Bon Secours Project

BUAD 5272: Database Management

Team Tau:
Ethan Hu
Tanner Reichard
Evin Smith
Phoebe Warren

Introduction:

Hospital readmissions cost healthcare providers millions of dollars each year. If a patient is on Medicare and is readmitted within 60 days, the hospital does not get that Medicare claim reimbursed. As a result, it is vital for healthcare providers to figure out the reasons for higher readmission rates. Although the importance of the medical diagnoses cannot be understated when analyzing readmissions, our team decided to focus on the traits of the patients who make up the claims. Specifically, we wanted to see if there were any correlations between patient demographics, readmissions, and months of the year. This line of thought led to our two research questions.

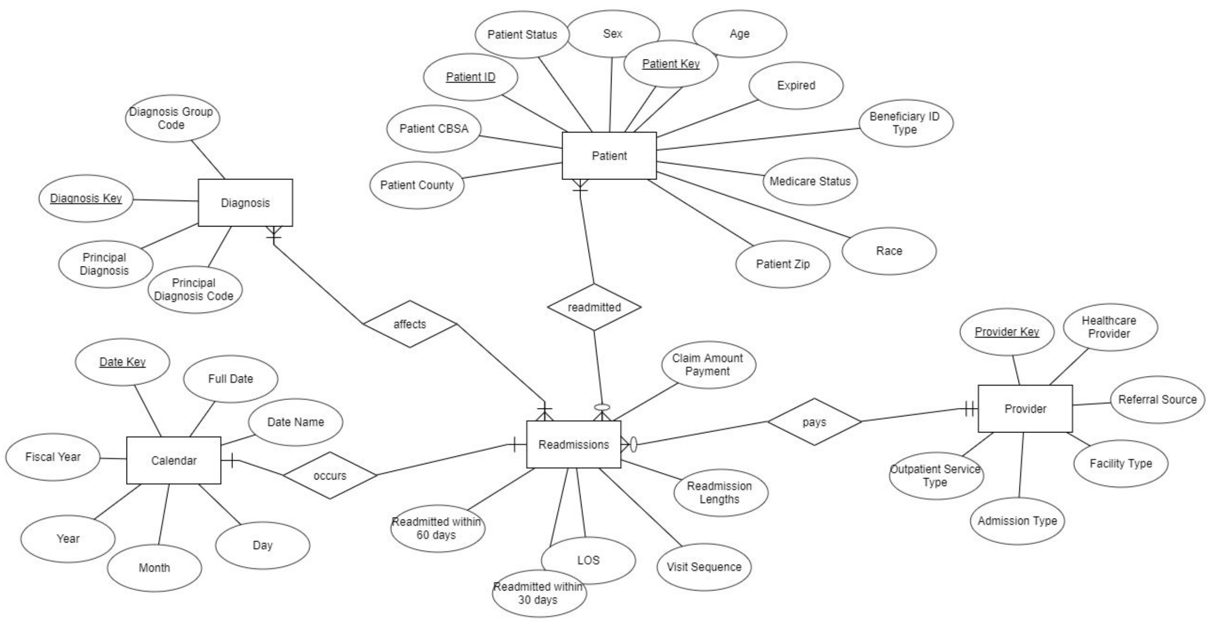
1. Do certain non-diagnosis patient characteristics, such as race, gender, and Medicare beneficiary status, have an impact on either readmission rates or claim amounts?
2. Are readmission rates higher during certain holidays or months of the year?

In order to reach our conclusions, we first had to clean our data and create a database. After reviewing our research questions, we decided that we needed only the high level claims data. We then extracted that data into Alteryx, cleaned it up by removing duplicate entries and incorrect data inputs, and loaded it into a mySQL database. Based on our research questions, we formed ten queries focusing on patient characteristics and timeframe. We then used Tableau to pull all ten of our queries and visualize them.

