**MEDICARE DYNAMICS**

**TRENDS, GEOGRAPHICAL VARIATIONS AND BENEFICIARY INSIGHTS**​

**Team 05**

Donetra Hemraj Lanjewar

Lakshmi Tatavarty

Sarvani Kodeboyina

Sowrabh Tirumala Vinjamuri

**G. Brint Ryan College of Business**

Table of Contents

[Table of Contents 4](#_Toc165130919)

[List of Figures 5](#_Toc165130920)

[List of Tables 6](#_Toc165130921)

[1. ACKNOWLEDGMENT 7](#_Toc165130922)

[2. ABSTRACT 8](#_Toc165130923)

[3. EXECUTIVE SUMMARY 9](#_Toc165130924)

[4. INTRODUCTION 10](#_Toc165130925)

[**4.1** **Stakeholders in Medicare Enrollment Research** 11](#_Toc165130926)

[**4.1.1****Primary Stakeholders** 11](#_Toc165130927)

[**4.1.2****Secondary Stakeholders** 11](#_Toc165130928)

[5 BUSINESS PROBLEM 13](#_Toc165130929)

[**5.1** **Research Problem** 13](#_Toc165130930)

[**5.2** **Research Objective** 13](#_Toc165130931)

[**5.3** **Research Questions** 14](#_Toc165130932)

[6 LITERATURE REVIEW 15](#_Toc165130933)

[**6.1** **Theories** 15](#_Toc165130934)

[**6.2** **Bias and Limitations** 15](#_Toc165130935)

[**6.3** **Other Research Study** 16](#_Toc165130936)

[**6.3.1****Title: Trends and Regional Variations in Hospital Use under Medicare** 16](#_Toc165130937)

[**6.3.2****Title: Medicare Advantage Enrollment Among Beneficiaries with End-Stage Renal Disease in the First Year of the 21st Century Cures Act** 17](#_Toc165130938)

[**6.3.3****Title: Trends in Cumulative Dis-enrollment in the Medicare Advantage Program, 2011-2020** 17](#_Toc165130939)

[**6.4** **Other Organizations** 18](#_Toc165130940)

[**6.4.1****Title: Medicare Advantage in 2023: Enrollment Update and Key Trends** 18](#_Toc165130941)

[7 RESEARCH DESIGN 19](#_Toc165130942)

[**7.1** **Research Design and Methods** 19](#_Toc165130943)

[**7.1.1****Correlation Analysis** 19](#_Toc165130944)

[**7.1.2****Feature Selection Techniques** 19](#_Toc165130945)

[**7.1.3****KNN** 19](#_Toc165130946)

[**7.1.4****Gradient Boosting** 20](#_Toc165130947)

[**7.1.5****Neural Networking** 20](#_Toc165130948)

[**7.2** **Dependent and Independent Variables** 21](#_Toc165130949)

[**7.2.1****Dependent Variable** 21](#_Toc165130950)

[**7.2.2****Independent Variable** 21](#_Toc165130951)

[**7.2.3****Relation between the dependent and independent variable** 21](#_Toc165130952)

[8 DATA 23](#_Toc165130953)

[**8.1** **About Data** 23](#_Toc165130954)

[9 ANALYSIS 28](#_Toc165130955)

[**9.1** **Data pre-processing 1** 28](#_Toc165130956)

[**9.1.1****Handling Missing Values and Special Characters** 28](#_Toc165130957)

[**9.1.2****Categorizing Variables** 28](#_Toc165130958)

[**9.1.3****Univariate Analysis** 29](#_Toc165130959)

[**9.1.4****Multivariate Analysis** 31](#_Toc165130960)

[**9.1.5****Pair Plots** 32](#_Toc165130961)

[**9.2** **Data pre-processing 2** 33](#_Toc165130962)

[**9.2.1****Normalizing the data** 33](#_Toc165130963)

[**9.2.2****Eradicating overfitting** 34](#_Toc165130964)

[**9.2.3****Improve interpretability** 34](#_Toc165130965)

[**9.2.4****Outliers' reduction** 34](#_Toc165130966)

[**9.3** **Meaning of the Results** 34](#_Toc165130967)

[**9.3.1****Trends in our data** 34](#_Toc165130968)

[**9.3.2****Factors affecting the dependent variable.** 36](#_Toc165130969)

[10 RESULTS 44](#_Toc165130970)

[**10.1** **Research Findings** 44](#_Toc165130971)

[11 DISCUSSIONS AND LIMITATIONS 45](#_Toc165130972)

[**11.1** **Discussions** 45](#_Toc165130973)

[**11.2** **Limitations** 45](#_Toc165130974)

[12 CONCLUSION AND RECOMMENDATIONS 46](#_Toc165130975)

[**12.1** **Conclusion** 46](#_Toc165130976)

[**12.2** **Recommendations** 46](#_Toc165130977)

[13 REFERENCE 48](#_Toc165130978)

[14 CONTRIBUTION 49](#_Toc165130979)

# **List of Figures**

[Figure 8. 1 Data variables information 25](#_Toc164978706)

[Figure 8. 2 Top 10 data records 26](#_Toc164978707)

[Figure 9. 1 Sample of dataset 28](#_Toc164978721)

[Figure 9. 2 Comparative count of beneficiaries across various groups 29](#_Toc164978722)

[Figure 9. 3 Comprehensive Dataset Statistics 30](#_Toc164978723)

[Figure 9. 4 A matrix of scatter plots 31](#_Toc164978724)

[Figure 9. 5 Applying MinMaxScaler to normalize the data 32](#_Toc164978725)

[Figure 9. 6 Comparison between Enrollment for MAPD and PDP between 2021 To 2022 33](#_Toc164978726)

[Figure 9. 7 Comparison between Total Beneficiaries and Prescription Drug Beneficiaries in 2021 to 2022 34](#_Toc164978727)

[Figure 9. 8 Distribution of Total Beneficiaries across Counties 35](#_Toc164978728)

[Figure 9. 9 Implementing Linear Regression in Python with Sklearn Library 36](#_Toc164978729)

[Figure 9. 10 Insights into Regression Analysis 37](#_Toc164978730)

[Figure 9. 11 OLS Regression results 38](#_Toc164978731)

[Figure 9. 12 Implementing PCA analysis with the Sklearn library 39](#_Toc164978732)

[Figure 9. 13 Evaluating Models 39](#_Toc164978733)

[Figure 9. 14 Python code and output for Decision Tree model 40](#_Toc164978734)

[Figure 9. 15 Implementing a k-NN Regressor to Predict and Evaluate Data 41](#_Toc164978735)

[Figure 9. 16 Python code for training and evaluating a Gradient Boosting Regressor model using scikit-learn 41](#_Toc164978736)

# **List of Tables**

[Table 8. 1 23](#_Toc164978782)

[Table 10. 1 43](#_Toc164978790)

# **ACKNOWLEDGMENT**

We express our sincere appreciation to Dr. Sameh Shamroukh for his invaluable mentorship and guidance throughout our research endeavours. Throughout the "Business Process Analytics" course, Dr. Shamroukh's support enabled us to develop essential skills and insights that were instrumental in the successful completion of our final research project.

Additionally, we extend our gratitude to every member of our team, "Group 5," for their significant contributions and dedication to our project, MEDICARE DYNAMICS: TRENDS, GEOGRAPHICAL VARIATIONS, AND BENEFICIARY INSIGHTS

We aspire that our research serves its intended purposes and makes meaningful contributions to Data Science. We hope our findings will help aspiring students navigate their path toward achieving their objectives.

# **ABSTRACT**

The analysis provides us a comprehensive understanding of enrollment dynamics trends across the states of the United States by examining the complex landscape of Medicare Monthly Enrollment Data in detail. The study highlights healthcare needs, demographic variations, and other common factors among Medicare enrollees. We reviewed some academic publications in the literature review section to identify shifts. Additionally, methods like Decision Trees, Gradient Boosting, and Linear Regression are utilized to extract some insightful information from the raw data. By identifying enrollment patterns in Prescription Drug Coverage Plans and other benefits through a rigorous process, the research shows how healthcare demands vary over various geographical regions. By finding the variables that impact the original Medicare beneficiaries, the report is significant in guiding policy discussions and bringing necessary changes in healthcare delivery systems.

# **EXECUTIVE SUMMARY**

This report consists of a deep analysis of Medicare enrollment trends and prescription drug coverage in the United States from the past decade 2013 to 2023. The data is from data.gov, We have explored the factors influencing beneficiaries and their characteristics and temporal and spatial variations. Based on the analysis, shows that there have been some changes in enrollment rates over the past years, which includes some changes in states and demographic groups.  
One of the important keys reveals is the ongoing downtrend enrollment in Medicare advantage prescription drug plans, especially among retirees and people residing in Georgia and North Carolina. These trends focus on the importance of targeted policies to make sure fair healthcare access for all demographics and regions.

Apart from enrollment trends, we have examined prescription drug coverage patterns, showing insights into the adoption of Medicare Advantage plans and the percentage of people who qualify for low-income subsidies. From the analysis we did, there are correlations between demographic factors and enrollment patterns, which shows some scope for future investigation.  
For this analysis, we did a range of statistical techniques, that includes analyses like univariate, bivariate, and multivariate and machine learning algorithms including linear regression, neural networks, and decision trees. The Data pre-processing steps are as follows standardizing formats and handling missing values are necessary to be sure about the accuracy and dependency of our results.  
Overall, this report offers valuable insights on Medicare enrollment and prescription drug coverage and offering policymakers, stakeholders, and healthcare providers to improve healthcare access across the nation so that it can benefit everyone.

# **INTRODUCTION**

In our analysis, we are exploring the Medicare Monthly Enrollment Data for October 2023. The dataset captures the complex terrain of Medicare enrolment across the United States.

The report has a comprehensive analysis that deals with Medicare Monthly Enrollment Data till 2023. The data provides us with a detailed picture of people enrolled in Medicare across the country US. We have analyzed the data and learned about the people relying on Medicare types and utilized these insights for decision-making in healthcare programs.

The dataset provides a granular view of complete medical Enrollment data, that divides into various categories of beneficiaries and their coverage types. The dataset included total beneficiaries (TOT\_BENES), which is the different combinations of beneficiaries enrolled in original Medicare (ORGNL\_MDCR\_BENES) and Medicare Advantage or other plan (MA\_AND\_OTH\_BENES). The difference is necessary for further analysis to observe the trends present in Medicare. This trend analysis will also help us to provide more plans that are helpful and improve the health plans.

