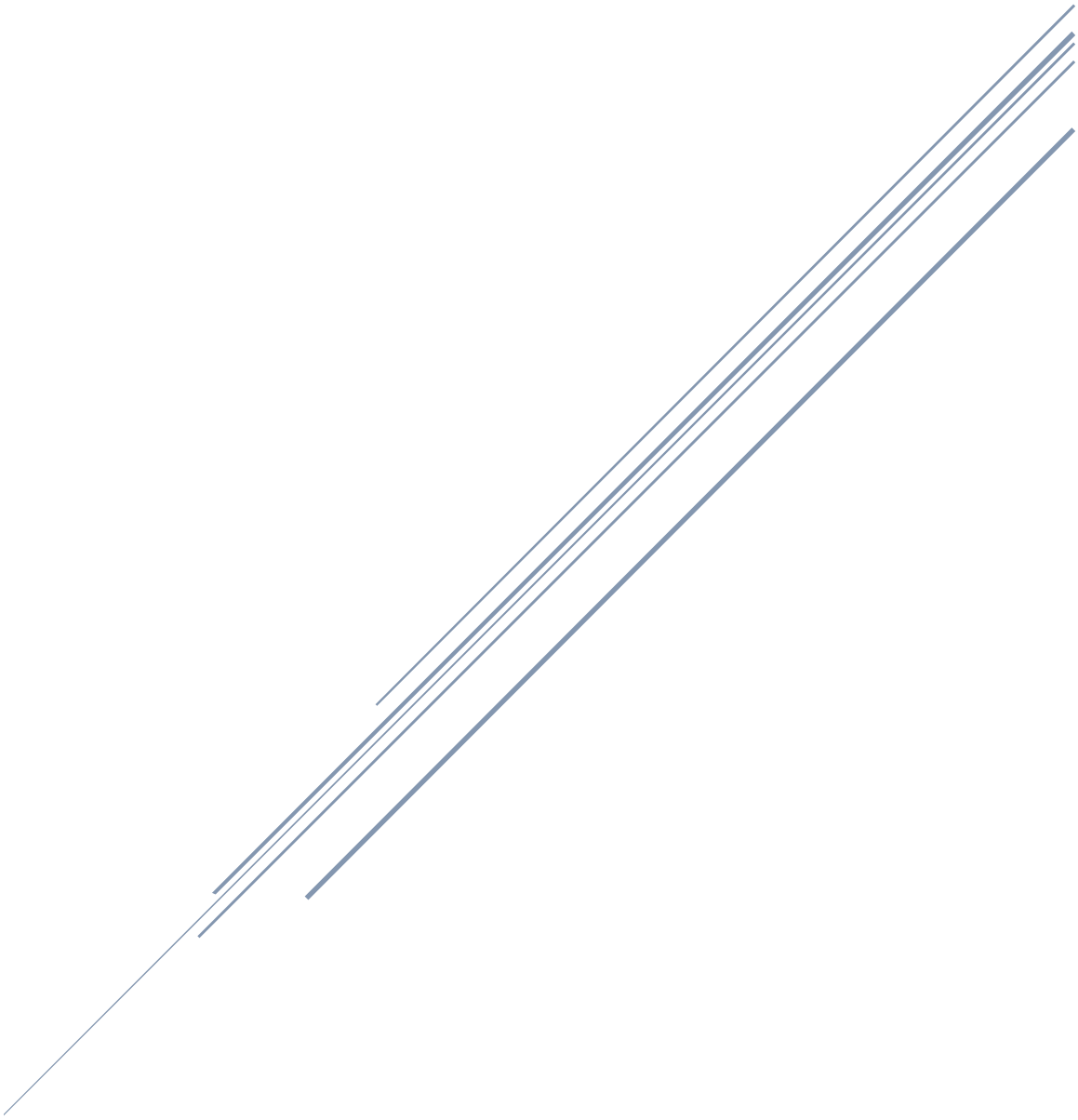
HERITAGE HEALTH PRIZE

Problem 1



Carpenter, Ryan
MGIS 355.01

-

Index

Problem Statement

a. More than 71 million individuals in the United States are admitted to hospitals each year, according to the latest survey from the American Hospital Association. Studies have concluded that in 2006 well over $30 billion was spent on unnecessary hospital admissions. Is there a better way? Can we identify earlier those most at risk and ensure they get the treatment they need? The Heritage Provider Network (HPN) believes that the answer is "yes".(*)

Mission Goal

b. The winning team will create an algorithm that predicts how many days a patient will spend in a hospital in the next year. Once known, health care providers can develop new care plans and strategies to reach patients before emergencies occur, thereby reducing the number of unnecessary hospitalizations. This will result in increasing the health of patients while decreasing the cost of care. In short, a winning solution will change health care delivery as we know it – from an emphasis on caring for the individual after they get sick to a true health care system.(*)

Mission Objectives

c. To achieve its goal of developing a breakthrough algorithm that uses available patient data to predict and prevent unnecessary hospitalizations, HPN is sponsoring the Heritage Health Prize Competition (the "Competition"). HPN believes that incentivized competition is the best way to achieve the radical breakthroughs necessary to begin fixing America's health care system.(*)

*Source: Heritage Health Prize. (2012, October 4). Retrieved from https://www.kaggle.com/c/hhp/

Key Variables

c.

a. Age	e. Frequency of visits
b. Gender	f. Days in Hospital
c. Procedure Group	g. Lab Count
d. Location	h. Drug Count

Key Relationships

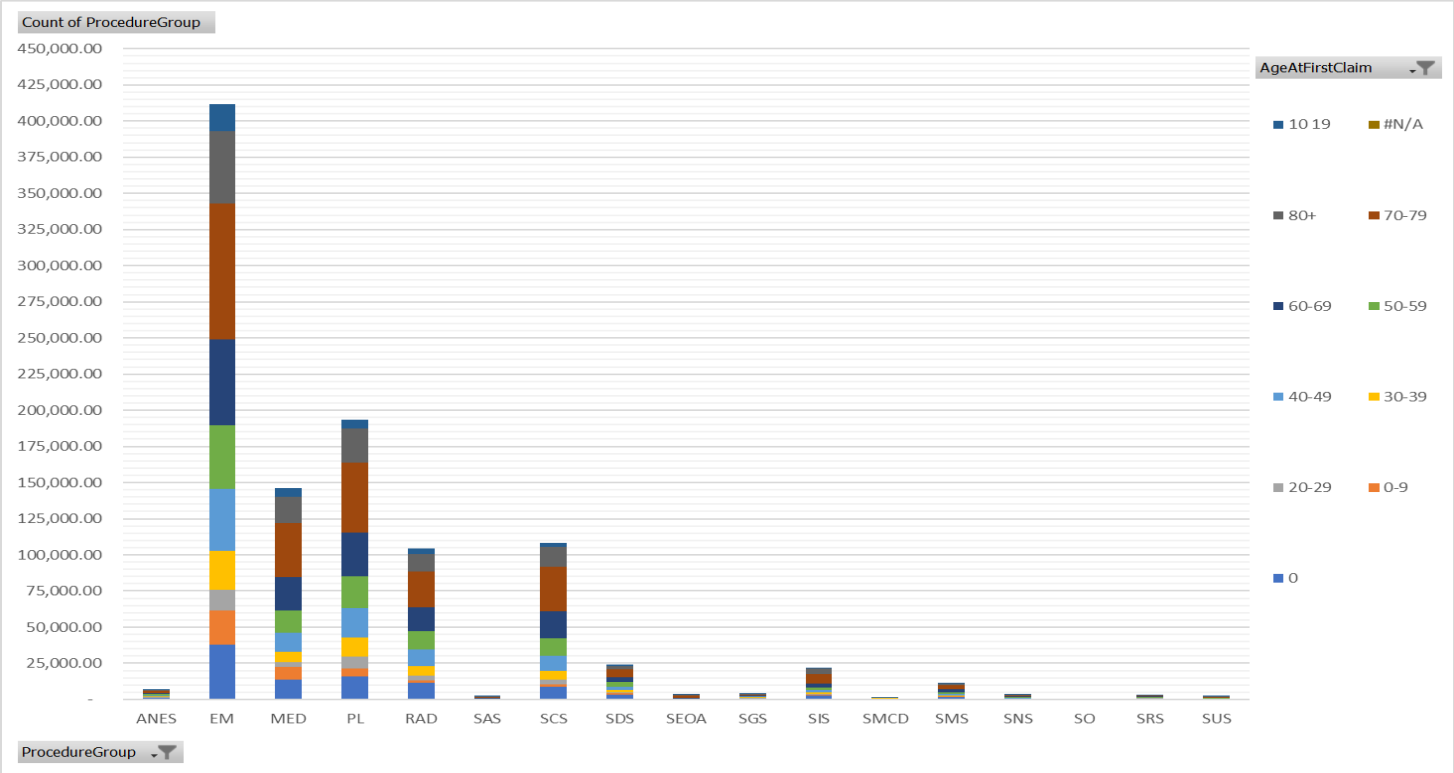
d.

