Analysis of degradation patterns in the clinical efficacies of anti-microbial drug usage

Shweta Arya, Oklahoma State University

ABSTRACT

In this paper we are analyzing the healthcare industry and primarily the anti-microbial drugs used to cure Urinary Tract infection and other diseases. We have provided analysis on how usage of drugs differ with different sexes and with different age groups. We also analyze how much amount is being spent on individual drugs and how many doses were sold. Additionally, we are predicting and identifying what factors lead to a patient to get re-admitted to a healthcare provider after getting diagnosed once. We are leveraging power of SAS with models such as Decision Trees, Regression and Neural Network to identify the aforementioned factors. Finally we are also forecasting the price of such drugs using SAS Forecasting tool and identifying if a patient can opt for any inexpensive substitute generic drug instead of spending a fortune on its expensive counterpart.

INTRODUCTION

Healthcare industry is an industry which had been largely off limits to the power of analytics. While analytics was powering other industries, healthcare was reeling behind. The changes in the industry came along due to available data, computational power and primarily due the acceptance of the stakeholders that analytics will power the industry and not weaken it. With onset of user friendly products by companies such as SAS, it became easier for professionals to apply the logic to derive hidden insights from the data.

According to Forbes', the healthcare industry in the United States is over 3 billion USD and it is set to double in coming years. There are ongoing vast technological advancements in the field and analytics is an integral part of that dynamically changing sphere.

It's scary how bacteria and other pathogens have evolved over time and increased their resistances to drugs and other medical advancements. Studies have revealed that the harmful effects of antimicrobial drug resistance and the patients' susceptibility to infection is growing at an alarming rate. It's projected that by 2050 this trend could not only lead to the reduction of GDP by 2 to 3.5% every year but also lead to death of over 10 million people.

According to Centers for Disease Control and Prevention (CDC), Antibiotic resistance infections in U.S alone costs an excess of $20 billion in health care and $35 billion societal costs every year. Cerner database has comprehensive medical records of over 50 million patients that are periodically refreshed with latest information. The data consists of information related to patient encounters, medication details and anti-microbial drug usage and their susceptibility results which will be used in determining the effectiveness of drug prescription.

We plan to analyze how success rate of each drug against patient encounters has changed over the past 10 years and then suggest the appropriate drug replacement measures such as Generic Substitution, as a prospective cost management opportunity which reduces the direct costs incurred in healthcare industry. The descriptive statistics, data cleaning & preparation will be done in SAS Enterprise Guide and SAS Enterprise Miner. Using SAS Forecast Studio, we will do time series modeling to forecast how the trend of the cost of these drugs will affect their usage in future.

DATA

The dataset used for this paper was obtained from the Cerner database with the help of Centre for Health Systems Innovation – Oklahoma State University (CHSI). It has records of over 50 million patients. In order to support this research, we merged 5 datasets namely, urban_clinical_event, urban_encounter, urban_diagonosis, urban_medication and urban_procedure. The scope is narrowed to the study of Urinary Tract Infection (UTI) patients and antimicrobial drugs Unasyn, Azithromycin, Ampicillin, Zithromax – used in the treatment of UTI along with drugs which might cause this infection if used in excess. The final dataset has around 10000 observations and 12 variables including age, gender, race, admission date-time, discharged date-time, caresetting, used etc.

EXPLORATORY ANALYSIS

The graph below demonstrates the amount spend on drugs by males and females to cure Urinary Tract Infection drugs. From the bar graph it is clear that females spend around three times the amount males spend on drugs. Through the data it was found that the females had spent around $120 million and males approximately $45 million for drugs. In total the amount spent was over $150 million.
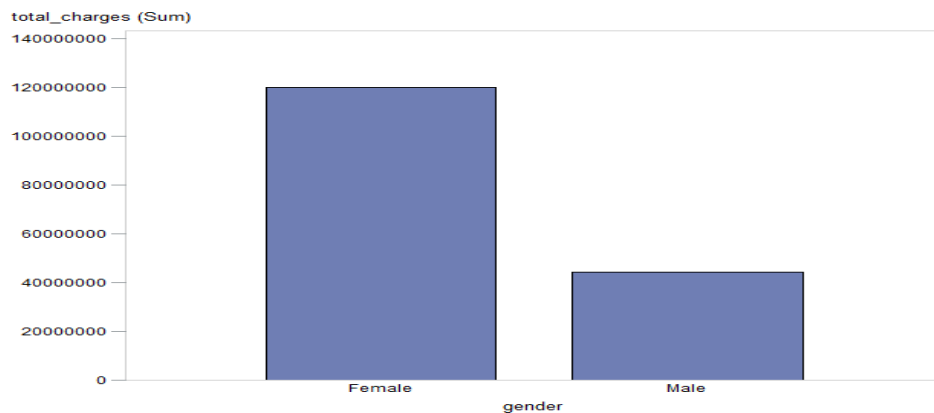

In the bar chart below, we have segmented drug usage and the respective amount spent on the individual drug. Furosemide was the drug on which maximum amount was spent with close to 50 million USD. The next most amount spent drug was Morphine Sulfate which was close to 40 million USD. Drugs such as Principen, Unasyn, Zithromax and Ampicillin are drugs on which the patients spent the least. It is interesting to know that UTI is found in among 2% of the people, mostly females with age 60+, taking Furosemide or Morphine Sulfate. The excess usage of these drugs might be one of the reasons for Urinary Tract Infection.