Furthermore, the dataset also has age and disability status parameters including total aged beneficiaries (AGED\_TOT\_BENES) for people with End-Stage Renal Disease (ESRD) among the aged (AGED\_ESRD\_BENES) and disabled beneficiaries (DSBLD\_TOT\_BENES). These helps understand demographic changes and healthcare needs in the Medicare population, which opens up aimed policy interventions and allocation of resources.

The major component of healthcare for all beneficiaries is prescription drug coverage. This dataset has different beneficiary plans along with prescription plans such as standalone prescription drug plan (PDP) and Medicare Advantage Prescription drug plans (MAPD), including the e low-income subsidy (LIS) eligibility criteria. This division highlights the potential to improve drug coverage and affordability for Medicare beneficiaries based on the accessibility and utilization of prescription drugs.

The data collected is based on different geographical regions based on national, state, and county, allowing in-depth analysis of Medicare Enrolment. This kind of data is beneficial in defining regional disparities in beneficiary demographics and Medicare coverage, improving healthcare policy and program development on a regional level.

Our goal is to discover the dynamics of Medicare Enrolment, analyzing the disparities, trends, and opportunities for accessing healthcare enhancement and maintaining it in the Medicare program. Our analysis looks to contribute to the progress of healthcare policy, to define and shape the well-being and healthcare outcomes of Medicare beneficiaries nationwide.

## **Stakeholders in Medicare Enrollment Research**

The research on Medicare Enrollment disparities and trends has a vital impact on stakeholders. Below is a breakdown segregating them as primary and secondary:

### **Primary Stakeholders**

* + - 1. Policymakers:
      2. Policymakers have the power to create and implement policies that direct Medicare Enrollment and its access. Our research findings can inform decisions related to program adjustments, resource allocation, and initiatives aimed at promoting equitable access.
      3. Medicare Beneficiaries:
      4. Enrollment trends and disparities are impacted by the people enrolled in Medicare programs. Our goal is to provide the affordable beneficiary plan that individuals deserve.

### **Secondary Stakeholders**

* + - 1. Researchers:

Influencing Medicare Enrollment is broadened with this research, and it is helpful in further development of access, program design, and patient outcomes.

* + - 1. Healthcare Providers:

An in-depth understanding of Enrollment patterns over different regions can improve the services provided by healthcare providers. These factors influence Enrollment choices and they can utilize these factors to reach most of the population’s needs.

* + - 1. Public Interest Groups:

These organizations help seniors, people with disabilities, or have healthcare access and can support our research findings to raise awareness about accessing Medicare with disparities

* + - 1. Private Insurance Companies:

Insights and outcomes obtained from our research can influence plan selection, introduce different offers, and improve marketing strategies in Medicare Advantage or Part D prescription drug plans.

# **BUSINESS PROBLEM**

* 1. **Research Problem**   
     This project aims to deal with the complexities involved in Medicare Enrollment and to improve inconsistency in accessing healthcare services among different regions. We are focusing on temporal and spatial dimensions where our study looks to deliver beneficial Medicare services at the state and national levels. Through data analysis and advanced statistical techniques, we study patterns in Enrollment rates over time and trends across various geographic regions.

The analysis helps us to track changes in Medicare Enrollment patterns for specific periods and enables us to study trends, fluctuations, and potential drivers. Studying these dynamics is important for expecting future shifts in Enrollment and informing proactive policy planning and allocation of resource strategies.

Our research methods include data collection from reliable sources, data cleaning, and using advanced statistical techniques such as trend analysis, time series modelling, and regression analysis. Moreover, we have used geographic information systems (GIS) to conduct spatial analysis and visualize spatial patterns, enhancing our understanding of Medicare Enrollment over geographical regions.

The goal of this research is to provide evidence for policy decisions and targeted interventions to provide equal access and allocation of resources to Medicare services for all demographic groups. We aspire to provide a wide range of improving health outcomes and foster a more inclusive healthcare system for all beneficiaries by studying the disparities and trends.

## **Research Objective**

The goal of this is to deal with the dynamics of Medicare Enrollment by focusing on prescription drug coverage and total beneficiaries. The outcomes from this research are expected to strengthen the areas providing Medicare plans and accessible prescription drug coverage for all beneficiaries.

The research objectives include:

* + 1. The initial focus is to determine the factors influencing the total number of beneficiaries, which serves as the dependent variable in our analysis. This results in inquiring about demographic, policy-related variables, and economics to improve their impact on Medicare Enrolment. This analysis will provide insights to forecast future trends and improve policy decisions in modifying healthcare coverage plans for elderly and disabled people.
    2. Secondly, the research involves studying the patterns that occurred in prescription drug coverage among Medicare beneficiaries. This analysis of various factors such as policy changes, demographic shifts, and market dynamics to understand how beneficiaries are influenced. This vital study is used for assessing the accessibility and adequacy of pharmaceutical care within the Medicare program.

## **Research Questions**

* + 1. What are the factors influencing dependent variable analysis? The variable TOT\_BENES represents the total number of beneficiaries. The question aims to find the many factors that impact this critical dependent variable.
    2. What trends exist in Medicare enrollees' prescription drug coverage? Examining the membership in Medicare Advantage prescription drug plans (MAPDs) as opposed to prescription drug plans (PDPs) and the percentage of beneficiaries judged eligible for low-income subsidies (LIS) could be part of this.

# **LITERATURE REVIEW**

## **Theories**

A Literature review is a comprehensive analysis of pre-existing literature and research works on a particular topic, incorporated into a united narrative. It involves the systematic search and review of related academic sources, analysis of their methodology, results, and theoretical structures, and identifying key topics, trends, and gaps in the literature. The Literature review provides a solid base for new research ventures by informing the existing landscape and hence advancing knowledge in their respective fields.

One of the research project papers is about Medicare Advantage Enrollment among beneficiaries with End-Stage Renal Disease during the first year of the 21st Century Cures Act (Nguyen KH, 2023). The project conducted a cross-sectional time-trend analysis to examine changes in the Medicare Advantage plan among Beneficiaries from 2019 to 2021 (Nguyen KH, 2023).

A different study, Trends in Cumulative Dis-enrollment in the Medicare Advantage Program, discovered disparities and performance differences amongst plans, as well as exceptionally high long-term churn in Medicare Advantage (Meyers DJ, 2023). This suggests re-evaluating our methods for measuring plan performance, going beyond variations in Enrollment yearly.

* 1. **Bias and Limitations**

After reviewing the literature, we have determined that there is a significant data availability challenge, specifically regarding specific characteristics and outcomes among people who will be enrolled in Medicare Advantage (MA) and Traditional Medicare (TM) plans in 2020 (Nguyen KH, 2023). This restriction may make it much more difficult for us to compare these groups fairly and identify common patterns in the data. For instance, it is challenging to create a thorough picture of coverage trends among Medicare enrollees if there is a dearth of information on prescription medication coverage for individuals who are dual enrolled in TM and MA. As a result, this deficit could cause the observed trends to be overestimated or underestimated. Our research can lessen this lack of overestimation and underestimation.

The study “Trends in Cumulative Medicare Advantage Dis-enrolment” did not consider how changes in Medicare Advantage beneficiary mortality rates could influence the results, potentially skewing the data (Meyers DJ, 2023). It is difficult to determine whether issues such as low patient satisfaction are a direct cause of Medicare Advantage program dis-enrolment. The study also did not fully explain why Medicare Advantage dis-enrollment has increased over the last few years. Our research aims to explore these factors in greater detail to understand better how they relate to the variables we are looking at.

Also, the applied models in the project no longer explicitly suggest the correlations among different factors and the modifications in dis-enrollment or insurance. As a result, those models are unable to conclude whether those elements directly affect the observed changes. Moreover, if there are massive elements influencing dis-enrollment or insurance traits that aren't accounted for inside the dataset, it may introduce bias into the outcomes generated with the aid of using the model. This highlights the significance of thinking about all applicable elements to ensure the accuracy and reliability of the findings.

## **Other Research Study**

### **Title: Trends and Regional Variations in Hospital Use under Medicare**

A survey conducted by Medicare and Medicaid inspected the dis-enrollment rates of patients with different chronic illnesses over distinctive states (M., 1982). With age, the predominance of chronic conditions increases, driving more hospitalizations, especially within the final years of life. The study focuses on regression analysis utilizing Medicare hospital discharge data from California and New York, two states with unmistakable characteristics.

Regression analysis can vary depending on the unit of investigation, whether it's people (an individual), hospitals, or geographic ranges. It's vital to recognize that this choice can affect the relationships between variables (M., 1982). For example, aggregated age data may not accurately predict hospital use in specific areas but at the same place, age in years may predict individual hospital use. Similarly, dissecting patterns in Medicare benefits over diverse geographic locales may not surrender exact accurate results. (M., 1982)

By utilizing compelling models such as neural networks and investigating different statistical strategies, the analysts were able to precisely degree dis-enrollment rates over states and understand regional variations in Medicare reimbursements (M., 1982). Similarly, for our research work, we plan to utilize diverse machine learning techniques to analyze trends over different topographical regions and recognize other factors impacting our dependent variable.

* + 1. **Title: Medicare Advantage Enrollment Among Beneficiaries with End-Stage Renal Disease in the First Year of the 21st Century Cures Act**

In this literature, we look at the effects of the 21st Century Cures Act on Medicare beneficiaries with ESRD and their Enrolments in Medicare Advantage plans (Nguyen KH, 2023). We look at the potential benefits of enrolling in Medicare Advantage plans for people with ESRD. Also, it talks about the need to monitor Enrollment patterns to ensure quality and equitable care. Finally, we look at why there are no estimates of the initial association between the 21st Century Cures Act and ESRD Enrollment choices, and why there is a need to understand the changes from traditional Medicare to Medicare Advantage plans across different beneficiary groups (Nguyen KH, 2023).

Besides, the literature also raises questions about the esteem of overseen care for people with genuine health conditions like ESRD. It emphasizes the noteworthiness of following dis-enrollment from MA plans as a marker of recipient inclinations and neglected needs (Nguyen KH, 2023). Understanding the move of recipients with ESRD from conventional Medicare to MA, particularly among specific beneficiary bunches, is vital for policymakers and MA plans to guarantee organized ampleness, optimize benefits, and advance even-handed care for people with ESRD.

* + 1. **Title: Trends in Cumulative Dis-enrollment in the Medicare Advantage Program, 2011-2020**

The methodology's qualities incorporate its longitudinal viewpoint, which sheds light on dis-enrollment patterns over a decade, and its population-based approach, upgrading the generalizability of the discoveries (Meyers DJ, 2023). In this case, as the study's nature is observational, it implies that it cannot build up causality between the inspected variables and dis-enrollment rates. Furthermore, potential perplexing variables might not be completely accounted for, and the findings' appropriateness may be constrained to the Medicare populace.

