Conclusion (in simple terms)
After performing data cleaning and feature engineering, the dataset is now well-prepared for further analysis or model building.
I handled missing values, normalized important numerical columns like Age, BMI, and HbA1c_Level, and created new features like Lifestyle_Index and Is_At_Risk. These additional features help the model better understand patient risk profiles and improve its ability to predict health outcomes like diabetes.

Documentation: Data Assumptions and Challenges

Category	Details
Missing Values	- Assumed that missing values in Smoking_History mean the information was not provided; filled them with "No Info".
Normalization Logic	- Applied Min-Max Normalization to scale numeric columns (Age, BMI, HbA1c_Level, Blood_Glucose_Level) between 0 and 1.
Categorical Mapping	- For new features like Lifestyle_Index, assigned 1 point to each risky behavior (e.g., smoking, high BMI, hypertension).
Risk Definition	- In Is_At_Risk, considered patients at risk if they had Hypertension, Heart Disease, or high Blood Glucose Level.
Data Challenges	- Dataset had inconsistent values (like decimals in age) - Some rows had missing or ambiguous entries in categorical fields
Smoking Category Logic	- Treated current and ever smokers as risky for the Lifestyle Index; never, not current, and No Info scored 0.