Time Series and Forecasting: Forecasting tool formulation for a walk-in emergency clinics

Name: Sagar Ramteke | Student ID: C23114494

Introduction

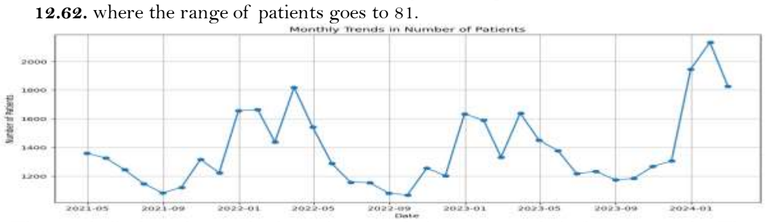
An emergency clinic tends to encounter the maximum number of urgent patients (ones without a prior appointment). Therefore, such medical providers need to be well resourced in advance to provide medical aids to n number of patients should there be an emergency. This report tends to create a forecasting tool for a walk-in emergency clinic to forecast the patient outpour one month ahead of its time.

The aim is to predict the number of walk-in patients in March 2024 by using the data collected from April 2021 to February 2024. Different time series models were used to derive the conclusion, starting from Baseline and Simple approach (Naive, Mean, Moving Average, and Simple Linear Regression) to Complex Approaches (SES, Holt Linear, Holt Winters, ARIMAs and SARIMAs).

Preliminary Analysis

mean	std	25%	50%	75%	range
45.62	12.62	37	44	52	81

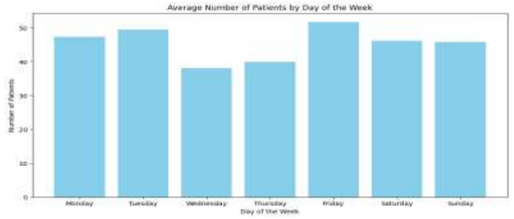
Moderate fluctuation is evident in the monthly walk-in patients , with a standard deviation of **12.62.** where the range of patients goes to 81.



The graph displays changes in the patient population over a period of many months. There are noticeable highs and lows, which point to variations in the number of patients.

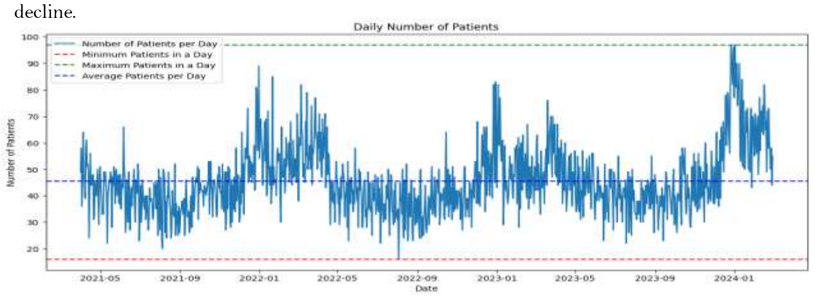
There's a notable surge in **March 2022**, when the patient count surpasses **2000** which in **66.67%** increase On the other hand, there are instances where there are less than **1200** patients.

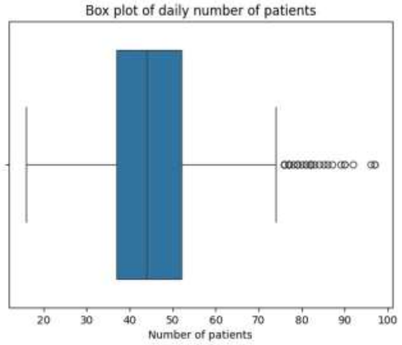
In total **153** weeks minimum number of patients were **187** which is in on Wednesday, while maximum number of patients were **586** was on Friday as per shown in graph. Healthcare facilities may need to allocate more resources on Monday, Tuesday and Wednesday to handle the high number of patients.



Compared to other days Monday, Tuesday and Friday have more patients while Saturday Sunday have a gradual decrease in patients which indicates more people seeks medical attention early in this week , possibly due to weekend recovery or scheduled appointments.

Below graph give the clear insight of seasonality which appears to be repeating pattern also growth or decline over the entire period in most recent around early 2024, followed by a decline.





