Observing the Effect of Various Indicators on Hospital Readmission Rates

Database Management

December 11, 2018

Table of Contents

Introduction	2
Modeling Approach	3
ETL Approach	3
Queries & Visualization	5
Research Question 1	6
Research Question 2	
Research Question 3	
Conclusions	16
Appendices	17
Appendix A: Relational Schema	
Appendix B: Cleaning the Data	
Appendix C: ETL Dimensions	
Appendix D: ETL Fact Table	
Appendix E: Combining the Databases	

Introduction

Utilizing data provided to us by medical company Bon Secours, we aimed to see if we could observe if any factors or grouping of factors have any effect on patient readmission rates in these hospitals. We had two main questions regarding these readmission rates that we aimed to answer: (1) are readmission rates affected by marital status, (2) are readmission rates affected by patient gender, and (3) are readmission rates affected by admission types.

After observing the data provided to us by Bon Secours, we interpreted the data to conclude that while patient age has a noticeable correlation with hospital readmission rates, it is more difficult to measure the effect of marital status on readmission rates.

We prepared the data for proper analysis using a multitude of methods. Firstly, we constructed an ERD diagram in order to fundamentally understand how the database function. We then drew a star schema draf to help us map the data in an ETL model later on. Next, we constructed tables using Alteryx and MySQL in order to connect the data appropriately and form the foundation on the database.

Using the SQL tables, we then brought our database to life using an ETL model built in Alteryx, and additionally added U.S. Census data as well in order to be able to better observe the effect of certain indicators on readmission rate. Once our database was built, we queried some of the data in MySQL and directly observed the relationship between certain factors and readmission rates. Lastly, we visualized our tables in Tableau in order to more easily confirm our initial observations.

Ultimately, we came to the conclusion that individuals who were either divorced or widowed, were readmitted to the hospital within 30 days in greater numbers than those who were married with the same disease. Additionally, we found that this widowed and divorced group had a longer length of stay, on average, and, the average patient age for this group was younger.

Addressing our question on patient gender, we found that for patients that when patients were grouped by disease type, males tended to be readmitted to the hospital far more likely than females with the same disease, on average.

We also discovered that a patient's admission type, whether they came in electively or during an emergency, is a good indicator of how cautious they are about their health, and how they are more likely to visit their primary care physician, be readmitted, stay longer in the hospital, but also less likely to expire.

Modeling Approach

Our star schema is shown in Appendix A: Relational Schema. Since we need to study factors affecting readmission within thirty days, we decided to build a analytical database (data warehouse) regarding to readmission. In every data warehouse, dimension tables and fact tables are involved. In this study, our fact table contains the readmission data, our subject of analysis. Each record shows a inpatient claim and relevant measurable data about readmission (details are explored in the ETL Approach section). For dimension tables, we include descriptions of location, patient information, diagnosis-related group, provider, admission, and time frame, to which readmission claim belongs.

ETL Approach

ETL stands for extraction-transformation-load infrastructure, including the tasks extracting useful data from operational data sources, transforming such data to conform to the structure of target data warehouse, and loading the transformed and quality assured data into the data warehouse.

Data source:

1. Five original data files from *: two high-level files and three detailed files for both from Richmond area and Hampton Roads area.
2. 2010 the United States Census

Extraction

In this step, we have already obtained the data from operational data sources, provided by Bon Secours. The data are from five original data files: two high-level files and three detailed files for both from Richmond area and Hampton Roads area, containing the relevant claims from 2016 to 2017.

Transformation

There are four main steps for us to implement the ETL:

1. Merge two high-level data files into a database and clean it by Alteryx.
2. Build dimensions that summarized from high-level data for this analytical database regarding to readmission by Alteryx and MySQL.
3. Build the basic fact table by Alteryx.

4. Supplement this fact table by principal care physician (PCP) related data extracted from the three detailed data files.

First of all, we merged the two high-level data files into a database and cleaned it because the they have the same types of information in the same structure. This is conducted in a single file named “clean.yxmd”, showing in Appendix B. We firstly renamed the variables, automatically changed the data type, created bins for lengths of stay, and deleted duplicate records. Next, We selected the inpatient claims as our target and add some relevant columns for analysis purpose. For example, we set a new column called “R30” to show whether this claim is a readmission within 30 days (1 for readmission within 30 days; 0 for else). In addition, we built another column called “R30” to show whether this claim is a readmission both within 30 days and with the same diagnosis as the previous one. The reason why we added those columns is that our main analysis is about readmission.

Second, we built the dimensions for readmission analytical database, including location dimension, patient dimension, diagnosis-related group dimension, provider dimension, admission dimension, and calendar dimension. This is conducted in a single file named “dimension.yxmd”, showing in Appendix C. For location dimension, it contains unique patient zip codes and the related information extracted from US Census 2010, including the %American Indian and Alaska Native alone percentage, Asian alone percentage, Black or African American alone percentage, White alone percentage, percentage of total female, percentage of total male, median age of total population, and total population for each zip code area. For patient dimension, it contains each patient’s patient gender, age, bins for age, expired date, death date, medicare status, visit number (how many time he or she was admitted by this hospital), bins for visit number, patient zip, patient county, patient state, average length of stay, number of times readmitted within 30 days, number of time readmitted within 30 days with the same diagnosis, bins for readmission number, the number of other types of claims he or she had during this period, and the bin for that. For diagnosis-related group(DRG) dimension, it contains both the DRG description and DRG codes. For provider dimension, it contains the provider names (centers or hospitals). For admission dimension, it shows the admission source and the admission type (elective, emergency, or urgent). For calendar dimension, it enables us to do time related analysis such as when this readmission happened, what were the starting and ending date of this claim, and in which time pattern does readmission happen more often. Specifically, only this dimension is built by existing SQL files rather than on Alteryx.

Third, we built the basic fact table by Alteryx. This is also conducted in a single file named “facttable1.yxmd”, showing in Appendix D. In this file, we selected useful data that should be included in fact table, including length of stay, payment, visit sequence for the patient incurring this claim, whether this claim is a readmission within 30 days, whether it’s because of the same diagnosis, and how many times the patient of the claim go to see his or her PCP in last six months (with the same diagnosis).

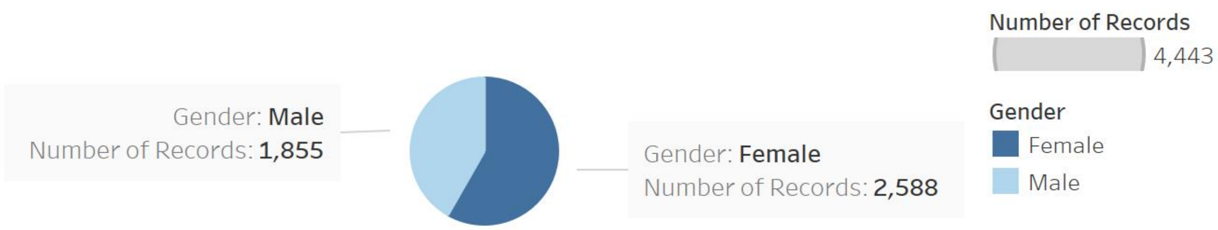
Finally, we supplemented this fact table by principal care physician (PCP) related data extracted from the three detailed data files in “facttable2.yxmd”, showing in Appendix E. The key step here is to use a “Append Fields” button in Alteryx to append the existing factable table (claim records) with PCP patients data transformed from the three detailed files, choosing “allows all the appends” rather than “allows for only 16 records”. Finally, we successfully add two columns, PCP6M and PCP6MSameDiagnosis, in the fact table to show how many times the patient of a claim went to see his or her principal care physician in the last 6 months, and whether they are under same diagnosis.

