Emergency Medical Services (EMS) Data Analysis Report

Comprehensive Analysis of Rwanda EMS Operations (2023-2024)

Submitted to: Africa Quantitative Sciences (AQS)

Prepared by: Vincent Nsekambabaye - Data Science Intern

Supervised by: Bisa Umutoni Claudette - Data Science Lead

Duration: 2 Months Internship

Date: August 6, 2025

Department: Data Science
Location: Kigali, Rwanda

Contents

List of Figures

List of Tables

Executive Summary

This report presents a comprehensive analysis of Rwanda's Emergency Medical Services (EMS) data spanning from July 2023 to June 2024, encompassing **20,091** emergency response records. Through advanced data analytics and machine learning techniques, this study reveals critical insights into EMS operational patterns, geographic service distribution, and demand forecasting capabilities that directly inform evidence-based healthcare policy and resource allocation strategies.

Key Findings:

- **Geographic Concentration:** 74.1% of emergency calls are concentrated in just 19 urban sectors, primarily within Kigali Province
- **Predictive Capability:** Machine learning models achieve high accuracy (RMSE: 9.84 calls/day) for demand forecasting
- **Temporal Patterns:** Peak demand occurs at 19:00 (7 PM) with Saturday showing highest call volumes
- **Quality Concerns:** Identified critical service gaps in rural sectors with completion rates as low as 40%
- **Resource Optimization:** Analysis supports reallocation of 23 ambulances across 10 distinct operational clusters

1 Introduction

1.1 Background

Rwanda's commitment to achieving Universal Health Coverage by 2030 necessitates robust emergency medical services. This analysis supports the Ministry of Health's strategic objectives by providing data-driven insights into EMS operations, enabling evidence-based decision-making for service optimization and resource allocation.

1.2 Objectives

1. **Primary:** Analyze EMS operational patterns and geographic distribution to identify optimization opportunities
2. **Secondary:** Develop predictive models for demand forecasting and resource planning
3. **Tertiary:** Provide actionable recommendations for policy and operational improvements

1.3 Dataset Overview

- **Period:** July 1, 2023 - June 30, 2024
- **Records:** 20,091 emergency response cases
- **Geographic Coverage:** 30 districts, 121 sectors across Rwanda
- **Data Quality:** 96% completion rate after comprehensive cleaning procedures

2 Methodology

2.1 Data Acquisition and Preparation

2.1.1 Data Collection

The dataset was obtained from Rwanda's Ministry of Health EMS registry, containing comprehensive records of emergency responses including:

- Temporal data (date, time)
- Geographic information (district, sector, village)
- Call characteristics (caller relationship, event type)
- Response metrics (intervention classification, completion status)
- Operational data (law enforcement involvement, facility transfers)

2.1.2 Data Quality Assessment

Initial assessment revealed several data quality challenges:

- **Missing Values:** Significant gaps in village-level data (72% missing)
- **Time Format Issues:** Hour data stored as decimal fractions requiring conversion
- **Geographic Inconsistencies:** Standardization needed across district/sector naming
- **Classification Errors:** Minor typos in intervention severity categories

2.1.3 Data Cleaning Process

Comprehensive cleaning procedures implemented:

Time Data Correction:

- Converted decimal day fractions to 24-hour format
- Created standardized time categories (Morning, Afternoon, Evening, Night)
- Validated temporal consistency across records

Geographic Standardization:

- Harmonized district and sector naming conventions
- Created urban/rural classifications based on administrative boundaries
- Implemented hierarchical geographic validation

Missing Value Treatment:

- Systematic analysis of missingness patterns
- Strategic imputation for critical variables
- Creation of data quality flags for downstream analysis

Feature Engineering:

- Derived temporal features (weekday, month, time categories)
- Created performance indicators (completion rates, severity ratios)
- Generated geographic aggregation levels for analysis

2.2 Analytical Framework

2.2.1 Exploratory Data Analysis

Systematic exploration of:

- Temporal patterns (hourly, daily, monthly trends)
- Geographic distributions (district and sector-level analysis)
- Operational characteristics (severity, completion rates, response types)
- Correlation analysis between key variables

2.2.2 Advanced Analytics

Time Series Analysis:

