Analyzing Healthcare Call Patterns During Emergencies: Trends & Insights

BDA 602 Statistics for Business Analytics Final Project

Group Members: Muhammad Yasir, Fatemeh Shenavar, Piyush Chowhan

Submission Date: 05/01/2025
Group Members: Muhammad Yasir, Fatemeh Shenavar, Piyush Chowhan
Group Number: 1

Table of Contents

Submission Date: 05/01/2025
Group Members: Muhammad Yasir, Fatemeh Shenavar, Piyush Chowhan
Group Number: 1

Analyzing Healthcare Call Patterns During Emergencies: Trends & Insights

1. Introduction:

1.1 Research Question:

This project analyzes healthcare-related emergency call trends to help optimize resource allocation by healthcare providers and policymakers.

1.2. Relevance and Significance:

Analyzing healthcare call patterns improves resource management and response strategies, especially during crises like COVID-19.

1.3 Expected Impact:

Findings can improve resource allocation, minimize bottlenecks, and enhance patient satisfaction by optimizing service availability during emergencies.

2. Data Collection Details:

2.1. Data Source:

The data is sourced from Kaggle, specifically from the dataset titled "Healthcare Call Data Analysis During Emergency Times" by (Basak, 2025).

2.2. Dataset Summary:

Number of Observations: The dataset contains monthly data spanning several years, with each row representing a specific month and year.

Number of Variables:

i) Year ii) Month iii) Total Number of Calls iv) Number of Total Health Information v) Number of Total Ambulance Information vi) Number of Total Complaints vii) Number of Calls to Know About the Service

2.3. Data Limitations:

- **Missing Values:** No missing data was found; the dataset is complete and reliable.
- **Biases:** Significant bias exists toward emergency periods, especially during COVID-19 (2020–2021), with call volumes spiking sharply (e.g., 2,329,457 calls in June 2020).

Submission Date: 05/01/2025
Group Members: Muhammad Yasir, Fatemeh Shenavar, Piyush Chowhan
Group Number: 1

- **Potential Errors:** No inconsistencies like negative values were detected. Extreme outliers during emergencies reflect real-world demand surges and require careful analysis.

2.4. Dataset:

Year	Month	Total Number of Calls	Total Number of Doctors Consultancy	Number of Total Health Information	Number of Total Ambulance Information	Number of Total Complaints	Number of Calls To Know About The Service
2025	Feb	155200	98964	3777	5773	3706	42980
2025	Jan	130982	87815	3217	4179	3178	32593
2024	Dec	133725	86832	3634	3735	3035	36489
2024	Nov	153698	93710	3010	4421	4133	48424
2024	Oct	214604	108207	6307	5792	8885	80804
2024	Sept	157686	94893	3683	5685	4286	40889
2024	Aug	168418	100902	6435	6628	5267	41964
2024	July	190861	121500	5444	7501	5040	51376
2024	June	169089	111573	6774	8330	5414	36998
2024	May	149502	102415	4982	8173	5037	28895

3. Processing Plan:

3.1. Handling Missing Data:

After inspecting the dataset, there are no missing values in any of the columns. All rows are complete.

3.2. Outlier Detection & Treatment:

Outliers were identified in emergency periods but are not errors. Statistical methods will be used to analyze their impact on trends.