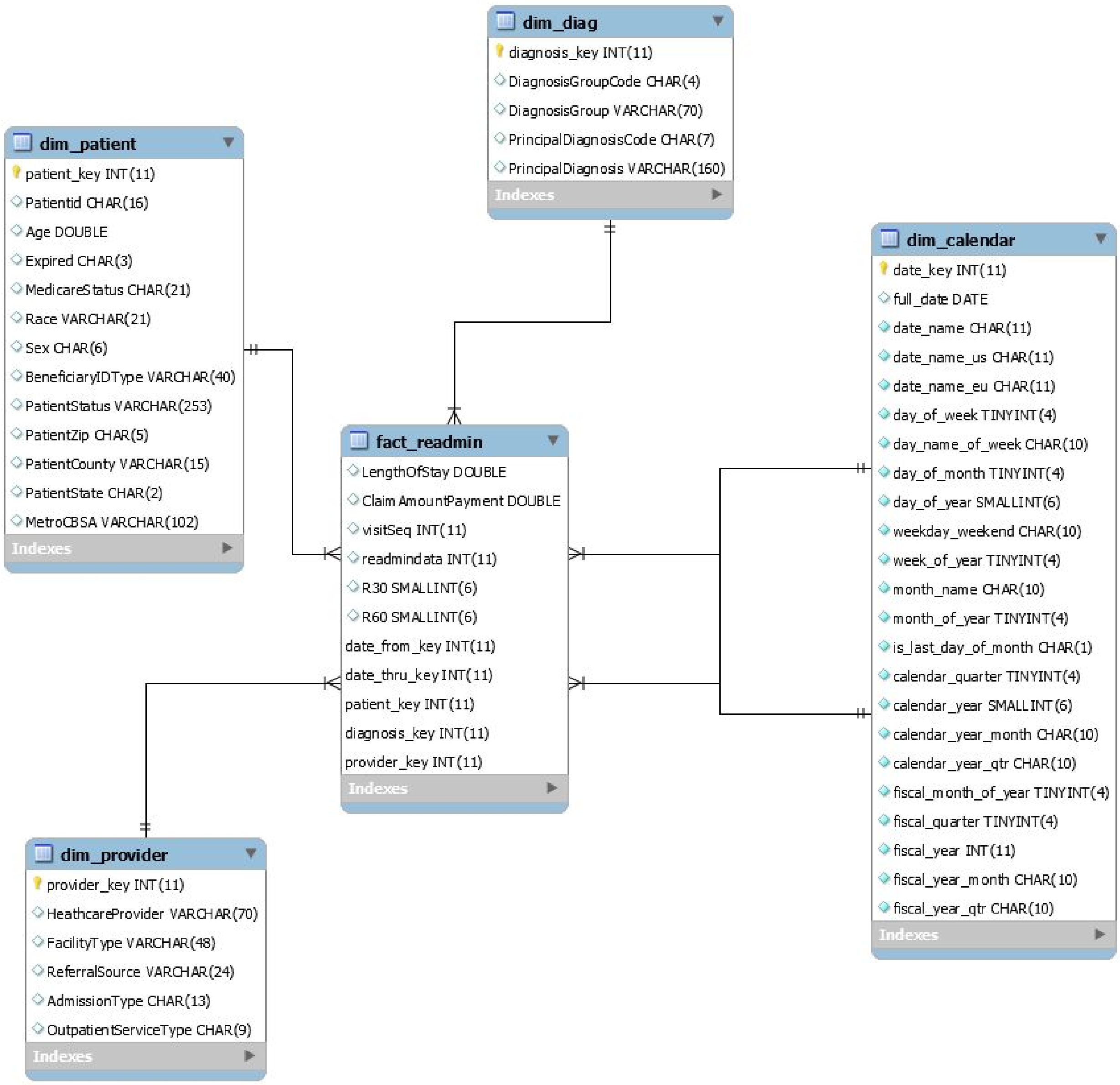
Based on our queries, we discovered that certain characteristics had a bigger impact on readmissions than others. Specifically, Medicare beneficiary ID type and week of the year seemed to be some of the most significant factors.

Modeling Approach:

We decided to focus on the high level claims data rather than the low level data. The high level data contained most of the patient data that we thought was needed for the project. Image 1 displays our conceptual design. We chose to divide the data into five categories: patient, calendar, provider, diagnosis, and a fact table that links them all together. We thought that these divisions would best complement our research questions.

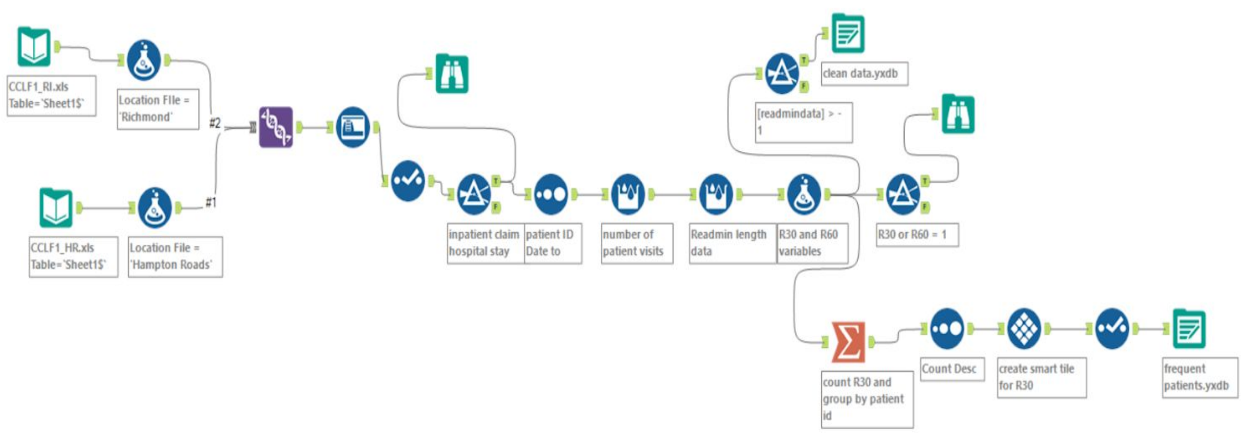


The image below shows our logical design, a star schema. We placed the following in our fact table: the claim amount, length of stay, length between readmissions, a binary variable for being readmitted within 30 days, and another binary variable for being readmitted within 60 days. The dimension tables consisted of a patient table, diagnosis table, provider table, and calendar table. Out of the four dimension tables, we ended up mostly consulting data found in the patient dimension, which was expected given our patient-focused research questions.