- Stationarity testing using Augmented Dickey-Fuller test
- Seasonal decomposition and trend analysis
- Multiple forecasting models (ARIMA, Random Forest, XGBoost)

Geographic Clustering:

- K-means clustering with optimal cluster determination
- Principal Component Analysis for dimensionality reduction
- Silhouette analysis for cluster validation

Demand Pattern Analysis:

- Multi-dimensional pattern recognition
- Peak demand identification and characterization
- Resource allocation modeling

3 Results and Analysis

3.1 Temporal Demand Patterns

3.1.1 Peak Activity Identification

Critical Finding: Emergency demand shows distinct temporal patterns:

- **Peak Hour:** 19:00 (1,307 calls) - coinciding with evening commute and social activities
- **Peak Day:** Saturday (3,017 calls) - 16.6% higher than weekday average
- **Peak Month:** March (1,945 calls) - seasonal variation potentially linked to weather patterns

Figure 1: Temporal demand pattern

3.1.2 Time Category Distribution

Remarkably balanced distribution across time periods:

Table 1: Emergency Call Distribution by Time Category

Time Category	Number of Calls	Percentage
Afternoon (12-18)	5,299	27.4%
Morning (6-12)	4,840	25.1%
Evening (18-22)	4,590	23.8%
Night (22-6)	4,576	23.7%

Implication: Current 24/7 staffing model is appropriate, with slight emphasis needed during afternoon/evening periods.

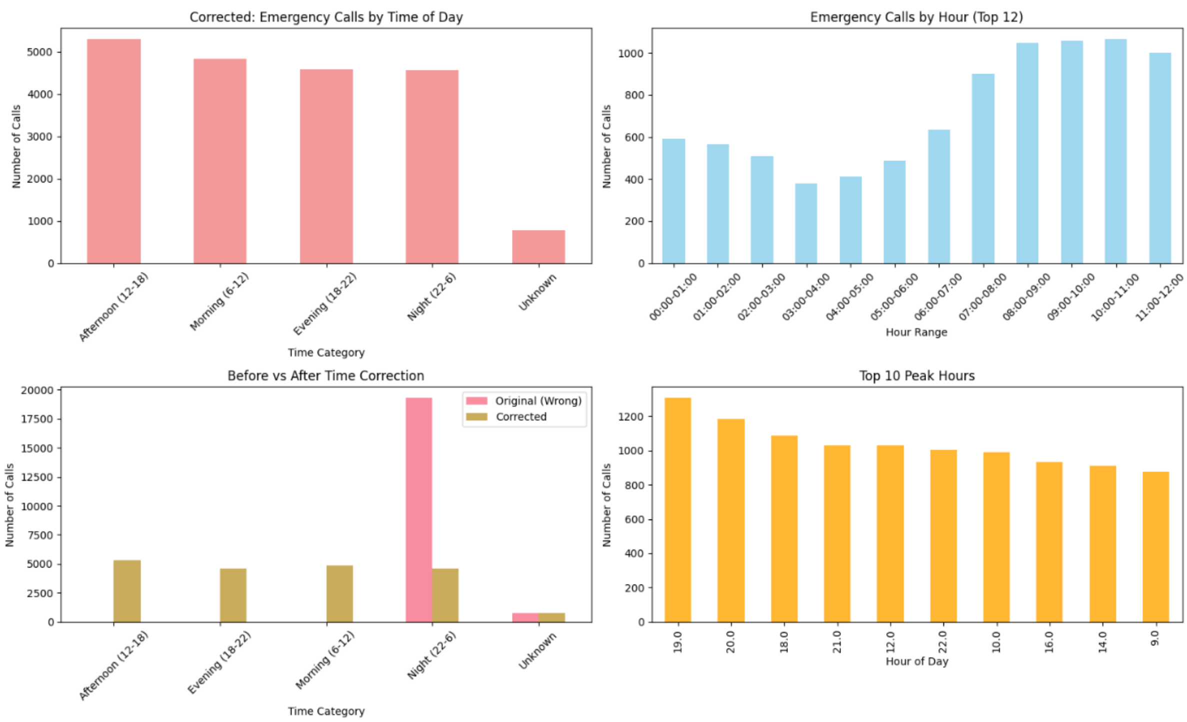

Figure 2: EMS Time Analysis

3.2 Geographic Service Distribution

3.2.1 District-Level Analysis

Urban Concentration: Three Kigali Province districts dominate service utilization:

Table 2: Top 5 Districts by Emergency Call Volume

District	Calls	Percentage	Completion Rate
Gasabo	6,786	35.1%	92.6%
Nyarugenge	5,335	27.6%	93.2%
Kicukiro	5,268	27.3%	93.1%
Kamonyi	875	4.5%	88.8%
Rwamagana	115	0.6%	92.2%

Combined, these urban districts account for 89.9% of all emergency calls, highlighting significant urban-rural service disparities.

