**PREDICTIVE ANALYTICS**

**HEALTHCARE DATA ANALYSIS**

DONE BY:

RANJITH LK 71762233035

MSc. DCS II YEAR

**Abstract: Healthcare Analysis Using Factor Analysis, Multiple Linear Regression, and F-test**

This project delves into the intricate landscape of healthcare analysis through the lens of statistical methodologies including Factor Analysis (FA), Multiple Linear Regression (MLR), and F-test. With a paramount focus on understanding and enhancing healthcare outcomes, these analytical tools were applied to a comprehensive dataset encompassing various health indicators.

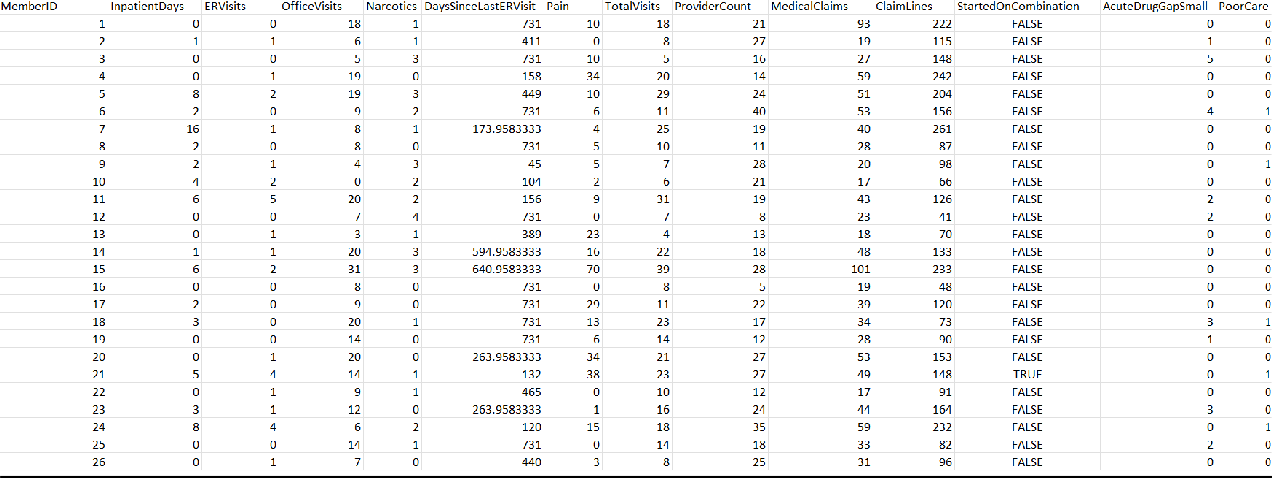
Factor Analysis was employed to unearth latent factors underpinning the observed variability in healthcare metrics, shedding light on underlying dimensions such as patient demographics, healthcare access, and quality. The extracted factors provide invaluable insights into the multifaceted nature of healthcare, enabling a nuanced understanding essential for targeted interventions and policy formulation.

Multiple Linear Regression was utilized to discern the intricate relationship between independent health variables and the dependent variable of interest, typically denoting healthcare quality. Through regression coefficients, intercepts, and mean squared errors, this analysis elucidated the impact of diverse healthcare parameters on overall healthcare outcomes. The identification of significant predictors guides strategic decision-making by pinpointing areas for intervention and resource allocation.

The F-test served as a robust statistical tool to gauge the collective significance of predictor variables within the MLR framework. By examining F-values and associated p-values, this analysis unveiled the pivotal contributors to healthcare quality, thereby informing prioritization strategies and resource optimization efforts.

Collectively, these analytical approaches offer a holistic understanding of healthcare dynamics, empowering stakeholders with actionable insights to drive positive change in healthcare delivery, policy formulation, and overall public health outcomes.

# **Dataset**



**Dataset Overview:**

This dataset appears to contain healthcare utilization and claims data for a group of patients, identified by unique MemberID. Here is an overview of the key columns and their contents:

