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

MemberID	InpatientDays	ERVisits	OfficeVisits	Narcotics	DaysSinceLastERVisit	Pain	TotalVisits	ProviderCount	MedicalClaims	ClaimLines	StartedOnCombination	AcuteDrugGapSmall	PoorCare
1	0	0	18	1	731	10	18	21	93	222	FALSE	0	0
2	1	1	6	1	411	0	8	27	19	115	FALSE	1	0
3	0	0	5	3	731	10	5	16	27	148	FALSE	5	0
4	0	1	19	0	158	34	20	14	59	242	FALSE	0	0
5	8	2	19	3	449	10	29	24	51	204	FALSE	0	0
6	2	0	9	2	731	6	11	40	53	156	FALSE	4	1
7	16	1	8	1	173.9583333	4	25	19	40	261	FALSE	0	0
8	2	0	8	0	731	5	10	11	28	87	FALSE	0	0
9	2	1	4	3	45	5	7	28	20	98	FALSE	0	1
10	1	2	0	2	101	2	6	21	17	86	FALSE	0	0
11	6	5	20	2	156	9	31	19	43	126	FALSE	2	0
12	0	0	7	1	731	0	7	8	23	91	FALSE	2	0
13	0	1	3	1	389	23	4	13	18	70	FALSE	0	0
14	1	1	20	3	991.9583333	16	22	18	48	133	FALSE	0	0
15	6	2	31	3	640.9583333	70	39	28	101	233	FALSE	0	0
16	0	0	8	0	731	0	8	5	19	48	FALSE	0	0
17	2	0	9	0	731	20	11	22	39	120	FALSE	0	0
18	3	0	20	1	731	13	23	17	34	73	FALSE	3	1
19	0	0	14	0	731	6	14	12	28	90	FALSE	1	0
20	0	1	20	0	263.9583333	34	21	27	53	153	FALSE	0	0
21	5	1	14	1	132	38	23	27	49	148	TRUE	0	1
22	0	1	9	1	465	0	10	12	17	91	FALSE	0	0
23	3	1	12	0	263.9583333	1	16	24	44	164	FALSE	3	0
24	8	4	6	2	120	15	18	35	59	232	FALSE	0	1
25	0	0	14	1	731	0	14	18	33	82	FALSE	2	0
26	0	1	7	0	440	3	8	25	31	96	FALSE	0	0

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.

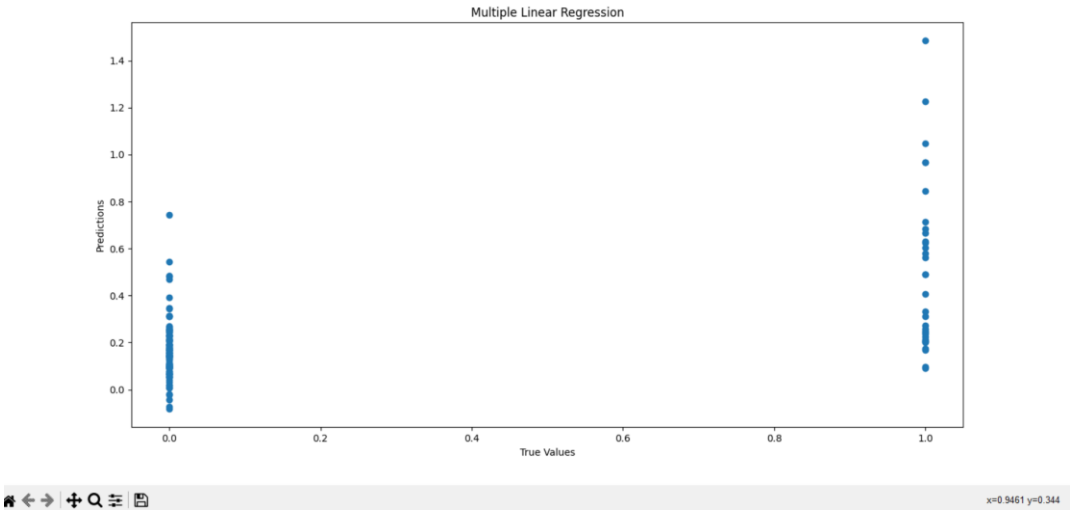
- **ProviderCount**: The number of different healthcare providers the patient has seen.
- **MedicalClaims**: The number of medical claims filed for the patient.
- **ClaimLines**: The number of lines or individual items/services listed on the patient's medical claims.
- **StartedOnCombination**: A boolean (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