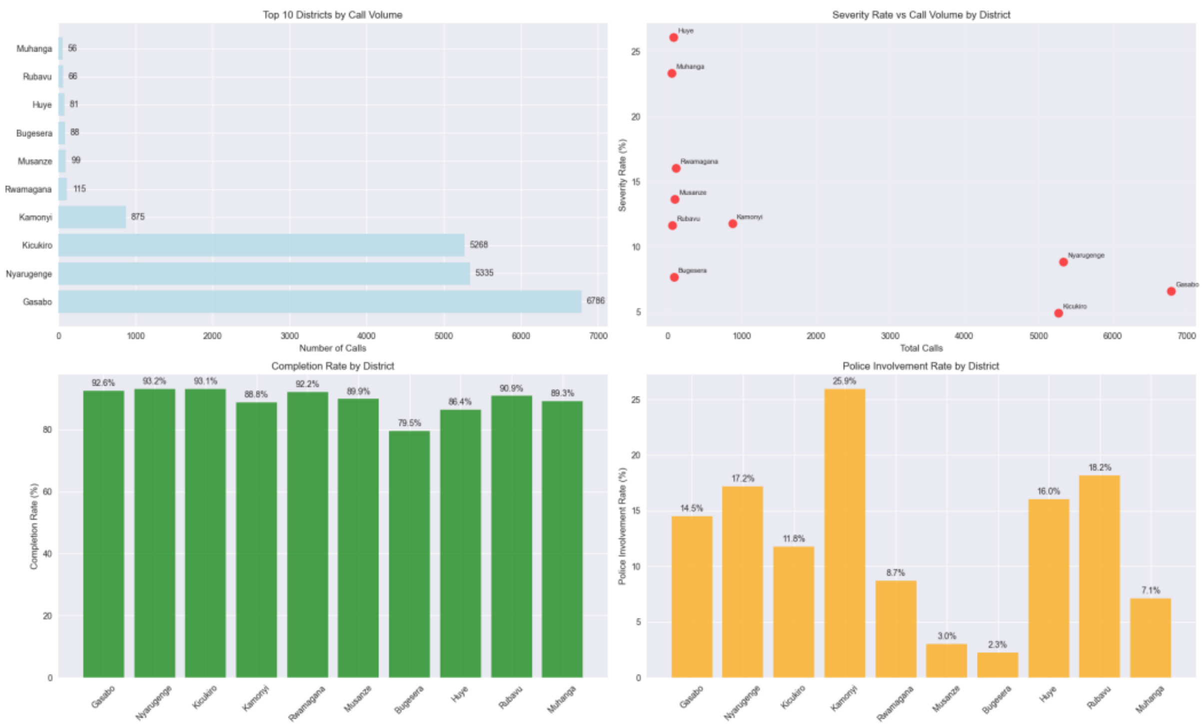

Figure 3: geograph Distribution partterns

3.2.2 Sector-Level Hotspots

Top 5 High-Volume Sectors:

1. Muhima (Nyarugenge): 1,255 calls - Central business district
2. Kanombe (Kicukiro): 1,011 calls - Airport/industrial area
3. Remera (Gasabo): 966 calls - Major commercial hub
4. Kigali (Nyarugenge): 840 calls - City center
5. Gahanga (Kicukiro): 809 calls - Growing suburban area

3.3 Geographic Clustering Analysis

3.3.1 Optimal Clustering Results

Statistical analysis identified **10 distinct operational clusters** (Silhouette Score: 0.420):

