

Healthcare Data Analytics for Improved Patient Care

Introduction

In today's healthcare landscape, data analysis plays a crucial role in understanding patient health patterns and optimizing treatment strategies. This report focuses on a series of comprehensive analyses derived from patient data, with the aim of identifying trends and providing insights into patient care, treatment effectiveness, and healthcare disparities. The analysis spans various aspects of patient health, such as diagnosis prevalence, medication usage, patient demographics, medical specialty performance, and lab procedures. Each section of this report addresses key healthcare questions that can guide clinical decision-making, resource allocation, and improved patient outcomes.

The primary goal of this report is to leverage data to highlight critical insights that can be used by healthcare providers to enhance patient care. These insights help prioritize high-risk patients, track the effectiveness of various treatments, and better understand the demographic factors that influence health outcomes. By using data-driven approaches, healthcare organizations can create more efficient, targeted interventions to improve the quality of care for all patients.

1. Analysis of Patients with Multiple Diagnoses

We want to identify patients who have multiple diagnoses across primary, secondary, and additional secondary categories. These patients might have more complex health conditions that require specialized care.

Why:

This analysis helps healthcare providers focus on patients with complex health profiles. By recognizing those with multiple diagnoses, providers can prioritize interventions, allocate resources efficiently, and offer more tailored care to patients at higher risk.

```

1 • SELECT patient_nbr,
2         COUNT(DISTINCT diag_1) AS primary_count,
3         COUNT(DISTINCT diag_2) AS secondary_count,
4         COUNT(DISTINCT diag_3) AS additional_secondary_count
5 FROM patient_data
6 GROUP BY patient_nbr
7 HAVING primary_count > 1 OR secondary_count > 1 OR additional_secondary_count > 1;
8

```

patient_nbr	primary_count	secondary_count	additional_secondary_count
135	2	2	2
1152	1	5	4
1314	3	2	3
5220	4	4	4
5337	1	2	2
6696	2	2	2

2. Ranking Primary Diagnoses by Frequency

We'll use a window function to rank primary diagnoses based on their frequency. This allows us to see which diagnoses are the most prevalent and track their distribution.

Why: This ranking will help us see which diagnoses have the highest number of cases and may help in setting priorities for healthcare management.

```

10 • SELECT
11     diag_1,
12     COUNT(*) AS count,
13     RANK() OVER (ORDER BY COUNT(*) DESC) AS rank_diag
14 FROM patient_data
15 GROUP BY diag_1;
16

```

diag_1	count	rank_diag
428	6862	1
414	6581	2
786	4016	3
410	3614	4
486	3508	5
427	2766	6

Key Insights:

1. Most Common Diagnoses:

- **ICD-9 Code 428** (Heart Failure) is the most frequent, followed by **ICD-9 Code 414** (Chronic Ischemic Heart Disease).
- **ICD-9 Code 786** (Symptoms related to circulatory/respiratory systems) and **ICD-9 Code 410** (Acute Myocardial Infarction) rank third and fourth.

2. Focus on Cardiovascular Issues:

- The top diagnoses highlight a strong prevalence of cardiovascular conditions, particularly heart failure and ischemic heart disease.

3. Identify High-Risk Patients

We want to create a stored procedure that flags patients who have multiple diagnoses, multiple medications, and long hospital stays. These patients are high-risk and may need additional care.

This procedure helps flag high-risk patients for further follow-up and intervention.

```
17 DELIMITER //
18 CREATE PROCEDURE identify_high_risk_patients()
19 BEGIN
20     SELECT patient_nbr, COUNT(DISTINCT diag_1) AS primary_count,
21           COUNT(DISTINCT diag_2) AS secondary_count,
22           COUNT(DISTINCT metformin) AS medication_count,
23           AVG(time_in_hospital) AS avg_hospital_stay
24     FROM patient_data
25     GROUP BY patient_nbr
26     HAVING primary_count > 1 AND secondary_count > 1 AND medication_count > 2
27     AND avg_hospital_stay > 5;
28 END;
29 //
30 DELIMITER ;
31 CALL identify_high_risk_patients();
```

patient_nbr	primary_count	secondary_count	medication_count	avg_hospital_stay
49860	4	4	3	6.2500
448875	5	3	4	6.0000
1812609	3	3	3	6.0000
3306798	4	4	3	5.2500
1406668	3	1	1	4.1411

Key Insights:

1. High-Risk Patients:

The output reveals patients with multiple diagnoses, multiple medications, and prolonged hospital stays, indicating they may be at higher risk for complications and require closer monitoring and care.

2. Treatment and Care Needs:

Patients who meet these criteria may benefit from more personalized treatment plans to manage both their medical conditions and hospitalizations effectively.