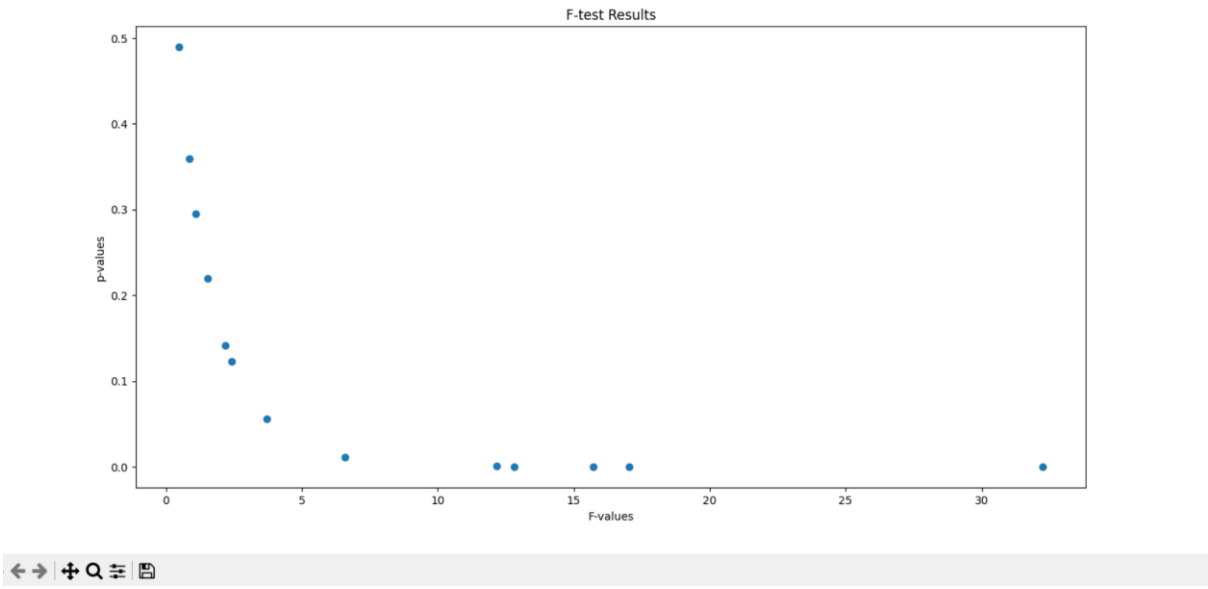
The MSE for the regression model indicates the average squared difference between the observed actual outcomes and the outcomes predicted by the model.

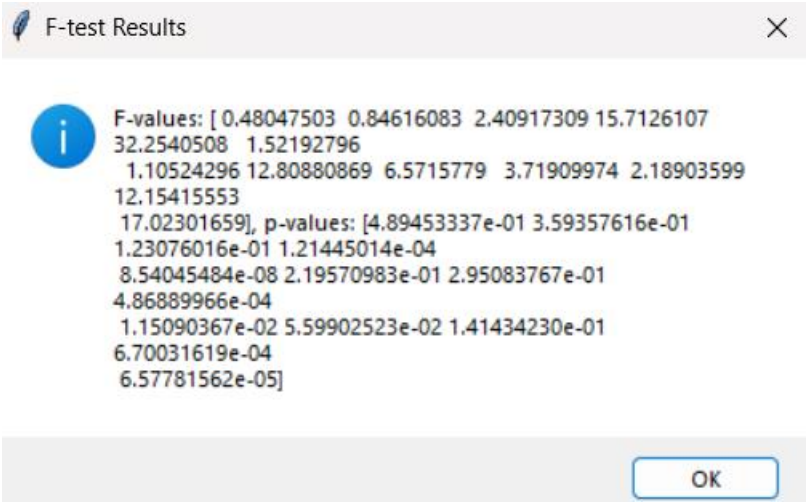
F-test

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

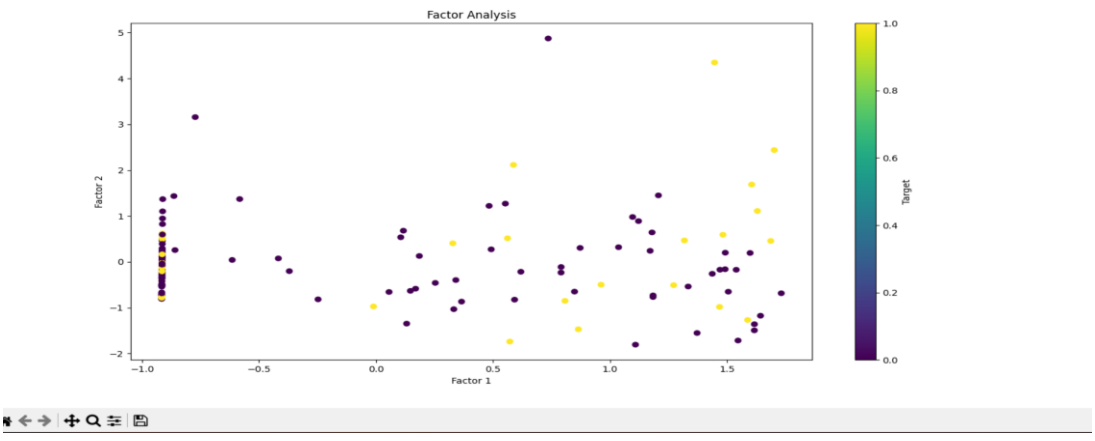
The F-values and corresponding p-values indicate the significance of each variable. Variables with low p-values (typically < 0.05) are considered significant predictors of the outcome variable.