i. Age vs Procedure	viii. Gender vs Lab Count
ii. Age vs. Location	ix. Count of Days in Hospital vs. Gender
iii. Count of Age at First Claim Ratio	x. Count of Days in Hospital by Age Range vs. Gender
iv. Age vs. Days in Hospital	xi. Count of Prescription per Gender
v. Age at First Claim vs. Lab Count	xii. Count by Gender vs. Age at First Claim
vi. Age vs. Prescription Percentage	xiii. Count of Primary Condition
vii. Gender Ratio	xiv. Count of Primary Conditions Served vs. Year

Graphical Displays

Key Variable: Age

i. Age vs. Procedure Count

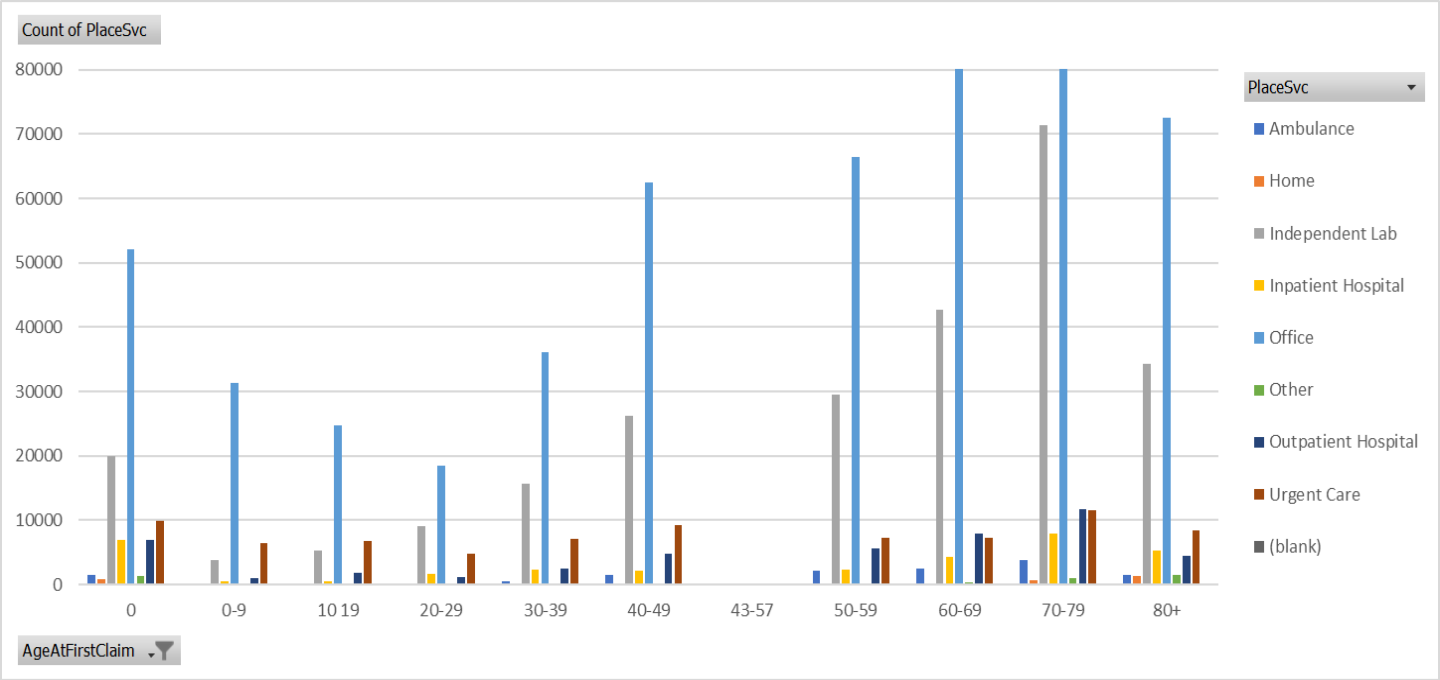


Count of ProcedureGroup	Age Range											
Procedure Name	0	0-9	20-29	30-39	40-49	50-59	60-69	70-79	80+	#N/A	10 19	Grand Total
ANES	1,114.00	103.00	353.00	566.00	740.00	670.00	831.00	1,425.00	591.00		168.00	6,561.00
EM	38,136.00	23,353.00	14,473.00	26,958.00	42,920.00	43,718.00	59,635.00	93,785.00	50,101.00	12.00	18,569.00	411,660.00
MED	13,741.00	8,815.00	3,441.00	6,906.00	13,305.00	15,092.00	23,154.00	37,366.00	18,197.00	2.00	6,404.00	146,423.00
PL	16,042.00	5,259.00	8,074.00	13,242.00	20,495.00	22,024.00	30,072.00	48,781.00	23,686.00	4.00	5,577.00	193,256.00
RAD	11,475.00	1,895.00	3,266.00	6,505.00	11,563.00	12,350.00	16,522.00	25,081.00	11,779.00	1.00	3,862.00	104,299.00
SAS	267.00	113.00	30.00	57.00	115.00	189.00	258.00	675.00	467.00		67.00	2,238.00
SCS	8,854.00	1,796.00	3,071.00	5,793.00	10,753.00	12,250.00	18,259.00	30,724.00	14,240.00	3.00	2,306.00	108,049.00
SDS	3,256.00	1,045.00	849.00	1,443.00	2,391.00	3,123.00	3,489.00	5,119.00	2,597.00		927.00	24,239.00
SEOA	283.00	8.00	9.00	31.00	77.00	183.00	544.00	1,353.00	796.00		6.00	3,290.00
SGS	698.00	19.00	265.00	455.00	546.00	287.00	280.00	609.00	592.00		35.00	3,786.00
SIS	2,660.00	354.00	733.00	1,068.00	1,475.00	1,666.00	3,046.00	6,353.00	4,115.00		687.00	22,157.00
SMCD	123.00		560.00	517.00	60.00						50.00	1,310.00
SMS	1,721.00	227.00	170.00	503.00	1,091.00	1,323.00	1,812.00	2,729.00	1,408.00		488.00	11,472.00
SNS	522.00	7.00	53.00	179.00	365.00	426.00	530.00	691.00	264.00		23.00	3,060.00
SO	54.00	7.00	1.00	8.00	17.00	17.00	10.00	17.00	9.00			140.00
SRS	644.00	76.00	92.00	160.00	348.00	324.00	462.00	633.00	187.00		133.00	3,059.00
SUS	479.00	41.00	23.00	75.00	155.00	183.00	277.00	617.00	280.00		22.00	2,152.00
Grand Total	**100,069.00**	**43,118.00**	**35,463.00**	**64,466.00**	**106,416.00**	**113,825.00**	**159,181.00**	**255,958.00**	**129,309.00**	**22.00**	**39,324.00**	**1,047,151.00**

Procedure	Description
ANES	Anesthesia
EM	Evaluation and Management
MED	Medicine
PL	Pathology and Laboratory
RAD	Radiology
SAS	Surgery-Auditory System
SCS	Surgery-Cardiovascular System
SDS	Surgery-Digestive System
SEOA	Surgery-Eye and Ocular Adnexa
SGS	Surgery-Genital System
SIS	Surgery-Integumentary System
SMCD	Surgery-Maternity Care and Delivery
SMS	Surgery-Musculoskeletal System
SNS	Surgery-Nervous System
SO	Surgery-Other
SRS	Surgery-Respiratory System
SUS	Surgery-Urinary System

THIS GRAPHIC DISPLAYS THE RELATIONSHIP BETWEEN AGE AND THE COUNT OF PROCEDURES EXPERIENCED. HERE WE CAN OBSERVE THE "EVALUATION & MANAGEMENT" CATEGORY CONTAINING THE HIGHEST CONTENT OF ALL AGE RANGES. THE DATA DISPLAYS THAT 70-79YR OLDS CONTAIN THE HIGHEST CONTENT OF EM, PL, RAD, SCS, SDS, SEOA, SIS, AND SMS PROCEDURES

