HealthCare Data Analyzer

# Team Members

Kriti Kankaria

Rukhaiah Shaik

[Soundarya Bachu](mailto:soundaryabachu@oakland.edu)

# Executive Summary of the Final Project

Our project outlines our goal of utilizing a healthcare dataset to gain insights into patients' histories, including their demographics, medical procedures, diagnoses, and associated costs. Through the analysis of this data, we aim to provide valuable insights for healthcare providers and policymakers, enabling them to make informed decisions that can improve healthcare delivery and cost-effectiveness.

Our project will also employ descriptive analytics techniques to uncover patterns and trends in the data. By doing so, we aim to not only understand the current state of healthcare but also to identify areas for improvement and optimization. The insights generated from our analysis have the potential to drive meaningful change in healthcare practices, ultimately leading to better patient outcomes and more efficient use of resources.

Overall, our executive summary highlights the importance of data-driven decision-making in the healthcare industry and underscores our commitment to using analytics to make a positive impact. We are excited about the potential of this project and look forward to the opportunities it presents for improving healthcare delivery and outcomes.

# Target Users or Analysis Consumers

The target users we aim to engage with for this project include healthcare providers, administrators, and insurance companies. We focus on these groups because healthcare providers are crucial for managing facilities, ensuring efficient operations, and directly interacting with patients to deliver medical services. Understanding patients' backgrounds enables providers to deliver tailored care and services, enhancing relevance and effectiveness.

Additionally, we target insurance companies because they assume risks and establish coverage policies. By utilizing our dataset, we aim to assist insurance companies in determining the coverage required. This dataset will help insurance companies in setting premiums, managing financial risks, and creating policies that meet patient needs. These policies will ensure patients receive value-based care while complying with regulations such as HIPAA and ACA.

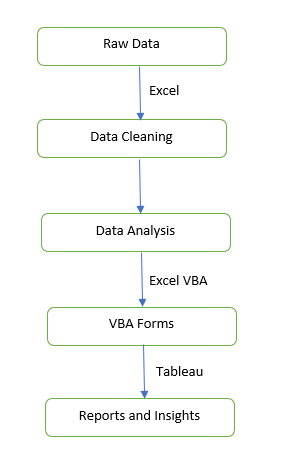
Our project aims to provide valuable insights to healthcare providers, administrators, and insurance companies, enabling them to enhance patient care, optimize operations, and create policies that benefit both patients and the healthcare industry as a whole.

# Technical Summary

For our project, we envision employing a dynamic blend of software tools and techniques that may evolve as we progress. Initially, we will utilize Excel for its versatility in data cleaning, manipulation, and basic analysis, leveraging its familiar interface and robust functionality. As we advance, we plan to incorporate Tableau for more advanced data visualization and dashboarding, enabling us to create interactive and insightful representations of our findings.

Our approach emphasizes adaptability, with a focus on tailoring our software tools and techniques to meet the evolving needs of the project and our growing skill set. This iterative process ensures that we leverage the most effective tools and techniques at each stage of our analysis.

Here is a diagram illustrating the overall architecture of our project:



The diagram elegantly depicts the comprehensive architecture of our project. Initially, we utilize Excel for data cleaning and basic analysis, leveraging VBA Code to implement advanced filtering techniques learned in class. The cleaned data is then seamlessly transferred into Tableau for advanced visualization and dashboard creation. This integrated approach optimizes the strengths of each tool, amplifying our capability to extract meaningful insights from the data.

By starting with Excel for tasks such as data cleaning and basic analysis, we lay a solid foundation for our analysis. Excel's familiar interface and feature-rich environment enable us to gain a comprehensive understanding of the data. As we progress, Tableau will play a pivotal role in transforming our insights into visually compelling and interactive visualizations and Tableau sheets. This approach not only enhances the clarity and impact of our findings but also facilitates a deeper understanding of the data among our audience.

# Data Needs and Sources

Our primary dataset is from Kaggle, a rich repository of providing insights into the patient’s demographics, medical conditions, hospitals, doctors, and associated costs. However, recognizing the multifaceted nature of healthcare research, we acknowledge the need of supporting our data analysis with additional data sources. This dataset will encompass the socio-economic factors, environmental factors, healthcare access metrics, patient outcomes data, and provider characteristics.

