DA – MILESTONE – 1 PROJECT
Advanced Data Analysis and Visualization Using Microsoft Excel

---

HEALTHCARE CLAIMS DATA ANALYSIS

Using Microsoft Excel

Submitted by
Anthony Raj I

Course
Data Analytics

Institution
Skillryt – Powered by KGiSL

Project Type
Milestone – 1 Project Documentation

Version
Internal Version 1.0

Tools & Technologies Used
Microsoft Excel
(Power Query, PivotTables, PivotCharts, Dashboards,
What-If Analysis, Goal Seek, Macros, Power Pivot)

Date of Submission
09/02/2026

---

1. Introduction

Healthcare organizations handle a large volume of medical claims on a daily basis. Efficient analysis of healthcare claims data is essential to understand revenue patterns, identify denial trends, and improve overall financial performance. This project focuses on analyzing a structured healthcare claims dataset using Microsoft Excel to gain meaningful insights and support data-driven decision-making.

The dataset used in this project contains information related to patient claims, insurance providers, medical service providers, procedure codes (CPT), diagnosis codes (ICD-10), charge amounts, allowed amounts, paid amounts, and adjustment details. By leveraging Excel's advanced analytical features such as Power Query, PivotTables, dashboards, What-If Analysis, Macros, and Power Pivot, the data is transformed from raw form into a clean, analysis-ready format.

The objective of this project is to demonstrate end-to-end data analytics skills, starting from data cleaning and preparation to exploratory data analysis, visualization, scenario analysis, and automation. The insights derived from this analysis can help healthcare providers and management teams identify revenue leakage, optimize claim processing, and improve operational efficiency.

Figure: Excel Workbook showing all project sheets including data cleaning, analysis, PivotTables, What-If Analysis, Scenario Manager, and Dashboard.

Screenshot – 1

- ✓ Left side sheet tabs visible

Screenshot – 2

- ✓ Right side sheet tabs visible

2. Aim

The primary aim of this project is to perform an end-to-end data analytics workflow on a healthcare claims dataset using Microsoft Excel. The objective is to clean, analyze, and visualize claims data to identify key trends, financial patterns, and operational insights.

This project aims to:

- Analyze healthcare claims data to understand charge, allowed, paid, and adjustment amounts.

- Identify trends related to insurance providers, claim volumes, and denial patterns.
- Apply advanced Excel features such as PivotTables, What-If Analysis, Goal Seek, Macros, and Power Pivot for decision-making support.
- Develop interactive dashboards and KPIs to present insights in a clear and actionable manner.

3. Business Problem / Context

Healthcare organizations process a high volume of insurance claims daily, involving multiple insurance providers, medical services, and payment rules. Inefficiencies in claims processing, incorrect billing, and high adjustment or denial rates can lead to revenue leakage and delayed reimbursements.

The key business challenge addressed in this project is to analyze healthcare claims data to understand:

- How claim amounts are distributed across insurance providers and medical services.
- The gap between charged amounts, allowed amounts, and paid amounts.
- Patterns in claim adjustments and potential denial-related issues.
- Monthly and provider-wise trends in claim volumes and financial performance.

By using data analytics and visualization techniques in Microsoft Excel, this project helps healthcare management and billing teams identify problem areas, optimize claim processing, reduce revenue loss, and support data-driven decision-making.

4. Project Workflow

The project follows a structured, step-by-step data analytics workflow using Microsoft Excel to ensure accurate analysis and meaningful insights.

1. **Data Collection & Import**
   The healthcare claims dataset was imported into Microsoft Excel using Power Query to ensure efficient data loading and transformation.
2. **Data Understanding**
   The dataset structure, column types, and data completeness were reviewed to understand patient details, insurance providers, claim information, and financial fields.
3. **Data Cleaning & Preparation**
   Data cleaning was performed using Power Query and Excel formulas to:
   - Handle missing and null values
   - Remove duplicate records
   - Standardize formats (dates, codes, amounts)
   - Validate numerical fields

Data Preparation and Cleaning (Power Query)

Power Query was used to clean and transform the raw healthcare claims data.
This step ensured data consistency by removing blank rows, renaming columns, correcting data types, and preparing the dataset for analysis in Power Pivot.