Loading
All the loading work is done by Alteryx.

Queries & Visualization

Research Question 1:
To analyze how marital status/living situation affected health, we grouped patients by BIC type and looked at statistics about each group. We started by looking at the total numbers for each group, and it became clear how the overwhelming majority of people fell into the retired workers category, and so whenever looking at the total number of a statistic, this group would dominate. We decided to examine the average statistics for each group, and one of the main results from this is the significance of the disabled widow group. Even though this group had the lowest average age (58), it had the highest 30 day readmission rate (0.35) and the longest average length of stay (6.9 days), as well as having a low average number of PCP visits in last six months (0.9). Also of note is the high average number of PCP visits in last six months for Part B Medicare only group (3.3), which is much higher than any other group, although there is nothing else especially noteworthy for that group.

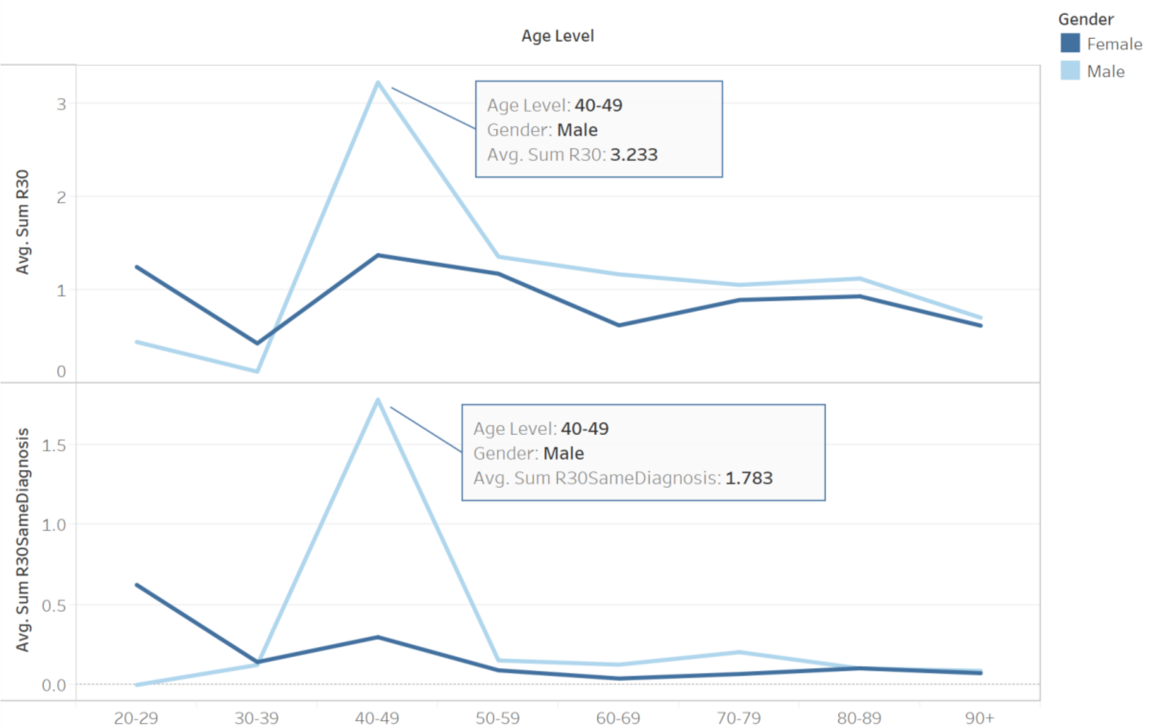
<u>Graph 1: Total number of in-patient claims by gender</u>



Gender (color) and sum of Number of Records (size).

Male patients incurred 1855 in-patient claims while female incurred 2588 admission claims. In other words, about 40% of the in-patient claims are incurred by male patients but about 60% of the in-patient claims are incurred by female.

Graph 2: Average number of Readmission among different age groups by gender



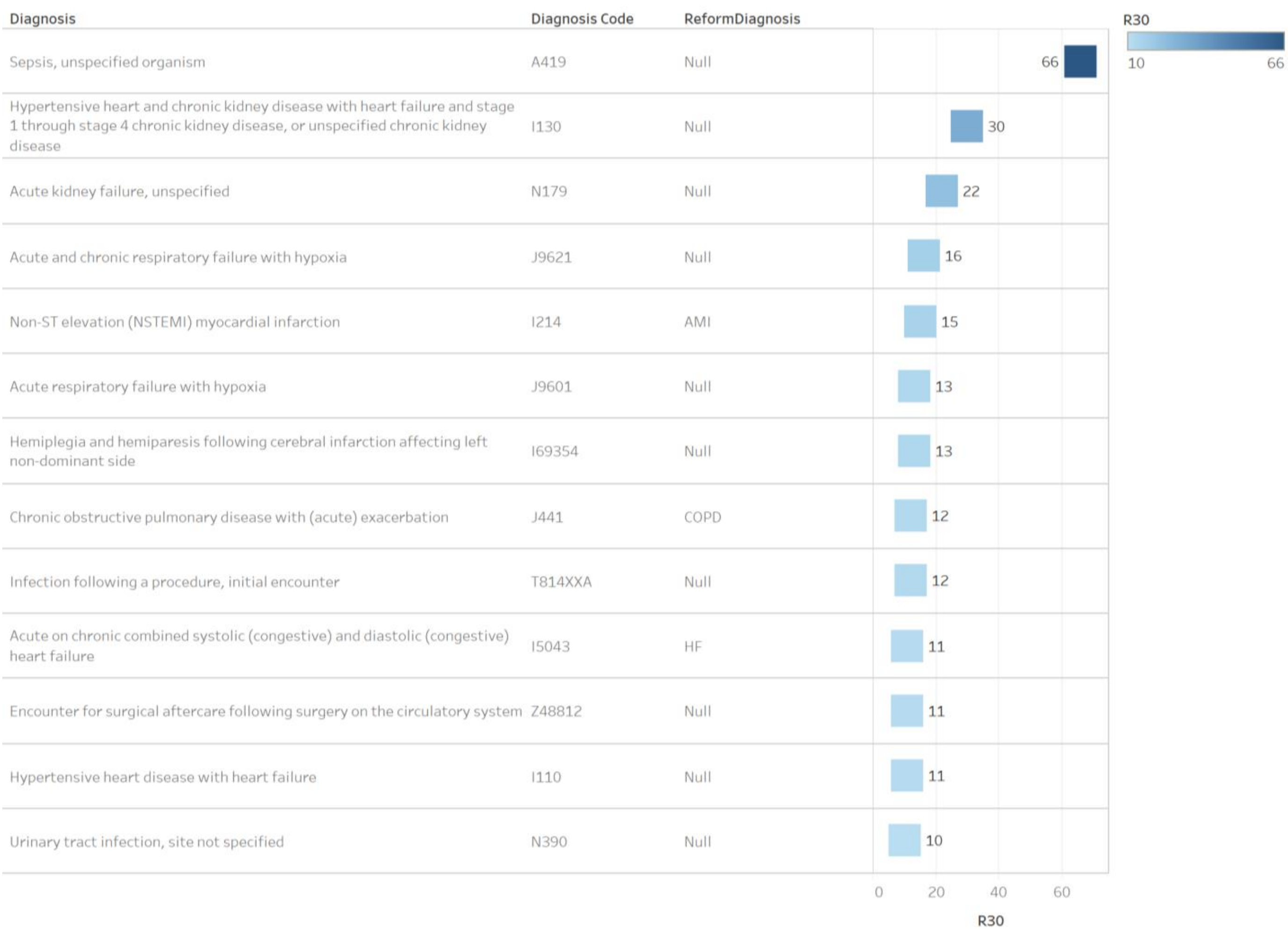
The trends of average of Sum R30 and average of Sum R30SameDiagnosis for Age Level. Color shows details about Gender.

