HEALTHCARE DATA ANALYSIS
PREDICTING BILLING AMOUNT FOR THE AGES

Presented by: J’yanni Ellis and Sunita Thakur

CONTENTS

IMPORTANCE OF DATA ANALYSIS IN HEALTHCARE

- Enhances decision-making
- Improves patient care
- Supports resource management
- Facilitates research and innovation
- Cost reduction
- Minimizing medical errors
- Financial risk control
- Disease prediction and prevention

DATA

This dataset collected from Kaggle.com contains information about patients admitted to a hospital, including demographic details, medical conditions, treatments, and billing information. The dataset comprises the following columns:

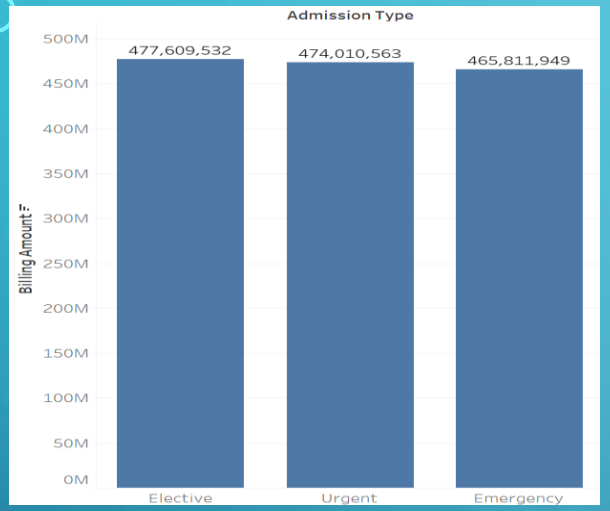
Name, Age, Gender, Blood Type, Medical Condition, Date of Admission, Doctor, Hospital, Insurance Provider, Billing Amount, Room Number, Admission Type, Discharge Date, Medication, Test Results

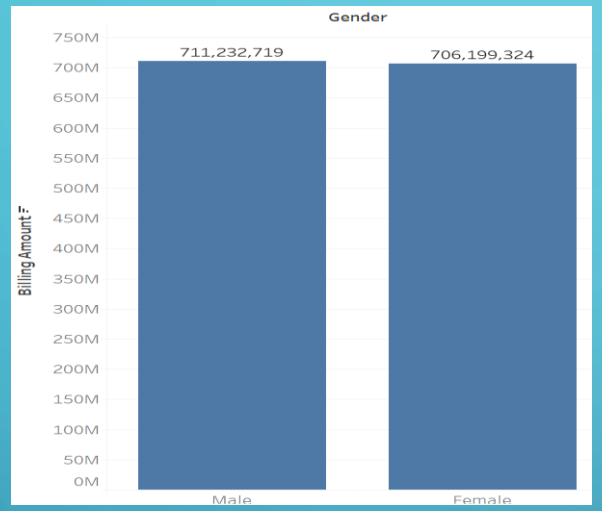
OBSERVATIONS

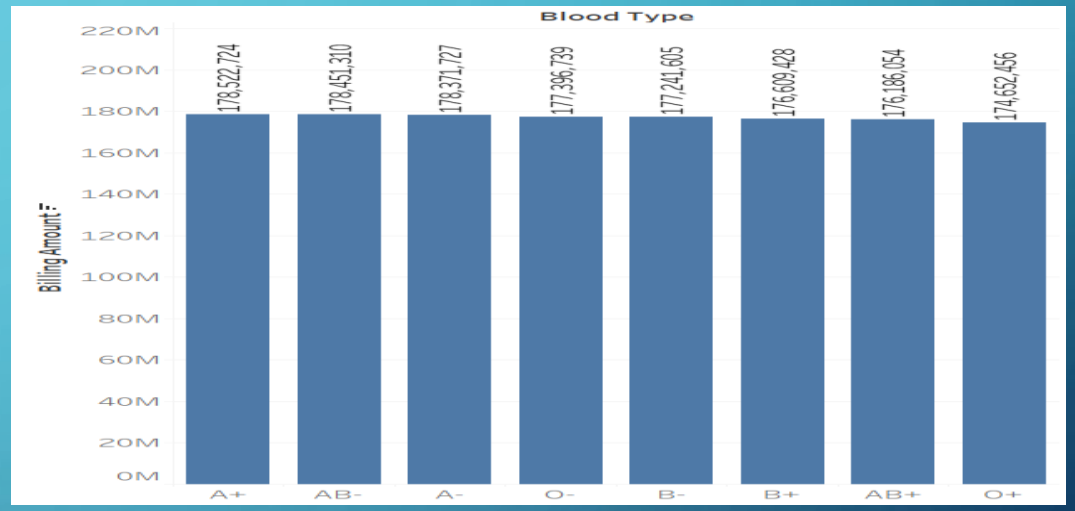
- Patients ages range from 13 to 89 years with the average of 52.
- Data spans from May 8, 2019, to May 7, 2024, providing a comprehensive five-year view of patient admissions.
- Admission types: Emergency, Elective, Transfer
- Blood types: A+, A-, B+, B-, O+, O-, AB+, AB-