4. Distribution of Patient Gender

Let's examine the gender distribution of patients in the dataset, which could provide insights into gender-specific health issues.

Why: Gender distribution is important for understanding how different health conditions and treatments might affect male and female patients differently.

```
36 SELECT gender, COUNT(*) AS patient_count
37 FROM patient_data
38 GROUP BY gender;
```

gender	patient_count
Female	54708
Male	47055

Key Insights

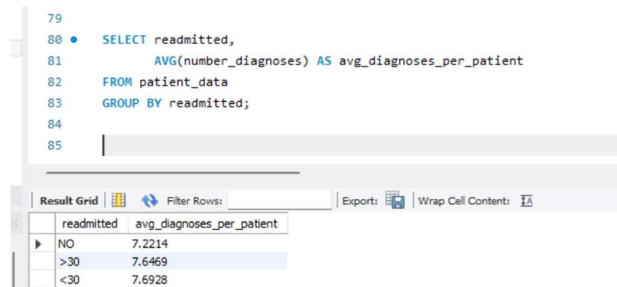
Higher Female Patient Count – The dataset shows a higher number of female patients (54,708) compared to male patients (47,055), indicating a potential trend in healthcare service utilization, disease prevalence, or demographic distribution.

Potential Healthcare Implications – The gender distribution might suggest differences in healthcare-seeking behavior, specific medical conditions, or accessibility factors that could be further analyzed for targeted healthcare strategies.

5. Average Diagnoses per Patient (Readmitted vs. Not Readmitted)

Do patients with more diagnoses have a higher chance of being readmitted within 30 days, after 30 days, or not at all?

Multiple diagnoses could indicate higher complexity, and we want to know if this leads to a higher likelihood of being readmitted sooner, later, or not at all.



```
79
80 • SELECT readmitted,
81           AVG(number_diagnoses) AS avg_diagnoses_per_patient
82 FROM patient_data
83 GROUP BY readmitted;
84
85
```


readmitted	avg_diagnoses_per_patient
NO	7.2214
>30	7.6469
<30	7.6928

Key Insights

Patients readmitted within 30 days have a slightly higher average number of diagnoses compared to those readmitted after 30 days or not readmitted at all. This suggests that patients with more complex conditions (i.e., more diagnoses) may require more frequent monitoring and are at a higher risk of readmission in the short term (within 30 days). It implies that comprehensive treatment plans, including closer follow-ups for such patients, might be necessary to reduce early readmissions.

6. Count of Different Medications per Patient

Count how many medications each patient is using, considering multiple medications may be prescribed.



```
125 • SELECT patient_nbr,
126           COUNT(DISTINCT CASE WHEN metformin = 'Steady' THEN 'metformin' END) +
127           COUNT(DISTINCT CASE WHEN insulin = 'Steady' THEN 'insulin' END) +
128           COUNT(DISTINCT CASE WHEN glimepiride = 'Steady' THEN 'glimepiride' END) AS medication_count
129 FROM patient_data
130 GROUP BY patient_nbr;
131
```

patient_nbr	medication_count
135	2
378	0
729	1
774	1
927	1

Key Insights

This helps track how many medications are prescribed for each patient, which could indicate the severity of the condition.

7.Percentage of Patients with Insulin

Calculate the percentage of patients who are using insulin. This shows how many patients are prescribed insulin, which can indicate the severity of diabetes in the population.

```
132 • SELECT
133     (COUNT(CASE WHEN insulin = 'Steady' THEN 1 END) * 100.0) / COUNT(*)
134     AS insulin_usage_percentage
135 FROM patient_data;
136
---
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

insulin_usage_percentage
30.31366

Key Insights

Approximately 30.3% of the patients in the dataset are on a steady insulin regimen. A **30% insulin usage rate** indicates that a significant portion of the patient population is being treated for diabetes or a related condition.

Given insulin is typically prescribed for patients with moderate to severe diabetes, this suggests that a considerable number of patients in your dataset may have advanced diabetes or other insulin-dependent conditions.

8.Medication Usage Based on Glucose Levels

Identify medication usage based on the glucose levels, with grouping for better insights.

```
137
138 • SELECT max_glu_serum,
139     COUNT(CASE WHEN metformin = 'Steady' THEN 1 END) AS metformin_usage,
140     COUNT(CASE WHEN insulin = 'Steady' THEN 1 END) AS insulin_usage
141 FROM patient_data
142 GROUP BY max_glu_serum
143 ORDER BY max_glu_serum DESC;
144
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

max_glu_serum	metformin_usage	insulin_usage
Norm	352	593
None	17628	29468
>300	157	298
>200	209	490

The query result shows the distribution of **Metformin** and **Insulin** usage across different **max_glu_serum (blood glucose levels)** categories:

- **Norm** (normal glucose): 352 metformin users, 593 insulin users.
- **None** (no abnormal glucose): 17,628 metformin users, 29,468 insulin users.
- **>300** (high glucose levels): 157 metformin users, 298 insulin users.
- **>200** (elevated glucose levels): 209 metformin users, 490 insulin users.