The graph above shows the average readmission number, and the graph below shows the average readmission number with the same diagnosis.

As shown in the graph, age group 40-49 obtains the highest average readmission both for man and women. Specially, Male obtains a unusual high average readmission number, twice as much as the average readmission level.

To support our results, we searched the total number of male and female who are in group 40-49. It is found that the total number of patients in 40-49 age group is 137, 60 of which are male patients. This proves that this finding is not because of some extreme cases, but a common phenomenon.
On the contrary, interestingly, age group 30-39 obtains the lowest average readmission both for man and women.
Therefore, we suggest that hospitals pay more attention to the age group 40-49 to reduce readmission rate. Further analysis can be conducted to find the true reasons, and better communication procedures should be employed because it's may due to psychological factors.

Graph 3: diagnoses with the highest total readmission number

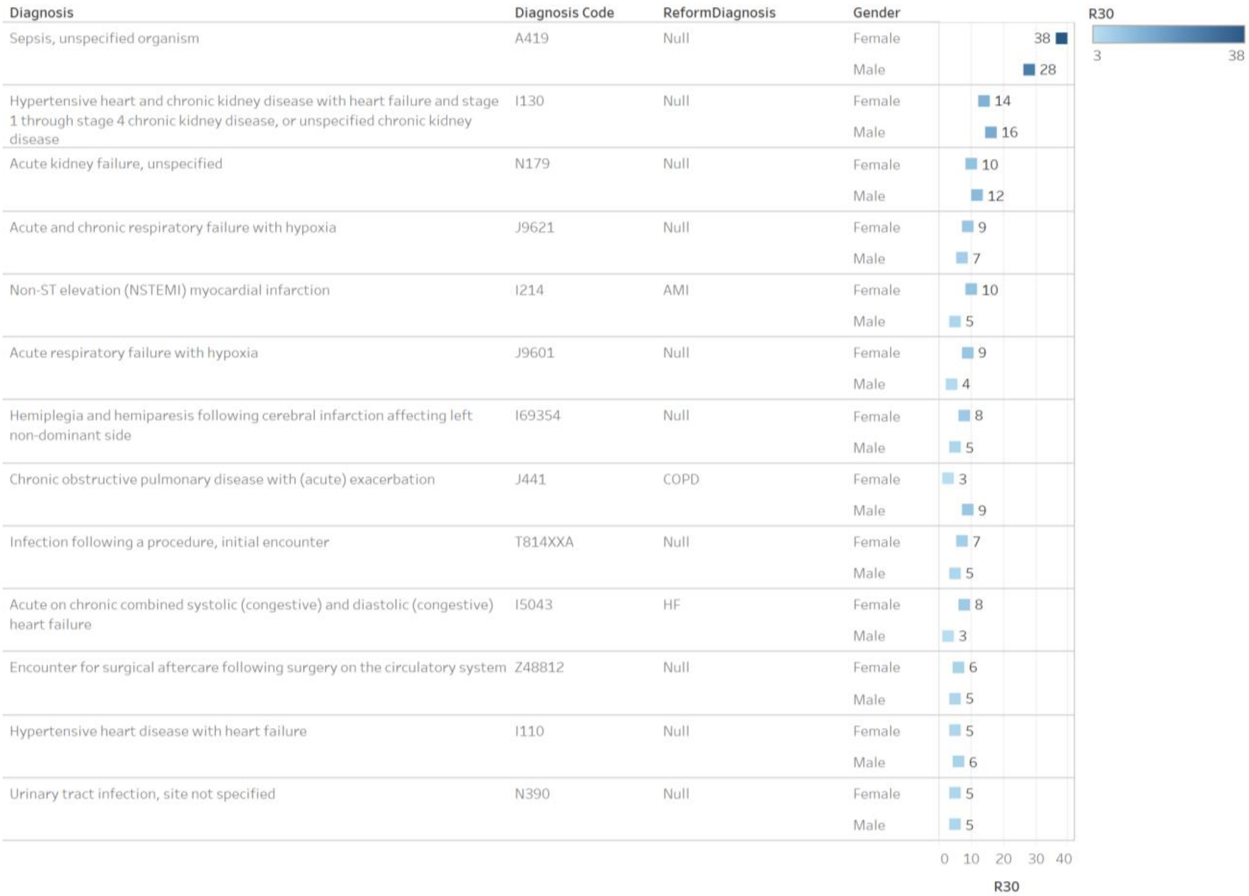


Sum of R30 for each ReformDiagnosis broken down by Diagnosis and Diagnosis Code. Color shows sum of R30. The marks are labeled by sum of R30. The view is filtered on ReformDiagnosis and Inclusions (Diagnosis,Diagnosis Code,ReformDiagnosis). The ReformDiagnosis filter keeps 6 of 6 members. The Inclusions (Diagnosis,Diagnosis Code,ReformDiagnosis) filter keeps 13 members.

We listed the top 10 diagnoses with the highest total readmission number in this graph. Special measures can be taken for the first few types of diagnoses.
Moreover, we contrast male patients and female patients in the following graph for the same diagnoses. We can see clear difference in certain diagnosis, such as sepsis, but generally either

both male and women have a high readmission number or both of them have a low readmission number.

