5/8/2025

Risk Stratification Dashboard with Streamlit

MGS 670: Healthcare Analytics

University at Buffalo
The State University of New York

Under the Guidance of:

Prof. Amin Serehali

(Senior Vice President - Chief Data & Analytics Officer at Independent Health)

Team 1

NADUVINKERI, SAGAR - 50592582
HOSKOTE VISHWATA, SANDHYA - 50571138
YADAV, VIBHOR KUMAR - 50602745
KOGILU NAGENDRA KUMAR, NAGARJUNA - 50559728
KUMBAR, PRATEEK LOKESH - 50559787

1. Introduction

The Risk Stratification Dashboard is an online application developed with Streamlit to support dynamic risk stratification of health members. The dashboard supports healthcare providers and care managers in detecting high-risk groups, analyzing risk factors, and making data-driven decisions to improve patient outcomes and reduce healthcare costs. Risk stratification is a key function of modern healthcare management that supports organizations in optimizing resource allocation, reducing avoidable hospitalization, and improving patient care quality.

Healthcare organizations are faced with the challenge of dealing with complex patient groups with differing medical needs. Sound risk stratification is imperative in enabling anticipatory care planning, intensive interventions, and efficient resource use. This dashboard addresses these challenges by providing a sound platform for the identification of high-risk members, analysis of clinical measures, and monitoring of cost trends, ultimately contributing to improved health outcomes and financial performance.

2. Project Objectives

The primary objectives of the Risk Stratification Dashboard are:

- Provide an interactive risk stratification platform for health members.
- Enable filtering, visualization, and export of risk segments.
- Enable simple-to-use visual components such as bar charts, risk score tables, and trend analysis.
- Support care management decision-making with intuitive data presentation.
- Enable proactive patient management by identifying high-risk members and identifying cost drivers.
- Enhance population health management by integrating multiple data sources and real-time analytics.

3. Data Sources and Preparation

The data used in this project includes member data, chronic disease counts, clinical metrics, financial costs, and prospective risk scores. The key data sources are:

- **risk.csv**: Contains member-level risk stratification data, including age, gender, chronic condition count, supporting data for chronic conditions, clinical measures, healthcare utilization metrics and prospective risk scores.

3.1 Data Loading and Preprocessing

- Data is loaded using the pandas library to facilitate quick data manipulation and analysis.
- Handling missing values, data type conversions, and outlier filtering are performed to offer uniform data.
- Key columns include Age, Gender, Current Risk Level, Count of Chronic Disease, Prospective Total Risk, and Allowed Medical Costs.
- Additional feature engineering is performed to construct meaningful variables such as risk segments, chronic disease counts, and cost measures.
- Scaling and normalization of data are employed to enhance performance of visualizations and statistical operations.

4. Dashboard Features

4.1 User Interface Design

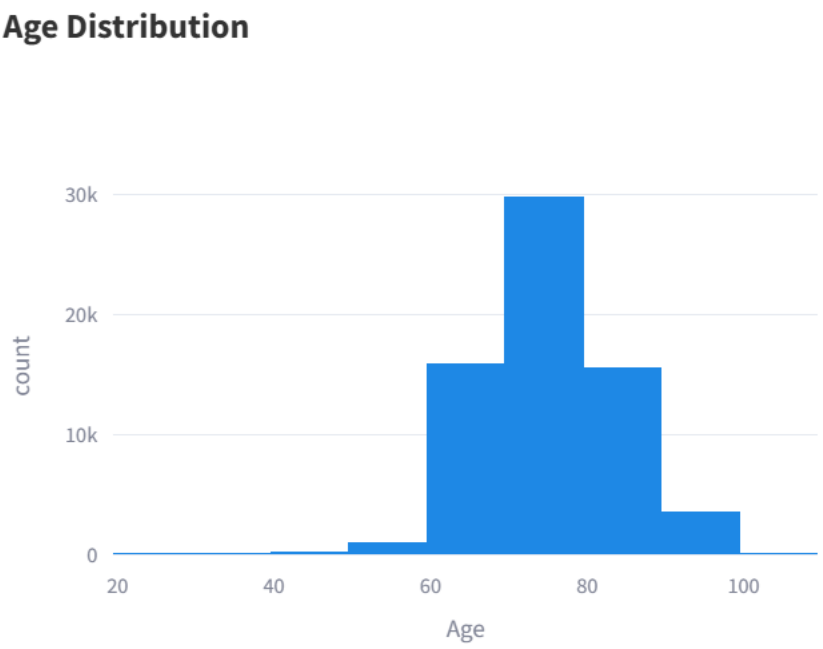
- Personalized styling of headers and sections for a professional appearance
- Friendly user interfaces with numerous tabs for diverse analyses, including Population Overview, Clinical Metrics, Financial Analysis, and Member-Level Analysis.
- Responsive design to accommodate different screen sizes and devices.

4.2 Key Functionalities

- **Dashboard Overview** - Provides an overview of the overall member population, including average age, number of chronic conditions, average potential risk, and increasing risk percentage.
- **Filters Panel** - Allows users to refine their analysis by selecting specific member attributes like risk level, age range, gender, and chronic disease count.