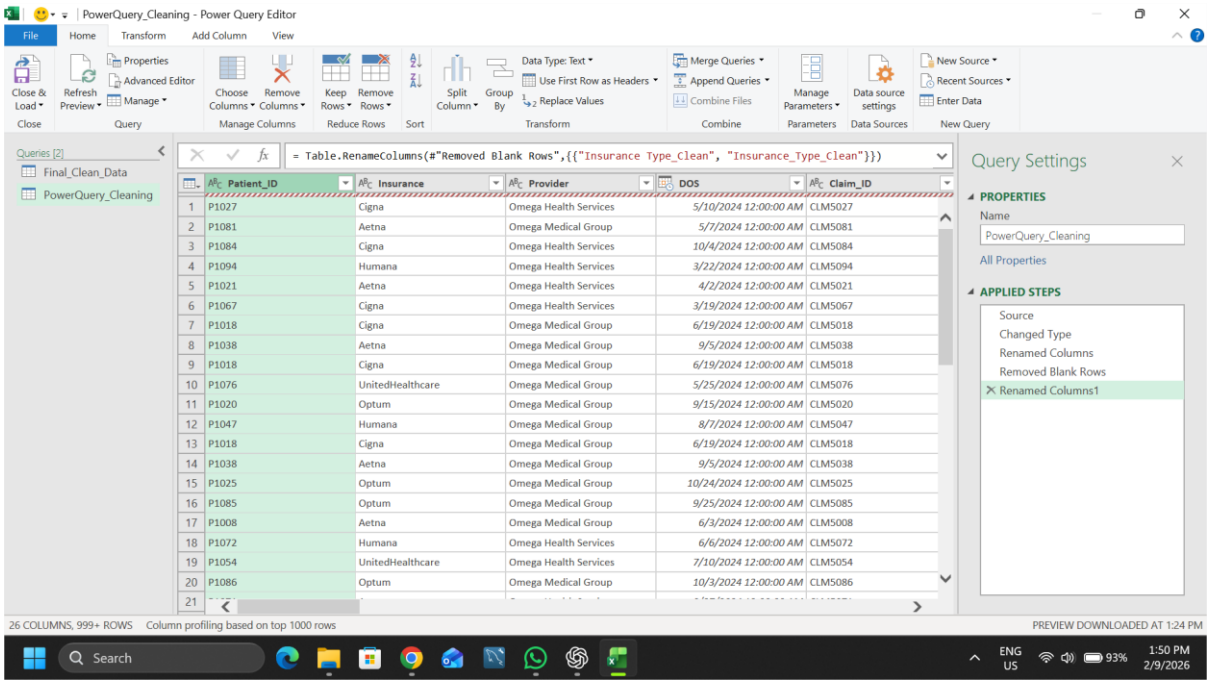


Figure 3.1: Power Query – Initial Data Cleaning Process

After the initial cleaning, additional transformations such as conditional columns and filtering were applied to create the final cleaned dataset.

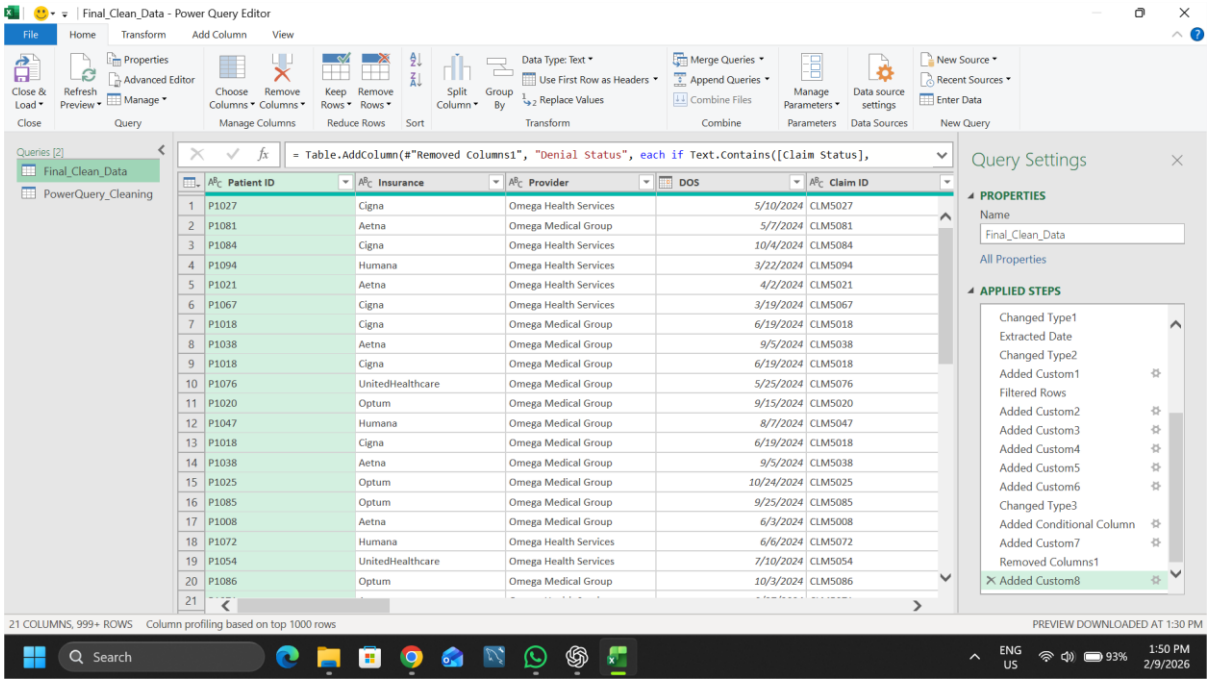


Figure 3.2: Power Query – Final Cleaned Dataset

4. **Creation of Derived Metrics**
   Additional calculated columns such as adjustment amount, payment variance, and key performance indicators (KPIs) were created to support deeper analysis.