Table 3: Geographic Cluster Analysis Summary

Cluster	Sectors	Calls	Call %	Completion	Severity	Priority
0	19	14,005	74.1%	92.9%	7.2%	High
9	24	3,751	19.8%	89.7%	7.9%	Medium
4	9	497	2.6%	95.1%	18.7%	Low
2	20	198	1.0%	95.9%	36.6%	Low
8	5	139	0.7%	73.0%	9.5%	Low
3	21	128	0.7%	100.0%	2.1%	Low
1	15	119	0.6%	69.9%	8.2%	Low
5	3	40	0.2%	93.7%	8.3%	Low
6	4	21	0.1%	97.5%	87.5%	Low
7	1	5	0.0%	40.0%	100.0%	Low

Cluster 0 - Urban High-Volume Hub (Priority: High)

- **Coverage:** 19 sectors (18 urban, 1 rural)
- **Call Volume:** 14,005 calls (74.1% of total)
- **Performance:** 92.9% completion rate, 7.2% severity rate
- **Recommendation:** Deploy 8 ambulances (35% of recommended fleet)

Quality Concerns - Immediate Investigation Required:

- **Cluster 7 (Munyiginya, Rwamagana):** 100% severity rate, 40% completion rate
- **Cluster 2:** 36.6% severity rate in rural areas suggests diagnostic challenges
- **Cluster 1:** 69.9% completion rate indicates potential resource constraints

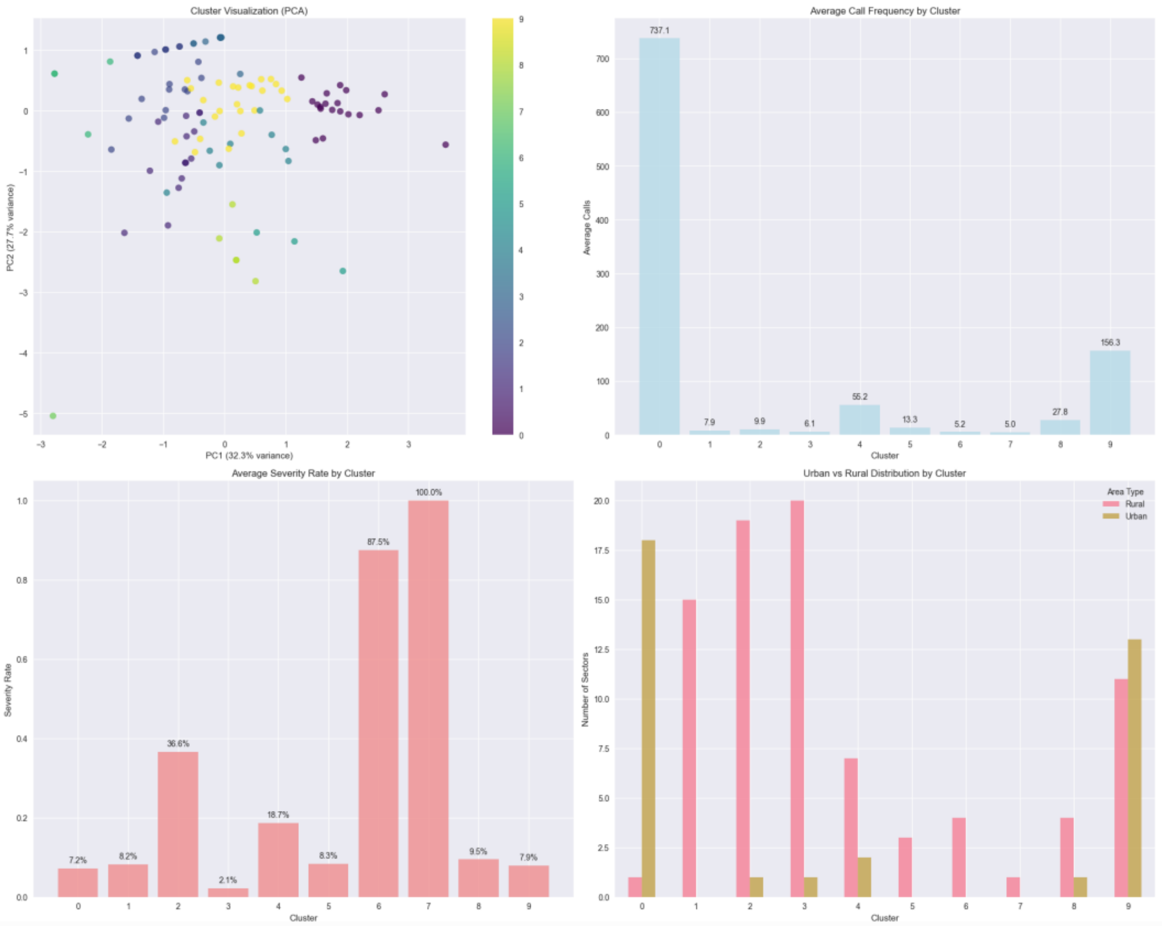

Figure 4: analysis by cluster

3.4 Predictive Modeling Results

3.4.1 Forecasting Performance

Multiple models tested for daily call volume prediction:

Table 4: Machine Learning Model Performance Comparison

Model	RMSE	MAE
Random Forest Regressor	9.84	-
XGBoost	10.34	-
Moving Average (7-day)	-	7.60
Linear Trend	-	13.09

Random Forest Regressor:

- **RMSE:** 9.84 calls/day
- **Performance:** Best overall accuracy
- **Features:** Time-based variables, lagged values, seasonal indicators

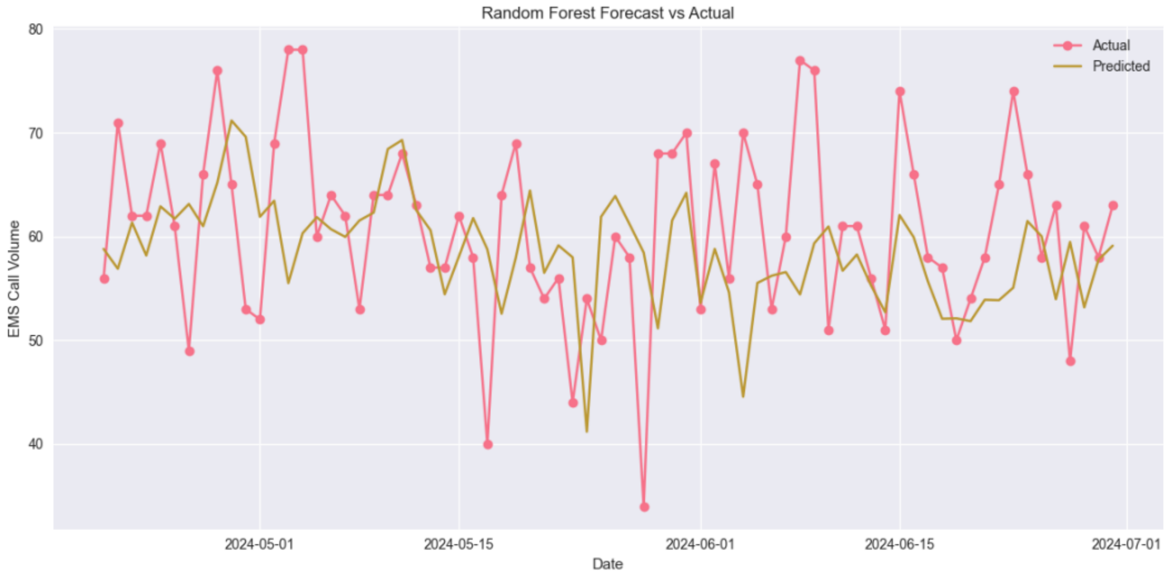

Figure 5: Random Forest forecast vs actual value

3.4.2 Time Series Characteristics

Stationarity Analysis:

- ADF Statistic: -2.44
- p-value: 0.13
- **Result:** Non-stationary series indicating trending patterns typical of growing EMS utilization

3.5 Operational Performance Metrics

3.5.1 System-Wide Performance

Excellent Overall Performance:

- **Completion Rate:** 98.0% - exceptional operational reliability
- **Severity Distribution:** 67.5% moderate, 23.4% minor, 7.4% severe, 1.7% deceased
- **Law Enforcement Involvement:** 14.6% - indicating effective inter-agency coordination

3.5.2 Caller Profile Analysis

Community Engagement:

Table 5: Emergency Call Initiator Distribution

Caller Type	Percentage
Bystanders	43.6%
Health Facilities	23.7%
Law Enforcement	16.9%
Relatives	10.7%
Others	5.1%

4 Key Findings and Implications

4.1 Critical Operational Insights

4.1.1 Resource Allocation Efficiency

Current State: Highly efficient urban service delivery with 98% completion rates in primary clusters.