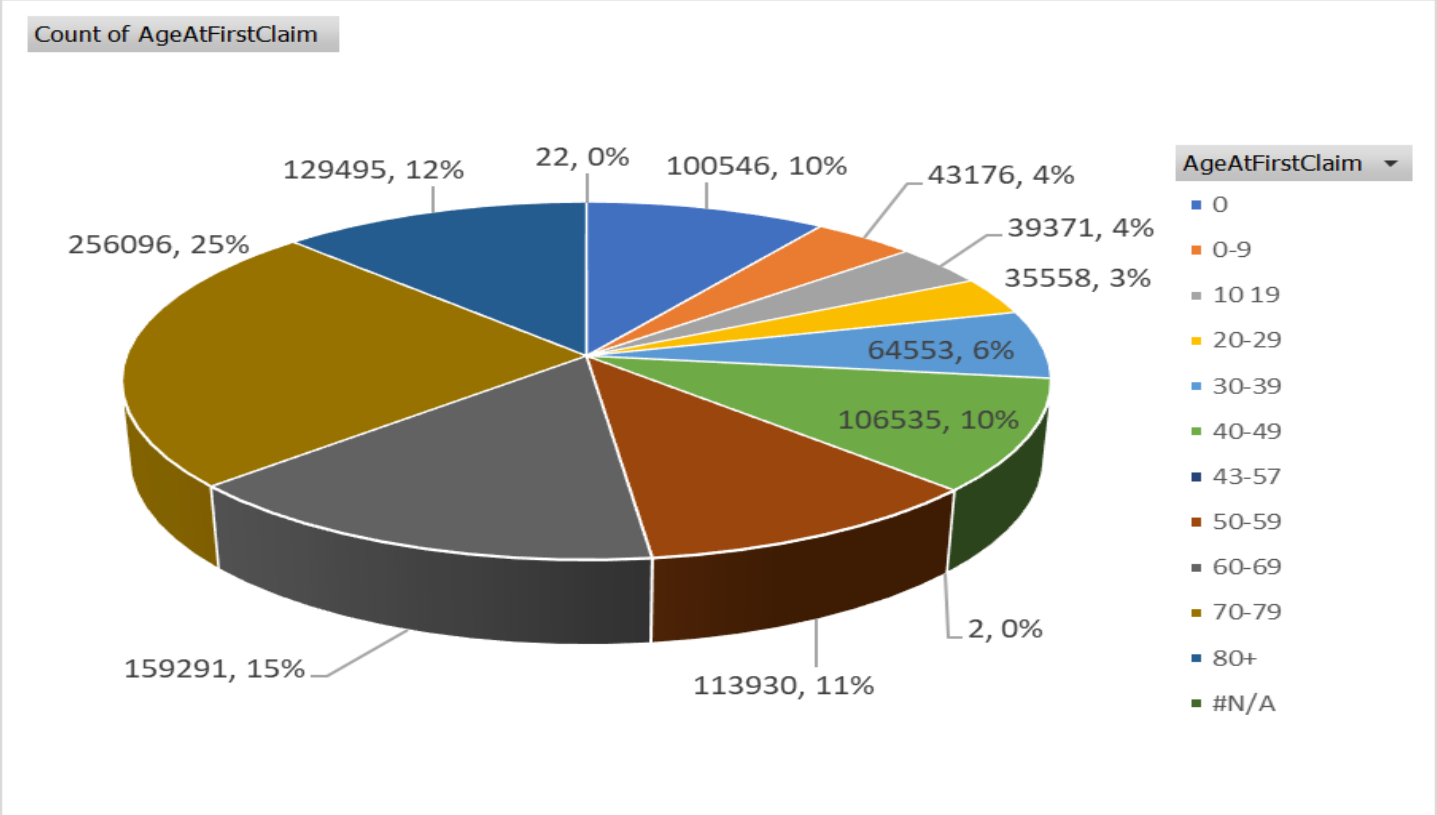
ii. Age vs. Location



Count of PlaceSvc	Service Locatio								
Age	Ambulance	Home	Independent Lab	Inpatient Hospital	Office	Other	Outpatient Hospital	Urgent Care	(blank)	Grand Total
0	1566	815	19993	6896	52111	1258	6865	9943		99447
0-9	104	7	3869	415	31238	1	987	6445		43066
10 19	147	6	5320	429	24690	15	1833	6796		39236
20-29	195	4	9059	1683	18400	21	1188	4860		35410
30-39	535	25	15726	2314	36070	52	2440	7152		64314
40-49	1437	64	26210	2153	62384	145	4710	9217		106320
43-57					2					2
50-59	2086	126	29490	2380	66498	247	5621	7290		113738
60-69	2434	191	42729	4213	94055	292	7845	7254		159013
70-79	3736	737	71361	7908	147735	993	11750	11597		255817
80+	1471	1316	34226	5344	72571	1529	4396	8330		129183
Grand Total	**13711**	**3291**	**257983**	**33735**	**605754**	**4553**	**47635**	**78884**		**1045546**

THIS GRAPHIC DISPLAYS THE RELATIONSHIP BETWEEN AGE AND THE LOCATIONS THESE PATIENTS ARE VISITING. ACROSS ALL AGE RANGES IT CAN BE OBSERVED THAT THE "OFFICE" LOCATION IS SUBJECT TO THE HIGHEST LEVEL OF PATIENTS. THIS TREND WITHIN THE DATA CAN BE INTERPRETED AS A POTENTIAL POSITIVE SIGN AS BOTH INPATIENT AND OUTPATIENT HOSPITAL COUNTS REMAIN AT THE LOWER LEVEL OF THE GRAPHIC.THE SECOND LOCATION SUBJECT TO HIGHER PATIENT DEMANDS IS THE INDEPENDENT LABS.

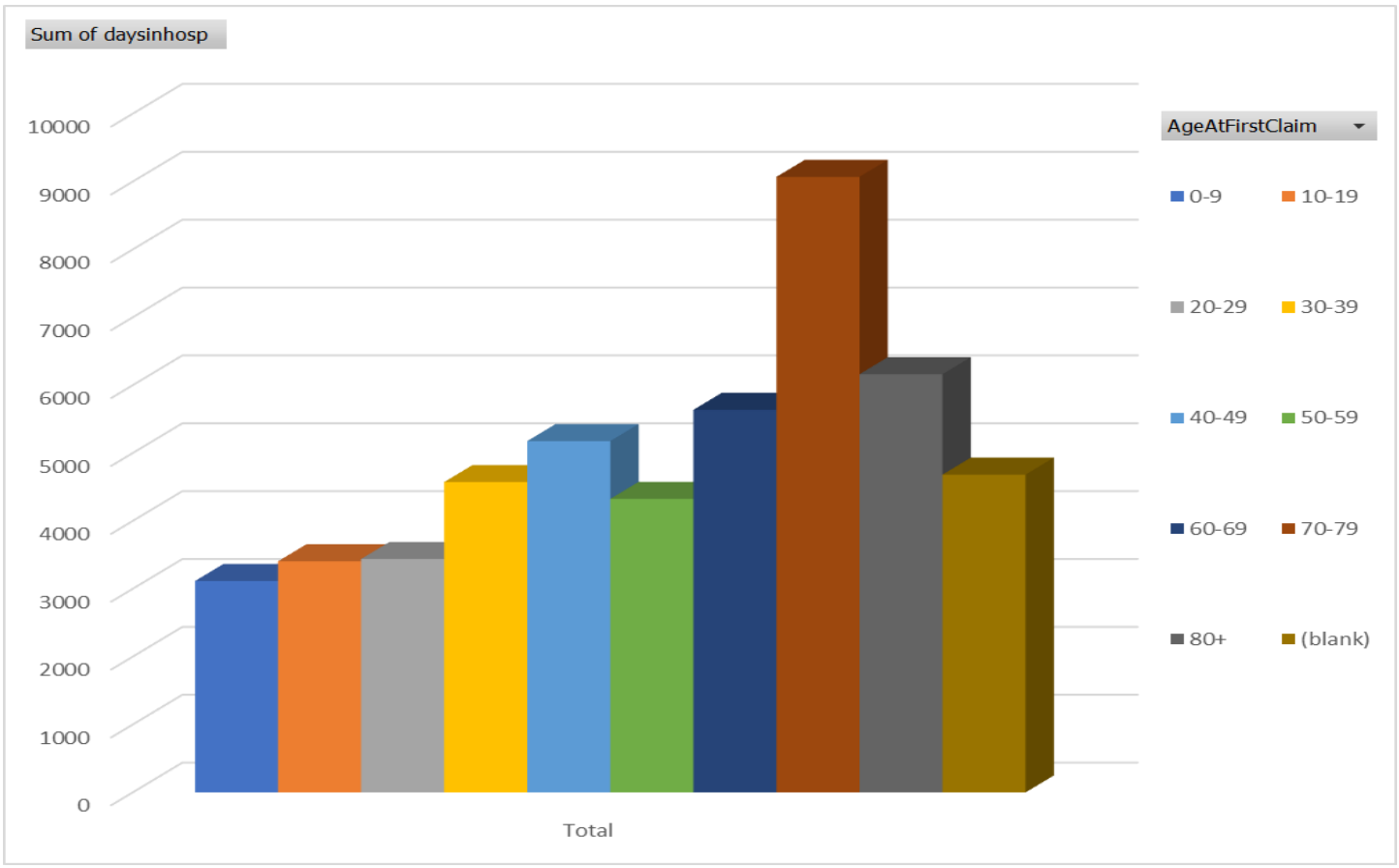
iii. Count of Age at First Claim Ratio



Age	Count of AgeAtFirstClaim
0	100546
0-9	43176
10 19	39371
20-29	35558
30-39	64553
40-49	106535
43-57	2
50-59	113930
60-69	159291
70-79	256096
80+	129495
#N/A	22
Grand Total	**1048575**

THIS PIE CHART REPRESENTS THE COUNT AND PERCENTAGE OF AGE AT FIRST CLAIM. THIS CHART SHOWS THAT 40% OF PATIENTS ARE WITHIN THE AGES 60-69 OR 70-79. CONVERSELY, YOUNGER GENERATIONS SUCH AS 10-19 YEAR OLDS OR 20-29 YEAR OLDS EACH ACCOUNTED FOR LESS THAN 10% OF THE PATIENT DATA.