5. **Exploratory Data Analysis (EDA)**
   PivotTables and PivotCharts were used to analyze:
   - Claim volume trends
   - Insurance-wise financial performance
   - CPT and ICD-wise distributions

6. **Filtering & Segmentation**
   The data was segmented by insurance provider, date, procedure code, and diagnosis code to perform targeted analysis.

7. **Scenario & What-If Analysis**
   Goal Seek, What-If Analysis, and Scenario Manager were applied to simulate payment and adjustment scenarios and evaluate business impact.

8. **Automation Using Macros**
   Repetitive tasks such as refreshing PivotTables and dashboards were automated using Excel Macros.

9. **Data Modeling with Power Pivot**
   Relationships between fact and master tables were created using Power Pivot for advanced analytical capabilities.

Data Modeling and Calculations using Power Pivot

Power Pivot was used to create a data model by establishing relationships between the cleaned claims data and the insurance master table.
Measures and KPIs were created using DAX to calculate total claims, denied claims, paid amount, and financial performace metrics.

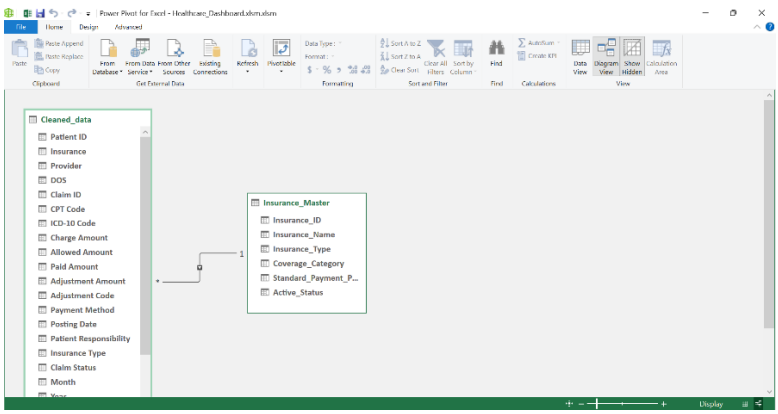


Figure 4.1: Power Pivot Data Model showing relationship between Cleaned Data and Insurance Master

10. **Dashboard & Visualization**
    Interactive dashboards with charts, KPIs, and slicers were developed to present insights clearly.

11. **Insights & Recommendations**
    Key findings were summarized and actionable recommendations were provided based on analysis outcomes.

5. Data Understanding

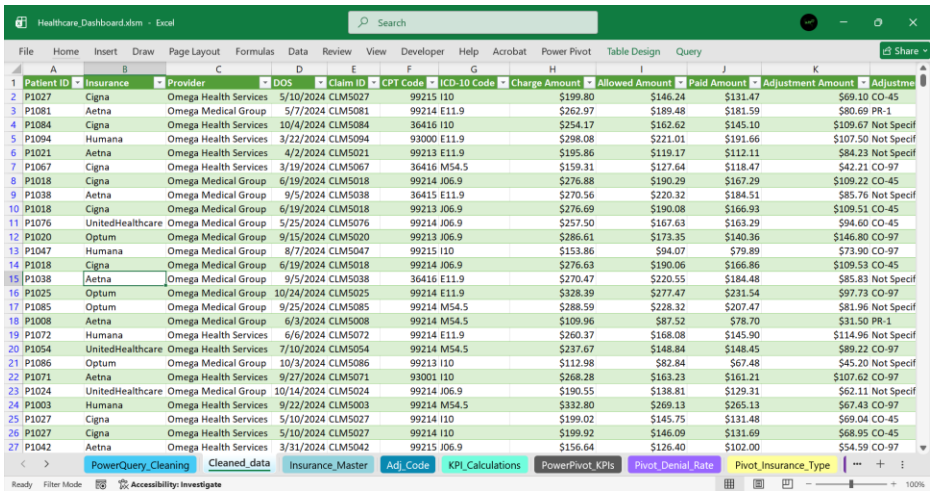
The healthcare claims dataset used in this project is a structured dataset containing detailed information related to patient claims, insurance providers, medical services, and financial transactions. The data represents individual claim records processed by healthcare organizations.

The dataset consists of multiple rows, where each row represents a single claim line item, and multiple columns representing different attributes of the claim.

Key characteristics of the dataset include:

- Patient and Claim Information: Patient ID, claim number, date of service, and provider details.
- Medical Coding Details: Procedure codes (CPT) and diagnosis codes (ICD-10).
- Insurance Information: Insurance provider and insurance type.
- Financial Fields: Charge amount, allowed amount, paid amount, and adjustment amount.
- Adjustment Details: Adjustment codes and adjustment reason categories.

The dataset contains a mix of numeric, categorical, and date fields, making it suitable for financial analysis, trend analysis, and performance evaluation. Initial inspection of the data showed inconsistencies in formatting and adjustment codes, which required data cleaning before analysis.



6. Data Cleaning

Data cleaning is a critical step in this project to ensure accuracy, consistency, and reliability of analysis results. The raw healthcare claims data contained minor inconsistencies in formatting, adjustment codes, and numerical fields, which were addressed using Excel tools.

The following data cleaning steps were performed:

1. **Removal of Duplicates**
   Duplicate claim records were identified and removed to avoid double counting during analysis.