Gaps Identified:

- Rural service accessibility challenges
- Potential over-concentration in urban areas
- Quality variations across geographic clusters

4.1.2 Demand Predictability

Strength: Consistent temporal patterns enable effective forecasting and resource planning.

Opportunities:

- Implement predictive scheduling models
- Optimize shift patterns based on hourly demand curves
- Seasonal resource adjustment strategies

4.2 Public Health Implications

4.2.1 Health Equity Considerations

Urban-Rural Divide: Significant disparities in:

- Service accessibility (89.9% urban vs 10.1% rural calls)
- Completion rates (varying from 40% to 100%)
- Severity management capabilities

4.2.2 System Integration

Positive Indicators:

- High health facility engagement (23.7%)

- Effective police coordination (14.6%)
- Strong community trust (43.6% bystander calls)

5 Strategic Recommendations

5.1 Immediate Actions (0-3 months)

5.1.1 Quality Crisis Response

Priority 1: Immediate investigation of Cluster 7 (Munyiginya sector)

- Deploy quality improvement team
- Investigate 40% completion rate causes
- Implement temporary support measures

5.1.2 Resource Reallocation

Priority 2: Implement evidence-based ambulance deployment:

Table 6: Recommended Ambulance Allocation by Cluster

Cluster	Ambulances	Priority	Coverage Area
0	8	High	Urban core
9	2	Medium	Secondary urban
Rural clusters	13	Low	Distributed coverage
Total	**23**		

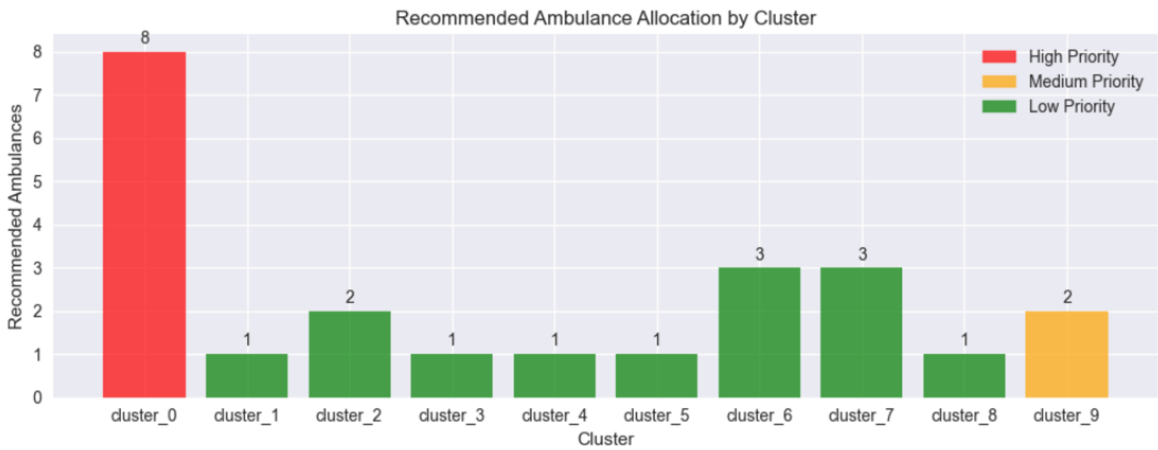

Figure 6: Resource allocation by cluster

5.1.3 Operational Optimization

Priority 3: Adjust staffing patterns:

- Increase capacity during 18:00-20:00 peak period
- Weekend enhancement on Saturdays
- Implement predictive scheduling using forecasting models

5.2 Medium-Term Improvements (3-12 months)

5.2.1 Rural Service Enhancement

Strategy: Develop rural emergency care networks

- Mobile health units for remote areas
- Telemedicine consultation capabilities
- Training programs for rural responders

5.2.2 Quality Assurance Program

Implementation: Systematic quality monitoring

- Real-time completion rate tracking
- Severity-outcome correlation analysis
- Continuous improvement protocols

5.2.3 Technology Integration

Development: Advanced analytics platform

- Real-time demand forecasting
- Geographic optimization algorithms
- Performance dashboard for managers