The key discoveries incorporate noteworthy dis-enrollment rates over five years duration, with varieties based on racial and comorbidity components, as well as the sort of Medicare Advantage (MA) contracts. The strategy included a cross-sectional ponder of Medicare recipients selected in MA between 2011 and 2020, focusing on dis-enrollment rates and their relationships with recipient and contract characteristics (Meyers DJ, 2023). Recognized crevices incorporate the associational nature of the work, which limits causation assessment, potential inclinations due to differential mortality, and a need for knowledge of the reasons behind dis-enrollment or consequent Enrolments.

## **Other Organizations**

### **Title: Medicare Advantage in 2023: Enrollment Update and Key Trends**

Extensive research was conducted by a non-profit organization on the growth and characteristics of the Medicare Advantage Program in 2023 (Nancy Ochieng, 2023). Advanced methods like decision trees and KNN models were utilized in the study to extract insights about the most influential factors for Enrollment rates in specific areas. Through these algorithms, they picked up bits of knowledge about the complicated elements influencing their dependent variable and picked up a more profound understanding of their impact on Medicare enrollees' patterns.

Moving ahead, the study utilized cluster analysis to segment counties based on their market characteristics, pinpointing locales with restricted competition. This analysis gave important bits of knowledge into market structures, supporting policymakers in distinguishing regions that if worked on would boost market competition and improve Enrolment (Nancy Ochieng, 2023).

Furthermore, predictive modelling techniques such as Support Vector Machines were utilized to look at the relationship between market structure and Enrollment results (Nancy Ochieng, 2023). This examination helped policymakers in distinguishing openings to upgrade showcase competitiveness and move forward Enrollment results.

Besides, the study recommends that advanced examination utilizing Natural Language Processing (NLP) methods could affect the approach to policy changes (Nancy Ochieng, 2023). NLP can extract key policies from policy documents and regulatory filings. This can provide policymakers with valuable insights for informed decision-making and further potential policy interventions.

# **RESEARCH DESIGN**

## **Research Design and Methods**

### **Correlation Analysis**

Relationship examination could be a measurable strategy to determine the strength, quality, and direction of the relationship between two or more variables. It gives a clear view of how closely two variables are related, without suggesting causation.

The most used measure of a relationship’s degree is the relationship coefficient, ordinarily indicated as "r" which ranges from -1 to 1.

* + - 1. When the relationship coefficient r is 1, it concludes a positive relationship, meaning that as one variable increments, the other variable moreover increments relatively.
      2. When the relationship coefficient r is -1, it shows an idealized negative relationship, meaning that as one variable increments, the other variable diminishes relatively.
      3. When the relationship coefficient r is 0, it demonstrates no straight relationship between the variables.

### **Feature Selection Techniques**

To recognize the most important free elements, this method utilizes feature selection procedures. Common strategies to proceed with in detecting the foremost factors are:

* + - 1. Filter Methods: The approach evaluates the relationship between each free variable and the target variable independent of the ML algorithm. Correlation-based feature selection and statistical tests (e.g., ANOVA) are examples of this method.
      2. Wrapper Methods: The method evaluates distinctive subsets of features by training and assessing the models iteratively. Forward selection, backward elimination, and recursive feature elimination (RFE) are some of the examples of this method.
      3. Embedded Methods: The method includes feature selection straightforwardly into training the model. LASSO (L1 regularization) and decision tree-based feature importance are two examples of this method.

### **KNN**

For problems including regression and classification, the supervised machine learning model K-Nearest Neighbours (KNN) algorithm is employed. In KNN classification model forecasts a new data point's class based on the majority class of feature space's k nearest neighbours. Similar to this, in KNN regression, the model uses the average of its k nearest neighbours’ values, or some other aggregation function, to forecast the value of a new data point. As a non-parametric, instance-based learning method, it keeps the complete training dataset for inference and makes no assumptions about the underlying data distribution. One crucial hyperparameter that influences the generalization and performance of the model is the selection of k, or the number of neighbours taken into account. KNN is easy to interpret, implement, and use, making it a popular choice for various classification and regression tasks, especially in the case of medium-sized datasets.

### **Gradient Boosting**

Gradient boosting is a powerful group of learning both regression and classification tasks. It is run by adding typically decision trees, and weak learners into a strong prediction model. This algorithm builds the combination in a particular order where every new model corrects the previous ones. The focus minimizes the loss function by adding fitting new trees to the residuals of the previously predicted outcomes. The idea of gradient boosting is to rearrange the performance of the group by gradient descent, where each learner is trained to minimize the loss of the model.

### **Neural Networking**

Artificial neural systems, or ANNs, are a computational model inspired by the engineering, function, structure, and operations of the neural systems found in the human brain. It is made up of layer-organized, interconnected nodes or neurons. Each neuron takes in approaching data as input signals, applies an activation function to process it, and after that transmits the result to neurons within the layer over. Expansive volumes of information are utilized to prepare neural systems so they can recognize complex designs and associations between inputs and yields. In the process of decreasing the disparity between the anticipated and genuine yields, the weights and predispositions of the associations between neurons are balanced amid this preparation stage. Once trained, neural systems can be prepared for various assignments, such as design acknowledgment, classification, regression, and reinforcement learning. Various fields such as image recognition, natural language processing, and autonomous driving have witnessed remarkable success of Neural Networks, making them fundamental tools in modern AI.

## **Dependent and Independent Variables**

### **Dependent Variable**

Total Number of Beneficiaries (TOT\_BENES): This variable represents the total number of beneficiaries enrolled in Medicare. It is the primary outcome variable that we seek to understand and explain based on various factors.

### **Independent Variable**

* + - 1. Location (BENE\_STATE\_DESC, BENE\_COUNTY\_DESC): Geographic location, such as state and county, serves as an independent variable. It can influence the total number of beneficiaries due to differences in population density, access to healthcare facilities, socio-economic factors, and regional healthcare policies.
      2. Age of Beneficiaries with End-Stage Renal Disease (AGED\_ESRD\_BENES): This variable represents the number of beneficiaries with End-Stage Renal Disease (ESRD) within a specific age group. ESRD prevalence may vary across different age cohorts, impacting the total beneficiary count.
      3. Presence of Disabilities and ESRD (DSBLD\_ESRD\_AND\_ESRD\_ONLY \_BENES ): This variable indicates the number of beneficiaries with disabilities and ESRD. Disabilities and chronic health conditions like ESRD can affect healthcare utilization patterns, potentially influencing the total number of beneficiaries.

### **Relation between the dependent and independent variable**

The independent variables, including location, age of beneficiaries with ESRD, and the presence of disabilities and ESRD, are hypothesized to have prescient control in clarifying varieties within the subordinate variable, the total number of beneficiaries (TOT\_BENES).

For instance, locales with the next extent of elderly people or the next predominance of ESRD and incapacities may show a higher total beneficiary count. Additionally, certain geographic regions may have aberrations in getting access to healthcare, resulting in differences in recipient Enrolment.

By analysing the connections between these dependent and independent factors through factual modelling strategies like regression analysis, we can get insights into how these variables collectively contribute to the entire number of Medicare recipients.

# **DATA**

## **About Data**

The Medicare Monthly Enrollment dataset provides a thorough analysis of trends in Medicare enrolment from 2013 to 2022 at various geographic levels including county, state, and national levels. Enrolments are broken down into monthly and annual categories according to different geographic areas, including the federal, state, and local levels. The count of total Medicare beneficiaries (including those with Medicare Advantage and Original Medicare plans), their age, and disability status are important factors considered in the analysis.

The dataset distinguishes between those with Stand-Alone Prescription Medication Plans (PDP) and Medicare Advantage Prescription Drug Plans (MAPD). Additionally, it gives information on which beneficiaries are partially or completely enrolled in Medicare Part D and further categorizes each based on their low-income subsidies (LIS).

The dataset has a total of 26 variables including total beneficiaries, aged beneficiaries, disabled beneficiaries, and other various prescription drug coverage plans beneficiaries enrolled in. This provides a clear understanding of the dynamics related to enrolees in Medicare plans over time and for different geographical locations. The further division of geographical segments allows us to compare the trends in data and get further insights at sub-levels.

The derived insights will act as a piece of treasure for policymakers and researchers who tend to understand the issues in equity and healthcare access among beneficiaries and ought to resolve them. This can be done by identifying the trends and patterns within the data and hence discovering the areas for improvement in delivering healthcare access to every person equally, allocating resources, and policy development to serve the diverse needs of Medicare enrolees nationwide.

The following are variables for the dataset:

Table 8. 1

Variables in the dataset

|  |  |
| --- | --- |
| ATTRIBUTE | DEFINITION |
| YEAR | Medicare Enrollment for calendar year |
| MONTH | Medicare Enrollment for calendar month |
| BENE\_GEO\_LVL | Aggregated data for each geographic level like National, State and county. |
| BENE\_STATE\_ABRVTN | Beneficiary residence based on State abbreviation |
| BENE\_STATE\_DESC | Beneficiary residence area. |
| BENE\_COUNTY\_DESC | Beneficiary residence county |
| BENE\_FIPS\_CD | County/State FIPS code of beneficiary residence |
| TOT\_BENES | Count of all the Medical Beneficiaries |
| ORGNL\_MDCR\_BENES | Count of all Original Medicare beneficiaries |
| MA\_AND\_OTH\_BENES | Count of all Medicare Advantage and Other Health Plan beneficiaries |
| AGED\_TOT\_BENES | Count of Medicare aged beneficiaries |
| AGED\_ESRD\_BENES | Count of Medicare aged beneficiaries with End Stage Renal Disease |
| AGED\_NO\_ESRD\_BENES | Count of Medicare aged beneficiaries without End Stage Renal Disease |
| DSBLD\_TOT\_BENES | Count of Medicare disabled beneficiaries |
| DSBLD\_ESRD\_AND\_ESRD\_ONLY\_BENES | Count of Medicare disabled beneficiaries with End Stage Renal Disease and beneficiaries with End Stage Renal Disease only |
| DSBLD\_NO\_ESRD\_BENES | Count of Medicare disabled beneficiaries without End Stage Renal Disease |
| A\_B\_TOT\_BENES | Count of Medicare beneficiaries enrolled in Hospital Insurance (or Part A) and Supplementary Medical Insurance (or Part B) |
| A\_B\_ORGNL\_MDCR\_BENES | Count of Original Medicare beneficiaries enrolled in Hospital Insurance (or Part A) and Supplementary Medical Insurance (or Part B) |
| A\_B\_MA\_AND\_OTH\_BENES | Count of Medicare Advantage and Other Health Plan beneficiaries enrolled in Hospital Insurance (or Part A) and Supplementary Medical Insurance (or Part B) |
| PRSCRPTN\_DRUG\_TOT\_BENES | Count of all Medicare Prescription Drug (or Part D) beneficiaries |
| PRSCRPTN\_DRUG\_PDP\_BENES | Count of Medicare Prescription Drug (or Part D) beneficiaries enrolled in a Prescription Drug Plan |
| PRSCRPTN\_DRUG\_MAPD\_BENES | Count of Medicare Prescription Drug (or Part D) beneficiaries enrolled in a Medicare Advantage Prescription Drug plan |
| PRSCRPTN\_DRUG\_DEEMED\_ELIGIBLE\_FULL\_LIS\_BENES | Specific count of certain Medicare beneficiaries for low-income subsidy |
| PRSCRPTN\_DRUG\_FULL\_LIS\_BENES | Count of Medicare beneficiaries with limited income and resources who do not fall into one of the deemed subsidy groups and are enrolled for full subsidy. These beneficiaries successfully applied for low-income subsidy. |
| PRSCRPTN\_DRUG\_PARTIAL\_LIS\_BENES | Count of Medicare beneficiaries with limited income and resources who do not fall into one of the deemed subsidy groups and are enrolled for partial subsidy. These beneficiaries successfully applied for low-income subsidy. |
| PRSCRPTN\_DRUG\_NO\_LIS\_BENES | Count of Medicare Part D beneficiaries with no low-income subsidy. |

