BUAN 6312.004 - Applied Econometrics and Time Series

Analyzing Medicare Data for Multiple Chronic Conditions

Project Report

Group – 7

Keerthana Panyala (KXP220025)

Vishal Kanna Natarajan Manohar (VXN220000)

Aaditya Srinivasan (AXS210464)

Kushagra Rastogi (KXR220031)

Mohana Krishna Venkatesan (MXK220077)

Sai Pavan Egiteela (SXE220021)

ABSTRACT:

This project aims to analyze the MCC dataset from CMS, comprising 44,974 records of Medicare beneficiaries with multiple chronic conditions. It seeks to uncover relationships between demographics, geography, conditions, healthcare utilization, and costs. Predictive models will estimate prevalence, total payments, and readmission rates. Utilizing advanced techniques like statistical modeling and machine learning, the research enhances understanding of managing chronic diseases within Medicare. Performance metrics like MSE will guide model selection. Findings can inform healthcare policy and drive improvements in care delivery for individuals with multiple chronic conditions.

INTRODUCTION:

The Multiple Chronic Conditions (MCC) dataset serves as a crucial resource for gaining in-depth understanding into the intricate healthcare requirements of Medicare beneficiaries dealing with two or more chronic conditions. Its extensive compilation of 44,974 records offers a robust foundation for comprehensive analysis. Sourced from the authoritative Centers for Medicare & Medicaid Services (CMS) Data website, the dataset likely encompasses a diverse array of variables. These variables span demographic factors such as age and gender, clinical data including diagnoses, and crucial healthcare utilization metrics like hospitalizations. By delving into these multifaceted aspects, researchers can conduct thorough examinations of healthcare outcomes and delivery mechanisms. This rich dataset holds promise in shedding light on the complexities of managing chronic conditions within the Medicare population, facilitating informed decision-making and targeted interventions to enhance healthcare quality and efficacy for this vulnerable demographic.

Overall, from the MCC dataset we are going to determine the hospital readmission rate which is determined by hospital rates. By analyzing this data, stakeholders can work towards developing more targeted and effective strategies for improving healthcare delivery and outcomes of individuals with multimorbidity.

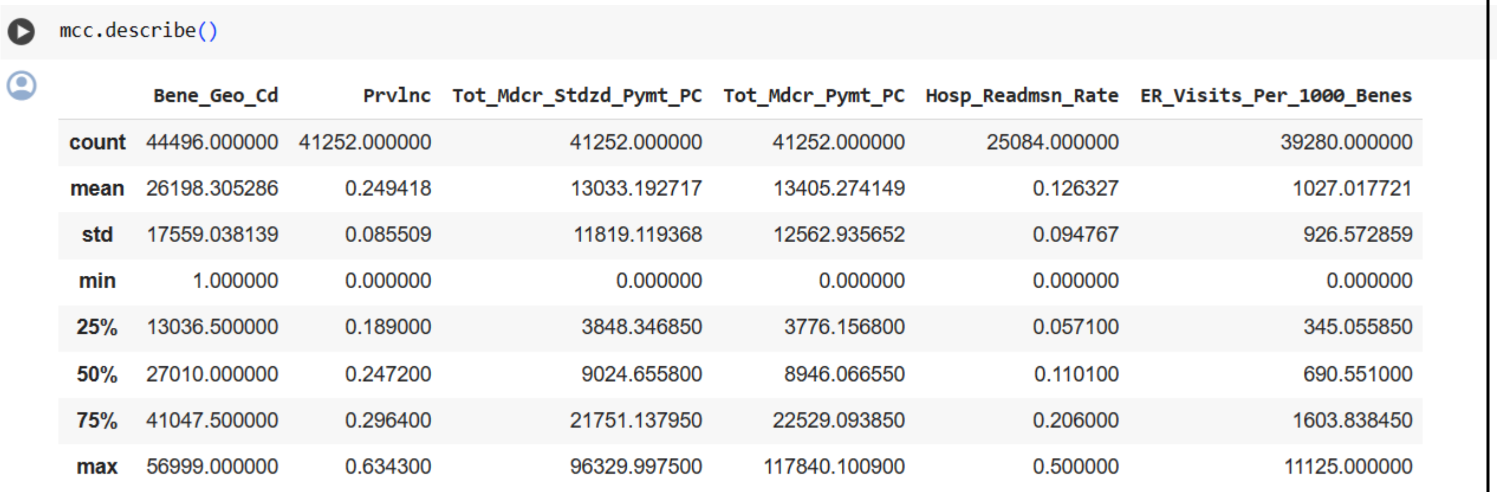
OBJECTIVE:

The objective of this project is to analyze the MCC dataset to understand the relationships between beneficiary demographics, geographic location, chronic conditions, and healthcare utilization and costs and build an efficient model to predict the readmission rate, total payments and hospital readmission rates with the variables available in the dataset thereby accurately predicting the future values.

DATA:

Summary Statistics:

The following table gives a summary of the variables of interest like the mean, median, standard deviation, min value, max value etc.



```
mcc.describe()
```

	Bene_Geo_Cd	Prvlnc	Tot_Mdcr_Stdzd_Pymt_PC	Tot_Mdcr_Pymt_PC	Hosp_Readmsn_Rate	ER_Visits_Per_1000_Benes
count	44496.000000	41252.000000	41252.000000	41252.000000	25084.000000	39280.000000
mean	26198.305286	0.249418	13033.192717	13405.274149	0.126327	1027.017721
std	17559.038139	0.085509	11819.119368	12562.935652	0.094767	926.572859
min	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	13036.500000	0.189000	3848.346850	3776.156800	0.057100	345.055850
50%	27010.000000	0.247200	9024.655800	8946.066550	0.110100	690.551000
75%	41047.500000	0.296400	21751.137950	22529.093850	0.206000	1603.838450
max	56999.000000	0.634300	96329.997500	117840.100900	0.500000	11125.000000

Data Visualization:

Analysis by Total Medicare Payments

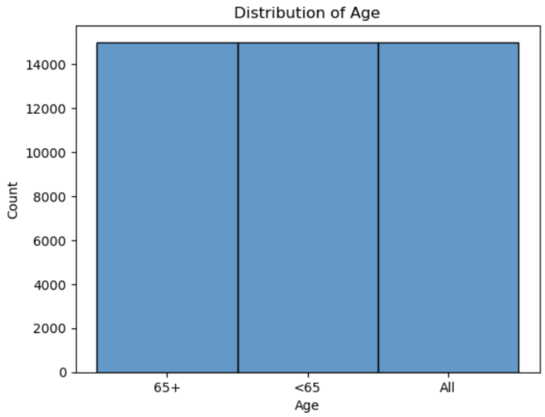


Fig 1: Distribution of Age

From the above figure, we can observe that the distribution of age is uniform. That is the number of people in the dataset with age greater than 65 is equal to the number of people with age less than 65.

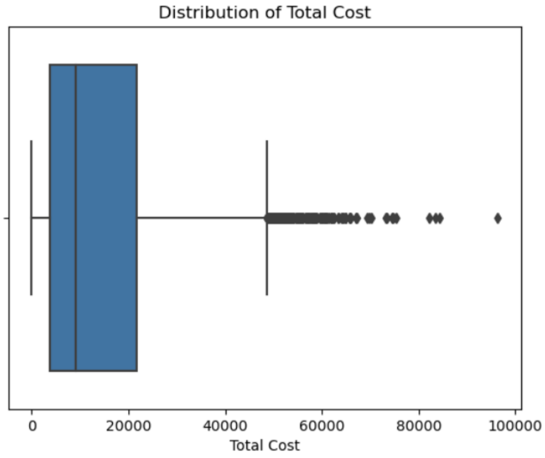


Fig 2: Distribution of Medicare Payments

Fig 2 represents the Total Medicare payments distribution. The box extends from the lower quartile (Q1) of about 50000 to the upper quartile (Q3) 20000, with a line inside the box representing the median i.e., around 10000.