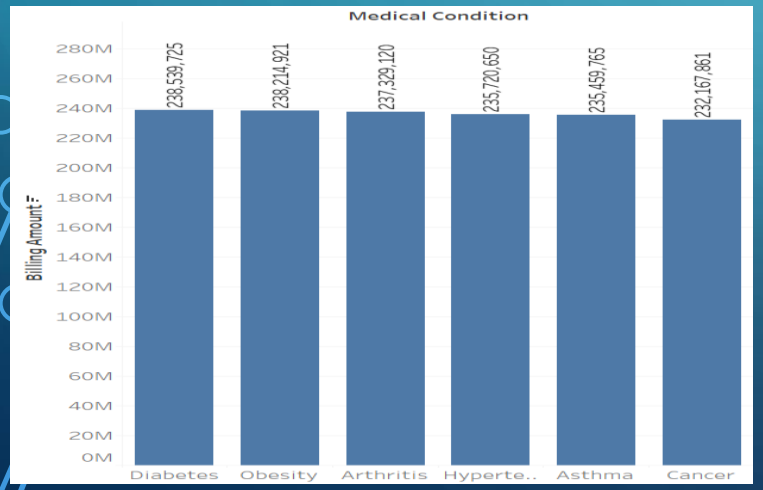
  A- blood group is the most prevalent.
- The dataset encompasses admissions from 44 hospitals, with LLC Smith being the most frequent.
- Total Doctors in the dataset are 27 and Michael Smith attends to the highest number of patients.
- Male patients are more than the female patients.
- Arthritis is the topmost disease in the hospitals
- Diabetes treatment costs are the highest.

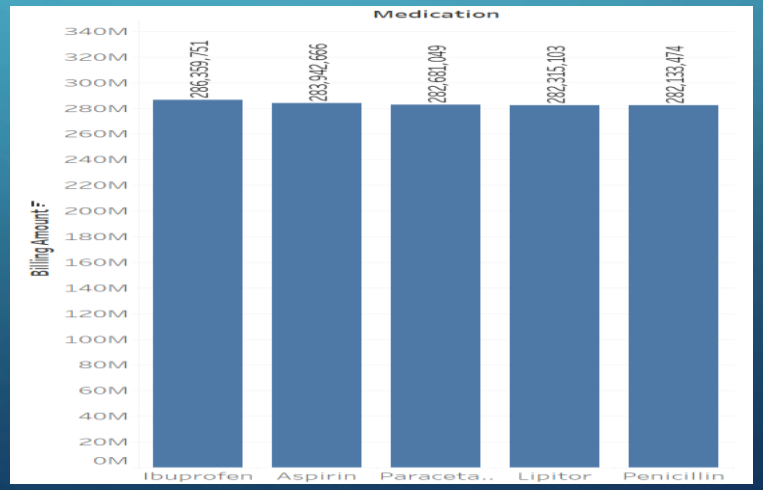
BILLING AMOUNT FOR ADMISSION TYPES, GENDER, BLOOD TYPE, MEDICAL CONDITION, MEDICATION, AND DISCHARGE DATE IN DESCENDING ORDER