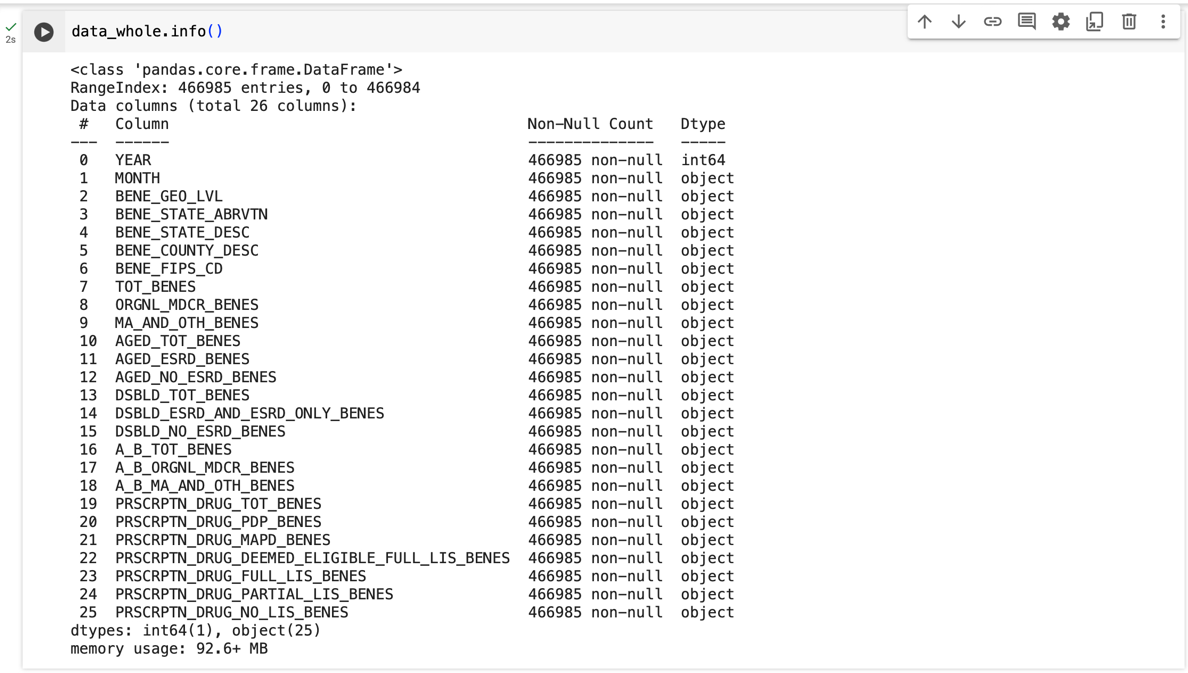


Figure 8. 1 Data variables information

The Data Frame consists of a total count of 466,985 data records and 26 columns, with each column representing a different unique feature of Medicare Enrolment. The above snap gives the names of all available columns and their data types.

All other columns have datatype as object (which usually indicates as string data), except for column ‘YEAR’ which has int datatype. It further says that the dataset has no null values, suggesting that there are no missing values in the dataset.

This gives us an overall comprehensive view of the data composition and datatypes and includes any missing values. Further analysis of deriving insights can be continued with this information as a base to explore the data in detail.

A sizable amount of the sample data in the dataset we're examining is now saved as an 'Object' data type. Typically, this sort of data represents categorized or textual data. But numerical data types like integers or floats are essential for our analytical needs, especially for variables that represent beneficiary counts. We first need to convert the datatype of these variables to numerical so that we can perform quantitative and statistical analysis on it. This conversion to integer or float will ensure that the count of beneficiaries and other quantitative variables are properly represented, allowing for a comprehensive exploration of trends, patterns, and relationships of variables within the data. As a result, the whole process will enhance the analysis process ultimately supporting informed decision-making and strategic planning in the realm of Medicare enrolment.

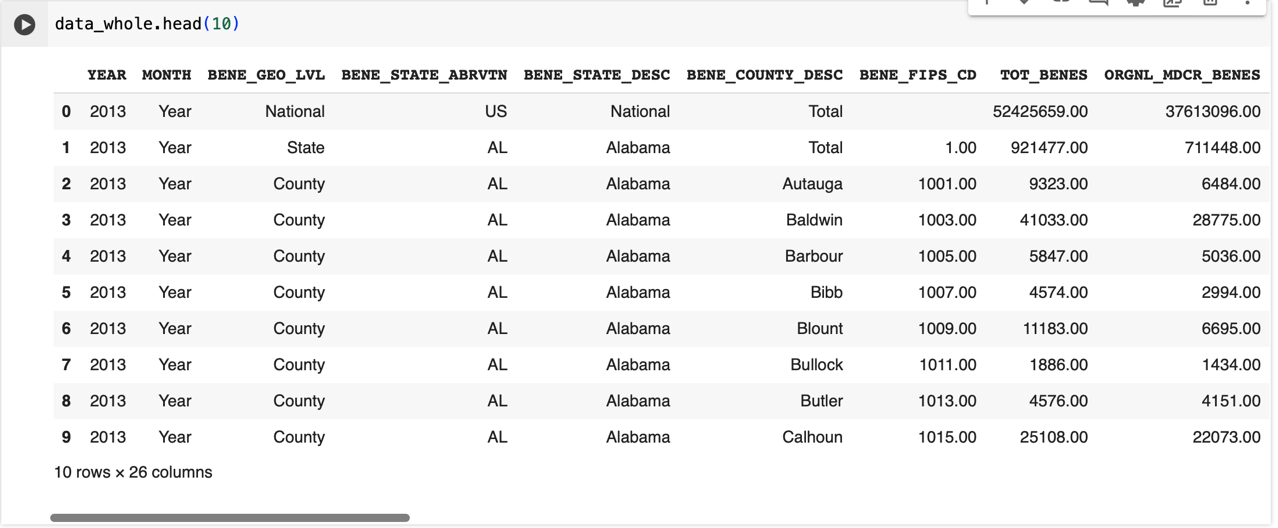


Figure 8. 2 Top 10 data records

# **ANALYSIS**

## **Data pre-processing 1**

### **Handling Missing Values and Special Characters**

The crucial step to begin data analysis is data cleaning. We’ve taken steps to handle missing data and clean it to maintain the integrity and accuracy of the data. The process starts with identifying the special characters, totals present in certain rows, and default values and later addressing them.

We started with identifying the rows containing total count, denoted by ‘Total’, and the rows with unknown values, represented by ‘Unknown’. We removed these rows from the dataset to eliminate any discrepancies arising from aggregate values or missing information. Additionally, special characters, such as '\*', have been replaced with 0 to standardize data.

By following these cleansing measures, we improved the dataset quality and consistency, hence maintaining the accuracy and proper interpretation of the analysis.

### **Categorizing Variables**

Following data cleansing, we categorized the data into numerical and categorical.

Month, BENE\_GEO\_LVL, BENE\_STATE\_ABRVTN, BENE\_STATE\_DESC, BENE\_COUNTY\_DESC, and BENE\_FIPS\_CD are into categorical side. They all represent the qualitative distinct characteristics such as geographic level, state abbreviation, state description, county description, and FIPS code.

Variables like YEAR, TOT\_BENES, ORGNL\_MDCR\_BENES, MA\_AND\_OTH\_BENES,AGED\_TOT\_BENES,AGED\_ESRD\_BENES, AGED\_NO\_ESRD\_BENES,DSBLD\_TOT\_BENES, DSBLD\_ESRD\_AND\_ESRD\_ONLY\_BENES,DSBLD\_NO\_ESRD\_BENES, A\_B\_TOT\_BENES,A\_B\_ORGNL\_MDCR\_BENES, A\_B\_MA\_AND\_OTH\_BENES,PRSCRPTN\_DRUG\_TOT\_BENES, PRSCRPTN\_DRUG\_PDP\_BENES,PRSCRPTN\_DRUG\_MAPD\_BENES, PRSCRPTN\_DRUG\_DEEMED\_ELIGIBLE\_FULL\_LIS\_BENES, PRSCRPTN\_DRUG\_FULL\_LIS\_BENES, PRSCRPTN\_DRUG\_PARTIAL\_LIS\_BENES, and PRSCRPTN\_DRUG\_NO\_LIS\_BENES are categorized into numerical side. They all represent quantitative measurements such as count of distinct beneficiary types, total beneficiaries, and statistics on prescription drug coverage.

The differentiation between categorical and numerical data types is vital for conducting all the analyses including visualization, statistical modelling, and machine learning.