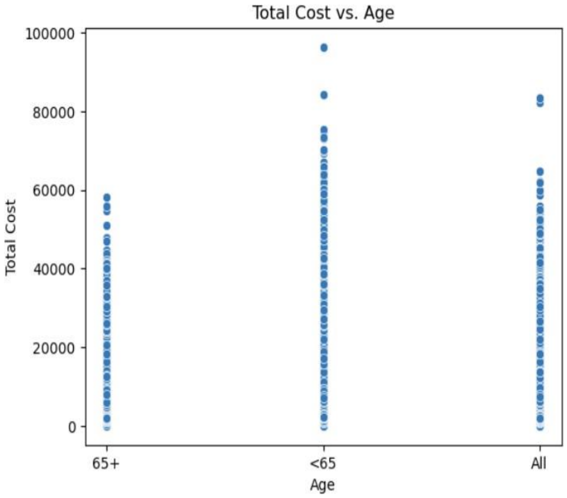


Fig 3: Total Cost vs Age

Fig 3 represents a scatter plot to visualize the relationship between the total cost of medical payments and the age of beneficiaries in the dataset. The x-axis represents age level and the y-axis represents the total cost of medical payments standardized by Medicare.

The scatter plot shows a slightly negative relationship between age and total cost. As age increases, the total cost of medical payments seems to decrease, although there is a lot of variability in the data.

Analysis by Prevalence

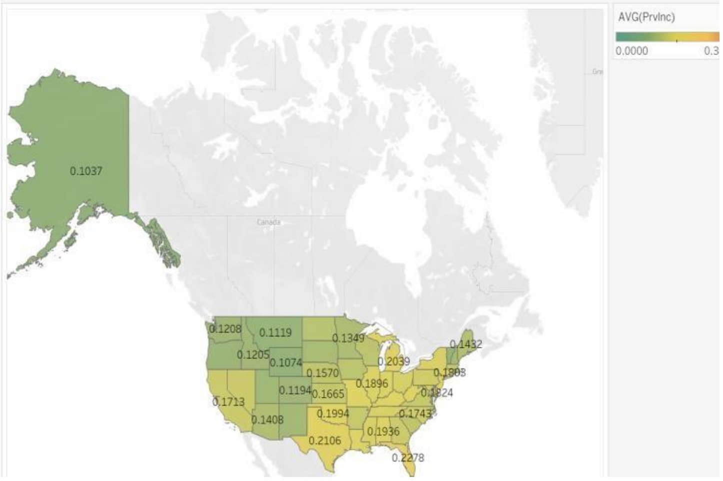


Fig 4: Total Cost vs Age

Fig 4 represents the geographical representation of prevalence for MCC> 6. The average prevalence was found to be highest in Florida with about 22.78% and lowest in Alaska with 10.37%

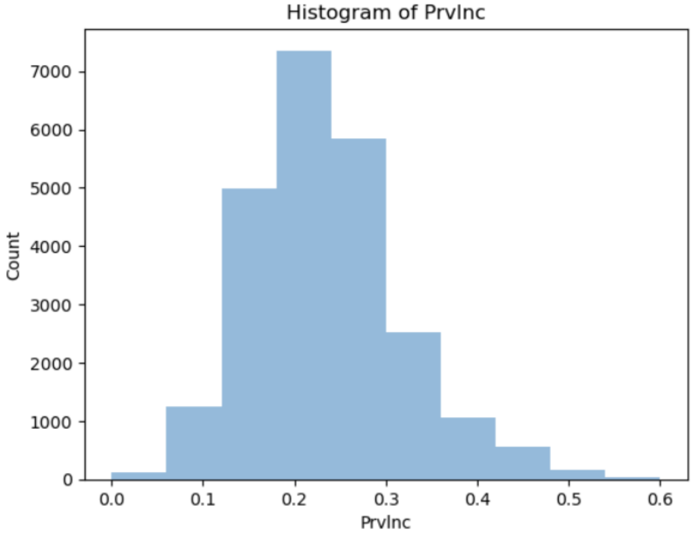


Fig 5: Histogram of Prevalence

Fig 5 represents histogram of prevalence. From the figure we can observe that the prevalence ranges from 0.2 to 0.3 for most of the states and counties.

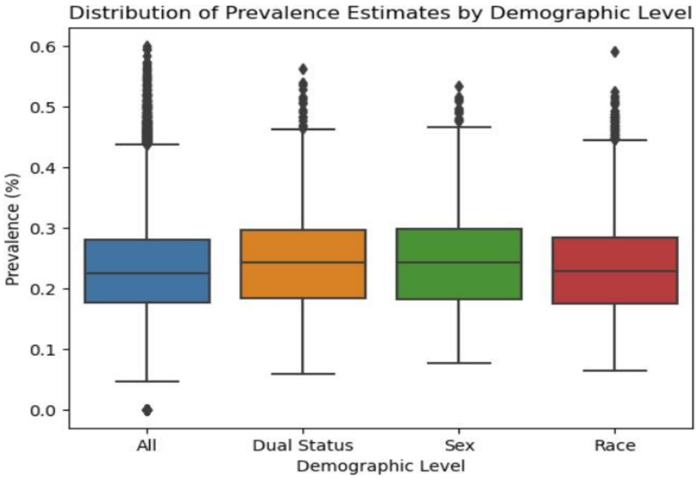


Fig 6: Prevalence vs demographic factors

Analysis by Hospital Readmission rate and ER visits

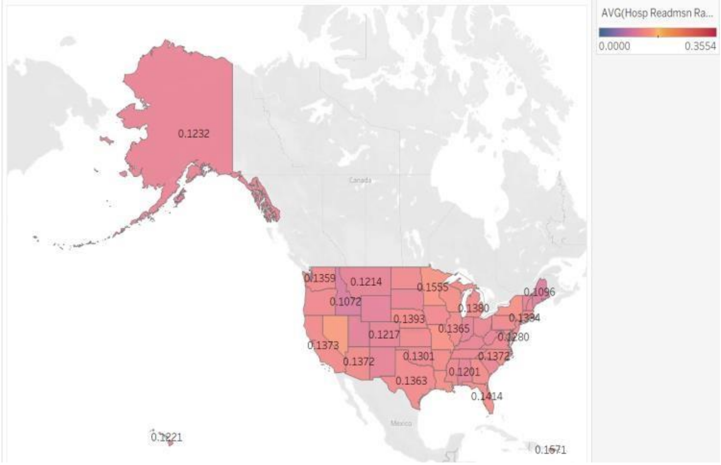


Fig 7: Prevalence vs demographic factors

Fig 7 represents the geographical representation of the average hospital readmission rate. From the figure it can be observed that the average hospital readmission rate is the highest for District of Columbia with about 16.32% and the lowest for Idaho with 10.7%.

Figure 6 represents side by side boxplots of prevalence vs various demographic factors Dual Status, Sex and race. It can be observed that the distributions are uniform irrespective of the demographic factors and the distribution is almost similar for the demographic factors Dual Status and Sex.