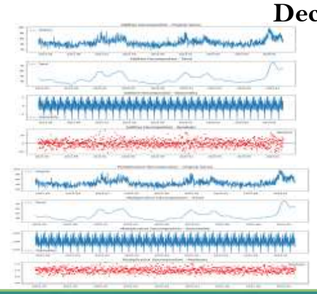
Box plot shows that usually, the typical range of patients tend to fall between **20** to **70**. The daily patient count exhibits a notable degree of variability, with prominently shown high outliers that suggest irregular peaks in patient counts. However, the outliers outside the graph have been kept in place as they represent the values of the real-time data that has been used.

Total days	Total patients	Minimum patients	Maximum patients	Avg. patients
1065	48446	16	97	45.48

48,446 walk-in patients, recorded in a span of **1065** days, demonstrate a high demand for service provided.

There were **16** patients in the leisurely day. On the busiest day, however, **97** patients were admitted, which might have placed a burden on the available resources. The daily average of the patients was **45.48**.

Decomposition



After comparison additive decomposition seems to be more appropriate by given residual's behavior it has slightly better as it shows more stability in the variance. The remaining constituents have a very steady behavior free of obvious trends, indicating that the breakdowns have effectively captured the main structures in the data. From the given graphs, it can also be assumed that seasonal variations are constant overtime. Furthermore, anomalies were detected and had been cleaned. However, both the decompositions are fairly good.

Stationarity Check

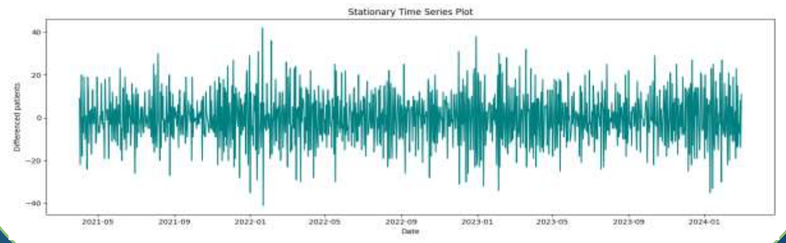
Test	Test statistic	P-value	Lags used	Critical val(1%)	Critical val(2.5%)	Critical val(5%)	Critical val(10%)
ADF	-2.89	0.046	22.00	-3.436	-	-2.86	-2.568
KPSS	0.5999	00.022	18.00	0.739	0.57	0.463	0.347

In order to determine whether the data is stationary or not, ADF and KPSS test were run. First Order Differencing was used to achieve stationarity because the preliminary testing showed that the data is not stationary and that a unit root exist.

Stationarity Check after first order

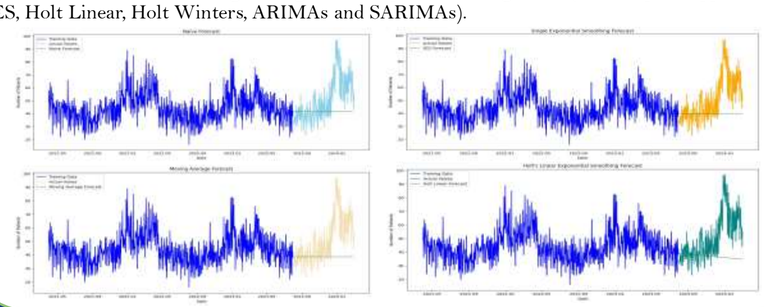
Test	Test statistic	P-value	Lags used	Critical val(1%)	Critical val(2.5%)	Critical val(5%)	Critical val(10%)
ADF	-7.45E+00	5.43E-11	2.20E+01	-3.436E+00	-	-2.86E+00	-2.56E+00
KPSS	0.014	0.100	15.00	0.739	0.57	0.463	0.347

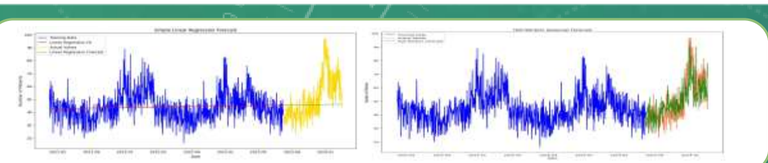
Below graph and above values shows that after successfully running ADF and KPSS test for first order differencing, the data is now stationary.