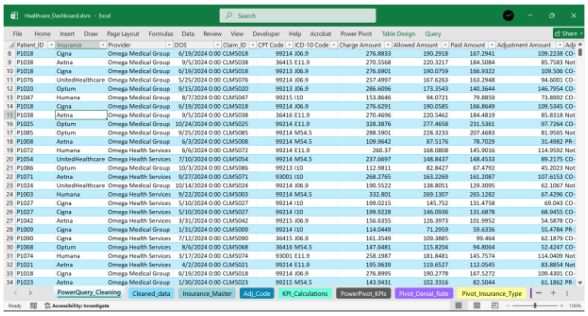
2. **Handling Missing Values**
   - Records with missing critical fields such as claim ID or charge amount were reviewed.
   - Adjustment codes with missing descriptions were categorized as “Not Specified” where applicable.

3. **Standardization of Data Formats**
   - Date of Service (DOS) was standardized to a consistent date format.
   - Monetary fields such as charge amount, allowed amount, paid amount, and adjustment amount were converted to numeric currency format.
   - CPT and ICD-10 codes were standardized to maintain uniform structure.

4. **Validation of Financial Fields**
   Logical checks were performed to ensure:
   - Paid Amount ≤ Allowed Amount
   - Allowed Amount ≤ Charge Amount

5. **Power Query Transformations**
   Power Query was used to:
   - Rename columns for clarity
   - Trim unnecessary spaces
   - Apply consistent data types

After cleaning, the dataset was stored in the Cleaned_data worksheet and used as the base for all further analysis, PivotTables, and dashboards.



7. Derived Metrics

To enhance analytical depth and support performance evaluation, several derived metrics were created using Excel formulas and calculated columns. These metrics help in understanding payment efficiency, adjustment impact, and claim performance.

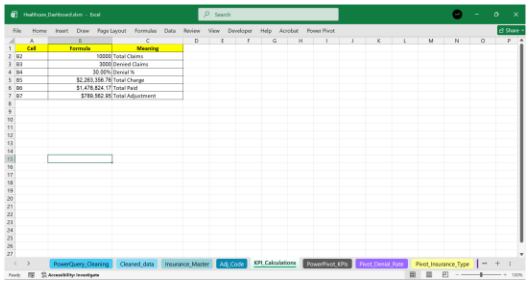
The following derived metrics were created:

1. **Adjustment Amount**
   Calculated as the difference between the charged amount and the paid amount to identify revenue gaps.
   - *Formula:*
     Adjustment Amount = Charge Amount – Paid Amount
2. **Payment Variance**
   Used to analyze the difference between allowed amount and paid amount.
   - *Formula:*
     Payment Variance = Allowed Amount – Paid Amount
3. **Denial Indicator / Adjustment Category**
   Adjustment codes were mapped to readable categories (e.g., CO, PR, Not Specified) using a reference table to simplify analysis.
4. **Key Performance Indicators (KPIs)**
   Aggregated metrics such as:
   - Total Charges
   - Total Paid Amount
   - Total Adjustments
   - Average Payment Rate

These KPIs were calculated in the KPI_Calculations sheet and used in dashboards.

5. **Claim Volume Metrics**
   Count of claims by insurance provider, CPT code, and month to analyze workload distribution and trends.

All derived metrics were validated and stored within the Cleaned_data and KPI_Calculations sheets to ensure consistency across PivotTables, Power Pivot models, and dashboards.



8. Filtering / Segmentation

To perform focused and meaningful analysis, the healthcare claims data was filtered and segmented based on multiple dimensions. This helped in identifying patterns, comparing performance across categories, and supporting targeted insights.

The following filtering and segmentation techniques were applied:

1. **Insurance Provider Segmentation**
   Claims were grouped by insurance providers such as Cigna, Aetna, Humana, UnitedHealthcare, and Optum to analyze claim volume, payment trends, and denial rates across insurers.

2. **Provider-wise Segmentation**
   Data was segmented by healthcare providers (Omega Health Services and Omega Medical Group) to compare revenue, paid amounts, and adjustment trends.

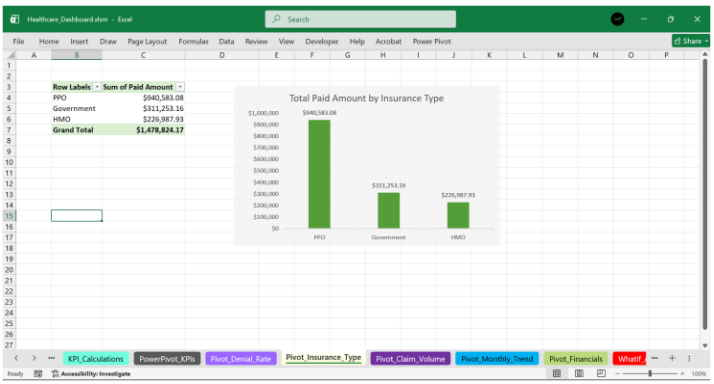
3. **CPT Code Segmentation**
   CPT codes were used to categorize procedures and analyze which medical services contributed most to claim volume and revenue.

