**INTRODUCTION TO DATA MANAGEMENT**

**PROJECT REPORT**

(Project Semester January-April 2025)

***Health Analysis Dashboard***

Submitted by

**Mayank Madhukar**

Registration No: 12307829

B.Tech CSE, Section: K23 SK

Course Code: INT217

Under the Guidance of

**Ms. Aashima Bansal (UID: 28968)**

**Assistant Professor, Data Science**

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

**DECLARATION**

### I, Mayank Madhukar, student of B.Tech Computer Science and Engineering, under the CSE/IT Discipline at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

### Date: Signature

### Registration No: 12307829 Mayank Madhukar

**CERTIFICATE**

### This is to certify that Mayank Madhukar, bearing Registration No: 12307829, has completed INT217 project titled, “Health Analysis Dashboard” under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

### Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### Ms. Aashima Bansal

### Assistant Professor – Data Science

### School of Computer Science and Engineering

### Lovely Professional University

### Phagwara, Punjab

### Date: \_\_\_\_\_\_\_\_\_\_\_

**ACKNOWLEDGEMENT**

I would like to express my heartfelt gratitude to **Ms. Aashima Bansal**, Assistant Professor, for her continuous support, valuable feedback, and expert guidance throughout the duration of this project. Her mentorship played a crucial role in shaping the analytical approach and structure of this work.

Lastly, I am thankful to **Lovely Professional University** for providing a platform that fosters practical learning, and to my peers for their motivation and constructive discussions.

TABLE OF CONTENT

- Cover Page ...................................................................................................................... 1

- Declaration ...................................................................................................................... 2

- Certificate ........................................................................................................................ 3

- Acknowledgement ........................................................................................................... 4

1. [Introduction ....................................................................................................................](#_1._Introduction) 6

2. [Source of Dataset ...........................................................................................................](#_2._Source_of) 7

3. [Dataset Preprocessing.....................................................................................................](#_3._Dataset_Preprocessing) 8

4. [Analysis on Dataset .......................................................................................................](#_4._Analysis_on) 10

4.1 [Track Long-Term Health Performance ………………………...........................](#_4.1_Track_Long-Term) 10

4.2 [Identify Top problems........................................................................](#_4.2_Identify_Top) 13

4.3 [Analyze Health Volume Patterns............................................................................](#_4.3_Analyze_Trading) 15

4.4 [Evaluate yearly cases…........................................](#_4.4_Evaluate_Monthly) 17

4.5 [Highlight](#_4.5_Highlight_Volatility)

4.6 [Dashboard Snapshots................................................................................................](#_4.6_Dashboard_Snapshots) 22

5. [Conclusion .....................................................................................................................](#_5._Conclusion) 23

6. [Future Scope ..................................................................................................................](#_6._Future_Scope) 24

7. [References .....................................................................................................................](#_7._References) 26

# **1. Introduction**

The **Health Analysis Dashboard** is a comprehensive and interactive Excel-based dashboard project that provides deep analytical insights into the health conditions of a population across a decade, covering the period from **1997 to 2010**. This project revolves around exploring public health trends and indicators such as disease prevalence, age-wise and gender-wise distribution, and seasonal patterns in health conditions.

This project was undertaken to fulfil the requirements of the **INT217 INTRODUCTION TO DATA MANAGEMENT** course, where the focus was on leveraging **Pivot Tables, Pivot Charts, Slicers, and other Excel features** to draw meaningful conclusions from large datasets. The dashboard examines health trends, disparities, and anomalies using intuitive visualizations and interactive components.

The suite is organized into two core dashboards:

* **Health Condition Overview (1997–2010):**  
  Provides a deep dive into the distribution, frequency, and severity of various health conditions, categorized by demographic factors such as age groups and gender over the decade.
* **Monthly Trends & Seasonality (1997-2010):**  
  Focuses on identifying recurring health patterns, seasonal outbreaks, and fluctuations in health indicators across months and years.

The project not only enhances understanding of public health movements but also demonstrates how Excel can be effectively utilized for **data-driven decision-making in healthcare analytics**. Each dashboard is connected through a **user-friendly Home Interface**, which acts as a navigation panel for ease of use.

With an emphasis on **interactivity, data pre-processing, and visual storytelling**, the **Health Analysis Dashboard** aims to empower users to explore health trends in a structured and insightful manner, making it suitable for both **academic research and public health planning**.

# **2. Source of Dataset**

The data utilized for the **Health Analysis Dashboard** project has been sourced from a publicly available repository on **Data.gov**, which hosts detailed health-related statistics and demographic information for children under age 18 in the United States. This dataset was selected for its reliability, completeness, and relevance to longitudinal health analysis.

**Data.gov** is a widely trusted platform for open government data sharing and serves as an authoritative source for public health statistics. The dataset provided a structured and clean format, making it highly suitable for trend analysis, demographic segmentation, and seasonality exploration using Excel.