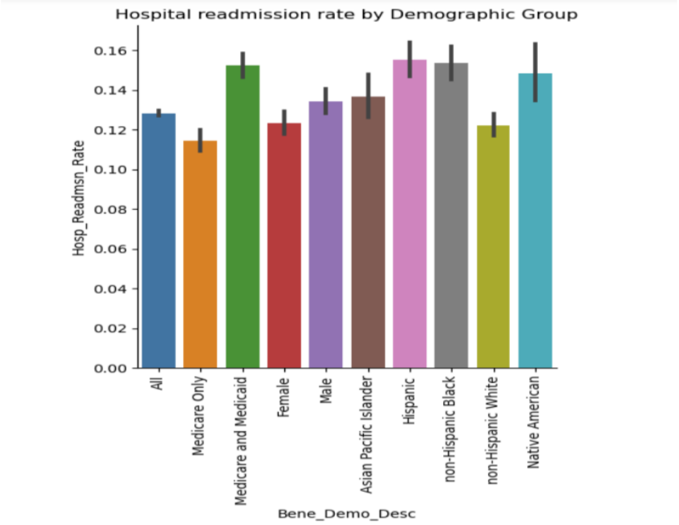


Fig 8: Hospital readmission vs demographic factors

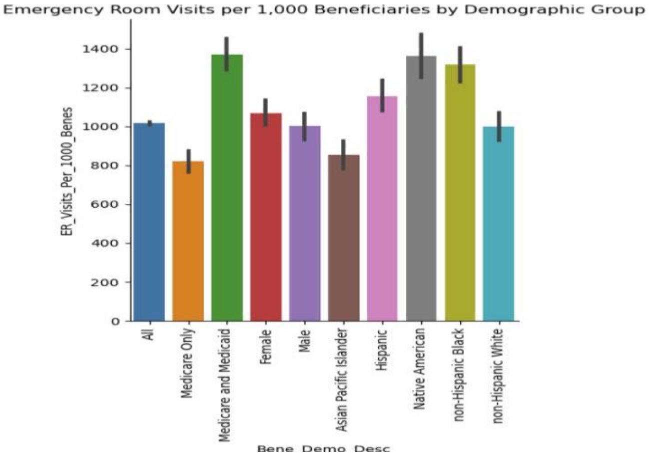
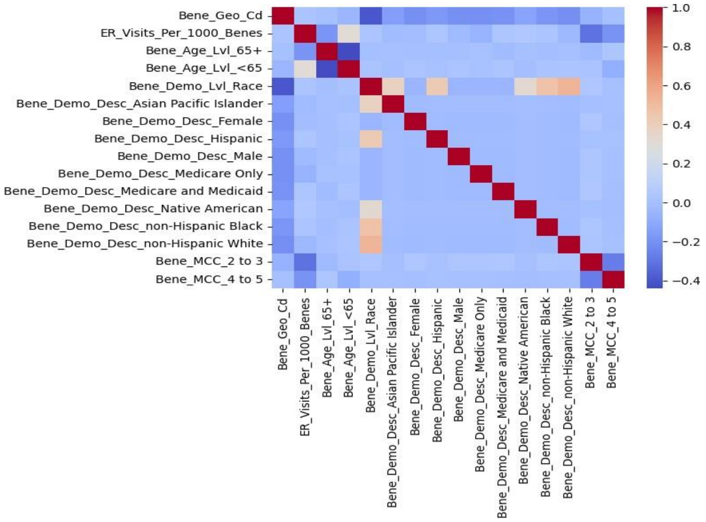


Fig 9: Emergency visits vs demographic factors

Fig 8 and 9 represent Hospital readmission rate and Emergency visits vs demographic factors. The hospital readmission rates are high for Hispanics and the lowest for Medicare only. Emergency visits are the highest for Native Americans and the lowest for Medicare people.

Data Cleaning and Preprocessing:



As a first step duplicate rows were removed. The next step was to remove any rows with missing values. To handle categorical variables, one-hot encoding is used. The get_dummies() method is used to create dummy variables for categorical variables such as Bene_Age_Lvl, Bene_Demo_Lvl, Bene_Demo_Desc, and Bene_MCC. The resulting dummy variables are concatenated with the original dataset. This step is essential to convert categorical variables into a format that can be processed by machine learning algorithms. Overall, data preprocessing is an essential step in any data analysis pipeline to ensure that the dataset is cleaned and ready for further analysis.

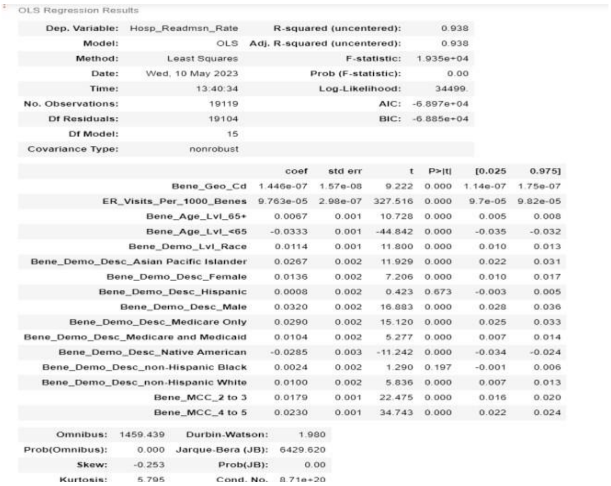
DATA MODELS:

Models to predict Hospital Readmission rate:

Before using models on the data, heatmaps were generated to test multicollinearity and to remove the highly correlated variables as it will result in incorrect predictions. Below are the heatmaps after removing the highly correlated variables.

Ordinary Least squares (OLS)

This output presents the results of an OLS regression model predicting the dependent variable, "Hosp_Readmsn_Rate", using 15 independent variables. Based on 19,119 observations, the model demonstrates high explanatory power, with an R-squared value of 0.938, indicating a significant proportion of variance explained. The adjusted R-squared also supports this, suggesting no overfitting. The F-statistic (1.935e+04) and its associated p-value (0.00) indicate the model's overall statistical significance. Coefficients, standard errors, t-statistics, and p-values for each independent variable are provided, showing their significance and direction of influence. For example, "ER_Visits_Per_1000_Benes" and "Bene_Age_Lvl_65+" show statistically significant effects on "Hosp_Readmsn_Rate". Assumptions regarding normality of residuals and independence of errors are assessed using various tests. Overall, the OLS results suggest a well-fitting model for predicting "Hosp_Readmsn_Rate" based on the included independent variables. Below is the OLS output generated.



