Healthcare

Advancing Healthcare Analysis through Data Insights

As a healthcare analyst at HealthStat Solutions, my goal is to use Power BI to dive deep into healthcare data. This means cleaning up the data and organising it in a way that makes sense. With advanced analytics tools like DAX, I'll create an interactive dashboard that tells a clear story about the data. This dashboard will help us see important trends, like how patient characteristics affect treatment outcomes and the costs of different medical procedures. By analyzing this data, we can help healthcare providers improve patient care and run hospitals more efficiently, putting HealthStat Solutions at the forefront of healthcare analytics.

Part-1:-

Data Importing and Initial Examination

During the initial examination, we found some issues in the data that needed fixing:

For Dataset1:Patient Medical Records

Missing Patient Names: Some records didn't have names for patients, so we needed to clean up the data to fill in these missing names.
Date Confusion: The dates when patients were admitted and discharged weren't in a consistent format. We fixed this so that all dates are clear and consistent, making sure our data is accurate.
Medication Details: The column with information about the medicines prescribed to patients was very detailed. To make sense of it, we needed to pick out the important parts for our analysis.

PatientID		PatientName		Age	
Valid	100%	Valid	0%	Valid	100%
Error	0%	Error	0%	Error	0%
Empty	0%	Empty	100%	Empty	0%
2		null		82	
8		null		13	
22		null		62	
31		null		12	
37		null		33	
65		null		47	
77		null		91	
131		null		79	
143		null		69	
149		null		61	
154		null		75	
160		null		25	
182		null		42	
240		null		63	
270		null		76	
311		null		85	
313		null		70	
316		null		29	
319		null		76	
351		null		31	
380		null		5	
394		null		53	
402		null		72	
404		null		5	

AdmissionDate		DischargeDate	
Valid	100%	Valid	100%
Error	0%	Error	0%
Empty	0%	Empty	0%
44198		44199	
44204		44205	
44218		44219	
44227		44228	
44233		44234	
44261		44262	
44273		44274	
44327		44328	
44339		44340	
44345		44346	
44350		44351	
44356		44357	
44378		44379	
44436		44437	
44466		44467	
44507		44508	
44509		44510	
44512		44513	
44515		44516	
44547		44548	
44576		44577	
44590		44591	
44598		44599	
44600		44601	

Full Prescription Details	
Valid	100%
Error	0%
Empty	0%
Losartan 50mg, twice a day for 7 days; Amoxicillin 500mg, twice a day ...	
Ibuprofen 400mg, twice a day for 7 days; Prednisone 20mg, twice a da...	
Tramadol 50mg, twice a day as needed; Doxycycline 100mg, once a da...	
Levothyroxine 50mcg, three times a day for 10 days; Meloxicam 15mg,...	
Amlodipine 5mg, three times a day for 10 days; Furosemide 40mg, thr...	
Tramadol 50mg, three times a day for 7 days; Fluoxetine 20mg, three t...	
Duloxetine 60mg, three times a day for 10 days; Simvastatin 40mg, twi...	
Pantoprazole 40mg, three times a day as needed; Losartan 50mg, thre...	
Hydrochlorothiazide 25mg, three times a day for 10 days; Warfarin 5m...	
Gabapentin 300mg, twice a day for 5 days; Pantoprazole 40mg, once a...	
Insulin Glargine 100 units/mL, twice a day for 10 days; Furosemide 40...	
Doxycycline 100mg, once a day as needed; Fluoxetine 20mg, three tim...	
Hydrochlorothiazide 25mg, once a day for 5 days; Prednisone 20mg, o...	
Azithromycin 250mg, twice a day as needed; Clopidogrel 75mg, three t...	
Amlodipine 5mg, once a day for 5 days; Ciprofloxacin 500mg, once a d...	
Clonazepam 2mg, twice a day for 5 days; Hydrochlorothiazide 25mg, t...	
Duloxetine 60mg, twice a day as needed; Omeprazole 40mg, once a da...	
Fluoxetine 20mg, once a day as needed; Azithromycin 250mg, once a ...	
Ciprofloxacin 500mg, twice a day as needed; Doxycycline 100mg, once...	
Escitalopram 10mg, three times a day for 5 days; Furosemide 40mg, th...	
Citalopram 20mg, once a day for 5 days; Loratadine 10mg, once a day ...	
Amlodipine 5mg, once a day as needed; Ibuprofen 400mg, three times ...	

For Dataset2:Hospital Treatment Details

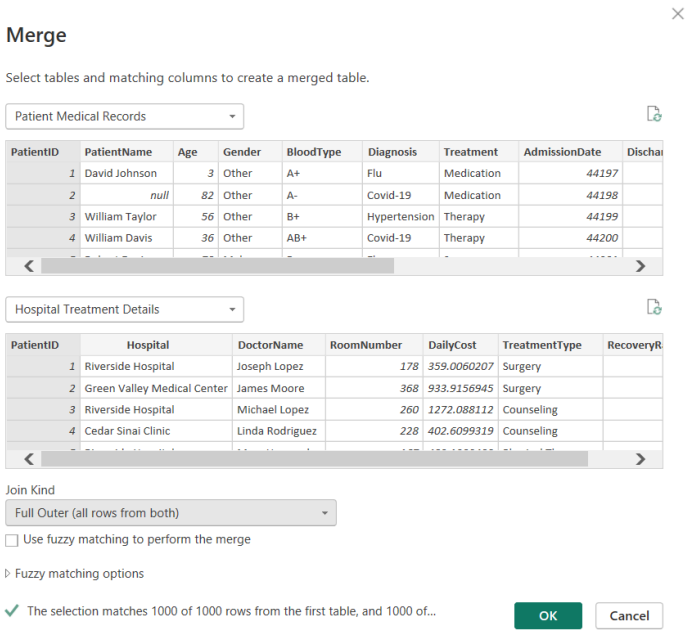
Hospital Name Column: Some records have missing values in the Hospital Name column. We need to fix these missing values for complete and accurate data.
Recovery Rating Column: there are null values in the Recovery Rating column. These null values might affect our analysis of patient recovery.

