**Western Governors University (WGU)**

**D211: Advanced Data Acquisition**

**Natallia Zimnitskaya | ID: 012247127**

**Master of Science, Data Analytics**

**Part 1.**

**A.** A copy of my dashboard, D211\_Dashboard.twbx, is attached to this submission in a D211\_PA.zip file.

**A1.** The WGU-provided “medical\_data” originates from a PostgreSQL database in pgAdmin on Virtual Machine. I used the medical\_clean.csv dataset for the D208, D209, and D210 Performance assessments. For this assessment, I needed two tabs from the medical\_data dataset: “patient” and “location” with common location\_id (please see the d211\_medical.csv file attached to this submission)

The external “CMS hospital readmission race/ethnicity 2012-2020” was found on Kaggle <https://www.kaggle.com/datasets/patysanchez/cms-hospital-readmission-raceethnicity-2012-2020>

It provides readmission percentage by race/ethnicity per state and county in the United States of America from 2012 to 2020. The “cms\_readmission.csv” file is attached to this submission in a D211\_PA.zip file.

**A2. Dashboard Installation.**

1. Take the D211\_PA.zip file from this submission, open File Explorer in the virtual machine, navigate to C:\Users\Public\Downloads and paste the D211\_PA.zip file.
2. Right-click on D211\_PA.zip and choose “Extract All...”. Select C:\Users\Public\Downloads as the destination for the extracted files.
3. Double-click pgAdmin on the desktop to open it. Then, in the left navigation pane, scroll down to the “medical\_data” database.
4. Right-click on the “medical\_data” database and choose Query Tool. When the Query Tool panel appears on the right, click the “Open File” button.
5. In the Select File dialog box, go to C:\Users\Public\Downloads and choose the readmission\_dataset\_setup.sql file.
6. After the Query Tool imports the contents of the readmission\_dataset\_setup.sql file, click the Play button in the top right corner to run the query.
7. Double-click Tableau 2021.4 to launch Tableau Desktop.
8. Click File at the top left, then select Open.
9. Navigate to C:\Users\Public\Downloads and choose the D211\_Dashboard.twbx file.
10. When asked to sign in, use the username "postgres" and the password "Passw0rd!".

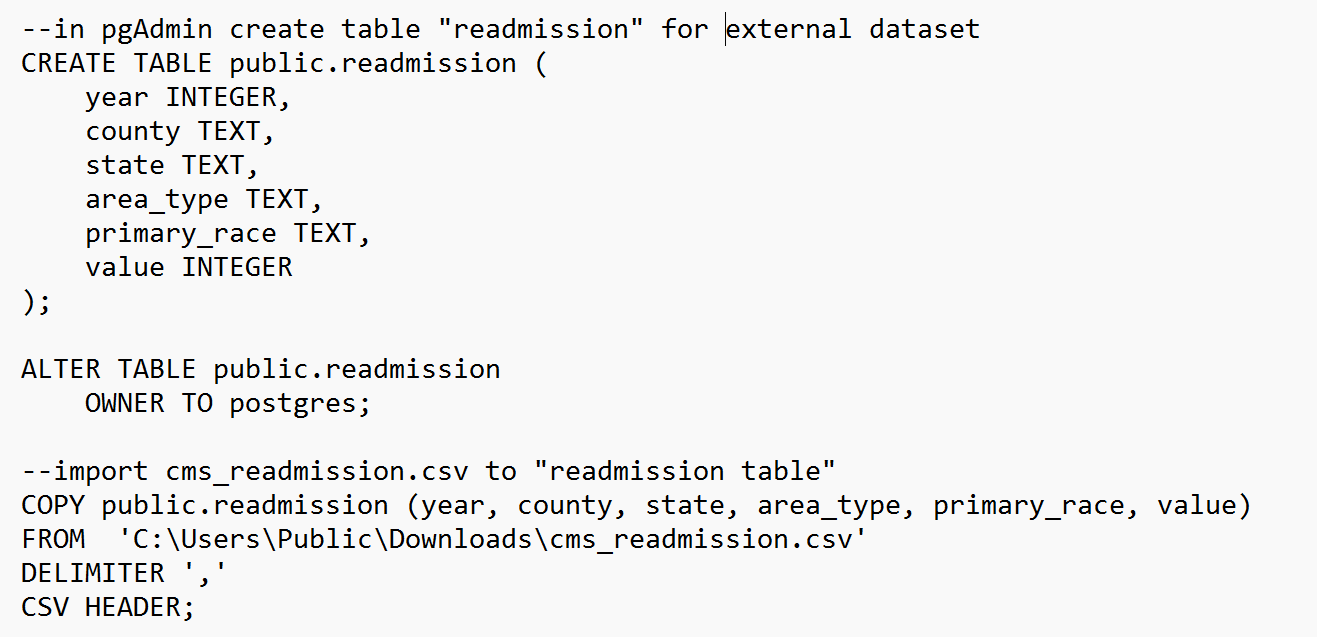
**A3. Navigating the Dashboard**

* Access the Dashboard: Start Tableau and open the “D211\_Dashboard.tbwx” file. All of the data should already be included.
* Enter presentation mode for a better view.
* Overview of the Dashboard Layout:
* KPI: Displays the average age of readmitted patients.
* Heatmap: Shows the average age of readmitted patients by state.
* Pie Chart: Illustrates the gender distribution of readmitted patients.
* Highlight Table: The readmission percentage by area type (rural, urban).
* Horizontal Chart: Displays the percentage of readmissions out of the total hospitalizations for each ethnicity.
* Using Filters:
* Year Filter: Located on the top right side. Use the drop-down menu to select the year(s) you want to analyze (2012-2020).
* Interactive Filtering: Click on any part of the visualizations to filter other charts by that selection. For example, clicking on a specific state in the heatmap will filter the other visualizations to show data related to that state.
* Interacting with Visualizations:
* KPI Average Age of Readmitted Patients: click on the state over the map, gender, area (urban/suburban), or ethnicity to see detailed information on how the average age of readmitted patients changes.
* Heatmap: Hover over a state to see the average age of readmitted patients. Click on a state to filter the other visualizations by that state.
* Pie Chart Patient Gender: Hover over a segment to see the percentage of male or female readmitted patients. Click on a segment to filter the other visualizations by gender.
* Highlight Table Readmission Percentage by Area Type: Click on a cell to filter the other visualizations by area type.
* Horizontal Chart Readmission Percentage by Ethnicity: Click on a bar to filter the other visualizations by ethnicity.

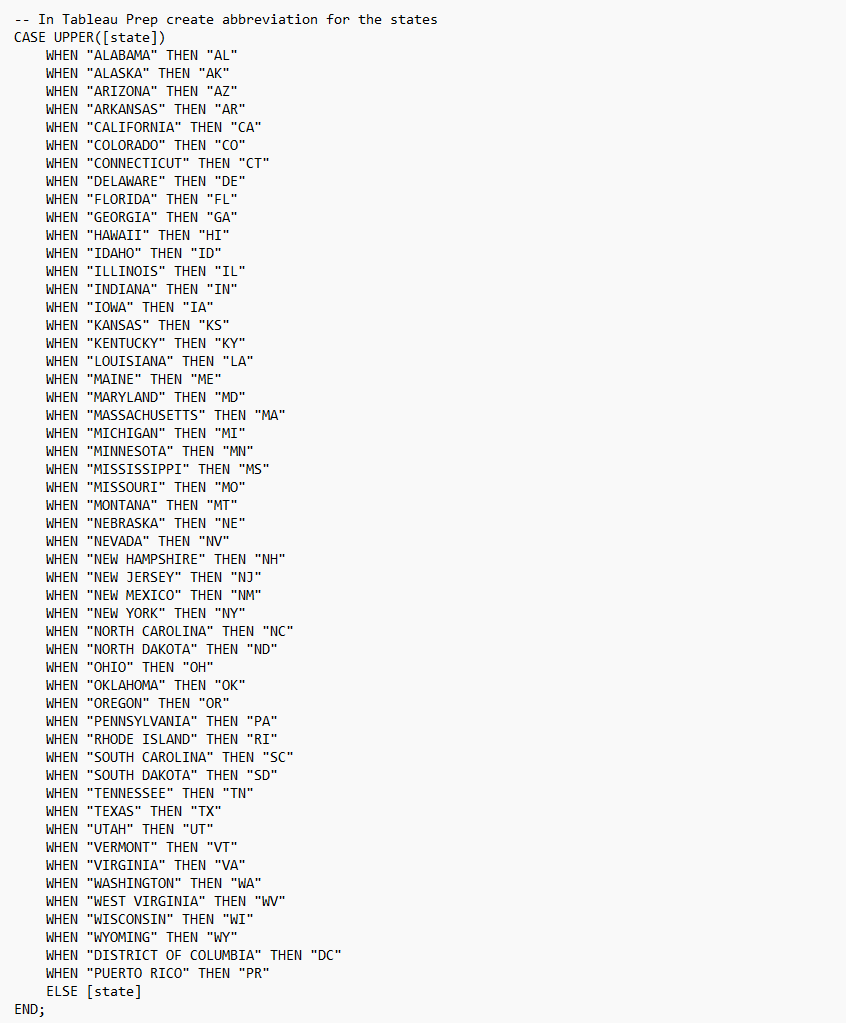
**A4.** **SQL code**

Please see d211\_sql.sql file attached to this submission in a D211\_PA.zip file.