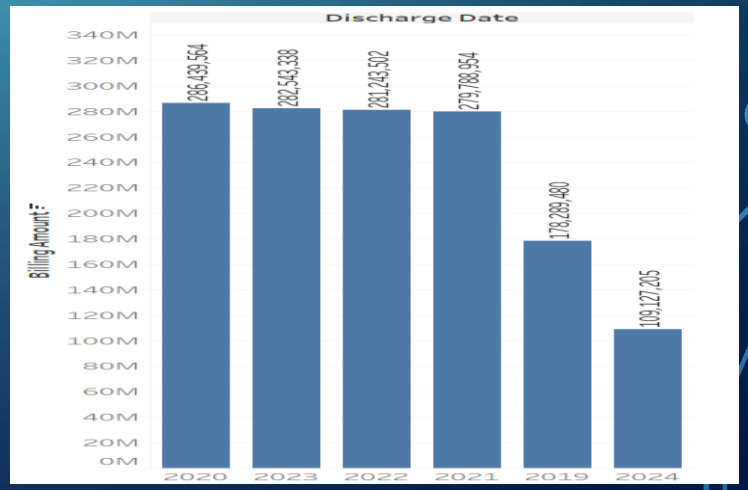




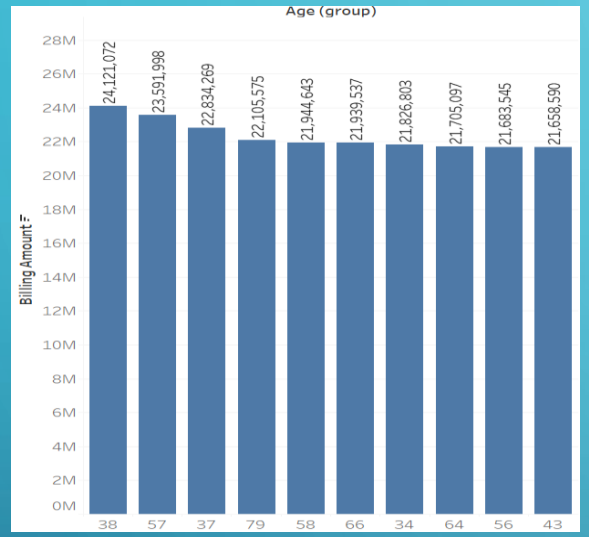


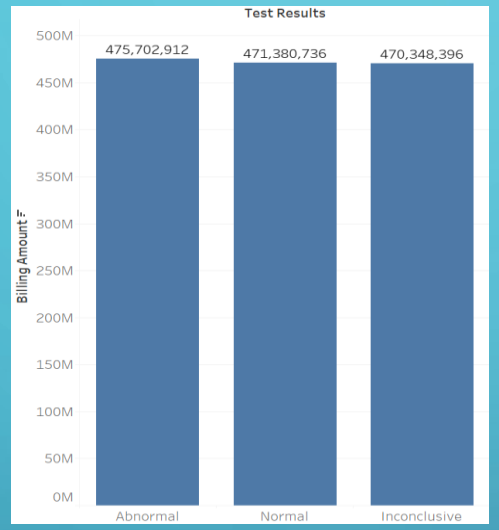


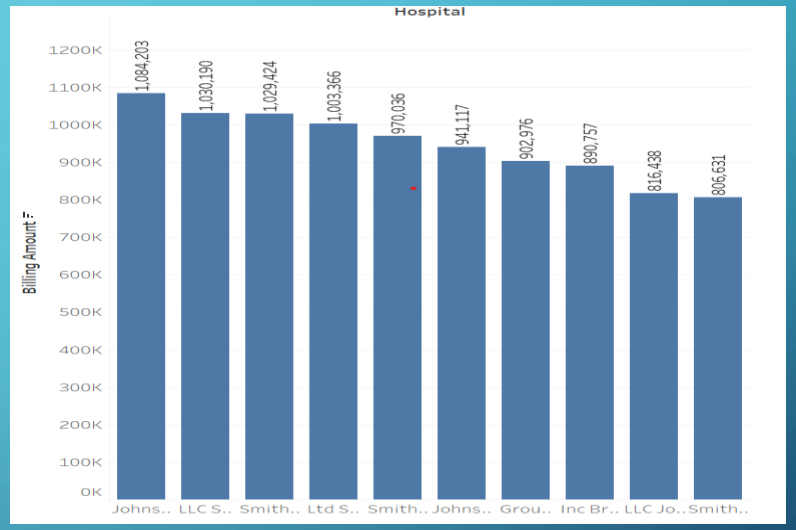


