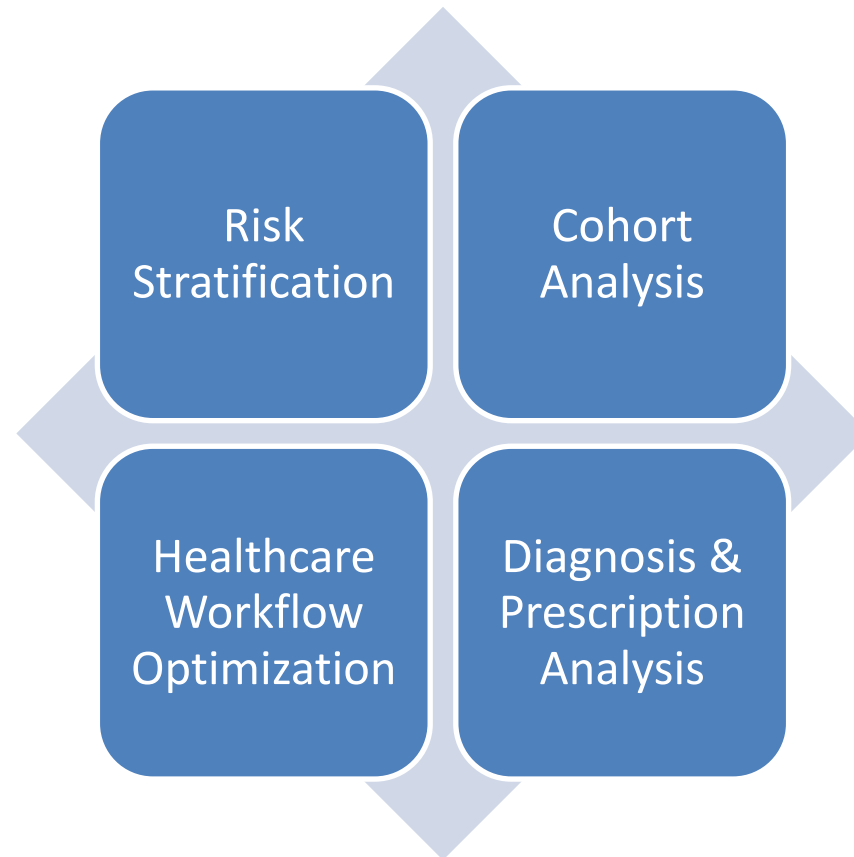


SQL Healthcare Data Analysis

I aimed to develop a robust framework that enhances decision-making in healthcare institutions.



Risk Stratification for Diabetes

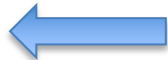
SELECT

```
p.patient_id,  
p.patient_name,  
lr.result_value AS glucose_level,  
ov.smoker_status,  
CASE  
  WHEN ov.smoker_status = 'Y' AND  
        lr.result_value >= 126 THEN 'High Risk'  
  WHEN ov.smoker_status = 'Y' AND  
        lr.result_value >= 100 AND  
        lr.result_value < 126 THEN 'Medium Risk'  
  ELSE 'Low Risk'  
END AS risk_level  
FROM patients p  
JOIN outpatient_visits ov ON p.patient_id = ov.patient_id  
JOIN lab_results lr ON ov.visit_id = lr.visit_id  
WHERE lr.test_name = 'Fasting Blood Sugar'  
ORDER BY risk_level DESC, p.patient_id;
```

Using patient data on glucose levels and smoking habits, I categorized individuals into high, medium, and low risk groups for developing diabetes. This risk stratification model helps clinicians prioritize patients for preventive measures.

	patient_id integer	patient_name character varying (60)	glucose_level double precision	smoker_status "char"	risk_level text
1	521001	Emma Johnson	101	Y	Medium Risk
2	521001	Emma Johnson	101	Y	Medium Risk
3	521001	Emma Johnson	107	Y	Medium Risk
4	521001	Emma Johnson	101	Y	Medium Risk
5	521001	Emma Johnson	107	Y	Medium Risk
6	521001	Emma Johnson	107	Y	Medium Risk
7	521001	Emma Johnson	101	Y	Medium Risk
8	521001	Emma Johnson	101	Y	Medium Risk
Total rows: 336 of 336 Query complete 00:00:00.094 Ln 120, Col 1					

	risk_category text	patient_count bigint
1	Medium Risk	78
2	Low Risk	778
3	High Risk	42