The dataset includes the following key fields:

* **Year** – The survey year
* **Health Conditions** – Various physical and mental health conditions reported
* **Age Group** – Age categorization of the children
* **Sex** – Gender of the child (Male/Female)
* **Race/Ethnicity** – Demographic classifications
* **Prevalence (%)** – Reported percentage of children affected by each health condition

The dataset spans a period from**1997 to 2010**, offering sufficient breadth and depth for both **trend analysis** and **comparative demographic insights**.

**Dataset Highlights:**

* **Time Period:** 1997 to 2010 (13 years)
* **Population:** Children under age 18 in the United States
* **Data Categories:** Health conditions, age group, sex, race/ethnicity
* **Format:** CSV (easily importable into Excel)
* **Source URL:** <https://catalog.data.gov/dataset/health-conditions-among-children-under-age-18-by-selected-characteristics-united-states-53b56>

The credibility of the dataset is assured through its publication by U.S. government health authorities. Any minor data inconsistencies (such as non-responses or aggregated reporting) were treated as typical in health survey data and were addressed during the data cleaning and transformation process in Excel.

By selecting a well-documented and publicly maintained government dataset, the project ensures a foundation of **accuracy and transparency**, allowing for more focus on **insight generation, visualization, and storytelling** through interactive dashboards.

# **3. Dataset Preprocessing**

To ensure consistency, reliability, and usability of the **health dataset spanning over a decade (1997-2010)**, a structured and meticulous preprocessing workflow was followed. The steps below outline the detailed process undertaken to prepare the dataset for interactive analysis in Excel:

**3.1 Consolidating Data for Unified Analysis**

The raw dataset included multiple entries based on various **health conditions, demographic characteristics (age, sex, race/ethnicity), and years**. To facilitate comprehensive and comparative analysis:

* The dataset was imported from **CSV format** into Excel using **Pivot Table**.
* A unified **master table** was created to combine all records into a single, structured dataset that supports both high-level summaries and detailed breakdowns.

**3.2 Standardizing Demographic Dimensions**

To enable demographic-based filtering and segmentation:

* Fields such as **Age Group**, **Sex**, and **Race/Ethnicity** were reviewed for consistency.
* Text inconsistencies and abbreviations were resolved using Excel’s **Find & Replace** and **Data Cleaning tools** to ensure uniform labels across records.
* Drop-down slicers were later applied to these columns for dynamic dashboard interaction.

**3.3 Parsing and Extracting Time Dimensions**

To support time-series and seasonal trend analysis:

* **Year** was extracted directly from the dataset to facilitate **year-on-year comparisons**.
* **Month** and **Quarter** fields were derived using formulas:
  + =TEXT([@Date], "mmm") for **Month**
  + =CHOOSE(MATCH(MONTH([@Date]),{1,4,7,10},1),"Q1","Q2","Q3","Q4") for **Quarter**
* These fields enabled structured seasonal trend visualizations and analysis of recurring health patterns.

To deepen insights, calculated fields were introduced:

* **Condition Prevalence Rate (%):** Calculated as the percentage of affected children in a given demographic category.
* **Year-over-Year Change (%):** Evaluates the annual trend for each health condition.
* **Age-wise Comparison Index:** A derived metric comparing prevalence across age groups to identify which age bracket is most affected.

# **4. Analysis on dataset**

**i. General Description**

This objective focuses on analyzing how various **health conditions among children under age 18** in the United States have evolved over the long term, specifically from **1997 to 2010**. By evaluating the average annual prevalence rates of selected health conditions, users can observe shifts in public health, detect emerging health concerns, and identify consistent trends over time. This analysis helps stakeholders understand how childhood health has responded to changes in healthcare, policy, and environment.

**ii. Specific Requirements**

To fulfill this objective, the dataset was grouped by **Health Condition** and **Year**. The key metric—**average annual prevalence rate (%)**—was calculated using a PivotTable:

* **Rows**: Health Condition
* **Columns**: Year
* **Values**: Average of Prevalence (%)

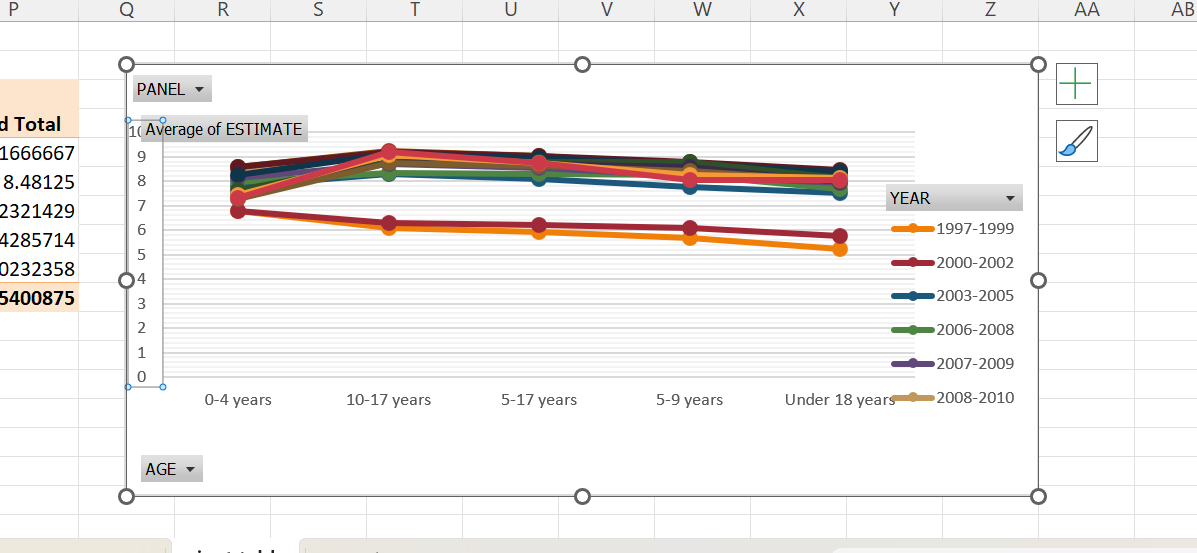
To enhance interactivity, **slicers** were added for **Year**, **Health Condition**, and **Demographic Characteristics** (e.g., Sex, Age Group, Race/Ethnicity), enabling users to filter and explore trends for specific groups or health outcomes. All percentage values were formatted to display up to **two decimal places** for readability.