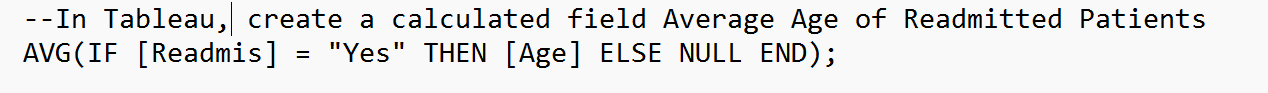
In pgAdmin: Create a " readmission " table for the external dataset and import the “cms\_readmission.csv” file.



In Tableau Prep Builder: Create a Calculated Field for state abbreviation



In Tableau: Create a calculated field Average Age of Readmitted Patients



**Part 2.**

**B.** Here is the link to the demonstration video: <https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=a32f438e-423f-4ec0-a745-b22200271370>

**Part 3.**

**C1. Purpose and Function of the Dashboard.**

The purpose of this dashboard is to provide a clear and interactive visualization of patient readmissions across the hospital chain, focusing on understanding and addressing the extent of the readmission problem. The dashboard aligns with stakeholder needs by highlighting critical trends and disparities in readmission rates, facilitating data-driven decision-making. It includes:

* Heatmap (Average Age of Readmitted Patient by State): Identifies geographic regions with higher readmission rates and older readmitted patients.
* Pie Chart (Patient Gender): Shows the gender distribution of readmitted patients, helping to identify gender-specific trends.
* Highlight Table (Readmission Percent by Area Type): Compares readmission rates between rural and urban areas.
* KPI (Average Age of Readmitted Patients): Provides a quick overview of the average age of readmitted patients.
* Horizontal Chart (Readmission Percent by Ethnicity): Highlights ethnic disparities in readmission rates.
* Year Filter (2012-2020): Allows stakeholders to analyze trends over time and assess the impact of interventions and policy changes.

Additionally, each visualization in this dashboard can be used as a filter, enabling stakeholders to drill down into the data and gain deeper insights through interactive exploration. This enhances the dashboard’s functionality, making it a powerful tool for strategic decision-making.

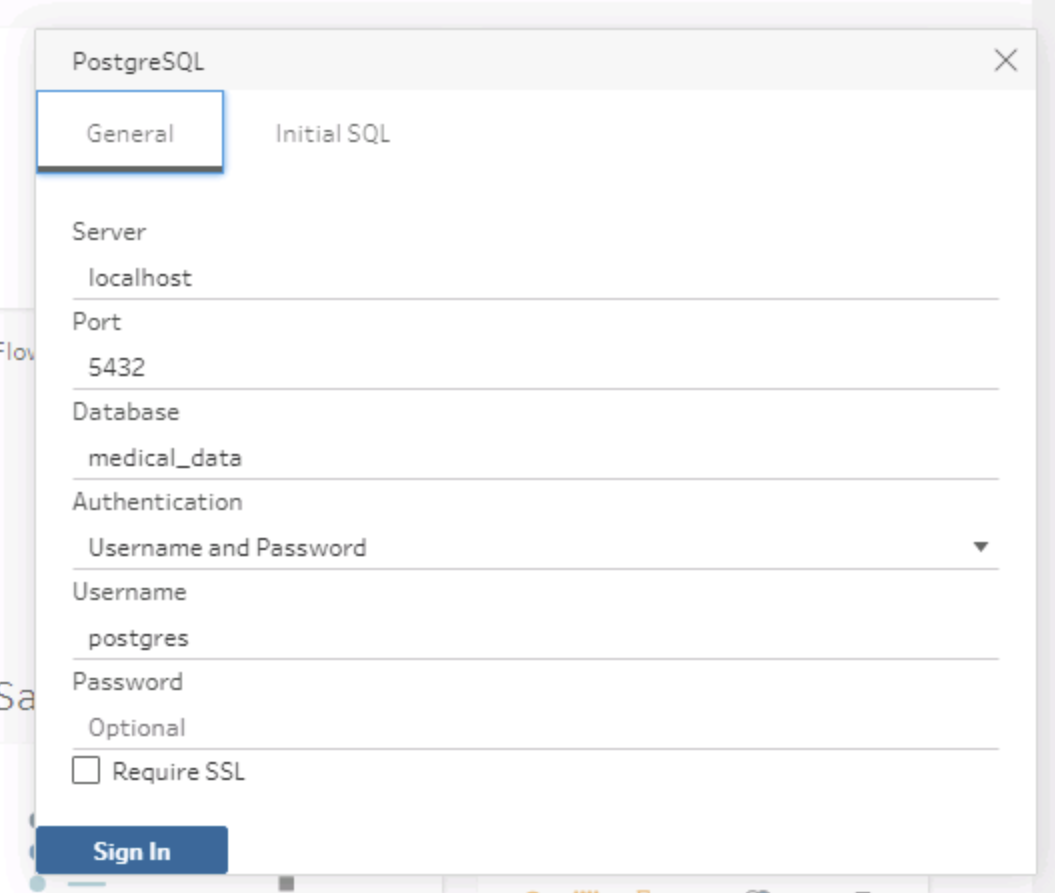
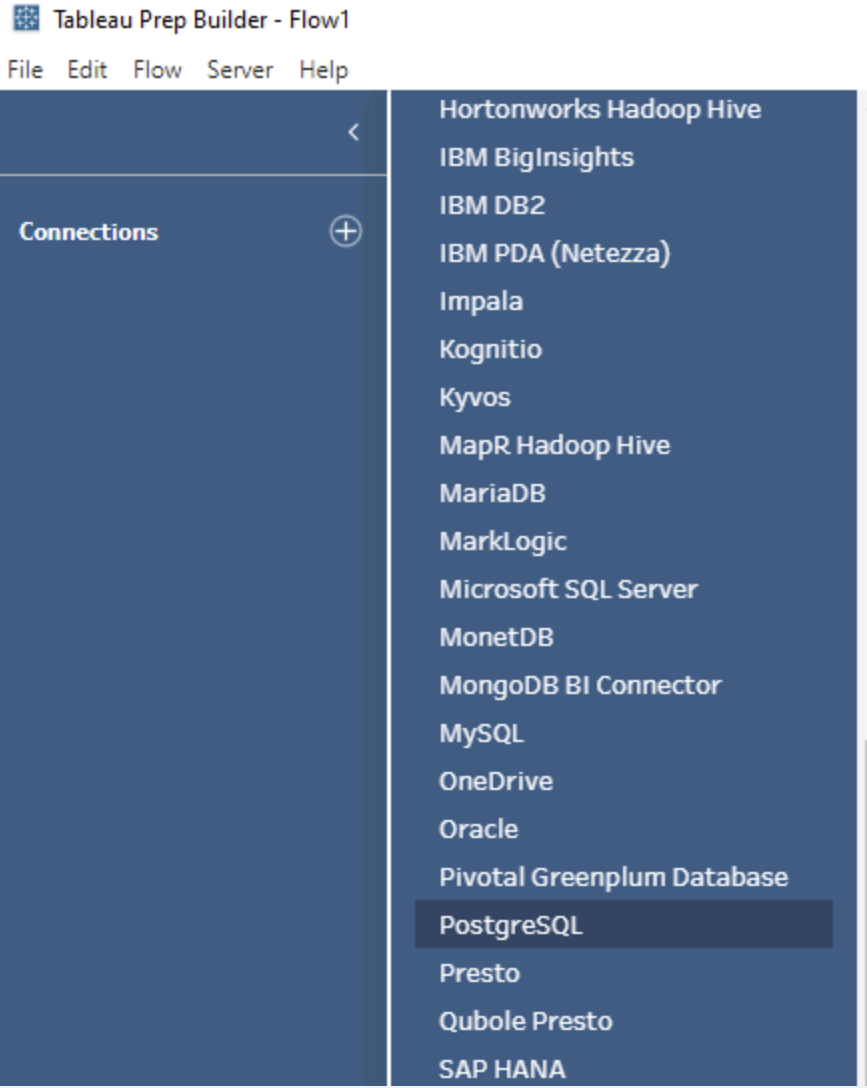
**C2. Justification for Selecting Tableau.**

Tableau was selected for its robust data visualization capabilities, ease of use, and ability to handle large datasets. It allows for:

* Interactive Visualizations: Users can filter and drill into the data to gain deeper insights.
* Real-Time Analysis: Stakeholders can interact with the dashboard and see real-time updates.
* Integration: Seamless integration with various data sources ensures comprehensive data analysis.
* Advanced Analytics: Supports complex calculations and visualizations, making it ideal for healthcare data analysis.