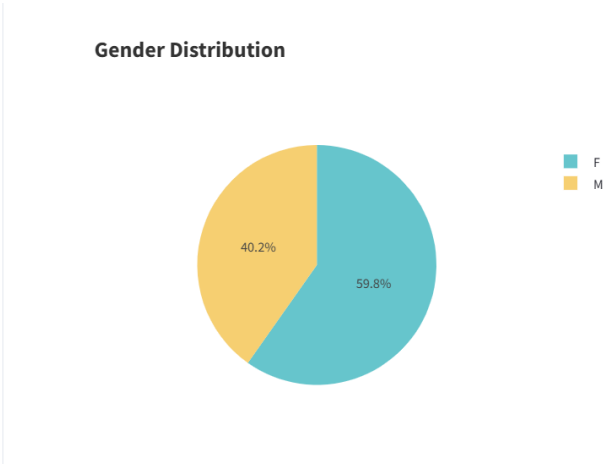
5. Population Overview

5.1 Age Distribution



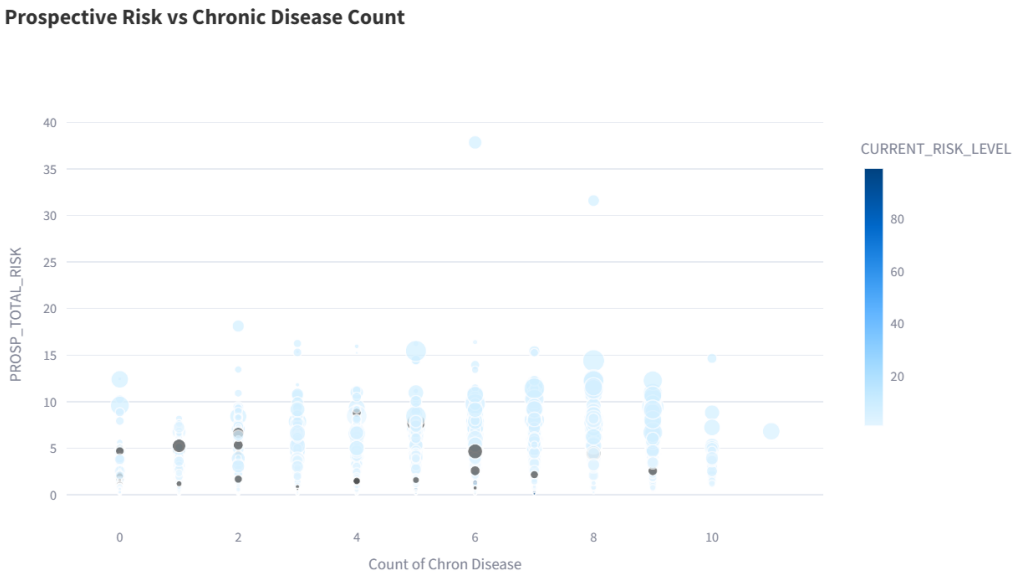
This histogram shows the distribution of the age of members. It shows the concentration of members in some age ranges, indicating the age composition of the population. Most of the members are concentrated between the ages of 60 and 80, reflecting the concentration of age characteristic of a healthcare plan for elderly people.

5.2 Gender Distribution



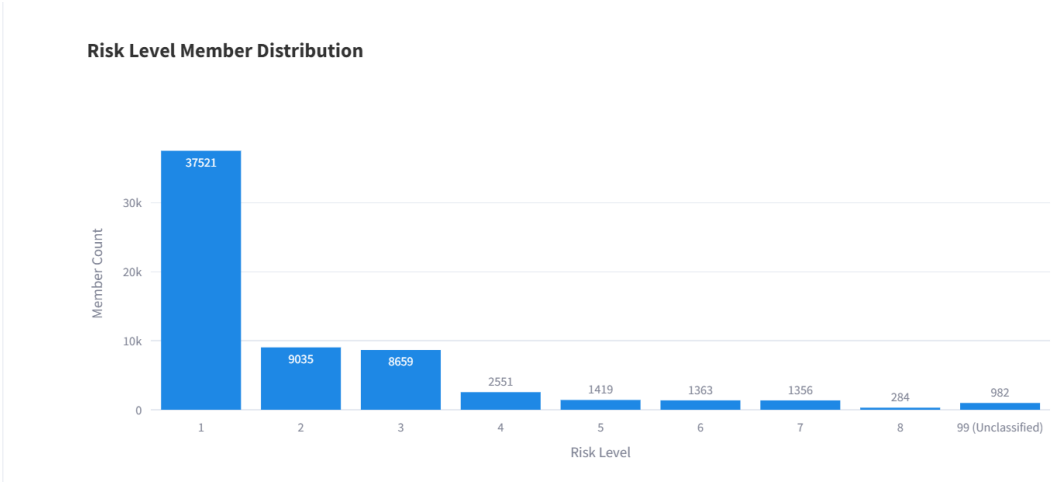
The following pie chart indicates the gender distribution of the members' population. The records show a gender imbalance with female and male membership being close to 59.8% and 40.2%, respectively, and offering potential for gender-related health needs and risks.

