**Phase 4: Visualization & Pipeline Deployment in ADF**

**Visualization:**

In this phase, the enriched data from the Gold layer is transformed into interactive dashboards to uncover meaningful insights. The goal is to support data-driven decision-making by visualizing health risk patterns across different demographic segments.

Two key dashboards were developed:

* **Risk Profile Explorer**: Analyzes diabetes risk across age groups, genders, and risk levels.
* **Lifestyle Impact Dashboard**: Highlights modifiable behaviors (obesity, inactivity, alcohol use) across demographic segments.

These visualizations serve as actionable tools for healthcare strategists, analysts, and public health teams to prioritize interventions and identify high-risk populations.

1. **Loading the data and Creating a View on top of it**

A screenshot of a computer

AI-generated content may be incorrect.

The data was loaded from the Databricks mount point corresponding to the Gold layer and stored into a Spark DataFrame named df\_gold. This DataFrame contains the final, aggregated, and enriched dataset ready for analysis and reporting. To enable SQL-based querying, a temporary view named **gold\_vw** was created from this DataFrame using the **createOrReplaceTempView**() method.

**2. Creating dashboard -1: Risk Profile Explorer**For this Dashboard the following Visualizations are used

**Visualization -1:**

Query Used:

A screenshot of a computer

AI-generated content may be incorrect.

A colorful circle with numbers and a white background

AI-generated content may be incorrect.  
The donut chart shows the percentage distribution of total patients across four categories, with the red group being the largest at 36.28%.

**Visualization -2:**

Query Used:

A screenshot of a computer

AI-generated content may be incorrect.

**Visualization -3:**

Query Used:

A screenshot of a computer

AI-generated content may be incorrect.

**Visualization -4:**

Query Used:

A screen shot of a graph

AI-generated content may be incorrect.

The scatter plot shows a **positive correlation** between Diabetes Rate and High Risk Rate for both males and females — as diabetes rates increase, high risk rates also tend to rise.

Visualization -5:

Query Used:



A graph of a bar chart

AI-generated content may be incorrect.

The bar chart shows that **Diabetes Rate increases with age for both males and females**, with males (orange) having slightly higher rates than females (blue), especially in the 65+ age group.

**Complete Dashboard-1**

**A screenshot of a medical report

AI-generated content may be incorrect.**

A screenshot of a medical report

AI-generated content may be incorrect.

1. **Creating Dashboard-2: Lifestyle Impact by Demographics**

Visualization -1:

Query Used:

A screenshot of a computer

AI-generated content may be incorrect.

A graph on a computer screen

AI-generated content may be incorrect.

The bar chart shows that **obesity rates are highest among patients in the High Risk group across all age groups**, especially between **18–49**, while the Low and Moderate risk groups show significantly lower obesity rates.

Visualization-2:

A screenshot of a graph

AI-generated content may be incorrect.

The bar chart shows that **low physical activity is highest in the Low and Moderate risk groups across all age groups**, while the **High Risk group consistently shows lower activity rates**, especially in older age groups.

Visualization-3:

A graph of different colored bars

AI-generated content may be incorrect.

The bar chart shows that **alcohol-related risk is highest in the Low and Moderate risk groups**, especially in the **50–64 age group**, while the **High Risk group consistently shows the lowest alcohol risk rates** across all age groups.

Complete Dashboard- 2

A screenshot of a graph

AI-generated content may be incorrect.

**Data Orchestration using Azure Data factory**

1. **Creating a Notebook Activity in our ADF (diabetes\_etl\_adf)**

A screenshot of a computer

AI-generated content may be incorrect.

1. Linked service called **ls\_diabets\_etl\_db1** is created
2. To access the Databricks workspace(**diabets\_etl\_db1)** heading to Databricks workspace> AccessControl

A screenshot of a computer

AI-generated content may be incorrect.

1. Giving Contributor access to ADF(diabetes-etl-adf) through Managed Identity

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

1. Verifying the access

A screenshot of a computer

AI-generated content may be incorrect.

After Linking azure data bricks with ADF we specify the location of our notebook called **bronze\_silver** to our **Notebook Activity**

A screenshot of a computer

AI-generated content may be incorrect.

Similarly attaching other notebooks(silver\_to\_gold, gold\_to\_dashboard)

1. **Creating Master Pipeline**

Making a pipeline with these Notebook activities and copy activity which copies data from GitHub to ADLS gen2

A screenshot of a computer

AI-generated content may be incorrect.

Pipeline is Debugged successfully.

A screenshot of a computer

AI-generated content may be incorrect.