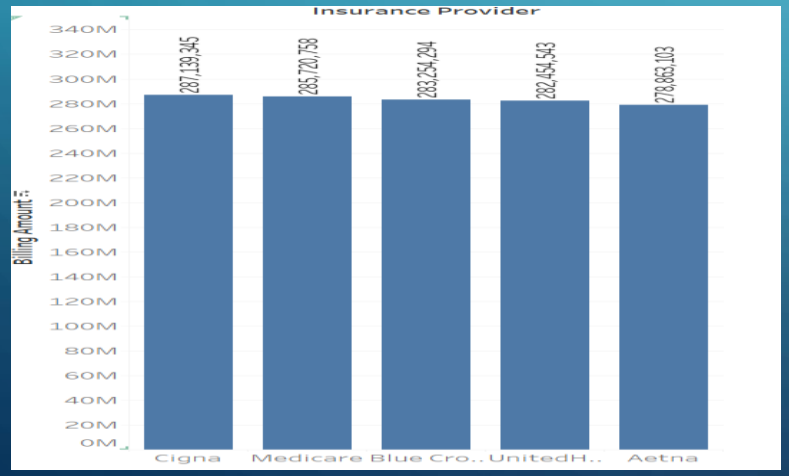


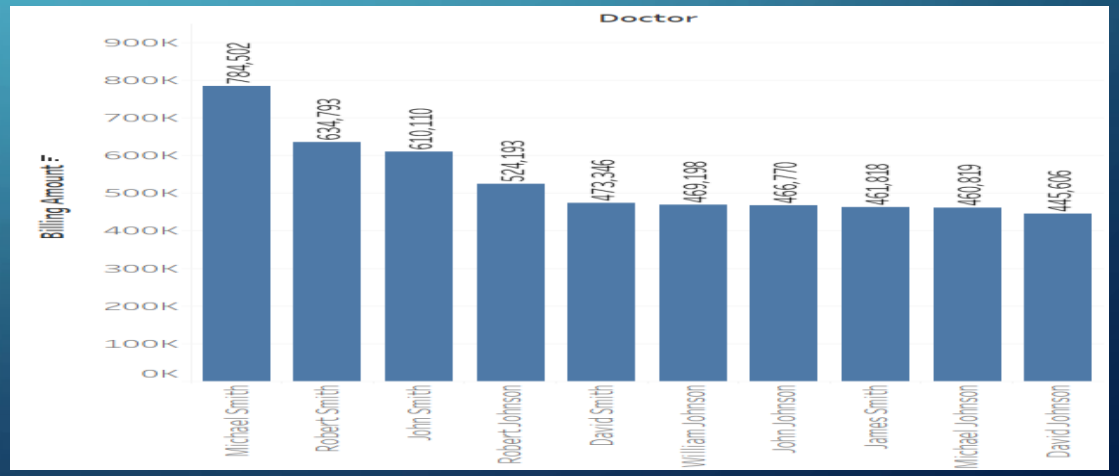
BILLING AMOUNT FOR AGE(GROUPS), TEST RESULTS, HOSPITAL, INSURANCE PROVIDER, AND DOCTOR IN DESCENDING ORDER