In terms of the primary dataset's overview, we have access to 6 years of healthcare data comprising 15 columns and 10,000 records. This dataset serves as a foundational element for our analysis, providing a solid base from which to build our insights.

By broadening our scope to include these diverse dimensions, we aim to cultivate a more holistic understanding of the complex interplay shaping healthcare dynamics. This expansive approach not only enriches our analysis but also enables us to unravel relationships and uncover hidden patterns that might otherwise remain obscured. Therefore, while the Kaggle dataset serves as our primary foundation, our commitment to exploring and integrating additional data sources underscores our dedication to thorough and comprehensive search.

# Outputs

Our project has successfully delivered an efficient data exploration tool and empowered users to make informed decisions. Central to this achievement is the development of a VBA Excel program, which offers a range of functions to enhance data analysis. The first form we designed enables users to select specific medical conditions, such as diabetes, prompting the VBA program to display all patient records related to that condition. This form not only filters the data based on the selected condition but also provides insights into patient demographics, facilitating a deeper understanding of the data.

In addition to the medical conditions form, we also created a form for selecting admission types. This form allows users to explore different types of admissions, providing valuable insights from the perspective of the hospital. It also aids in calculating the average cost of treatment for different admission types, which is crucial for decision-making processes. Furthermore, we developed a form specifically tailored for insurance providers, named "Select Insurance Providers." This form enables users to select insurance policies and determine if a hospital supports those policies. It also provides information on average billing costs, assisting insurance providers in making well-informed decisions.

These tools and forms enhance data exploration and decision-making capabilities, ultimately benefiting healthcare providers, administrators, and insurance companies alike.

Another significant outcome of our project is the implementation of data visualization techniques using Tableau, which has enabled us to present data in a highly interactive and insightful manner. By leveraging Tableau, we have created Tableau sheets that enhance the user's ability to comprehend complex datasets comprehensively and intuitively. Through a variety of graphs and charts, these Tableau sheets facilitate the communication of key insights and trends, thereby aiding in understanding and decision-making processes.

Moreover, our visualization approach incorporates a high level of interactivity, allowing users to explore the data dynamically. For example, one of our Tableau sheets, the "Number of Patients - Age" report, provides a detailed breakdown of the number of patients admitted at different ages. This dashboard reveals significant insights, such as the average age of patients being 59 years old, which is crucial information for healthcare providers and policymakers.

Additionally, we have developed another report focused on "Test Results," which illustrates the probability of obtaining abnormal results. This report is particularly valuable as it provides a percentage value indicating the likelihood of abnormal test outcomes. Such insights are instrumental in understanding patient health trends and can guide decision-making processes related to healthcare services and treatments.

Our use of Tableau for data visualization has significantly enhanced the presentation of our findings, making complex healthcare data more accessible and actionable for stakeholders.

# Benefits to Target Audience

This project promises several benefits for its target audience, particularly healthcare providers. One key advantage is optimized resource allocation within healthcare facilities. By leveraging the dataset, providers can identify high-demand services, allocate staff accordingly, and streamline processes, leading to smoother operations.

Another significant benefit is the enhancement of patient care. The dataset enables providers to gain a deeper understanding of patients' needs and tailor care plans accordingly. By identifying existing health conditions and anticipating patient requirements, providers can deliver personalized care that is more effective and meaningful.

Additionally, the project offers benefits in terms of cost containment. Through thorough analysis of patterns and associated costs, providers can identify opportunities to save money and improve efficiency. This analysis enables the implementation of strategies to reduce unnecessary procedures, minimize waste, and optimize resources, leading to cost savings without compromising the quality of care.

Insurance companies stand to gain several benefits from the dataset, enhancing their ability to conduct risk assessments and management. One significant advantage is the dataset's utility in helping companies identify and manage risks more effectively. By analyzing patient histories, insurers can pinpoint areas of risk and adjust policies to mitigate financial exposure.

Moreover, the dataset enables actuarial analysis, allowing insurance companies to develop precise pricing models and set premiums based on the costs associated with different groups and medical conditions. This level of analysis ensures that premiums accurately reflect the expected costs of providing coverage, leading to more fair and sustainable pricing structures.