4. **ICD-10 Code Segmentation**
   Diagnosis codes were used to understand claim distribution by medical condition and identify high-frequency diagnoses.

5. **Time-based Segmentation**
   Date of Service (DOS) was used to group claims by month to identify seasonal trends and monthly claim patterns.

6. **Denial and Adjustment Category Filtering**
   Claims were filtered using adjustment codes (CO, PR, Not Specified) to analyze denial patterns and financial impact.

Excel filters, PivotTable filters, and slicers were extensively used to enable interactive exploration and dynamic segmentation of the data.



9. Statistical / Formula-Based Analysis

Statistical and formula-based analysis was performed using Microsoft Excel functions and PivotTables to evaluate financial performance and claim behavior.

Key statistical measures and calculations include:

1. **Total Claims Count**
   Calculated using COUNT / COUNTA functions to determine the total number of claims processed.

2. **Total Paid Amount**
   Calculated using the SUM function to measure total reimbursement received.
   - *Formula Example:*
     =SUM(Paid_Amount)

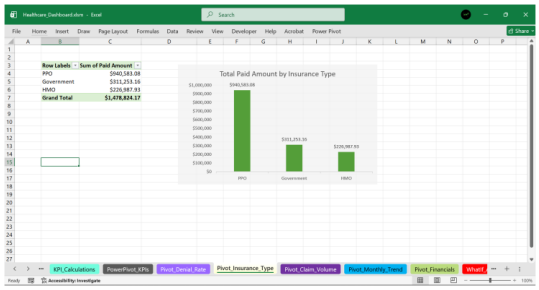
3. **Total Charge Amount**
   Used to analyze overall billed revenue.
   - *Formula Example:*
     =SUM(Charge_Amount)

4. **Total Adjustment Amount**
   Calculated to identify revenue leakage.
   - *Formula:*
     Adjustment Amount = Charge Amount – Paid Amount

5. **Denial Percentage**
   Calculated using the ratio of denied claims to total claims.
   - *Formula:*
     Denial % = (Denied Claims / Total Claims) × 100

6. **Insurance Type-wise Paid Analysis**
   PivotTables were used to calculate the sum of paid amount by insurance type (PPO, HMO, Government) to identify high-revenue segments.

These statistical calculations enabled comparison across insurance types, identification of payment gaps, and evaluation of claim processing efficiency.



10. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was conducted using PivotTables, PivotCharts, and visualizations in Microsoft Excel to identify trends, patterns, and relationships within the healthcare claims dataset.

The following analyses were performed:

1. Insurance Type-wise Analysis

PivotTables were created to analyze total paid amount, claim volume, and adjustment amount across different insurance types such as PPO, HMO, and Government plans.
This helped identify which insurance category contributes the highest revenue.

2. Claim Volume Analysis

Claim counts were analyzed by:

- Insurance Provider
- CPT Code
- ICD-10 Code

This enabled identification of high-frequency procedures and diagnoses.

3. Monthly Trend Analysis

Claims were grouped by Date of Service (Month-wise) to observe trends over time.
This analysis helped detect seasonal patterns and fluctuations in claim volumes and payments.

4. Financial Analysis

PivotTables were used to summarize:

- Total Charges
- Total Paid Amount
- Total Adjustments

These insights helped assess revenue realization and financial performance.

5. Denial Rate Analysis

Denial rates were analyzed using PivotTables based on insurance type and adjustment codes, helping to identify high-risk categories contributing to revenue loss.

6. Visualizations

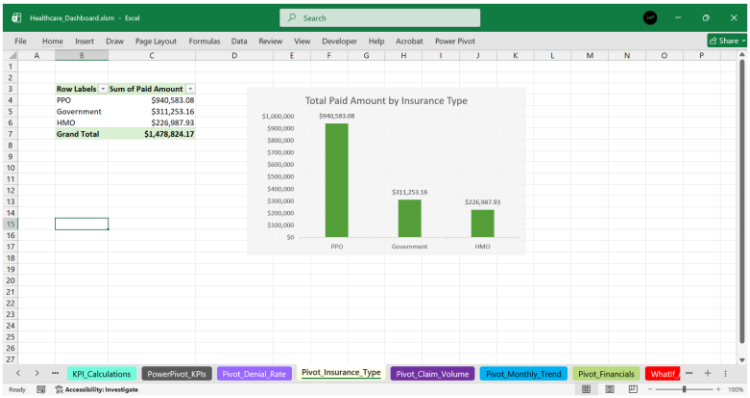
PivotCharts such as:

- Column Charts
- Bar Charts
- Line Charts

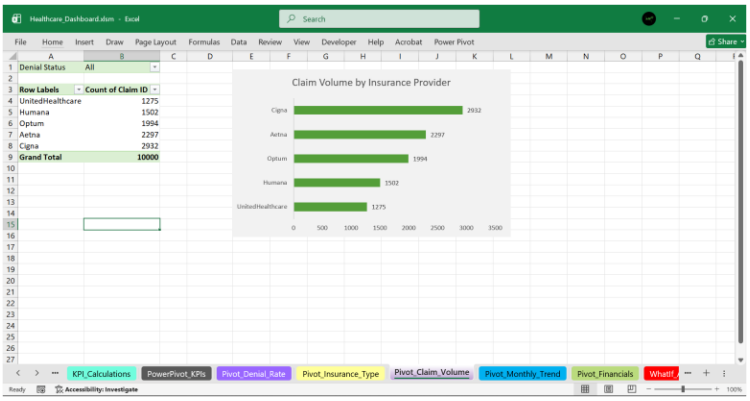
were created to visually represent trends and comparisons, making insights easy to interpret for stakeholders.

