# USA Healthcare Analysis – Excel-Based Data Project

## 1. Outline

This report provides a structured data analysis of healthcare performance across the United States using Microsoft Excel. It follows a clear analytical flow, covering:

- Introduction
- Story of Data
- Data Splitting and Preprocessing
- Pre-Analysis Insights
- In-Analysis Observations
- Post-Analysis Recommendations
- Visualizations & Charts
- Conclusion

The purpose is to identify key patterns in healthcare expenditure, patient distribution, and outcomes across different regions and healthcare services, highlighting trends that can support data-driven health policy decisions.

## 2. Introduction

### Objective of the Project

The goal of this project is to analyze how healthcare costs, patient numbers, and medical service types vary across states in the USA. It aims to uncover where resources are most utilized and where improvements may be needed.

### Problem Being Addressed

Healthcare in the U.S. is complex, with large differences in cost, accessibility, and patient outcomes between states. This analysis uses Excel to transform raw healthcare data into insights that can help identify disparities and guide better allocation of resources.

### Key Datasets and Methodologies

The dataset includes records of patient demographics, hospital categories, total cost of healthcare services, and regional data.
All analysis was conducted using Microsoft Excel, applying:

- **PivotTables** for summarization by region and category
- **IF and COUNTIFS** formulas for segmentation

- **Descriptive analytics** for average and distribution insights
- **Charts and dashboards** for visualization of patterns

# 3. Story of Data

This dataset captures the activities of healthcare facilities in the United States for the year 2016. It includes information on facility performance, patient discharges, case severity, medical and surgical classifications, and associated costs. The data helps paint a clear picture of healthcare delivery patterns, operational efficiency, and cost implications across different facilities.

## Stakeholders of the project

The key stakeholders include healthcare administrators, policymakers, medical directors, and financial analysts responsible for resource allocation, cost management, and improving patient care outcomes across all facilities.

## Value to the Industry

It demonstrates how Excel can support data-backed healthcare management without requiring complex programming, making it practical for non-technical analysts in health institutions.

## Data Source

The dataset was downloaded from [Kaggle.com](Kaggle.com)

## Data Structure

Each record represents healthcare service or patient data for a particular category or region. Key columns include:

- **State/Region**
- **Hospital Type**
- **Number of Patients**
- **Average Cost per Service**
- **Total Expenditure**
- **Mortality Rate**
- **Readmission Rate**

## Important Features and Their Significance

- **Cost & Expenditure:** Reflect financial efficiency and system strain.
- **Patient Count:** Highlights population health needs.
- **Readmission & Mortality Rates:** Indicators of healthcare quality.

## Data Limitations

Some regions had incomplete data, and values were generalized per state rather than per hospital. Despite that, the dataset remains useful for regional-level analysis.

# 4. Data Splitting and Preprocessing

## Data Cleaning

Steps taken included:

- Removing blank or duplicate rows
- Converting cost and rate columns to numeric values
- Standardizing state names and hospital types

## Handling Missing Data

Missing values were filled using averages from similar regions to preserve data consistency.

## Data Transformation

New calculated fields were introduced to enhance interpretation:

- **Cost per Patient = Total Expenditure ÷ Number of Patients**
- **Efficiency Score = (1 – Readmission Rate) × 100**
- **Mortality Classification:** Low, Medium, High

# 5. Pre-Analysis Insights

- Discharges have increased steadily over the years in most facilities, indicating rising patient inflow.
- Facilities with higher APR severity levels tend to have higher mean costs per discharge.
- Certain APR descriptions are linked to significantly higher costs across all facilities.
- Some facilities maintain relatively stable costs despite fluctuating discharge volumes.
- APR medical cases show a tendency toward higher costs than surgical cases in most facilities.
- A few facilities consistently handle the majority of discharges, dominating patient care delivery.
- Higher discharge volumes do not always lead to lower mean costs, suggesting efficiency gaps.
- Changes in APR codes over the years align with variations in cost and severity patterns.
- Facilities with frequent severe APR cases incur the highest costs per patient on average.
- Facilities with more complex APR descriptions often report longer patient stays and higher bills.
- Mean cost differences are wider in surgical cases compared to medical ones across facilities.

- Facilities with low discharge volumes sometimes have surprisingly high mean costs per case.
- Some facilities show remarkable cost efficiency, managing high discharges at below-average costs.
- The distribution of APR codes suggests that a small number of codes dominate facility reporting.
- Severity levels directly influence discharge costs, with critical cases being the most expensive.
- There are significant year-on-year fluctuations in costs for certain APR codes, indicating changing treatment dynamics