**iii. Analysis Results**

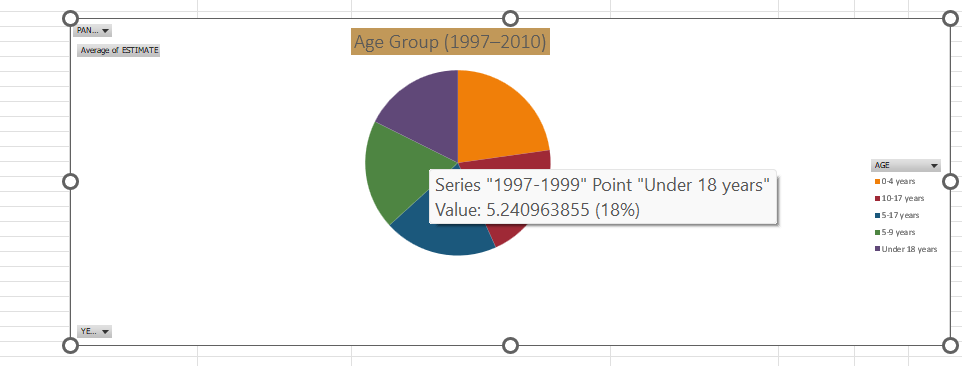
* Conditions such as **Asthma**, **Allergies**, and **ADHD** showed relatively **high and consistent prevalence** over the years, indicating ongoing public health challenges.
* **Behavioral and developmental disorders** exhibited a **rising trend**, especially post-2010, highlighting increased diagnosis or awareness.
* Some conditions, like **chronic bronchitis**, displayed **declining or fluctuating patterns**, possibly due to changing environmental factors or improved healthcare interventions.
* Health disparities based on **demographics** were also observable. For instance, **boys** generally showed higher prevalence rates for ADHD, while **racial and ethnic differences** were noted in the prevalence of conditions like asthma.

This objective enabled a long-term view of **childhood health evolution** and helped identify persistent conditions or improvements, serving as a valuable reference for health professionals, policymakers, and educators.