5.3 Prospective Risk vs Chronic Disease Count



This scatter plot illustrates the relationship between the number of chronic diseases and future risk scores. It highlights that members with more chronic diseases have higher future risk scores, emphasizing the importance of managing multiple chronic diseases to reduce overall healthcare risk.

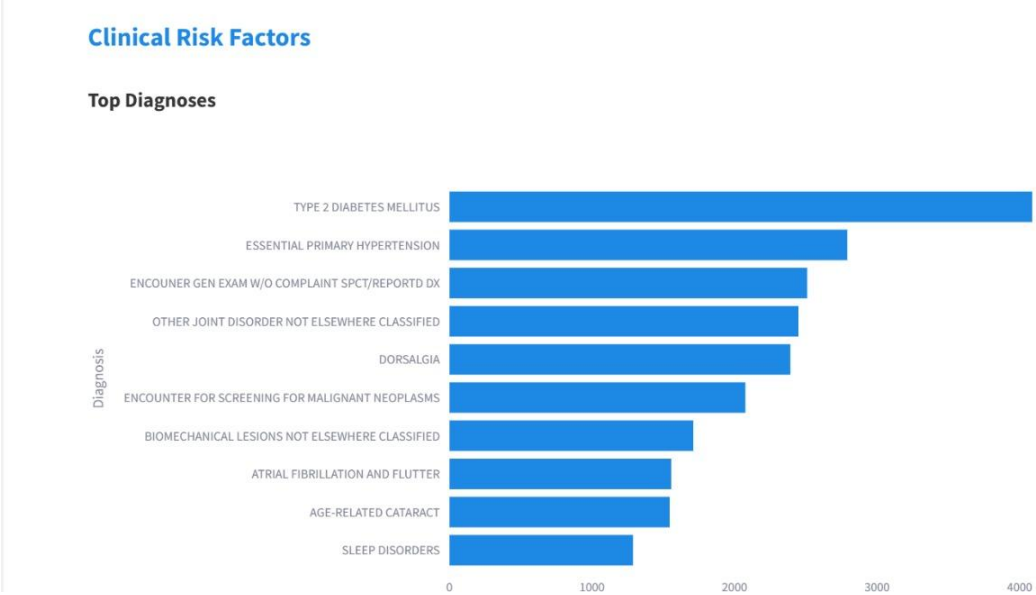
5.4 Risk Level Member Distribution



This bar chart shows the distribution of members by risk levels. It clearly shows that most members are in the lower risk groups, with a sharp decline in member numbers as the risk level goes up. This information is vital for targeting high-risk groups for proactive care management.