4. Descriptive Analysis:

4.1. Summary Statistics Table:

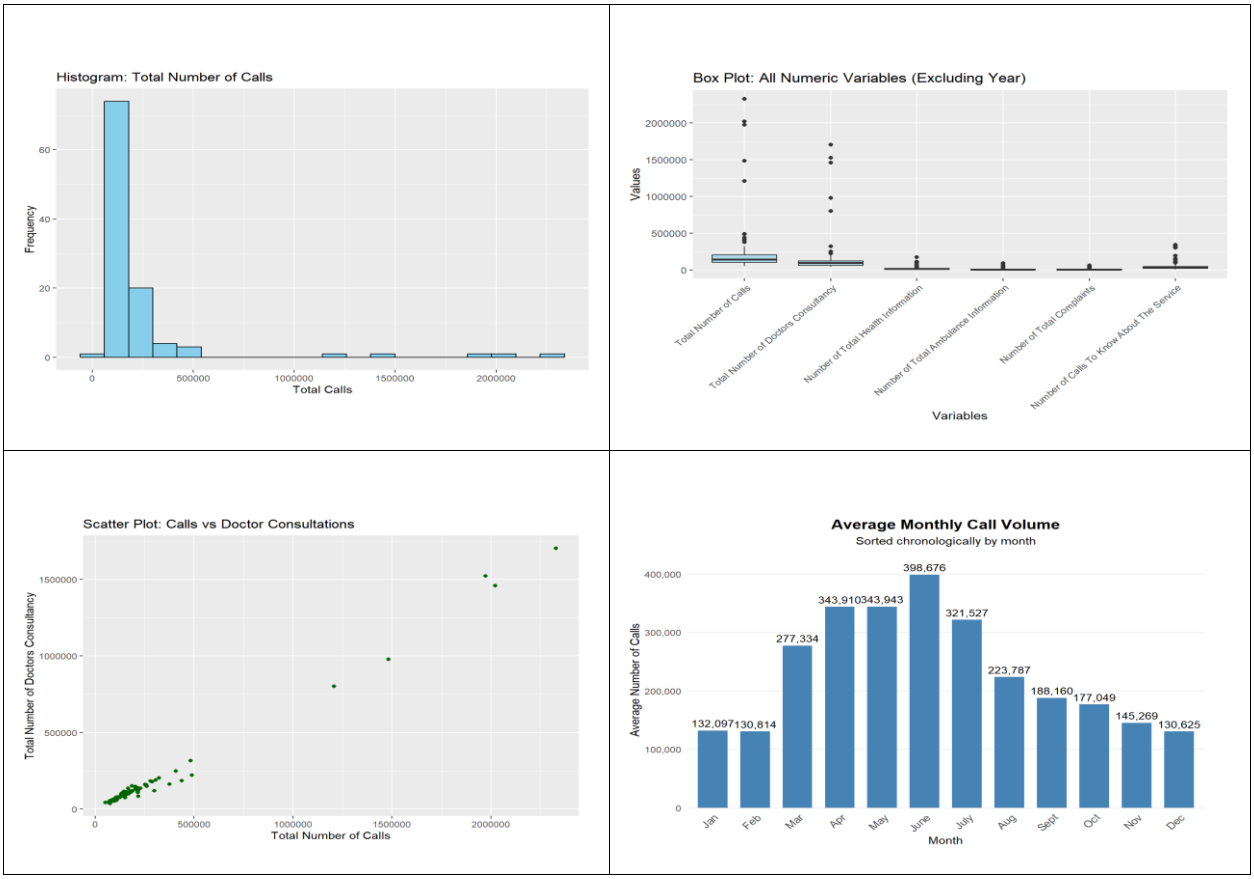
Summary Statistics Table

	Variable	Minimum	Maximum	Mean	Std_Dev
Total Number of Calls	Total Number of Calls	51432	2329457	234031.626	369700.08
Total Number of Doctors Consultancy	Total Number of Doctors Consultancy	38162	1703100	154455.346	268135.14
Number of Total Health Information	Number of Total Health Information	3010	173306	15714.729	23104.17
Number of Total Ambulance Information	Number of Total Ambulance Information	220	91751	9240.318	15330.68
Number of Total Complaints	Number of Total Complaints	77	59760	6924.458	10484.20
Number of Calls To Know About The Service	Number of Calls To Know About The Service	4075	342434	47509.103	61227.80

Submission Date: 05/01/2025
Group Members: Muhammad Yasir, Fatemeh Shenavar, Piyush Chowhan
Group Number: 1

4.2. Graphical Visualization:

4.2.1. Histogram, Box Plot, Scatter Plot & Bar Chart:



4.3. Interpretations:

4.3.1. Summary Statistics Table:

- Minimum/Maximum: Reveals the range of values for each metric.
- Mean: Indicates average values (e.g., average call volume).
- Standard Deviation: Shows how spread out the data is.

Large standard deviations relative to means suggest high variability in call volumes/ service usage.

Histogram: • Right-skewed distribution. • Most month had <500K calls but rare surge exceed 1.5M.	**Box Plot:** • Doctor Consultancy has the largest IQRs, confirming high variability.
Scatter Plot: • Points cluster diagonally – higher calls = more consultations.	**Bar Chart:** • Peak in April/May: ~344K calls. • Trough in Nov/Dec: ~130K calls. • Secondary Peak: August (322K call)

Submission Date: 05/01/2025
Group Members: Muhammad Yasir, Fatemeh Shenavar, Piyush Chowhan
Group Number: 1

5. Handling Data Issues:

- Confirms COVID-19 period as legitimate extreme values.
- Recommends log transformation for modeling.
- Suggests time series decomposition for seasonality.