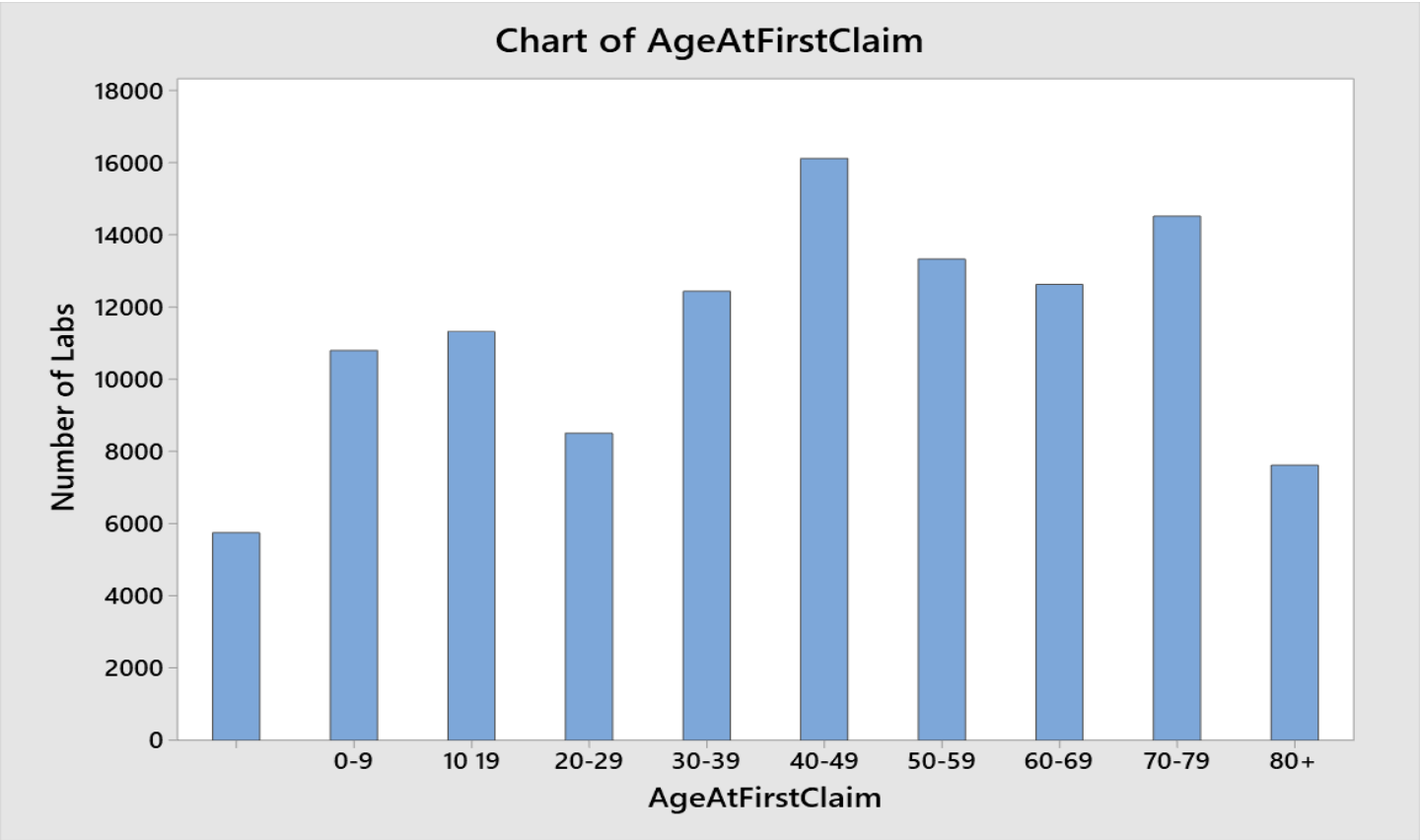
iv. Age vs Days in hospital



	Age										
	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80+	(blank)	Grand Total
Sum of daysinhosp	3115	3406	3440	4573	5177	4325	5637	9070	6162	4683	49588

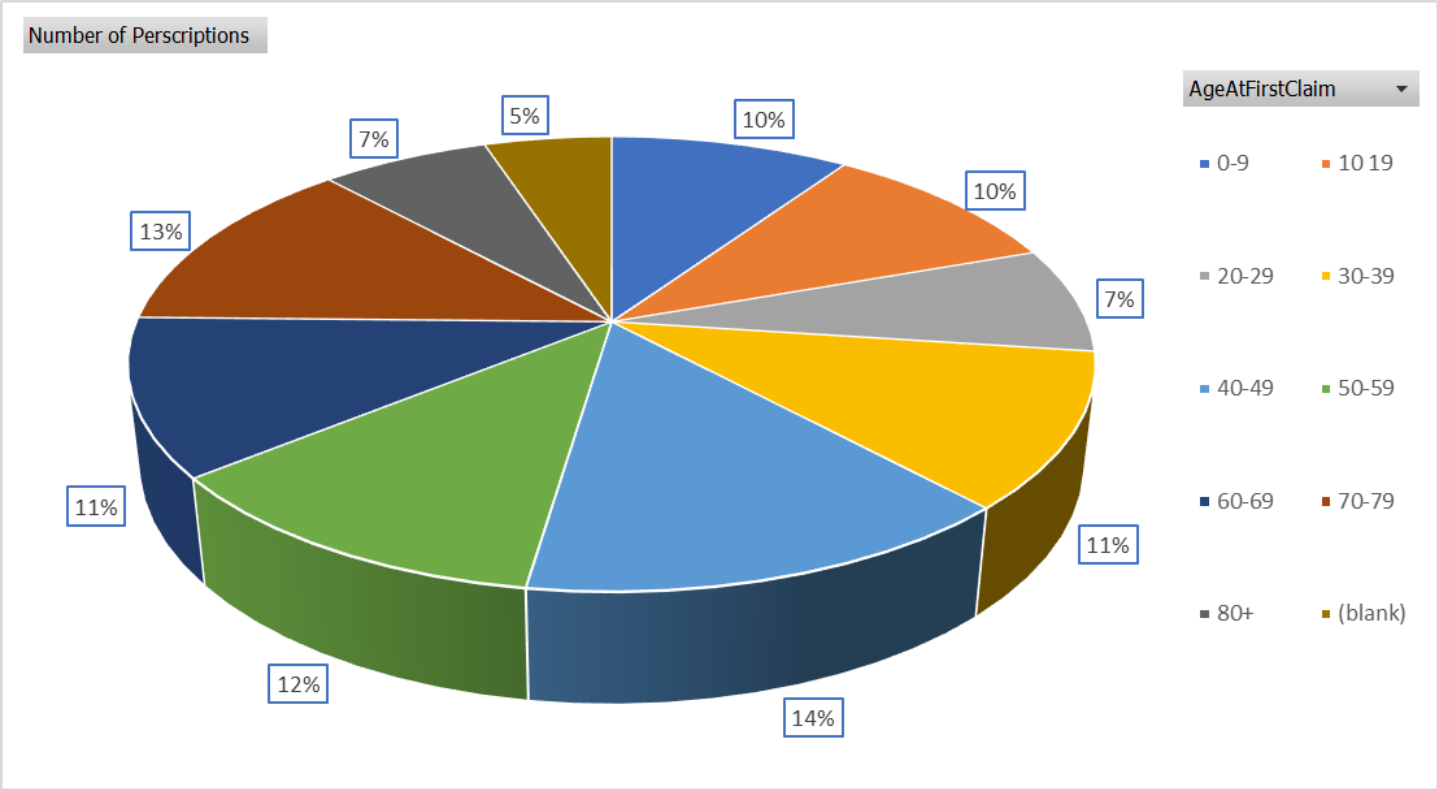
THE GRAPHIC ABOVE DISPLAYS THE PATIENT AGE RANGE VS THE SUM OF DAYS SPENT IN THE HOSPITAL. HERE WE CAN OBSERVE THAT 70-79 YEARS OF AGE ONCE AGAIN RATED HIGHEST AMONG DAYS SPENT IN THE HOSPITAL.