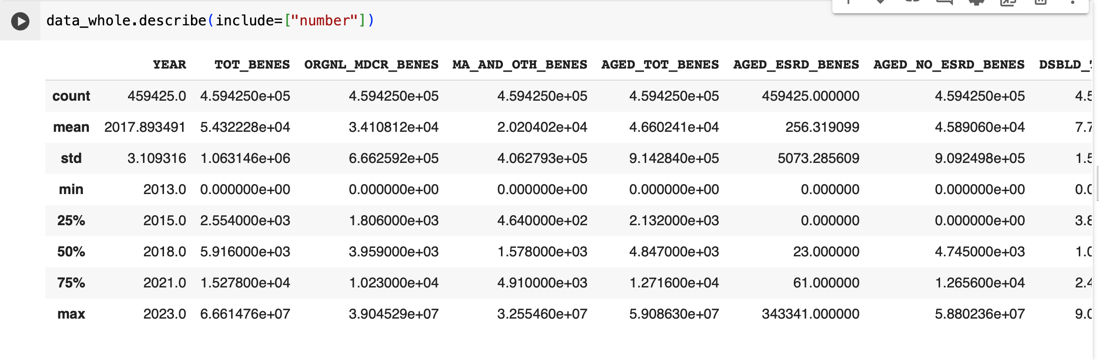


Figure 9. 1 Sample of dataset

### **Univariate Analysis**

Univariate analysis is a basic statistical analysis technique that examines one variable at a time. Doing so provides us with a thorough grasp of its distribution, central tendency, and variability.

Histogram out of all other tools used in univariate analysis, is one of the most useful and perceptive visualization tools. It also offers a lot of crucial information into the nature of data under investigation.

The shape of data distribution is the primary insight provided by the histogram. The data skewness, whether left or right or any other feature is depicted by the histogram graphically. Skewed distribution implies that the data is concentrated towards one side of the spectrum. Whereas a symmetric distribution (often in a bell-shaped curve), implies a normal distribution, i.e. the data is distributed uniformly around the mean.

Histograms are unable to provide precise measures of the central tendency, but they give important hints about the frequent values within the dataset. The tallest bar in the graph, indicating the mode of distribution, gives a clear picture of the central tendency. This mode is the value that occurs most frequently in the dataset, thus providing insight into the value that appears in the dataset with the highest frequency.

The Spread of the data is another important aspect that the histogram talks about. The extent to which data points deviate from the central tendency is depicted via the distribution of bars over the range of histogram. A widespread indicates that the data points are dispersed over a broader range of values. On the other hand, a closer cluster of data points towards the central tendency is indicated by a narrow range of bars, denoting lower variability. Researchers and analysts can refer to these insights to make relevant inferences, and defensive choices and draw meaningful insights from the data.

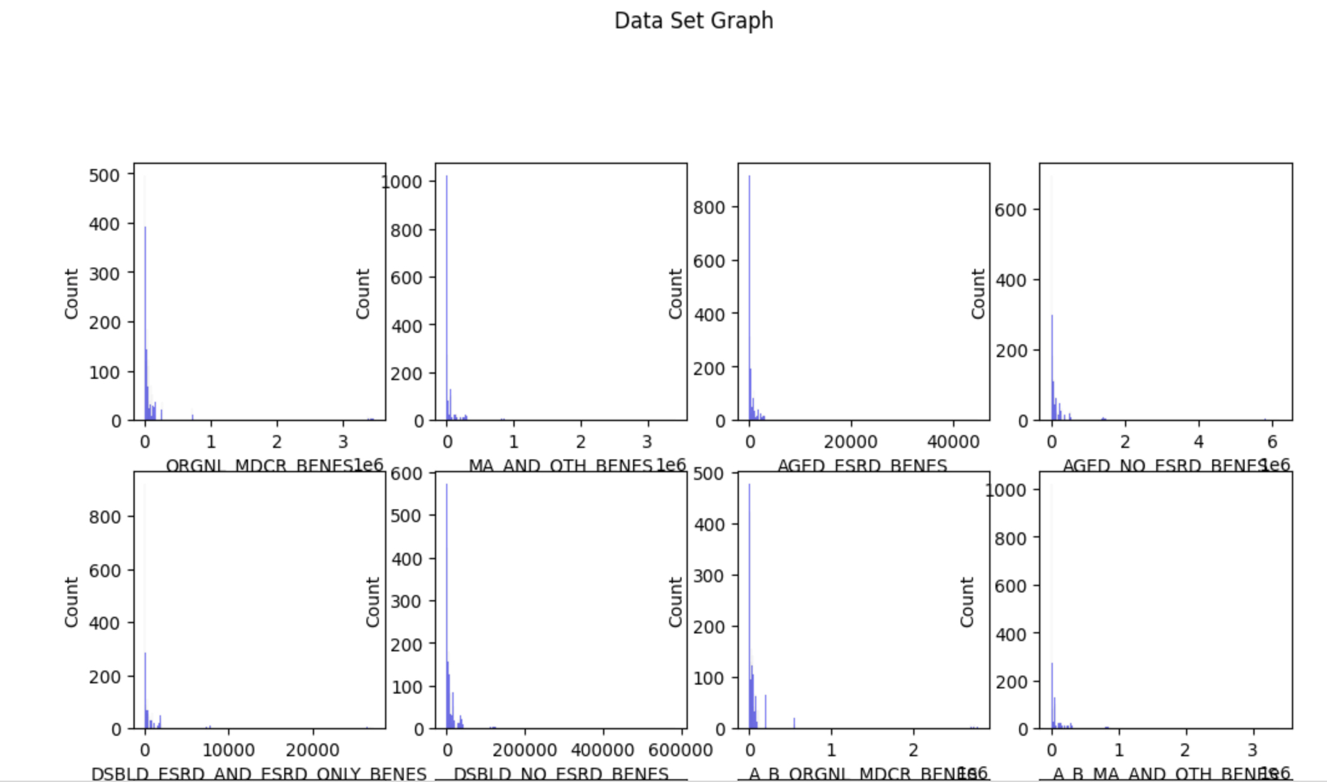


Figure 9. 2 Comparative count of beneficiaries across various groups

Although our dataset covers over 50 states and the years 2013 through 2023, the skewness is a valid and real occurrence in light of various demographic and geographical traits of various regions. States like California and Florida, well known for being attractive retirement destinations, naturally tend to have high rates of Medicare enrolment as compared to other states. The primary reason for this act is the pleasant weather and senior-focused amenities. In the same way, areas with denser populations tend to have higher beneficiary counts as the demand for healthcare services increases in such regions. This remains the primary reason for the skewness in Tt.

To focus on our analysis and mitigate the complexities associated with working with a large dataset, we have chosen to limit our examination to data from 202 onwards. By doing so, we can obtain a more current and relevant picture of the trends and developments in Medicare enrolment, providing clear insights into the factors driving enrollees', and minimizing the influence of historical data.

### **Multivariate Analysis**

To analyze data, we have used statistical techniques which have multiple variables, this analysis is known as multivariate analysis. Apart from univariate analysis or bivariate analysis that focuses on single or double variables. These multivariate studies the relationship between three or more than three variables to study and understand the relationship between them.

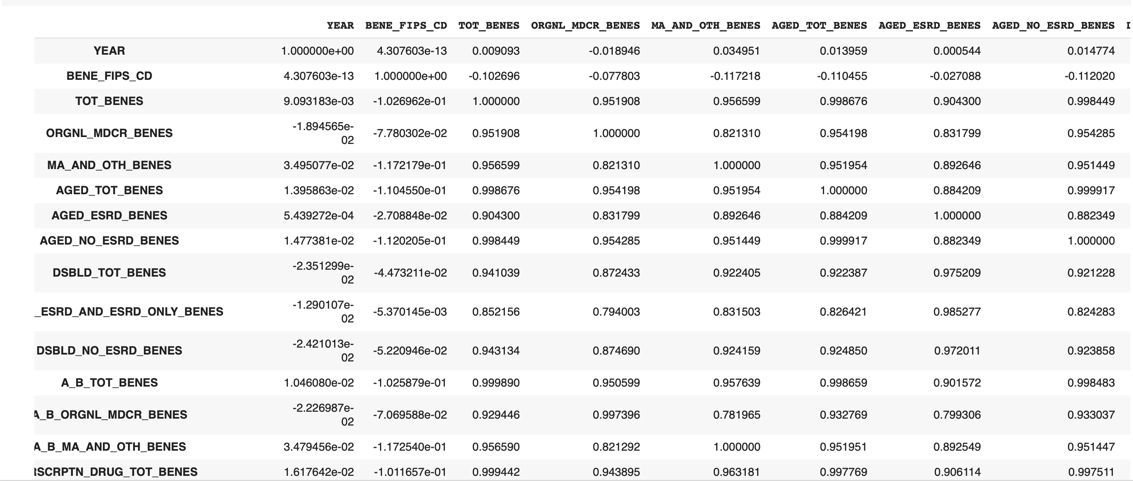


Figure 9. 3 Comprehensive Dataset Statistics

The correlation between variables is excessively high, resulting in issues of multicollinearity and further distorted results. To solve the high correlation, we have implemented logarithmic and cube root transformations.

Reducing correlation remains a vital task to identify and interpret the influencing factors for dependent variables accurately. These transformations reduce the correlation between independent and dependent variables, hence improving analysis robustness. Multivariate analysis thus helps us delve into the complex relation between the variables and reveal significant patterns that won’t be uncovered when analyzing only one or two variables at a time.

### **Pair Plots**

Graphical representations (as scatterplot matrices) that help visualize the relationship in a pairwise manner between different variables in a dataset. Generally used for getting insights into the relationships and patterns present in the dataset.

Along the diagonal of each pair plot, there is a histogram or kernel density plot for each variable in the analysis. Each of these plots represents the distribution of values for variables. Histograms provide details on the density of values within each variable with the spread and central tendency of the data.

For the off-diagonal cells of the pair plot, the scatterplots represent the relationship between each pair of variables. A plot of one variable against another illustrates how one of the variables will change corresponding to the other, while each point on the plot represents one single observation or data point. One can go for the details of strength, direction, and shape of the relation between variables, identifying clusters, outliers, and other patterns of linear or non-linear relations.



Figure 9. 4 A matrix of scatter plots

As Pari plots provide a comprehensive overview of the relationships between factors, they are generally used in Exploratory Data Analysis (EDA). Inspecting these plots gives valuable insights into the underlying structure of data, potential relations that may exist, correlations, etc. that guide further modelling efforts and subsequent analytical approaches.

The plots attached above clearly say that many variables are exhibiting a linear correlation against each other with each graph representing a different pair of variables.

Linear correlation implies that a change in one variable will affect and show up a corresponding change in another variable.