PatientID	Hospital
Valid 100%	Valid 0%
Error 0%	Error 0%
Empty 0%	Empty 100%
14	null
29	null
30	null
37	null
38	null
46	null
57	null
76	null
109	null
139	null
152	null
197	null
220	null
229	null
243	null
253	null
254	null
255	null
275	null
291	null
301	null
303	null
315	null
318	null

TreatmentType	RecoveryRating
Valid 100%	Valid 0%
Error 0%	Error 0%
Empty 0%	Empty 100%
Counseling	null
Physical Therapy	null
Surgery	null
Counseling	null
Physical Therapy	null
Physical Therapy	null
Counseling	null
Physical Therapy	null
Physical Therapy	null
Counseling	null
Physical Therapy	null
Surgery	null
Physical Therapy	null
Physical Therapy	null
Physical Therapy	null
Medication	null
Surgery	null
Physical Therapy	null

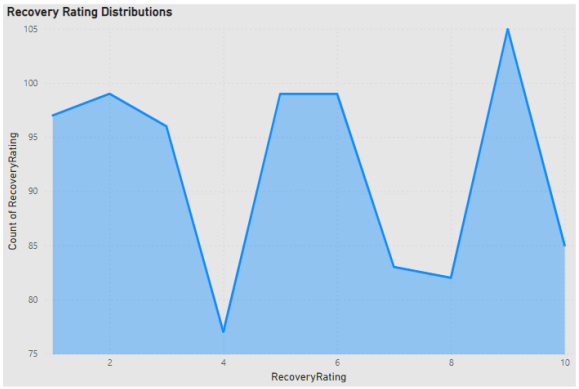
Merging and Relating Datasets

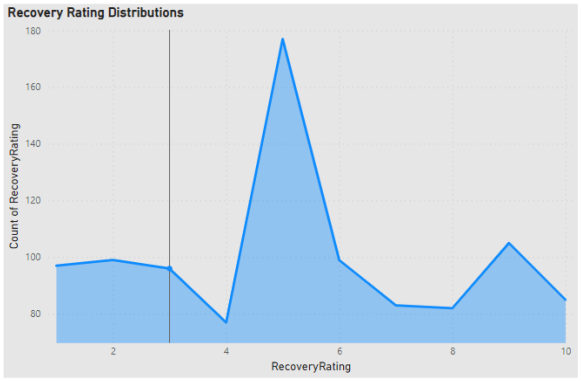
Both datasets were successfully merged using the PatientID column with a full outer join in this step. This merging technique ensured that all patient records from both datasets were included in the unified dataset.

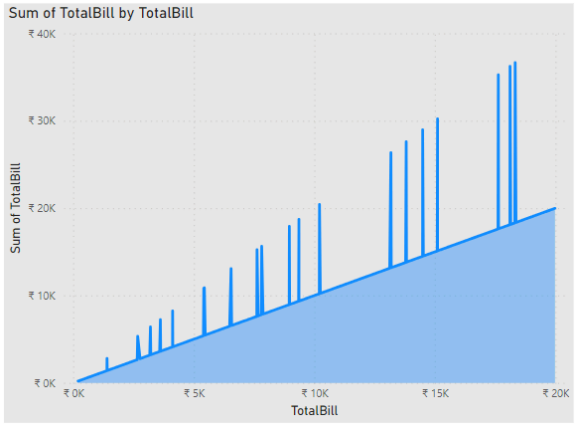


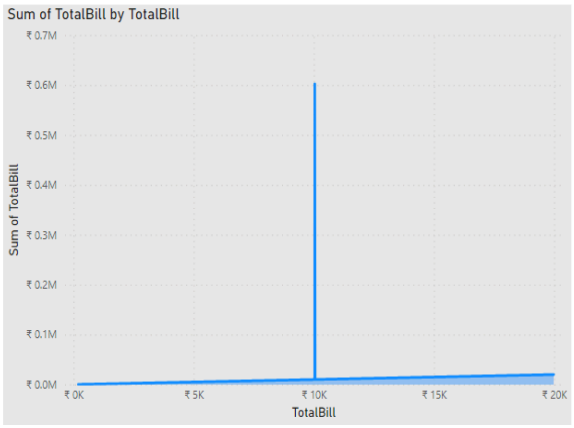
Cleaning: Handling Missing and Irrelevant Data and Data Type Conversion

Standardized Date formats for 'AdmissionDate' and 'DischargeDate'.
We removed duplicate entries.
Imputed null values with the mean for normally distributed data.









Categorizing Age Groups and Length of Stay

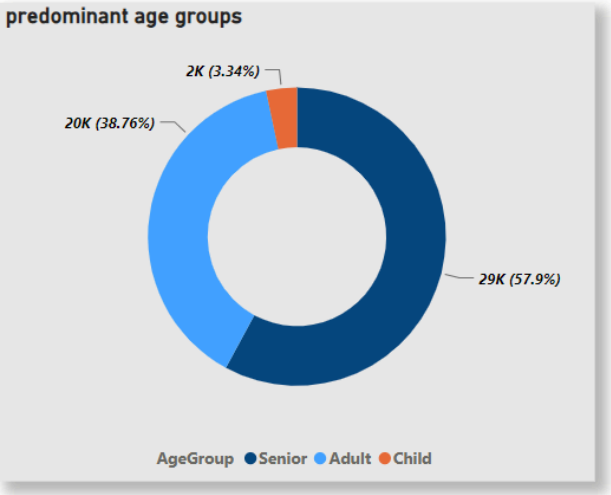
Length of Stay Calculation: Created a "LengthOfStay" column to determine the duration of each patient's stay.
Age Categorization: Introduced an "AgeGroup" column to classify patients into three groups: "Child," "Adult," and "Senior" based on their age.