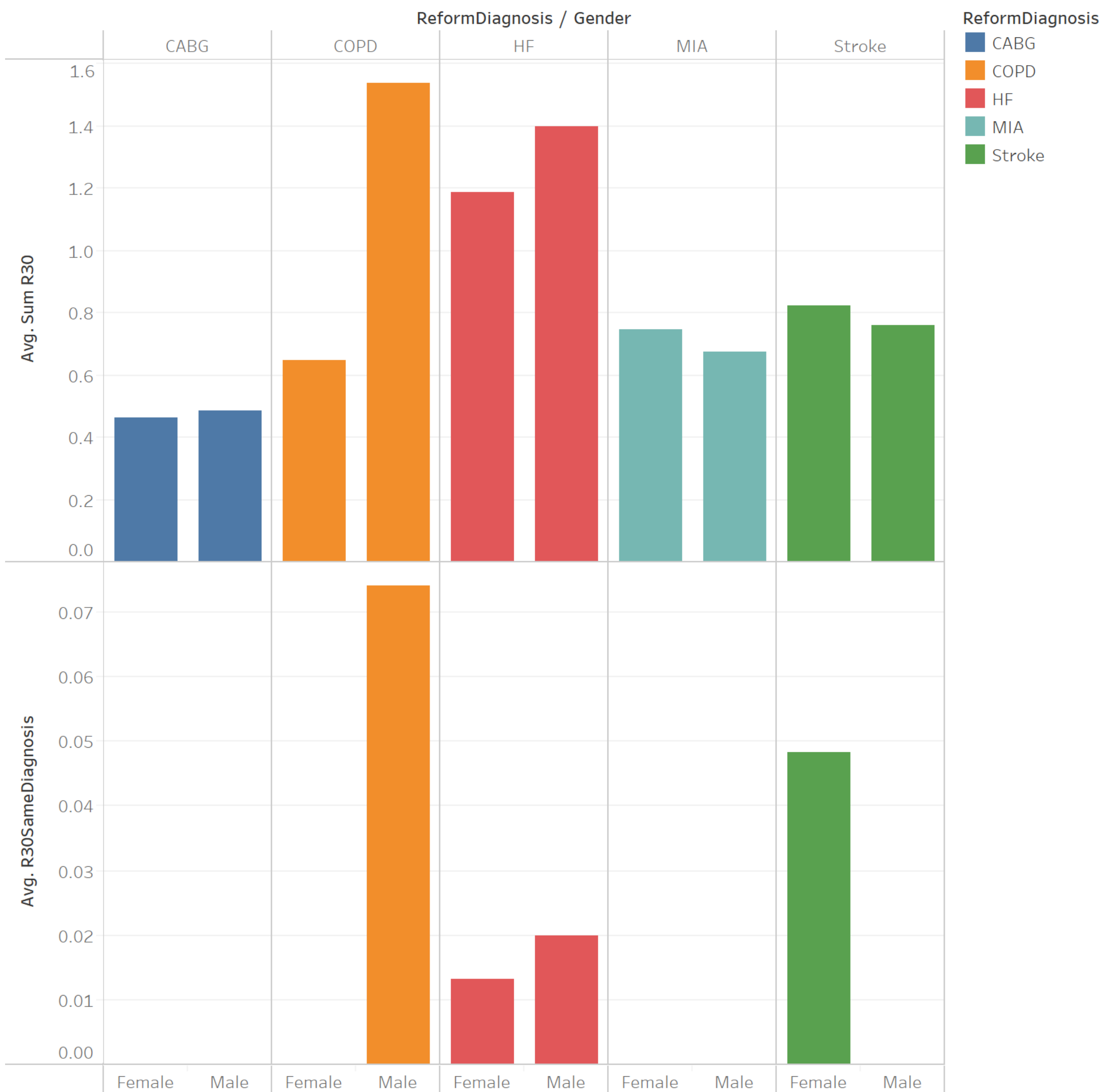
Graph 4: diagnoses with the highest total readmission number by gender



Sum of R30 for each Gender broken down by Diagnosis, Diagnosis Code and ReformDiagnosis. Color shows sum of R30. The marks are labeled by sum of R30. The view is filtered on ReformDiagnosis and Inclusions (Diagnosis,Diagnosis Code,ReformDiagnosis). The ReformDiagnosis filter keeps 6 of 6 members. The Inclusions (Diagnosis,Diagnosis Code,ReformDiagnosis) filter keeps 13 members.

Graph 5: average readmission number for certain diagnoses by gender

Gender



Average of Sum R30 and average of R30SameDiagnosis for each Gender broken down by ReformDiagnosis. Color shows details about ReformDiagnosis. The view is filtered on ReformDiagnosis, which excludes Null.

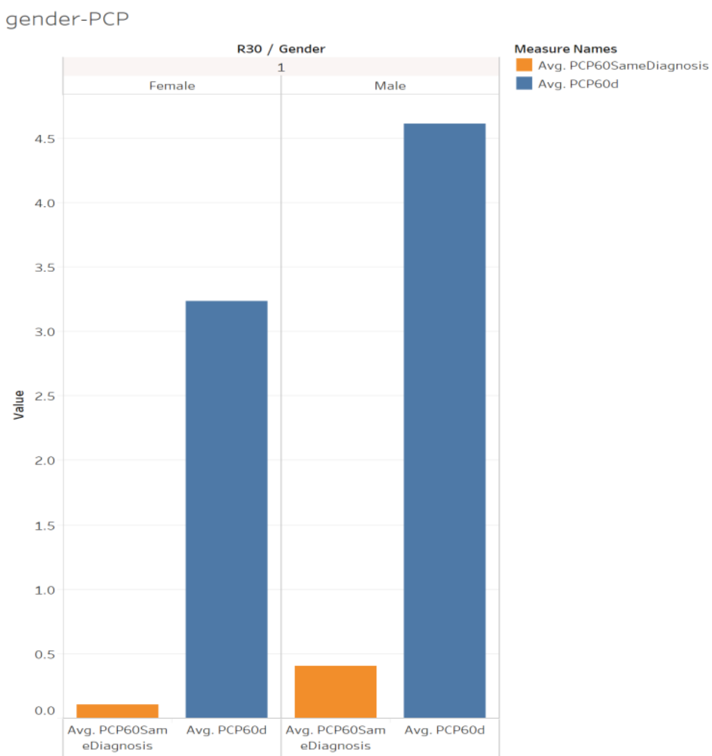
The corresponding diagnosis name for each label:

1. Acute myocardial infarction – AMI
2. Heart failure – HF
3. Chronic obstructive pulmonary disease – COPD
4. Coronary artery bypass grafting – CABG
5. Stroke – Stroke

This graph well contrasts women patient readmission to male patient readmission for the 5 certain diagnoses. For example, female patients are much easier to get readmitted to hospital for CABG and Stroke diagnoses. However, male patients are more likely to get readmitted for

COPD diagnosis . Most importantly, current record shows that only male patients may get admitted for the same diagnosis as previous under COPD diagnosis. Similarly, only female patients may get admitted for the same diagnosis as previous under stroke diagnosis. However, both may be readmitted for the same diagnosis of heart failure. This may due to the different attitudes or different physiological constitution between men and women.
In summary, hospitals are doing well to prevent the readmission with same diagnoses of AMI and CABG. Tailored measures to different gender should be done to reduce readmission with same diagnoses of COPD, HF, and Stroke.

<u>Graph 6: Average PCP (principal care physician) visit number for certain diagnoses in last six months by gender</u>



Avg. PCP60SameDiagnosis and Avg. PCP60d for each Gender broken down by R30. Color shows details about Avg. PCP60SameDiagnosis and Avg. PCP60d. The view is filtered on R30, which keeps 1.

This table reveals the average number of times the patient visited their PCP (primary care physician) within six months before this readmission and those PCP visits with the same diagnoses. Is illustrate male patient are more likely to have visited their PCP six months before a readmission for both general PCP visits and PCP visits with the same diagnosis.