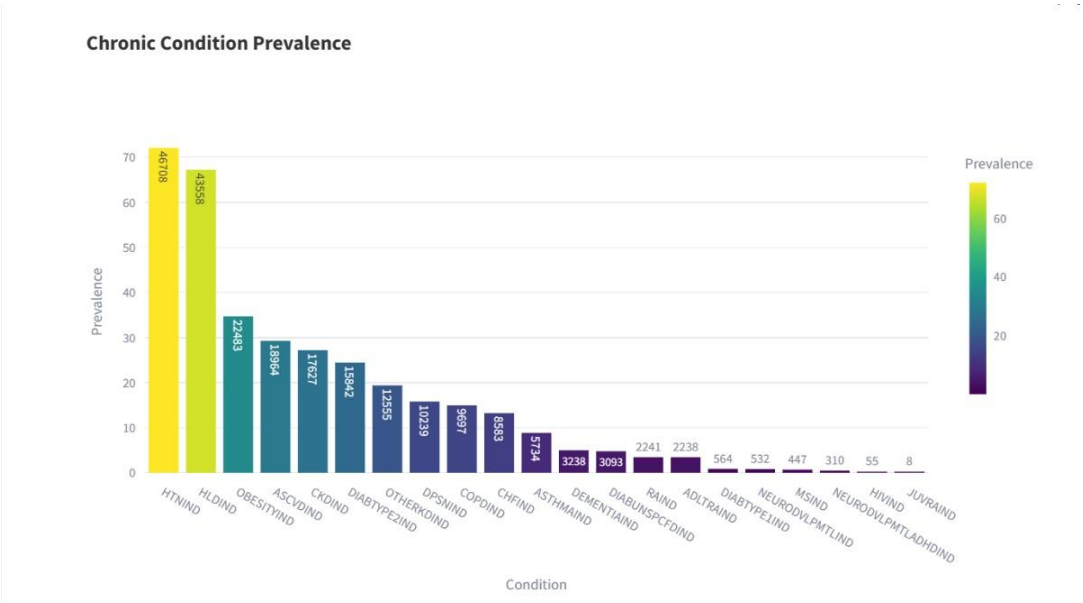
6. Clinical Metrics

6.1 Top Diagnoses



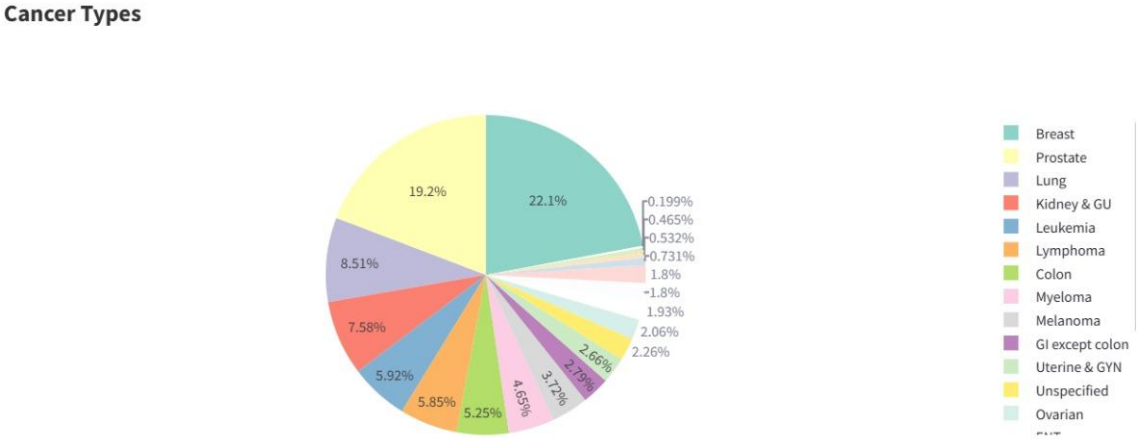
The "Top Diagnoses" bar chart provides an overview of the most common primary diagnoses within the member population. The chart shows that Type 2 Diabetes Mellitus is the most common condition, followed closely by Essential Primary Hypertension. These two conditions alone account for a significant proportion of the chronic disease burden in the population, reflecting the widespread impact of metabolic and cardiovascular disease. The other notable diagnoses include Joint Disorders, Dorsalgia (Back Pain), and Atrial Fibrillation, all of which are significant drivers of healthcare utilization and costs. These common diagnoses guide healthcare organizations in prioritizing interventions and resources accordingly.

6.2 Chronic Condition Prevalence



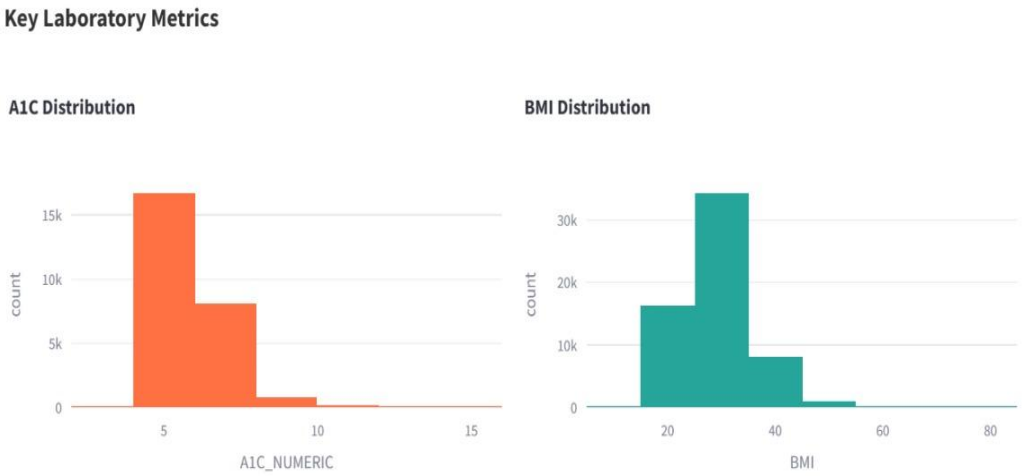
The "Prevalence of Chronic Conditions" bar chart also provides an overall picture of the most prevalent chronic conditions among members. The top chronic conditions are Hypertension (HTNIND) and Hyperlipidemia (HLDIND) with 46,708 and 43,558 members, respectively. Others are Obesity (OBESITYIND), Atherosclerotic Cardiovascular Disease (ASCVDIND), and chronic kidney disease (CKDIND). Prevalence of chronic conditions highlights the imperative necessity of ongoing management of cardiovascular risk factors, which are major determinants of long-term healthcare costs and adverse health outcomes.

6.3 Cancer Types



The "Cancer Types" pie chart provides a close-up look at the distribution of various cancer diagnoses within the member population. **Breast Cancer** is the most common type, accounting for **22.1%** of all cancer diagnoses, followed by **Prostate Cancer (19.2%)**, **Lung Cancer (8.51%)**, and **Kidney & GU (7.58%)** cancers. The chart also presents a wide range of less common cancers, which reflect the diverse oncology needs in the population. Such a distribution is valuable in structuring targeted oncology treatment and deployment of resources.