AdmissionDate	DischargeDate	length_of_stay
Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%
01-01-2021	02-01-2021	1
02-01-2021	03-01-2021	1
03-01-2021	04-01-2021	1
04-01-2021	05-01-2021	1
05-01-2021	06-01-2021	1
06-01-2021	07-01-2021	1
07-01-2021	08-01-2021	1
08-01-2021	09-01-2021	1
09-01-2021	10-01-2021	1
10-01-2021	11-01-2021	1
11-01-2021	12-01-2021	1
12-01-2021	13-01-2021	1
13-01-2021	14-01-2021	1
14-01-2021	15-01-2021	1
15-01-2021	16-01-2021	1
16-01-2021	17-01-2021	1
17-01-2021	18-01-2021	1
18-01-2021	19-01-2021	1
19-01-2021	20-01-2021	1
20-01-2021	21-01-2021	1
21-01-2021	22-01-2021	1
22-01-2021	23-01-2021	1
23-01-2021	24-01-2021	1

Age	AgeGroup
Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%
3	Child
82	Senior
56	Adult
36	Adult
78	Senior
18	Child
62	Adult
13	Child
34	Adult
31	Adult
93	Senior
31	Adult
65	Adult
64	Adult
6	Child
47	Adult
27	Adult
40	Adult
91	Senior
24	Adult
6	Child
62	Adult

Categorizing Age Groups

In the hospital stay duration analysis, it's notable that David Johnson and John Moore experienced the longest stays, with David Johnson leading the count. Moreover, seniors emerge as the predominant age group among hospitalizations, suggesting a higher frequency of hospital visits within this demographic.



Length Of Stay for each Patient

unknown=null Values

PatientName	Sum of LengthOfStay
unknown	997
David Johnson	988
John Moore	958
John Anderson	950
Joseph Davis	950
David Hernandez	937
Jennifer Smith	934
Linda Thomas	923
Susan Wilson	923
David Taylor	921
Patricia Anderson	910
Barbara Brown	901
Jessica Rodriguez	899
Barbara Gonzalez	897
Thomas Miller	897
Michael Davis	894
Richard Johnson	891
Barbara Miller	885
Thomas Anderson	884
David Wilson	883
Patricia Hernandez	883
Total	**129909**

Analysis of Treatment Costs

On average, the costliest treatment category is medication, averaging around $10,195, closely followed by therapy. Additionally, the highest total treatment cost was observed for Jennifer Wilsen, amounting to approximately $98,530.

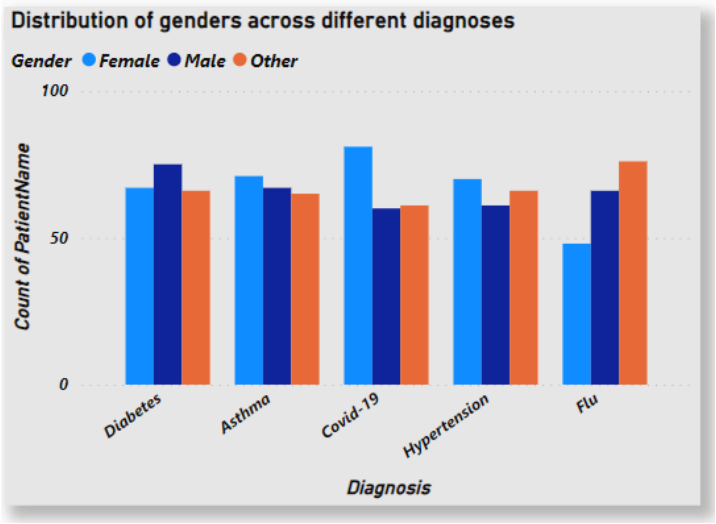
Treatment	Average of TotalBill
Medication	$10,195.13
Therapy	$10,027.20
Surgery	$9,875.96
Total	**$10,038.88**

Total cost of Treatment for each Patient

unknown=null values

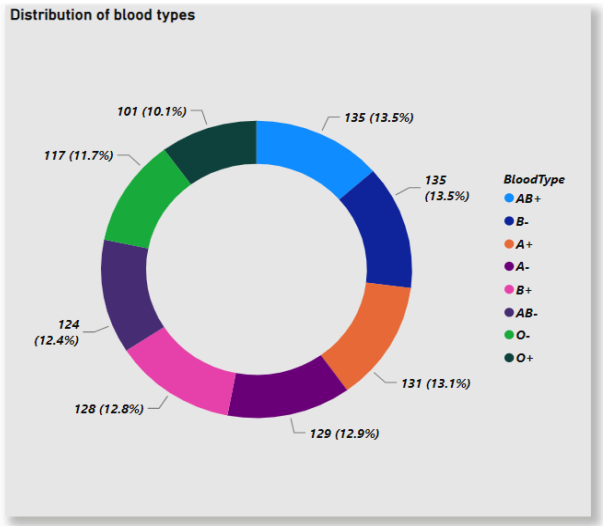
PatientName	Sum of TotalBill	Sum of DailyCost
unknown	$6,07,886.00	$59,920
Jennifer Wilson	$98,530.00	$5,100
Robert Rodriguez	$92,716.00	$6,774
Elizabeth Lopez	$85,988.00	$7,524
Jessica Rodriguez	$78,684.00	$4,831
Michael Garcia	$74,124.00	$6,349
Jennifer Anderson	$71,355.00	$2,238
Mary Miller	$69,290.00	$4,478
Barbara Miller	$69,041.00	$4,166
Barbara Brown	$68,948.00	$4,664
David Gonzalez	$68,911.00	$6,136
Joseph Davis	$68,894.00	$6,288
Michael Lopez	$67,594.00	$3,579
James Martin	$65,645.00	$5,242
Patricia Gonzalez	$65,127.00	$5,076
Richard Johnson	$63,072.00	$5,043
David Rodriguez	$61,383.00	$8,625
Jessica Brown	$59,008.00	$4,182
Michael Gonzalez	$56,839.00	$5,362
Sarah Moore	$53,900.00	$1,424
Patricia Hernandez	$53,861.00	$5,446
Jennifer Jackson	$53,522.00	$4,241
Jennifer Miller	$53,188.00	$1,513
Elizabeth Hernandez	$52,563.00	$3,600
Total	$1,00,38,877.00	$9,98,358