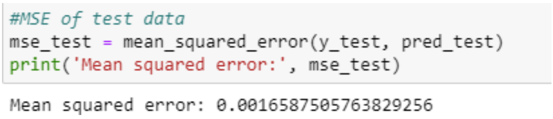
OLS Regression Results

Dep. Variable:	Hosp_Readmsn_Rate	R-squared (uncentered):	0.938
Model:	OLS	Adj. R-squared (uncentered):	0.938
Method:	Least Squares	F-statistic:	1.935e+04
Date:	Wed, 10 May 2023	Prob (F-statistic):	0.00
Time:	13:40:34	Log-Likelihood:	34499.
No. Observations:	19119	AIC:	-6.897e+04
Df Residuals:	19104	BIC:	-6.885e+04
Df Model:	15		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Bene_Geo_Cd	1.446e-07	1.57e-08	9.222	0.000	1.14e-07	1.75e-07
ER_Visits_Per_1000_Benes	9.763e-05	2.98e-07	327.516	0.000	9.7e-05	9.82e-05
Bene_Age_Lvl_65+	0.0067	0.001	10.728	0.000	0.005	0.008
Bene_Age_Lvl_<65	-0.0333	0.001	-44.842	0.000	-0.035	-0.032
Bene_Demo_Lvl_Race	0.0114	0.001	11.800	0.000	0.010	0.013
Bene_Demo_Desc_Asian Pacific Islander	0.0267	0.002	11.929	0.000	0.022	0.031
Bene_Demo_Desc_Female	0.0136	0.002	7.206	0.000	0.010	0.017
Bene_Demo_Desc_Hispanic	0.0008	0.002	0.423	0.673	-0.003	0.005
Bene_Demo_Desc_Male	0.0320	0.002	16.883	0.000	0.028	0.036
Bene_Demo_Desc_Medicare Only	0.0290	0.002	15.120	0.000	0.025	0.033
Bene_Demo_Desc_Medicare and Medicaid	0.0104	0.002	5.277	0.000	0.007	0.014
Bene_Demo_Desc_Native American	-0.0285	0.003	-11.242	0.000	-0.034	-0.024
Bene_Demo_Desc_non-Hispanic Black	0.0024	0.002	1.290	0.197	-0.001	0.006
Bene_Demo_Desc_non-Hispanic White	0.0100	0.002	5.836	0.000	0.007	0.013
Bene_MCC_2 to 3	0.0179	0.001	22.475	0.000	0.016	0.020
Bene_MCC_4 to 5	0.0230	0.001	34.743	0.000	0.022	0.024

Omnibus:	1459.439	Durbin-Watson:	1.980
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6429.620
Skew:	-0.253	Prob(JB):	0.00
Kurtosis:	5.795	Cond. No.	8.71e+20

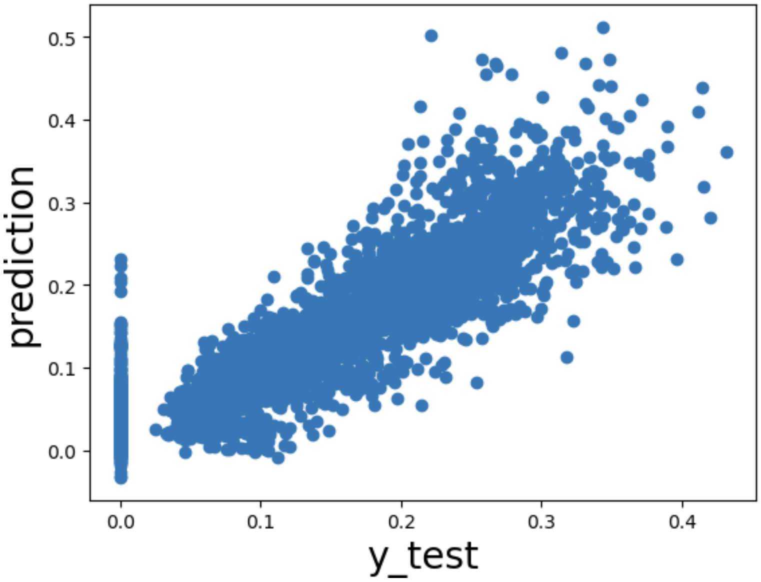
Prediction and Accuracy Measures



```
#MSE of test data
mse_test = mean_squared_error(y_test, pred_test)
print('Mean squared error:', mse_test)
```

```
Mean squared error: 0.0016587505763829256
```

The mean square error for OLS model for test data is 0.0016587.
Below is plot of predicted values vs actual values.

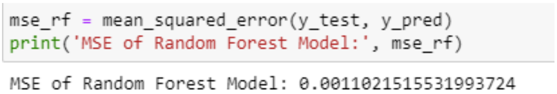


Random Forest Model

The random forest model is a machine learning algorithm used to predict hospital readmission rates based on a set of predictor variables. The model works by creating multiple decision trees and then combining their predictions to arrive at a final prediction. The model is highly accurate and robust, making prediction tasks.

In this particular model, the predictor variables used include beneficiary geographic code, emergency room visits per 1000 beneficiaries, beneficiary age level, beneficiary demographic level, and beneficiary comorbidity level. The model was trained on a dataset consisting of 19,119 observations, with the hospital readmission rate as the dependent variable.

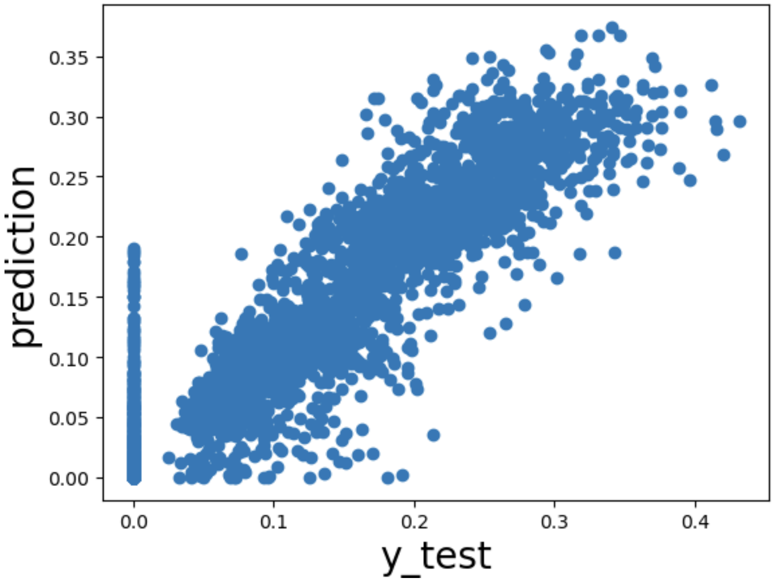
Prediction and Accuracy Measures



```
mse_rf = mean_squared_error(y_test, y_pred)
print('MSE of Random Forest Model:', mse_rf)
```

```
MSE of Random Forest Model: 0.0011021515531993724
```

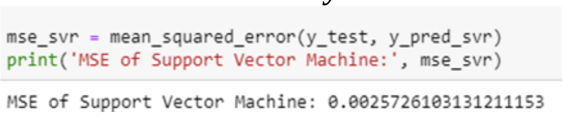
The mean square error for Random Forest model for test data is 0.00110215. Below is plot of predicted values vs actual values.



Support Vector Regression

Support Vector Regression (SVR) is a popular machine learning technique for regression problems. In this case, it is being used to predict hospital readmission rates. The code first scales the data using the StandardScaler function from the preprocessing module of the scikit-learn library. This is done to ensure that all the features have the same scale and are centered around zero. Next, an SVR model is created with a radial basis function kernel (kernel='rbf') and hyperparameters of C=10, gamma=0.1, and epsilon=0.1. These hyperparameters are used to control the tradeoff between the model's ability to fit the training data and its ability to generalize to new data. Finally, the model is trained on the scaled training data using the fit() function. Once trained, the model was used to make predictions on new data. Overall, SVR is a powerful technique that can be used to accurately predict hospital readmission rates based on a variety of patient and demographic features.