v. Age at First Claim vs. Lab Count



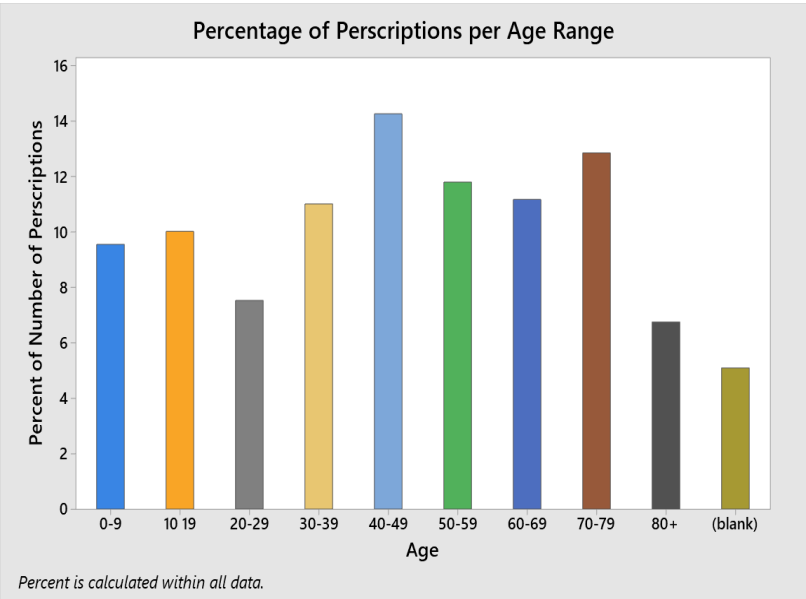
THIS CHART DISPLAYS THE NUMBER OF LABS PROVIDED BY THE GENERATIONAL AGE RANGES. IT CAN BE OBSERVED THAT THE AGE RANGE 40-49 CONTAINS THE HIGHEST NUMBER OF LABS PROVIDED. 70-79 YEARS OF AGE FOLLOWS WITH THE SECOND HIGHEST VALUE, THIS COMES AS A BIT OF A SURPRISE AS THE AGE RANGE OF 70-79 WAS HIGHEST IN DAYS IN HOSPITAL AND SECOND HIGHEST IN INDEPENDENT LAB UTILIZATION DISPLAYED ON PAGE 5.

vi. Age vs Prescription Percentage



Age	Number of Perscriptions
0-9	10791
10 19	11319
20-29	8505
30-39	12435
40-49	16111
50-59	13329
60-69	12622
70-79	14514
80+	7621
(blank)	5753
Grand Total	**113000**

THE PIE CHART ABOVE PROVIDES A VISUAL REPRESENTATION OF AGE RANGE VS THE PERCENTAGE OF PRESCRIPTIONS WRITTEN. HERE WE CAN OBSERVE A REALITIVELY EVEN DISTRIBUTION WITH 40-49 YEAR OLDS CONSUMING THE MOST MEDICATIONS THROUGHOUT THE STUDY.

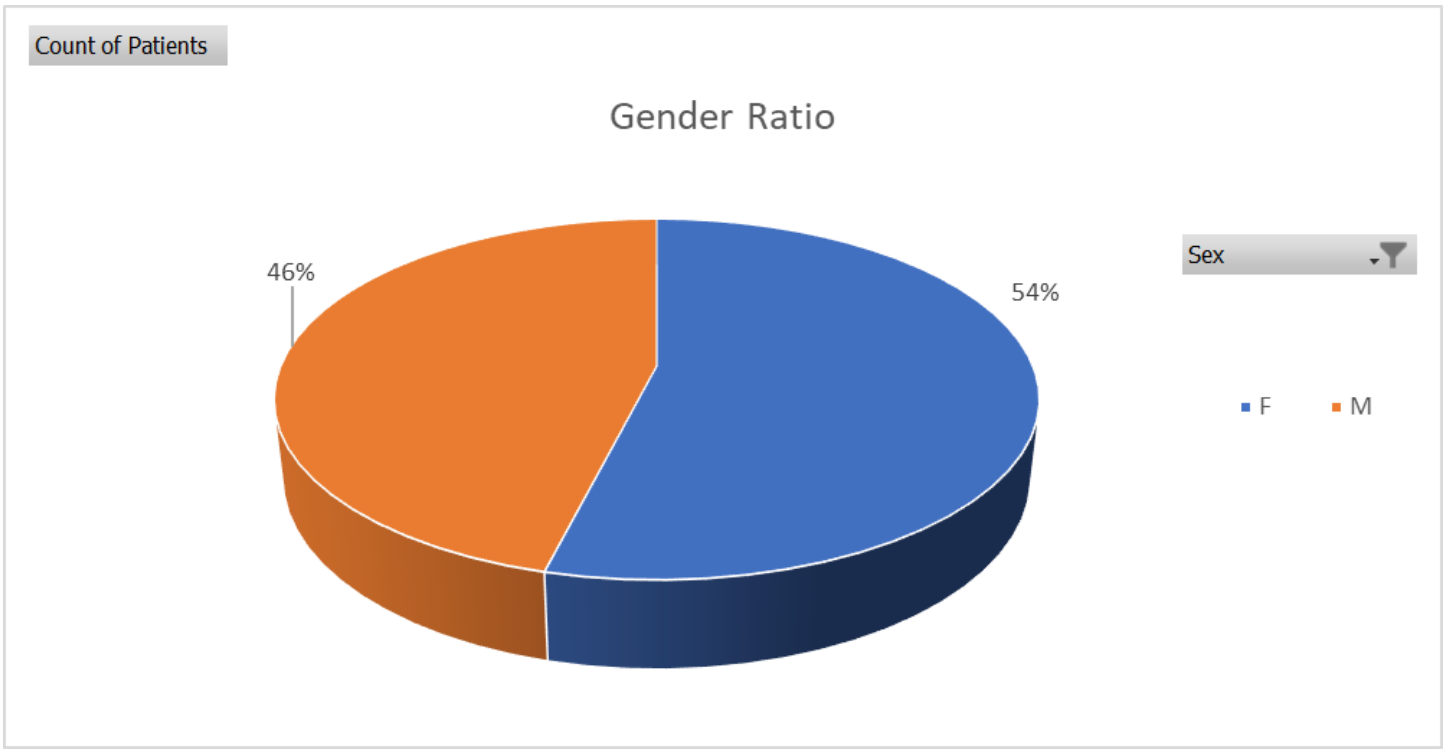


THIS BAR CHART DISPLAYS THE SAME DATA AS ABOVE. HERE IT IS RIENFORCED THAT 40-49 YEAR OLDS CONSUMER THE HIGHEST NUMBER OF PRESCRIPTIONS. 70-79 YEAR OLDS RANK SECOND AMONG THE POPULATION.