**C3.** The data cleaning and preparation were meticulously carried out using Tableau Prep Builder. Tableau Prep Builder is a great data preparation tool that makes working with source data easy. Its visually appealing user interface and modularized flow-based design make the process not only simple but also enjoyable (<i>Tableau Community Forums</i>, n.d.). Please see the d211\_tableau\_prep\_flow.tflx file in the attached D211\_PA.zip file. The following steps were involved:

* Open Tableau Prep Builder 2021.4. Click Connections, scroll down to PostgreSQL, and click on it.



* Sign in to PostgreSQL:

Server: localhost

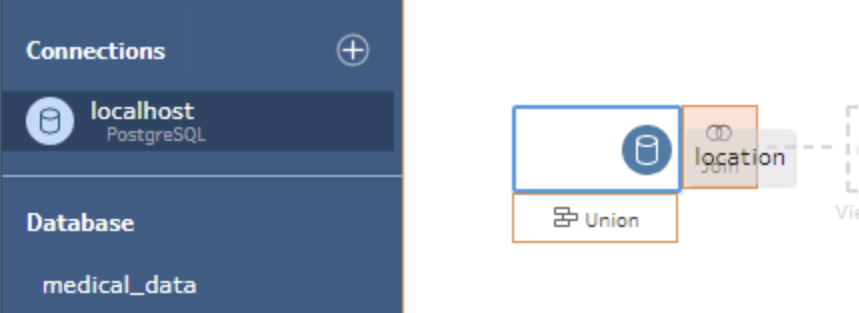
Port: 5432

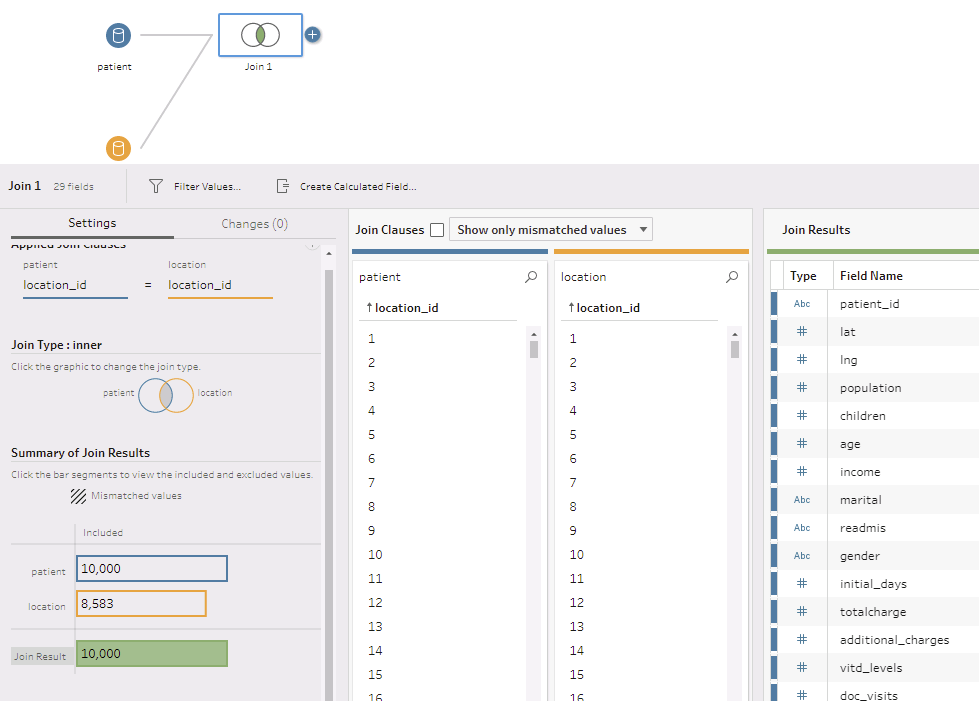
Database: medical\_data

Username: postgres

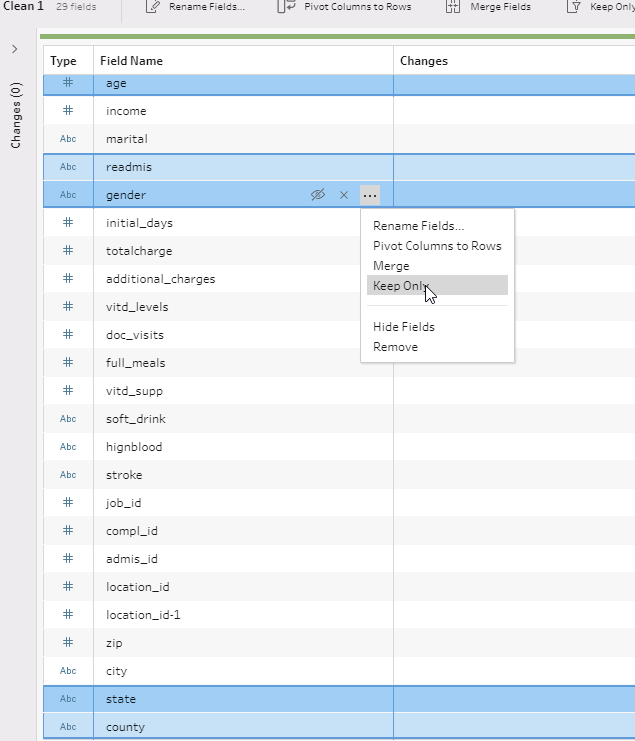
Password: Passw0rd!

* Drag the “patient” tab to Canva and then drag the “location” tab onto the “patient” and perform Join.

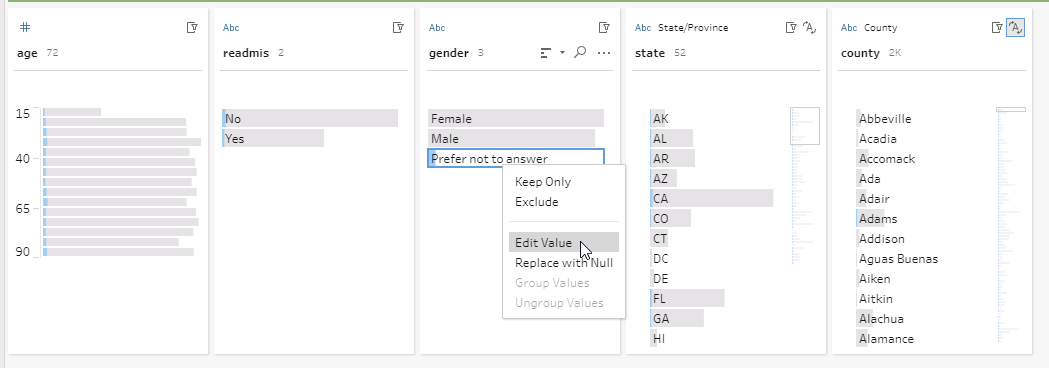




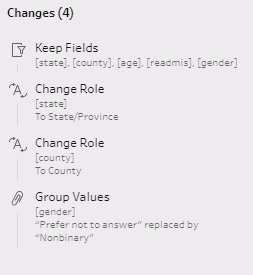
* Keep only fields relevant to dashboard creation: age, gender, readmis, state, and county



