**Vaccination Data Analysis and Visualization**

**Project Title:** Vaccination Data Analysis and Visualization  
**Domain:** AIML  
**Date:** 07-09-2025  
**Submitted by:** GUNTUR RIDHI

**1. Project Overview**

The project focuses on analyzing global vaccination data to understand trends in vaccination coverage, disease incidence, and effectiveness. The data was cleaned, stored in a normalized SQL database, and visualized using Power BI to provide actionable insights for public health strategies.

**2. Project Approach**

**2.1 Data Cleaning**

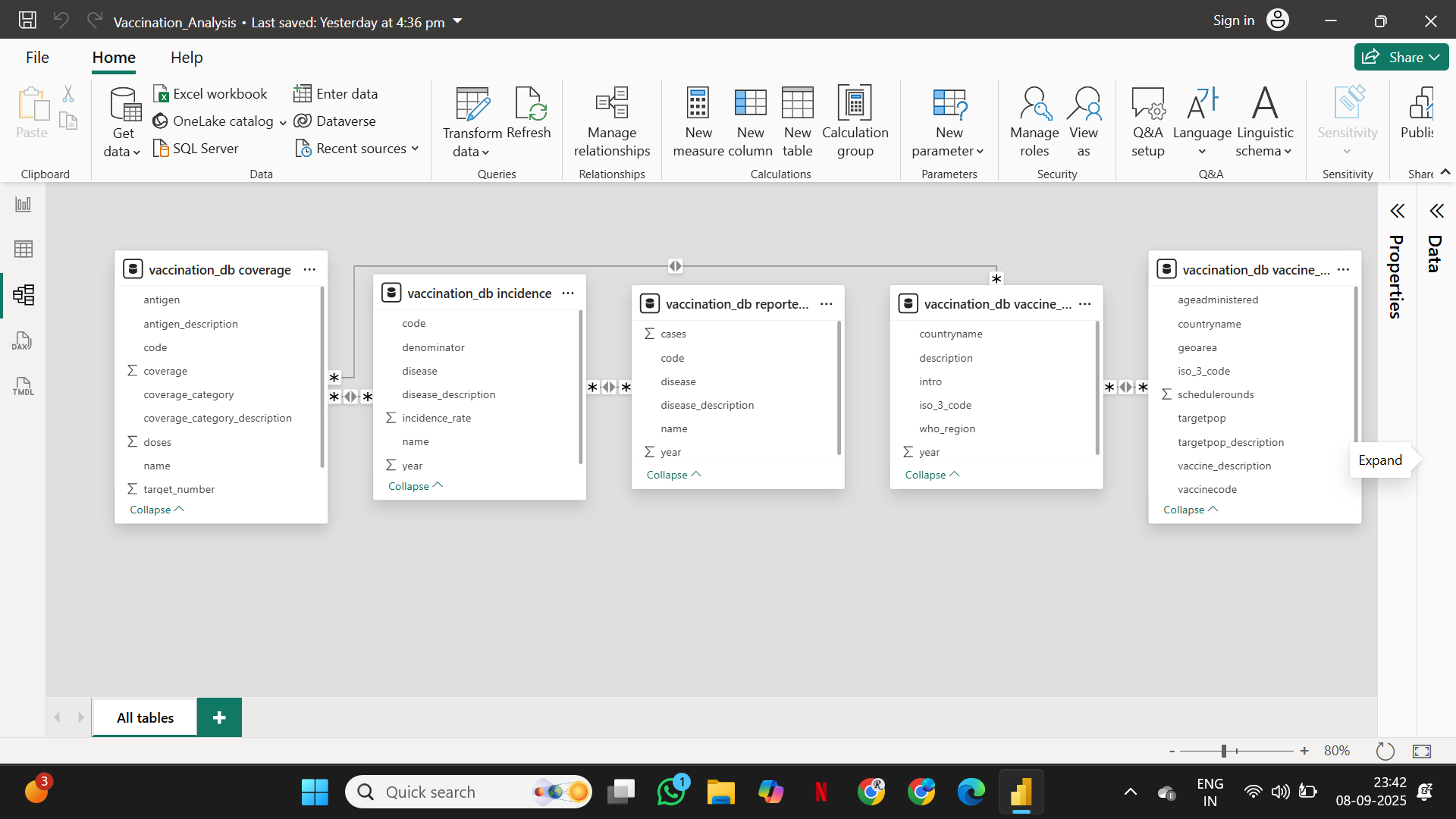
* Handle Missing Data: Imputed missing values or removed incomplete records.
* Normalize Units: Ensured consistency in coverage percentages, doses, and target population.
* Date Consistency: Uniform formatting of year fields across tables.

