B. JOSHNA

Healthcare Re-admission Analysis Report

1. Dataset Description

1.1 Source:
The dataset used for this project is Healthcare Readmissions.csv containing patient records.
It captures hospital stays, medical procedures, medications, visits, and readmission status.

1.2 Columns:
- readmitted: Whether the patient was readmitted (yes/no)
- time_in_hospital: Length of hospital stay
- n_lab_procedures: Number of lab procedures performed
- n_procedures: Number of medical procedures performed
- n_medications: Number of medications prescribed
- n_outpatient: Outpatient visits
- n_inpatient: Inpatient visits
- n_emergency: Emergency visits
- age: Age group of the patient
- medical_specialty: Specialty of the admitting physician

1.3 Data Quality:
- Missing values (e.g., '?') were handled and converted to null.
- Numeric fields were cast into integers.
- Target column 'readmitted' was encoded as 1 (yes) and 0 (no).
- The dataset supports statistical aggregation and visualization.

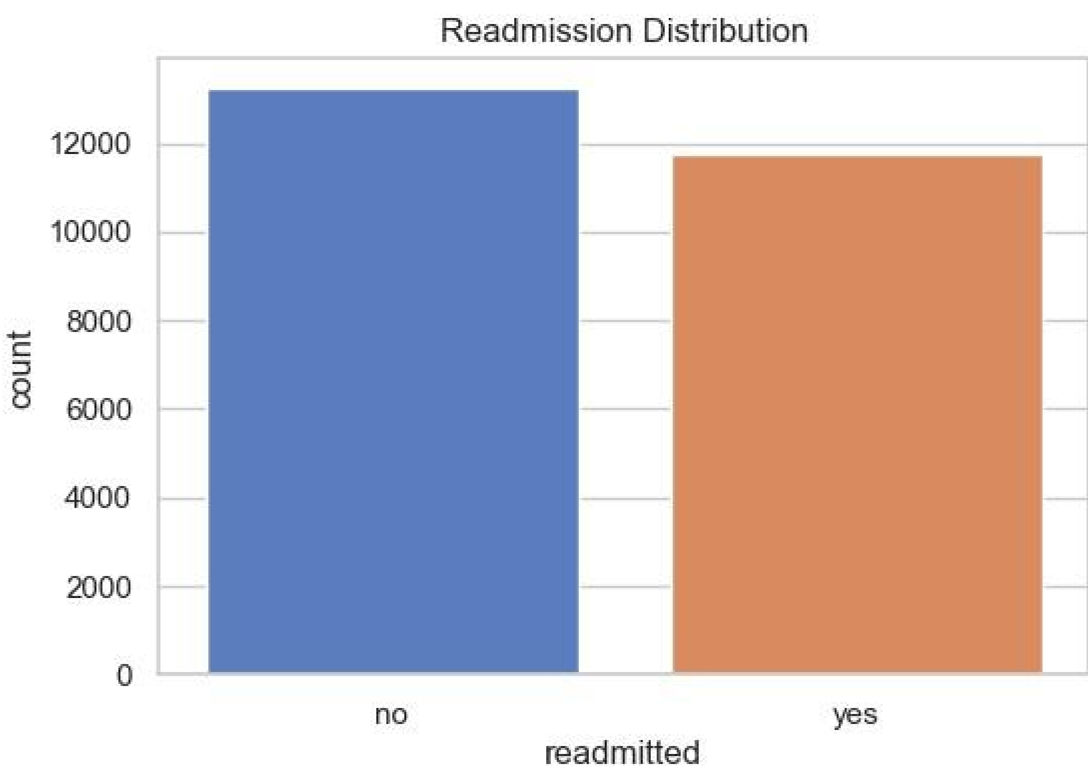
2. Operations Performed

2.1 Data Loading and Inspection
- Loaded the dataset using PySpark for scalability.
- Replaced missing values and casted numeric columns.
- Verified schema and previewed records.
- Converted Spark DataFrame to Pandas for visualization.