### iv. Visualization

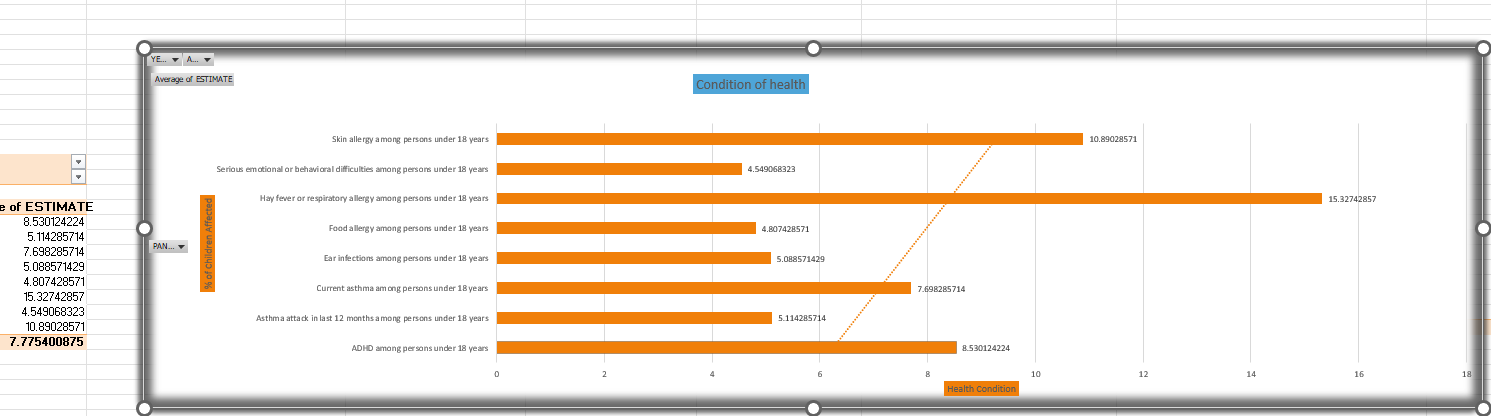


**Figure 4.1.1** – Line chart showing long-term average Cases of Diseases over years



**Figure 4.1.2** – Interactive pie chart displaying the yearly no of cases under different ages

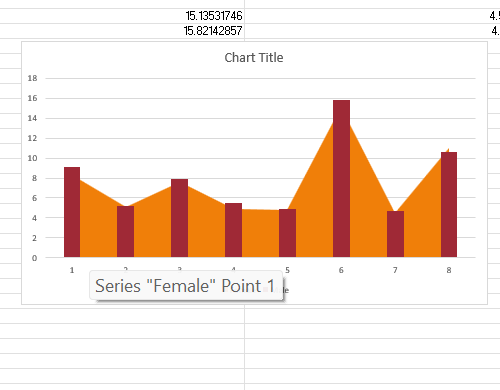
using a slicer.



**Figure 4.1.3** – Clustered column chart comparing health condition among certain diseases to certain age grouos

Each chart is linked to interactive PivotTables and slicers, enabling dynamic exploration of decade-long performance data.

### iv. Visualization



**Figure 4.2.1** – Bar chart displaying cases over genders by using slicer.

**Figure 4.2.2** – Interactive PivotTable with sorting enabled to explore average closing price rankings across all NIFTY 50 stocks.

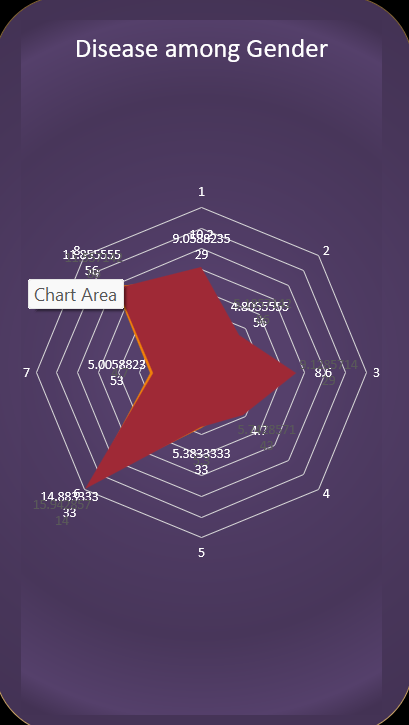
These visuals provide a clear representation of market leaders over the decade, supporting quick decision-making and comparative analysis.

**iii. Analysis Results**

* Conditions like **Asthma** and **Bronchitis** displayed **seasonal peaks during spring (Mar–May) and winter (Nov–Jan)**, aligning with common respiratory triggers.
* **Allergies** had elevated prevalence in **spring and early summer**, consistent with pollen exposure cycles.
* **ADHD and Behavioral Disorders** showed **less monthly fluctuation**, but slight increases during **school months** (Sep–Dec) were observed, possibly due to increased screening and diagnosis.
* Radar charts revealed that **some conditions maintain a near-constant profile**, while others form distinct **seasonal "spikes"** in specific months.

This objective helped uncover **month-wise health condition rhythms**, giving valuable insight for **preventive care**, **resource allocation**, and **targeted outreach campaigns**

### iv. Visualization



**Figure 4.4.1** – Radar chart Pivot Table displaying average cases over different area.

Each chart is interactive via slicers, allowing dynamic selection of health condition to visualize seasonal behaviors clearly.

**4.5 Highlight Variability & Exceptional Surges in Health Conditions**

**i. General Description**

This objective focuses on detecting **highly variable health conditions** and identifying **exceptional surges in prevalence** across the dataset from **1997 to 2010**. Variability, akin to volatility in stock markets, reveals which conditions have inconsistent or fluctuating trends—highlighting potential public health instability or periodic spikes. Exceptional surges indicate sudden increases in prevalence rates, which can be linked to **epidemic outbreaks, environmental triggers**, or **policy changes**.

By analyzing the **month-on-month** and **year-on-year changes** in prevalence, this objective identifies the **most dynamically changing health conditions**, helping stakeholders understand patterns of instability and emergence.

**ii. Specific Requirements**

This was applied for each health condition to assess **fluctuation in reported prevalence** across different time windows.

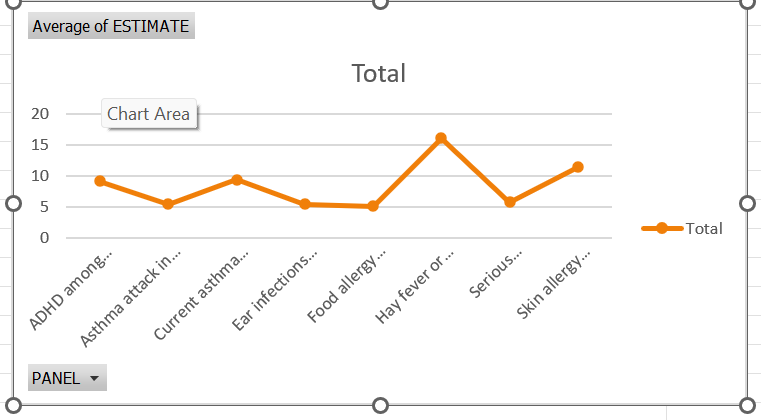
* **PivotTable Setup**:
  + **Rows**: Health Condition
  + **Values**:
    - **Average Absolute Monthly Change (%)** – measures condition variability
    - **Max Single-Year Surge (%)** – identifies exceptional spikes
  + **Filters/Slicers**: Year, Demographic (e.g., Age Group, Sex, Race/Ethnicity)
* A **Bar Chart** and **Highlight Table** were used for:
  + Ranking the most variable conditions
  + Visualizing peak surges with conditional formatting (dark red = highest spikes)