Baseline, Extrapolation and regression models

Given below are the various time series forecasting models which were applied on the training set (80%) and assessed on the test set(20%) in order to accurately reflect the changing patterns of the walk in patients. The following models were used: starting from Baseline and Simple approach (Naive, Mean, Moving Average, and Simple Linear Regression) to Complex Approaches (SES, Holt Linear, Holt Winters, ARIMAs and SARIMAs).





Error Statistics

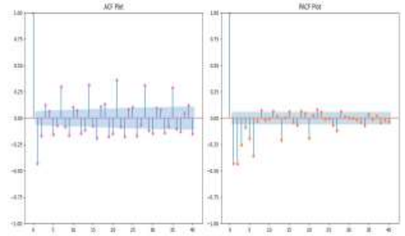
Model	Naive	Moving Avg.	Linear Reg.	Single Exp.	Holt's Linear	Holt Winters	ARIMA	SARIMA
MSE	346.74	419.34	287.90	400.28	444.38	179.40	393.48	450.54

Of all the models examined, the Holt-Winters Model has the best fit, as evidenced by its lowest MSE (179.40) and lowest MAPE (0.200).

Linear regression is a good alternative, but not as good as Holt-Winters, as it has the second lowest MSE (287.90) and second lowest MAPE (0.247).

The Naive Model fares marginally better than Holt-Winters and Linear Regression, but not by much.

ARIMA Models & ACF, PACF Analysis



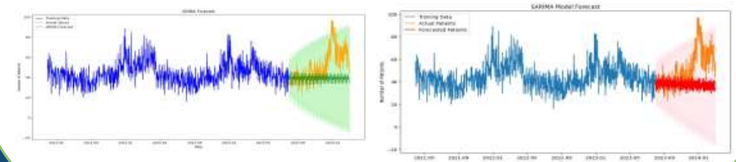
Seasonality in the data is indicated by regular intervals of spikes in the ACF plot. The ACF plot, which shows a steady decrease in correlations, verified the non-stationarity found in the earlier ADF tests.

In order to achieve steady data, first order differencing was applied. A notable peak first emerges in the PACF plot at lag 1, after which it surpasses the relevance threshold.

Now hypnosis which has a unit root is non-stationary. PACF graph indicates the p-value and q-value. While ACF plot indicates the auto-corelations between the data-time and day time series, the PACF indicates the partial corelations between the same.

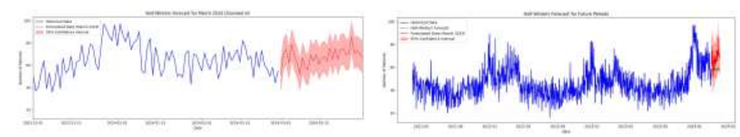
Model	ARIMA	SARIMA
AIC	5943.87	5932.38
BIC	6010.32	6003.36

Upon comparing the results of the ARIMA and SARIMA models, it was noted that SARIMA model's MSE value was higher as compared to that of ARIMA. However, when comparing their BIC value, the SARIMA model showed a lower count making it a better fit for the forecasting as compared to the ARIMA model, in this case.



Observations & Conclusions

With values varying between the upper 50s and low 60s, the anticipated patient count exhibits a comparatively constant pattern. This suggests that the quantity of patients is consistently in demand. Furthermore, the number of patients each day varies somewhat; on 2024-03-26, the greatest expected number is around 59.67, while on 2024-03-09, the lowest forecast figure is approximately 56.88. The data appears to show neither a significant rising nor decreasing trend over the course of this time, indicating that a steady patient population is anticipated.



Stable Demand: Throughout March 2024, the hospital or clinic can anticipate a steady and regular daily patient volume. Planning resources and making sure there is enough staff and inventory to meet patient demands depend on this stability.

Resource Allocation: More efficient planning of resources can be done in light of the forecast's relative stability. Hospital management can make sure there are neither shortages nor excesses of resources because they are aware that patient numbers will likely remain in the upper 50s.

No Significant Peaks or Troughs: Since the data do not indicate any significant peaks or troughs, no unusual steps are needed to address abrupt increases or decreases in the number of patients.

Forecast Reliability: The prediction indicates that the forecasting model is quite trustworthy for short-term forecasts, as evidenced by the forecast's minimal daily deviations. Operational planning can be done immediately with this reliability.