ETL Approach:

Our ETL approach started by joining the Richmond and Hampton Roads datasets together in Alteryx. We then figured out what data each column represents, as well as renamed columns to make it more understandable for the user. After filtering the data to only include inpatient admissions, we then sorted by patient and created a visit sequence column that counted the number of times the patient appeared in the data set. If the patient appeared more than once, that means that he/she was readmitted. We then created another column that tracked the days in between hospital visits for patients who were readmitted. After that, we created two more columns: one that tracked readmissions within 30 days and another that tracked readmissions within 60 days. After filtering out readmission data that was negative, we loaded the cleaned data set into mySQL to create our database.

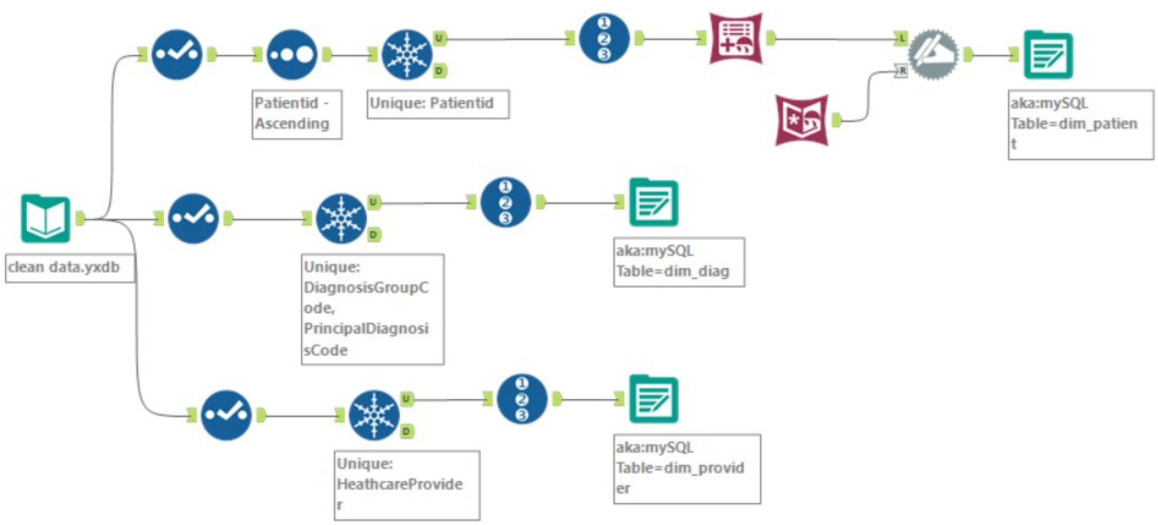


We created our calendar dimension in mySQL and populated it with all the dates of 2016 and 2017. We then created the rest of the dimensions in Alteryx by uploading the cleaned data and selecting fields for either patient, provider, or diagnosis.

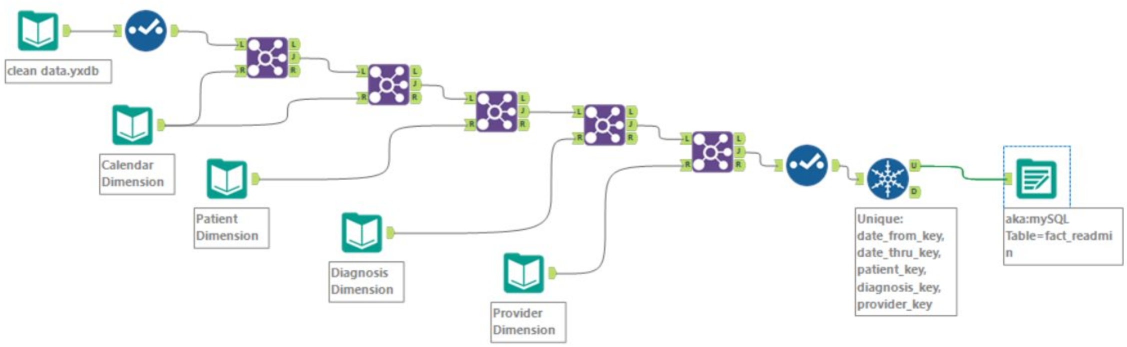
For dim_patient, we chose to include Age, Patient ID, Expired, Medicare Status, Race, Sex, Beneficiary ID Type, Patient Status Patient Zip, Patient County, and Patient State. We then created a patient key based on the unique patient ID and merged the patient data with the Metro Core Based Statistical Area (CBSA) in the US Census.

For dim_diag(diagnosis), we chose Diagnosis Group Code, Diagnosis Group, Principal Diagnosis Code, and Principal Diagnosis. We created a diagnosis key based on the diagnosis group code and the principal diagnosis code.

For Dim-provider, we chose healthcare provider, Facility Type, Referral Source, Admission type, and Outpatient Service Type. We created a provider key based on each healthcare provider name.