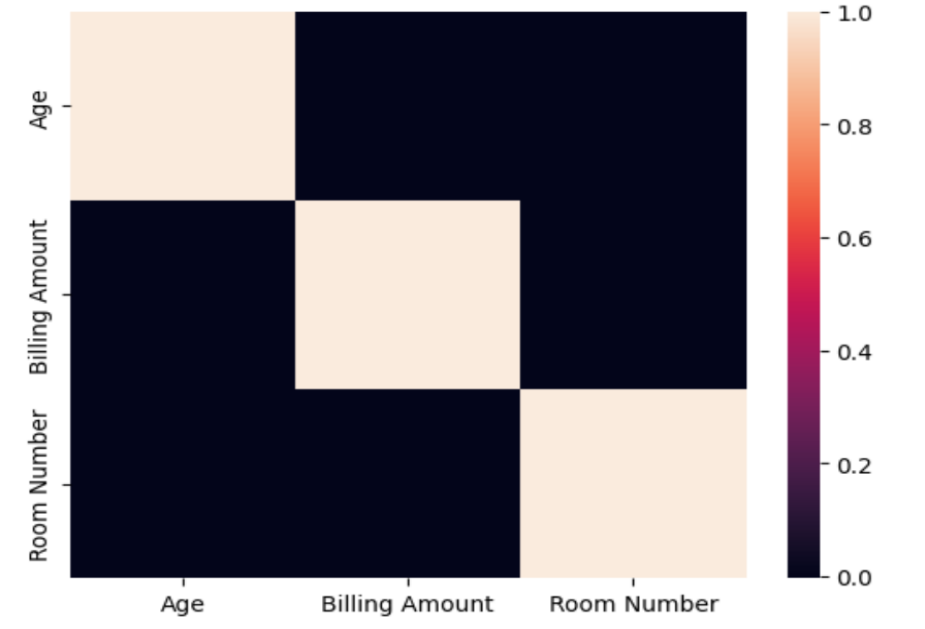


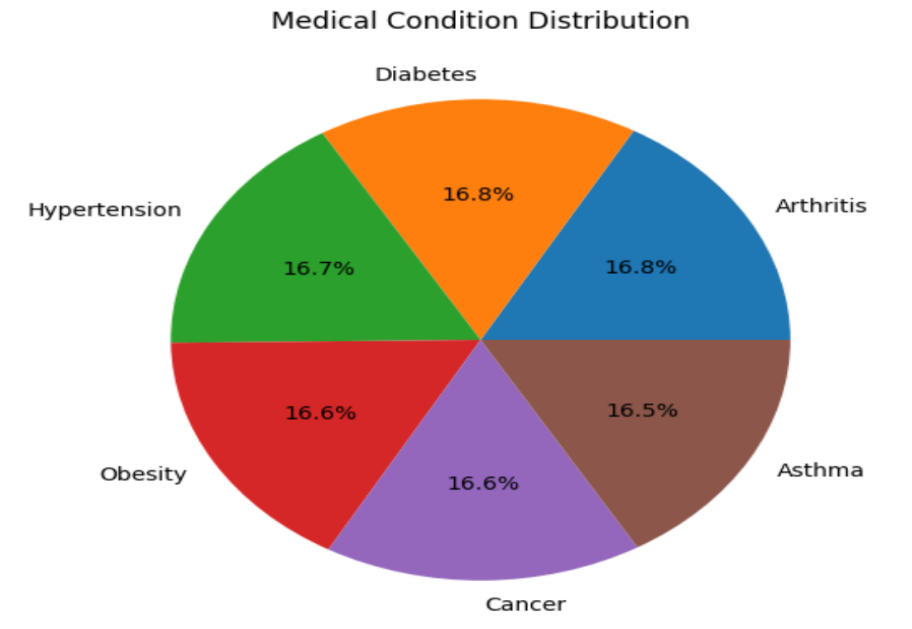






CORRELATION BETWEEN BILLING AMOUNT AND AGE
MEDICAL CONDITIONS





- There is no significant correlation between billing amount and age.

- Arthritis is the topmost disease in the hospitals.

METHODOLOGY

- Dependent variable: Billing amount
- Independent variable: Age
- Linear regression
- Lasso Regression
- Classification
- Time series forecasting: ARIMA, SARIMA, and AUTO ARIMA models

BILLING AMOUNT PREDICTIONS FOR THE AGES BY LINEAR REGRESSION AND LASSO REGRESSION

Linear Regression Predictions

Age	Billing amount
38	25559.39
51	25532.55
78	25476.82
19	25598.61
50	25534.62
61	25511.91
71	25491.27
58	25518.1
49	25536.68
74	25485.08
50	25534.62
34	25567.64
61	25511.91
50	25534.62
38	25559.39
72	25489.21
24	25588.29
21	25594.48
40	25555.26
39	25557.32
32	25571.77
42	25551.13
39	25557.32
80	25472.69
39	25557.32
81	25470.63
56	25522.23
20	25596.54
43	25549.07
66	25501.59

20	25596.54
29	25577.96
75	25483.01
60	25513.98
76	25480.95
48	25538.75
67	25499.53
85	25462.37
26	25584.16
71	25491.27
27	25582.09
58	25518.1
64	25505.72
73	25487.14
24	25588.29
58	25518.1
50	25534.62
41	25553.19
71	25491.27
26	25584.16
24	25588.29
23	25590.35
78	25476.82
30	25575.9
55	25524.3
65	25503.66
40	25555.26
42	25551.13
77	25478.89
24	25588.29

37	25561.45
67	25499.53
64	25505.72
52	25530.49
82	25468.56
41	25553.19
30	25575.9
44	25547
73	25487.14
22	25592.41
78	25476.82
46	25542.87
57	25520.17
73	25487.14
72	25489.21
46	25542.87
84	25464.44
22	25592.41
19	25598.61
26	25584.16
63	25507.78
19	25598.61
59	25516.04
56	25522.23
38	25559.39
40	25555.26
58	25518.1
61	25511.91
79	25474.76
18	25600.67

26	25584.16
25	25586.22
36	25563.52
72	25489.21
48	25538.75
36	25563.52
60	25513.98
31	25573.84
62	25509.85
55	25524.3
82	25468.56
42	25551.13
87	25458.24
60	25513.98
34	25567.64
23	25590.35
38	25559.39
39	25557.32
65	25503.66
43	25549.07
71	25491.27

Lasso Regression Predictions

Age	Billing amount
38	25559.39
51	25532.55
78	25476.82
19	25598.61
50	25534.62
61	25511.91
71	25491.27
58	25518.1
49	25536.68
74	25485.08
50	25534.62
34	25567.64
61	25511.91
50	25534.62
38	25559.39
72	25489.21
24	25588.28
21	25594.48
40	25555.26
39	25557.32
32	25571.77
42	25551.13
39	25557.32
80	25472.69
39	25557.32
81	25470.63
56	25522.23
20	25596.54
43	25549.07
66	25501.59

