**PERFORMANCE ASSESSMENT**

**D211 – Advanced Data Acquisition**

Michael Lawson

Western Governors University

D211 – Advanced Data Acquisition

David Gagner

09/01/2022

Michael C. Lawson | Student ID: #001479557 | Program: MS Data Analytics (5/1/21) |

Program Mentor: Emil Stoica | 859-265-1888 (EST) | [mlaws77@wgu.edu](mailto:mlaws77@wgu.edu)

**Part 1: Interactive Dashboard**

1. **Data sources:**
   1. WGU Medical Data Set: Medical\_clean.csv
   2. External data source: Data.NY.gov. Mental Health Readmission: Beginning 2014. <https://data.ny.gov/Human-Services/Mental-Health-Readmission-Beginning-2014/ke9m-imcz>
2. **Dashboard Access**:

To navigate to the dashboard, type the following web address into any browser address bar

(You can also hold the “Cntrl” key and click the link below to open directly from this document):

<https://public.tableau.com/app/profile/inigomontoya/viz/FauxReadmissionData/Dashboard1>

1. **Dashboard Instructions:**

**“% Readmitted Patients by Region” map:**

* The color spectrum legend in the lower left corner reveals how the color gradient of the map corresponds to the percentage of patients readmitted per the four US regions.
* Hovering over each region will display information for that area. Click on the region to change other tiles in the dashboard to information for that region.
* When a region is clicked with the left mouse button, the states in the data map tile to the right that are in that region will be highlighted. The “Patient Survey Answers” at the bottom will change to only contain survey answers from that region. The “Services Administered to All Patients in Area” data tile will change to only data from that region. Notice that the New York mental health data tile only shows up when the “North” region is highlighted.

**“% Readmitted Patients by State” map:**

* Like the previous map, the gradient color bar at the bottom reveals how the percentage of readmitted patients in each state corresponds to the colors of each state.
* Hovering the mouse over the map will automatically reveal an informational box with the state’s abbreviation and the actual percentage of patients readmitted in that state.
* Clicking on a state with the left mouse button will change the data in the “Services Administered to Patients” and “Patient Survey Answers” data tiles.
* Hold down “Cntrl” and left click on multiple states to reveal data for multiple states. This allows for multiple state selections within a region.
* An interactive box in the lower right of the map can be used to search or choose a state from a list by left clicking the mouse in the search box labeled “Highlight State”.

**“Readmission of New York Mental Health Patients”:**

This is an informational tile from the state of New York’s open data source Data.NY.gov. This tile is interactive and only appears when New York is included in the selection on the map tiles. The data is open for research and education at no cost at the following link which is displayed at the bottom of the dashboard tile:

<https://data.ny.gov/Human-Services/Mental-Health-Readmission-Beginning-2014/ke9m-imcz>

The line graph reveals the percentage of readmitted mental health patients from 2014-2019 in hospitals in New York. It can be used to compare the percentage of New York mental health patients against data from hospitals in the internal data.

**“Services Administered to Patients”:**

This tile is a count on services administered for all patients, both initial and readmits, that changes with the chosen areas on the maps above. When no area is chosen, the entire data set is used.

**“Patient Survey Answers”:**

This tile shows the answers (8 being the least important and 1 being most important) to a survey taken by patients in the data set. The data is separated by gender for even more insight.