## **Data pre-processing 2**

### **Normalizing the data**

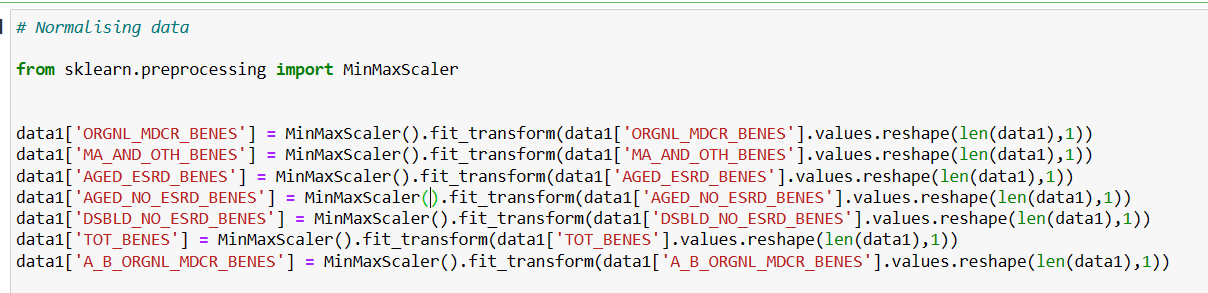


Figure 9. 5 Applying MinMaxScaler to normalize the data

Transforming data to achieve a specific distribution where the mean is 0 and the standard deviation is 1 is called Normalization. To achieve this process, we use the MinMaxScaler function from the sci-kit-learn library that fits the data within the interval [0, 1]. This subtracts the minimum value from each data point and divides the range of data to get normalization.

The advantages of Normalization or normalizing the data are enhancement of machine learning algorithms, prevention of gradient explosion, and variable consistency. It improves visualization and data analysis which results in better performance of machine learning algorithms.

In Medicare Monthly Enrollment data performing Normalization contributes below benefits:

Normalizing data helps the ML model play a vital role in understanding the relationship between variables and influencing them by differing scales.

### **Eradicating overfitting**

Overfitting is defined as the data point that excessively fits to train the data and how to generalize future data that can be minimized with normalization.

### **Improve interpretability**

The data is simplified to bring interpretability of output obtained through machine learning models by bringing all the features to same scale which is achieved through normalization.

### **Outliers' reduction**

Performing Normalization has an impact on machine learning models which reduces the outliers and improves accuracy of predictions.

The beneficial effect of performing Normalization as first step in Medicare Monthly Enrollment data provides forecasting, outputs provided by machine learning models and interpretability is improved.

## **Meaning of the Results**

### **Trends in our data**

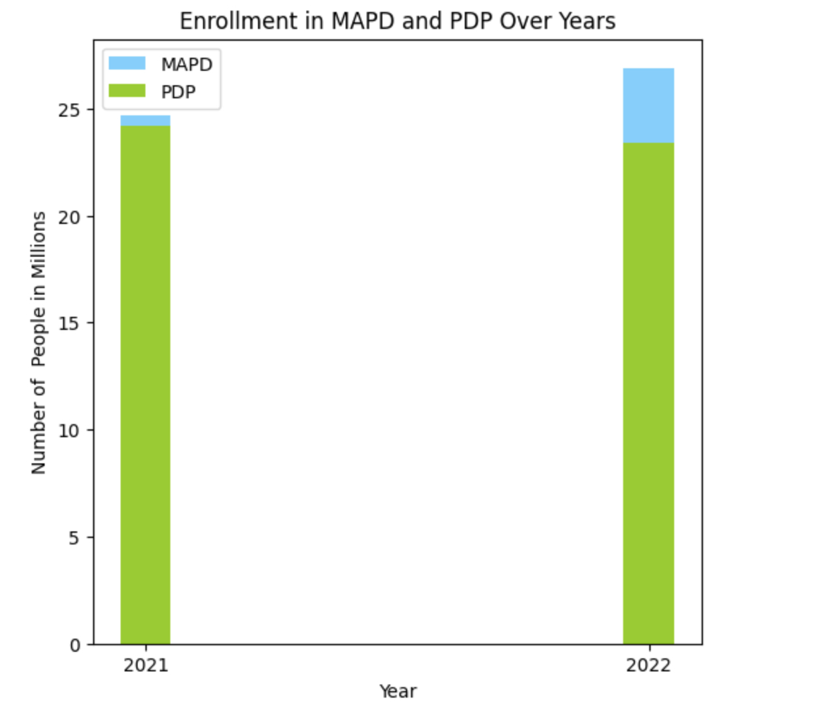


Figure 9. 6 Comparison between Enrollment for MAPD and PDP between 2021 To 2022

Considering two years' data and comparing them, we could observe the fluctuations in the Enrollment of beneficiaries in MAPD and PDP. MAPD has the highest Enrollment of beneficiaries than PDP beneficiary Enrolment. Enrollment increased in 2022 for MAPD, and PDP Enrollment decreased in 2022. With these observations, we can observe the beneficiary's behavior on how they are influenced in enrolling in different Medicare plans and observe that there is a decrease in PDP Enrollment for over years 2021 and 2022.

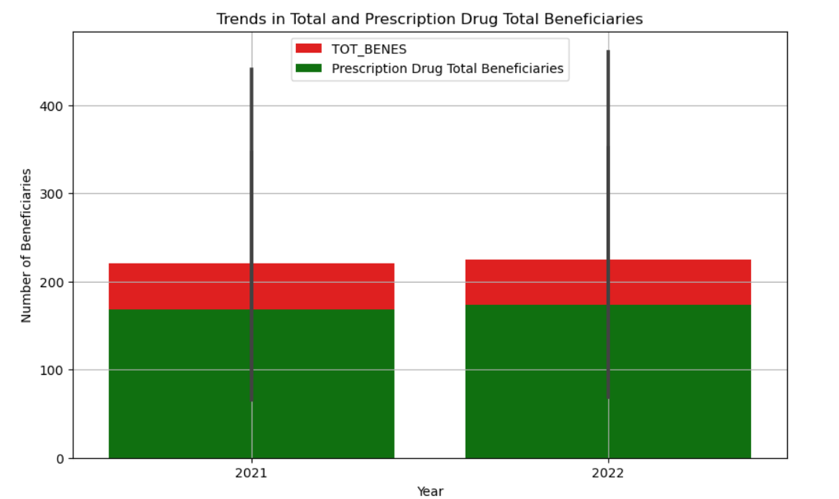


Figure 9. 7 Comparison between Total Beneficiaries and Prescription Drug Beneficiaries in 2021 to 2022

The bar graph depicts trends of the total beneficiaries and prescription drug total beneficiaries. Here the x axis shows the comparison of data between two cumulative years 2021 and 2022. The y-axis depicts the total number of beneficiaries. The graph has red and green lines where the red line displays the total number of beneficiaries, and the green line shows the total prescription drug beneficiaries. The results show that the prescription drug total beneficiaries are lower than the total beneficiaries in the years 2021 and 2022 with 2.2 and 2.3 million beneficiaries.

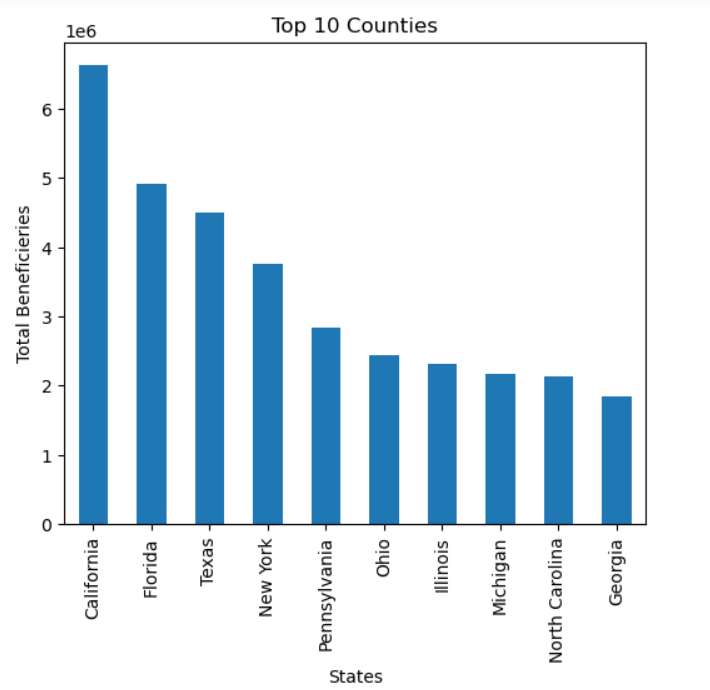


Figure 9. 8 Distribution of Total Beneficiaries across Counties

The above bar graphs show that the top ten states on the x-axis and the total number of beneficiaries on the y-axis provide us with a negative trend. with this analysis, we can observe the total number of beneficiaries enrolled in prescription drug plan Medicare benefits at each state and how the market dynamics at that place act accordingly based on the Enrolment. From the above analysis, we have observed that California state has the highest Enrollment of beneficiaries of about seven million.

### **Factors affecting the dependent variable.**

To find what variables are affecting the dependent variable we ran a few models to understand the significance of the variables.

#### **Linear Regression**

The statistical method Linear regression is a technique used to understand how the independent variable is influenced by the dependent variable. Linear regression draws a straight line and fits the cluster of data points. This slope provides the change in the average dependent variable for the rise in every independent variable the y-intercept shows us the value predicted and the cause of zero. This technique is used to make predictions and estimate the future beneficiary Enrolment, figure out our marketing trends, and predict which Medicare plan has the highest Enrolment. It is important to know that applying linear regression provides us with the data points that fall in a straight line and the other data points that do not fall on the line are preferred as outliers and we need to deal with the outliers and their reasons why they are occurring inside our data. If the data is more inappropriate or if we observe a curved pattern represents that linear regression might not be a good idea to approach.

Below we have applied Linear Regression on Pennsylvania state people who enrolled in Medicare Monthly Enrollment plans.

