Project Proposal

Predictive Disease Risk Assessment

NAME	ROLE	LINKEDIN
Ziad Hesham Mohamed	Data Modelling (Team Leader)	PROFILE
Albaraa Ehab Elzeftawy	Data Analyst	PROFILE
Esraa Hamada	Viz Wizard	PROFILE
Mohamed Sherif Mohamed	Data Curator	PROFILE

1. Project Title

Predictive Disease Risk Assessment: Healthcare Resource Utilization and Trend Forecasting.

2. Objective

To analyze historical public health data to identify patterns in disease incidence, forecast future outbreaks, and model healthcare resource utilization, such as hospital admissions and vaccine needs. The goal is to provide actionable insights for proactive public health interventions and efficient resource allocation.

3. Dataset Description

The project utilizes a synthetic/public dataset containing:

Time Series Data: Weekly/Monthly records of disease case counts by region.

Demographic Data: Population statistics by region, age group, and vaccination rates.

Resource Data: Hospital admissions, ICU bed occupancy, and vaccine inventory levels.

Metadata: Region names, dates, and disease codes.

4. Methodology

Week 1: Build Data Model, Data Cleaning, and Preprocessing

Data Modeling: Design a star schema for analytics focusing on health events, dates, regions, and diseases.

Data Preprocessing: Handle missing values, standardize region names, and create calculated fields such as incidence rate per 100k population.

Tools: SQL for modeling and Python for cleaning.

Deliverables: Cleaned dataset in analysis-ready format and a Jupyter notebook documenting logic.

Week 2: Analysis Questions Phase

Core Questions: Determine historical trends, seasonality, and how incidence rates correlate with demographic factors like population density.

Correlation: Analyze the relationship between vaccination rates and subsequent disease case numbers.

Resource Impact: Evaluate how hospital resource utilization varies with disease outbreaks.

Deliverables: A documented list of 5–7 prioritized analysis questions and initial visualizations.

Week 3: Forecasting Questions Phase

Forecasting: Predict case counts for the next 3–6 months for specific diseases using SARIMA or Prophet.

Predictive Modeling: Identify high-risk periods for hospital overload based on early signals.

Scenario Analysis: Model the impact of a 10% increase in vaccination coverage on future case loads.

Deliverables: Forecasting models with performance metrics (MAE, RMSE) and prediction interval plots.

Week 4: Visualization Dashboard and Final Presentation

Dashboard Build: Create an interactive Tableau dashboard including an executive summary and national hotspot alerts.

Deep Dives: Include interactive filters for demographics and 6-month forecasts with confidence intervals.

Resource Analysis: Link case forecasts to simulated resource needs.

Deliverables: Published Tableau dashboard with a user guide and a final presentation deck.

5. **Expected Outcomes**

   Faster, more targeted public health responses leading to better outbreak containment.

A demonstrable end-to-end data analysis project showcasing skills in data engineering, analytics, forecasting, and visualization.