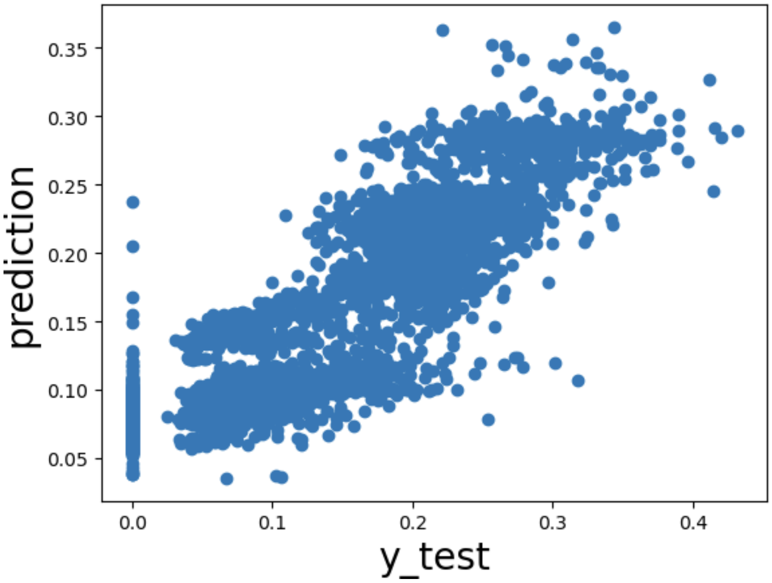
Prediction and Accuracy Measures



```
mse_svr = mean_squared_error(y_test, y_pred_svr)
print('MSE of Support Vector Machine:', mse_svr)
```

```
MSE of Support Vector Machine: 0.002572610313121115З
```

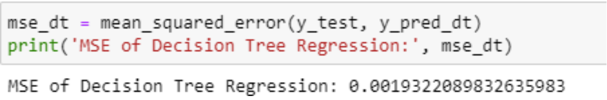
The mean square error for Support vector regression for test data is 0.002576. Below is plot of predicted values vs actual values.



Decision Tree Regression

Decision tree regression is a machine learning technique used for both classification and regression problems. It works by recursively partitioning the data into subsets based on the values of the independent variables. The algorithm selects the variable that provides the best split and then splits the data accordingly. This process is repeated until a stopping criterion is met. In the case of regression, the decision tree algorithm generates a tree where each node represents a decision rule and each leaf node represents a prediction.

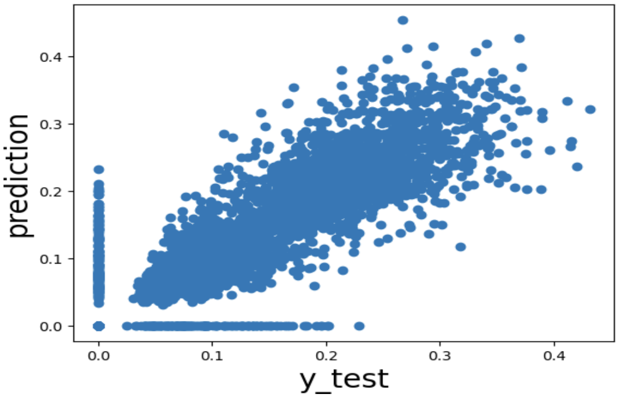
Prediction and Accuracy Measures



```
mse_dt = mean_squared_error(y_test, y_pred_dt)
print('MSE of Decision Tree Regression:', mse_dt)
```

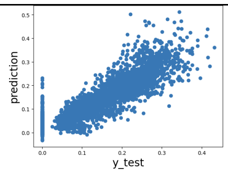
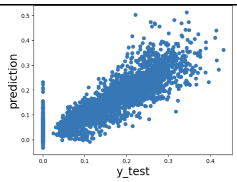
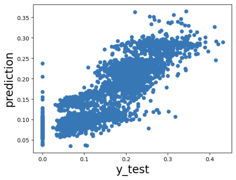
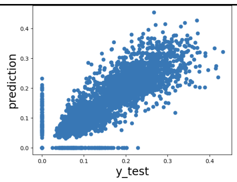
```
MSE of Decision Tree Regression: 0.0019322089832635983
```

The mean square error for Decision tree regression for test data is 0.001932. Below is plot of predicted values vs actual values.



Comparison of models to predict Hospital readmission rate:

Based on the above results Random Forest Model gives the best result for the prediction of Hospital Readmission rate.

Model name	**OLS**	**Random Forest regression**	**Support Vector regression**	**Decision Tree regression**
MSE	0.0016587	0.00110215	0.002576	0.001932
Actual vs predicted plot				

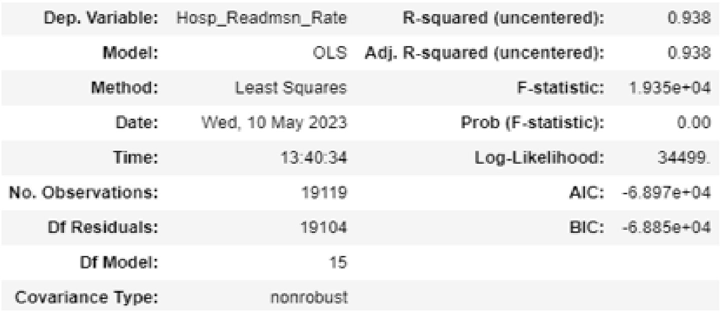
METHODOLOGY:

Endogeneity Considerations: Empirical Challenges: There are several empirical challenges in the regression analysis. One of the main challenges is endogeneity, which occurs when independent variables are correlated with the error term. For example, ER_Visits_Per_1000_Benes may be endogenous if it is influenced by unobserved factors that also affect Hosp_Readmsn_Rate.

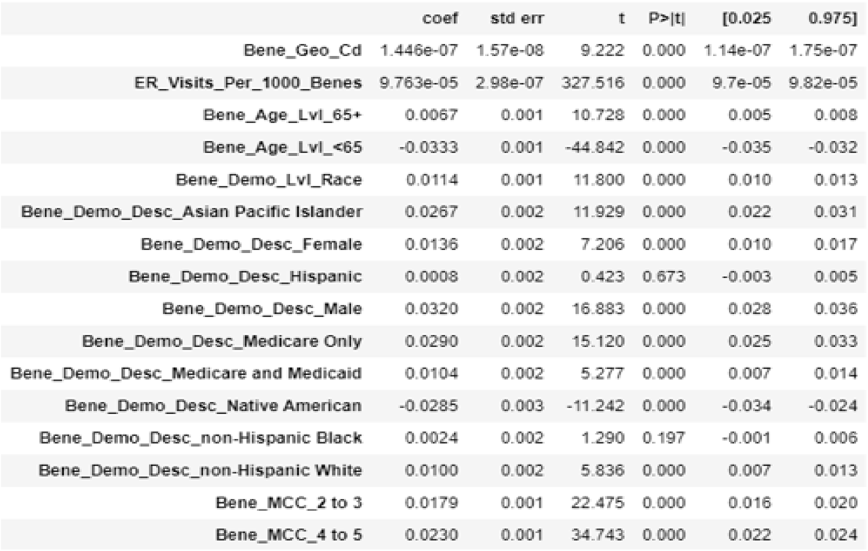
Reasons for Endogeneity: Endogeneity can arise due to various reasons. For instance, ER_Visits_Per_1000_Benes may be influenced by factors such as patients' health status, accessibility to healthcare facilities, and healthcare policy changes. These factors may also affect hospital readmission rates, leading to a correlation between the independent variable and the error term.

Instrumental Variables (IV): Choice of Instruments: In the regression analysis, ER_Visits_Per_1000_Benes is used as an instrument for itself (ER_Visits_Per_1000_Benes). The rationale for using this variable as an instrument is that it is correlated with ER_Visits_Per_1000_Benes but is assumed to be uncorrelated with the error term in the regression model. However, it is essential to ensure that the instrument satisfies the relevance and exclusion restrictions.