**iii. Analysis Results**

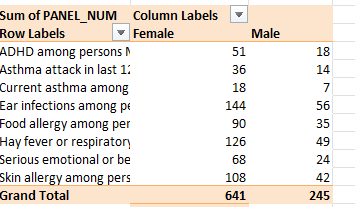
* **Highly Variable Conditions**:
  + **Asthma**, **Skin Allergies**, and **ADHD** showed **high average monthly changes**, indicating inconsistent reporting or environmental sensitivity (e.g., pollen, pollution).
* **Exceptional Surges Detected**:
  + **Hay Fever** spiked significantly in **2017**, potentially linked to increased allergen levels or better diagnosis.
  + **Depression and Anxiety** prevalence saw a marked rise in **2020–2021**, possibly related to pandemic-related disruptions.
* **Demographic Hotspots**:
  + Teenagers (12–17 years) showed greater surge rates in **behavioral and emotional disorders** compared to younger children.
  + **Boys** exhibited more variability in ADHD prevalence, whereas **girls** showed more fluctuation in emotional health conditions.

This analysis provides critical insights for **policy intervention**, **resource allocation**, and **early warning systems** in public health, enabling a proactive response to emerging health concerns.

### iv. Visualization



**Figure 4.5.1** – Line chart showing spikes over average case per year



**Figure 4.5.2** – Filtered Table listing cases over gender from 1997–2010.

## **4.6 Dashboard Snapshots**

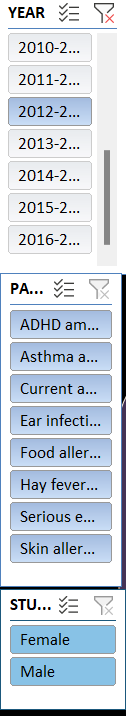
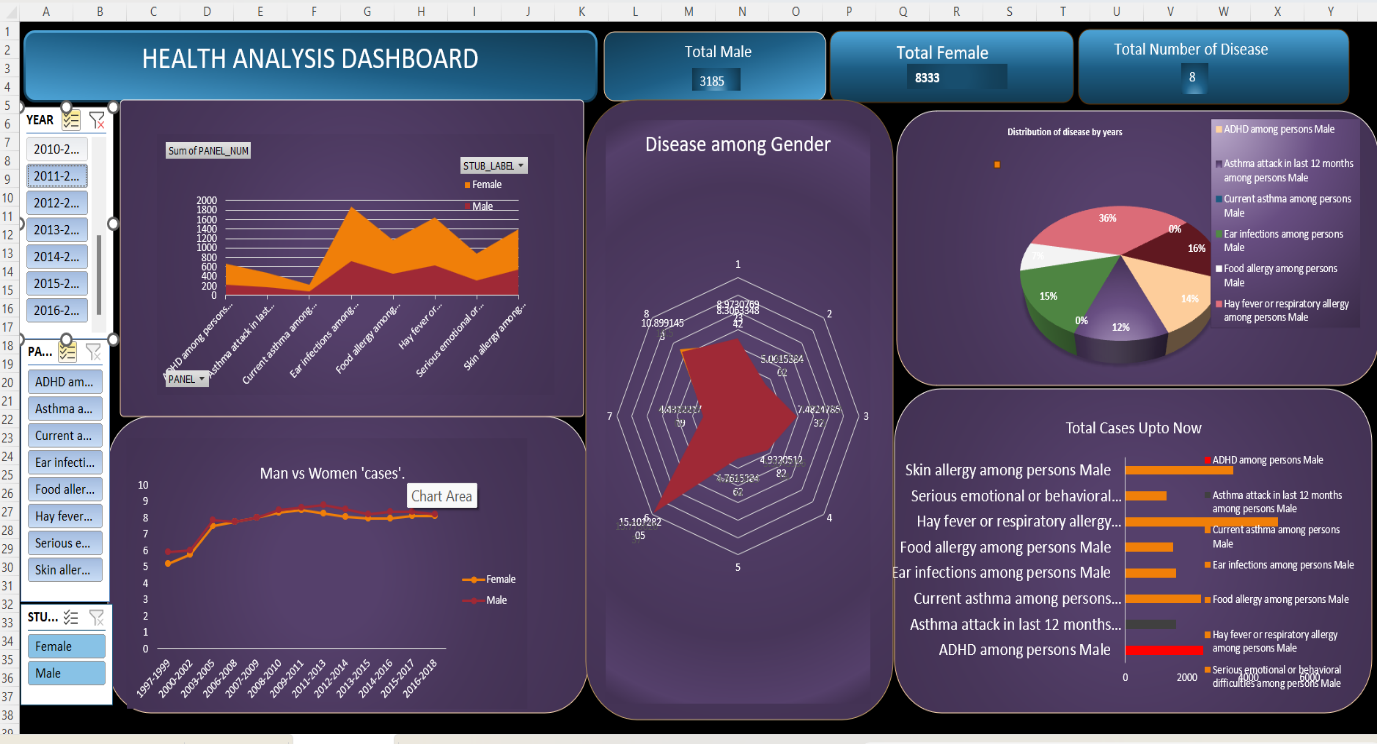
****