Bar Chart
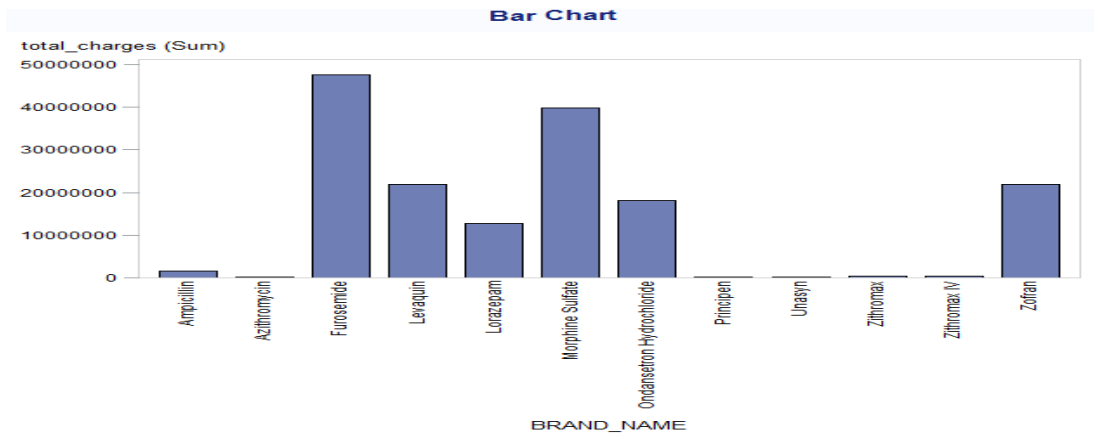
Interestingly, in the age group of under 15 in male category, Zofran is the drug on which the most amount is spent. Rest all the drugs have minimal amount spent. Alternatively, we can see that in females under 15 category, most of the amount is spent on Ondansetron Hydrochloride and Lorazepam.
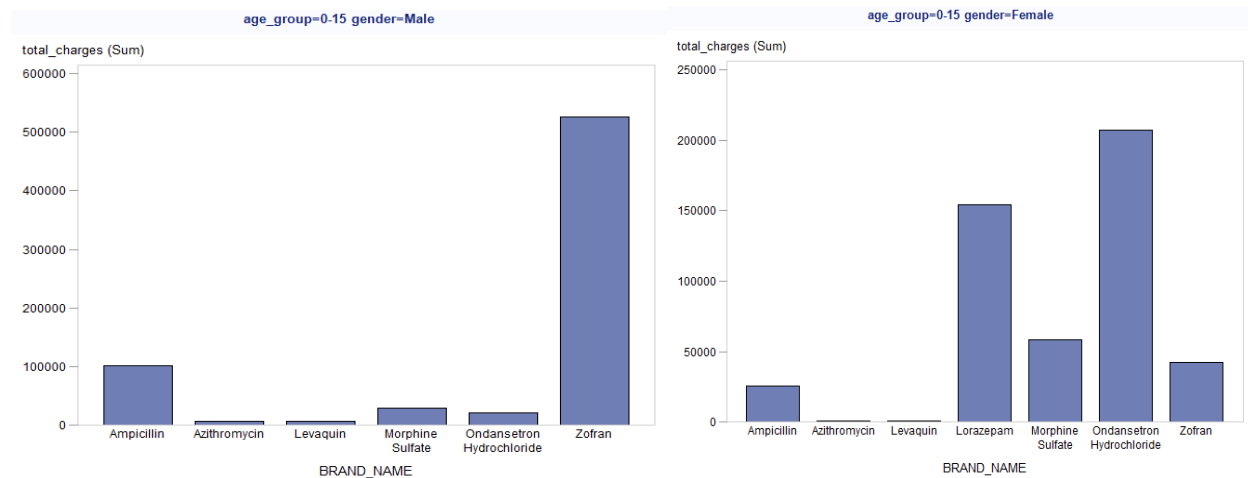

In the age group 16-30 for females, the bar graph below shows that the Ondansetron Hydrochloride had the most amount spent on them. In the same age group, males spent most of the amount in Morphine Sulfate. This shows how different genders spent different amounts on different drugs. Another drug Furosemide was the next drug on which the 2nd largest amount was spent.

age_group=16-30 gender=Female

total_charges (Sum)


age_group=16-30 gender=Male

total_charges (Sum)
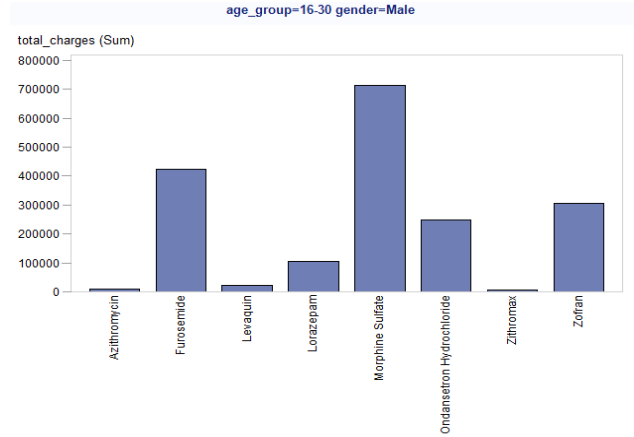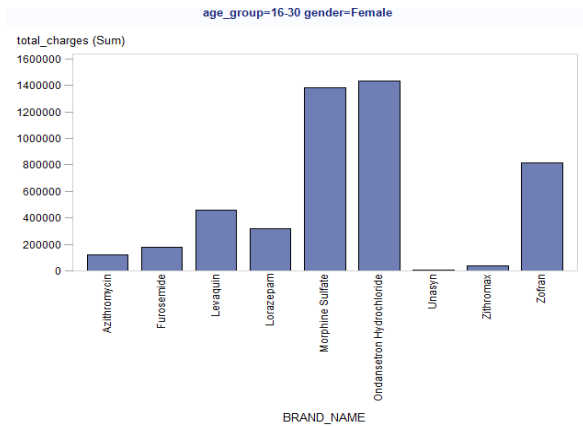
In the age group 61-75, Furosemide is the most spent drug by females and males. However females spent around 20 million on it, and males around 3 million. This ratio is really small as compared to original ratio of 3:1.