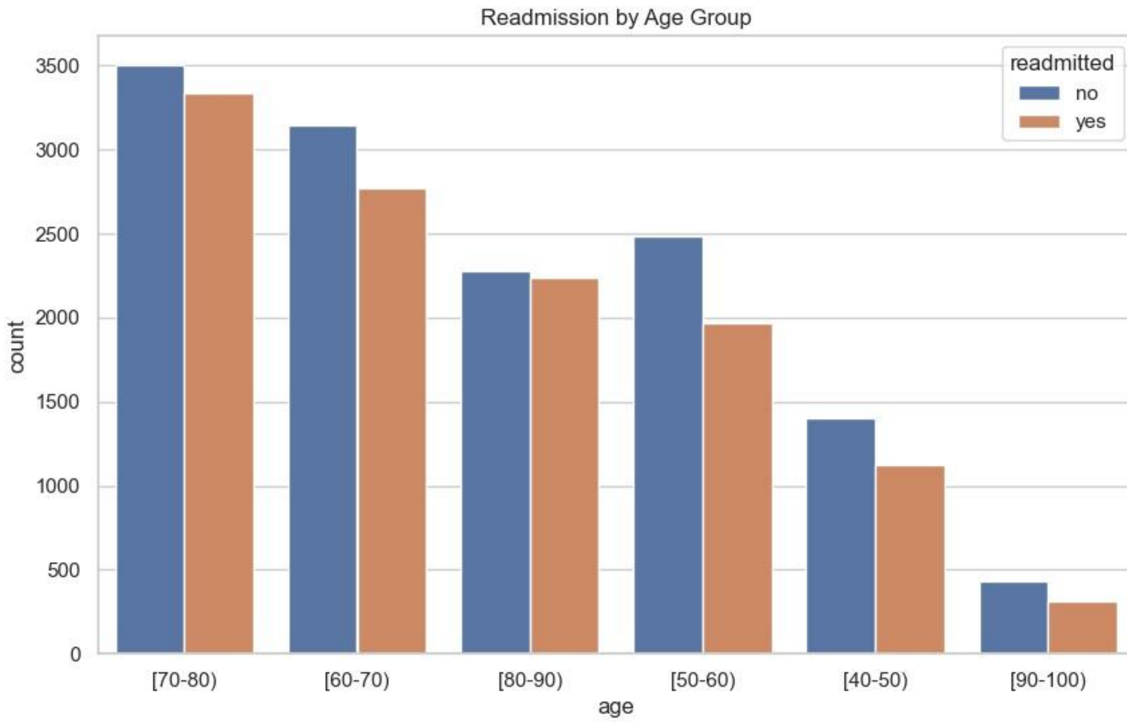
2.2 Aggregations and Visualizations
- Computed averages and grouped statistics by readmission status.
- Generated visualizations using Seaborn and Matplotlib:
• Readmission Distribution
• Readmission by Age Group
• Hospital Stay Duration vs Readmission
• Medications vs Readmission
• Medical Specialty vs Readmission (Top 10)
• Correlation Heatmap

3. Key Insights

3.1 Readmission Volume
- The dataset shows a significant proportion of patients are readmitted. - Distribution varies across age groups and treatment characteristics.