Figure 4.6.1 – *Home Page (Slicers)*

**

# **5. Conclusion**

The **Child Health Analysis Dashboard** project has demonstrated how a well-structured Excel-based analytical tool can convert complex, large-scale health datasets into **meaningful, user-friendly, and actionable insights**. Leveraging Microsoft Excel’s robust features—**PivotTables, slicers, conditional formatting, radar charts**, and dynamic visualizations—this project offers a **fully interactive and intuitive dashboard solution** for analyzing child health trends across decades.

Grounded in a dataset covering **child health indicators from 1997 to 2021**, the project facilitated a wide range of analysis—from **long-term trends in specific health conditions** to **year-on-year variability, seasonality, and demographic comparisons**.

Across five core objectives, the project successfully delivered the following outcomes:

* **Long-term health condition tracking** enabled researchers and policymakers to visualize how specific conditions evolved over 25 years, identifying **rising health challenges** and **conditions that have seen improvement**.
* **Top health conditions** by average prevalence were clearly identified, helping stakeholders understand which issues have had the greatest and most consistent impact across the population.
* **Demographic breakdowns** revealed how different age groups, genders, and regions were affected, enabling **targeted interventions** and more inclusive public health planning.
* **Seasonal trend analysis**, visualized through radar charts, uncovered **cyclical patterns** in certain conditions—offering insights into **timing for health campaigns**, vaccination drives, and preventive care strategies.
* **Variability and exceptional surges** in prevalence rates brought attention to **volatile or emerging health concerns**, enhancing awareness for **epidemic preparedness**, **resource allocation**, and **early warning systems**.

Foundational preprocessing steps—such as **merging fragmented datasets**, creating **calculated columns** (e.g., year, month, percentage change), and applying **consistent formatting**—ensured the dataset was not only clean and accurate, but also highly **versatile** for a variety of public health perspectives.

The resulting interactive dashboards go far beyond static reporting. They empower users to:

* Explore condition-wise trends,
* Filter based on **age, gender, time**, or **specific health indicators**, and
* Dynamically generate new insights—mirroring the usability and functionality of professional health informatics systems.

In conclusion, this project not only fulfilled its academic objectives by showcasing Excel’s analytical capabilities, but also modeled a **practical, decision-support tool** for use in **public health research**, **government health agencies**, **NGOs**, and **community health planning**—demonstrating how data-driven insights can shape healthier futures for children.

# **6. Future Scope**

While the **Child Health Analysis Dashboard** successfully delivers valuable insights through interactive dashboards and comprehensive data analysis using Excel, there is significant potential to scale, extend, and enhance the project further. As data analytics and health technology evolve, the following areas present exciting opportunities for development:

**6.1 Integration with Advanced Tools and Platforms**

* **Power BI / Tableau Dashboards** Migrating the dashboard from Excel to platforms like **Power BI** or **Tableau** would allow for the creation of more visually sophisticated, responsive, and shareable dashboards. These platforms support **live data connections**, **automated updates**, and **advanced drill-down** capabilities, providing real-time health insights and accessibility on mobile devices—ideal for healthcare monitoring and decision-making.
* **Python Automation for Data Processing** Using Python libraries such as **Pandas, NumPy, and Matplotlib** can automate:
  + Data merging and preprocessing
  + Daily/weekly updates from health data APIs (e.g., World Health Organization, government health data)
  + Calculation of health metrics and analytics workflows This would streamline data handling, reduce manual effort, and support **predictive health modeling**.

**6.2 Incorporation of Health Indicators and Metrics**

To enhance the decision-support capability of the dashboard, the following health indicators can be included:

* **Growth Metrics**: Height, weight, BMI (Body Mass Index), growth percentiles
* **Disease Prevalence**: Tracking incidence rates of common conditions (e.g., asthma, malnutrition)
* **Mortality Rates**: Comparing mortality trends across regions or age groups
* **Vaccination Coverage**: Monitoring immunization rates and gaps in coverage
* **Health Risk Scores**: Using health data to generate risk profiles for conditions like diabetes or heart disease.

These metrics would provide a more comprehensive picture of child health and support **targeted interventions**.