age_group=61-75 gender=Female

total_charges (Sum)

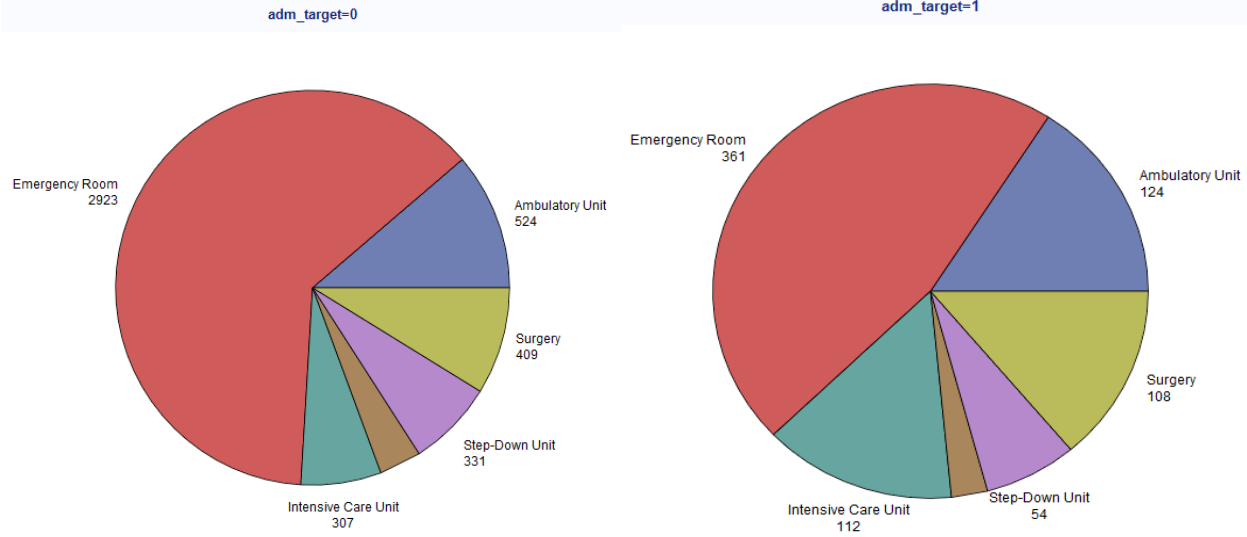age_group=61-75 gender=Male

total_charges (Sum)
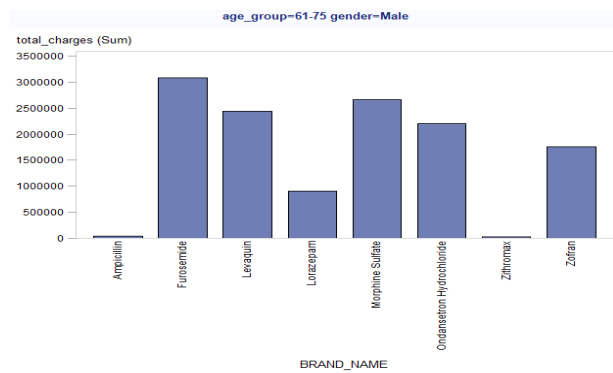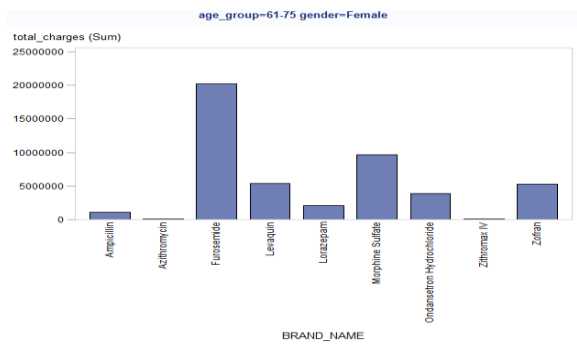
We see similar trends in age group 76-90 where in the most amounts are spent on Furosemide and Morphine Sulfate.
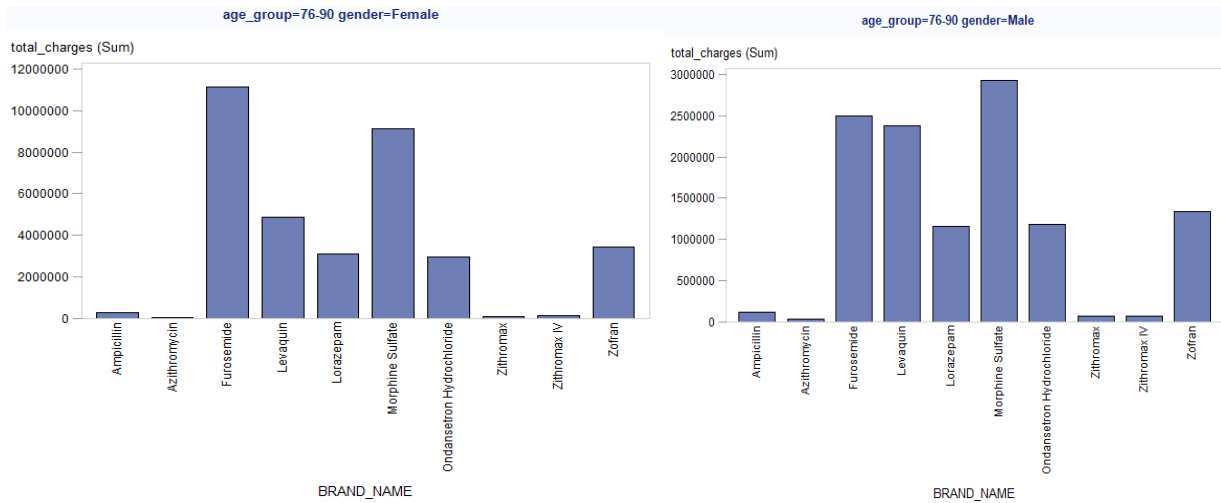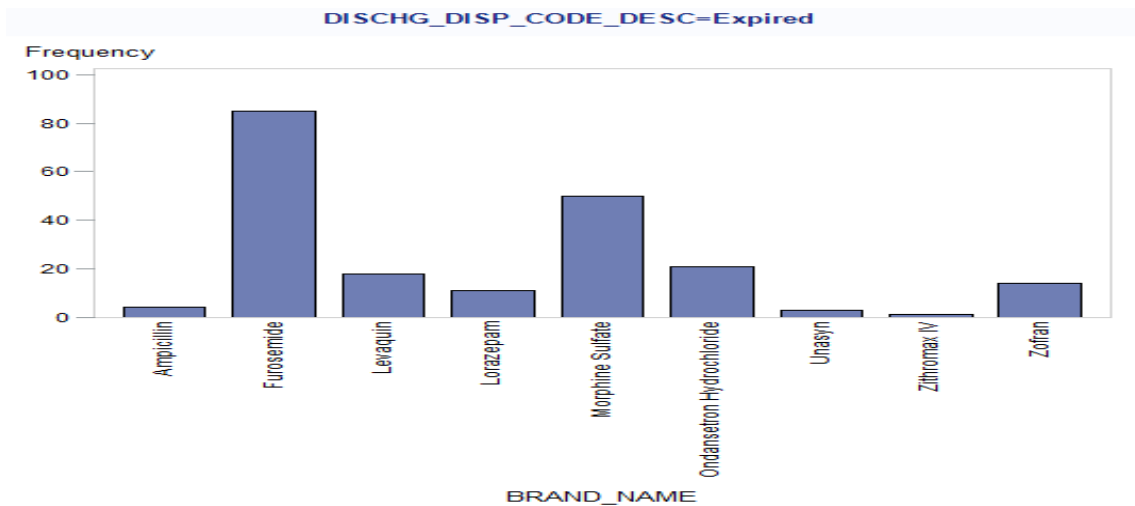
In addition to this, we also have the total amount of doses taken in Morphine Sulfate had the maximum amount of sales of its doses. Furosemide and Ondansetron Hydrochloride had the next most sales of their doses.
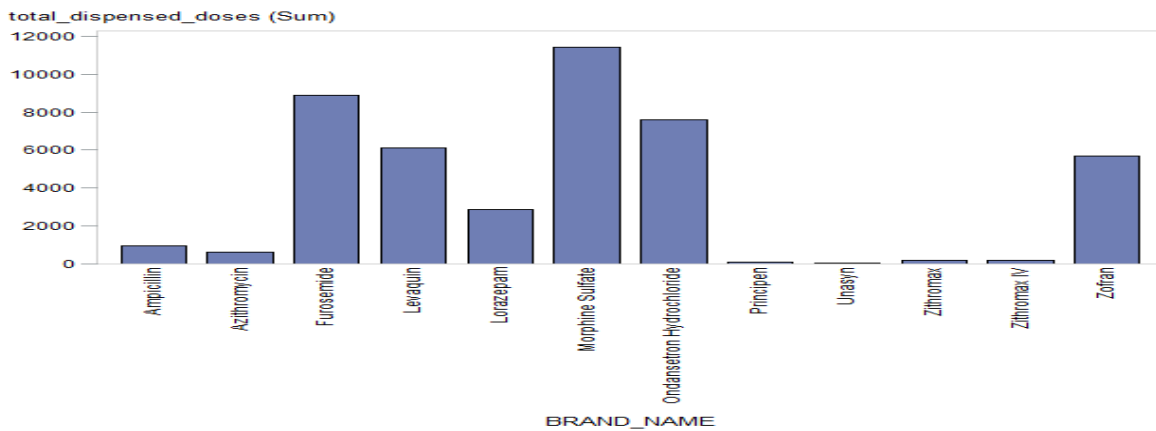
In the pie charts below show the distribution of caresetting based on the patients that came back to hospitals within 30 days (adm_target =0) and those that came after 30 days (adm_target =1). The patients that came back to the hospital within 30 days had 2923 emergency visits which is over 800% more than the visits of the people after the 30 day period. Similarly other fraction of visits also increased in the patients coming after 30 days and their visits were more diverse in nature.