6.4 Key Laboratory Metrics



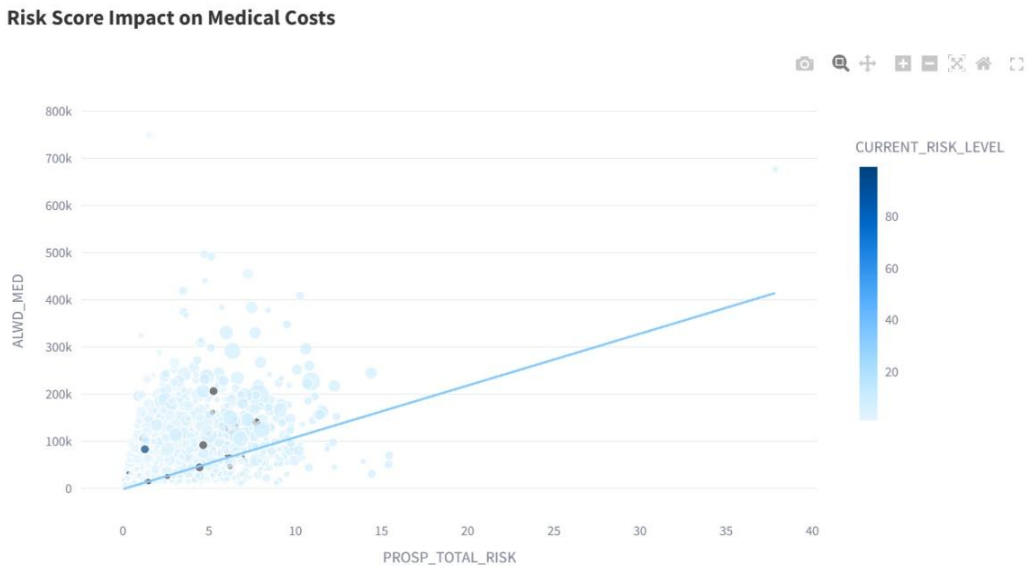
This page includes two of the most important laboratory measures, **A1C Distribution** and **BMI Distribution**, which are important markers of long-term health risk:

- **A1C Distribution**: The graph illustrates that most members have an A1C of less than 7.0, consistent with ideal targets for diabetic control. Yet many members have elevated A1C levels, indicating uncontrolled diabetes.

- **BMI Distribution**: The distribution chart of BMI reflects that the greater part of the population is represented in the obese and overweight sections, reflecting an important risk factor for a range of chronic disorders, including cardiovascular disease, diabetes, and hypertension.

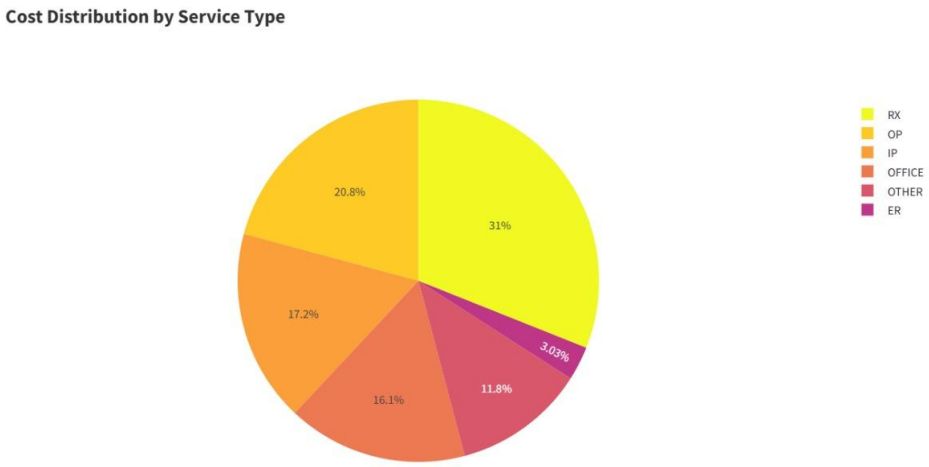
7. Financial Metrics

7.1 Risk Score Impact on Medical Costs



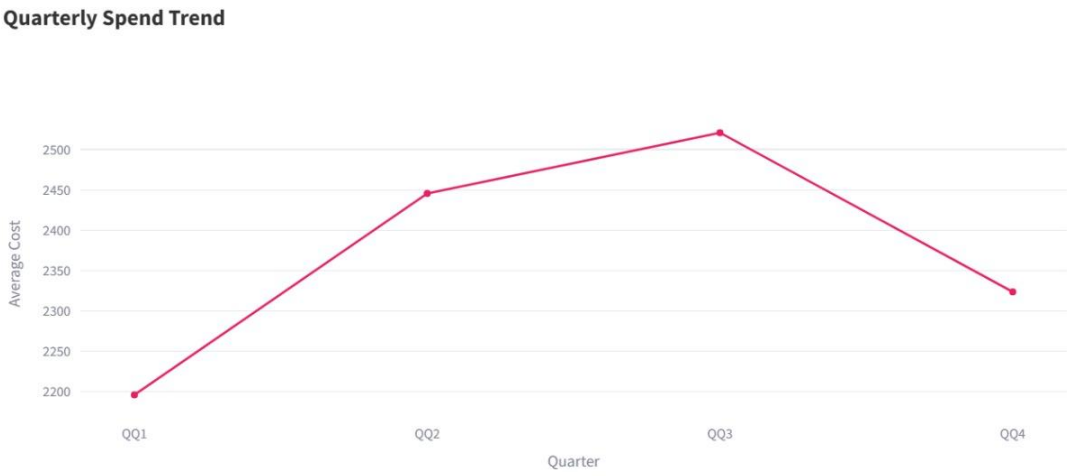
The "**Risk Score Impact on Medical Costs**" scatter plot nicely shows the relationship between a member's prospective total risk and total medical costs. The dots represent individual members, with the x-axis being the prospective risk score and the y-axis being the total allowed medical costs. The trend line reveals a positive correlation, which shows that individuals with rising risk scores have considerably more medical expenses. This pattern is relevant to health providers as it demonstrates the expense consequence of escalating risk levels and validates the need for targeting interventions to lower costs.