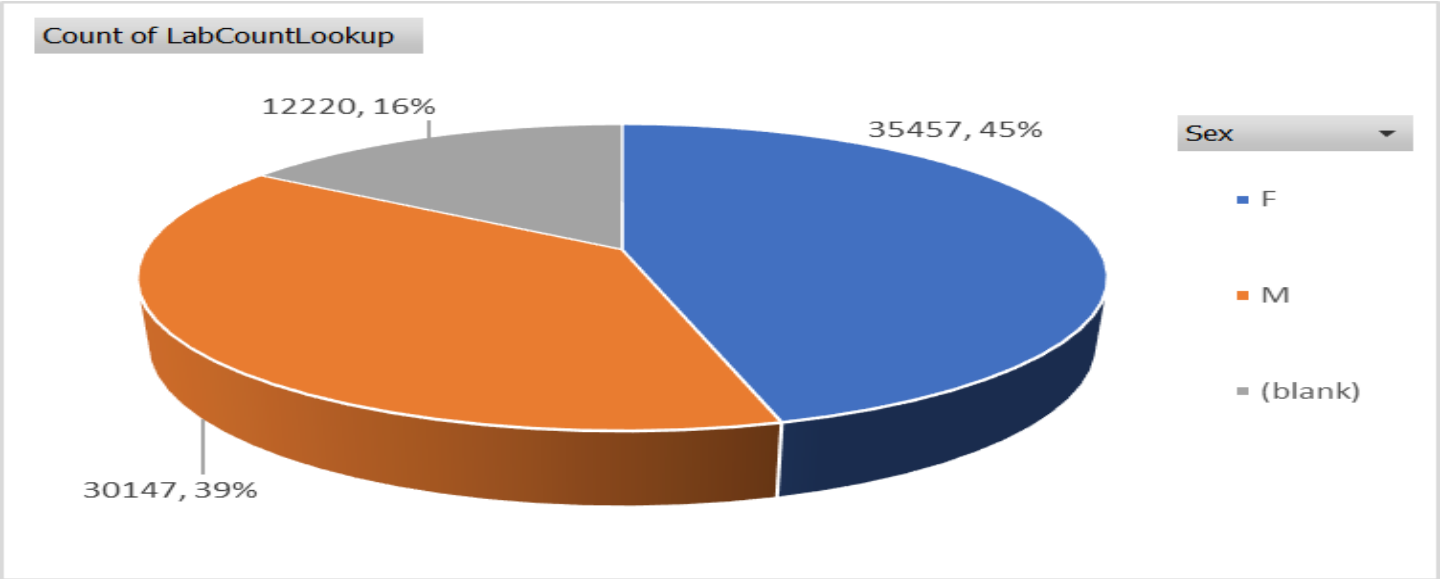
Key Variable: Gender

vii. Gender Ratio

THE PIE CHART BELOW PROVIDES A BASIC PERCENTAGE IN REGARD TO GENDER. HERE WE CAN OBSERVE THAT THE DATA IS BROKEN UP BETWEEN 54% FEMALE AND 46% MALE WITH DISREGARD TO INCORRECT RESEARCH DATA ENTRY.

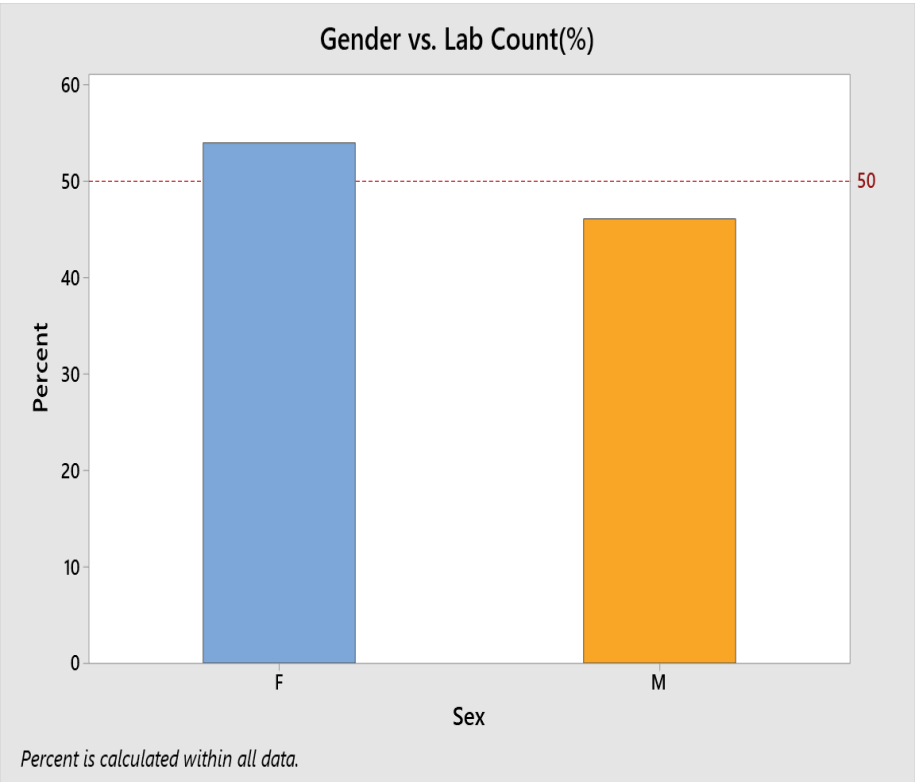


viii. Gender vs Lab Count



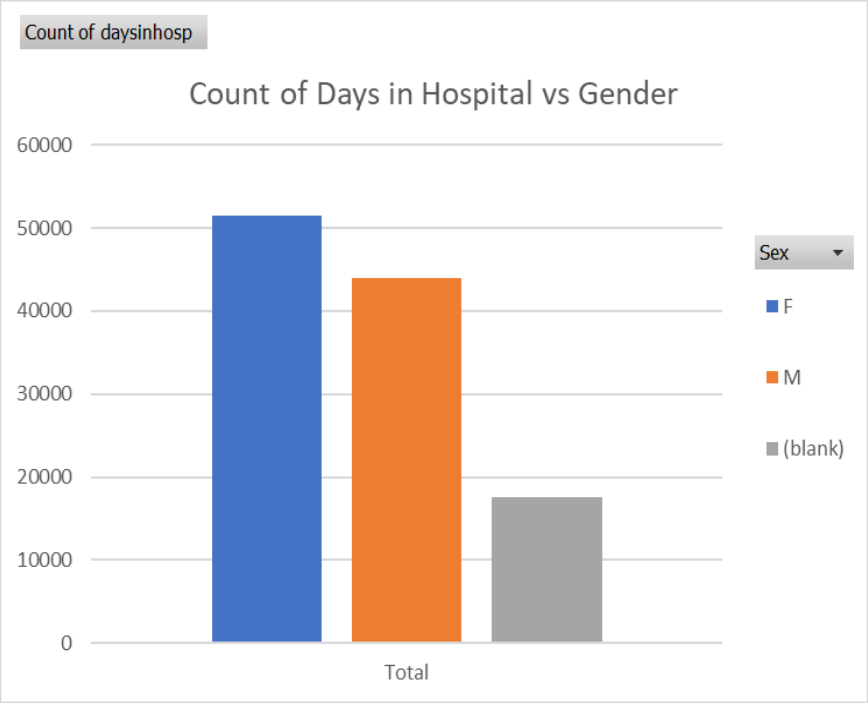
Row Labels	Count of LabCountLookup
F	35457
M	30147
(blank)	12220
Grand Total	77824

THE ABOVE PIE CHART DISPLAYS BOTH THE COUNT AND POPULATION PERCENTAGE OF GENDERS VS LAB COUNTS. WITHIN THE VISUAL IT CAN BE OBSERVED THAT THE FEMALE SEX ACCOUNTS FOR ROUGHLY 45% OF THE LAB WORK ISSUED. WITHIN THIS CHART WE HAVE BLANK DATA DUE TO INCORRECT RESEARCH DATA ENTRY.