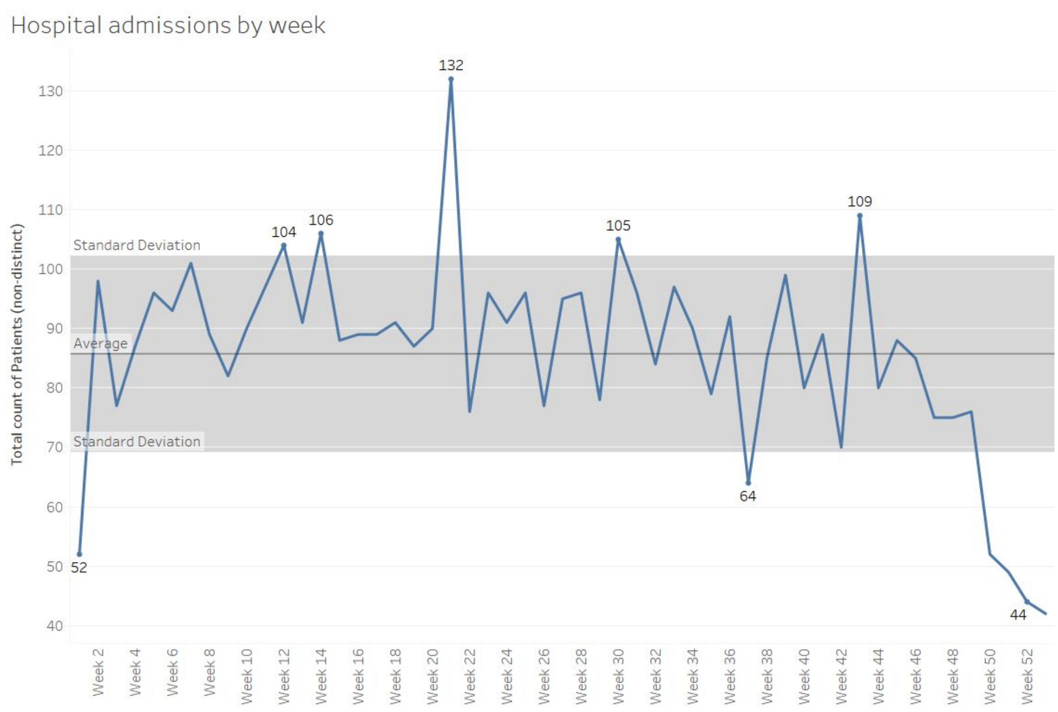
For the fact table, we selected most of the clean data and then joined it with each dimension table. We joined the calendar dimension twice in order to create keys for both date admitted and date released. The last thing we did was to filter out only unique keys in order to remove any duplicates.



Queries and Visualizations:

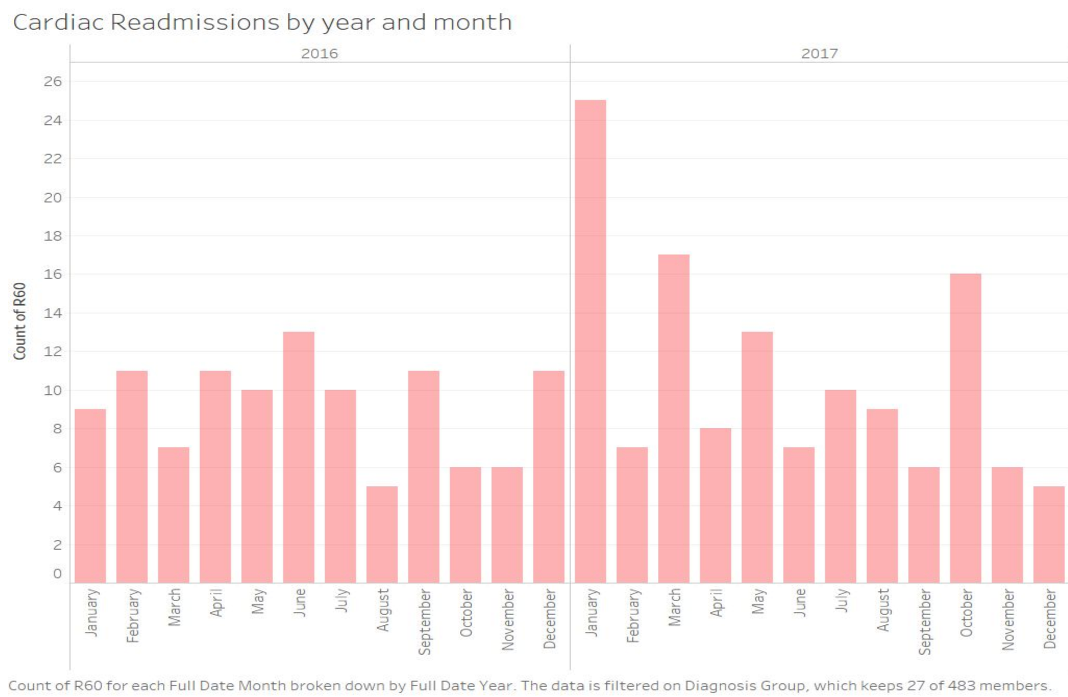
We elected to use Tableau to visualize our cleaned dataset and to identify trends and outliers within different subsets. We used subgrouping to examine potential non-diagnostic traits that would make patients more likely to be readmitted. We also created a week by week timeline to see which weeks out of the year displayed spikes that were more than one standard deviation away from the average readmission amount. We then cross-referenced these weeks with a calendar to see if there were any significant holidays during these timeframes. To simplify our method, we decided that any week containing a holiday would be significant.