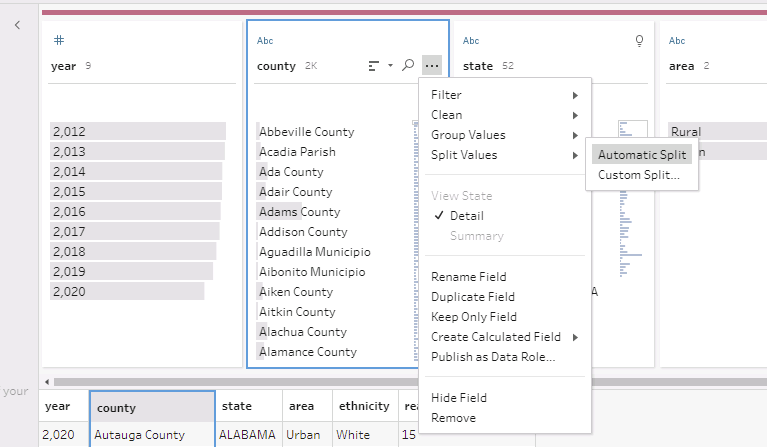
* In the “gender” edit value “Prefer not to answer” to “Nonbinary”



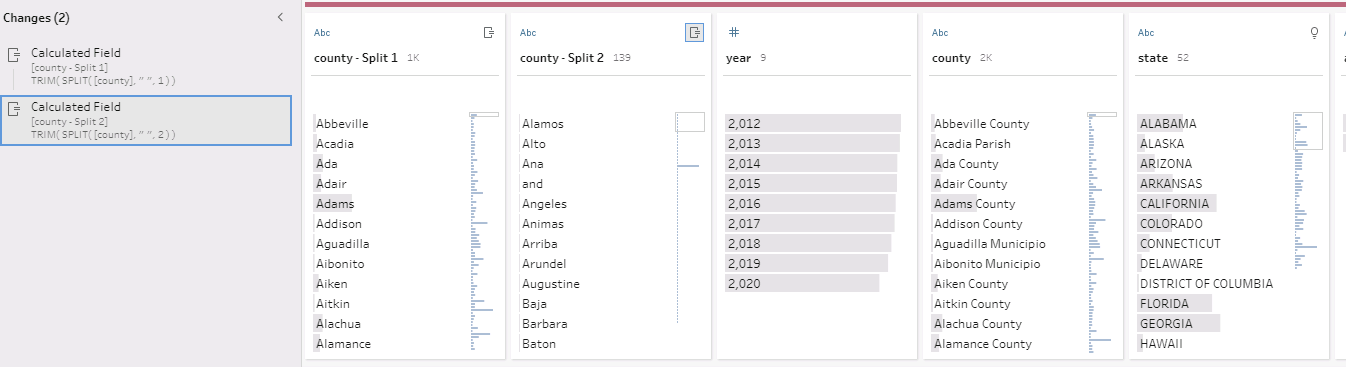
* Change role for “state” to State/Province and for “county” to County
* Please see the changes applied to the WGU medical dataset during the data cleaning and data preparation process in the screenshot below.



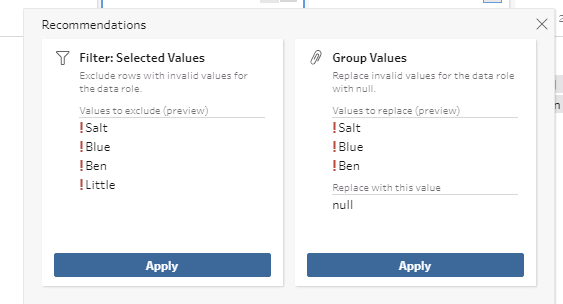
* Drag the “readmission” tab onto Canvas
* Split field “county”



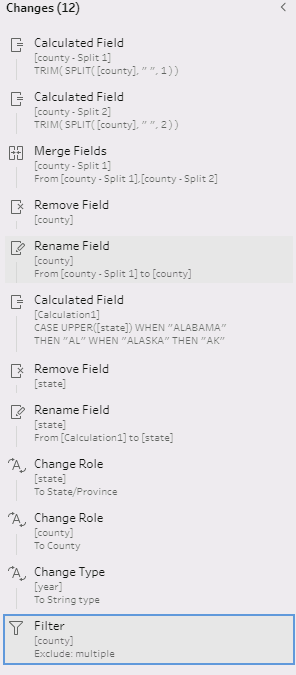
* Merge fields “county- Split 1” and “county – Split 2”, remove field “county”, and rename field “county – Split” to “county”



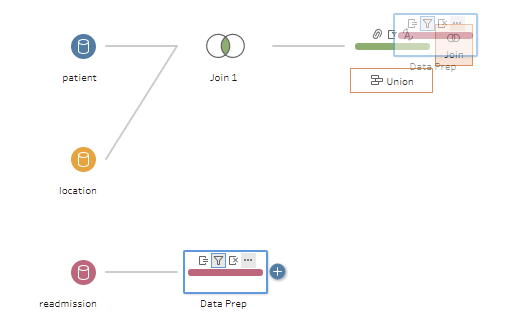
* Create a Calculated Field “Calculation 1” for the state abbreviations
* Remove the field “state”
* Rename the field “Calculation 1” to “sate”
* Change role “state” to State/Province
* Change role “county” to County
* Change type “year” to String
* Exclude rows with invalid values
* Replace invalid values for the data role with null



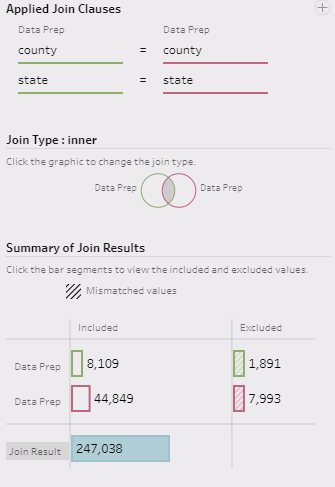
* Please see the changes applied to the external “readmission” dataset during the data cleaning and data preparation process in the screenshot below.



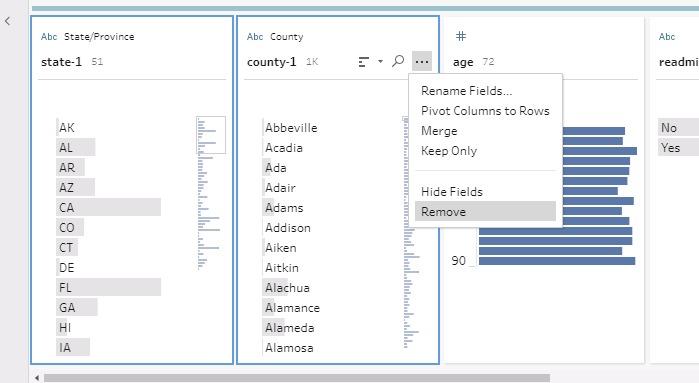
Drag cleaned “readmission” dataset to previously prepared medical dataset and perform Join.



Perform inner Join on “state” and “county”



Drop “county- 1” and “state- 1” fields



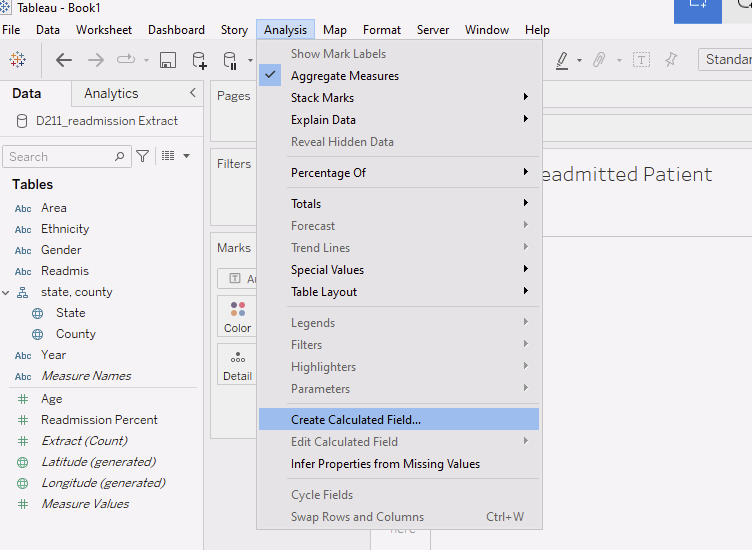
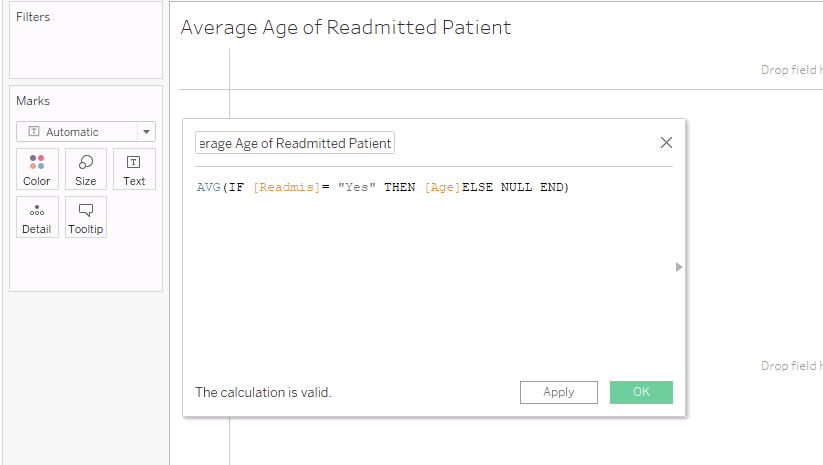
* Create Output “D211\_readmission.hyper” Tableau Prep flow.
* Save “D211\_readmission.hyper”