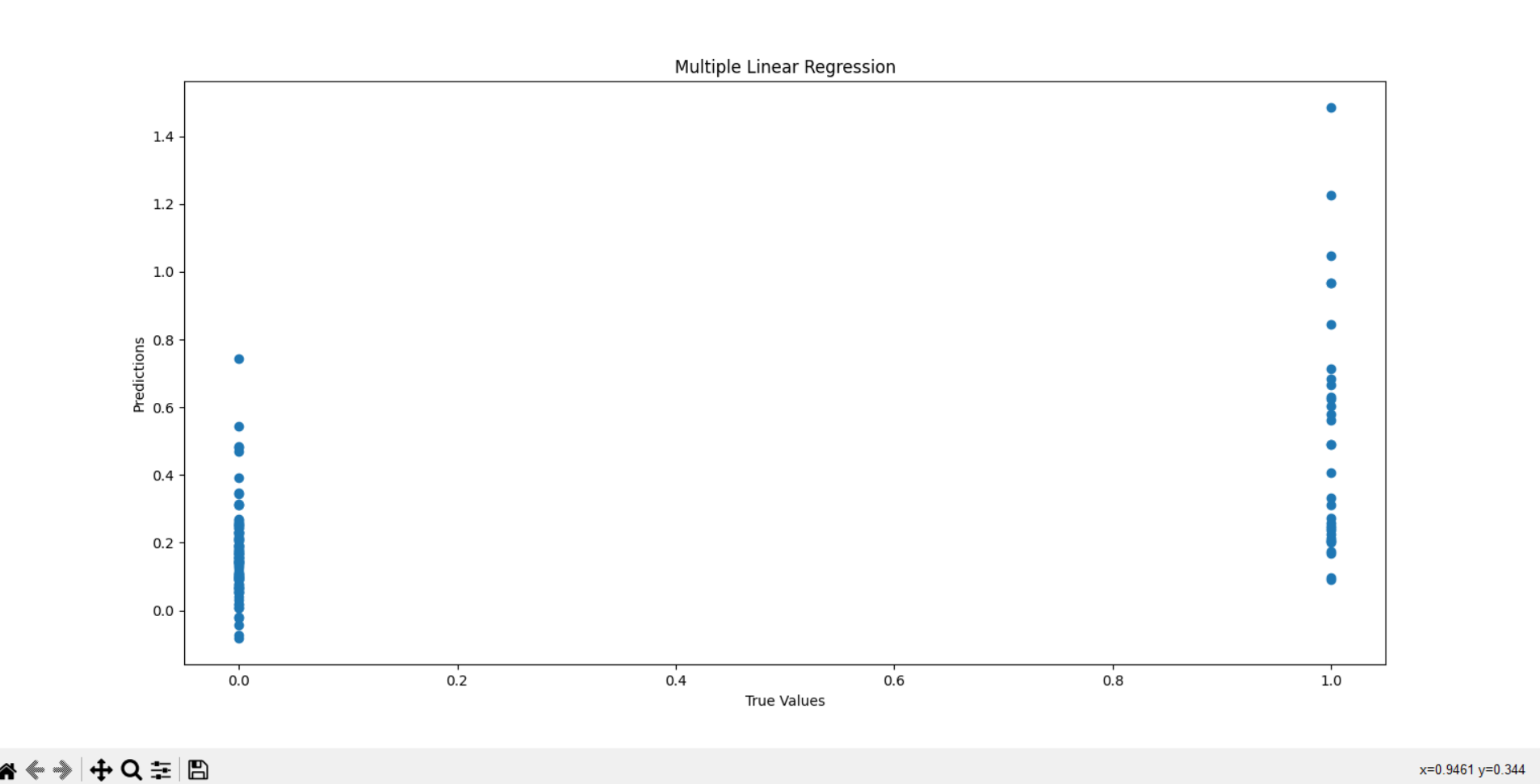
* **MemberID**: A unique identifier for each patient.
* **InpatientDays**: The number of days the patient spent in the hospital as an inpatient.
* **ERVisits**: The number of visits the patient made to the emergency room.
* **OfficeVisits**: The number of visits the patient made to a doctor's office.
* **Narcotics**: Indicates whether the patient was prescribed narcotics (1 for yes, 0 for no).
* **DaysSinceLastERVisit**: The number of days since the patient's last emergency room visit.
* **Pain**: A metric related to the patient's pain level, possibly on a scale from 0 to 100.
* **TotalVisits**: The total number of healthcare visits, which could be a sum of inpatient, ER, and office visits.
* **ProviderCount**: The number of different healthcare providers the patient has seen.
* **MedicalClaims**: The number of medical claims filed for the patient.
* **ClaimLines**: The number of lines or individual items/services listed on the patient's medical claims.
* **StartedOnCombination**: A boolean (TRUE/FALSE) indicating if the patient started on a combination of treatments or medications.
* **AcuteDrugGapSmall**: A count or indicator of small gaps in acute drug prescriptions.
* **PoorCare**: A count or indicator of instances of poor care or suboptimal outcomes for the patient.