1. **SQL Code:**
2. CREATE TABLE public."NY\_MH\_Readmits"
3. (
4. "Metric ID" numeric,
5. "Metric Description" character varying,
6. "Grouping Description" character varying,
7. "Region or Coverage Category" character varying,
8. "Age Inpatient Type" character varying,
9. "Year" numeric,
10. "Quarter" character varying,
11. "Numerator" numeric,
12. "Denominator" numeric,
13. "Rate" numeric,
14. "YTD Numerator" numeric,
15. "YTD Denominator" numeric,
16. "YTD rate" numeric,
17. "State" text
18. );
19. – *import external csv file to database*
20. --command " "\\copy public.\"NY\_MH\_Readmits\" (\"Metric ID\", \"Metric Description\",
21. \"Grouping Description\", \"Region or Coverage Category\", \"Age Inpatient Type\", \"Year\",
22. \"Quarter\", \"Numerator\", \"Denominator\", \"Rate\", \"YTD Numerator\", \"YTD
23. Denominator\", \"YTD rate\", \"State\") FROM 'C:/Users/LabUser/Desktop/MENTAL~1.CSV'
24. OIDS DELIMITER ',' CSV QUOTE '\"' ESCAPE '''';""
25. CREATE TABLE public.admission
26. (
27. admins\_id integer NOT NULL,
28. initial\_admission text,
29. PRIMARY KEY (admins\_id)
30. );
31. CREATE TABLE public.complication
32. (
33. complication\_id integer NOT NULL,
34. complication\_risk text,
35. PRIMARY KEY (complication\_id)
36. );
37. CREATE TABLE public.job
38. (
39. job\_id integer NOT NULL,
40. job\_title text,
41. PRIMARY KEY (job\_id)
42. );
43. CREATE TABLE public.location
44. (
45. location\_id integer NOT NULL,
46. zip integer,
47. city text,
48. state text,
49. county text,
50. PRIMARY KEY (location\_id)
51. );
52. CREATE TABLE public.patient
53. (
54. patient\_id text NOT NULL,
55. lat numeric,
56. lng numeric,
57. population integer,
58. children integer,
59. age integer,
60. income numeric,
61. marital text,
62. readmis text,
63. gender text,
64. initial\_days numeric,
65. totalcharge numeric,
66. additional\_charges numeric,
67. vitd\_levels numeric,
68. doc\_visits integer,
69. full\_meals integer,
70. vitd\_supp integer,
71. soft\_drink text,
72. hignblood text,
73. stroke text,
74. job\_id integer,
75. compl\_id integer,
76. admis\_id integer,
77. location\_id integer,
78. PRIMARY KEY (patient\_id)
79. );
80. CREATE TABLE public.servicesaddon
81. (
82. patient\_id text NOT NULL,
83. services text,
84. overweight text,
85. arthritis text,
86. diabetes text,
87. hyperlipidemia text,
88. backpain text,
89. anxiety text,
90. allergic\_rhinitis text,
91. reflux\_esophagitis text,
92. asthma text,
93. PRIMARY KEY (patient\_id)
94. );
95. CREATE TABLE public.survey\_responses\_addon
96. (
97. patient\_id text NOT NULL,
98. item1 integer,
99. item2 integer,
100. item3 integer,
101. item4 integer,
102. item5 integer,
103. item6 integer,
104. item7 integer,
105. item8 integer,
106. PRIMARY KEY (patient\_id)
107. );
108. ALTER TABLE public.patient
109. ADD FOREIGN KEY (admis\_id)
110. REFERENCES public.admission (admins\_id)
111. NOT VALID;
112. ALTER TABLE public.patient
113. ADD FOREIGN KEY (compl\_id)
114. REFERENCES public.complication (complication\_id)
115. NOT VALID;
116. ALTER TABLE public.patient
117. ADD FOREIGN KEY (location\_id)
118. REFERENCES public.location (location\_id)
119. NOT VALID;
120. ALTER TABLE public.survey\_responses\_addon
121. ADD FOREIGN KEY (patient\_id)
122. REFERENCES public.patient (patient\_id)
123. NOT VALID;
124. ALTER TABLE public.servicesaddon
125. ADD FOREIGN KEY (patient\_id)
126. REFERENCES public.patient (patient\_id)
127. NOT VALID;
128. ALTER TABLE public."NY\_MH\_Readmits"
129. ADD FOREIGN KEY ("State")
130. REFERENCES public.location (state)
131. NOT VALID;
132. END;
133. SELECT "patient"."additional\_charges" AS "additional\_charges",
134. "patient"."admis\_id" AS "admis\_id",
135. "patient"."age" AS "age",
136. "patient"."children" AS "children",
137. "patient"."compl\_id" AS "compl\_id",
138. "patient"."doc\_visits" AS "doc\_visits",
139. "patient"."full\_meals" AS "full\_meals",
140. CAST("patient"."gender" AS TEXT) AS "gender",
141. CAST("patient"."hignblood" AS TEXT) AS "hignblood",
142. "patient"."income" AS "income",
143. "patient"."initial\_days" AS "initial\_days",
144. "patient"."job\_id" AS "job\_id",
145. "patient"."lat" AS "lat",
146. "patient"."lng" AS "lng",
147. "patient"."location\_id" AS "location\_id",
148. CAST("patient"."marital" AS TEXT) AS "marital",
149. CAST("patient"."patient\_id" AS TEXT) AS "patient\_id",
150. "patient"."population" AS "population",
151. CAST("patient"."readmis" AS TEXT) AS "readmis",
152. CAST("patient"."soft\_drink" AS TEXT) AS "soft\_drink",
153. CAST("patient"."stroke" AS TEXT) AS "stroke",
154. "patient"."totalcharge" AS "totalcharge",
155. "patient"."vitd\_levels" AS "vitd\_levels",
156. "patient"."vitd\_supp" AS "vitd\_supp"
157. FROM "public"."patient" "patient";