Gender Distribution in Diagnosis



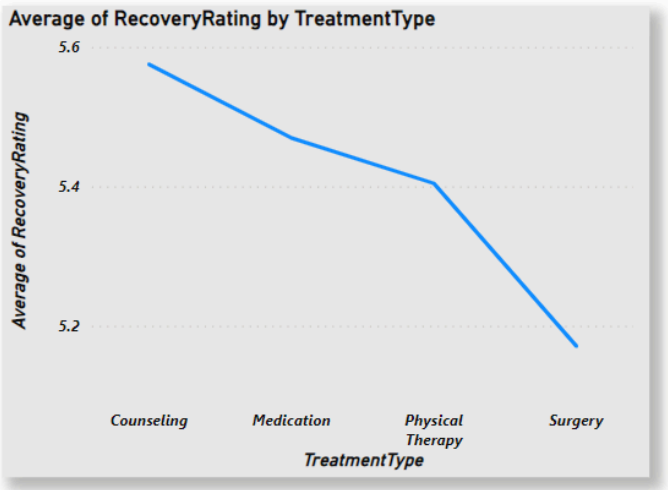
Upon examining the gender distribution across various diagnoses, a discernible pattern emerges. Firstly, females exhibit a higher vulnerability to a wide range of diagnoses, indicating a heightened susceptibility to illness. Secondly, in the case of COVID-19, females appear to be the most affected gender, as evidenced by the graph. Additionally, flu is prevalent among other gender types, while asthma emerges as a common diagnosis across genders.

Blood Type Analysis



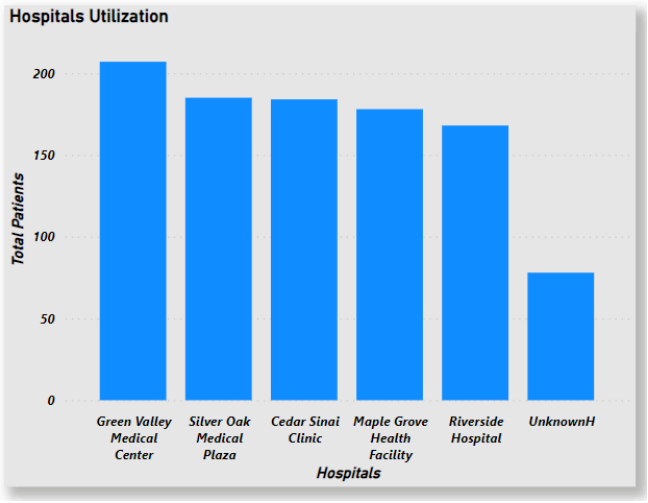
The distribution of blood types among patients appears to follow a normal distribution. However, when identifying the predominant blood types, AB+ and B- emerge as the leading types, while O+ and O- appear to be less prevalent.

Recovery Rating Analysis



Upon analyzing the average recovery ratings across different treatment types, it becomes evident that counselling treatment exhibits the highest average recovery rating, while surgery shows the lowest.