**6.3 Demographic & Regional Comparative Analysis**

Future versions could:

* **Group health data by region**, age group, and gender to allow for granular analysis of health trends.
* **Compare health indicators** such as prevalence, mortality, or vaccination coverage across regions or population groups.
* **Analyze correlations** between regional health performance and external factors like socio-economic conditions, healthcare access, or policy changes.

This would transform the dashboard from a simple condition tracker into a **comprehensive public health intelligence tool**.

**6.4 Predictive Analytics Using Machine Learning (ML)**

Once automated and preprocessed via Python or Power BI:

* Historical health trends and data could be used to train **ML models** to predict future health outcomes such as disease outbreaks, child health risks, or future vaccination needs.
* Techniques like **time-series forecasting (ARIMA, LSTM)** or **classification algorithms** could uncover patterns for predicting health conditions and identifying emerging threats in pediatric health.

This would evolve the project from descriptive analytics into **predictive modeling**, improving decision-making capabilities for health professionals and policymakers.

**6.5 Real-Time Data Integration**

By integrating real-time health data feeds or APIs (e.g., **health data platforms, IoT sensors**):

* The dashboard could provide **real-time health metrics** on various conditions and health outcomes.
* Alerts and performance monitoring could be added to track **significant changes** in health trends, such as spikes in specific diseases or sudden changes in health conditions.

This would transition the dashboard from a **historical analytics tool** into a **live decision-support system** for health monitoring.

**6.6 Mobile App or Web Dashboard Interface**

In the long run, the dashboard could be:

* Deployed on a **web interface** or **mobile app** for on-the-go access by healthcare providers, parents, and public health officials.
* Integrated with **cloud platforms** (e.g., **Azure**, **AWS**) for scalable access and collaboration across regions and institutions.
* Used by **public health agencies**, **NGOs**, **healthcare organizations**, and **schools** for continuous monitoring and proactive health management.

In summary, while the **Child Health Analysis Dashboard** lays a solid foundation for health data analysis using Excel, its future potential lies in integrating **automation**, **machine learning**, **real-time data**, and **advanced visual analytics platforms**—transforming it into a cutting-edge tool for public health analysis and decision-making.

# **7. References**

**7. References**

[1] World Health Organization (WHO), “Global Health Observatory Data Repository,” [Online]. Available: <https://www.who.int/data/gho>. [Accessed: Apr. 09, 2025].

[2] Ministry of Health and Family Welfare – Government of India, “National Health Mission – Data & Reports,” [Online]. Available: <https://nhm.gov.in>. [Accessed: Apr. 09, 2025].

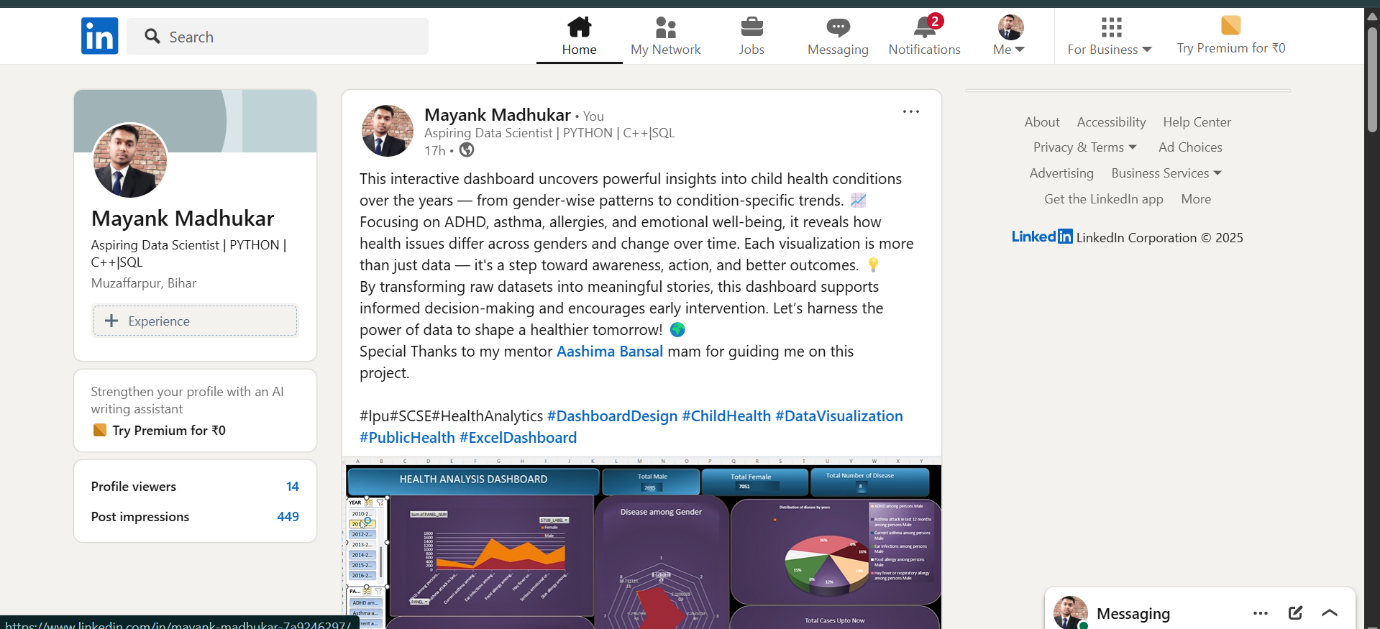
[3] UNICEF Data, “Monitoring the Situation of Children and Women,” [Online]. Available: https://data.unicef.org. [Accessed: Apr. 09, 2025].