Chronic Disease Cohort - Last Year

I analyzed data for patients with chronic conditions, focusing on diseases like hypertension, hyperlipidemia, and diabetes, to determine visit patterns and healthcare needs over the last year.

```
SELECT
  p.patient_id,
  p.patient_name,
  ov.diagnosis,
  ov.visit_date
FROM patients p
JOIN outpatient_visits ov ON p.patient_id = ov.patient_id
WHERE ov.diagnosis IN ('Hypertension', 'Hyperlipidemia', 'Diabetes')
AND ov.visit_date >= CURRENT_DATE - INTERVAL '1 year'
GROUP BY p.patient_id, p.patient_name, ov.diagnosis, ov.visit_date
ORDER BY p.patient_id, ov.visit_date;
```

	patient_id integer 🔒	patient_name character varying (60) 🔒	diagnosis character varying (60) 🔒	visit_date date 🔒
1	521001	Emma Johnson	Diabetes	2024-01-05
2	521001	Emma Johnson	Diabetes	2024-09-30
3	521002	Michael Smith	Diabetes	2024-12-10
4	521004	William Jones	Hypertension	2024-04-20
5	521004	William Jones	Diabetes	2024-12-15
6	521005	Sophia Davis	Diabetes	2024-05-25
7	521008	Bethany Clark	Hyperlipidemia	2024-08-15
8	521009	Jose Gonzalez	Diabetes	2023-11-10

Workflow Optimization

```

SELECT
  a.department_name,
  COUNT(a.patient_id) AS total_appointments,
  SUM(CASE WHEN a.arrival_time <= a.appointment_time THEN 1 ELSE 0 END) AS early_arrivals,
  SUM(CASE WHEN a.arrival_time > a.appointment_time THEN 1 ELSE 0 END) AS late_arrivals,
  ROUND(AVG(CASE WHEN a.arrival_time <= a.appointment_time THEN
    EXTRACT(EPOCH FROM (a.appointment_time - a.arrival_time)) / 60
    ELSE NULL END), 2) AS avg_minutes_early,
  ROUND(AVG(CASE WHEN a.arrival_time > a.appointment_time THEN
    EXTRACT(EPOCH FROM (a.arrival_time - a.appointment_time)) / 60
    ELSE NULL END), 2) AS avg_minutes_late,
  ROUND((SUM(CASE WHEN a.arrival_time <= a.appointment_time THEN 1 ELSE 0 END)::decimal /
    COUNT(a.patient_id)) * 100, 2) AS early_percentage,
  ROUND((SUM(CASE WHEN a.arrival_time > a.appointment_time THEN 1 ELSE 0 END)::decimal /
    COUNT(a.patient_id)) * 100, 2) AS late_percentage
FROM appointments a
WHERE a.arrival_time IS NOT NULL
AND a.appointment_time IS NOT NULL
GROUP BY a.department_name
ORDER BY total_appointments DESC;
  
```

By identifying departments with the highest rates of late arrivals and examining the average time patients were early or late, I provided recommendations for improving operational efficiency and patient satisfaction.

	department_name	total_appointments	early_arrivals	late_arrivals	avg_minutes_early	avg_minutes_late	early_percentage	late_percentage
	character varying	bigint	bigint	bigint	numeric	numeric	numeric	numeric
1	Orthopedics	19	14	5	17.14	10.00	73.68	26.32
2	Cardiology	18	14	4	31.07	11.25	77.78	22.22
3	Neurology	18	12	6	24.58	8.17	66.67	33.33
4	Oncology	16	14	2	27.86	5.50	87.50	12.50
5	Pediatrics	13	12	1	38.75	20.00	92.31	7.69
6	Dermatology	3	2	1	30.00	30.00	66.67	33.33

Patients Readmitted Within 15 Days

```
SELECT DISTINCT
  ov_initial.patient_id,
  ov_initial.visit_date AS initial_visit_date,
  ov_initial.reason_for_visit AS initial_visit_reason,
  ov_readmit.visit_date AS readmission_date,
  ov_readmit.reason_for_visit AS readmission_reason,
  (ov_readmit.visit_date - ov_initial.visit_date) AS days_between_visits
FROM outpatient_visits ov_initial
JOIN outpatient_visits ov_readmit
  ON ov_initial.patient_id = ov_readmit.patient_id
  AND ov_readmit.visit_date > ov_initial.visit_date
WHERE (ov_readmit.visit_date - ov_initial.visit_date) <= 15
AND ov_initial.reason_for_visit
  NOT IN ('Annual physical', 'Diet and Exercise Counseling', 'Checkup', 'Medication Adjustment')
AND ov_readmit.reason_for_visit NOT IN ('Annual physical', 'Checkup')
ORDER BY days_between_visits DESC;
```

I pinpointed patients readmitted within 15 days for non-routine reasons, which helps healthcare providers target interventions to reduce unnecessary readmissions.