**Part II: Storytelling with Data:**

Panapto Video Link: <https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=4d9caa3b-0cda-4e90-8d00-af03017ed658>

**Part III: Reflection Paper**

1. The data dictionary was not provided in the lab environment, so the data dictionary was used from previous courses that matched the medical\_data set. The task given in the data dictionary “to investigate the extent to which readmission is a problem for this chain of hospitals… to build a data dashboard to enable executive leaders to explore the data, identify trends, and compare key metrics”. All visualizations in the dashboard are controlled by two maps, which reveal statistics on patient readmission rates. The external data from mental patients of New York gives insight to how the internal data compares to readmission patients outside of the organization.

The Senior Vice President (SVP) represents all locations, so the map was not limited to one region or state. The dashboard allows the vice president to interactively explore all states and geographical regions where the company has hospitals. The SVP is interested in patient treatments and the outcomes of those treatments, so the tile for “Services Administered to Patients” was added, and the interactive map allows the SVP to have insight into how the balance of services administered compares to the readmission rates for each state or region. The survey answer results are separated by gender, and the SVP can choose a state or region and see how survey results differed by region and by gender at the same time.

The “Patient Survey Answers” tile was specifically chosen for the Vice President of Research (VP) who is looking to identify patterns in patient care. The VP can also refer to the “Services Administered to Patients” visual to identify the trend in services administered by region or state. Regional VPs can separate the data with the click of a mouse to report findings on their specific region.

The executives want to know how the hospital’s staff and procedures affect readmission, so the variables that contributed the most to this information were the patient survey results and the services the hospital administered. Adding in information like patient conditions, income, or hospital charges would not have helped executives make decisions on services and treatments, so those variables were excluded. The external data, “Readmission of New York Mental Health Patients” was used to help executives compare the readmission rates for multiple years to the readmission rates of internal data. The data dictionary gave insights into what the audience required, including regional data on readmission rates, data specific to services, treatment and staff. Only data that gave insight into these requirements was used.

1. Tableau is the business intelligence tool chosen to present the medical data set. Tableau is simple to use because of features like the ability drag and drop tables into presentations, and Tableau automatically chooses the best visualizations for the chosen data. Tableau offers a wide variety of options to analyze and present data. Tableau can “handle copious amounts of data” (Dewan, 2019). The program has no issues handling millions of rows of data. Tableau integrates scripting languages commonly used by data analysts such as Python and SQL. Users can import data from virtually any data source. Tableau was also chosen because it offers public access to dashboards that do not require the stakeholders to download any special software.
2. Only the external data brought to the data set needed to be cleaned. The external data brought in did not have nulls, but the column “State” in the original data set spelled out the full name of “New York” rather than the state’s abbreviation “NY”. This was changed in Excel before importing the table into PostgreSQL to match the abbreviations used in the location table of the medical\_data set. Every value in the “State” column was for New York, so making the change to the entire column in Excel was the easiest method for this data set. The new data was then connected as a RELATIONSHIP rather than a JOIN to the location table that already existed in medical\_data. RELATIONSHIP is a more flexible method than JOIN in SQL and can tie various data types and allow tables to be sourced without the same restrictions of JOIN statements. This method was chosen because the data was coming from an external source, and it was the most efficient method if the columns did not perfectly match.