Emergency Room
2923

Ambulatory Unit
524

Surgery
409

Step-Down Unit
331

Intensive Care Unit
307

Emergency Room
361

Ambulatory Unit
124

Surgery
108

Step-Down Unit
54

Intensive Care Unit
112

The bar graph below shows what drugs did people took who eventually died and Furosemide was the drug which topped the graph. This can also be attributed to the amount of people who were consuming the drug was large. In comparison of the number of doses sold, it is interesing to see that the Furosemide has low amount of doses sold yet more ratio of the people who eventually died.

DISCHG_DISP_CODE_DESC=Expired

Frequency

| 100 |
| 80 |
| 60 |
| 40 |
| 20 |
| 0 |

Ampicillin · Furosemide · Levaquin · Lorazepam · Morphine Sulfate · Ondansetron Hydrochloride · Unasyn · Zithromax IV · Zofran

BRAND_NAME

METHODOLOGY

Cross Industry Standard Process for Data Mining methodology is used for building predictive models. According to CRISP-DM methodology, the process is divided into 6 stages, namely- business understanding, data understanding, data cleaning, modeling, evaluation and deployment.

The flow chart below provides a vivid display of all the processes involved in the analysis of readmission of patients in this paper. The data is prepared using Base SAS and SAS Enterprise Guide, and then it was processed for data validation and imputing missing values. Regression, decision tree and neural network are built to predict the binary variable adm_target. This variable has been created to predict the factors affecting readmission of a patient. The target variable is created with the help of admitted_dt and discharged_dt variables. If the patient is readmitted within 30 days of discharge, then the adm_target value is 0. If the patient doesn't encounter again or is readmitted after 30 days, then the adm_target value is 1.
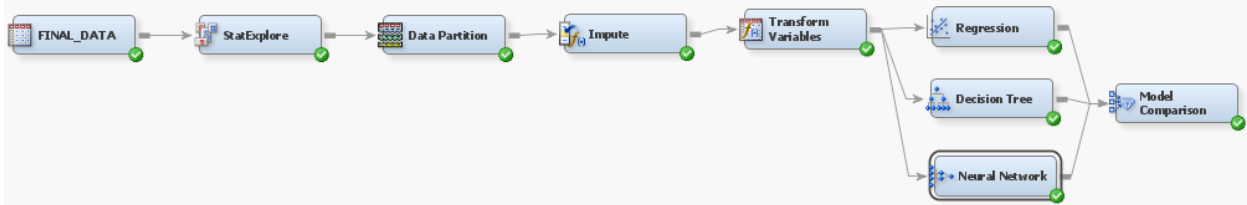

The data was partitioned into 70% training, 15% validation and 15% test data. Training data is used to train the model so that it performs well when the data is unknown. Validation data is used to avoid overfitting of models and the test data is used to check how well our model predicts in the real world scenario.

The created models were compared using the model comparison node to find the best model among those. It is seen that the decision tree was chosen as the best model on the basis of valid misclassification rate.

| Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Valid: Average Squared Error	Selection Criterion: Train: Average Squared Error	Valid: Misclassifica tion Rate	Train: Average Squared Error	Train: Misclassifica tion Rate
Y	Tree	Tree	Decision Tr...	adm_target	0.117924	0.109592	0.140882	0.109592	0.129678
	Neural	Neural	Neural Net...	adm_target	0.129058	0.113788	0.159032	0.113788	0.139311
	Reg	Reg	Regression	adm_target	0.128213	0.123074	0.155575	0.123074	0.149315
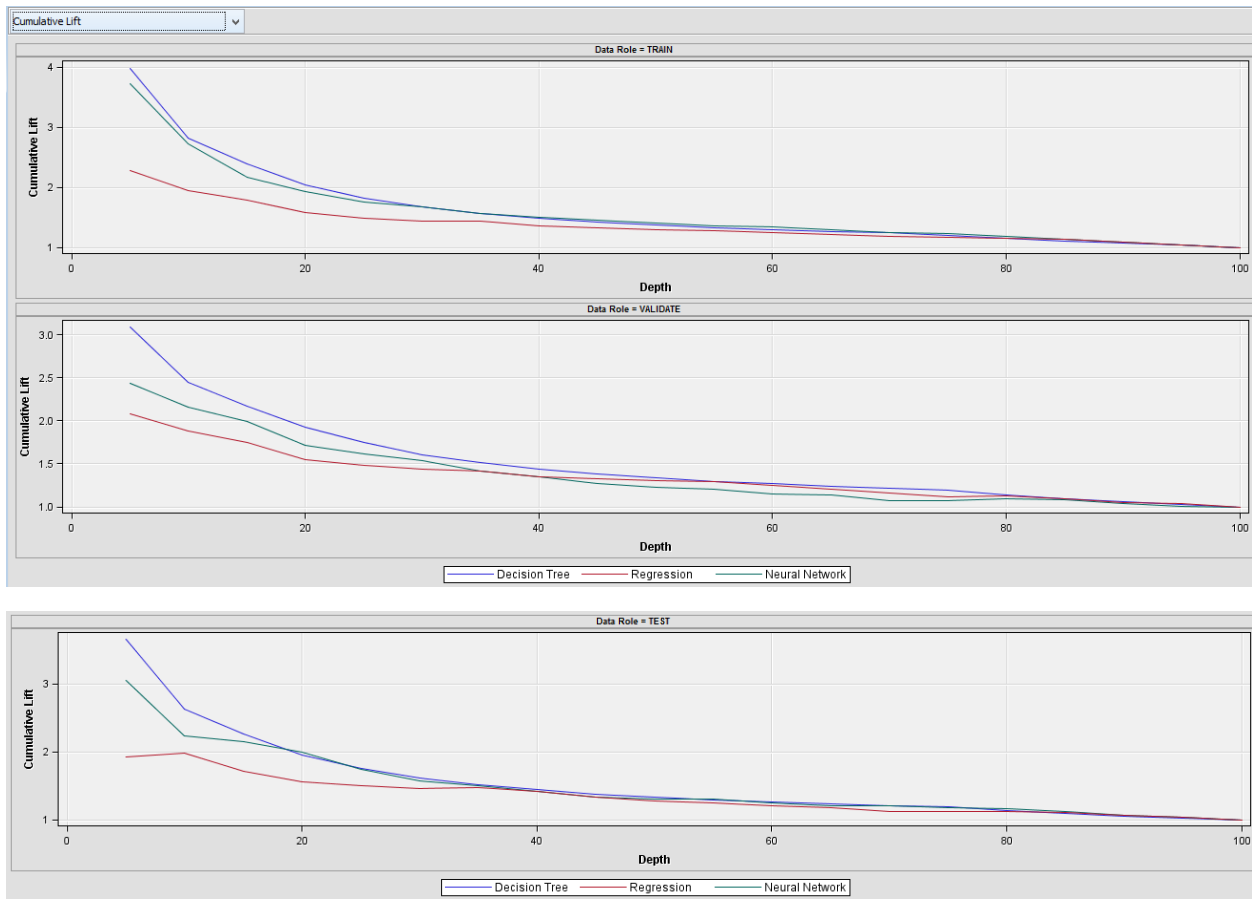
The below graph is the Receiver Operating Characteristic (ROC) curve for Training, Validation and test data set.
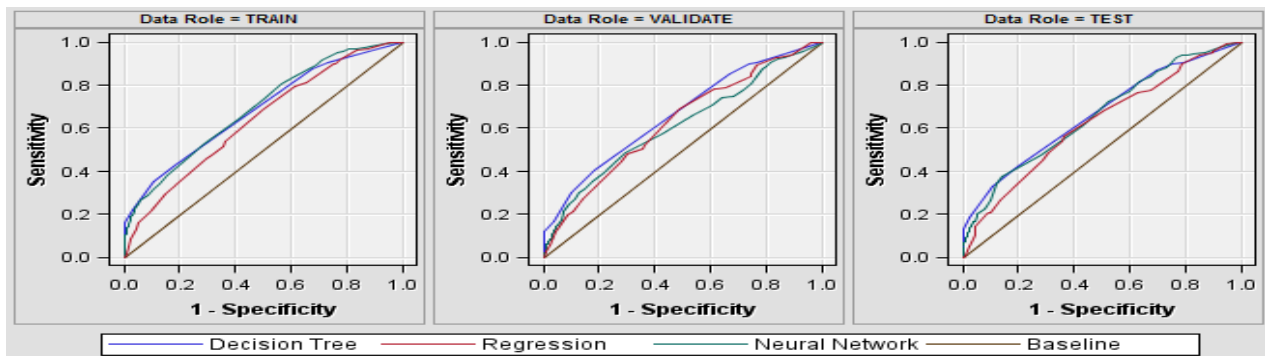
We have plotted the Receiver Operating Characteristic (ROC) curve for the three models developed on the data. From the curve we can see how much better results they provide over the baseline model and after analyzing the ROC curve, we find that Decision Tree has the largest area under the curve and highest cumulative lift amongst all models. Decision tree's ROC curve is demonstrated by the blue line and it provides the best balance between Specificity and Sensitivity.