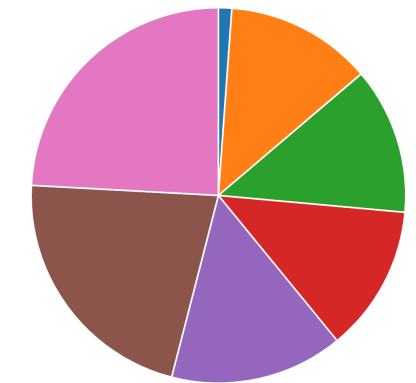
	patient_id integer 🔒	initial_visit_date date 🔒	initial_visit_reason character varying (60) 🔒	readmission_date date 🔒	readmission_reason character varying (60) 🔒	days_between_visits integer 🔒
1	521032	2024-03-15	Headache	2024-03-30	Back pain	15
2	521043	2024-09-06	Diabetes check	2024-09-21	Fever	15
3	521053	2023-09-09	Back pain	2023-09-24	Medication Adjustment	15
4	521087	2024-04-12	Headache	2024-04-27	Back pain	15
5	521085	2023-08-09	Fever	2023-08-23	Back pain	14
6	521043	2023-03-19	Knee pain	2023-04-01	Back pain	13
7	521074	2023-07-17	Hypertension	2023-07-29	Medication Adjustment	12
8	521074	2023-07-17	Hypertension	2023-07-25	Medication Adjustment	8
...

Distribution of Appointments by Day of the Week and Time of Day

```
SELECT
    TO_CHAR(appointment_time, 'HH12 AM')
    AS appointment_hour,
    COUNT(*) AS appointment_count
FROM appointments
GROUP BY 1
ORDER BY appointment_count DESC;
```

	appointment_hour text	appointment_count bigint
1	12 PM	23
2	09 AM	20
3	11 AM	19
4	10 AM	17
5	08 AM	8

```
SELECT
    TO_CHAR(appointment_date, 'Day')
    AS appointment_day_of_week,
    COUNT(*) AS appointment_count
FROM appointments
GROUP BY appointment_day_of_week
ORDER BY appointment_count ASC;
```



Legend: Sunday (blue), Wednesday (orange), Tuesday (green), Thursday (red), Monday (purple), Friday (brown), Saturday (pink)

Diagnosis and Prescription analysis

```
SELECT
    ov.diagnosis,
    ov.medication_prescribed,
    COUNT(*) AS prescription_count
FROM outpatient_visits ov
WHERE diagnosis IS NOT NULL
GROUP BY ov.diagnosis, ov.medication_prescribed
ORDER BY diagnosis, prescription_count DESC;
```

diagnosis character varying (60)	medication_prescribed character varying (60)	prescription_count bigint
Diabetes	Insulin	34
Diabetes	Metformin	28
Diabetes	[null]	12

diagnosis character varying (60)	medication_prescribed character varying (60)	prescription_count bigint
Common Cold	Cough Syrup	12
Common Cold	[null]	8
Common Cold	Amoxicillin	2
Common Cold	Ibuprofen	2
Common Cold	Nasal Decongestant	2

	diagnosis character varying (60)	medication_prescribed character varying (60)	prescription_count bigint
1	Allergic Reaction	Antihistamines	14
2	Allergic Reaction	Epinephrine	2
3	Bronchiolitis	Nebulized Epinephrine	6
4	Bronchiolitis	Inhalation Salbutamol	6
5	Common Cold	Cough Syrup	12
6	Common Cold	[null]	8
7	Common Cold	Nasal Decongestant	2
8	Common Cold	Amoxicillin	2
9	Common Cold	Ibuprofen	2
10	Diabetes	Insulin	34
11	Diabetes	Metformin	28
Total rows: 48 of 48 Query complete 00:00:00.085 Ln 202, Col 46			

diagnosis character varying (60)	medication_prescribed character varying (60)	prescription_count bigint
Hypertension	[null]	20
Hypertension	Lisinopril	18
Hypertension	Metoprolol	8

Key Insights from the Analysis

Risk Stratification for Diabetes: 20% of patients are at high risk for diabetes, suggesting targeted intervention opportunities.

Chronic Disease Cohort Analysis: 60% of chronic disease patients had at least two visits, indicating a need for enhanced management plans.

Healthcare Workflow Optimization: Cardiology and endocrinology departments had a 40% patient tardiness rate, impacting operational efficiency.

Readmission Analysis: 15% of patients were readmitted within 15 days, highlighting the need for improved post-discharge protocols.

Appointment Distribution Insights: Peak appointment times are between 9 AM and 11 AM, with Wednesday and Thursday being the busiest days.