The “IS NULL” statement was used in Postgres SQL to check each column for missing values. “SELECT DISTINCT” was used to check for spelling and capitalization errors. The medical\_data database that was already loaded in to PostgresSQL in the lab environment did not have nulls or errors to be cleaned. Relationships between the tables was checked using the EDR tool in PostgresSQL. The city column was deleted from the external data set, and only the tables and columns needed were pulled into Tableau so that storage space was not improperly used for data that was not needed.

1. The dashboard created in Tableau started with an import of the live database from PostgreSQL into Tableau. When Tableau is initially opened, a “Connect” menu is displayed, and the option for “More…” was chosen from the “To a Server” group on the page. The data previously prepared in PostgreSQL was selected, and the tables of the Medical\_data set were imported. Only the tables that were needed to create the dashboard were dragged into the data source area starting with the “patient” table and then joining the “location”, “servicesaddon” and “survey\_responses\_addon” tables. The external data was joined to the “location” table. Individual worksheets were created and explored using the tables chosen, and then those worksheets were dragged into the Dashboard page of Tableau.
2. The “Patient Survey Answers” visualization reveals responses to questions regarding staff and treatment. Interestingly, the ratio of gender changes when the region or state changes, and the skew of the answers changes by state and region chosen Executives can compare regions and states to see what areas need attention according to the answers patients gave to the survey questions. The ”Services Administered to Patients” tile allows executives to see what services are more common per region or state. When compared to readmission rates given on the map visualizations, executives can make decisions on whether different services are affecting readmission rates. They can also compare the variation in services to the answers to the survey questions. A good example of how this visualization can be used is selecting Kentucky vs Michigan. The “Intravenous” service increases disproportionately from other services from Kentucky to Michigan. Michigan also has a much higher readmission rate. Utah has a much higher CT scan count than Kentucky, but the readmission rate is close to the same. Both visualizations will help the executives make insightful decisions.
3. The data is limited by the information that was collected. The survey questions only cover a very specific part of the patient experience, and only request information on what is important to patients rather than specific information about the stay. The data is limited to this specific healthcare chain’s collection of information and would be more accurate if data could be shared from outside sources to compare statistics. The external data set was limited to the state of New York. It may be helpful to locate data for the remaining states for the external data set. Other factors not in the data set might contribute to readmission rates, and it would be useful to ask internal employees what other measurements or information could be useful to collect to discover what affects readmission rates. The data is not live feed or continually updated, and without dates it is impossible to know what time period the data covered. The data could be old or during a time when something influenced readmission rates that no longer applies.

**References**

Knaflic, Cole Nussbaumer. November 2nd, 2015. *Storytelling with Data : A Visualization Guide for Business Proffesionals*. <https://ebookcentral.proquest.com/lib/westerngovernors-ebooks/detail.action?docID=4187267>

The New York State Office of Mental Health, Office of Performance Measurement and Evaluation (OPME). July 12th, 2021. *Mental Health Readmission: Beginning 2014.* <https://data.ny.gov/Human-Services/Mental-Health-Readmission-Beginning-2014/ke9m-imcz>

Dewan, Smriti. 2019. *Top 5 Reasons Why Tableau is Leading the Business Intelligence Industry.* https://www.grazitti.com/blog/top-5-reasons-why-tableau-is-leading-the-business-intelligence-industry/

Sewell, Dr. Willliam. *SQL Sunday*. <https://westerngovernorsuniversity-my.sharepoint.com/:p:/g/personal/william_sewell_wgu_edu/Ebd2E_F3aRNLkmkIB0aCETIBB356vANHTfA9_c3ImIyo5w?rtime=U7L7TpGL2kg>

PGAdmin. *ERD Tool*. https://www.pgadmin.org/docs/pgadmin4/latest/erd\_tool.html