As a result, we discovered that the week of Memorial Day contained a higher level of readmissions than any other holiday. The week of Halloween had the second highest amount.

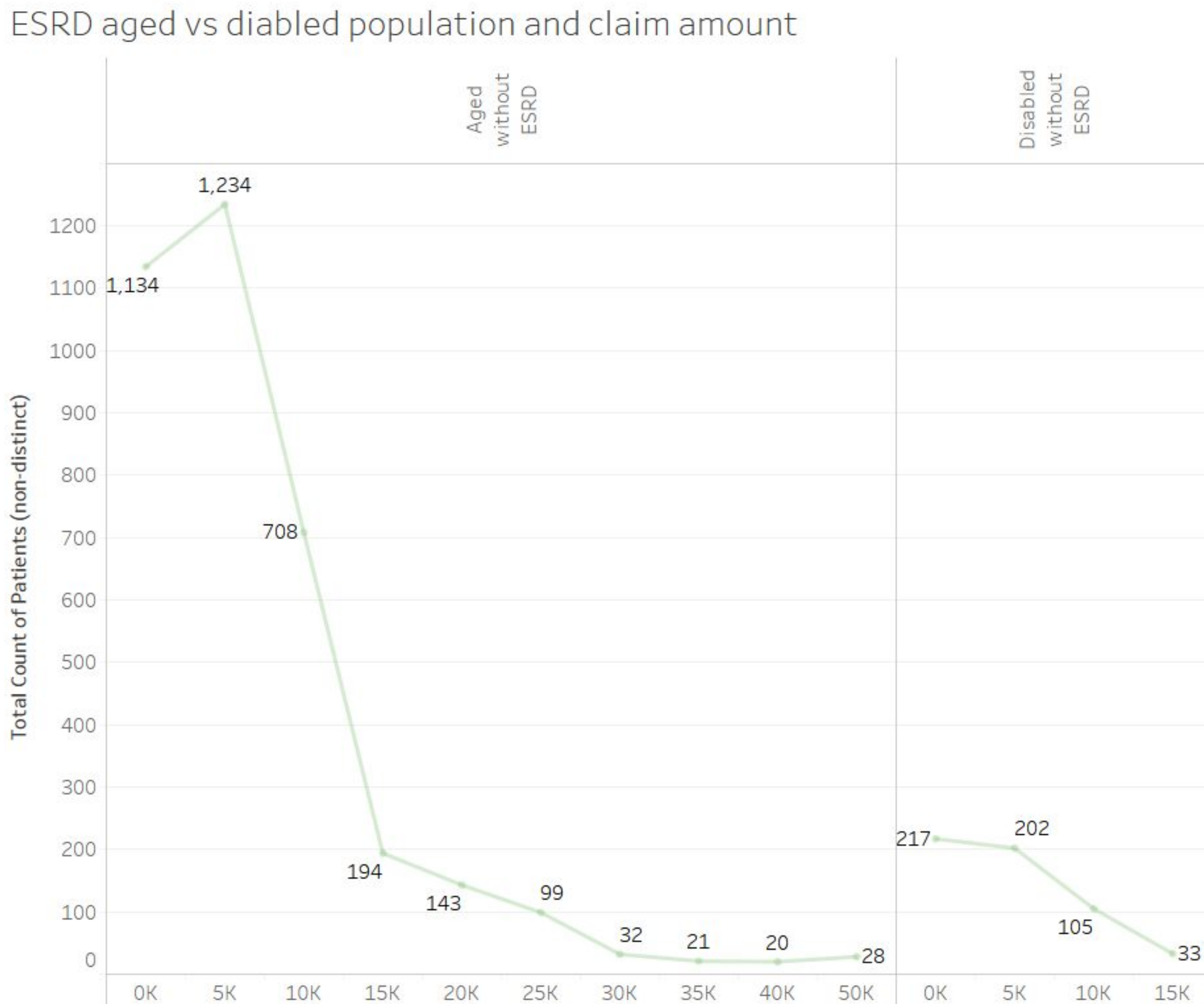


The trend of count of Patientid for Full Date Week.

Because cardiac cases seemed to be one of the most common reasons for hospital admissions, we decided to see if cardiac readmissions within 60 days varied by month. We discovered that January 2017 had the largest amount of cardiac patient readmissions. We filtered on only cardiac related cases.