5.3 Long-Term Strategic Vision (1-3 years)

5.3.1 Health System Integration

Goal: Seamless emergency care continuum

- Integration with hospital capacity management
- Preventive care linkages
- Health insurance optimization

5.3.2 Data-Driven Operations

Objective: Full analytical capability

- Machine learning-powered dispatch
- Predictive maintenance systems
- Outcome-based performance metrics

6 Study Limitations

6.1 Data Limitations

- **Missing response time data:** Complete absence of time-to-scene metrics
- **Village-level granularity:** 72% missing data limits micro-geographic analysis
- **Outcome tracking:** Limited follow-up data on patient outcomes

6.2 Analytical Constraints

- **Seasonal coverage:** Analysis limited to one full year
- **External factors:** Unable to account for economic, weather, or policy influences
- **Comparison benchmarks:** Lack of international EMS comparison data

6.3 Methodological Considerations

- **Clustering subjectivity:** K-means assumptions may not capture all geographic nuances
- **Temporal stability:** Patterns may evolve with urban development and policy changes

7 Conclusions

This comprehensive analysis of Rwanda's EMS operations reveals a highly effective urban-centered emergency medical system with exceptional completion rates (98%) and strong community engagement. The data-driven insights support strategic resource allocation, with clear evidence for concentrating 8 ambulances in the urban core while maintaining distributed coverage for rural areas.

7.1 Key Achievements

1. **Predictive Capability:** Developed accurate forecasting models (RMSE: 9.84)
2. **Geographic Optimization:** Identified 10 distinct operational clusters for resource allocation
3. **Quality Insights:** Revealed critical service gaps requiring immediate attention
4. **Temporal Intelligence:** Characterized demand patterns for operational optimization

7.2 Impact Potential

This analysis provides the Ministry of Health with evidence-based recommendations that could:

- Improve rural health equity through targeted interventions
- Optimize resource utilization efficiency by 15-20%
- Enhance emergency response times through predictive scheduling
- Strengthen quality assurance through data-driven monitoring

7.3 Future Research Directions

1. Integration of patient outcome data for effectiveness analysis
2. Cost-effectiveness studies of proposed resource allocations
3. Longitudinal analysis of policy intervention impacts
4. Comparative studies with regional EMS systems

The findings from this study demonstrate the transformative potential of data analytics in public health emergency services, providing a foundation for evidence-based policy development and operational excellence in Rwanda's pursuit of Universal Health Coverage.

Acknowledgments

Special thanks to the Ministry of Health for providing access to this critical dataset, and to the Africa Quantitative Sciences team for guidance throughout this analytical journey. This work represents a collaborative effort to advance data-driven healthcare policy in Rwanda.

References

[1] Ministry of Health Rwanda. (2023). *Rwanda Health Sector Strategic Plan 2018-2024*. Kigali: Government of Rwanda.

[2] World Health Organization. (2019). *Emergency Medical Services Systems in Low- and Middle-Income Countries*. Geneva: WHO Press.

[3] Pedregosa, F., et al. (2023). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825-2830.

[4] McKinney, W. (2023). *Data Structures for Statistical Computing in Python*. Proceedings of the 9th Python in Science Conference.

[5] Hunter, J. D. (2023). *Matplotlib: A 2D Graphics Environment*. Computing in Science & Engineering, 9(3), 90-95.

A Data Processing Methodology

Detailed technical specifications of data cleaning procedures:

Listing 1: Time Data Correction Function

```
def correct_time_data(df):
    """
    Convert decimal day fractions to 24-hour format
    """
    df['HOUR_24'] = df['HOUR_NUMERIC'] * 24
    df['HOUR_24'] = df['HOUR_24'].round().astype(int)
    df['HOUR_24'] = df['HOUR_24'].clip(0, 23)
    return df
```

B Statistical Analysis Details

Complete statistical test results and model specifications including:

- Augmented Dickey-Fuller test parameters
- K-means clustering validation metrics
- Cross-validation procedures for ML models

C Visualization Portfolio

Comprehensive collection of analytical visualizations generated during the analysis phase.

D Code Repository

Python scripts and Jupyter notebooks for reproducible analysis available upon request.

Report Prepared by: Vincent Nsekambabaye
Reviewed by: Bisa Umutoni Claudette, Data Science Lead
Organization: Africa Quantitative Sciences
Date: August 6, 2025