This suggests that insulin usage increases with higher blood glucose levels, while metformin usage appears more prevalent across the entire glucose spectrum.

9.Analysis of Average Procedures by Medical Specialty

This query identifies medical specialties that perform a high average number of procedures per patient. By focusing on specialties with over 50 patients and an average of more than 2.5 procedures, we gain insights into the specialties that are most procedure-intensive.

What we get: Specialties like **Surgery-Thoracic** and **Surgery-Cardiovascular/Thoracic** lead in average procedures, while **Cardiology** handles the largest patient volume with a slightly lower procedure rate.

```
145 • SELECT medical_specialty,
146         AVG(num_procedures) AS avg_procedures,
147         count(*) as count
148 FROM patient_data
149 GROUP BY medical_specialty
150 having count > 50 and avg_procedures > 2.5
151 ORDER BY avg_procedures DESC;
152
153
154
```

medical_specialty	avg_procedures	count
Surgery-Thoracic	3.5046	109
Surgery-Cardiovascular/Thoracic	3.2454	652
Radiologist	3.2421	1140
Cardiology	2.7044	5352
Surgery-Vascular	2.5704	533
Radiology	2.5283	53

The query provides the **average number of procedures** performed and the **number of patients** for each medical specialty, with the conditions that the specialty has more than 50 patients and an average of more than 2.5 procedures per patient.

Interpretation:

- **Surgery-Thoracic** has the highest average at 3.50 procedures per patient with 109 patients.
- **Surgery-Cardiovascular/Thoracic** follows with 3.25 procedures, but with a much larger patient base of 652.
- **Radiologist** and **Radiology** show similar averages around 2.53-3.24, with **Radiologist** handling the most patients (1140).
- **Cardiology** has an average of 2.70 procedures with a very high patient count (5352).
- **Surgery-Vascular** has an average of 2.57 procedures, with 533 patients.

In short, specialties like **Surgery-Thoracic** and **Surgery-Cardiovascular/Thoracic** are performing more procedures on average, but specialties like **Cardiology** have a significantly larger patient base.

10.Lab Procedures by Race

This query examines the number of lab procedures performed across different races. By analyzing the average number of lab procedures per patient for each race, we can identify which racial groups are associated with a higher frequency of lab tests, providing insight into healthcare patterns or potential disparities in treatment.

The analysis reveals which racial groups undergo more frequent lab procedures on average, helping us to understand healthcare practices and identify any potential disparities in access to or need for lab tests across different races.

154 • SELECT race,

155 AVG(num_lab_procedures) AS avg_lab_procedures,

156 COUNT(*) AS count

157 FROM patient_data

158 GROUP BY race

159 HAVING count > 50 AND avg_lab_procedures > 5

160 ORDER BY avg_lab_procedures DESC;

161

Result Grid

Filter Rows:

Export:

Wrap Cell Content:

	race	avg_lab_procedures	count
▶	?	44.1047	2273
	AfricanAmerican	44.0851	19210
	Other	43.4369	1506
	Caucasian	42.8329	76099
	Hispanic	42.7933	2037
	Asian	41.2122	641

Conclusion

This analysis reveals several valuable findings that are crucial for improving healthcare delivery and patient outcomes. Key insights include a high prevalence of cardiovascular issues such as heart failure and chronic ischemic heart disease, which should prompt healthcare providers to focus more on these conditions in treatment protocols. Furthermore, the identification of high-risk patients-those with multiple diagnoses, medications, and prolonged hospital stays-enables targeted interventions that can reduce complications and readmissions. The gender and race-based breakdowns show important demographic patterns, emphasizing the need for tailored healthcare strategies to address specific needs.

The examination of average diagnoses, lab procedures, and medical specialties provides a deeper understanding of resource allocation and how different medical fields engage with patients. By identifying specialties with the highest average number of procedures, healthcare facilities can better allocate resources and ensure efficient care delivery. The insights derived from medication usage and lab tests, particularly with regards to glucose levels and insulin use, highlight areas where preventive care can be improved.

In conclusion, this report showcases the power of data analysis in healthcare. The insights gathered not only reflect the current state of patient care but also point toward opportunities for improvement. By applying these findings, healthcare providers can refine their strategies, optimize resource allocation, and ultimately provide more effective and personalized care for their patients.