**2.2 SQL Database Setup**

* Created tables: coverage, incidence, reported\_cases, vaccine\_introduction, vaccine\_schedule.
* Normalization: Structured data to avoid redundancy; separate tables for vaccines, diseases, countries, and years.
* Data Integrity: Added primary and foreign keys where applicable.

**2.3 Power BI Integration**

* Connected Power BI to SQL database.
* Imported relevant tables and created relationships.
* Set up filters and slicers for interactivity.



**2.4 Data Visualization**

* Created interactive dashboards for: coverage by antigen, disease incidence trends, vaccine introduction timelines, and schedules.
* Visualizations include bar charts, line charts, scatter plots, geographical maps, and KPI indicators.

**2.5 Exploratory Data Analysis (EDA)**

* Analyzed vaccination coverage, disease incidence trends, regional disparities.
* Visualized patterns to identify low-coverage areas and assess vaccination impact.

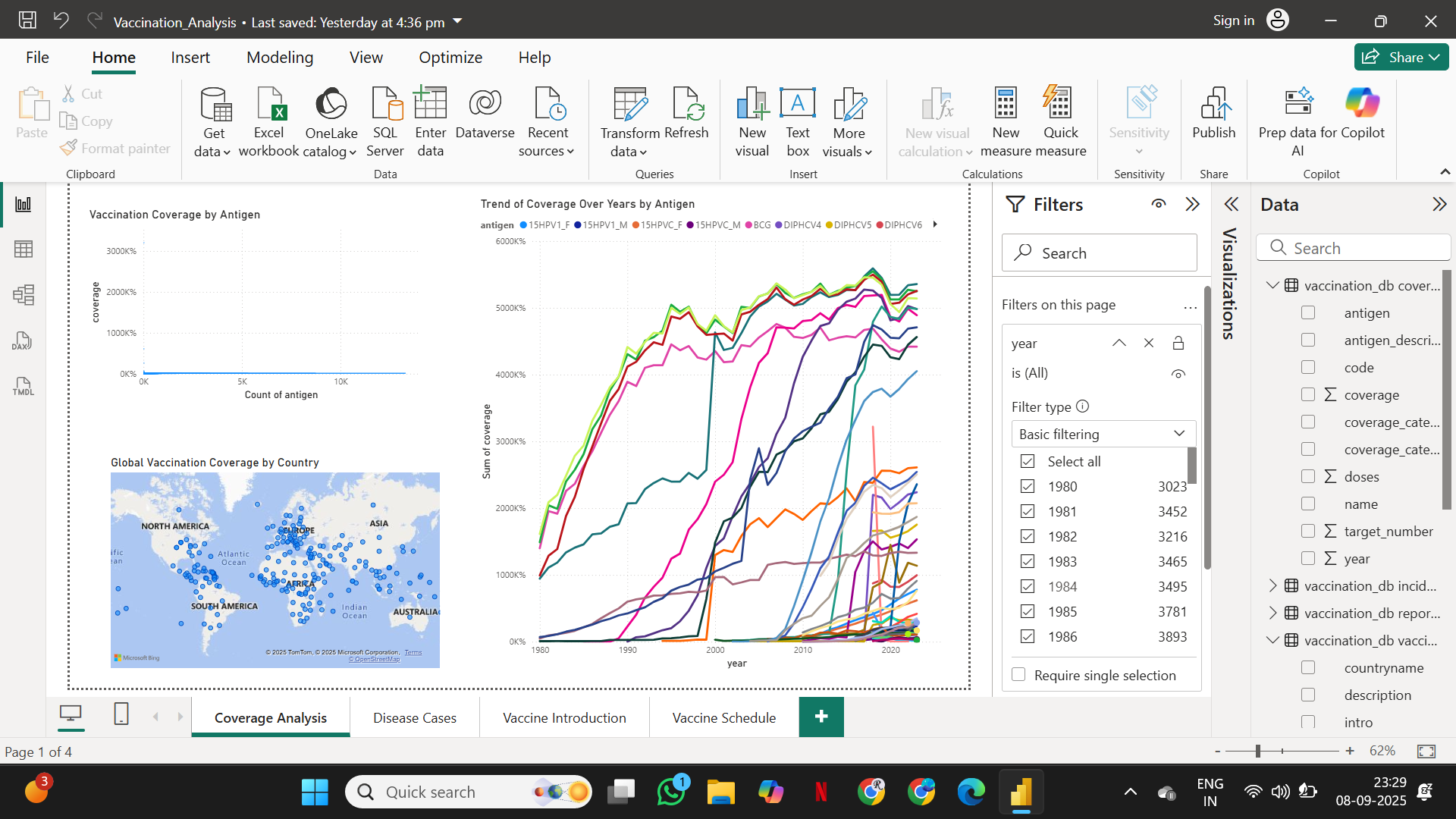
**3. Database Schema**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table | Columns | Data Type | Primary Key | Foreign Key |
| coverage | code, name, year, antigen, antigen\_description, coverage\_category, coverage\_category\_description, target\_number, doses, coverage | VARCHAR, INT, FLOAT | code, year, antigen | code → vaccine\_schedule.iso\_3\_code, year → vaccine\_schedule.year, antigen → vaccine\_schedule.vaccinecode |
| incidence | code, name, year, disease, disease\_description, denominator, incidence\_rate | VARCHAR, INT, FLOAT | code, year, disease | code → coverage.code |
| reported\_cases | code, name, year, disease, disease\_description, cases | VARCHAR, INT, FLOAT | code, year, disease | code → coverage.code |
| vaccine\_introduction | iso\_3\_code, countryname, who\_region, year, description, intro | VARCHAR, INT | iso\_3\_code, year, description | iso\_3\_code → coverage.code |
| vaccine\_schedule | iso\_3\_code, countryname, who\_region, year, vaccinecode, vaccine\_description, schedulerounds, targetpop, targetpop\_description, geoarea, ageadministered | VARCHAR, INT, FLOAT | iso\_3\_code, year, vaccinecode, schedulerounds | iso\_3\_code → coverage.code |