**C4. Dashboard Creation Steps.**

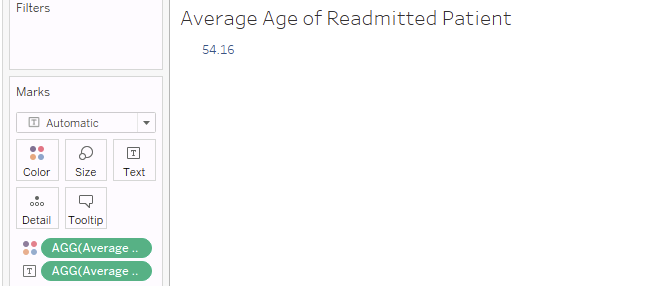
1. Create a table “readmission” in PostgreSQL: import the cms\_readmission.csv file into PostgreSQL.
2. Join the “patient” and “location” tabs from medical\_data dataset in Tableau Prep Builder.
3. Clean and prepare this dataset, keeping only the fields needed for dashboard creation: age, gender, readmission, county, and state.
4. Clean and prepare the external dataset “readmission”: ensure the readmission dataset is cleaned and standardized.
5. Join datasets in Tableau Prep Builder: join the medical dataset with the readmission dataset using “county” and “state” as join clauses.
6. Check the newly created dataset and remove any duplicated fields.
7. Create Output and save it as a d211\_join.csv file.
8. In Tableau 2021.4: open the d211\_join.csv file in Tableau 2021.4.
9. Create sheets with visualizations:

KPI: Average Age of Readmitted Patients:

* + - Create a calculated field Average Age of Readmitted Patients using the calculation: AVG(IF [Readmis] = "Yes" THEN [Age] ELSE NULL END)

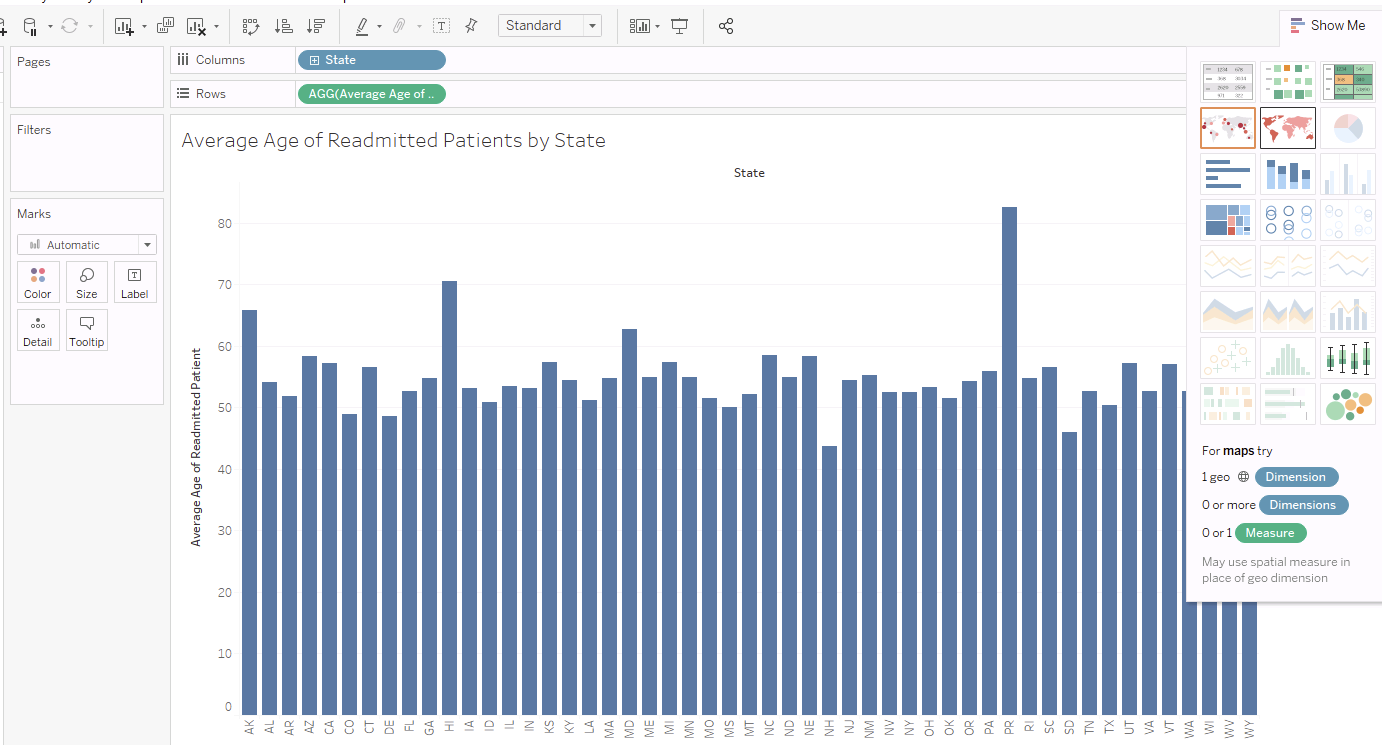
 

* + - Drag this field to the Text and Color labels in the Marks panel.

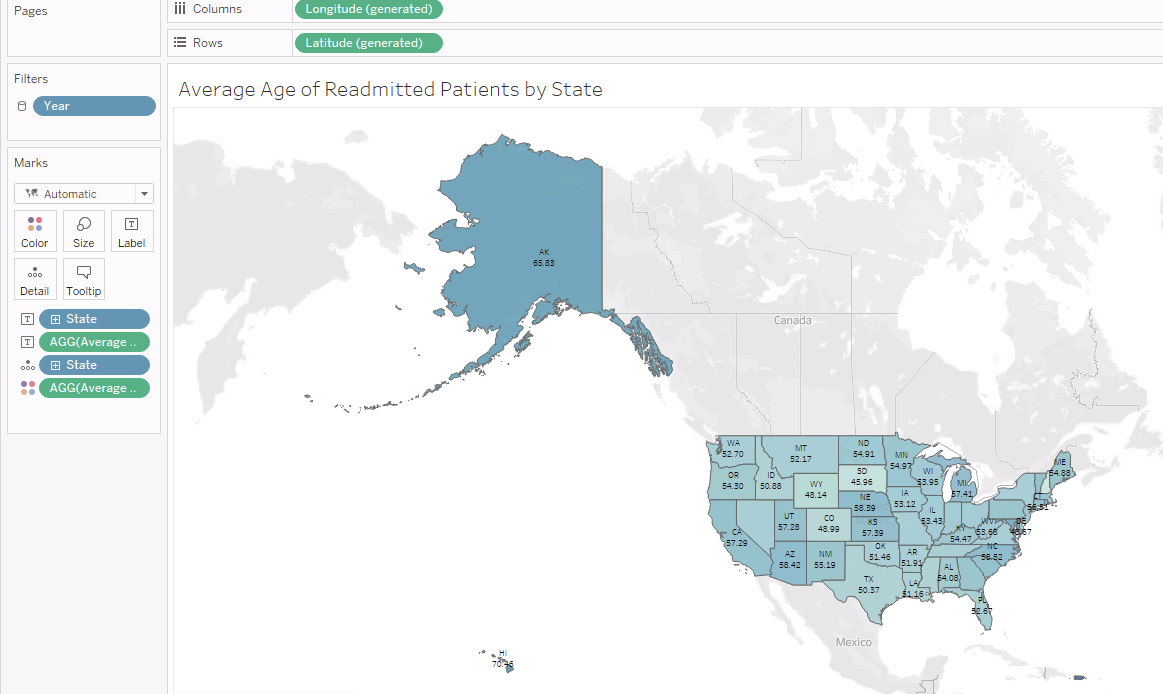


Heatmap: Average Age of Readmitted Patients by State:

* + - Drag State to the Columns shelf.
    - Drag Average Age of Readmitted Patients to the Rows shelf.