Graph 7: Average PCP (principal care physician) visit number for certain diagnoses in last six months by gender

Avg. PCP6M and Avg. PCP6MSameDiagnosis for each Gender broken down by ReformDiagnosis. Color shows details about Gender. The data is filtered on R30, which keeps 1. The view is filtered on ReformDiagnosis, which keeps AMI, CABG, COPD, HF and Stroke.

The corresponding diagnosis name for each label:

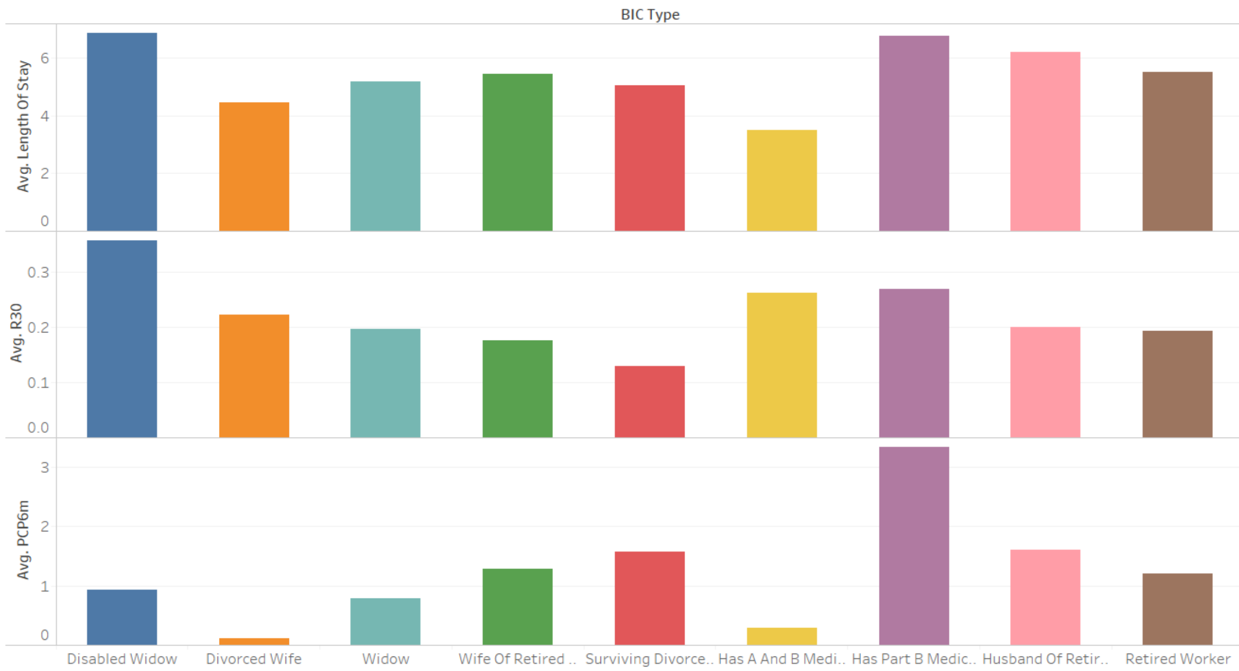
1. Acute myocardial infarction – AMI
2. Heart failure – HF
3. Chronic obstructive pulmonary disease – COPD
4. Coronary artery bypass grafting – CABG
5. Stroke – Stroke

The graph above shows, for readmissions within 30 days, what's the average number of times those patients visit their principal care physician(PCP) in last 6 months by different types of diagnoses and gender. The graph below shows the corresponding average number of times those patients visit their principal care physician(PCP) in last 6 months for the same diagnosis as this readmission.

Interestingly, for COPD, male patients visited PCP for average 1.75 times for the same diagnosis in last 6 months, but female patients never did that. Similarly, for stroke, male patients visited PCP for average 2.3 times for the same diagnosis in last 6 months but female patients visited for average only 0.25 times.

Graph 8: Readmission Rate by Marital Status

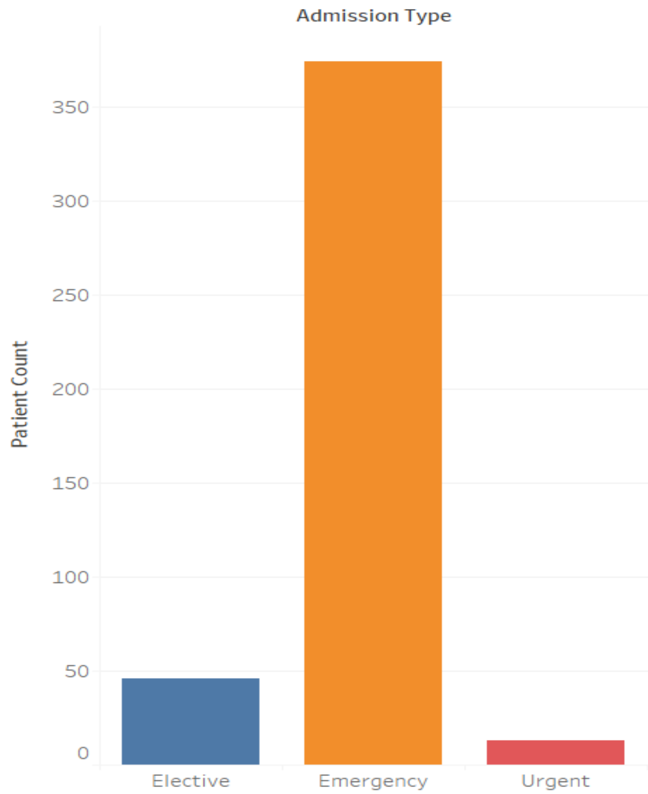
After looking at gender and seeing the relationship between gender and readmission rates and pcp visit raet, we decided to break it down a little bit more and look at marital status.



The three columns are the far left of the graph above are all groups of divorced or widowed women, and one can easily see that they have significantly lower pcp visit rates than most of the other groups. The disabled widow group is also especially noteworthy, because it has the highest average readmission rate, the highest average length of stay, but it also has the youngest average age at 58. To examine the effect of being single, we can compare these groups to the retired workers group. It is clear that being divorced or widowed is a good indicator of higher readmission rates and lower pcp visit rates, and more data could be gathered about why this is the case.