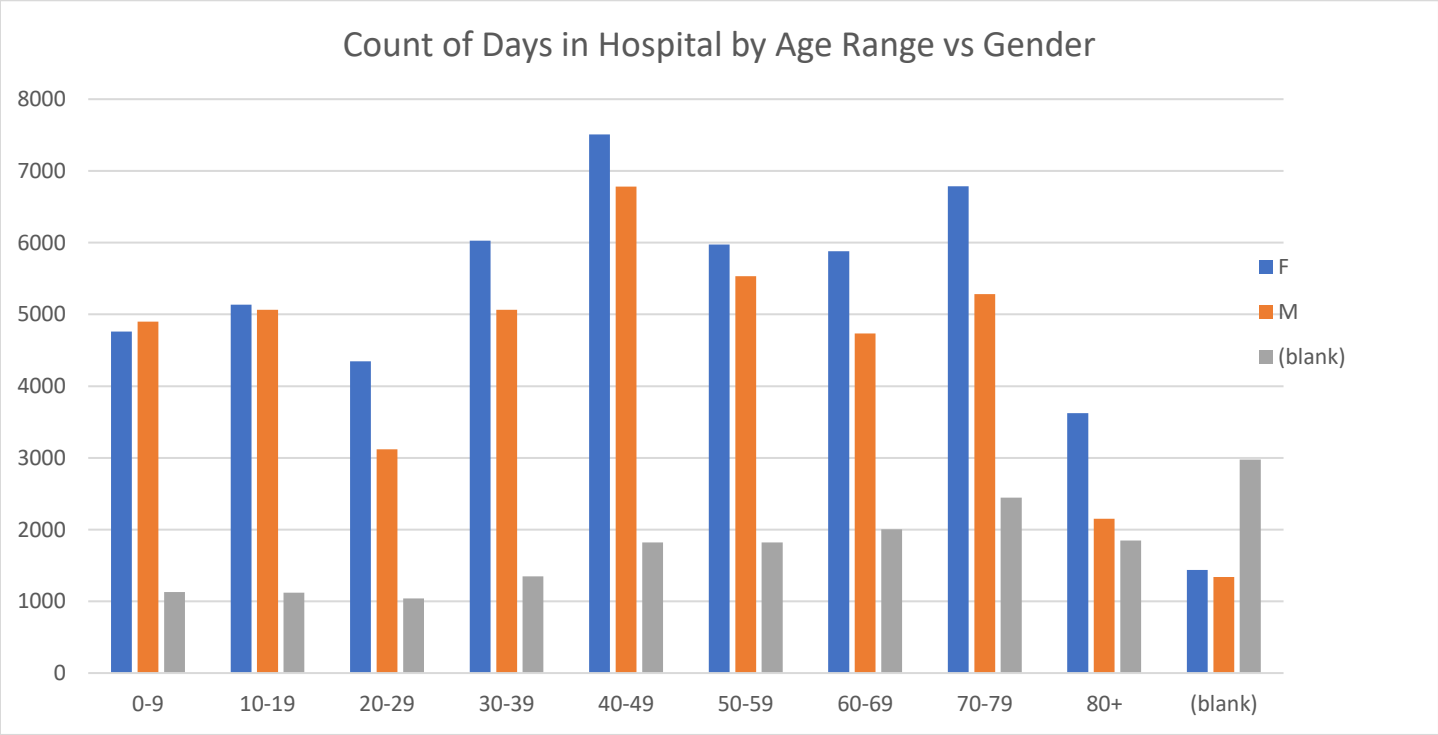


THE BAR CHART TO LEFT DISPLAYS THE CHANGE IN PERCENTAGE ONCE THE "(BLANK)" DATA IS REMOVED. HERE IT CAN BE OBSERVED THAT BY REMOVING THE INCCORRET DATA, THE FEMALE GENDER ACCOUNTS FOR MORE THAN 50% OF THE LAB WORK ISSUED.

ix. Count of Days in Hospital vs Gender

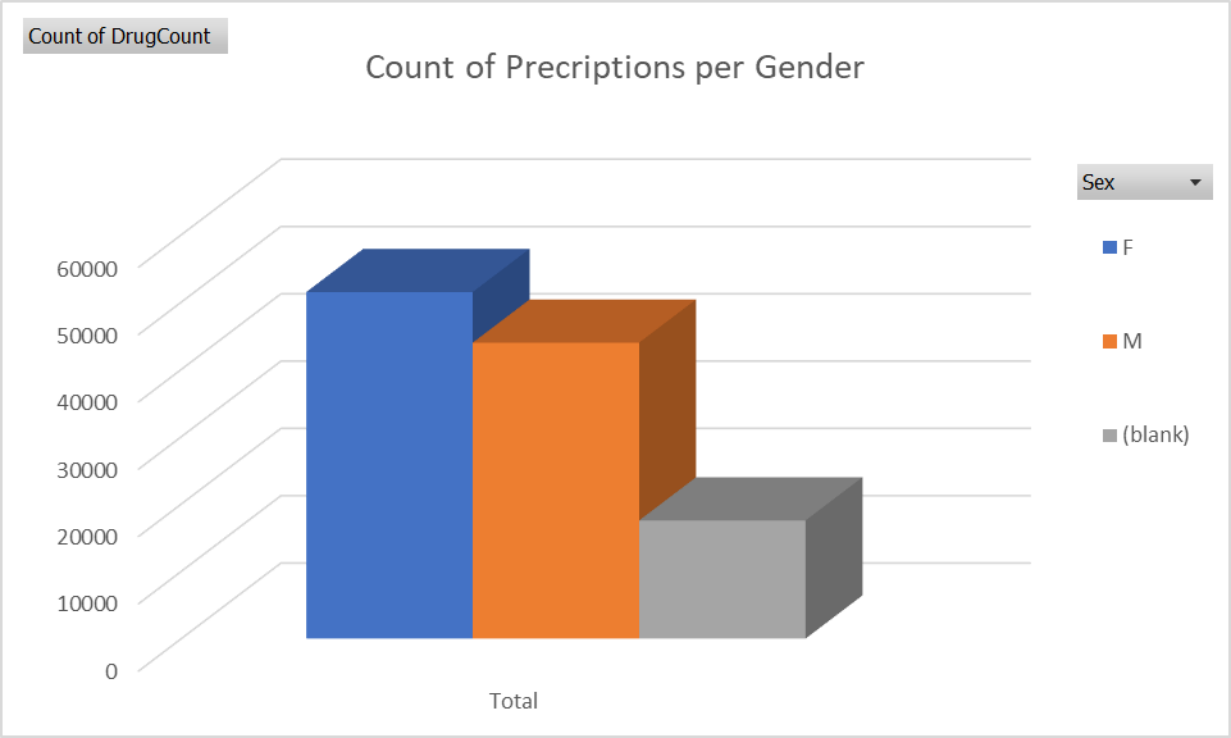


THIS BARCHART DISPLAYS THE COUNT OF DAYS IN HOSPITAL VS GENDER. WE CAN CONCLUDE THAT FEMALE PATIENTS HAVE SPENT MORE DAYS IN THE HOSPITAL THAN THE MALE SEX.

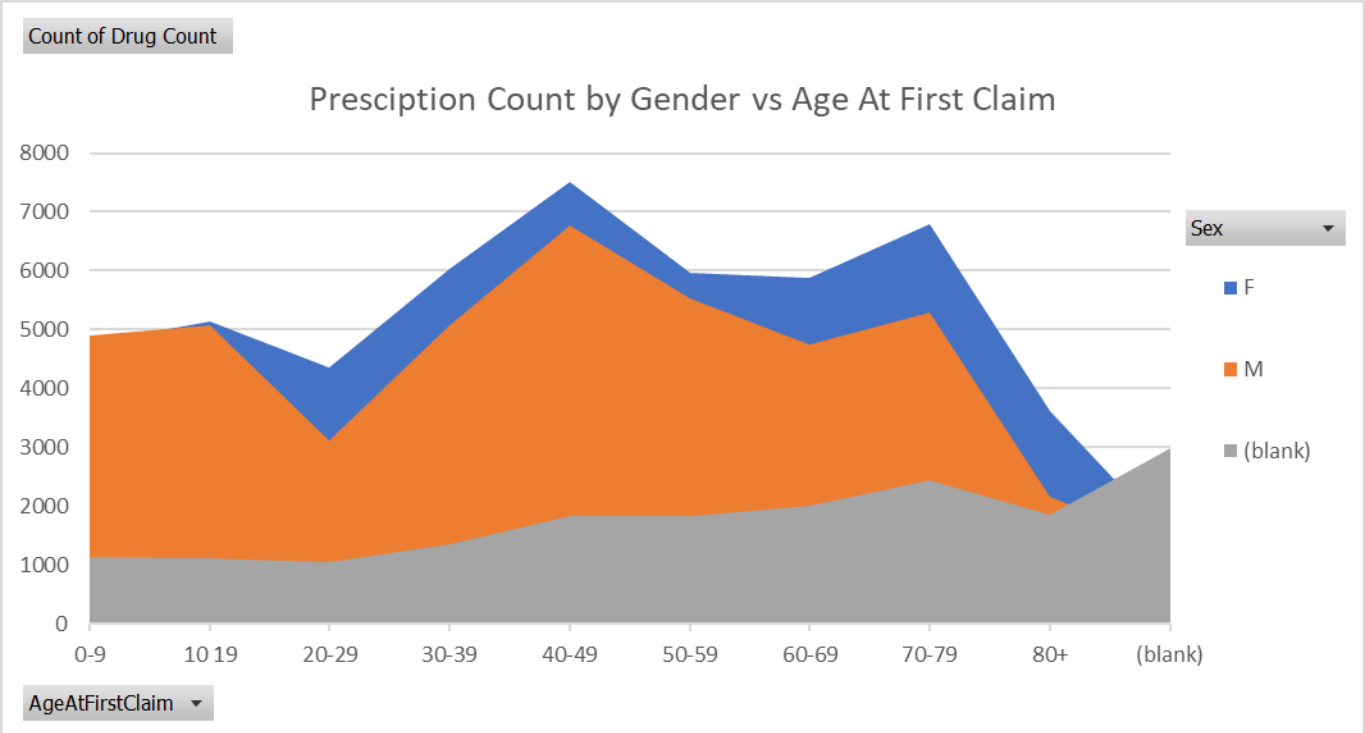


THE BARCHART ABOVE IS AN EXPANSION OF THE INITIAL CHART ON THIS PAGE. HERE WE BREAK DOWN THE DAYS IN HOSPITAL BY THE AGE RANGES AND GENDER. AS EXPECTED, THE DATA IS CONSISENT WITH ABOVE IN DISPLAYING THAT FEMALE PATIENTS SPEND LONGER TIMES IN THE HOSPITAL. THIS DISPLAY SHOWS THAT THE AGE RANGES OF 30-39, 40-49, AND 70-79 ALL CONTAIN THE HIGHEST VALUES OF FEMAL HOSPITAL STAY.