# 6. In-Analysis Observations

- Diabetes consistently shows up as the most prevalent illness, the most severe, and among the most discharged cases, signaling a triple burden on healthcare facilities.
- Cardiomyopathy has the lowest prevalence and discharge rates, yet its severity level indicates it might need specialized care despite fewer cases.
- Bipolar Disorders dominate the discharge chart, suggesting either frequent admissions or effective treatment protocols leading to quick discharges.
- Fever and Abdominal Pain appear as mid-level illnesses in both prevalence and severity, representing common but manageable hospital admissions.
- Top hospitals such as New York Presbyterian Hospital handle both the highest patient volumes and the highest treatment costs, indicating their central role in healthcare delivery.
- Treatment cost differences across the top three hospitals remain small, pointing toward standardized pricing or similar case complexities.
- Hospitals with high patient counts also show equally high discharge numbers, reflecting efficiency in handling patient inflow and outflow.
- The close-range in-patient discharges across hospitals suggests a balanced capacity distribution among top facilities.
- Acute Leukemia shows moderate severity but low discharge rates, hinting at prolonged treatment durations for such cases.
- The dashboard shows a consistent pattern where high prevalence often aligns with high discharge rates, except in rare illnesses.
- Cost data indicate that high patient inflows are directly linked to high financial expenditures in top facilities.
- Illnesses with low prevalence rates also tend to have lower discharges, except for bipolar disorders which defy this trend.
- Patient counts for the top hospitals stay above the 1,000 marks, showing high patient dependence on these facilities.
- Minimal variation in severity percentages among middle-tier illnesses suggests shared treatment complexity levels.
- Some illnesses like Fever, though not severe, still create a moderate load due to their frequency in hospital admissions.

- The repeated dominance of New York Presbyterian facilities across multiple metrics highlights them as leading care providers.
- Cost efficiency may vary slightly among hospitals, but the gap is too small to indicate major operational differences.
- Low discharge numbers for Cardiomyopathy may reflect critical care needs or longer recovery periods for heart-related illnesses.
- The consistent cost patterns suggest possible nationwide healthcare billing standards for similar treatment types.
- Hospitals with slightly lower patient counts may have untapped capacity to handle additional cases if demand increases.
- The narrow cost range across hospitals points toward cost uniformity regardless of patient volume differences.
- The data provides a holistic view of illness severity, prevalence, cost, and discharge trends in U.S. healthcare facilities

## Methods Used in Excel

- **PivotTables and Slicers** for filtering by region and category
- **COUNTIFS** for counting conditions
- **AVERAGEIFS** for state-level cost comparison
- **Charts and conditional formatting** for clarity

# 7. Post-Analysis Recommendations

- Healthcare administrators should prioritize diabetes management programs given its dominance in prevalence, severity, and discharge metrics.
- -Mental health facilities need further support since bipolar disorders account for disproportionately high discharge numbers.
- -Cost management strategies could target hospitals with the highest expenditures to ensure efficiency without compromising care quality.
- -Preventive care for mid-prevalence illnesses like Fever and Abdominal Pain can reduce hospital admission rates and overall costs.
- -More specialized facilities could be established for illnesses like Cardiomyopathy and Acute Leukemia due to their severity and longer treatment cycles.
- -Policymakers should consider balancing patient loads by directing more cases to hospitals with slightly lower patient counts.
- -Investment in mental health infrastructure could help manage the high turnover rates in Bipolar Disorder cases effectively.
- -Standardized treatment protocols across facilities may help maintain the already narrow cost variations among top hospitals.
- -Data analytics should be applied continuously to detect emerging illness trends early for proactive healthcare planning.
- -Hospitals with high discharge efficiency should be studied to replicate their practices in lower-performing facilities.

- -Preventive healthcare campaigns targeting diabetes could reduce both prevalence and treatment costs significantly over time.
- -More research is needed into why some illnesses have high severity but low discharge rates, as seen with Cardiomyopathy.
- -Mid-level severity illnesses like Fever could be managed through telemedicine to free up hospital capacity for critical cases.
- -Expansion of outpatient care services can reduce hospitalization costs for illnesses with shorter treatment cycles.
- -Resource allocation should be dynamic, responding to yearly shifts in illness prevalence and severity patterns.
- -Mental health awareness campaigns could reduce frequent hospitalizations for disorders like Bipolar Disorders.
- -High-cost hospitals should explore cost-reduction strategies through technology adoption and operational efficiency measures.
- -Real-time health data dashboards could help decision-makers act quickly on rising illness trends or cost spikes.
- -Collaboration between high- and low-patient-volume hospitals can ensure better distribution of healthcare resources.
- -Incentives could be provided to hospitals achieving both high discharge efficiency and low average treatment costs.
- -Chronic disease prevention programs should be expanded to reduce long-term healthcare costs nationwide.
- -Training healthcare workers on data-driven decision-making can improve both patient outcomes and operational efficiency

## Comparison with Initial Insights

The early hypotheses were partly confirmed while some regions indeed had high costs, efficiency was more linked to patient management practices than pure spending.

# 8. Data Visualizations & Charts

The analysis featured multiple visuals to make patterns easier to interpret:

- **Column Chart:** Total healthcare expenditure by region
- **Pie Chart:** Distribution of patients across hospital types
- **Line Chart:** Yearly trend of average cost per service
- **Dashboard:** An interactive Excel view summarizing cost, patients, and outcomes

Each visual was used to validate a specific pattern or hypothesis.

# 9. Conclusion

This project highlights how Excel can provide clear, data-driven insight into healthcare efficiency and performance across U.S. states.
It confirms that spending alone does not guarantee better outcomes, and efficient management plays a larger role in healthcare quality.

Overall, Excel proves to be an accessible yet powerful tool for public health analytics and policy reporting.