Model Estimation: OLS with Instrumental Variables: The ordinary least squares (OLS) regression model is estimated using instrumental variables (IV) to address endogeneity concerns. The instrument parameter is specified to include the instrumental variable (ER_Visits_Per_1000_Benes).



Dep. Variable:	Hosp_Readmsn_Rate	R-squared (uncentered):	0.938
Model:	OLS	Adj. R-squared (uncentered):	0.938
Method:	Least Squares	F-statistic:	1.935e+04
Date:	Wed, 10 May 2023	Prob (F-statistic):	0.00
Time:	13:40:34	Log-Likelihood:	34499.
No. Observations:	19119	AIC:	-6.897e+04
Df Residuals:	19104	BIC:	-6.885e+04
Df Model:	15		
Covariance Type:	nonrobust		



	coef	std err	t	P>|t|	[0.025	0.975]
Bene_Geo_Cd	1.446e-07	1.57e-08	9.222	0.000	1.14e-07	1.75e-07
ER_Visits_Per_1000_Benes	9.763e-05	2.98e-07	327.516	0.000	9.7e-05	9.82e-05
Bene_Age_Lvl_65+	0.0067	0.001	10.728	0.000	0.005	0.008
Bene_Age_Lvl_<65	-0.0333	0.001	-44.842	0.000	-0.035	-0.032
Bene_Demo_Lvl_Race	0.0114	0.001	11.800	0.000	0.010	0.013
Bene_Demo_Desc_Asian Pacific Islander	0.0267	0.002	11.929	0.000	0.022	0.031
Bene_Demo_Desc_Female	0.0136	0.002	7.206	0.000	0.010	0.017
Bene_Demo_Desc_Hispanic	0.0008	0.002	0.423	0.673	-0.003	0.005
Bene_Demo_Desc_Male	0.0320	0.002	16.883	0.000	0.028	0.036
Bene_Demo_Desc_Medicare Only	0.0290	0.002	15.120	0.000	0.025	0.033
Bene_Demo_Desc_Medicare and Medicaid	0.0104	0.002	5.277	0.000	0.007	0.014
Bene_Demo_Desc_Native American	-0.0285	0.003	-11.242	0.000	-0.034	-0.024
Bene_Demo_Desc_non-Hispanic Black	0.0024	0.002	1.290	0.197	-0.001	0.006
Bene_Demo_Desc_non-Hispanic White	0.0100	0.002	5.836	0.000	0.007	0.013
Bene_MCC_2 to 3	0.0179	0.001	22.475	0.000	0.016	0.020
Bene_MCC_4 to 5	0.0230	0.001	34.743	0.000	0.022	0.024

Interpretation: The regression results show that several independent variables are statistically significant predictors of Hosp_Readmsn_Rate. For instance, ER_Visits_Per_1000_Benes has a statistically significant positive coefficient, suggesting that an increase in emergency room visits per 1000 beneficiaries is associated with an increase in hospital readmission rates.

However, it's important to interpret the coefficients cautiously, considering the potential endogeneity of the variables and the validity of the instrumental variable approach.

The OLS regression model shows a strong explanatory power (R-squared = 0.825) in predicting "Hosp_Readmsn_Rate", but the variable "Bene_Age_Lvl_65+" lacks statistical significance, suggesting potential model refinement. Further investigation into alternative model specifications or additional variables may enhance predictive accuracy.

The robust regression model demonstrates statistical significance for "Prvlnc", "Tot_Mdcr_Stdzd_Pymt_PC", and "Tot_Mdcr_Pymt_PC" predictors in predicting "Hosp_Readmsn_Rate", with coefficients of -0.0700, 5.797e-06, and 7.46e-07 respectively.

A recommendation for further analysis could involve investigating potential multicollinearity among predictors to ensure the stability of the model results.

Endogeneity Assessment:

Potential Issues: Despite using instrumental variables, there may still be endogeneity issues if the instrument is weakly correlated with the endogenous variable or if there are unobserved confounding factors.

Weaknesses of the Approach: The instrumental variable approach relies on the assumption that the instruments are exogenous and satisfy the relevance and exclusion restrictions. Violation of these assumptions can lead to biased estimates and undermine the validity of the results.

FURTHER CONSIDERATIONS:

Multicollinearity: The regression results indicate the presence of strong multicollinearity problems, as evidenced by the large condition number. Multicollinearity can affect the stability of the estimates and the interpretation of the coefficients.

Robustness Checks: Conducting sensitivity analyses and robustness checks can help assess the robustness of the results to different model specifications and assumptions.

RESULTS:

Different Model Specifications:

Alternative Specification 1

The R-squared value is 0.846, indicating that the model explains about 84.6% of the variance in the dependent variable.

All three independent variables (Prvlnc, Tot_Mdcr_Stdzd_Pymt_PC, ER_Visits_Per_1000_Benes) are statistically significant ($p < 0.05$) in predicting Hosp_Readmsn_Rate.

The coefficient for Tot_Mdcr_Stdzd_Pymt_PC is 4.301e-06, indicating that a one-unit increase in standardized Medicare payments per capita is associated with an increase in hospital readmission rate by 4.301e-06 units.

Alternative Specification 2

The R-squared value is 0.838, slightly lower than the previous model.

Similar to the first alternative, all three independent variables are statistically significant.

The coefficient for Tot_Mdcr_Pymt_PC is 3.676e-06.

Both specifications provide similar insights, indicating the robustness of the results across different model specifications.

Subset Analysis:

Subset Analysis - Age Group: 65+

The R-squared value is 0.868, indicating a good fit for the model.

Both Prvlnc and ER_Visits_Per_1000_Benes are statistically significant predictors of Hosp_Readmsn_Rate.

Subset Analysis - Age Group: Under 65

The R-squared value is 0.789, indicating a slightly lower explanatory power compared to the 65+ age group.

Both predictors (Prvlnc and ER_Visits_Per_1000_Benes) are statistically significant.

Heteroscedasticity-Robust Standard Errors:

The model with heteroscedasticity-robust standard errors provides similar coefficients and significance levels for the variables compared to the non-robust model.

Bootstrapping:

The model with bootstrapped standard errors also yields similar coefficients and significance levels compared to the non-robust model.

Robust Regression Model:

The robust regression model, which is less sensitive to outliers, shows similar coefficient estimates but slightly different standard errors compared to the ordinary least squares (OLS) models.

Specification Tests (Ramsey RESET Test):

The Ramsey RESET test statistic is 4326.65 with a p-value of 0.0, indicating that there is evidence to reject the null hypothesis of correct specification. This suggests that there may be some misspecification in the model.