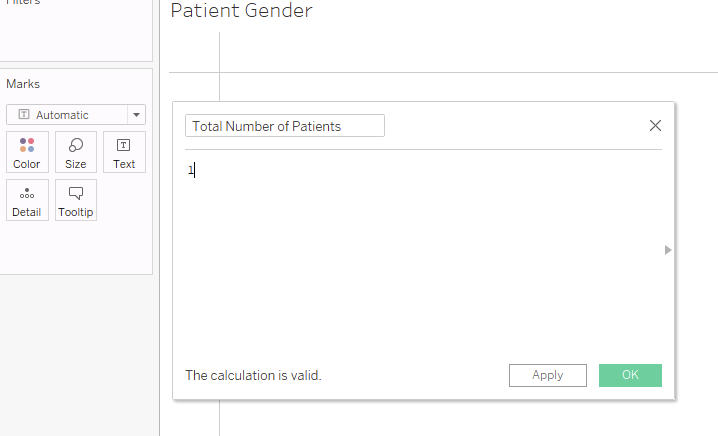


* + - Click "Show Me" and choose "Map".
    - Drag the State and Average Age of Readmitted Patients to the text label.
    - Click on Year and choose "Show filter".

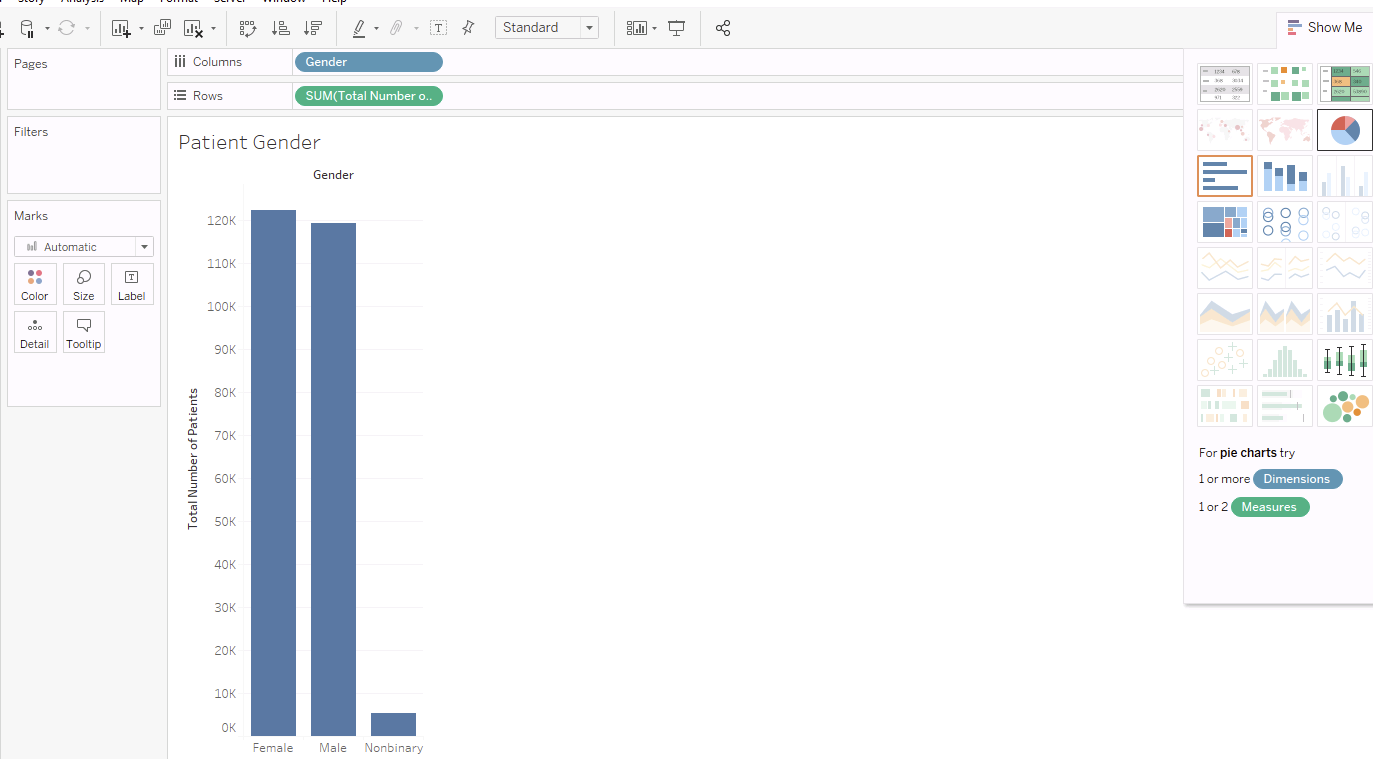


Pie Chart: Gender Distribution of Patients:

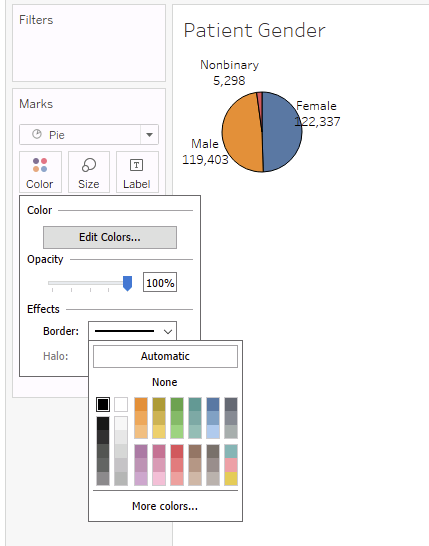
* + - Create a calculated field for the total number of patients using 1 as the calculation.



* + - Drag Gender to the Columns shelf.
    - Drag the Total Number of Patients to the Rows shelf.
    - In "Show Me", choose the pie chart.