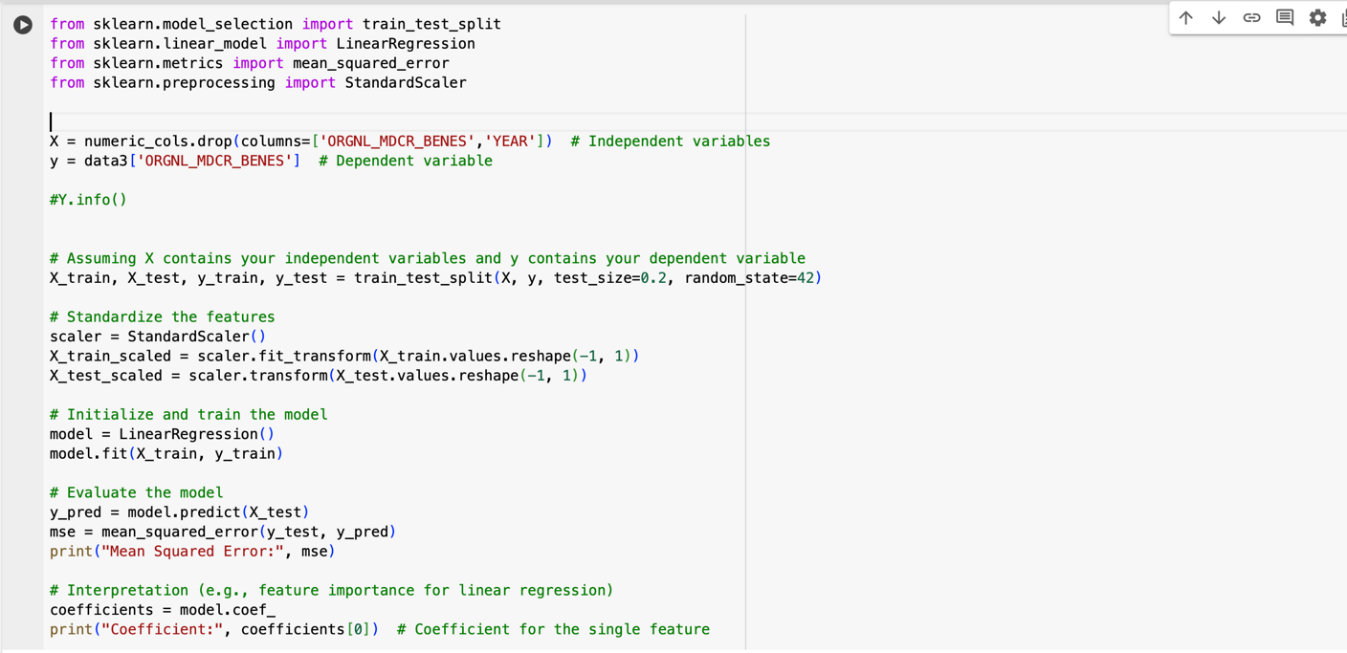


Figure 9. 9 Implementing Linear Regression in Python with Sklearn Library

From the above linear regression results we observe that the error is 1.55\*e^-10 and the coefficient is 2.38\*e^-13. The results appeared to be pretty low and the model doesn't seem to have errors for the data provided.

Checking the regression components and significance using the Ordinary least squares method: 

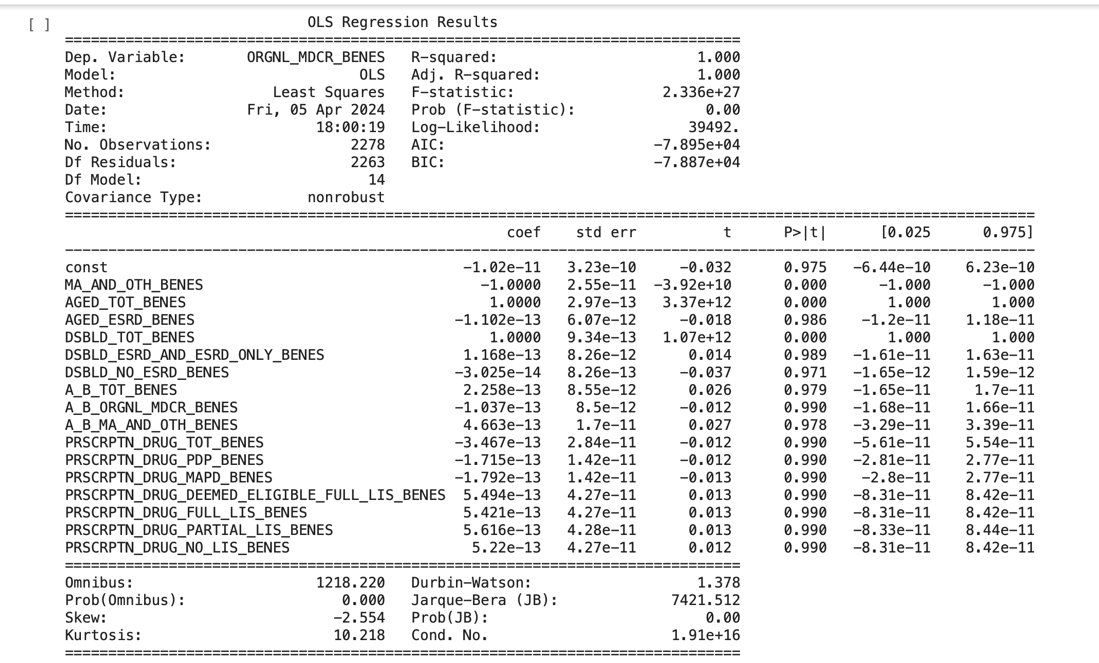


Figure 9. 10 Insights into Regression Analysis

By Checking the R Square the adjusted R Square result is 1. This 1 over here represents the model is 100% variance in data. From the data, we could observe few insignificant variables can be dropped from the model and improve the model and this condition shows that there is the existence of multicollinearity in the data.

The model is rerun to observe whether the performance is improved post dropping those few insignificant variables.

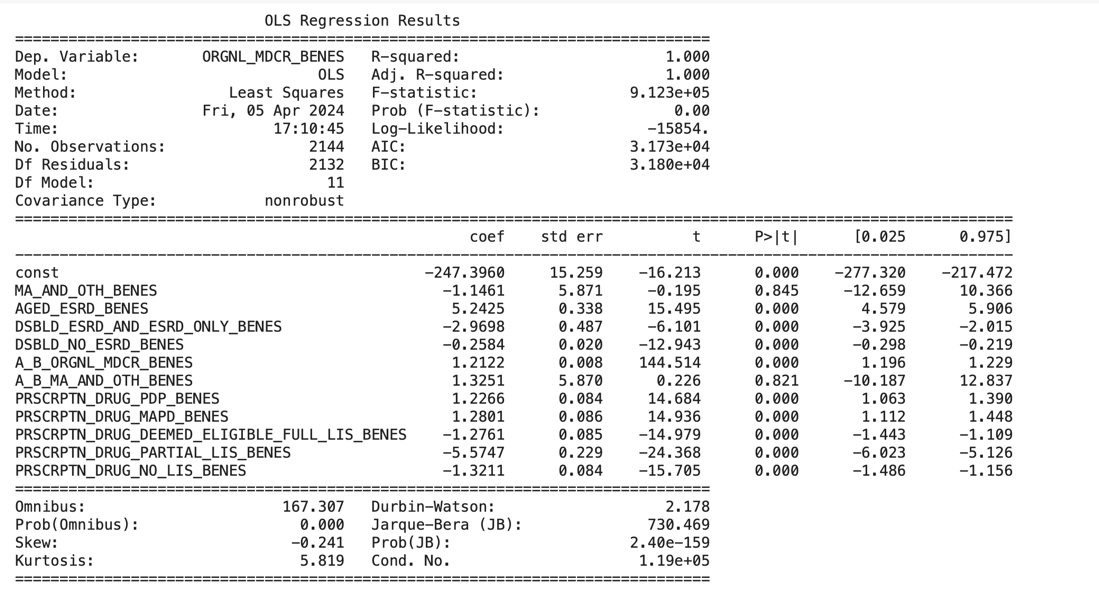


Figure 9. 11 OLS Regression results

#### **Linear Regression with Principal Component Analysis**

Principal Component Analysis, in short, known as PCA, helps to analyze complex datasets with more variables. It helps to summarize the data known as principal components which trap the main features of data. PCA identifies the main data points from the huge data sets and then it simplifies and helps to focus on core features by minimizing the data redundancy between components. The benefits of PCA are: It helps in visualization, exploring data, and noise reduction and plays a vital role in reducing dimensionality for machine learning algorithms. Although PCA has many benefits, it is not the best fit for Highly nonlinear scenarios as it assumes the data to be linear in structure.

We have applied Principal Component Analysis for Medicare Monthly Enrollment data to observe Dimensionality Reduction.



Figure 9. 12 Implementing PCA analysis with the Sklearn library

From the above results, we can observe there is still regression interpreted and the R square value is 1 which represents 100% variance in data.

Now, let us work with the other model and observe whether the model is fit.

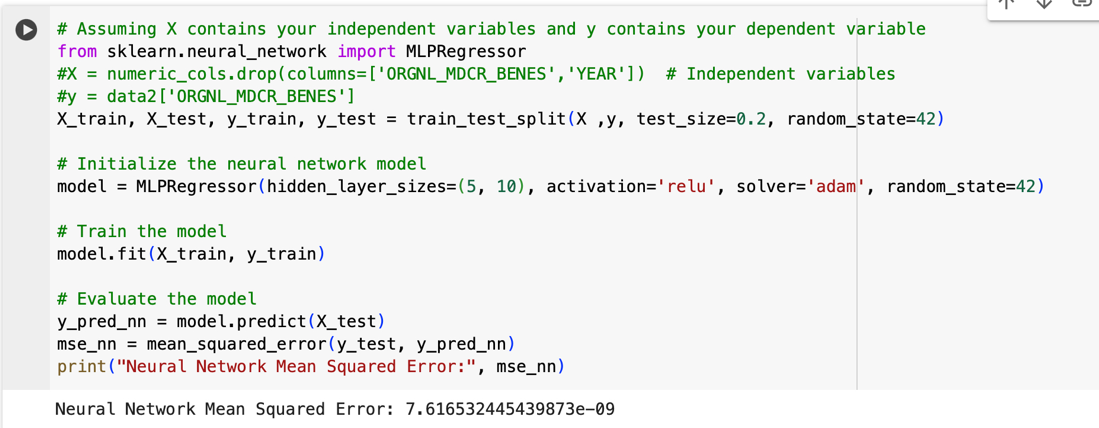


Figure 9. 13 Evaluating Models

Based on the observations above, it seems the error in the Medicare Monthly Enrollment dataset is relatively insignificant. However, this error can lead to overfitting of the model, thereby representing a limitation of the dataset. To mitigate this issue, alternative models such as decision trees or gradient boosting could be explored. Additionally, further analysis may be warranted, particularly if more effective feature selection techniques are implemented.