CONCLUSION:

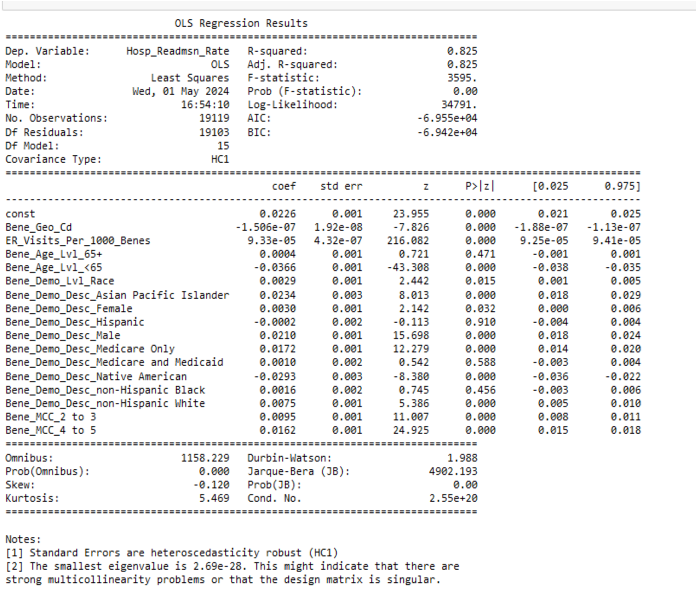
The instrumental variable regression provides insights into the relationship between healthcare utilization factors and hospital readmission rates while addressing potential endogeneity concerns. However, researchers should remain cautious of the assumptions underlying the instrumental variable approach and conduct thorough diagnostics to ensure the validity of the results. Additionally, addressing multicollinearity issues and conducting robustness checks are essential steps in validating the findings of the regression analysis.

REFERENCES:

https://data.cms.gov/medicare-chronic-conditions/multiple-chronic-conditions
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8940207/

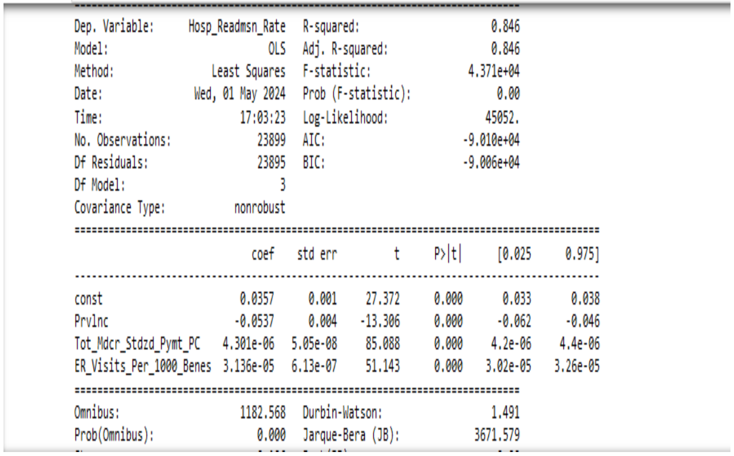
APPENDIX:

1A Endogenity Test



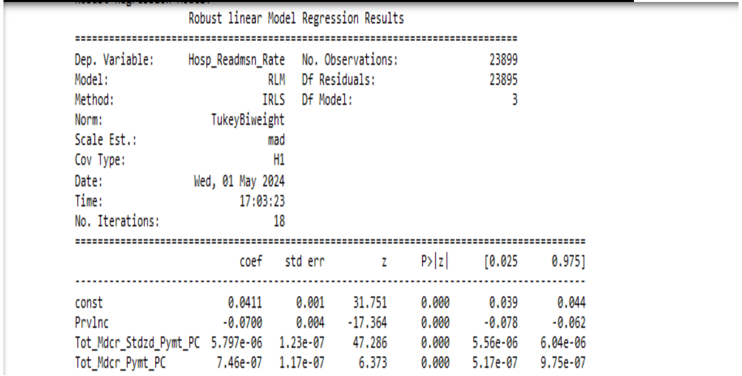
```
                            OLS Regression Results
==============================================================================
Dep. Variable:      Hosp_Readmsn_Rate   R-squared:                       0.825
Model:                            OLS   Adj. R-squared:                  0.825
Method:                 Least Squares   F-statistic:                     3595.
Date:                Wed, 01 May 2024   Prob (F-statistic):               0.00
Time:                        16:54:10   Log-Likelihood:                 34791.
No. Observations:               19119   AIC:                        -6.955e+04
Df Residuals:                   19103   BIC:                        -6.942e+04
Df Model:                          15
Covariance Type:                  HC1
=========================================================================================================
                                           coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------------------------------
const                                    0.0226      0.001     23.955      0.000       0.021       0.025
Bene_Geo_Cd                          -1.506e-07   1.92e-08     -7.826      0.000   -1.88e-07   -1.13e-07
ER_Visits_Per_1000_Benes               9.33e-05   4.32e-07    216.082      0.000    9.25e-05    9.41e-05
Bene_Age_Lvl_65+                         0.0004      0.001      0.721      0.471      -0.001       0.001
Bene_Age_Lvl_<65                        -0.0366      0.001    -43.308      0.000      -0.038      -0.035
Bene_Demo_Lvl_Race                       0.0029      0.001      2.442      0.015       0.001       0.005
Bene_Demo_Desc_Asian Pacific Islander    0.0234      0.003      8.013      0.000       0.018       0.029
Bene_Demo_Desc_Female                    0.0030      0.001      2.142      0.032       0.000       0.006
Bene_Demo_Desc_Hispanic                 -0.0002      0.002     -0.113      0.910      -0.004       0.004
Bene_Demo_Desc_Male                      0.0210      0.001     15.698      0.000       0.018       0.024
Bene_Demo_Desc_Medicare Only             0.0172      0.001     12.279      0.000       0.014       0.020
Bene_Demo_Desc_Medicare and Medicaid     0.0010      0.002      0.542      0.588      -0.003       0.004
Bene_Demo_Desc_Native American          -0.0293      0.003     -8.380      0.000      -0.036      -0.022
Bene_Demo_Desc_non-Hispanic Black        0.0016      0.002      0.745      0.456      -0.003       0.006
Bene_Demo_Desc_non-Hispanic White        0.0075      0.001      5.386      0.000       0.005       0.010
Bene_MCC_2 to 3                          0.0095      0.001     11.007      0.000       0.008       0.011
Bene_MCC_4 to 5                          0.0162      0.001     24.925      0.000       0.015       0.018
==============================================================================
Omnibus:                     1158.229   Durbin-Watson:                   1.988
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             4902.193
Skew:                          -0.120   Prob(JB):                         0.00
Kurtosis:                       5.469   Cond. No.                     2.55e+20
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC1)
[2] The smallest eigenvalue is 2.69e-28. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

1B (i) Robustness Check – Different Model Specifications



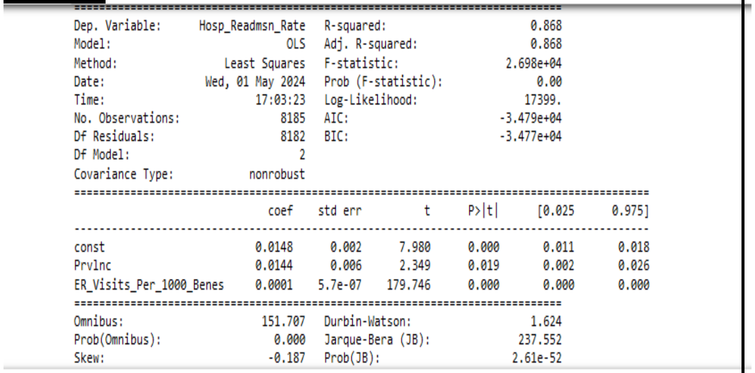
```
==============================================================================
Dep. Variable:      Hosp_Readmsn_Rate   R-squared:                       0.846
Model:                            OLS   Adj. R-squared:                  0.846
Method:                 Least Squares   F-statistic:                 4.371e+04
Date:                Wed, 01 May 2024   Prob (F-statistic):               0.00
Time:                        17:03:23   Log-Likelihood:                 45052.
No. Observations:               23899   AIC:                        -9.010e+04
Df Residuals:                   23895   BIC:                        -9.006e+04
Df Model:                           3
Covariance Type:            nonrobust
============================================================================================
                               coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------
const                        0.0357      0.001     27.372      0.000       0.033       0.038
Prvlnc                      -0.0537      0.004    -13.306      0.000      -0.062      -0.046
Tot_Mdcr_Stdzd_Pymt_PC    4.301e-06   5.05e-08     85.088      0.000     4.2e-06     4.4e-06
ER_Visits_Per_1000_Benes  3.136e-05   6.13e-07     51.143      0.000    3.02e-05    3.26e-05
==============================================================================
Omnibus:                     1182.568   Durbin-Watson:                   1.491
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3671.579
```

1B (ii) Robustness Check – Robust Regression



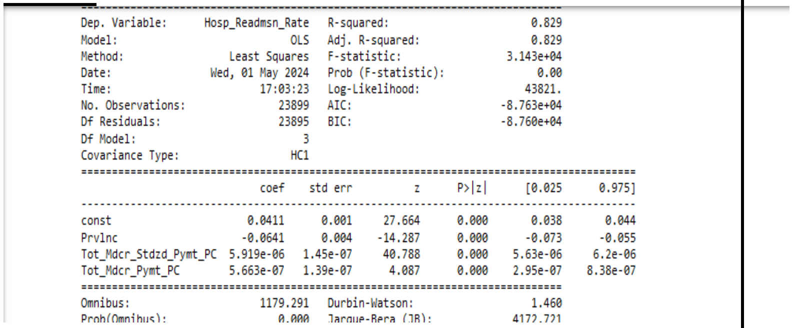
```
                    Robust linear Model Regression Results
==============================================================================
Dep. Variable:      Hosp_Readmsn_Rate   No. Observations:                23899
Model:                            RLM   Df Residuals:                    23895
Method:                          IRLS   Df Model:                            3
Norm:                   TukeyBiweight
Scale Est.:                       mad
Cov Type:                          H1
Date:                Wed, 01 May 2024
Time:                        17:03:23
No. Iterations:                    18
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0411      0.001     31.751      0.000       0.039       0.044
Prvlnc                    -0.0700      0.004    -17.364      0.000      -0.078      -0.062
Tot_Mdcr_Stdzd_Pymt_PC  5.797e-06   1.23e-07     47.286      0.000    5.56e-06    6.04e-06
Tot_Mdcr_Pymt_PC          7.46e-07   1.17e-07      6.373      0.000    5.17e-07    9.75e-07
```

1B (iii) Robustness Check – Subset Analysis (Age Group)



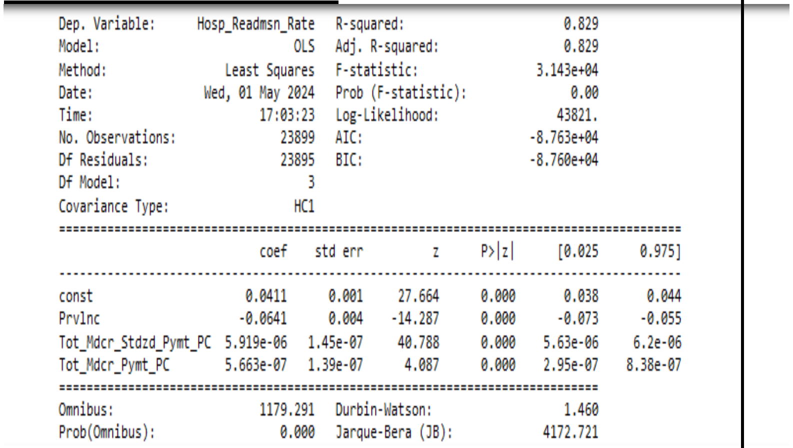
```
==============================================================================
Dep. Variable:      Hosp_Readmsn_Rate   R-squared:                       0.868
Model:                            OLS   Adj. R-squared:                  0.868
Method:                 Least Squares   F-statistic:                 2.698e+04
Date:                Wed, 01 May 2024   Prob (F-statistic):               0.00
Time:                        17:03:23   Log-Likelihood:                 17399.
No. Observations:                8185   AIC:                        -3.479e+04
Df Residuals:                    8182   BIC:                        -3.477e+04
Df Model:                           2
Covariance Type:            nonrobust
============================================================================================
                               coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------
const                        0.0148      0.002      7.980      0.000       0.011       0.018
Prvlnc                       0.0144      0.006      2.349      0.019       0.002       0.026
ER_Visits_Per_1000_Benes     0.0001    5.7e-07    179.746      0.000       0.000       0.000
==============================================================================
Omnibus:                      151.707   Durbin-Watson:                   1.624
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              237.552
Skew:                          -0.187   Prob(JB):                     2.61e-52
```

1B (iv) Robustness Check – Bootstrapping Standard Errors



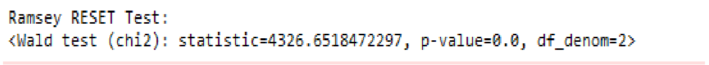
```
==============================================================================
Dep. Variable:      Hosp_Readmsn_Rate   R-squared:                       0.829
Model:                            OLS   Adj. R-squared:                  0.829
Method:                 Least Squares   F-statistic:                 3.143e+04
Date:                Wed, 01 May 2024   Prob (F-statistic):               0.00
Time:                        17:03:23   Log-Likelihood:                 43821.
No. Observations:               23899   AIC:                        -8.763e+04
Df Residuals:                   23895   BIC:                        -8.760e+04
Df Model:                           3
Covariance Type:                  HC1
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0411      0.001     27.664      0.000       0.038       0.044
Prvlnc                    -0.0641      0.004    -14.287      0.000      -0.073      -0.055
Tot_Mdcr_Stdzd_Pymt_PC  5.919e-06   1.45e-07     40.788      0.000    5.63e-06     6.2e-06
Tot_Mdcr_Pymt_PC        5.663e-07   1.39e-07      4.087      0.000    2.95e-07    8.38e-07
==============================================================================
Omnibus:                     1179.291   Durbin-Watson:                   1.460
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             4172.721
```

1B (v) Robustness Check – Heteroskedasticity-Robust Standard Errors



```
Dep. Variable:      Hosp_Readmsn_Rate   R-squared:                       0.829
Model:                            OLS   Adj. R-squared:                  0.829
Method:                 Least Squares   F-statistic:                 3.143e+04
Date:                Wed, 01 May 2024   Prob (F-statistic):               0.00
Time:                        17:03:23   Log-Likelihood:                 43821.
No. Observations:               23899   AIC:                        -8.763e+04
Df Residuals:                   23895   BIC:                        -8.760e+04
Df Model:                           3
Covariance Type:                  HC1
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0411      0.001     27.664      0.000       0.038       0.044
Prvlnc                    -0.0641      0.004    -14.287      0.000      -0.073      -0.055
Tot_Mdcr_Stdzd_Pymt_PC  5.919e-06   1.45e-07     40.788      0.000    5.63e-06     6.2e-06
Tot_Mdcr_Pymt_PC        5.663e-07   1.39e-07      4.087      0.000    2.95e-07    8.38e-07
==============================================================================
Omnibus:                     1179.291   Durbin-Watson:                   1.460
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             4172.721
```

1B (vi) Robustness Check – Ramsey Reset Test



```
Ramsey RESET Test:
<Wald test (chi2): statistic=4326.6518472297, p-value=0.0, df_denom=2>
```