EDA enabled deeper understanding of claim behavior, revenue patterns, and operational inefficiencies within the healthcare claims process.

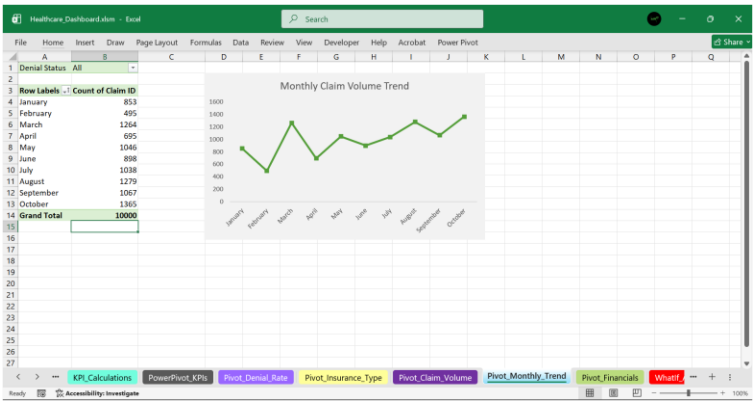
Pivot_Insurance Type



Pivot_Claim_Volume



Pivot_Monthly_Trend



11. Scenario Analysis

Scenario Analysis was performed using Excel's **What-If Analysis tools**, including **Scenario Manager** and **Goal Seek**, to evaluate the impact of changes in key variables on healthcare financial performance.

1. What-If Analysis

What-If Analysis was used to simulate different business scenarios by changing important inputs such as:

- Denial Rate
- Claim Approval Rate
- Paid Amount Percentage

Multiple scenarios were created to understand how changes in denial percentage affect:

- Total Paid Amount
- Total Adjustments
- Net Revenue

This helped management evaluate best-case, worst-case, and most-likely outcomes.

2. Scenario Manager

Scenario Manager was used to store and compare different combinations of assumptions such as:

- High Denial Scenario
- Moderate Denial Scenario
- Low Denial Scenario

Each scenario provided insight into financial risk and revenue optimization opportunities.

3. Goal Seek Analysis

Goal Seek was used to determine the required denial rate reduction needed to achieve a target paid amount.

For example, Goal Seek calculated:

- The denial percentage required to reach a desired total paid revenue target.

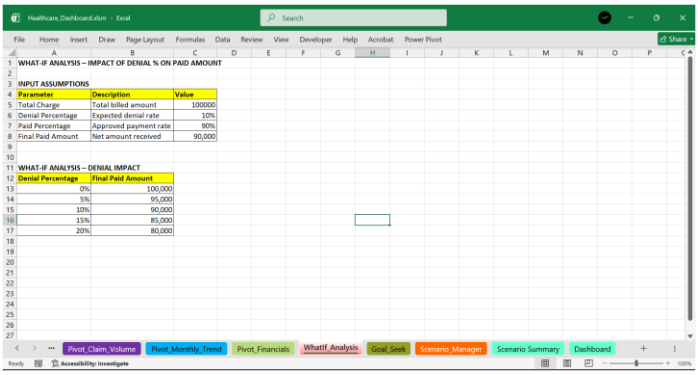
This analysis helps decision-makers set realistic performance goals and improve claims management strategies.

4. Business Impact

Scenario Analysis enables:

- Better financial forecasting
- Risk assessment
- Data-driven decision making

It supports proactive planning to reduce revenue loss and improve operational efficiency.



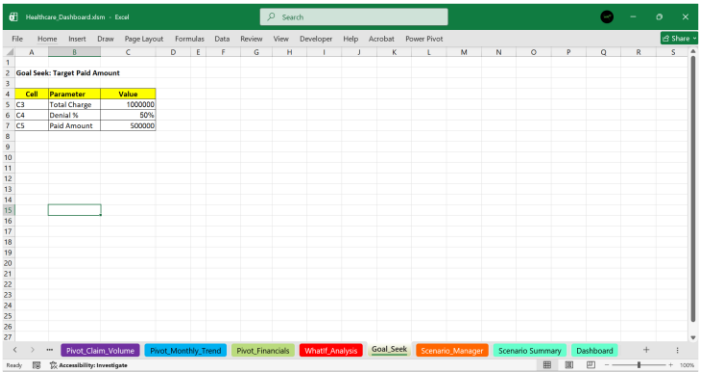
WHAT-IF ANALYSIS – IMPACT OF DENIAL % ON PAID AMOUNT