**4. Power BI Dashboard Overview**

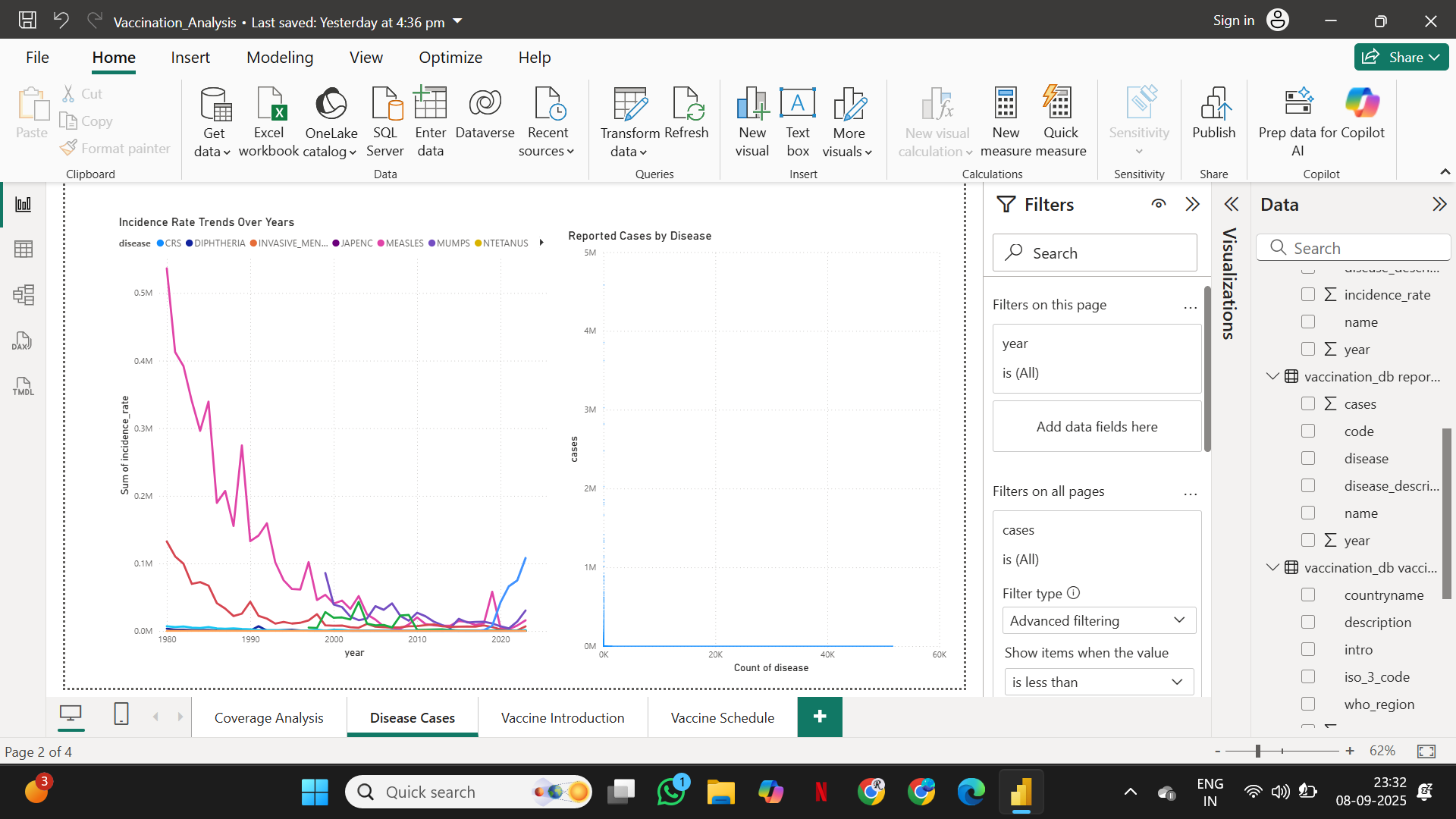
**Page 1: Coverage Analysis**

* Clustered Bar Chart: Coverage by Antigen
* Line Chart: Coverage Over Years
* Map: Coverage by Country