RESULTS

Decision tree is chosen as the best model on the basis of valid misclassification rate of 14%.The most important variables chosen by decision tree are payer_code_desc (how was the amount paid), caresetting_desc, age_in_years, total_charges, dose_units_desc and several more. According to this model, if a patient's age is greater than 78, paying by Medicaid or Medicare and has been admitted in the emergency room then there is 86% probability that he/she will be readmitted.
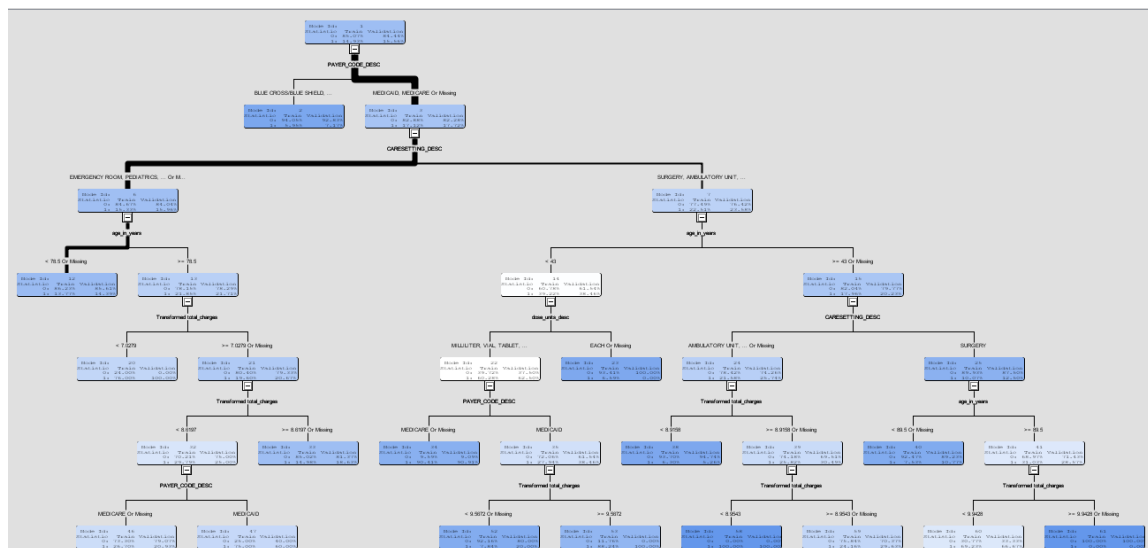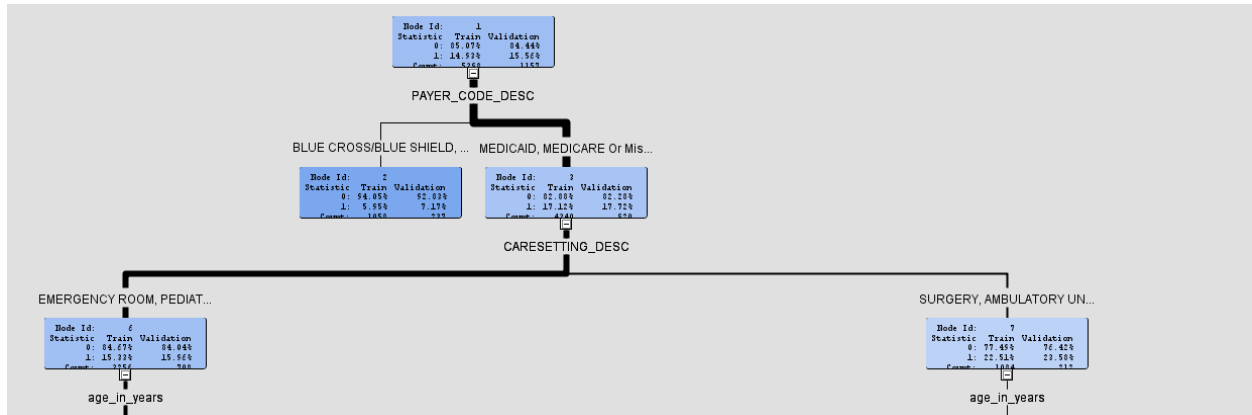

FORECASTING

SAS Forecast Studio 13.2 has been used to perform the time series analysis and forecast the trend of drug usage.

We have chosen the admitted date as the time variable and total amount spent on each drug as the target. The total MAPE is 8.43 for the different forecast models built.