Hospital Utilization Analysis

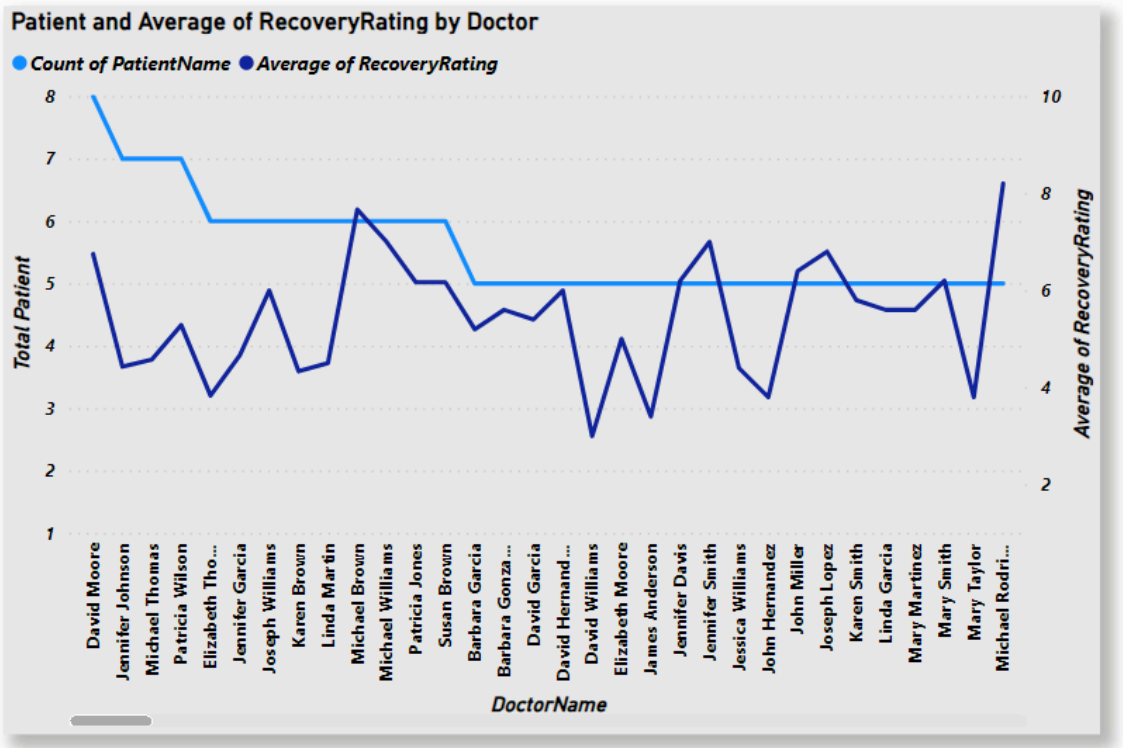


Hospital	RoomNumber	Count of PatientName
Cedar Sinai Clinic	54	3
Cedar Sinai Clinic	143	3
Cedar Sinai Clinic	237	3
Cedar Sinai Clinic	336	3
Green Valley Medical Center	138	3
Green Valley Medical Center	211	3
Green Valley Medical Center	255	3
Green Valley Medical Center	326	3
Green Valley Medical Center	359	3
Green Valley Medical Center	360	3
Green Valley Medical Center	385	3
Green Valley Medical Center	451	3
Riverside Hospital	233	3
Riverside Hospital	243	3
Riverside Hospital	383	3
Riverside Hospital	411	3
Silver Oak Medical Plaza	385	3
Silver Oak Medical Plaza	481	3
UnknownH	211	3
Cedar Sinai Clinic	2	2
Cedar Sinai Clinic	3	2
Cedar Sinai Clinic	28	2
Cedar Sinai Clinic	69	2
Total		1000

The Green Valley Medical Centre stands out with the highest number of admissions among hospitals. While the analysis suggests an even distribution of patients across room numbers, a closer examination reveals that Room Numbers 54, 143, and 237 at the Cedar Sinai Clinic accommodate the maximum number of patients.

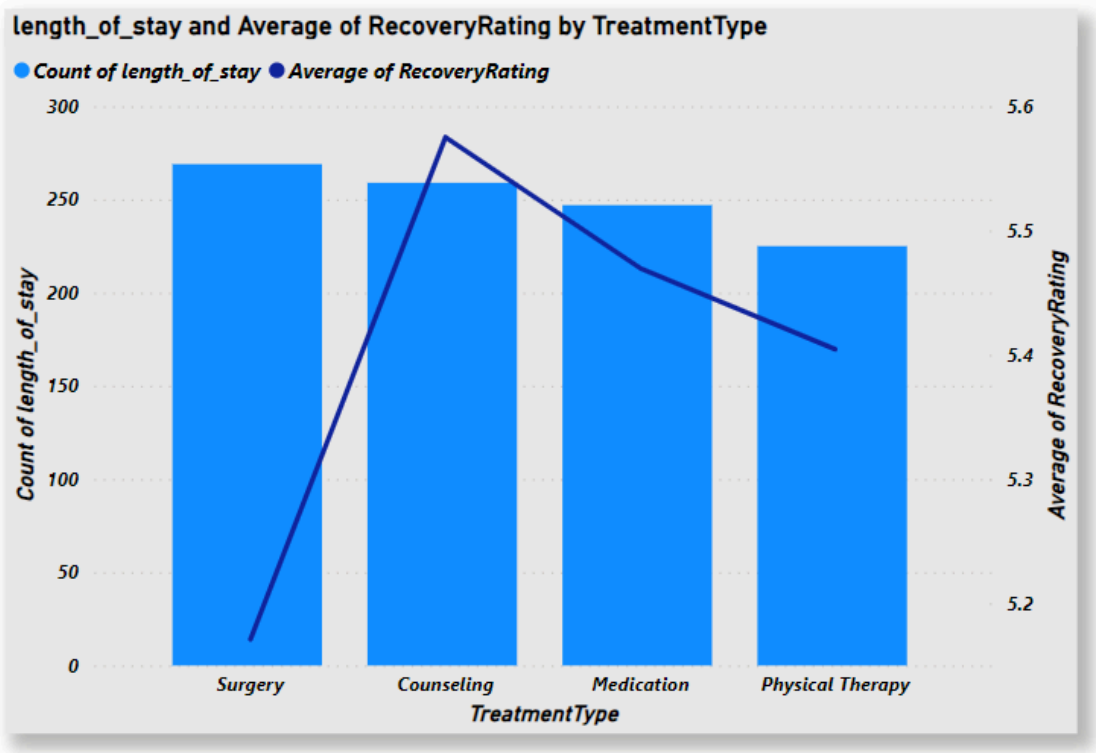
Doctor's Patient Load

Upon reviewing the graph depicting the total patient count and average recovery rating for different doctors, it appears that patient load could potentially influence the recovery rating. However, it's worth noting that this correlation varies among different doctors, suggesting individual differences in patient care and treatment efficacy.