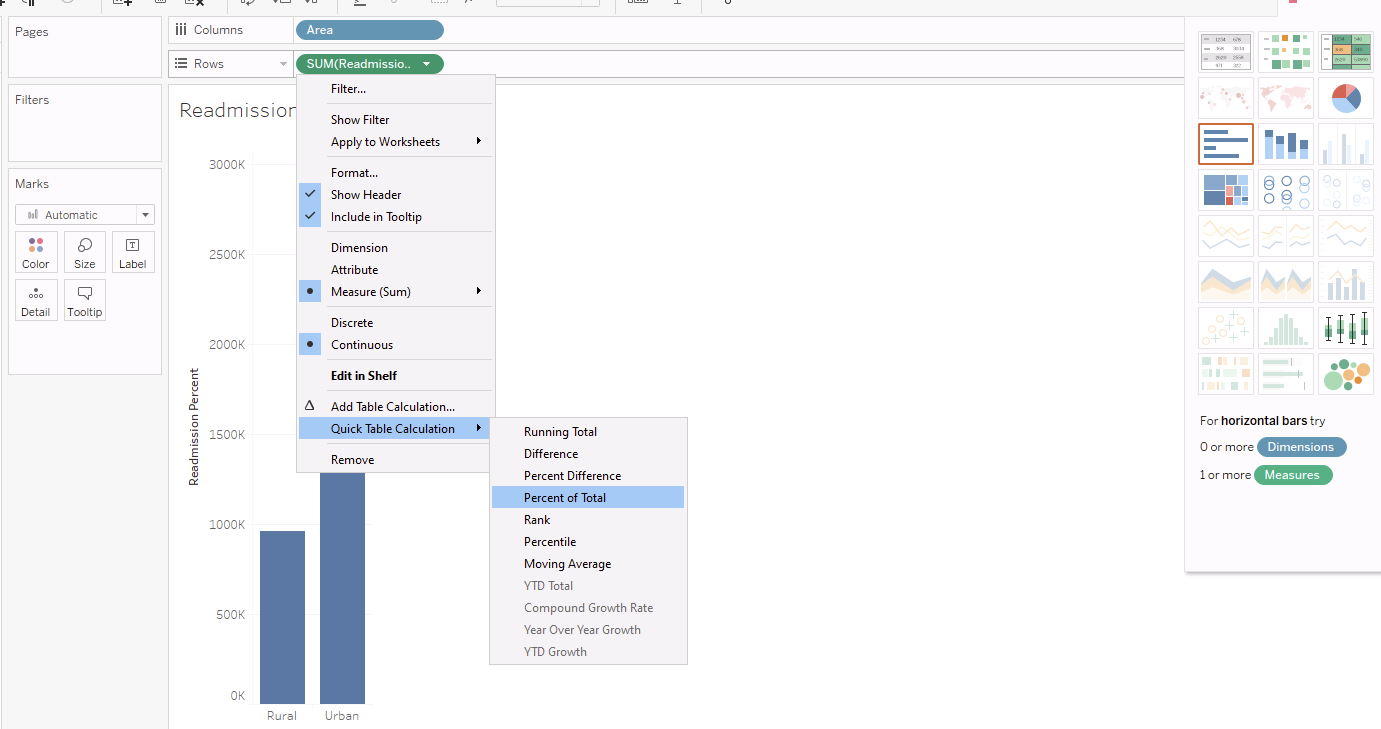


* + - Drag Gender and Total Number of Patients to the text label.
    - Click on "Color" and choose "Edit color," select a color-blind palette, click "Assign palette," and then click "OK,” then click on “Color” again and add border.

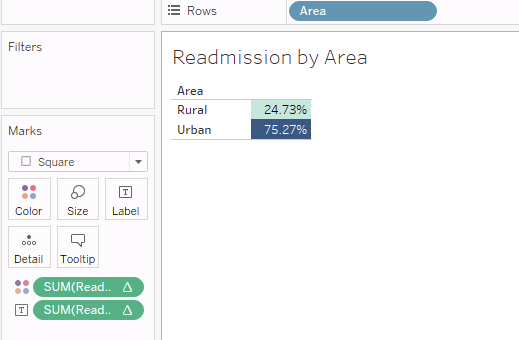


Highlight Table: Readmission Percentage by Area Type:

* + - Drag the Area to the Columns shelf.
    - Drag Percentage of Readmission to the Rows shelf.
    - Click on the arrow on Percentage of Readmission, choose "Quick Table Calculation," and select "Percent of Total."

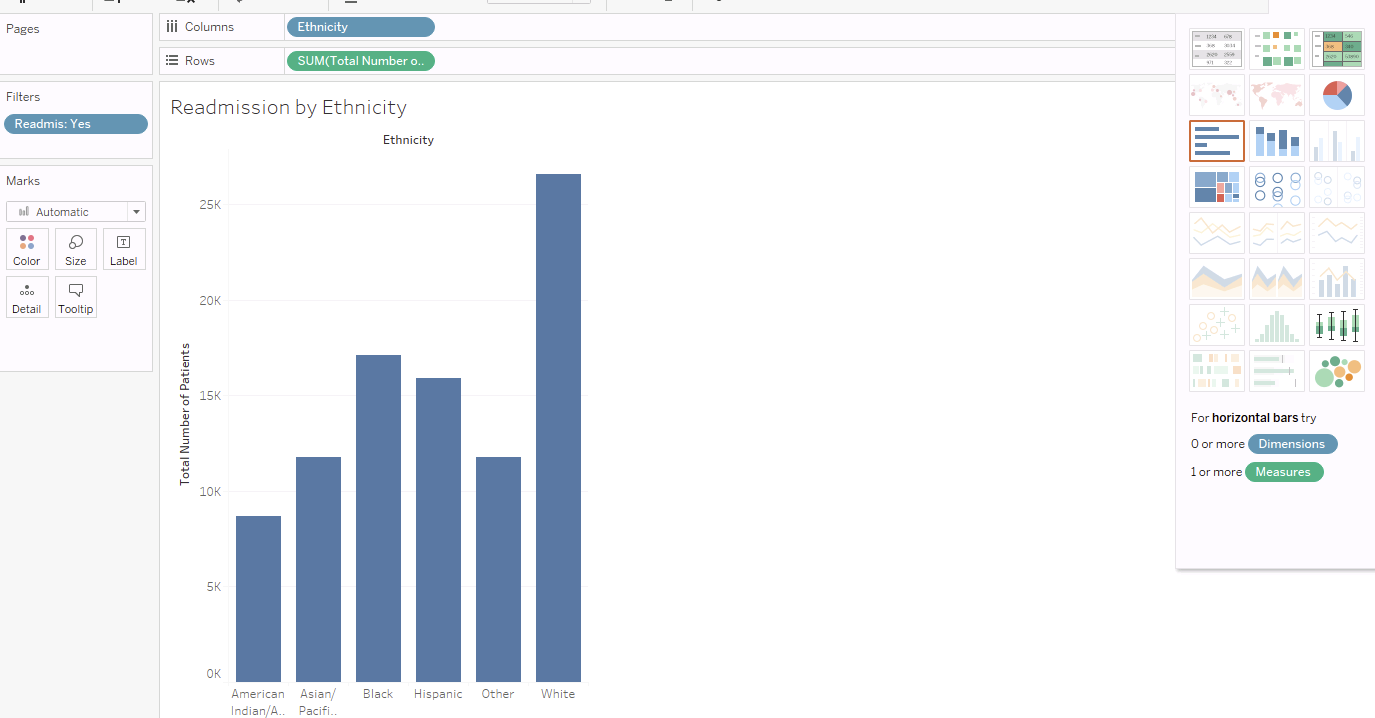


* + - In "Show Me", choose the highlight table.

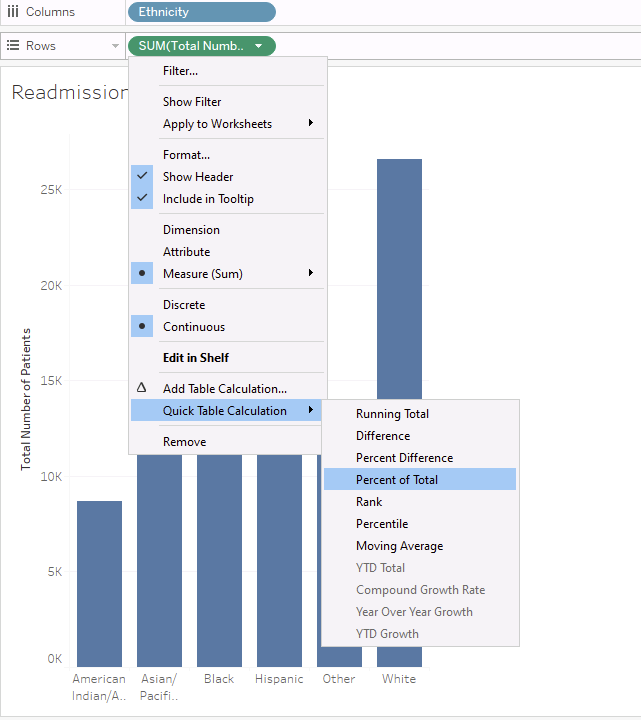


Horizontal Chart: Readmission by Ethnicity:

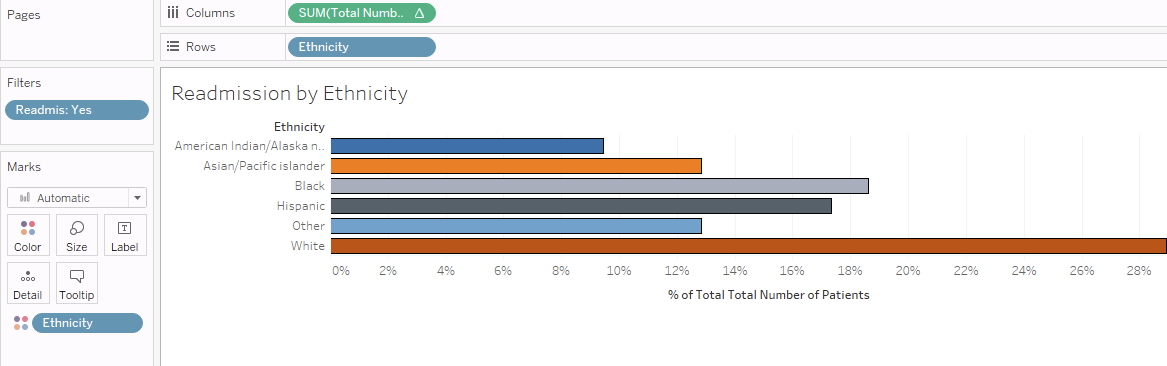
* + - Drag Ethnicity to the Columns shelf.
    - Drag the Total Number of Patients to the Rows shelf.
    - Drag Readmis to the Filters shelf and choose "YES".