Graph 9: Diagnosis Based on Admission Type

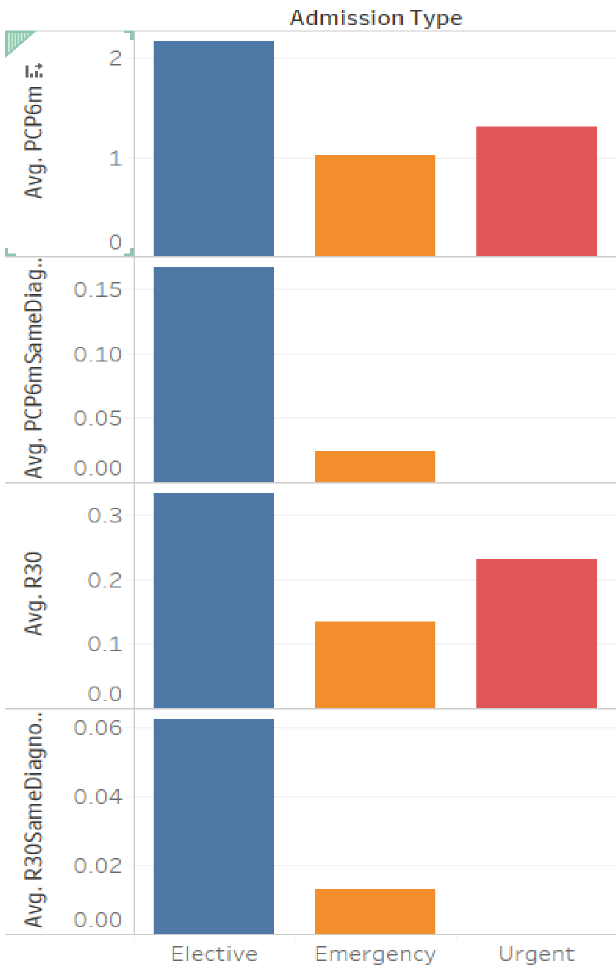
To analyze the 5 diagnoses above even more (AMI, HF, COPD, CABG, and Stroke), we grouped them together and analyzed how people with these diagnoses were admitted, and also, how many people that were admitted expired.



The results show that over 90% of people that receive this diagnoses were admitted after an emergency or urgent care, which we can compare to the average over all patients, which is about 75%. The graph also shows that as the urgency of a patient increasing, the expiration rate also increases; the difference between elective and emergency expirations is about 10%.

Graph 10: R30/PCP6m Based on Admission Type

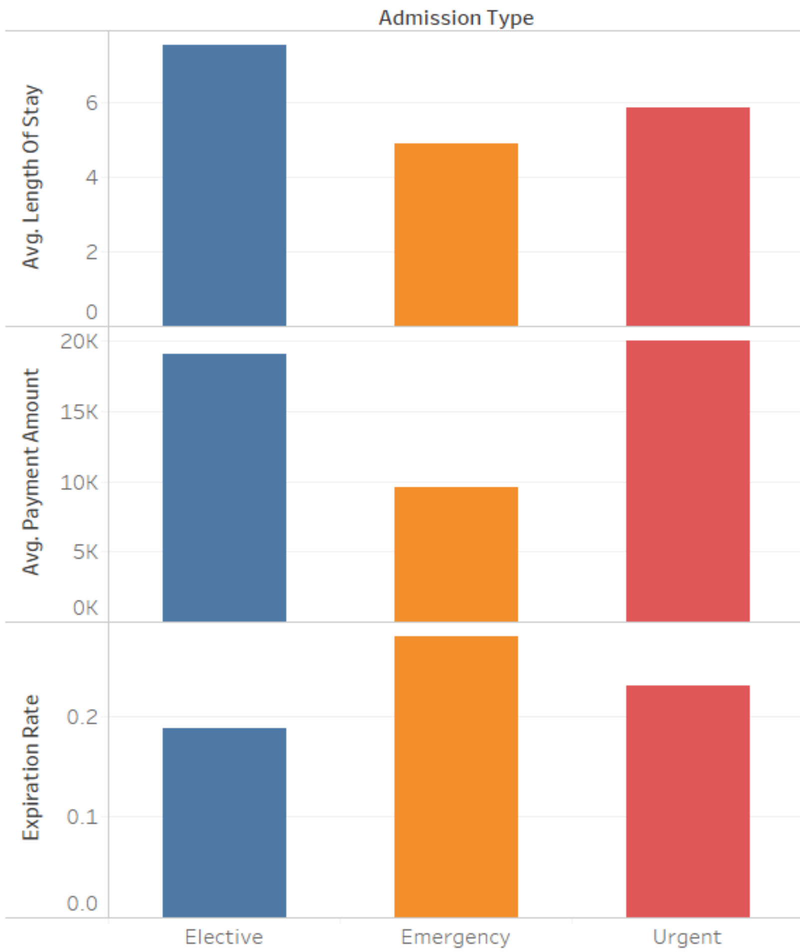
We continued looking at how people within this diagnoses group were admitted by examining how many of them were readmitted within 30 days and how many had seen their primary care physician within the last 6 months.



The results show that patients who suffered AMI, HF, COPD, CABG, or a Stroke and had been admitted electively had a much higher average number of PCP visits within the last 6 months and also a much higher average readmission rate within the next 30 days compared to those who had been brought in because of an emergency

Graph 11: End Result Based on Admission Type

We wanted to compare what happened to each patient who suffered from one of AMI, HF, COPD, CABG, or a Stroke after they were admitted. To do this, we analyzed the average length of stay in the hospital, average cost of the hospital stay, and expiration rate.



The results show that compared to those brought in electively, those who were brought in because of an emergency had an average of a 2 day shorter length of stay, an almost $10,000 cheaper bill, but also an almost 10% rate of expiring.

Looking at these three graphs, we can see that those who are admitted electively seem to be much more cautious; they visit their pcp and are readmitted more often, they stay longer and pay more at hospitals, and the way someone is admitted can tell us a lot about them. One issue with this analysis is that we used the hospitals records of how people were admitted, and we do not know exactly what qualifies an admission to be elective rather than urgent, and more data on admission would help us clarify our results. To continue this analysis, we tried to look at any identifying characteristics of those that were admitted electively, but after looking at age, race, and zip code, we could not find any significant patterns, and more data about each patient, such as income, would be very useful.

