER JOIN servicesaddon AS services ON
pat.patient_id=services.patient_id
        INNER JOIN complication ON
pat.compl_id=complication.complication_id);

CREATE TABLE public.hcup_compare (
        order_id numeric, state_1 text, hospitalization_type text, expected_payer
text, cost numeric);
ALTER TABLE public.hcup_compare
        OWNER TO postgres;
COPY hcup_compare
FROM 'C:\Users\Public\Downloads\hcup_data.csv'
DELIMITER ','
CSV HEADER;

ALTER TABLE public.hcup_compare ADD COLUMN source_1 text;
UPDATE public.hcup_compare SET source_1='HCUP';

CREATE TABLE wgu_hcup_compare AS (
        SELECT *
        FROM wgu_compare
        FULL OUTER JOIN hcup_compare
        ON wgu_compare.state = hcup_compare.state_1);
```

Part Two - Demonstration

B. Video

I have provided the link to the Panopto video here as well as separately with the submission of this PA.

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=d2bebad6-c85d-49e4-a8fb-b17800530785

Part Three - Report

C1. *Purpose and Functionality*

The purpose of my dashboard is to illustrate the difference of analysis between having too vague of data in the case of the WGU dataset and more specific, decision-level data as is the case of the HCUP data. My intention to to show the stakeholders of the WGU Hospital System that a restructuring of the data keeping process should be undertaken to align with the standards of the Healthcare Cost and Utilization Project. In obtaining better data that records things like who is paying for hospitalization and more specific categories for hospitalization type I believe that the WGU Hospital System can better optimize performance for both the hospital and the patients.

By exploring the contrasting visualizations between the WGU data and HCUP data it becomes clear that analysis on the HCUP data leads to better executive decision making to find ways that the hospitals can improve costs and performance.

C2. *Tableau*

Tableau is an intuitive visualization program that uses many behind the scenes SQL queries to allow an analyst to both explore their data and create tailored dashboards. Along with allowing for beautiful visuals Tableau has easy to create filter and sorting systems so that nearly every part of the dashboard can be used by analysts and stakeholders to see the data and make clear stories and conclusions from the data.

C3. *Clean and Prepare*

To prepare the dataset for use in tableau I ensured that each of the "medical_data" tables had appropriate foreign keys to keep referential integrity. I also added a source column so that I would know where the data came from, although due to the nature of the analysis and the difference in data sets this step was not necessary however I believe important for any further analysis. I then created a table with only the columns I needed to use from the "medical_data" database. For use in this analysis further cleaning of the data was not needed.

For the HCUP data I added a column in the CSV file of merely a sequential order since the HCUP data has already been thoroughly cleaned and anonymized I wanted to ensure some kind of primary key for the data. I then created a new table in pgAdmin and populated it with the CSV file of the HCUP data. I also added a source column as well.

The last step was to combine the WGU table I had created and the HCUP table. I did this using a FULL OUTER JOIN because I needed to maintain all of the columns from both tables.

C4. *Dashboard Creation*

I started the dashboard by creating the two map visuals. Using tableau's function to automatically create a map when given a column consisting of places I merely dragged the State columns onto the work space then used color to chart the average cost of hospitalization of each state. To create the filters I dragged the columns I wanted to use as filters into the filter section of the workspace and made those filters visible.

Since I knew I would use bar charts for the age breakdowns I decided to create heat maps for the breakdown of cost in each state by the type of hospitalization. These were created by putting the state and "hospitalization_type" or "initial_admission" columns in the row and column sections of the work space and then sorting by average cost by color. Again to add the filters I put the columns I wanted to filter by in the filter section of the workspace and made them visible.

The bar charts for age and cost were created by putting the age and cost columns into the row and column sections. The one further action I did for age was to create an in/out sort of the WGU ages so that I could compare age groups similar to those in the HCUP data. This was done by right clicking on the age column and creating an in/out group with the in group being ages 18 to 64 and the out group being ages 65 and older.

To create the two dashboards all I had to do was to drag each sheet containing the visuals I wanted to compare to the dashboard space. I made the filters float so that they didn't take up as much space and organized them in a way that was intuitive to the user.

C5. *Results*

The results of my analysis is that the HCUP data collecting standard leads to better executive decision making. By being able to filter the cost of hospitalization by who the expected payer is allows for a better understanding of where the hospital systems are getting their money from. Having this understanding will allow the hospital systems to utilize their strongest form of income and come up with solutions to pad up revenue streams that are lacking.

Furthermore the HCUP data keeps more specific values for hospitalization type. While the WGU data only have broad categories, the HCUP data gives us actionable data for working on increasing or decreasing certain hospitalization types, focusing on who pays the most for which type of hospitalization and so on. Having this level of detail cuts down the time needed for analysis and allows for quicker decisions to improve hospital functions.

Overall, the dashboards serve as an illustration for why the WGU hospital system needs to revamp their data collection and work to have better records for patients, payouts, and services.

C6. *Limitations*

There are several limitations to the data analysis. The first is that the WGU data has no timeframe attached to it and therefore cannot accurately be compared to the HCUP data which is very specifically from quarter one of 2019. Having a timeframe in the data is important to most analysis and therefore is a huge limit to this analysis. Although I would argue that this is just one more area that the WGU data should be improved on.

Another limitation is that the WGU data has 10,000 patients while the HCUP data has a little under 2000. 10,000 points of data is much more comprehensive than 2,000, especially when considering that the HCUP data consists of many hospital systems and in theory the WGU data is only one hospital system. Therefore we have much more information about the average of the WGU patients than the HCUP patients.

Sources

No in-text citations were used for this project.

This project borrowed heavily from my previous project assessment in course D210.

The Healthcare Cost and Utilization Project is open source data which can all be found on their website. https://www.ahrq.gov/data/hcup/index.html

The course DataCamp classes provided the framework for my SQL queries.

Additionally the webinar for D211 was a great source.

Bibliography

AHRQ. "Healthcare Cost and Utilization Project (HCUP) | Agency for Healthcare Research & Quality." *Ahrq.gov*, Dec. 2022, www.ahrq.gov/data/hcup/index.html. Accessed 21 May 2024.

Datacamp. "D211 Advanced Data Aquisition." *Datacamp*, app.datacamp.com/learn/custom-tracks/ custom-d211-advanced-data-acquisition. Accessed 21 May 2024.

Salesforce. "What Is Tableau Public?" *Tableau.com*, 2022, public.tableau.com/app/about. Accessed 21 May 2024.

Sewell, Dr. William. "D211_Webinar." *Wgu.hosted.panopto.com*, wgu.hosted.panopto.com/Panopto/ Pages/Viewer.aspx?id=ccb02dcf-8d56-4340-82d3-b0bb00e4de5c. Accessed 21 May 2024.