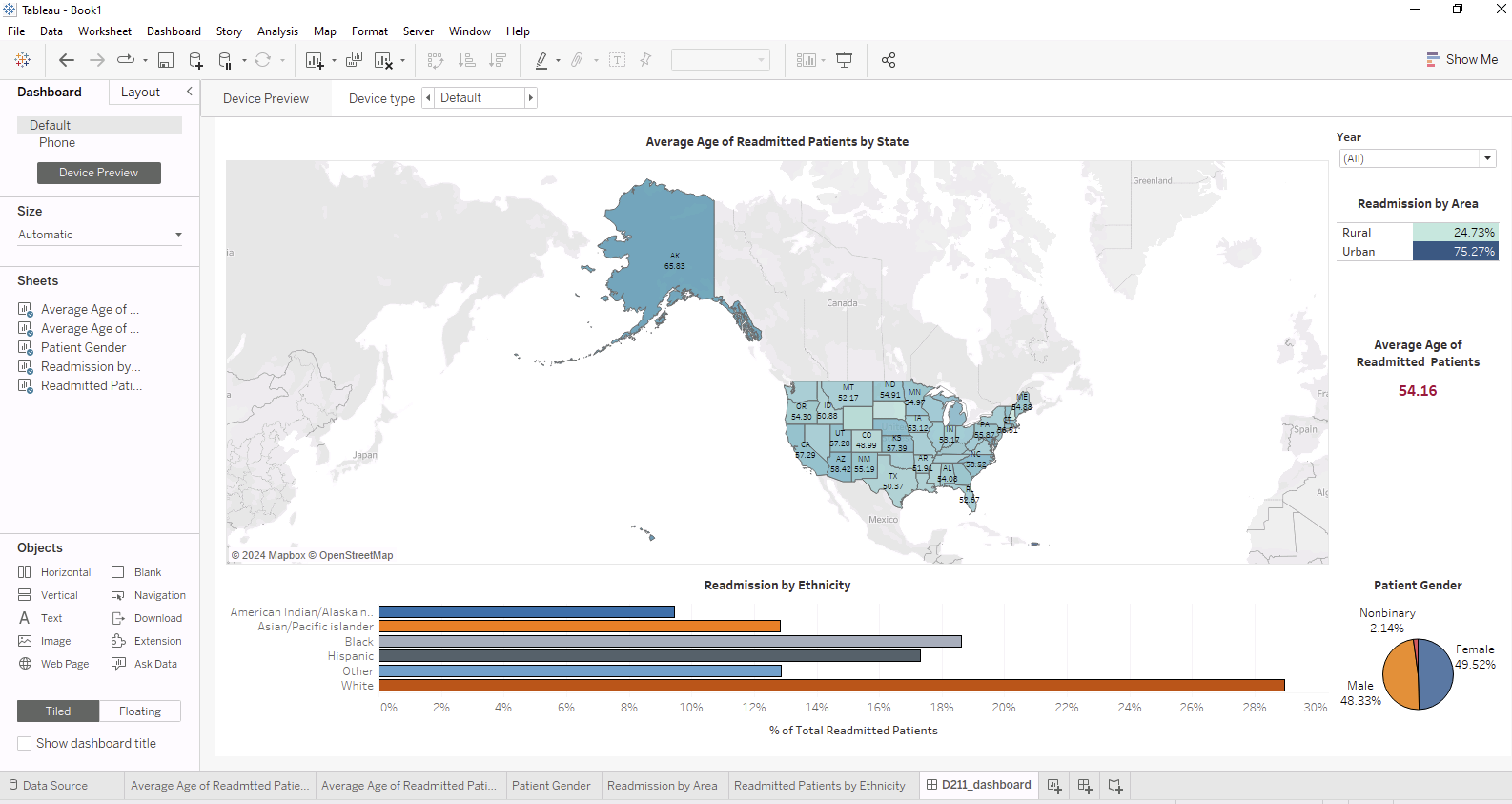
* + - In the Rows shelf, click the arrow on Total Number of Patients, choose "Quick Table Calculation," and select "Percent of Total."



* + - Drag Ethnicity to "Color".
    - Click on "Color," choose "Edit Color," and select a color-blind palette.
    - In "Show Me," choose "Horizontal Bars."



1. Drag and arrange the sheets to form the dashboard and experiment with various layouts.



1. Add Filters: implement a year filter (2012-2020), apply it to all worksheets using this dataset, and enable all visualizations to act as filters for interactive analysis.

**C5. Results and Insights.**

The data analysis revealed several critical findings that align with the purpose and function of the dashboard.

* Geographic Insights: Certain states have notably higher average ages of readmitted patients (HI 70.46 and AK 63.85 vs NH 43.77 and DE 48.67), indicating potential regional trends that require targeted interventions. These insights help stakeholders understand where to focus their efforts geographically to address readmission issues effectively.
* Demographic Patterns: Gender and ethnicity analyses highlighted groups with higher readmission rates. This information is crucial for developing gender-specific and culturally sensitive interventions, ensuring that the hospital's strategies are inclusive and effective for all patient demographics.
* Area-Specific Trends: Urban and rural areas showed differing readmission percentages (Rural 24.73% vs Urban 75.27%), suggesting the need for tailored intervention strategies based on area type. This insight supports creating customized programs that address the unique challenges and needs of urban versus rural healthcare settings.
* The KPI for the average age of readmitted patients provides a quick and essential snapshot of the demographic most affected by readmissions. This metric allows stakeholders to track changes over time and evaluate the effectiveness of age-specific interventions.

These insights provide a clear, data-driven understanding of readmission trends, aiding in strategic decision-making. The dashboard's visualizations enable stakeholders to drill down into specific areas, demographics, and time periods, making it a powerful tool for reducing readmission rates and improving patient outcomes.

**C6. Limitations of Data Analysis.**

The data analysis faced several limitations that should be acknowledged for accurate interpretation of the results:

* Inner Join on County and State: An inner join on the county and state fields excluded some entries from both datasets due to mismatches or missing values. Specifically, the WGU dataset included 8109 entries and excluded 1891 entries, while the external dataset included 44849 entries and excluded 7993 entries. This exclusion may lead to potential biases or gaps in the analysis, as it only considers matched entries and disregards unmatched ones.
* Data Coverage and Completeness: The analysis is based on data from 2012 to 2020. Any trends or changes after 2020 are not captured, potentially affecting the current relevance of the insights.
* Geographic Discrepancies: Variations in data quality and consistency between counties and states could impact the accuracy of the analysis. Some counties may have had more complete or detailed records, leading to uneven representation across regions.
* Demographic Specificity: While the analysis includes age, gender, ethnicity, and area type, other demographic factors (e.g., socioeconomic status, medical history) were not considered, which could provide additional context to the readmission rates.
* External Factors: The analysis does not account for external factors (e.g., policy changes and healthcare innovations) that might have influenced readmission rates during the analyzed period. These factors could play a significant role in the observed trends and should be considered in further studies.
* Use of Aggregate Data: Aggregated data (e.g., percentage of readmissions) can mask individual variations and outliers. A more granular analysis at the patient level might provide deeper insights into the factors driving readmissions.

**D. Code source:**

*CMS hospital readmission race/ethnicity 2012 2020*. (2022, July 8). Kaggle. <https://www.kaggle.com/datasets/patysanchez/cms-hospital-readmission-raceethnicity-2012-2020/data>

*Centers for Medicare & Medicaid Services data*. (n.d.). https://data.cms.gov/tools/mapping-medicare-disparities-by-population

**E. Sources:**

*D211 - Advanced Data Acquisition*. (n.d.). datacamp.com. Retrieved November 1, 2024, from https://app.datacamp.com/learn/custom-tracks/custom-d211-advanced-data-acquisition

Fry, J. (n.d.). *A practical course in data preparation and cleansing using tableau prep* [Video]. Udemy.com. Retrieved October 30, 2024, from <https://wgu.udemy.com/course/tableau_prep/learn/lecture/10447974#overview>

*Tableau Community Forums*. (n.d.). https://community.tableau.com/s/news/a0A8b00002Gy0wvEAB/beginners-guide-to-tableau-prep-top-5-feature-highlights