6. Hypothesis Test: One-Sample T-test

6.1. Business Question:

Is the average number of complaints significantly different from 5000?

6.2. Hypothesis:

1. H0 (Null): $\mu = 5000$
2. H1 (Alternative): $\mu \neq 5000$

6.3. Test Results:

1. Sample Mean: 6924.46
2. t-statistic: 1.90
3. p-value: 0.060
4. 95% Confidence Interval for the Mean: (4937.91, 8911.01)

6.4. Interpretation:

1. The p-value for the one-sample t-test is 0.060, which is slightly above the common significance level of 0.05.
2. Therefore, we fail to reject the null hypothesis.
3. This means there is not enough statistical evidence to conclude that the average number of complaints is significantly different from 5000.

6.5. Business Implication:

1. Based on this test, we cannot confidently say the number of complaints is significantly higher or lower than 5000.
2. More data or further segmentation might be needed to explore complaint patterns in specific time periods or conditions.

```
## One Sample t-test
## data:  data$Number.of.Total.Complaints
## t = 1.8987, df = 106, p-value = 0.06032
## alternative hypothesis: true mean is not equal to 5000
## 95 percent confidence interval:
## 4915.004 8933.912
## sample estimates:
## mean of x
## 6924.458
## Result: p-value = 0.06031817 → Fail to reject the null hypothesis. No significant difference from 5000.
```

7. Hypothesis Test: Two-Sample T-test

7.1. Business Question:

Is there a significant difference in the average number of healthcare calls between the COVID period and non-COVID years?

7.2. Hypothesis:

1. H0 (Null): There is no difference in the mean number of calls between COVID and non-COVID periods.
2. H1 (Alternative): The mean number of calls is different between the two periods.

7.3. Test Results:

1. Test Used: Welch's t-test (unequal variances)
2. T-statistics: 3.43
3. p-value: 0.0023
4. 95% confidence interval for difference in means: (200,907 to 735,220)

7.4. Interpretation:

1. Since the p-value < 0.05, we reject the null hypothesis.
2. There is a statistically significant increase in the number of calls during the COVID period.

7.5. Business Implication:

1. Healthcare providers should prepare for major increases in service demand during emergency periods like pandemics.

2. Resource allocation and staffing plans should reflect such spikes to avoid system overload.

```
## Welch Two Sample T-test
## data:  covid_calls and non_covid_calls
## t = 3.434, df = 23.061, p-value = 0.002258
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 186137.4 749989.3
## sample estimates:
## mean of x mean of y
## 597108.8 129045.5
## Result: p-value < 0.002258007 → Reject the null hypothesis (H₀).
```

8. Regression Model Setup 1:

8.1. Business Question:

Which type of healthcare inquiries contribute most to the total call volume received by the healthcare call center?

8.2. Dependent Variable (DV):

Total Calls: Total number of calls received in each period.

8.3. Independent Variable (IV):

IV Name	Description	Expected Impact on Total Calls
Doctors Consultancy	Calls seeking doctor consultation	Positive — core reason for calling
Health Info	Calls asking about health information	Positive — common public concern
Ambulance Info	Calls about ambulance services	Positive — may be less frequent
Complaints	Calls reporting complaints	Positive — adds to total volume
Service Info Calls	Calls to ask about the service itself	Positive — reflects awareness needs

9. Regression Results:

Predictor	Coefficient	Std. Error	t-value	p-value	95% Confidence Interval	Significance
Intercept	41.69	156.23	0.27	0.79	[-268.23, 351.61]	✗
Doctors Consultancy	0.9991	0.0013	786.87	< 2e-193	[0.9966, 1.0016]	✓

Health Info	0.9769	0.0124	78.91	< 1.4e-92	[0.9523, 1.0014]	✓
Ambulance Info	0.9503	0.0268	35.45	< 8.4e-59	[0.8971, 1.0035]	✓
Complaints	1.0112	0.0315	32.07	< 9.3e-55	[0.9487, 1.0738]	✓
Service Info Calls	1.0217	0.0089	115.39	< 4.7e-109	[1.0041, 1.0393]	✓

- R^2: 0.99999 (nearly perfect fit)
- Adjusted R^2: 0.99999
- Standard Error of the Model: 26.05

9.1. Independent Variable:

- All service types significantly predict total call volume.
- Each service call increases total calls by approximately one unit.
- Doctor consultations have the strongest impact, with a coefficient near 1.0 and a very high t-value.
- Service info calls slightly exceed a 1-to-1 contribution, possibly due to multi-purpose calls.