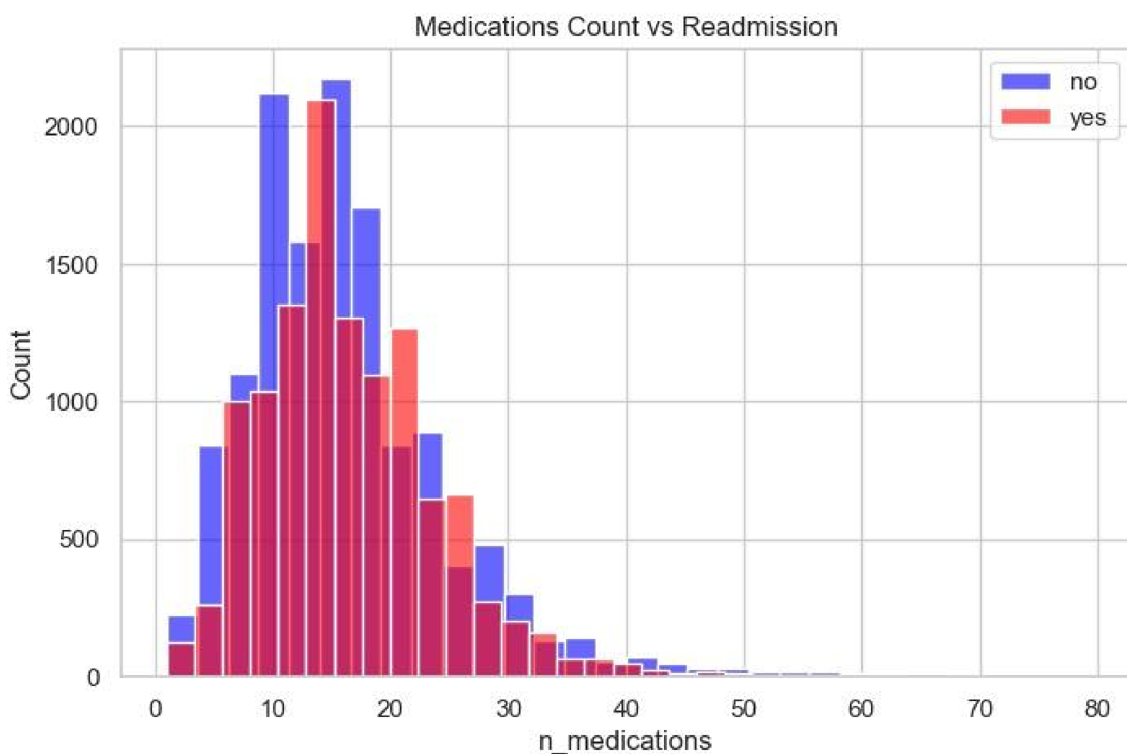


3.2 Age Group Trends
- Certain age groups show higher readmission rates.
- Elderly patients tend to have longer hospital stays and higher readmission likelihood.



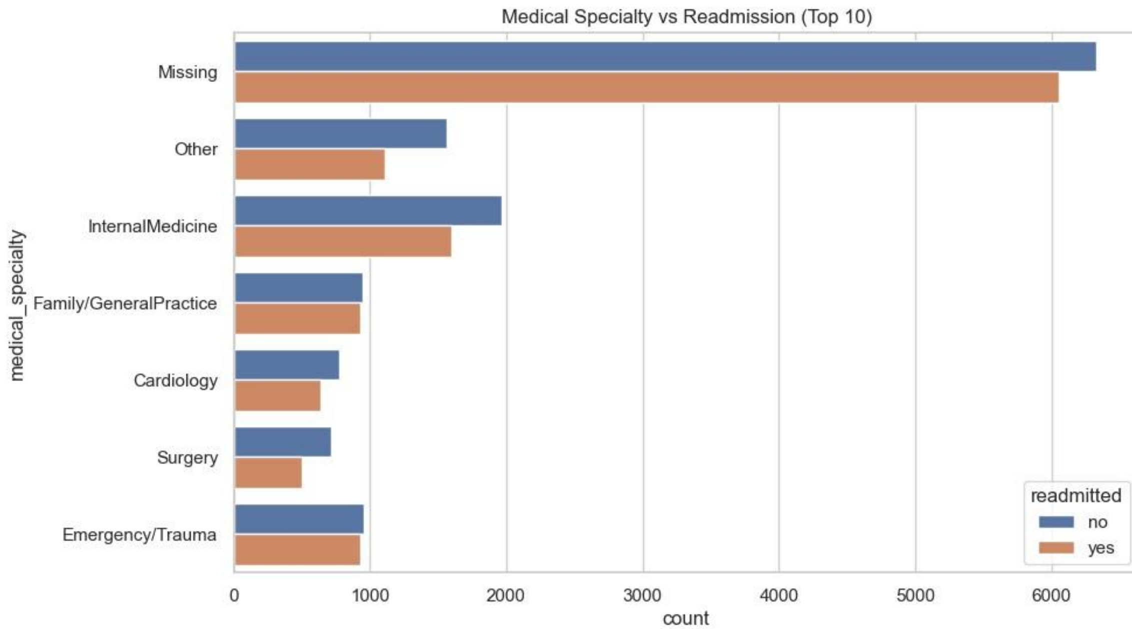
3.3 Medication and Procedures

- A higher number of medications is linked to increased readmission.
- Complex procedures and more hospital visits correlate with higher risk.