7.2 Cost Distribution by Service Type



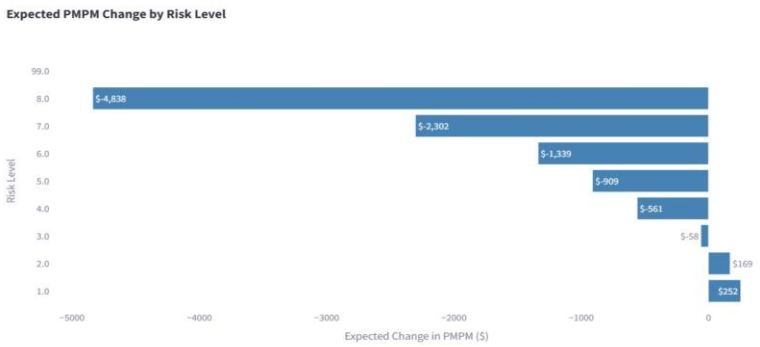
The "Cost Distribution by Service Type" pie chart divides healthcare expenditure into different service types. The pie chart indicates that most of the healthcare expenses are due to **RX** (31%), **OP** (20.8%), and **IP** (17.2%) services, with smaller contributions from **OFFICE** (16.1%), **OTHER** (11.8%), and **ER** (3.03%). This distribution emphasizes the significant contribution of pharmacy and outpatient services to overall healthcare expenditure, where providers need to focus on cost reduction in these high-expense areas.

7.3 Quarterly Spend Trend



This line chart tracks four quarters' average healthcare spending, revealing seasonal or periodic trends in expenditures. The graph indicates a peak in Q3, which could reflect higher healthcare use during this quarter. By identifying these trends, healthcare organizations can rationalize their budgeting and resource allocation, linking financial planning to seasonal demand.

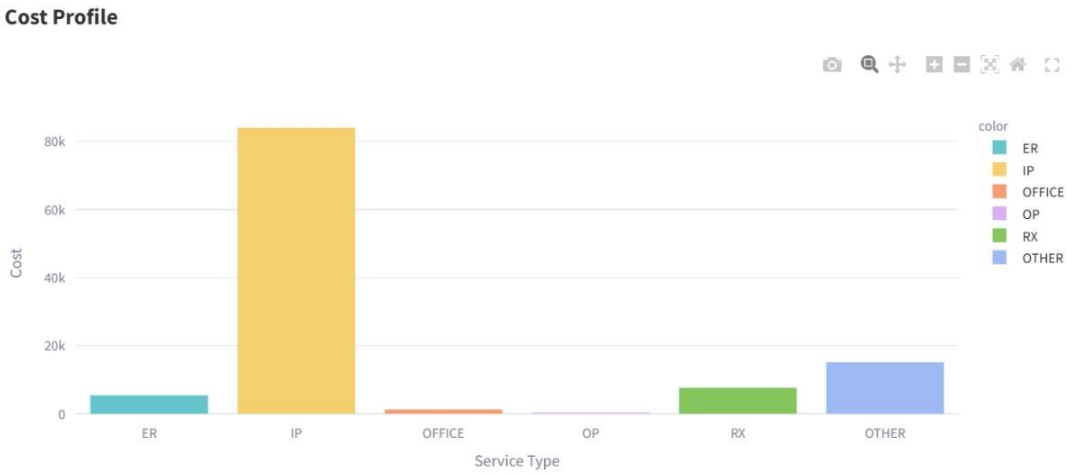
7.4 Expected PMPM Change by Risk Level



The bar graph of "**Expected PMPM Change by Risk Level**" presents the projected change in per-member-per-month (PMPM) costs by risk levels. It reflects the considerable cost savings possible for higher-risk members, with **Risk Level 8** experiencing the highest PMPM decrease of **$4,838**. The data can be used to target care management at high-risk populations, which would lead to significant cost savings.

8. Member Details (Enhanced)

8.1 Cost Profile

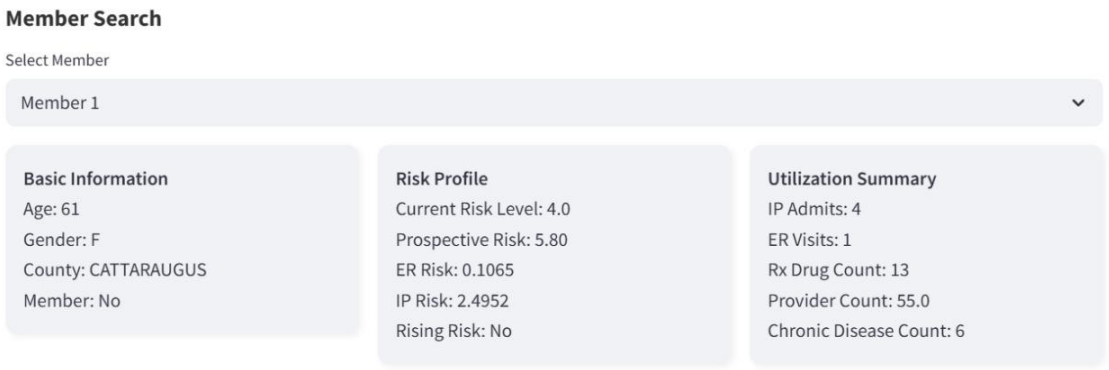


This bar chart shows the Cost Profile of the selected member by type of healthcare services consumed. It categorizes the total healthcare expenditure into different categories of services such as:

- **ER (Emergency Room)**
- **IP (Inpatient)**
- **Office (Office Visits)**
- **OP (Outpatient Services)**
- **RX (Prescription Drugs)**
- **OTHER (Miscellaneous Services)**

In this example, **Inpatient (IP)** spending is the dominant type, indicating that most of the this member's healthcare expenditure is in the form of hospitalization, which can significantly affect overall healthcare spending.

8.2 Basic Information, Risk Profile, and Utilization Summary



The first snapshot captures a snapshot of the Basic Information, Risk Profile, and Utilization Summary of a selected healthcare member. The snapshot provides a general overview of the members' demographic and healthcare utilization data, including:

Basic Information: In this category, the member's age (61), sex (Female), and county (**CATTARAUGUS**) are stated, as well as whether the member is an active member at present, with this illustration labeled as "No.".

Risk Profile: Includes key risk metrics such as **Current Risk Level** (e.g., 4.0), **Prospective Risk** (e.g., 5.80), **ER Risk** (e.g., 0.1065), **IP Risk** (e.g., 2.4952), and whether the member is a Rising Risk case (No). These metrics are critical to predicting future healthcare spending and utilization patterns.

Utilization Summary: This segment gives the summary of the members' healthcare events, i.e., **IP Admits** (e.g., 4), **ER Visits** (e.g., 1), **Rx Drug Count** (e.g., 13), **Provider Count** (e.g., 55.0), and **Chronic Disease Count** (e.g., 6). These metrics suggest how many and how intense the healthcare services the member has consumed.

8.3 Clinical Profile and Lab Results

The second snapshot focuses on the **Clinical Profile** and **Lab Results** of the selected members, providing a detailed view of their primary health conditions and recent lab findings:

- **Primary Diagnosis:** Displays the diagnosis code (e.g., I63), description of diagnosis (e.g., CEREBRAL INFARCTION), and critical health indicators such as **Behavioral Health SPMI** (e.g., Yes) and **Cancer Active** status (e.g., No). This information is crucial for understanding the primary medical challenges faced by the members.
- **Lab Results:** Shows key laboratory results that can indicate overall health status, including **BMI** (e.g., 39.7), **A1C** (e.g., 6.6), **GFR** (e.g., 42.2), and **Creatinine** (e.g., 1.42). These values help assess the risk of chronic conditions like diabetes, kidney disease, and cardiovascular problems.

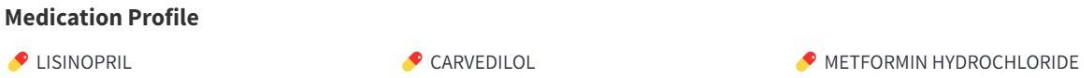
8.4 Chronic Conditions

This section lists the **Chronic Conditions** that have been identified for the selected member. It includes multiple chronic conditions such as:

- **ASCVDIND (Atherosclerotic Cardiovascular Disease)**
- **COPDIND (Chronic Obstructive Pulmonary Disease)**
- **DPSNIND (Diabetic Peripheral Neuropathy)**
- **DIABTYPE2IND (Type 2 Diabetes)**
- **HTNIND (Hypertension)**
- **OBESITYIND (Obesity)**
- **OTHERKIND (Other Kidney Disorders)**

This concise list provides a quick overview of the long-term health challenges that may require ongoing management and monitoring.

8.5 Medication Profile



This image highlights the **Medication Profile** for the selected member, listing the primary medications currently being used to manage chronic conditions. Examples include:

- **Lisinopril** - Commonly used to treat hypertension and heart failure.
- **Carvedilol** - Used for managing high blood pressure and heart failure.
- **Metformin Hydrochloride** - Typically prescribed for managing Type 2 Diabetes.

Understanding the medication profile is essential for coordinating care, preventing adverse drug interactions, and managing chronic conditions effectively.

9. Data Analysis and Insights

Data analysis is the core component of the Risk Stratification Dashboard, offering actionable information on the member population's health status, risk levels, and financial contribution. This chapter presents the major findings from the clinical, financial, and demographic data, emphasizing the most notable patterns and correlations.

9.1 Population Health Insights

- **Age Distribution**: Most of the population are between 60 and 80 years of age, which aligns with the high prevalence of chronic conditions characteristic among the elderly, such as hypertension, diabetes, and diseases of the joints.
- **Gender Distribution**: There is a higher proportion of women (59.8%) to men (40.2%), which reflects broader trends in healthcare use, as women live longer and thus have higher lifetime healthcare needs.
- **Prospective Risk Correlation**: The scatter plot of **Prospective Risk vs Chronic Disease Count** indicates a **positive** correlation, with members who have more chronic diseases tending to have higher prospective risk scores. This illustrates the compounding effect of multiple chronic diseases on health risk.