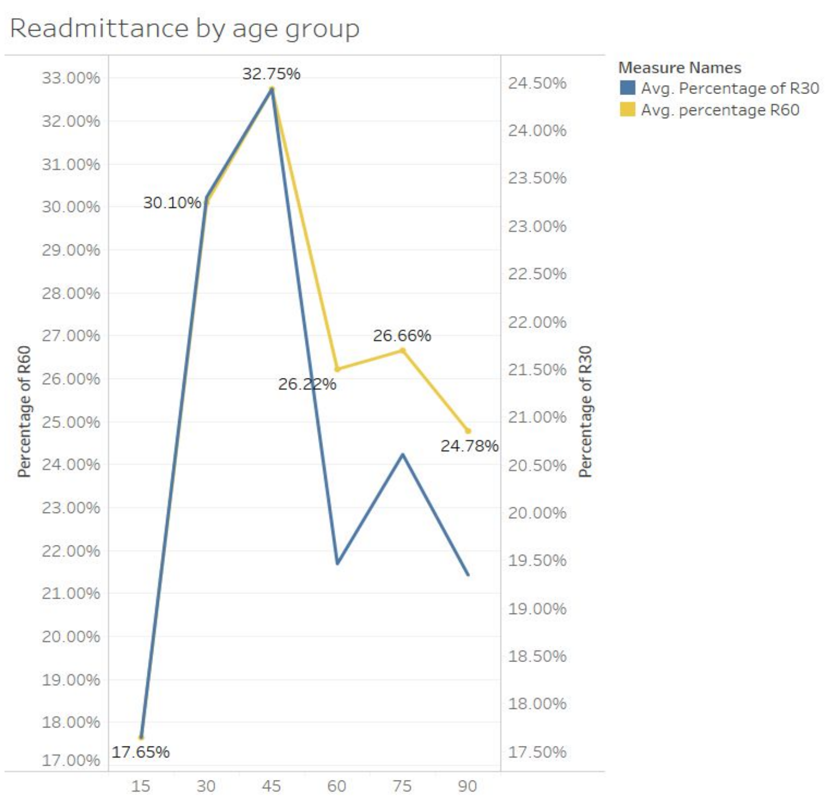


We decided to look at the impact of Medicare status on the claim payment amount. Our findings show that the most common claim amount for patients who were aged without ESRD was around $5,000. We only looked at claims with at least 20 records and only patients with ESRD.



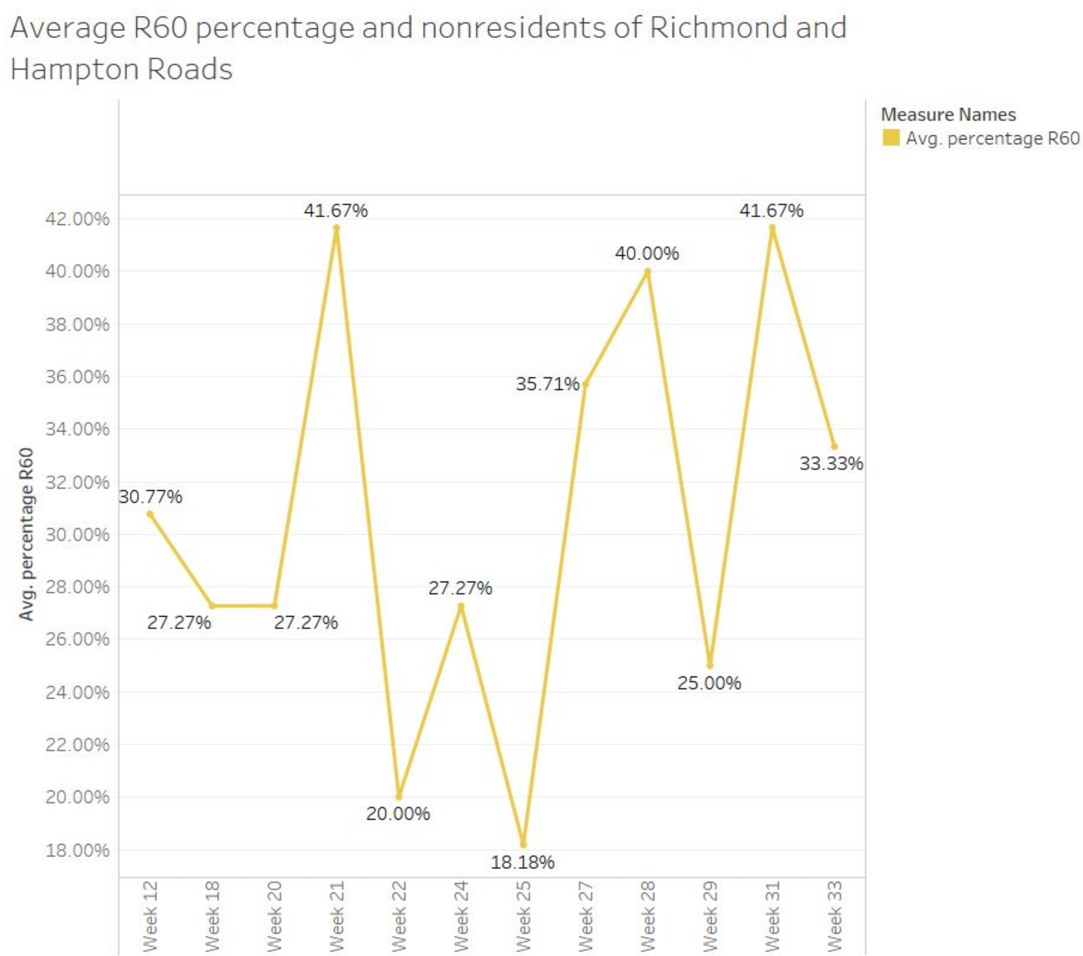
The trend of count of Patientid for ClaimAmountPayment (dim patient) (bin) broken down by Medicare Status. The view is filtered on count of Patientid and Medicare Status. The count of Patientid filter ranges from 20 to 1,274. The Medicare Status filter keeps Aged without ESRD, Disabled without ESRD and ESRD only.