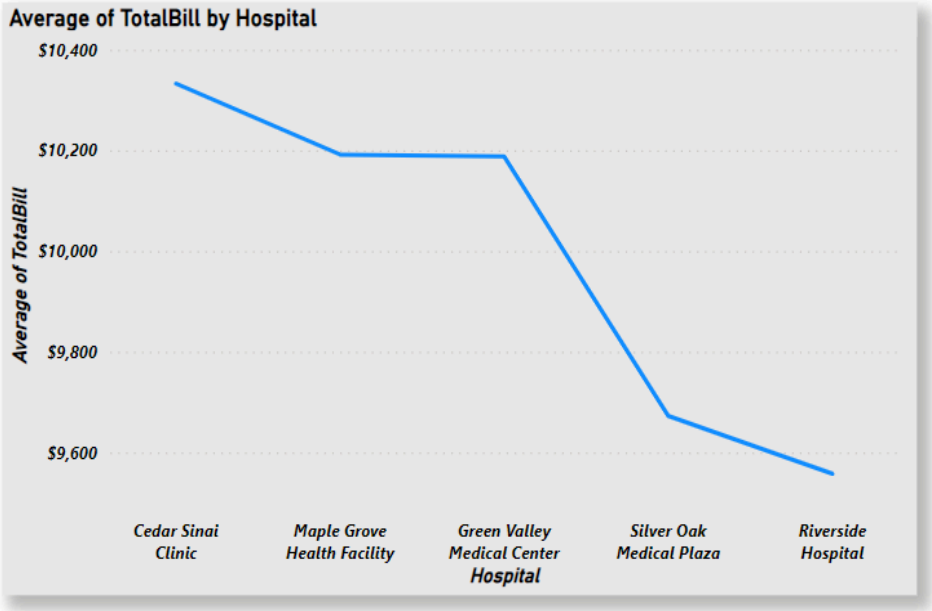


Treatment Effectiveness

Upon analyzing the recovery rating alongside the length of stay for various treatment types, the graph highlights noteworthy findings. While surgery exhibits the longest length of stay, it paradoxically displays the lowest recovery rating. Conversely, counselling, with its notably longer duration of treatment, showcases the highest recovery rating among the treatment types examined.

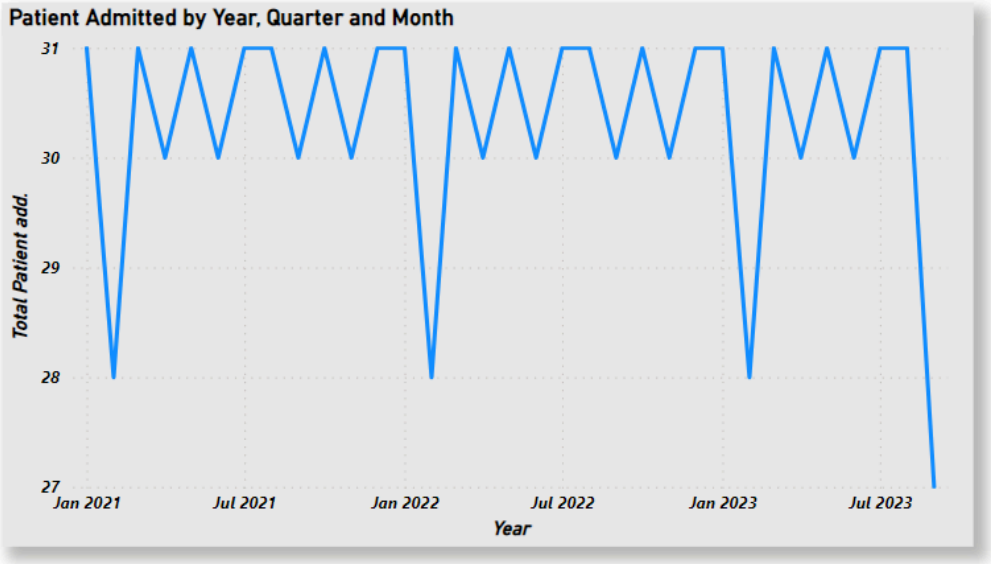


Cost Analysis by Hospital



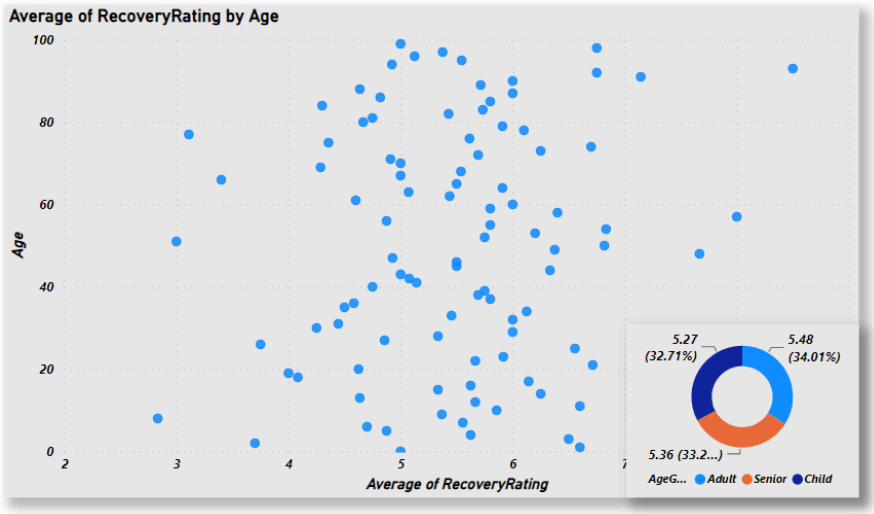
Upon analysis of the average treatment costs across various hospitals, Cedar Sinai Clinic emerges as the most expensive, showcasing comparatively higher treatment costs. Conversely, Riverside Hospital exhibits the lowest treatment costs among the hospitals examined.

Patient Admission Trends Over Time



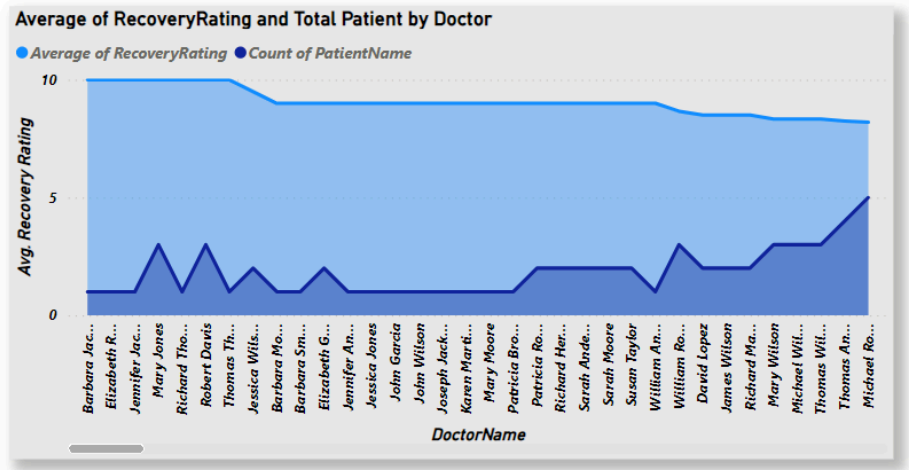
The patient admissions throughout the year generally follow a consistent pattern, with no significant deviations observed. However, it's notable that in February of each year, there is a gradual decrease in patient admissions.