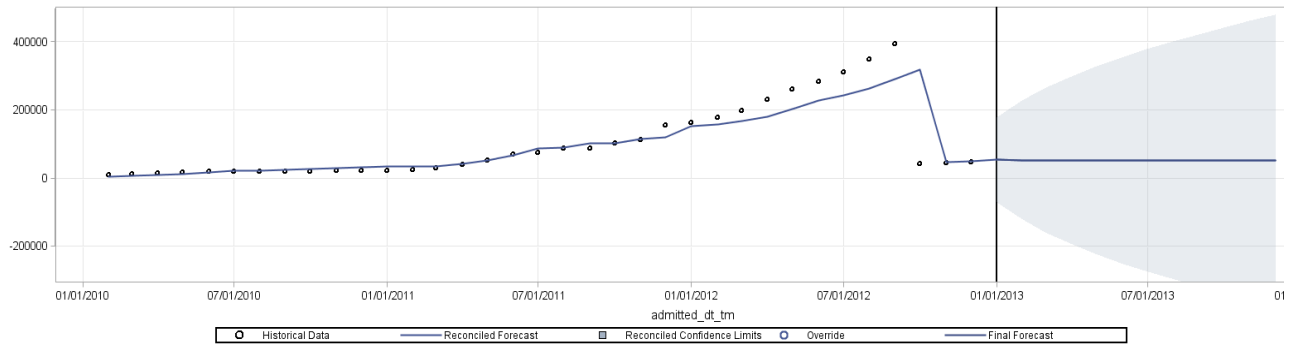
The data for Azithromycin has peculiarities that we see in the graph above but unable to explain because of our lack of domain expertise. For instance, it seems the amount spend was increasing slowly but steadily from 2010 to 2012 for Azithromycin and then there was a sudden drop. Post-2012 it seems to have flat-lined as forecasted by SAS Forecast Studio.
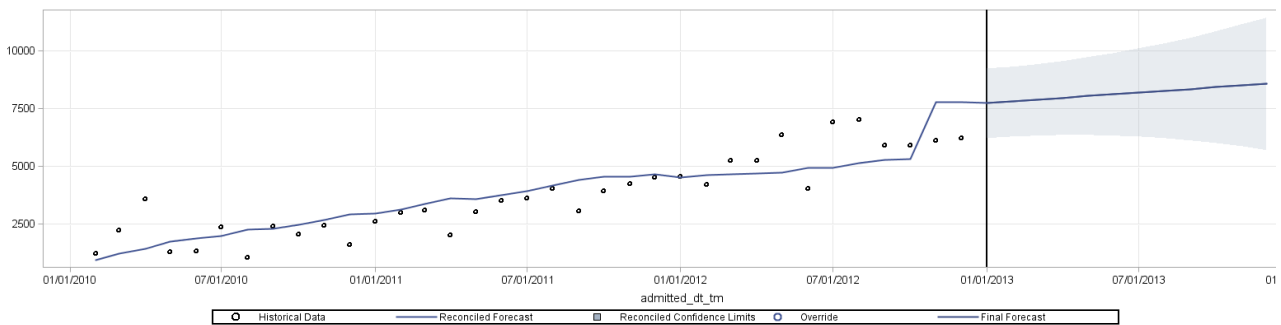
Ampicillin Sulbactum has a linear positive trend.
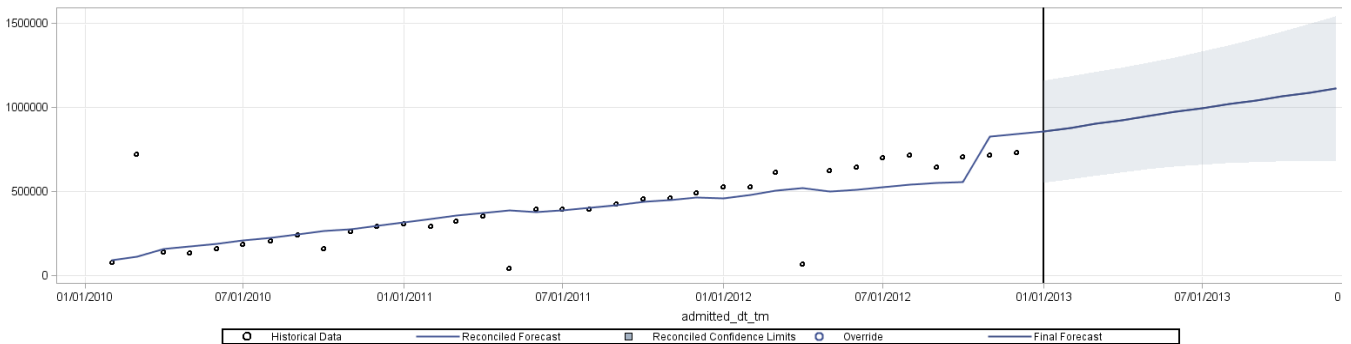
Unasyn spending (graph above) has been increasing gradually over the years. Also, this drug is the brand name for a generic name Ampicillin.
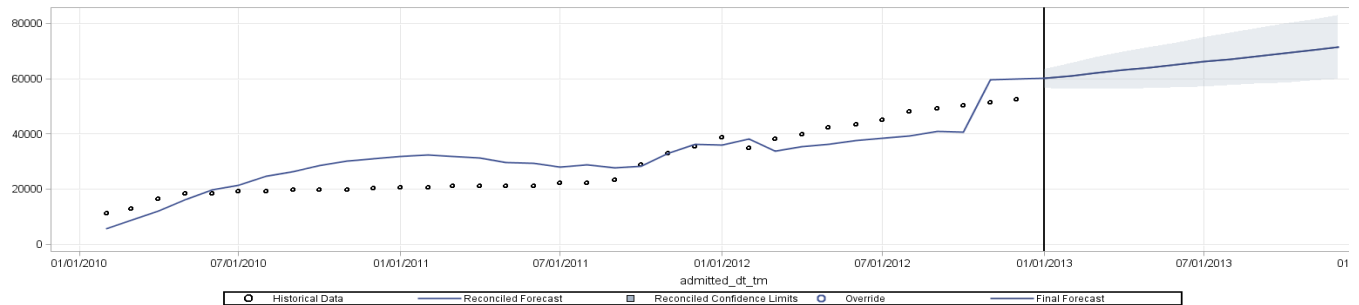
Zithromax has an increasing trend.