We then decided to sort readmissions within 30 days by age group. Interestingly, we discovered that ages 30 to 45 had the highest readmission rates. Initially, this was confusing to us because Medicare is typically provided to patients over the age of 65. However, upon further investigation, we discovered that those under 65 can receive Medicare if they have a legal disability. If they receive Social Security Disability Insurance payments (SSDI) for more than 24 months, they are eligible for Medicare despite not reaching the age minimum (source). To get the percentage of readmittance, we took the total number of records multiplied by the binary variables R60 and R30, respectively. We then divided this amount by the total number of records and took the average of it when we graphed it.



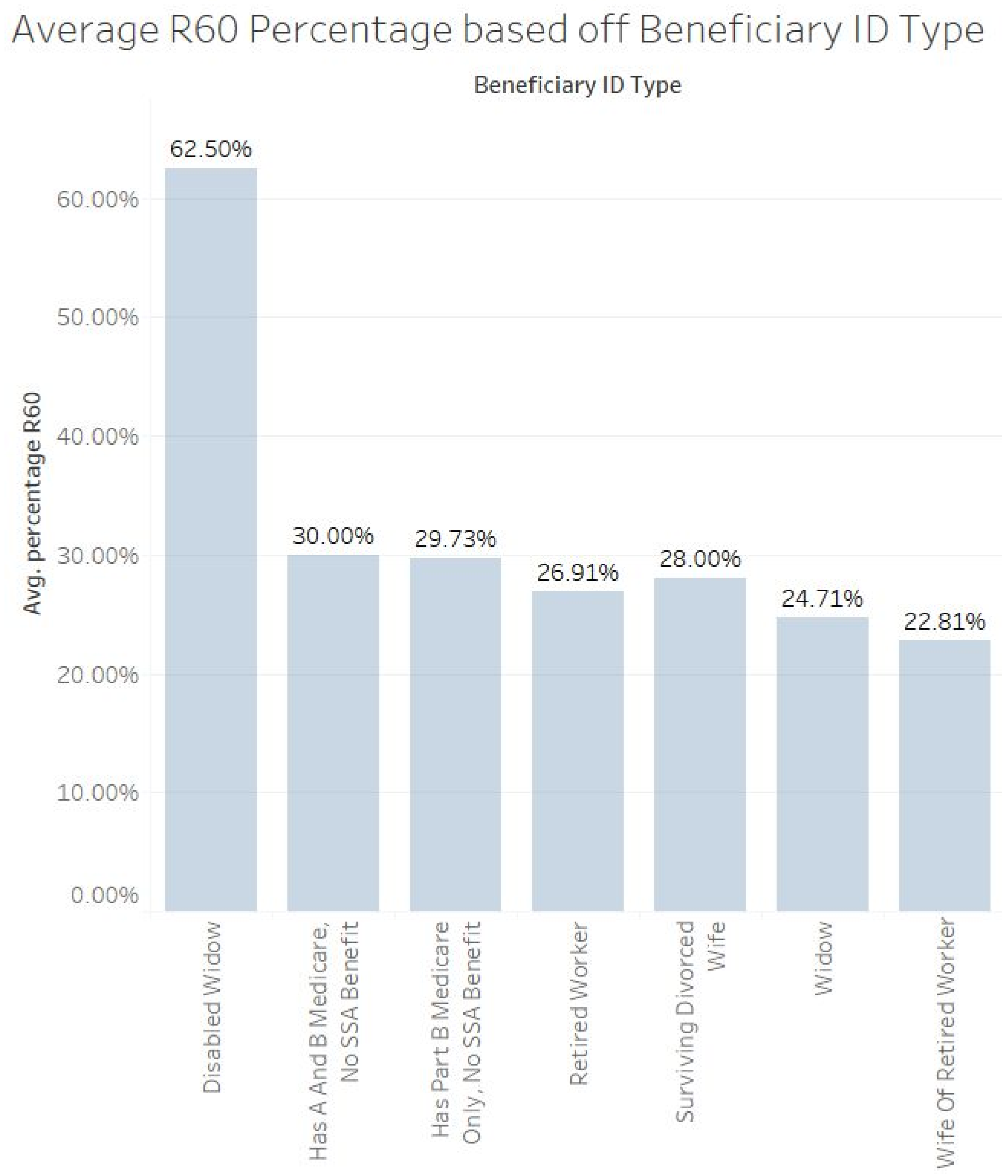
The trends of Avg. percentage R60 and Avg. Percentage of R30 for Age (bin). Color shows details about Avg. percentage R60 and Avg. Percentage of R30.

We then decided to look into R60 rates for people who lived outside of the Richmond and Hampton Roads metropolitan areas. To do this, we uploaded the Metropolitan Core Based Statistical Area (CBSA) from the US Census when we were uploading the data into Alteryx. For this query, we omitted all patients whose zip codes were located in Richmond and Hampton Roads, the locations of the two Bon Secours hospitals. We then deemed the other patients as "tourists" because they did not reside in either of these areas. Overall, we discovered that the data did not show much seasonality for tourists. This was surprising, given that tourists usually travel during the summer months.