20	25596.54
29	25577.96
75	25483.01
60	25513.98
76	25480.95
48	25538.75
67	25499.53
85	25462.37
26	25584.16
71	25491.27
27	25582.09
58	25518.1
64	25505.72
73	25487.14
24	25588.28
58	25518.1
50	25534.62
41	25553.19
71	25491.27
26	25584.16
24	25588.28
23	25590.35
78	25476.82
30	25575.9
55	25524.3
65	25503.66
40	25555.26
42	25551.13
77	25478.89
24	25588.28

37	25561.45
67	25499.53
64	25505.72
52	25530.49
82	25468.57
41	25553.19
30	25575.9
44	25547
73	25487.14
22	25592.41
78	25476.82
46	25542.87
57	25520.17
73	25487.14
72	25489.21
46	25542.87
84	25464.44
22	25592.41
19	25598.61
26	25584.16
63	25507.78
19	25598.61
59	25516.04
56	25522.23
38	25559.39
40	25555.26
58	25518.1
61	25511.91
79	25474.76
18	25600.67

26	25584.16
25	25586.22
36	25563.52
72	25489.21
48	25538.75
36	25563.52
60	25513.98
31	25573.84
62	25509.85
55	25524.3
82	25468.57
42	25551.13
87	25458.24
60	25513.98
34	25567.64
23	25590.35
38	25559.39
39	25557.32
65	25503.66
43	25549.07
71	25491.27

MODEL EVALUATION

Linear regression

- Mean Absolute Error: 12295.712152020222
- Mean Squared Error: 202584280.70492148
- Root Mean Squared Error: 14233.210484810568

Lasso regression

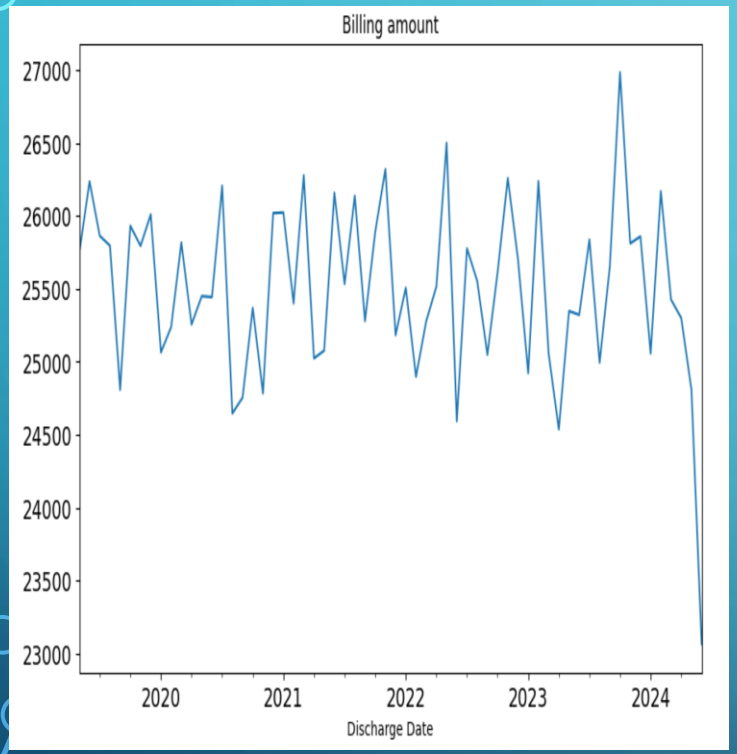
- Mean Absolute Error: 12295.712152108861
- Mean Squared Error: 202584280.7764024
- Root Mean Squared Error: 14233.210487321629

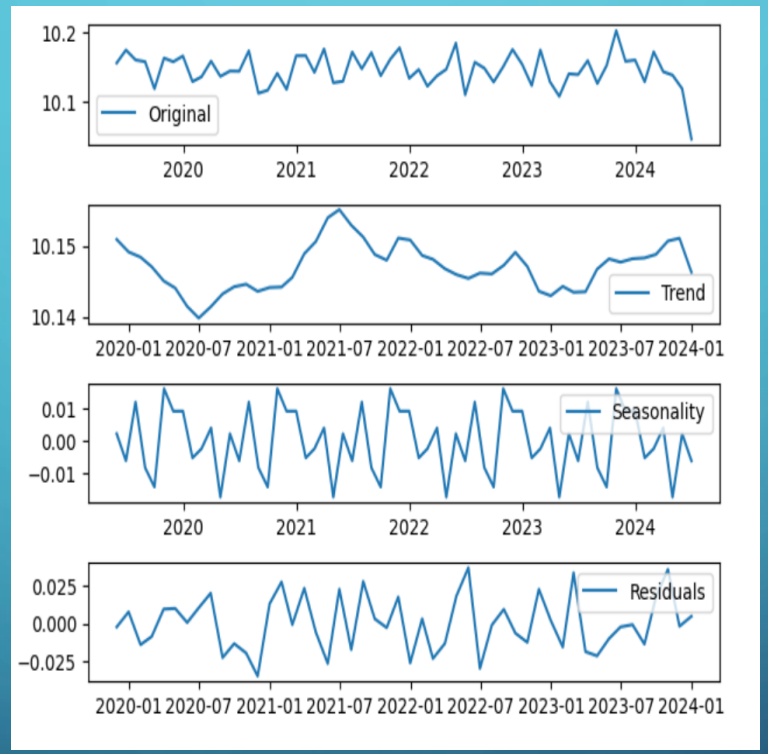
Billing amount prediction for the ages by Linear regression is better than Lasso regression because of smaller root mean squared error.