9.2 Clinical Insights

- **Top Diagnoses**: The prevalence of **Type 2 Diabetes Mellitus** and **Essential Primary Hypertension** as the most common diagnoses suggests that metabolic and cardiovascular health should be primary focuses for care management programs.
- **Chronic Condition Prevalence**: The most common chronic conditions are hypertension and hyperlipidemia, reflecting the importance of cardiovascular disease management in this population.
- **Cancer Distribution**: The prevalence of cancer types, with breast and prostate cancers being most prevalent, shows the necessity for specialized oncology services and preventive screening interventions.
- **Lab Metrics**: The variation of A1C and BMI levels suggests ample opportunity for intervention on both metabolic and weight management.

9.3 Financial Insights

- **Cost Drivers**: The scatter plot of **Risk Score Impact** on Medical Costs clearly depicts a trend such that increased risk scores are accompanied by significantly greater medical costs, indicating a straightforward cost effect arising from unmanageable health risk.
- **Service Utilization**: The **cost distribution by service type** pie chart shows that **RX** and **IP** services account for the greatest portions of medical expenditures, and this highlights the costliness of inpatient treatment and the use of prescription drugs.
- **PMPM Variations**: The **Expected PMPM Change** by **Risk Level** bar chart shows that there is a higher risk associated with adverse PMPM change since it reflects the cost of high-risk members on healthcare spending.

10. Use Cases and Applications

The Risk Stratification Dashboard is designed to support a wide range of healthcare use cases, including:

- **Population Health Management**
  Healthcare providers can use the dashboard to identify high-risk patients, prioritize interventions for care, and allocate resources more effectively. This minimizes hospital admission, improves patient outcomes, and maximizes total healthcare costs.

- **Predictive Modeling for Risk Reduction**
  Predictive analytics can be employed for high-risk members to forecast likely adverse health events and implement preventive care practices, reducing long-term care costs and the burden on patients' quality of life.

- **Care Coordination and Management**
  The dashboard facilitates coordinated care planning through the provision of a holistic overview of member health, including chronic conditions, risk scores, and cost effects. An integrated care approach is needed in managing the complex patient who has multiple comorbidities.

- **Financial Forecasting and Budgeting**
  The financial insights provided by the dashboard can guide budget planning and financial forecasting, helping organizations identify cost-saving opportunities and optimize financial performance.

- **Personalized Patient Engagement**
  Member-level information supports patient-level patient engagement, improving care plan adherence and patient satisfaction with tailored communications that are derived from individualized health profiles.

11. Future Improvements

To enhance the utility and effectiveness of the Risk Stratification Dashboard, several improvements are recommended:

- **Enhanced Data Integration**
  Integrate more data sources, such as social determinants of health (SDOH), real-time clinical data, and patient-reported outcomes, to develop a more comprehensive view of patient risk.

- **Machine Learning and AI Integration**
  Integrate advanced machine learning algorithms to improve risk prediction, identify emerging health risks, and enable automated member segmentation.

- **Real-Time Data Processing**
  Incorporate features of real-time data processing to provide minute-to-minute insight to facilitate faster decision-making and preemptive care intervention.

- **Customizable Alerts and Notifications**
  Add functionality for real-time alerts based on significant changes in member risk profiles, facilitating immediate care team intervention.

- **Enhanced Visualization and User Experience**
  Refine the dashboard's user interface to improve navigation, filter capabilities, and overall user experience, making it more intuitive for healthcare professionals.

12. Challenges and Lessons Learned

Developing the Risk Stratification Dashboard presented several challenges, including:

- **Balancing Complexity and Usability**
  Crafting an inclusive yet usable dashboard involves designing carefully to prevent inundating users with excessive information while still offering actionable insights.

- **Scalability and Performance**
  With increasing healthcare data volumes, the dashboard needs to be scalable enough to process greater datasets without impacting performance.

- **Continuous Improvement**
  Healthcare data is dynamic and calls for constant updates and improvement to keep the dashboard relevant and effective.

13. Conclusion

Risk Stratification Dashboard is a significant leap forward in healthcare analytics, providing actionable information regarding population health, clinical performance, and financial performance. Through combining powerful data sources and innovative visualizations, the dashboard facilitates proactive care management, reduces healthcare costs, and improves patient outcomes. Improvements such as machine learning integration and real-time analytics in the future will continue to augment the platform's impact on healthcare decision-making.

Appendix

Appendix A: Data Dictionary

- **Age**: Member's age in years.
- **Gender**: Member's gender (Male/Female).
- **Chronic Disease Count**: Total number of chronic conditions per member.
- **Prospective Total Risk**: Estimated future health risk score.
- **Allowed Medical Costs**: Total allowed medical expenses for each member.

Appendix B: Abbreviations

- **HTNIND** - Hypertension Indicator
- **HLDIND** - Hyperlipidemia Indicator
- **CKDIND** - Chronic Kidney Disease Indicator
- **A1C** - Hemoglobin A1C, a measure of blood sugar levels
- **BMI** - Body Mass Index, a measure of body fat based on height and weight