xi. Count of Prescriptions per Gender



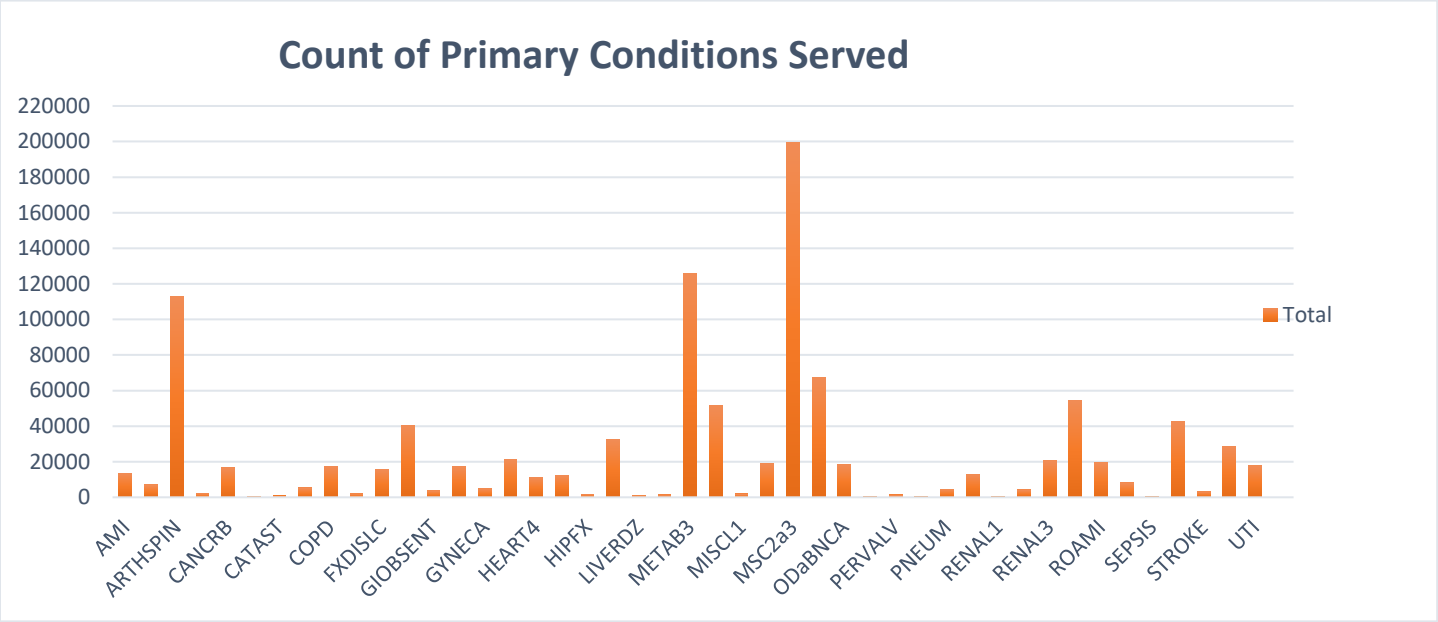
xii. Prescription Count by Gender vs. Age at First Claim



GRAPHIC 1 DISPLAYS THE TOTAL COUNT OF PRESCRIPTIONS PER GENDER. IT CAN BE OBSERVED THAT THE FEMALE SEX CONSUMES MORE PRESCRIPTIONS THAN MEN.

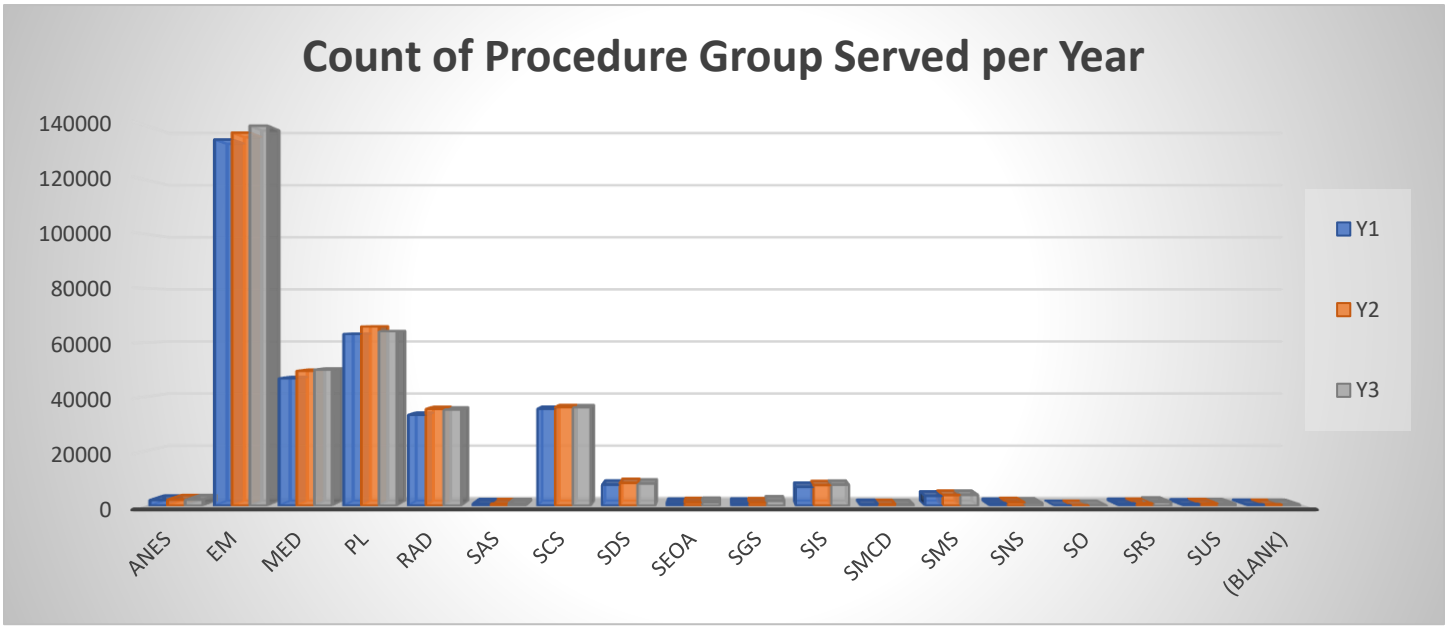
THE AREA CHART ABOVE GIVE GRAPHICAL REPRESENTATION TO THE INCREASES IN PRESCRIPTIONS PER AGE RANGE. THIS GRAPHIC ALSO SUPPORTS THE ASSUMPTIONFROM GRAPHIC 1.

xiii. Count of Primary Conditions Served



THE ABOVE CHART PROVIDES REFERENCE TO THE COUNT OF PRIMARY CONDITIONS SERVED. HERE WE CAN OBSERVE THAT THE MISCELLANEOUS 2 AND OTHER METABLIC PROCEDURES ARE THE MOST COMMON.

xiv. Count of Procedure Group Served vs. Year



THE ABOVE GRAPHIC DISPLAYS THE COUNT OF PROCEDURE GROUP SERVED CATEGORIZED BY YEAR. HERE IT CAN OBSERVED THAT THROUGHOUT THE THREE YEARS OF THE STUDY EVALUATION AND MANAGEMENT PROCEDURES REMAINED THE HIGHEST.

Conclusion

Through initial evaluation it can be observed that the highest deficiencies are coming from the amount of miscellaneous procedures and extremely high yearly rate of evaluation and management procedures. These two main categories are more than likely costing the health care industry the highest amounts of money and resource allocation.

Secondly, it can be concluded that the female sex requires more services than men. With this conclusion it is understandable that women would maintain a higher amount of prescriptions, labs, and days in hospital.

We can also conclude that once patients begin to transition out of their 30's they fall into a "high risk" age bracket. Within this "high risk" evaluation, the data displays that these patients require a higher level of treatment across all healthcare treatments.

After initial research, the data does not necessarily surprise in any way. The distribution of prescriptions seems normal and tendencies within age range across various tests show consistencies throughout all tests. With that being said, it is expected that elderly and female patients require more attention than younger or less frequent patients.

Upon concluding my research, I believe the best way to combat this issue of unnecessary hospital admission would be to conduct even more detailed data discovery. The utilization of a race data entry may provide socio-economic information that may identify possible trends to better predict a patient's health requirements as they age. Secondly, the development of an "early warning" system may be beneficial in identifying procedure groups or primary condition groups that will lead to higher prescription, lab, and hospital stay requirements.