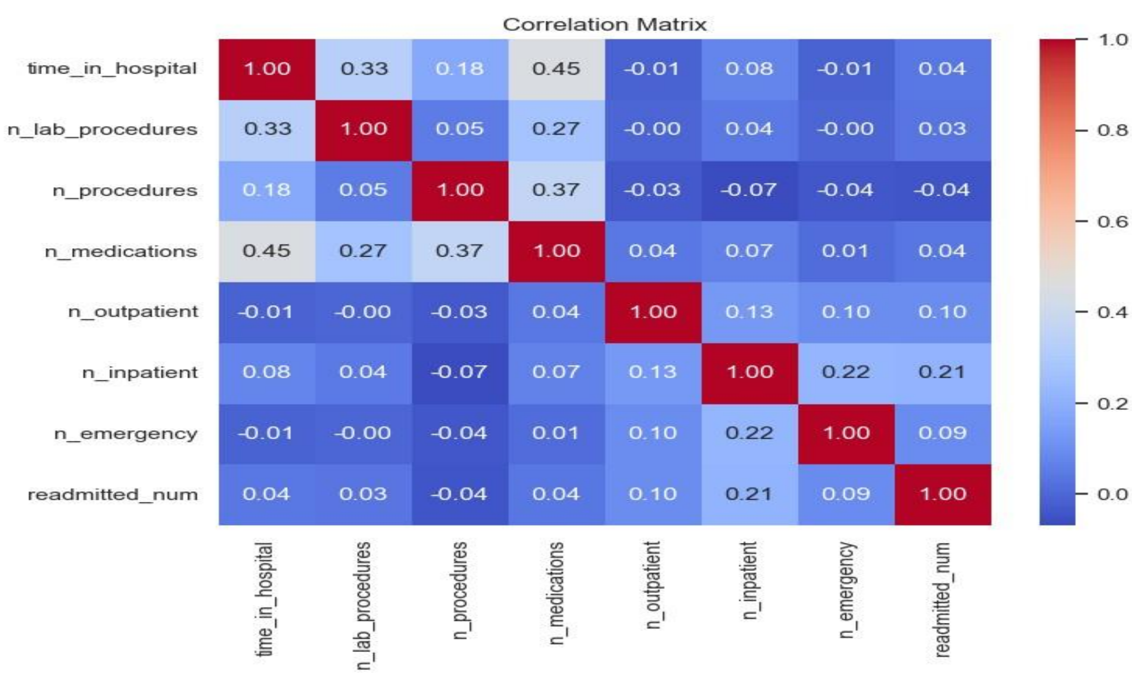
3.4 Medical Specialty

- Some medical specialties have a higher readmission ratio.
- Targeted interventions in high-risk departments could reduce readmissions.



3.5 Correlations

- Inpatient visits, emergency visits, and time in hospital are positively correlated with readmission.
- Correlation matrix highlights strong relationships among clinical variables.



4. Recommendations

- Implement targeted care management programs for high-risk age groups.
- Optimize medication plans to avoid over-prescription.
- Strengthen post-discharge follow-ups in specialties with high readmission.
- Monitor patient visits (inpatient/emergency) to predict readmission risk.
- Build dashboards for real-time hospital monitoring.

5. Future Analytics Opportunities

- Develop predictive machine learning models for readmission forecasting.
- Perform clustering of patients based on risk profiles.
- Integrate external data (lifestyle, comorbidities, socioeconomic factors).
- Build hospital management dashboards with real-time alerts.

6. Conclusion

The analysis highlights clear relationships between hospital stays, medications, age groups, and readmissions. Cleaned and structured data enabled reliable aggregation and visualization. Next steps include implementing predictive models and dashboards to assist hospitals in reducing readmissions and improving patient care.