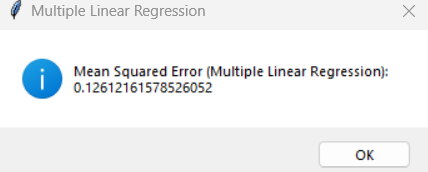
**Multiple Linear Regression**

**Definition**

Multiple Linear Regression (MLR) is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. It is used to understand the relationship between the dependent variable and one or more independent variables.

**Screenshot of Output**

****

****

**Inference**

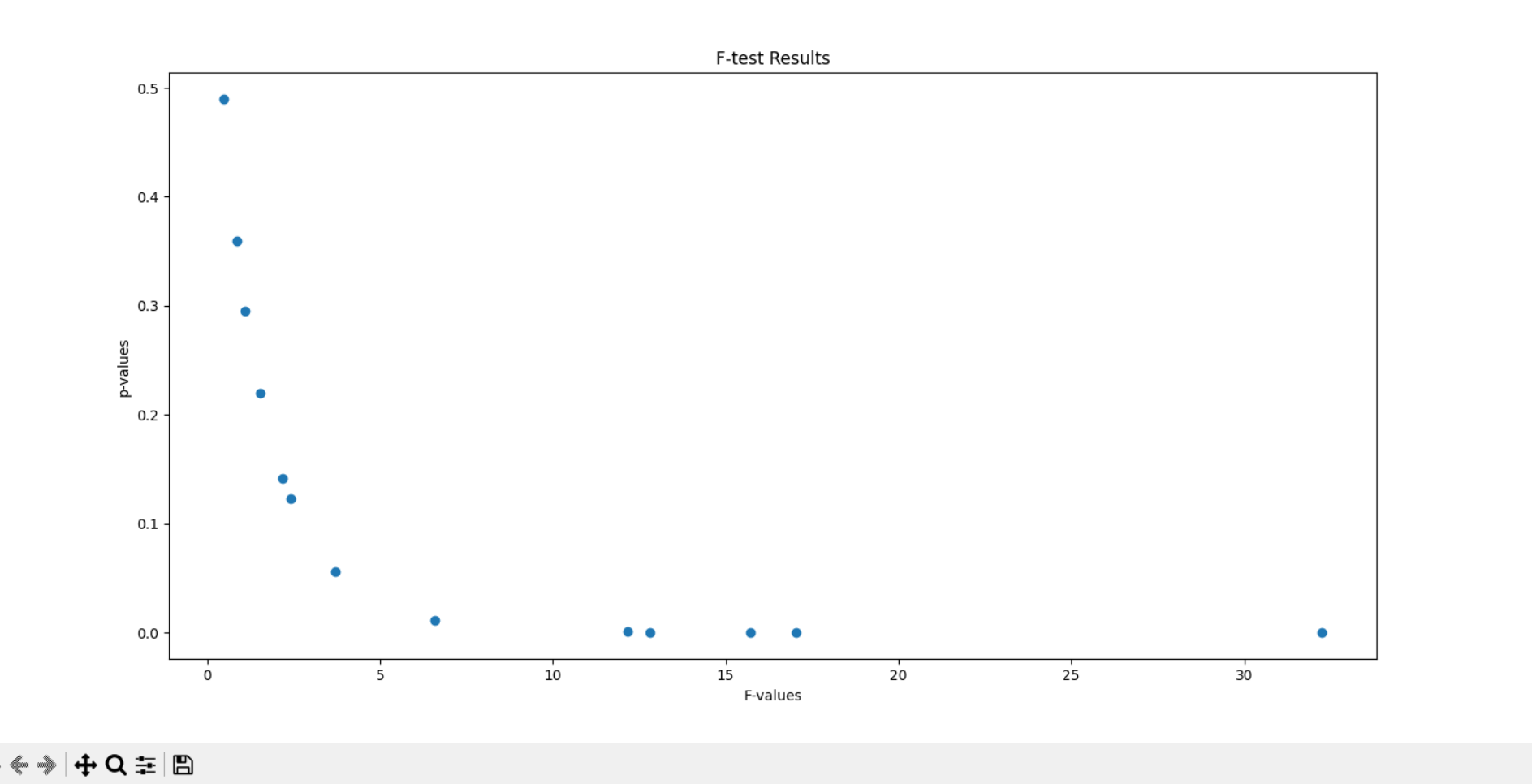
The MSE for the regression model indicates the average squared difference between the observed actual outcomes and the outcomes predicted by the model.