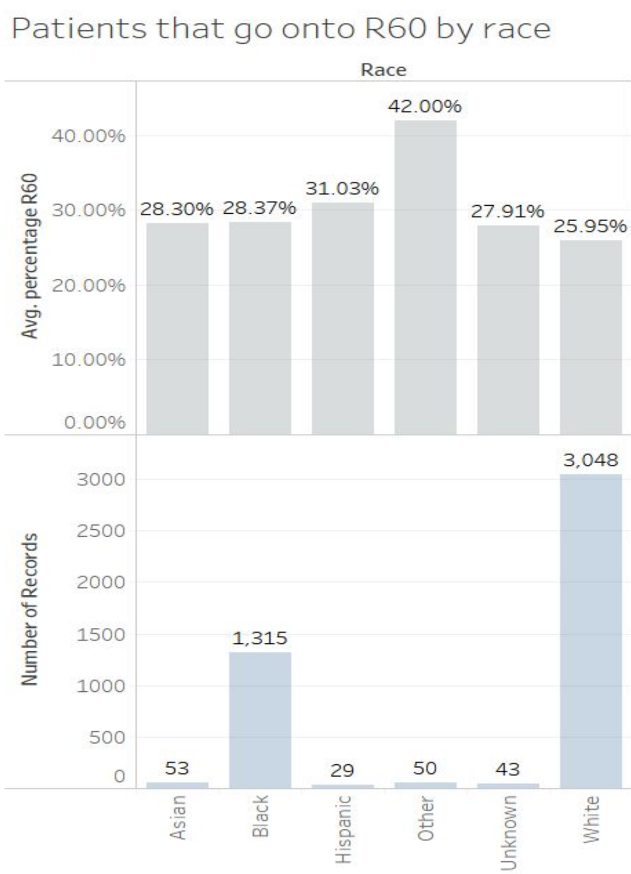
The trend of Avg. percentage R60 for Full Date Week broken down by Name. Color shows details about Avg. percentage R60. The data is filtered on sum of Number of Records, which ranges from 10 to 15. The view is filtered on average of percentage R60 and Name. The average of percentage R60 filter ranges from 1.00% to 66.67%. The Name filter keeps 56 of 45 members.

We also examined R60 rates by Medicare beneficiary ID type. We used the same average percentage calculation as before, and we discovered that about 62.5% of disabled widows get readmitted within 60 days. This statistic is more than double the average.



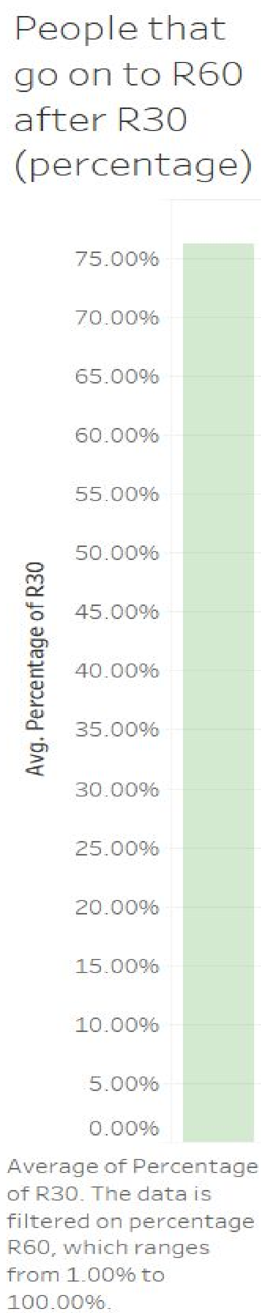
Average of percentage R60 for each Beneficiary ID Type. The data is filtered on sum of Number of Records, which ranges from 10 to 3,879. The view is filtered on Beneficiary ID Type, which keeps 10 of 13 members.

After that, we decided to sort R60 to see average percentage by race. We found out that 42% of those identifying their race as “other” (not solely black, white, Hispanic, Asian, or Native American) are readmitted within 60 days, which is above average. However, when the number of unique patients is displayed, there are only 50 patients in the “other” race category. On the subject of the higher rate of the “other” category, our group speculated that this discrepancy may exist from population distributions within the United States. Indeed, according to US census data, North Virginia has a higher than average population of people identifying as multiracial--which fits in our “other” category-- than the rest of the country (source).



Average of percentage R60 and sum of Number of Records for each Race. The view is filtered on Race and sum of Number of Records. The Race filter keeps 6 of 6 members. The sum of Number of Records filter ranges from 29 to 3,048.

After examining the R60 data, we wanted to see how many patients are readmitted within 60 days given that they were readmitted within 30 days of the original appointment. We discovered that 76.17% of patients readmitted within 30 days are readmitted again within 60 days.



Our last query was to see the counties with the most admissions. Not surprisingly, Henrico County, which is located near some of the Bon Secours hospitals in Richmond, Virginia. We then realized that the Hampton Roads data contained incomplete county names, which Tableau did not register. As a result, the map mainly shows only Richmond counties.