Conclusion

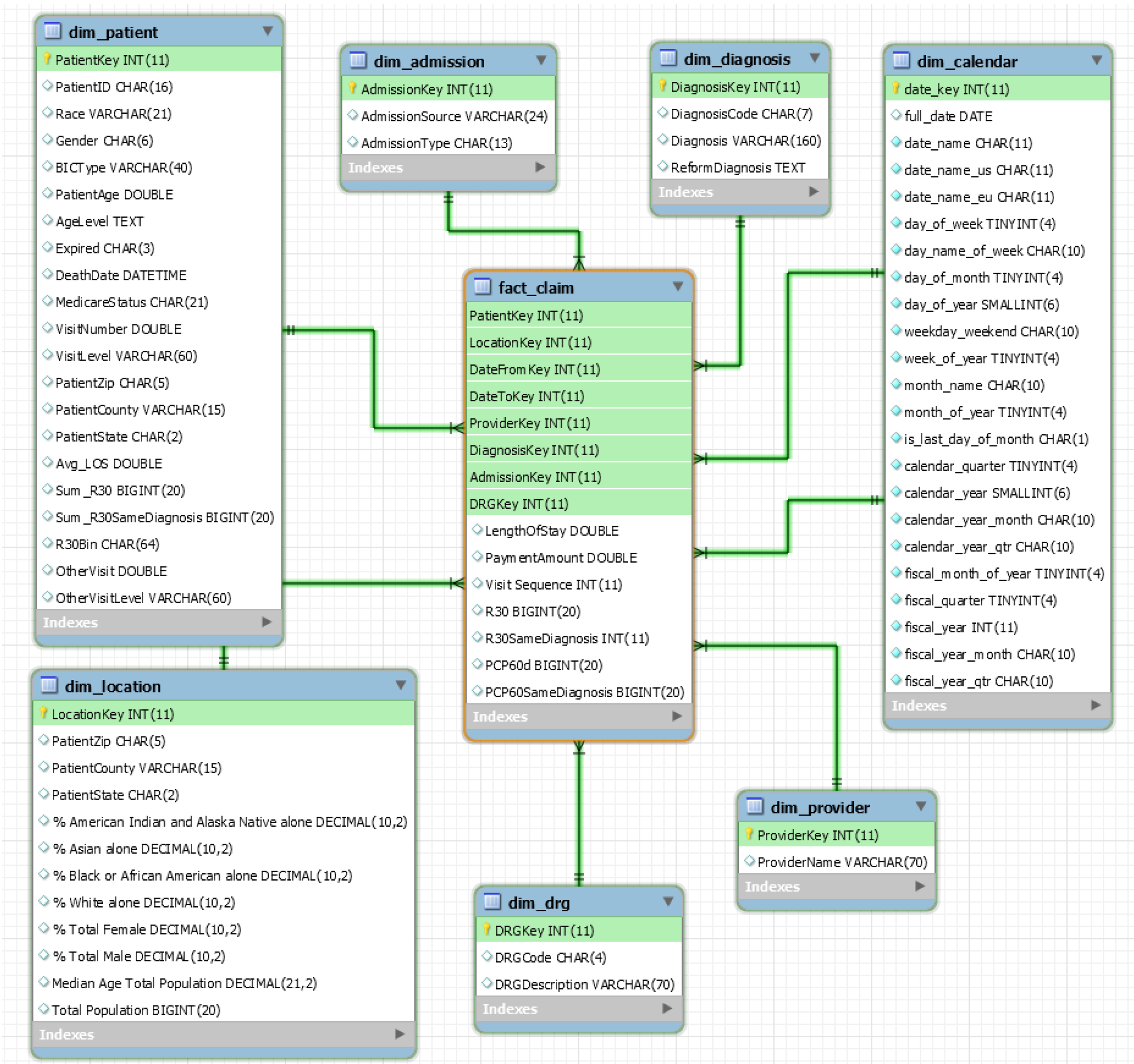
Ultimately, after analyzing and visualizing the data we came to several interesting conclusions regarding readmission rate. In answering our initial three research questions, we found that:

1) Those who were divorced or widowed were far more likely to be readmitted within 30 days than those who were married. Additionally, the widowed group also had a larger average length of stay in addition to being younger (70 - average age) on average than most other patients readmitted
2) Men and women display different patterns regarding their readmission and frequency; women are more likely to be readmitted for stroke, men are more likely to be readmitted for COPD for heart failure. Men and women also display different habits in regards to communication with their primary care providers:
    a) Men are more likely to see their PCP 6 months prior to readmission
    b) Women are more likely to file an in-patient claim
3) Patients can be broken up into two groups; cautious and less cautious. Cautious patients go to the hospital before an emergency, visit their PCP more often, and stay longer. Less cautious patients are less likely to visit their PCP and try to stay away from the hospital. It is hard to differentiate these two groups, but admission type is a good indicator.

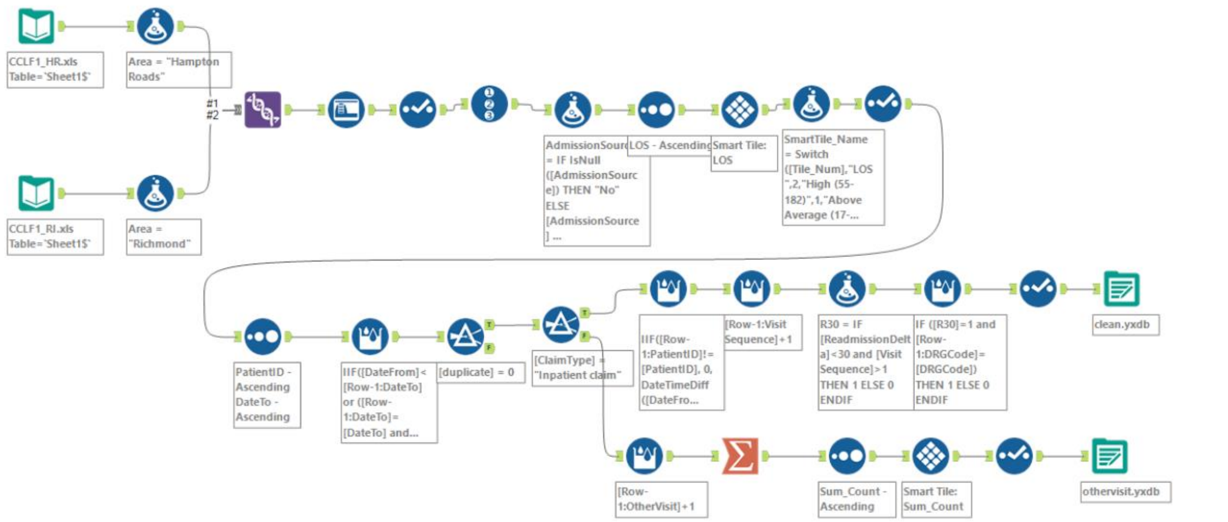
While we learned a lot of interesting things through this project and observation of the data, we still believe there is more to explore. If given the opportunity to attempt this project again, we would try to obtain data regarding patient communication with their primary care provider (PCP) and see how that factor affects readmission. Additionally, we would also try to find another external source of data in order to measure the effect of income level on readmission rate as we believe both of these factors could have an interesting effect on readmission rates.