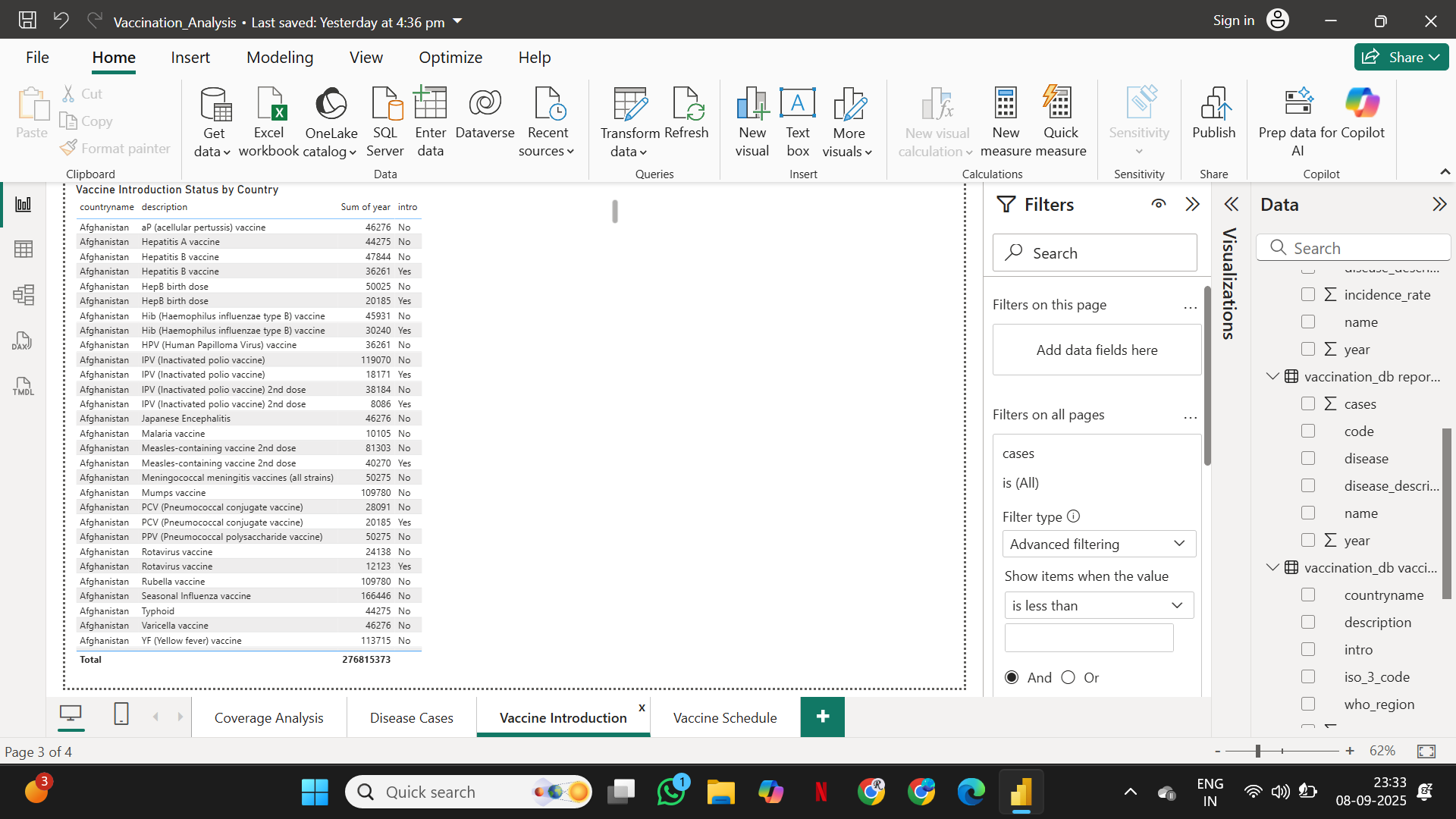
**Page 2: Disease Incidence Analysis**

* Line Chart: Incidence Trends Over Years
* Bar Chart: Disease Cases by Country



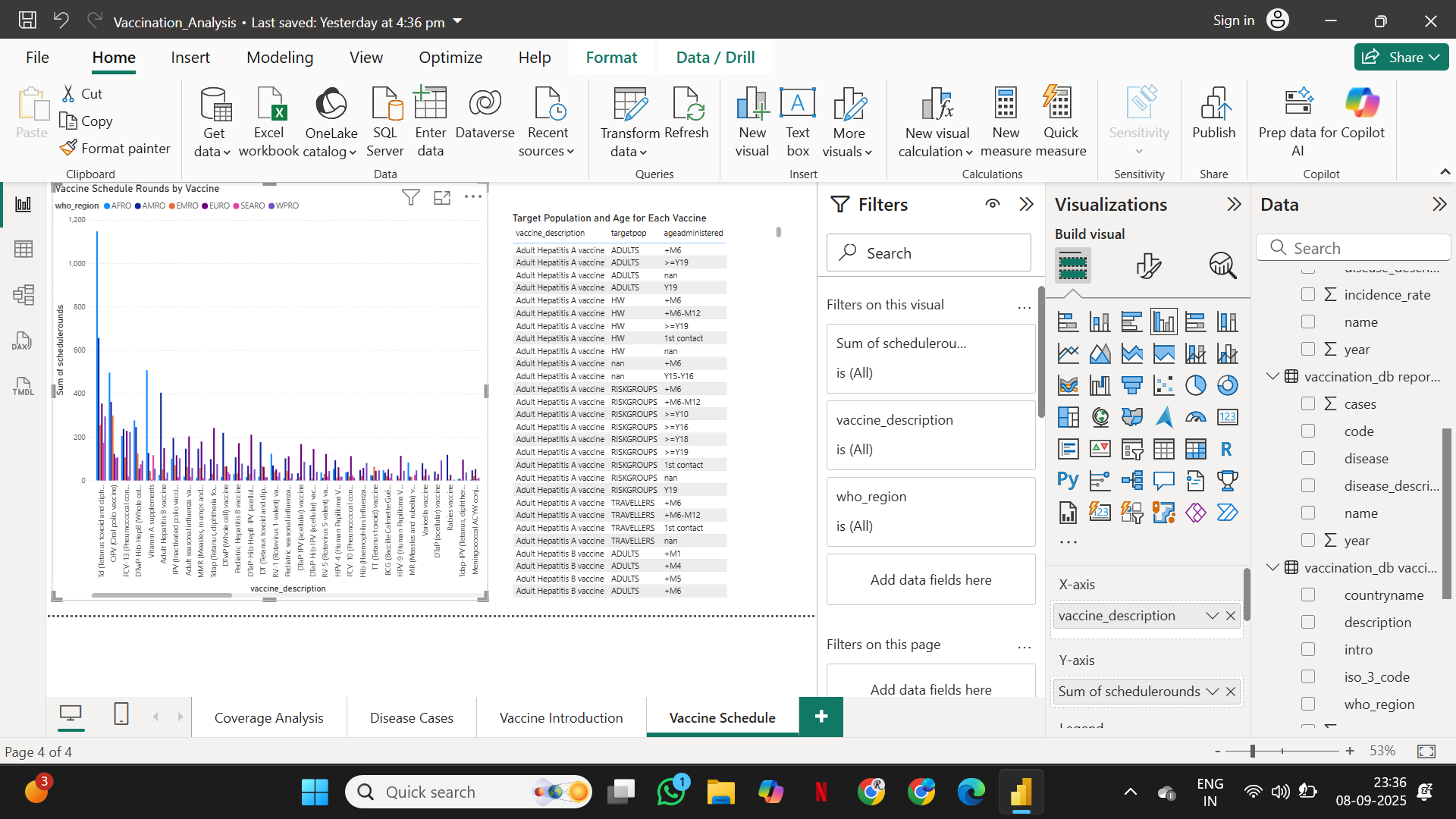
**Page 3: Vaccine Introduction**

* Table Chart: Vaccine Introduction by Country



**Page 4: Vaccine Schedule Analysis**

* Clustered Column Chart: Target Population Coverage per Vaccine
* Table Chart: Target population and age for each vaccine



**6. Challenges and Solutions**

|  |  |
| --- | --- |
| **Challenge** | **Solution** |
| Duplicate entries preventing primary key creation | Removed duplicates before adding primary keys |
| Foreign key constraints failing due to missing reference data | Cleaned coverage data to match vaccine\_schedule references |
| Power BI map warnings | Ensured country names matched SQL table data; enabled map visuals |
| Handling large CSV files | Loaded in chunks and verified integrity |

**7. Project Deliverables**

* **Source Code:** Python scripts for data cleaning and SQL loading.
* **SQL Database:** Normalized tables with primary and foreign keys.
* **Power BI Reports:** Interactive dashboards.
* **Documentation:** This document detailing methodology, challenges, and insights.

**8. Project Evaluation Metrics**

* Data Cleaning: Completeness, consistency, missing value handling.
* SQL Database Quality: Normalization, relationships, primary/foreign keys.
* Power BI Visualizations: Clarity, interactivity, relevance.
* Insights & Actionability: Accuracy of trends, regional disparities, and actionable recommendations.

**9. Conclusion**

The project successfully analyzed global vaccination data, created a structured SQL database, and visualized insights using Power BI. The dashboards provide valuable information for public health planning, vaccine allocation, and monitoring disease trends.

**10. References**

* WHO Vaccination Data
* Python, SQL, Power BI Documentation

*End of Document*