CONCLUSION

- We are not medical professionals and have no domain expertise. Therefore, our conclusions are purely based on patterns in the data and common knowledge obtained via Googling brand names and drug names.

- Females and people under the age group of 56-75 are more prone to having urinary tract infection and hence spend a lot on these antimicrobial drugs in comparison to others.

- Zithromax, being one of the many brands of the generic drug Azithromycin costs 7.16 times more than the generic drug. Replacing the brand name drug with the generic medicine may reduce the annual amount spent on treating urinary tract infections by $294 K.

- Similarly Unasyn, being one of the many brands of the generic drug Ampicillin sulbactam costs 2.56 times more than the generic drug. The brand name drug with the generic medicine may reduce the annual amount on treating urinary tract infections using antimicrobial drugs by $2 million.

FUTURE WORK AND LIMITATIONS

- Due to the constraints in data collected, we were unable to include the cost factor as time variant which is very essential in predicting the future trend. Incorporating time variant cost details can add great value to the model.

- If we get the data till 2015, then the results will be more accurate.

- We can find cases in which the drugs were effective and in which they were not. Along with this, we can find out if there were any similarity in the living habits of these patients which can be improved in the future.

REFERENCES

https://chsi.okstate.edu/
http://www.mayoclinic.org/diseases-conditions/urinary-tract-infection/basics/treatment/con-20037892
https://www.drugs.com/pro/unasyn.html
http://emedicine.medscape.com/article/231574-overview
http://www.scsug.org/SCSUGProceedings/2011/repalli1/Prediction%20on%20Diabetes%20Using%20Data%20mining%20Approach.pdf

http://www.sv-europe.com/crisp-dm-methodology/
http://www.ehealthme.com/ds/morphine%20sulfate/uti/
http://pharmacybook.net/furosemide/
http://www.ehealthme.com/ds/furosemide/urinary%20tract%20infection/

ACKNOWLEDGEMENT

CONTACT INFORMATION

Shweta Arya
Oklahoma State University
Email: shweta.arya@okstate.edu
Shweta Arya is a Masters student in Business Analytics at Spears School of Business, Oklahoma State University. She is currently working as Product Analysis and Research intern at CreditXpert Inc. Baltimore. She is a SAS certified Base Programmer 9 and SAS certified Advanced Programmer for SAS 9.