Appendix

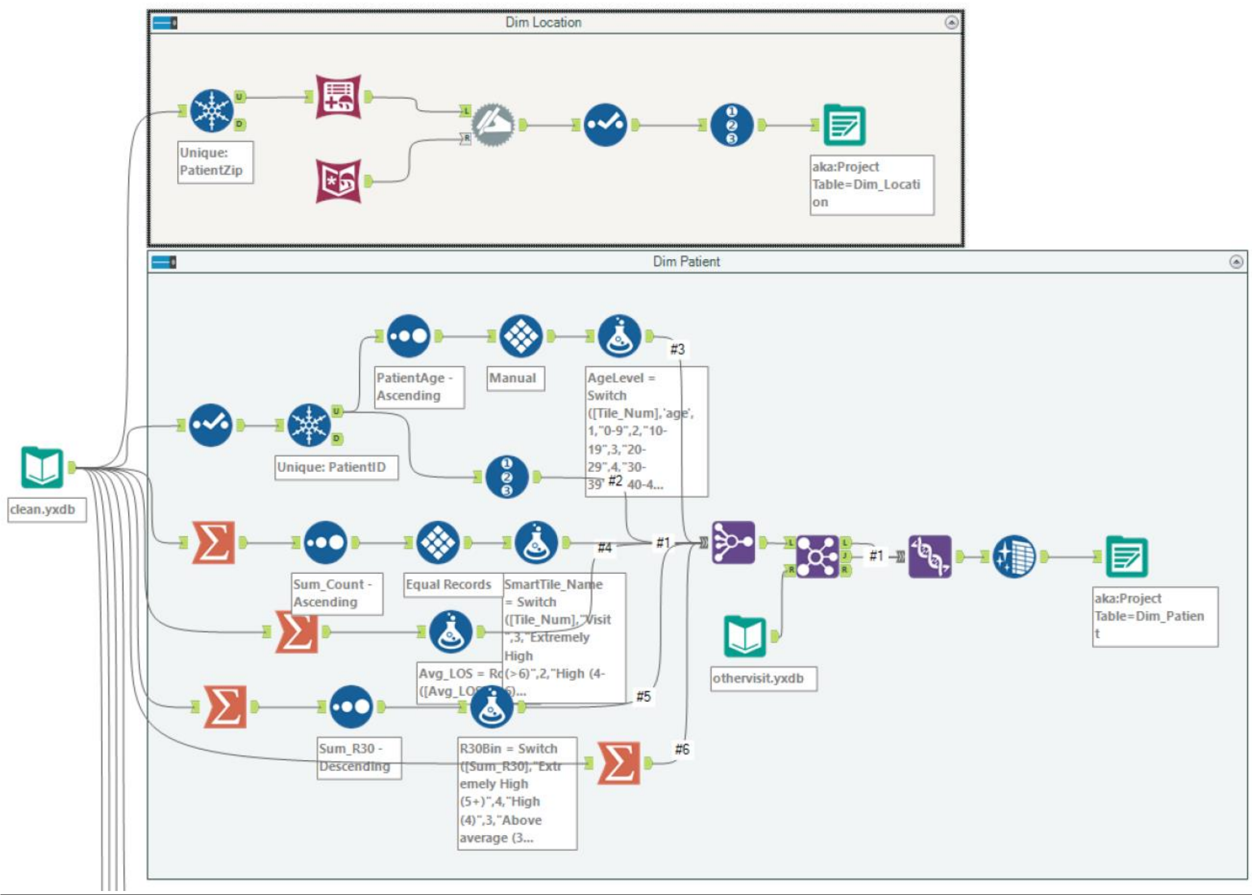
Appendix A: Relational Schema

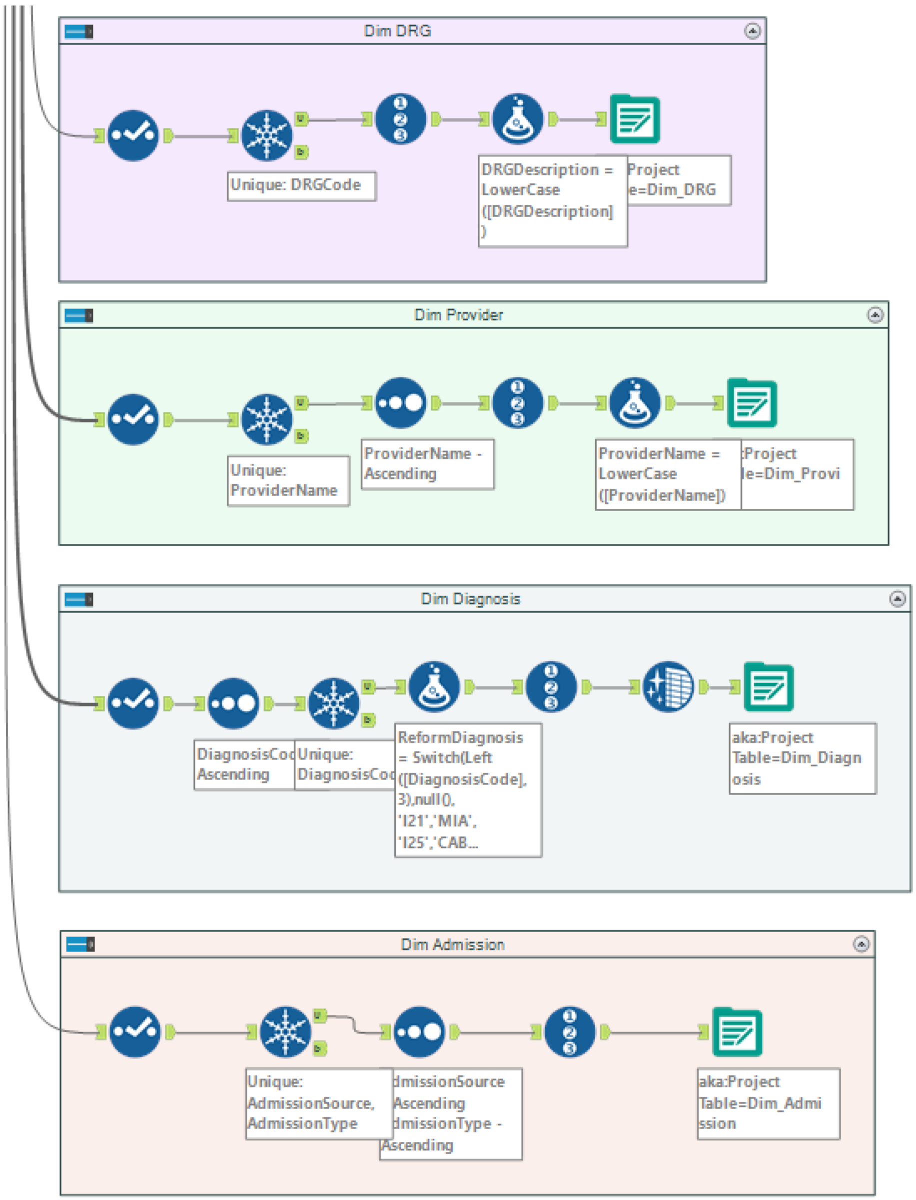


Appendix B: Cleaning Data in ETL

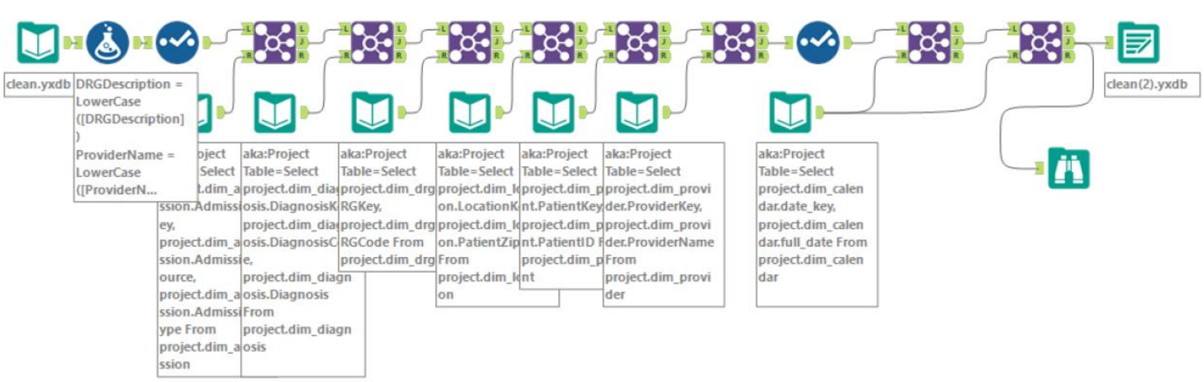


Appendix C: ETL Dimensions





Appendix D: Building the Fact table in ETL



Appendix E: Combining the Databases