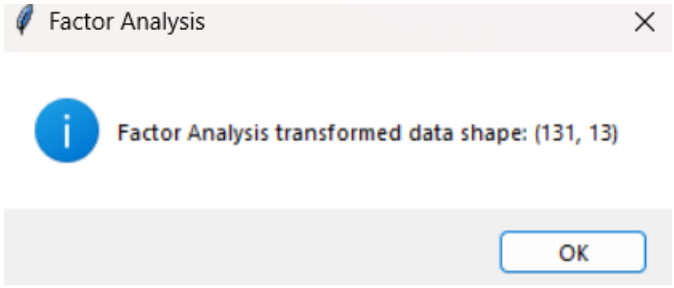
Factor Analysis

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

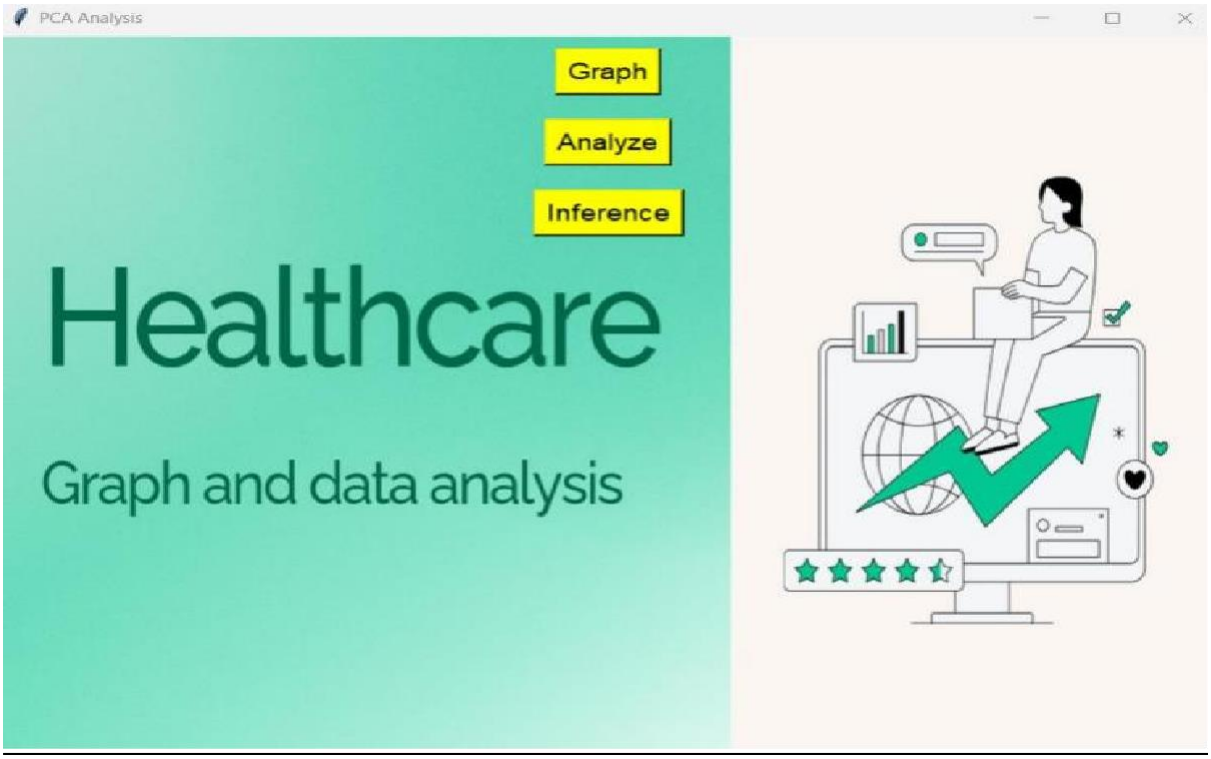




Inference

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.