Correlation Between Age and Recovery



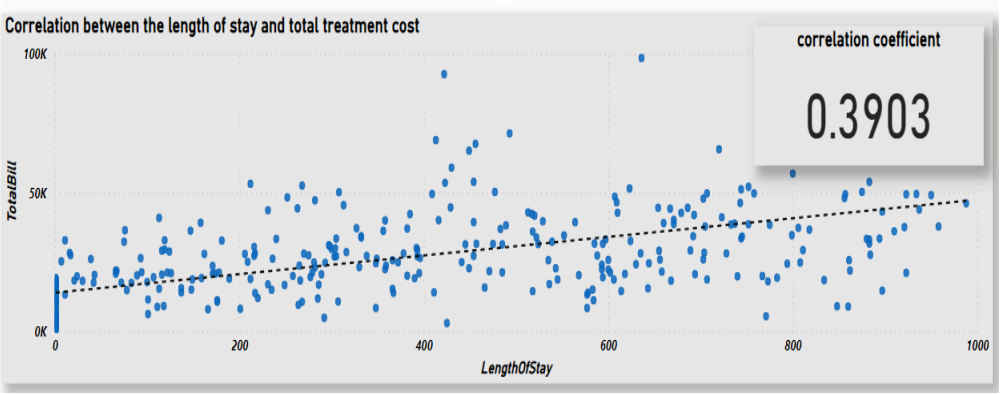
After a thorough examination of the scatter plot and pie chart, it becomes apparent that there isn't a direct correlation between age and recovery rating.

Impact of Doctor on Recovery



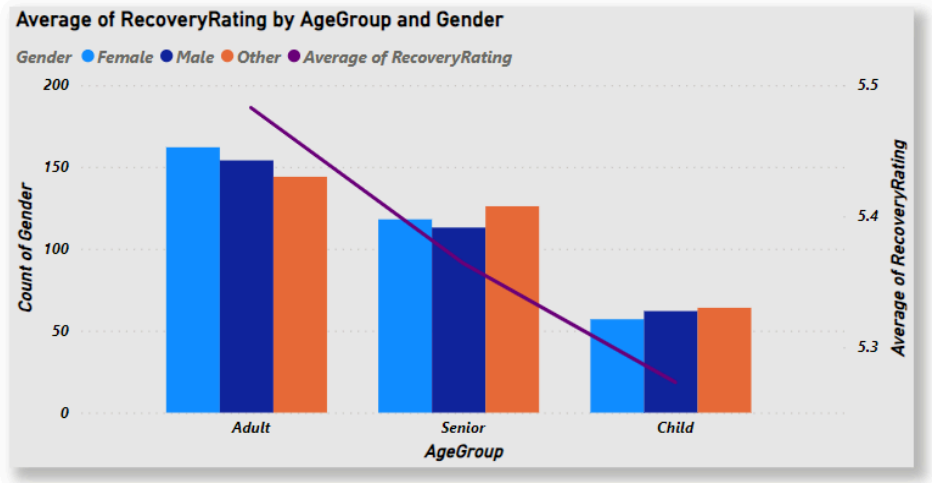
Upon analyzing the graph depicting the impact of doctors on recovery ratings, it appears that a doctor's treatment does indeed influence patient recovery ratings. However, it's important to note that recovery ratings are influenced by various other factors as well.

Advanced DAX: Length of Stay and Cost Correlation



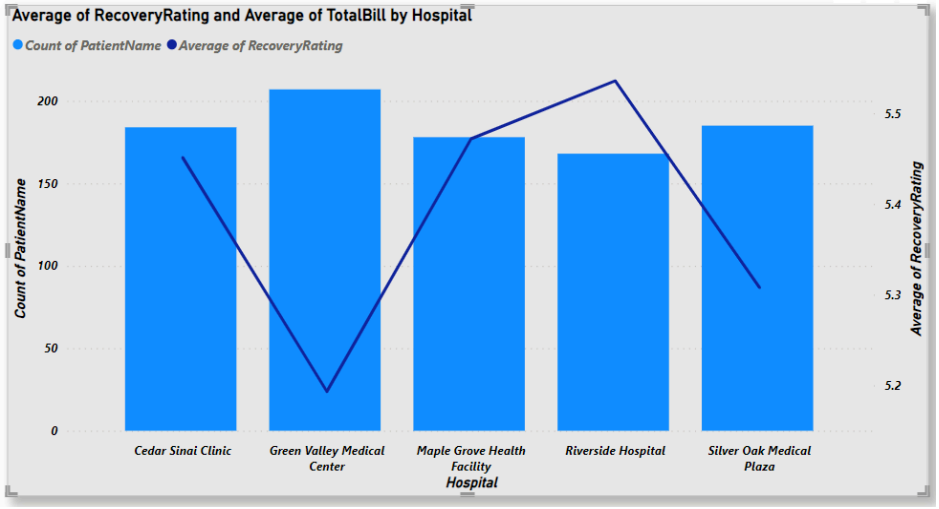
A correlation coefficient of 0.3905 indicates a moderate positive correlation between the two variables being analyzed. In this case, it suggests that there is a tendency for the length of stay and the total bill to increase together, but the relationship is not extremely strong.