INPUT ASSUMPTIONS

Parameter	Description	Value
Total Charge	Total billed amount	100000
Denial Percentage	Expected denial rate	10%
Paid Percentage	Approved payment rate	90%
Final Paid Amount	Net amount received	90,000

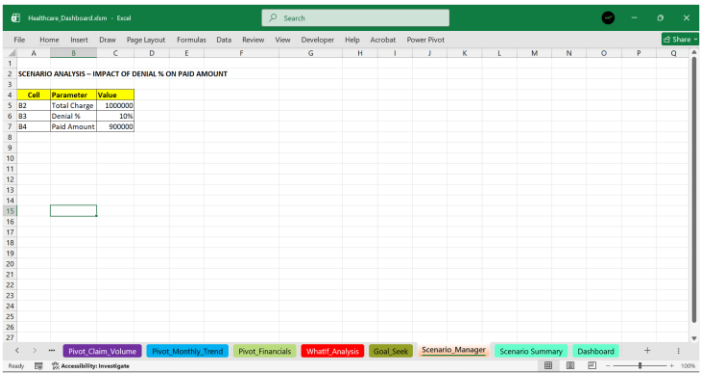
WHAT-IF ANALYSIS – DENIAL IMPACT

Denial Percentage	Final Paid Amount
0%	100,000
5%	95,000
10%	90,000
15%	85,000
20%	80,000



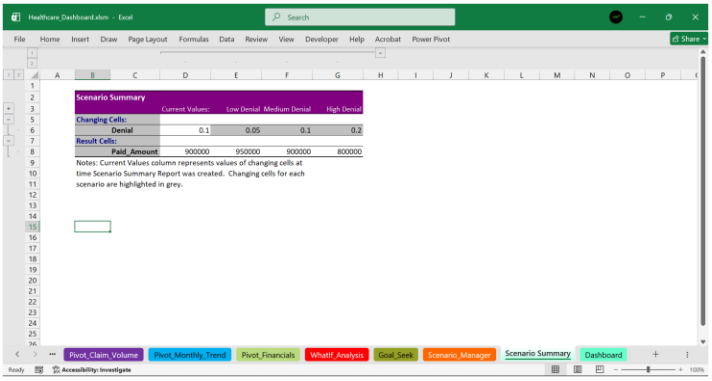
Goal Seek: Target Paid Amount

Cell	Parameter	Value
C3	Total Charge	1000000
C4	Denial %	50%
C5	Paid Amount	500000



SCENARIO ANALYSIS – IMPACT OF DENIAL % ON PAID AMOUNT

Cell	Parameter	Value
B2	Total Charge	1000000
B3	Denial %	10%
B4	Paid Amount	900000



Scenario Summary

	Current Values:	Low Denial	Medium Denial	High Denial
Changing Cells:				
Denial	0.1	0.05	0.1	0.2
Result Cells:				
Paid_Amount	900000	950000	900000	800000

Notes: Current Values column represents values of changing cells at time Scenario Summary Report was created. Changing cells for each scenario are highlighted in grey.

12. Dashboard Design & Key Insights

An interactive **Healthcare Claims Dashboard** was created using **Excel PivotTables, PivotCharts, slicers, and KPIs** to present key financial and operational metrics in a clear and visually intuitive format.

1. Dashboard Components

The dashboard includes the following elements:

- **Total Claims Count**
- **Denied Claims & Denial Percentage**
- **Total Charged Amount**
- **Total Paid Amount**
- **Total Adjustments**
- **Insurance-wise Paid Amount**
- **Monthly Claim & Revenue Trends**
- **Claim Volume by Insurance Type**

Slicers were added to allow dynamic filtering by:

- Insurance Type
- Date / Month

2. Visualizations Used

The dashboard uses multiple chart types for better insight:

- Bar charts for insurance-wise paid amount comparison
- Line charts for monthly trends
- KPI cards for quick performance tracking

These visuals help users quickly identify patterns, outliers, and performance gaps.