#### **Decision Tree Algorithm**



Figure 9. 14 Python code and output for Decision Tree model

Based on the above result the code evaluates the performance of the model on the testing set using the mean\_squared\_error function from sklearn.metrics. The mean squared error (MSE) is a common metric for regression problems that measures the average squared difference between the predicted values and the actual values, and the error is said to be high with a value of 15472 which can be considered as the model does not fit with the data.

#### **KNN Algorithm**



Figure 9. 15 Implementing a k-NN Regressor to Predict and Evaluate Data

From the implementation of KNN algorithm, we can say that this can be better model than Decision Tree algorithm with the less error but not the model that can be used for the data as the error of the model stands out to be 4552 which is considered as the very high value.

#### **Gradient Boosting Algorithm**

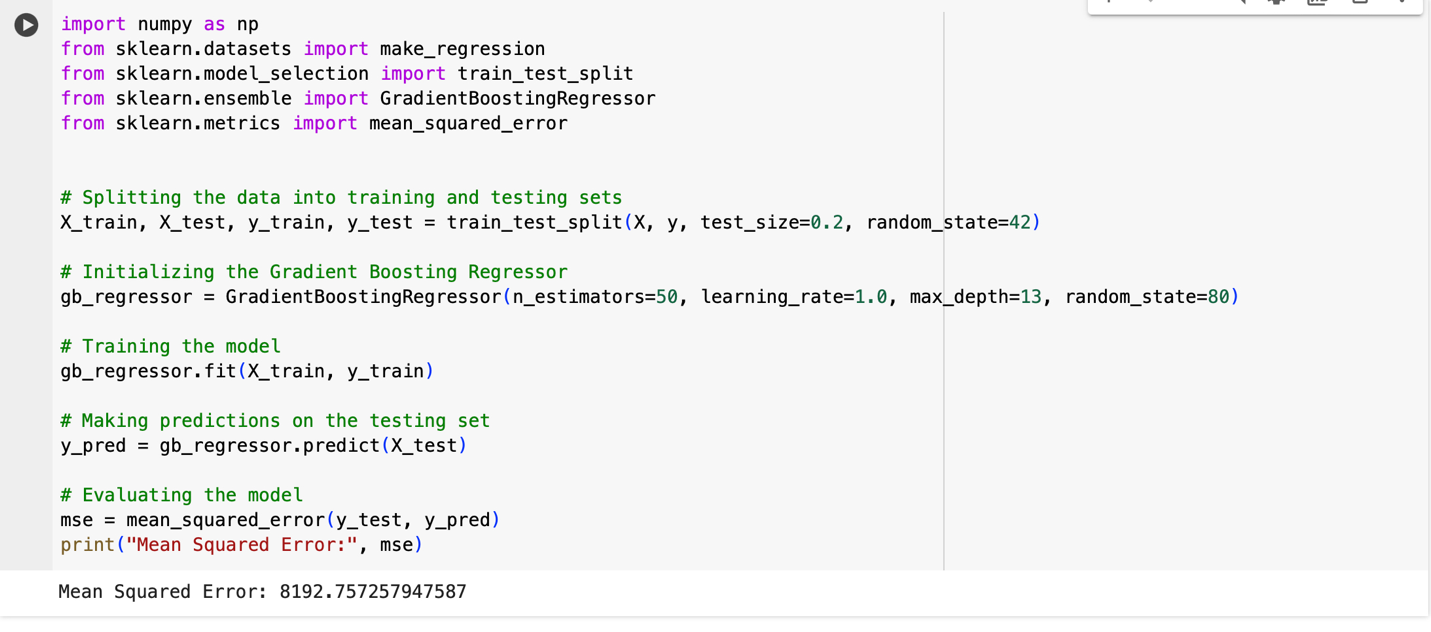


Figure 9. 16 Python code for training and evaluating a Gradient Boosting Regressor model using scikit-learn

The above model interprets that the error obtained is same as that of KNN algorithm but slightly greater than Decision tree algorithm. However, we will not be considering this model as the error for this model is obtained as 8192 which is very high for the model to be fit for the dataset.

# **RESULTS**

## **Research Findings**

Table 10. 1

Comparison of Mean Square Error (MSE) for various Machine Learning Algorithms

|  |  |
| --- | --- |
| **Algorithm** | **Mean Square Error** |
| Linear Regression (Without PCA analysis) | 2.454 |
| Linear Regression (With PCA analysis) | 1.323 |
| Decision Tree Algorithm | 15472.465 |
| KNN Algorithm | 4552.40 |
| Gradient Boosting Algorithm | 8192.34 |

Among all the models, based on least mean square errors, Linear Regression with PCA is performing better for the data. Other models having much higher errors, are not a good fit. The above tables implies that linear regression is a better model for predicting Medicare enrollments in any location.

# **DISCUSSIONS AND LIMITATIONS**

## **Discussions**

The results show that linear regression is the exact model for this dataset to predict the factors that affect beneficiaries enrolled in orginal\_medicare\_benefits. However, after considering all the results through various machine learning algorithms like decision tree, neural network models It clarifies a clear picture of how different factors are changing according to the different variables. The evidence from previous research which was done by other organizations was observing the trends on how the enrollment rate was affected in different geographical regions. Additionally, in our research, we included the top counties with the highest enrollment rate of specific Medicare benefits. The breakdowns of these demographics and the use of these machine learning models to predict future trends are helpful for policymakers. They can use this information to get ahead of the curve and ensure everyone in Medicare has the support they need.

## **Limitations**

The Constraints in our data include:

* + 1. In the process of finding the factors that can affect the dependent variable, it seems our data is only fit for the linear regression model and the regression results range to 100% accuracy. However, In Practicality, no data is completely accurate, and our model suggests potential overfitting.
    2. Our data spans a specific time frame from 2019 to 2021 which is limited.
    3. The data primarily addresses more about Prescription drug coverage rather than other healthcare benefits.
    4. Since the data is in real time, the data did not disclose any information regarding the cost of the healthcare benefits.

# **CONCLUSION AND RECOMMENDATIONS**

## **Conclusion**

This research delved deep into Medicare enrollment data in detail, which led the path to understanding how the country's healthcare system is changing. This research offers useful insight for creating various healthcare initiatives by thoroughly analyzing the various demographics, trends, and healthcare benefits among the recipients. By utilizing the methods of several machine learning models, the results not only provide insight into the enrollment patterns of the present but also provide predictions for upcoming trends. This makes it possible for policymakers to meet the demands of the expanding Medicare population.

With the identification of states with the highest enrolment rates for various Medicare programs, the analysis proceeded forward from the mere national trends. Moving ahead with such granularity helped us to have a clear perspective on the disparities in accessibility and varied regional healthcare needs. The report, by integrating predictive results, provides policymakers with targeted solutions and tools that specifically address the challenges faced by different regions.

The comprehensive analysis delivers an important advancement towards the data-driven healthcare policy. To ensure the well-being of Medicare enrollees, the research work provides a national roadmap that combines cutting-edge modeling approaches with empirical findings. This strategy paved the path for sensible policy decisions and observable improvements to the healthcare delivery system.

## **Recommendations**

From the conclusions, we can have a clear idea of how the advancements of healthcare policies and Policymakers tailoring interventions to address the needs of diverse regions within Medicare landscape by different Predictive modeling techniques. The focused strategy aims to provide fair access to healthcare services by allocating resources optimally and improving overall effectiveness in the healthcare delivery system.

Additionally, the change in healthcare policy establishes a foundation for a more comprehensive perspective in shaping the healthcare system by incorporating the most advanced model techniques. In the Future, these kinds of findings can provide a roadmap for improving the well-being of beneficiaries, but also provide a foundation for complex challenges of modern times which are faced by different stakeholders utilizing the power of data analytics.

# **REFERENCE**

Nguyen, K. H., Oh, E. G., Meyers, D. J., Kim, D., Mehrotra, R., & Trivedi, A. N. (2023). Medicare Advantage Enrollment Among Beneficiaries With End-Stage Renal Disease in the First Year of the 21st Century Cures Act. *JAMA*, *329*(10), 810–818. https://doi.org/10.1001/jama.2023.1426

Meyers, D. J., Ryan, A. M., & Trivedi, A. N. (2023). Trends in Cumulative Dis-enrollment in the  
Medicare Advantage Program, 2011-2020. *JAMA Health Forum*, *4*(8), e232717. https://doi.org/10.1001/jamahealthforum.2023.2717

Nancy Ochieng, Jeannie Fuglesten Biniek, Meredith Freed, Anthony Damico, and Tricia Neuman. Medicare Advantage in 2023: Enrolment Update and Key Trends, 2023. *KFF*​. [https://www.kff.org/medicare/issue-brief/medicare-advantage-in-2023-Enrolment-update-and-key-trends/](https://www.kff.org/medicare/issue-brief/medicare-advantage-in-2023-Enrollment-update-and-key-trends/)

Gornick M. (1982). Trends and regional variations in hospital use under Medicare. *Health care financing review*, *3*(3), 41–73. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4191259/>

# **CONTRIBUTION**

Our team, "Group 5" had put a lot of effort and attention to detail to make this research on "Medicare Dynamics: Trends, GEOGRAPHICAL VARIATIONS AND BENEFICIARY INSIGHTS" a success. To overcome hurdles and reach expectations, every team member was essential.

**Donetra Lanjewar**   
Your commitment to understanding the data, cleaning the data, and analyzing the data was crucial and much appreciated by everyone. you have made the data more understandable. Your Report writing skills are very rich and highly professional. Once again, our team appreciates all your hard work and for being an awesome teammate.

**Lakshmi Tatavarty**   
You have proactively solved one of our research questions and actively guided all the team members to examine the patterns in Medicare enrollment across different regions. The analysis you provided helped us to dig more into the data and see the reasons behind the enrollment of beneficiaries. We appreciate you for being an excellent team member.

**Sarvani Kodeboyina**    
You are the reason behind for us to select this dataset. The way you think about healthcare policies and idealogy to provide healthcare services for maximum reach and the way you have analyzed the results and interpreted what model best fits our data was commendable. Our Team appreciates your interpretation skills and thank you for your contribution.

**Sowrabh Tirumala Vinjamuri**   
Your research in delving into data understanding what the factors are affecting Beneficiary Enrollment and focusing on those factors for improving the future Medicare services is awesome. Our team appreciates you for being super enthusiastic and for being the pillar by maintaining an inclusive and cooperative atmosphere by encouraging everyone.

**To every member in Group – 5**   
To every Team Member, you have made yourself a strong team and made this project successful by enthusiastically moving out of your comfort zones. Each member's contribution and effort made our research, providing a positive approach for future Medicare services.