Recovery Trends by Gender and Age Group



Analysis of the average recovery ratings by gender and age group reveals that adults have the highest admission rates to hospitals, coinciding with comparatively higher recovery ratings. Conversely, the child age group exhibits lower recovery ratings compared to other age groups.

Hospital Performance Analysis

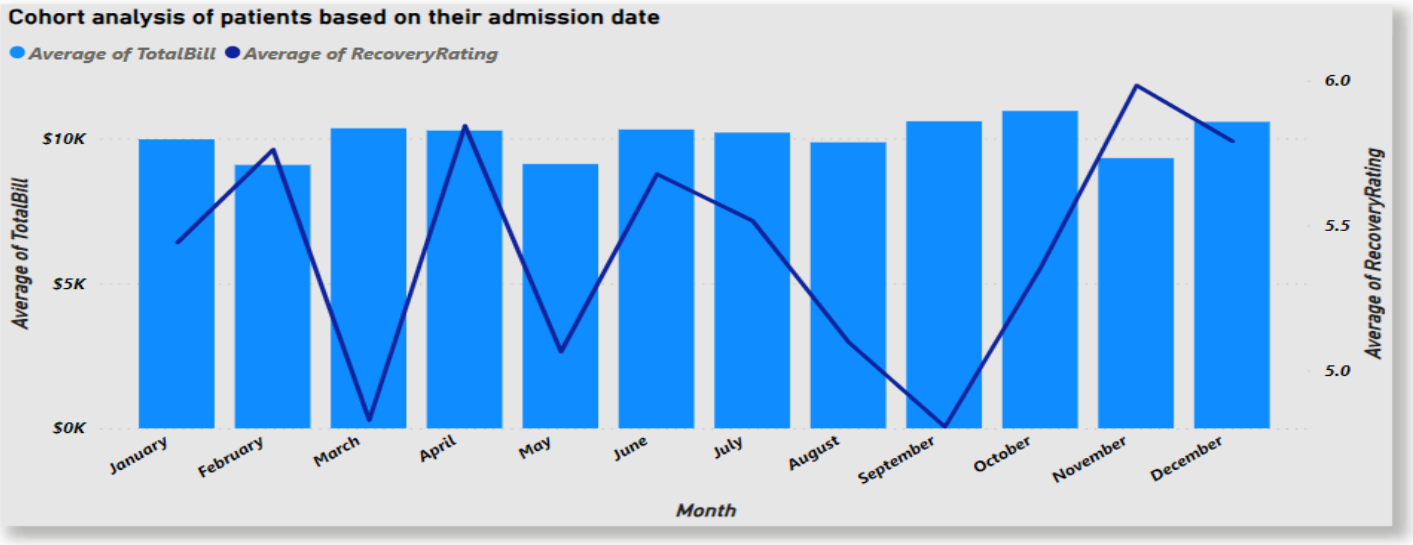


In the analysis of hospital performance, the Green Valley Medical Center stands out with the highest patient admissions, yet it demonstrates the lowest recovery rating. This suggests a potential impact of patient burden on recovery rates. Conversely, the Riverside Hospital, despite having fewer patient admissions compared to others, exhibits the highest recovery rating. This correlation suggests that patient burden may indeed influence recovery rates.

Extracting Key Information

Medications&Doses	
Valid	100%
Error	0%
Empty	0%
Clonazepam 2mg,Furosemide 40mg, Hydrochlorothiazide 25mg	
Amoxicillin 500mg,Losartan 50mg, Prednisone 20mg	
Gabapentin 300mg,Amlodipine 5mg,	
Simvastatin 40mg,Azithromycin 250mg,	
Amoxicillin 500mg,Duloxetine 60mg, Doxycycline 100mg	
Fluoxetine 20mg,Escitalopram 10mg, Duloxetine 60mg	
Fluoxetine 20mg,Sertraline 100mg, Doxycycline 100mg	
Prednisone 20mg,Ibuprofen 400mg,	
Prednisone 20mg,Duloxetine 60mg, Hydrochlorothiazide 25mg	
Tramadol 50mg,Fluoxetine 20mg, Furosemide 40mg	
Clopidogrel 75mg,Levothyroxine 50mcg, Lisinopril 20mg	
Lisinopril 20mg,Escitalopram 10mg, Furosemide 40mg	
Loratadine 10mg,Furosemide 40mg, Levothyroxine 50mcg	
Sertraline 100mg,Ibuprofen 400mg,	
Ibuprofen 400mg,Citalopram 20mg,	
Gabapentin 300mg,Ciprofloxacin 500mg, Duloxetine 60mg	
Gabapentin 300mg,Levothyroxine 50mcg,	
Simvastatin 40mg,Meloxicam 15mg, Amlodipine 5mg	
Sertraline 100mg,Azithromycin 250mg, Levothyroxine 50mcg	
Amoxicillin 500mg,Clonazepam 2mg, Simvastatin 40mg	
Fluoxetine 20mg,Atorvastatin 10mg, Amlodipine 5mg	
Doxycycline 100mg,Tramadol 50mg,	
Insulin Glargine 100 units/mL,Gabapentin 300mg, Escitalopr…	
Sertraline 100mg,Ciprofloxacin 500mg, Omeprazole 40mg	

Data Modeling: Cohort Analysis Based on Admission Date



Analyzing the month-wise cohort, we observe that the recovery ratings in March and September are notably lower. This trend coincides with higher admissions during these months, potentially attributing to strained resources and consequently impacting recovery outcomes negatively. Interestingly, despite higher admissions, the revenue from bills appears to peak during these periods, indicating a possible surge in medical services rendered. Conversely, November stands out with the highest average recovery rating, reaching approximately 6. This outlier suggests effective treatment protocols or favorable patient outcomes during this month, warranting further investigation into the underlying factors contributing to this positive trend.