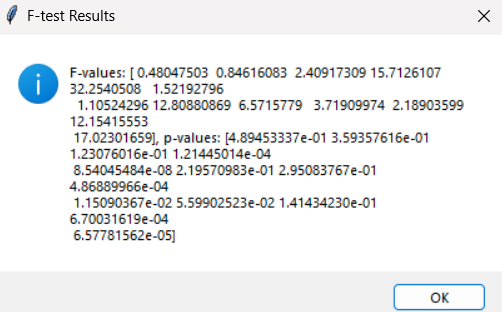
**F-test**

**Definition**

An F-test is used to determine whether there are significant differences between the variances of two or more groups. In regression, it's used to compare models and check if a group of variables significantly predicts the dependent variable.

**Screenshot of Output**

****

****

**Inference**

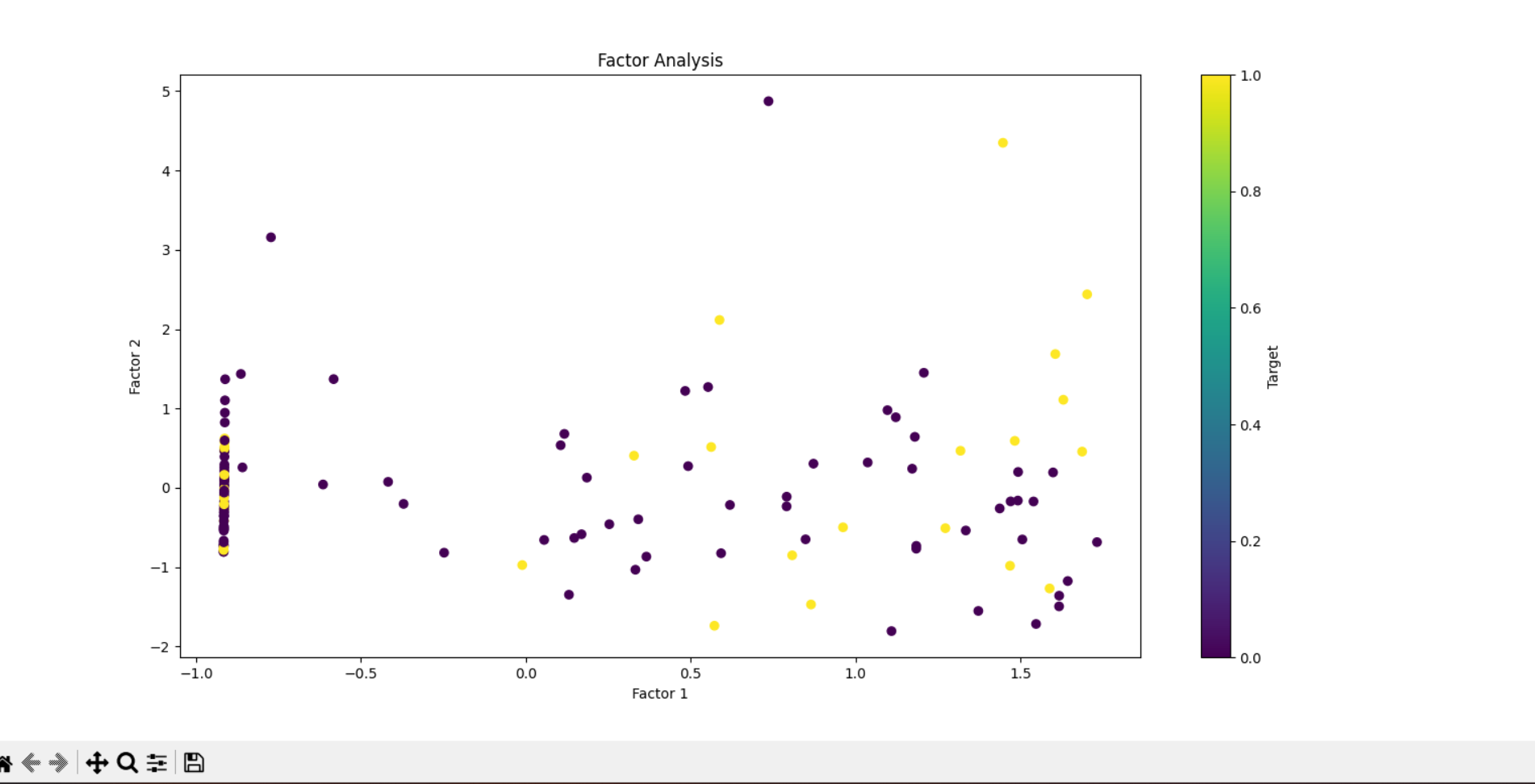
The F-values and corresponding p-values indicate the significance of each variable. Variables with low p-values (typically < 0.05) are considered significant predictors of the outcome variable.