9.2. Business Impact:

- Staff and planning should prioritize doctor consultations and service info lines.
- The near-linear contribution of each service type makes the model useful for forecasting.

10. Model Strengths and Limitations:

- $R^2 \approx 1$: The model explains virtually all the variation in total call volume.
- All predictors are statistically significant.
- The relationship is intuitive and easy to communicate.
- The model is descriptive and omits time trends, seasonality, and external factors.
- Some service categories may overlap.
- Residual analysis and multi-collinearity checks have not yet been performed.

10.1. Refinements to Consider:

- Add Month and Year as predictors to capture time trends.
- Use interaction terms to detect service-specific seasonal effects.

- Test for multicollinearity with VIF.
- Visualize residuals to identify non-linearity and outliers.

11. Regression Model Setup 2:

11.1. Business Question:

What factors influence the number of complaints received by the healthcare call center over time?

11.2. Dependent Variable (DV): Number of Total Complaints

IV Name	Description	Expected Impact on Complaints
Doctors Consultancy	Indicates interaction with doctors	Positive - if doctor-related issues exist
Health Info	General info calls	Neutral
Ambulance Info	Emergency-related interactions	Positive - delays/availability could cause frustration
Total calls	Overall workload	Positive - higher load may strain quality
Service Info Calls	Reflects understanding/ confusion about offerings	Positive - could reflect unmet expectations

11.3. Independent Variable (IV):

12. Regression Results:

Predictor	Estimate	Std. Error	t value	p-value	Significance
Intercept	-25.64	147.46	-0.174	0.862	✗
Doctors Consultancy	-0.9	0.028	-32.26	< 2e-16	✓
Health Info	-0.87	0.032	-27	< 2e-16	✓
Ambulance Info	-0.82	0.045	-18.27	< 2e-16	✓
Total Calls	0.9	0.028	32.07	< 2e-16	✓
Service Info Calls	-0.92	0.031	-29.45	< 2e-16	✓

- R^2: 0.990 Adjusted R^2: 0.9898
- Residual Std. Error: 1061 on 101 df
- F-statistic: 2050 ($p < 2.2e-16$)
- The model is statistically significant and explains 99% of the variation.

Submission Date: 05/01/2025
Group Members: Muhammad Yasir, Fatemeh Shenavar, Piyush Chowhan
Group Number: 1

12.1. Model Comparison: Total Calls Vs. Complaints

Feature	Model 1: Total Calls	Model 2: Complaints
Dependent Variable	Total number of calls	Number of complaints
Purpose	Understand drivers of volume	Understand drivers of dissatisfaction
Model Fit (R^2)	0.99999 (extremely high)	0.990 (still excellent)
Predictor Significance	All significant	All significant
Best Use Case	Forecasting demand and staffing	Improving quality and complaint management
Interpretation Simplicity	High (coefficients ~1)	Moderate (mixed signs, nuanced)
Sensitivity	Stable, high linearity	Sensitive to call type mix

13. Discussion & Recommendations:

13.1. Business Implications:

- Negative coefficients (doctoral consultations, health info, service info) indicate reduced complaints and higher satisfaction.
- A positive total calls coefficient links complaint increases to system strain.
- Staff training, education, and better service visibility could lower complaints.
- Ambulance support performs well and could serve as a quality benchmark.

13.2. Future Research Directions:

- Add temporal features (month, day, event flags) to capture trends or anomalies.
- Use sentiment analysis or customer satisfaction scores to measure qualitative experience.
- Test for multicollinearity and explore dimensionality reduction (PCA) if needed.
- Expand analysis with categorical predictors like service type or call reason tags

References

Basak, S. K. (2025). *Healthcare Call Data Analysis DuringEmergencyTimes.* Kaggle. doi:10.34740/KAGGLE/DSV/11067536