BI Project Second Draft – Group 3
Synthetic Healthcare Dataset on Kaggle:
https://www.kaggle.com/datasets/prasad22/healthcare-dataset

1. Introduction

This BI project is a deep dive into how data analytics and dashboarding might be beneficial in the healthcare sector's operations at UnitedHealthcare. The organization is a host to a mountain of claim and patient data. Our main concern is how Business Intelligence can be of help in these initiatives; fight fraud, reduce costs, and enhance the decision-making process.

2. Company Overview

UnitedHealthcare is one of the most prominent health insurance providers in the USA, offering a diversity of plans such as individual, employer, Medicare, and Medicaid to a variety of members. The company deals with various massive datasets involving claims, billing, procedures, and outcomes and that makes it a perfect candidate for the transformation through BI.

3. What problems does the company have?

- The company is burdened with the high costs of managing claims and administrative overhead.
- The company is faced with difficulties in detecting fraudulent billing or provider abuse.
- Due to the scattering of data sources, there is uncoordinated care.
- There is a large volume of data that can be categorized as underused.
- There is no definite real-time tracking of claim trends or risk flags.

4. What are the company's overall goals?

- The company aims to reduce unnecessary costs which still has to be accompanied by care quality maintenance.
- One of the company's objectives is to utilize data in a manner that will facilitate implementation of value-based care models.
- The company has set a goal to be able to detect fraud risks and act on them earlier.
- The company intends to give the internal teams the power by means of dashboards and real-time insights.
- By consolidating data into one platform, the company wants to advance its decision-making process.

5. How could BI help solve these problems?

With the Business Intelligence tools, the company manages to:

- Create visual layers for examining claim trends and regional usage.
- Finding anomalies in the data of high-value claims to be used in fraud detection.
- Allow drill-down analytics to reveal more about procedures and payer behaviour.
- Assess the performance of the provider and thus, improve patient outcomes.
- Give quick and easy access to insights for operations teams.

6. Software Tools Used.

- Excel is mainly used in the initial formatting and basic profiling stages.
- SQL is involved in creating structured queries on cleaned tables.
- Python is used for exploratory data analysis (EDA), anomaly checks, and joins.
- Tableau serves the purpose of dashboard design and BI storytelling.

7. What are the logical components of the star schema?

Fact Table	Description
Claims Fact	Stores individual claim details, including patient ID, provider ID, service date, diagnosis and procedure codes, billed amount, and claim status.

Dimension Tables	Description
Patient Dimension	Patient demographics: age, gender, location, and insurance type.
Provider Dimension	Information on doctors, specialists, and medical facilities.
Time Dimension	Date information allowing analysis by day, month, quarter, and year.
Service Dimension	Medical procedures, diagnosis codes, and treatment details.
Payer Dimension	Insurance plan types and payer information.

8. What Else Can You Do to Help the Company Achieve Their Goals?

We are still fine-tuning this, but some ideas are:

- Employing cost driver analysis for finding costly services or providers.
- Building predictive models to detect fraudulent claims in advance.
- Implementing real-time alerting for unusually high claim activities.
- Aligning referral steps to both improve patient outcomes and reduce cost.

As we complete the visualizations, there will be more text to explain these.

9. Who Is the Consumer of BI Insights?

Initial stakeholders:
- Claims processing teams - require summaries and exception flags.
- Fraud analytics teams - require anomaly dashboards.
- Executives and finance managers - require cost-control and performance trends.
- Healthcare quality teams - require patient outcomes and procedure tracking.

We are going to clarify these roles even more when we prepare the final report.

10. What Might You Recommend to Improve the BI System? (Work in Progress)

We are likely to recommend:
- Tableau dashboards that are centralized and hosted on a server for access by multiple roles.
- Data refresh pipelines on a daily basis to keep the insights up to date.
- Staff training on the use of self-service BI for data exploration.
- Merging EHR data with other data sources for cross-system analysis.

The final remarks will be based on the visual and statistical insights.

11. Work Status

The initial data cleaning in Excel is done, and we have also formatted the main columns and tables.
At the moment, we are engaged in:
- Exploratory Data Analysis (EDA) with Python (pandas, seaborn, matplotlib) to discover main trends.
- SQL queries for dataset aggregation and filtering for visual analysis (e.g., grouping by diagnosis, summing claim amounts by region or provider).
- Tableau dashboards creation to make these data accessible and interactive.

Intended chart areas:
- Claims by Procedure
- Top Diagnoses by Region
- Total Claims by Insurance Plan
- Monthly Claims Trend
- High-Claim Providers Heatmap

The dashboards and the findings will be completed in the next weeks.

12. Presentation Plan

We will present a 15 - 20 minute video with:

- Company overview and BI context.
- Data challenges and our proposed architecture.
- Demonstrations of dashboard functionalities (Tableau).
- In-progress insights and stakeholder discussion.
- Brief Q&A prep section.

The presentation will be done by three members, and to attract the audience, we'll employ simple visuals.

13. Conclusion

Our project continues to move forward. We have designed the schema, started cleaning and exploring the data, and connected business challenges with BI tools. We will be finalizing our dashboards, refining our recommendations, and completing our report and presentation during the next weeks.