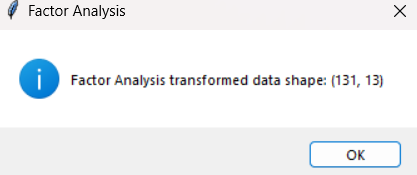
**Factor Analysis**

**Definition**

Factor Analysis (FA) is a technique used to identify underlying relationships between variables. It reduces data by grouping variables that share common variance, which can be interpreted as latent factors.

**Screenshot of Output**

****

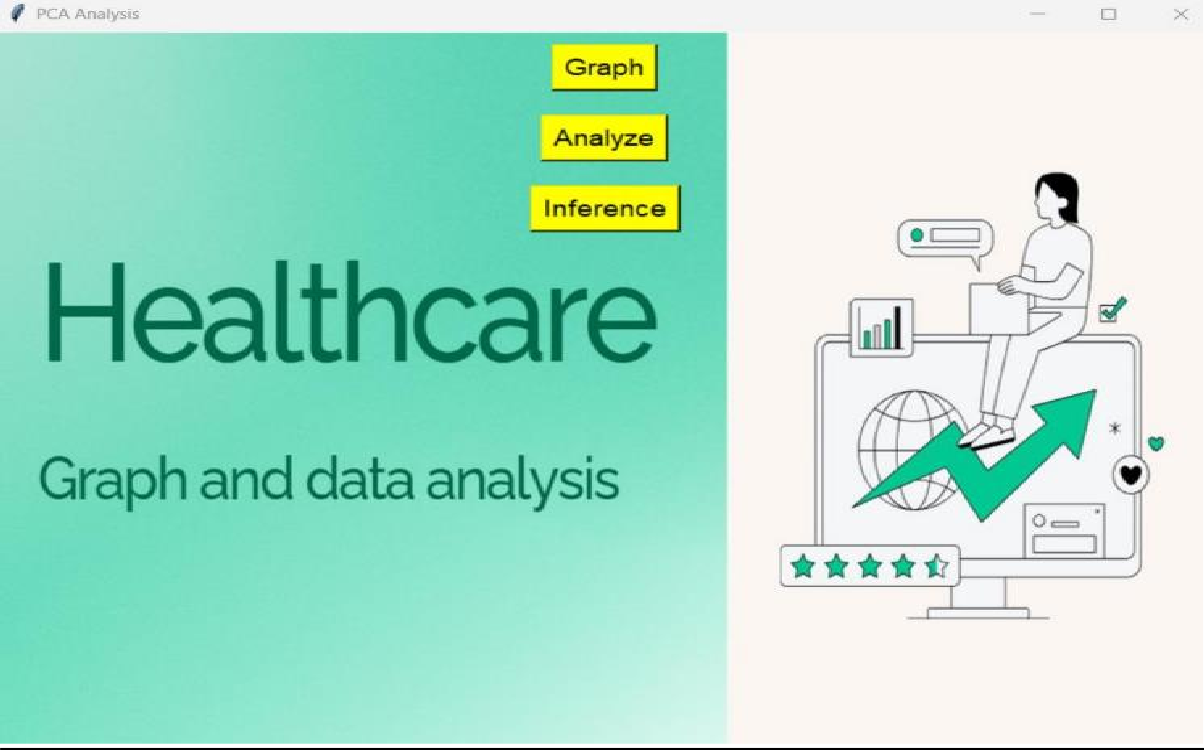
****

**Inference**

FA revealed latent factors that account for the variance in healthcare indicators. Understanding these factors can help in designing targeted interventions for improving healthcare outcomes.

# **GRAPHICAL USER INTERFACE(GUI):**





**Conclusion:**

In conclusion, the application of Factor Analysis, Multiple Linear Regression, and F-test in healthcare analysis has provided a multifaceted understanding of the determinants of healthcare quality.

Overall, the integration of these analytical techniques has enriched the understanding of healthcare dynamics. By identifying key factors, significant predictors, and validating their importance, this comprehensive analysis provides a strong foundation for improving healthcare policies and practices. The findings empower stakeholders to implement data-driven interventions, optimize resource utilization, and ultimately enhance the quality of healthcare services delivered to the population.