Additionally, the dataset can be used to detect and prevent healthcare fraud. By analyzing patterns in the data, insurers can identify suspicious activities and implement measures to prevent fraud. This proactive approach helps insurance companies protect against financial losses and maintain the integrity of their operations.

In conclusion, this project holds immense promise for healthcare providers, offering tangible benefits that can enhance resource allocation, improve patient care, and contain costs, ultimately leading to more efficient and effective healthcare delivery and the dataset offers insurance companies valuable tools for conducting risk assessments, developing pricing models, and combating fraud, ultimately leading to more efficient and effective operations.

# Challenges

Throughout this project, we encountered several challenges, some of which were unexpected. One of the initial hurdles we faced was data cleaning in Excel. Cleaning the data required meticulous attention to detail and problem-solving skills. We had to identify and rectify inconsistencies, outliers, and missing values present in the dataset. Despite the majority of the data coming from a single source, Kaggle, we had to remain vigilant, as data anomalies demand a keen eye for detail and a methodical approach to error detection and correction.

After completing the data cleaning process in Excel, we encountered challenges with the VBA program. While we anticipated that programming in VBA would be challenging, the extent of the difficulty caught us off guard. One specific challenge arose when coding for the admission type button. Storing the admission type data and ensuring that the code worked seamlessly for all buttons proved to be particularly challenging. We iteratively wrote and debugged the code, using the debugging tool to identify and fix errors. This iterative process required careful attention to detail and a methodical approach to ensure the functionality of the program.

The next challenge we encountered was creating the Tableau sheets in Tableau. While we had successfully uploaded the data into Tableau, we struggled with determining how to effectively convey the insights from the dataset. This challenge caught us off guard, as we initially believed that preparing reports and visualizations would be straightforward. However, as we delved deeper into the project, we realized the complexity of effectively presenting the data.

After extensive discussion over a Zoom call, we decided to showcase various aspects of the data, ranging from patients' histories to the billing amounts received by insurance companies. With such a wide array of data points, selecting a starting point proved challenging. Ultimately, we chose to begin with the patients' information and then branch out from there.

Starting with the patients' histories, we expanded the Tableau sheets to include information such as the average length of stay, the type of admission, admission dates, and discharge dates. We also included details about the patients' medical conditions. For example, we created a dashboard focusing on medical conditions from 2018 to 2023. This dashboard provided insights into the prevalence of various conditions and the number of admissions associated with each condition, helping us understand trends over time.

# Personal Learning

Throughout this project, we have gained valuable insights and learned several key lessons. One of the most significant lessons was our deepened understanding of Tableau. As we created Tableau sheets, we explored its various functions and capabilities, enhancing our skills in data visualization and storytelling. A crucial takeaway was the importance of simplicity in visualizations. While Tableau offers numerous features for creating visually appealing sheets, we learned that clear, concise visualizations are more impactful than cluttered ones.

Another lesson we learned was the significance of understanding our audience. We realized the importance of tailoring visualizations to meet the specific needs of our stakeholders, ensuring that our insights resonate and drive action. Additionally, we learned the value of continuous learning. Working with Tableau has shown us that there is always something new to discover and learn. By staying curious and actively seeking out opportunities to expand our skills, we have been able to create more visually appealing Tableau sheets.

In conclusion, this project has not only enhanced our Tableau skills but has also taught us valuable lessons in data visualization, audience targeting, and continuous learning. We look forward to applying these lessons in future projects to create even more impactful and insightful visualizations.

# Closing Thoughts

This project has been a rewarding journey, enabling us to deepen our skills and knowledge in the healthcare domain. Leveraging our expertise in Tableau and Excel, we successfully created new buttons to extract meaningful insights from the data and developed interactive Tableau sheets. This hands-on experience has not only enhanced our technical skills but also enriched our understanding of how data can drive decision-making in healthcare.

As we reflect on the project, we are proud of the results we have achieved through our hard work and dedication. This project has not only expanded our skill set but also reaffirmed our passion for using data analytics to make a positive impact. We look forward to applying these skills in future projects and continuing our journey of learning and growth in the field of data analytics.