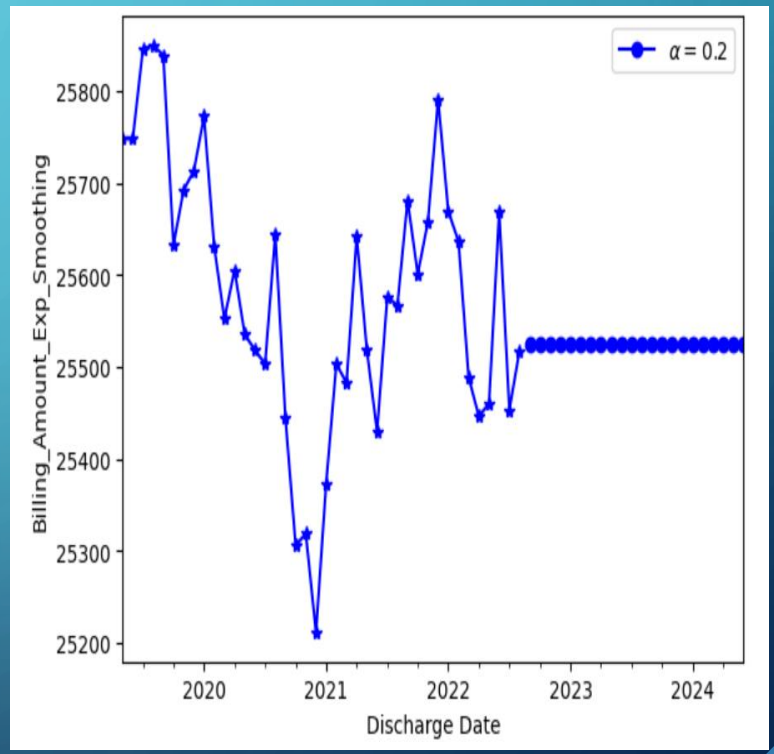
Classification

- Logistic Regression classifier with accuracy 0.50599 is better than Decision tree classifier with accuracy 0.50595.