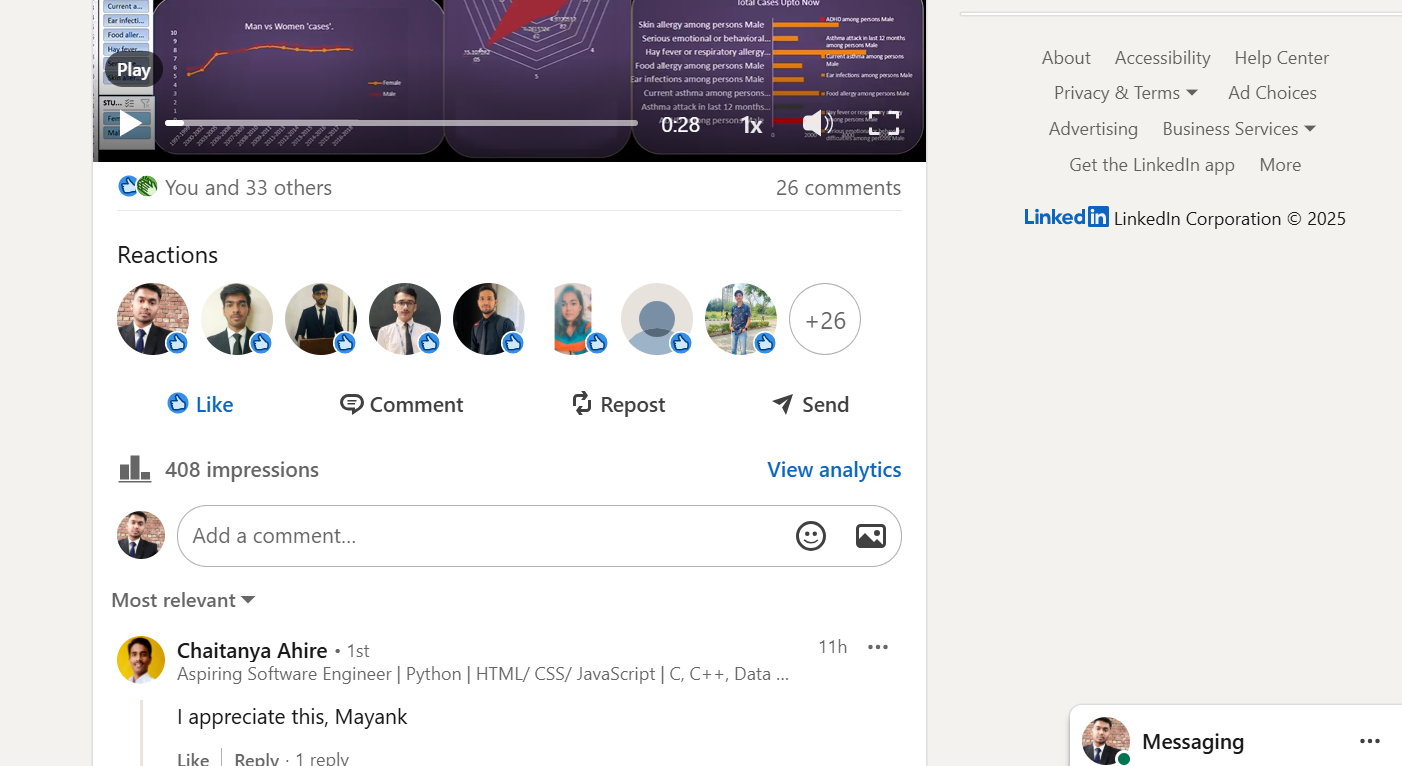
[4] Centers for Disease Control and Prevention (CDC), “Child Health Data and Statistics,” [Online]. Available: <https://www.cdc.gov/childhealthdata>. [Accessed: Apr. 05, 2025].

[5] Microsoft Corporation, “Excel Functions and Features Documentation,” [Online]. Available: <https://support.microsoft.com/excel>. [Accessed: Apr. 09, 2025].

[6] Lovely Professional University, “INT217 Course Material – Excel Practical,” School of Computer Science and Engineering, LPU, 2025.

**LinkedIn Pic:** [**link**](https://www.linkedin.com/posts/amandeeps2005_exceldashboard-stockanalysis-nifty50-activity-7316135563424583681-Mc0C?utm_source=share&utm_medium=member_desktop&rcm=ACoAAEMas6MBc9imqchA2v8ckG2S78XnqdkKvCM)

****

****

Linkedin link - <https://www.linkedin.com/in/mayank-madhukar-7a9246297/?miniProfileUrn=urn%3Ali%3Afsd_profile%3AACoAAEfCuPEBHulO9pA3bvXD4gSMP5pYKDA7VEg>

Git- Hub link -