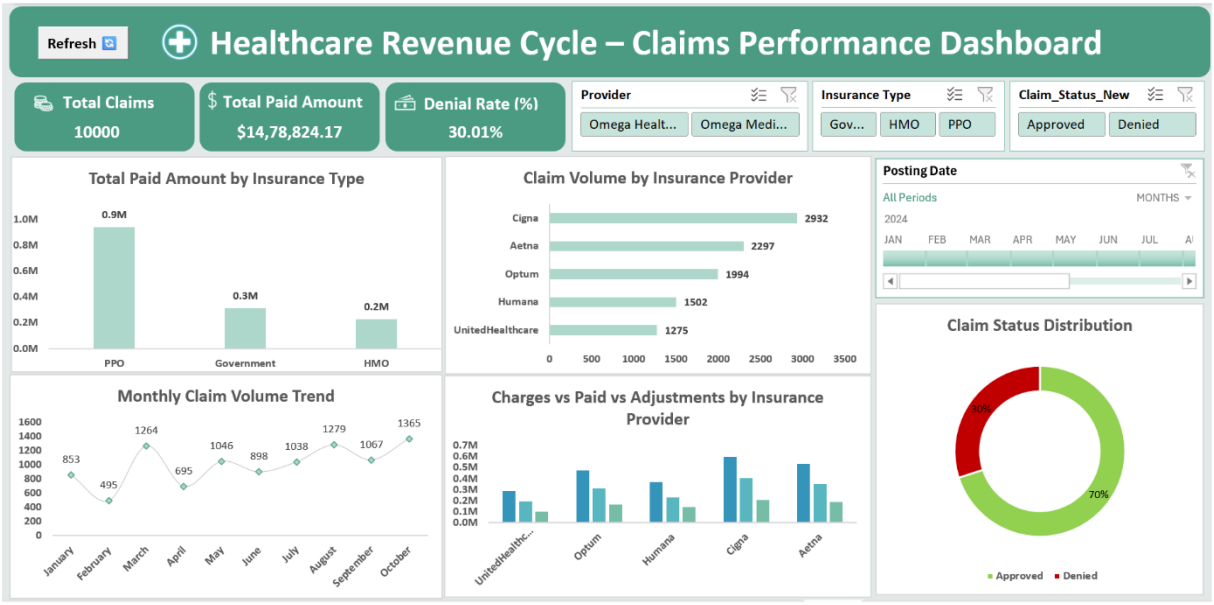
3. Key Insights from Dashboard

- PPO insurance contributes the highest paid amount compared to Government and HMO plans.
- Denial rate has a significant impact on total revenue.
- Adjustments form a major portion of revenue leakage.
- Monthly trends highlight periods of high claim volume and payment fluctuations.

4. Business Value

The dashboard supports:

- Faster decision-making
- Improved revenue cycle management
- Identification of denial reduction opportunities
- Enhanced financial transparency



Appendix A: Automation Using Macros and VBA

Purpose of Macros

- **Automatically refresh all PivotTables and PivotCharts**
- **Reset all slicers and timelines to default state**
- **Reduce manual effort and improve dashboard usability**

Macro Implemented

- **Macro Name: Reset_All_Filters_Final**

Macro Functionality

- **Refreshes all PivotTables with a single click**
- **Clears all applied slicer and timeline filters**
- **Ensures dashboards always load in a clean, consistent state**

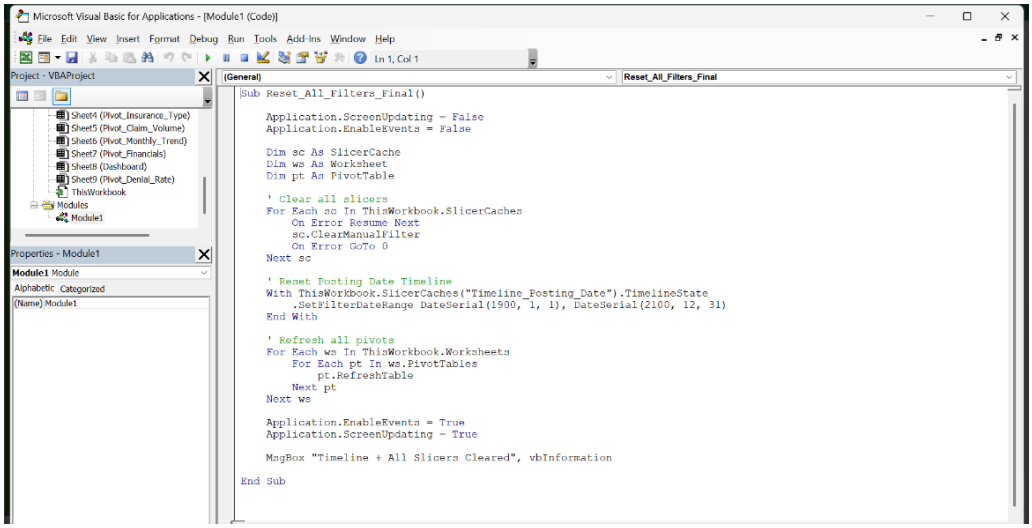
Business Value

- **Saves time during repeated analysis and reporting**

- **Minimizes manual errors**
- **Improves user experience for stakeholders**

Tools Used

- **Excel VBA (Macros)**



13. Conclusion & Future Scope

Conclusion

This project demonstrates an end-to-end **Healthcare Claims Data Analysis** workflow using **Microsoft Excel**. Starting from raw data extraction and cleaning through **Power Query**, the project applies structured data preparation, KPI calculations, PivotTable analysis, What-If Analysis, Scenario Manager, and dashboard creation to generate meaningful business insights.

The analysis helped identify key financial metrics such as claim volumes, denial rates, paid amounts, and adjustments. Interactive dashboards enabled easy monitoring of performance across insurance types and time periods, supporting data-driven decision-making and improved revenue cycle management.

Overall, this project showcases strong analytical skills, business understanding, and effective use of Excel's advanced features for real-world healthcare data analysis.

Future Scope

This project can be further enhanced in the following ways:

- Automating data refresh using Power Query with live data sources
- Integrating the dataset with **Power BI** for advanced visualizations

- Applying predictive analysis to forecast claim denials and revenue trends
- Enhancing dashboards with advanced slicers and drill-down capabilities
- Implementing VBA macros for report automation