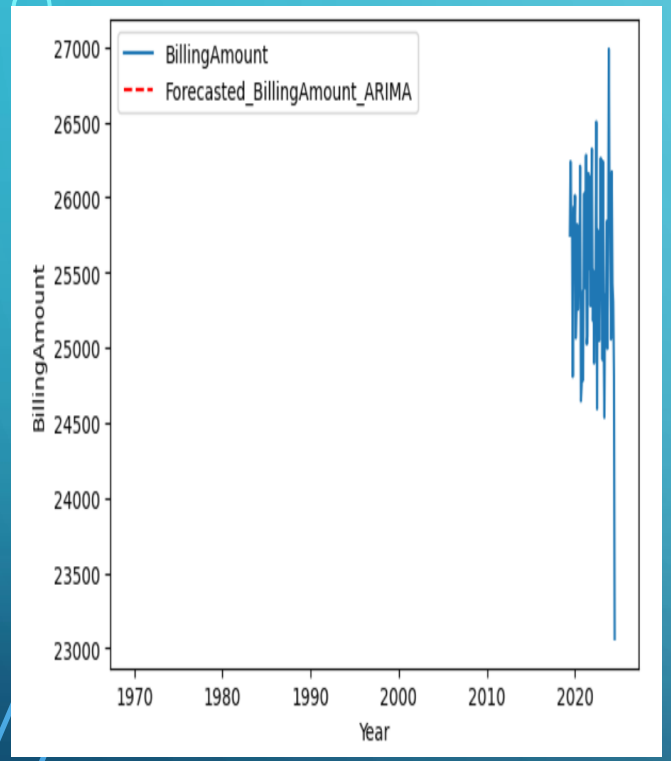
TIME SERIES PLOTTING, DECOMPOSING, AND BILLING AMOUNT FORCASTING

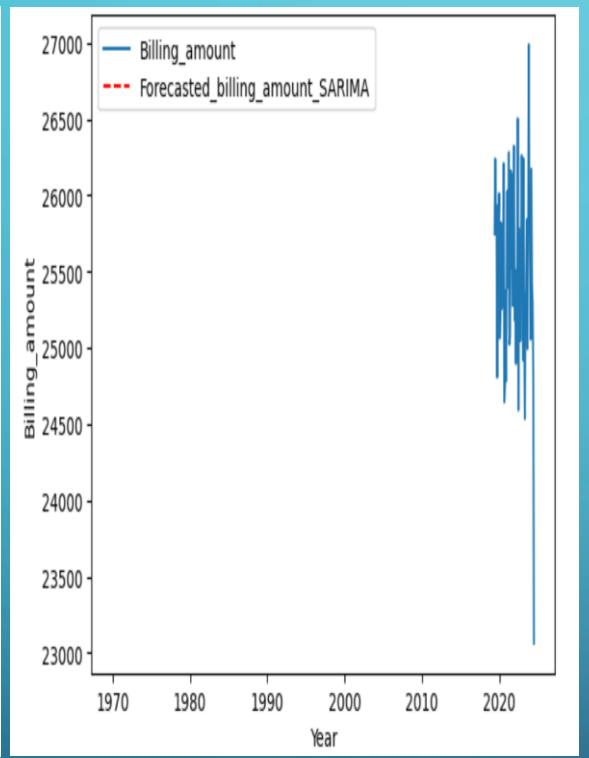


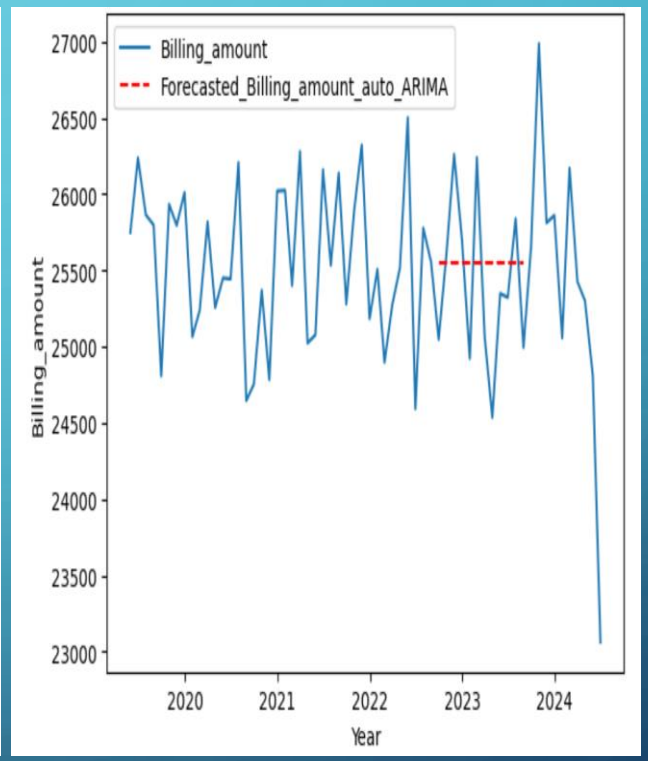




BILLING AMOUNT TIME SERIES FORCASTING AND MODEL EVALUATION







- ARIMA model's AIC = 632.49
- SARIMA model's AIC = 463.29

- Model with smaller AIC is the better model.
- SARIMA model is better in forecasting than ARIMA model.

CONCLUSION

- Maximum billing amount is for the ages: 38, admission type: elective, gender: male, blood type: A+, medical condition: diabetes, medication: ibuprofen, discharge date: 2020, test results: abnormal, hospital: John’s, Doctor: Michael Smith, insurance provider: Ciagna.
- Arthritis and diabetes are the topmost diseases in the hospitals.
- Future trends for the billing amount is constant in the hospitals.
- This dataset provides valuable insights into patient demographics, medical conditions, treatment patterns, and hospital operations. It can be used for various analyses, including better patient outcome prediction, more efficient resource management, informed decision-making and healthcare quality improvement initiatives.

REFERENCES

[1] Analysis: https://www.kaggle.com/code/muhammadfurqan0/unlocking-healthcare-trends-data-analysis

[2] Analysis: https://www.kaggle.com/code/youssefayman22/simple-neural-network-algorithm

[3] Dataset: https://www.kaggle.com/datasets/prasad22/healthcare-dataset

THANK YOU

QUESTIONS?