

Week 9 Deliverable

Group Name: Individual Project - Healthcare Persistency

Name: Kirtoria Ward

Email: kirtoria@gmail.com

Country: United States

College/Company: North Central College

Specialization: Data Science

Problem Description

This project involves predicting drug persistency to help pharmaceutical companies optimize strategies and improve patient adherence to therapy

GitHub Repository Link

https://github.com/KTW04/Data_Glacier_Internship_LISUM38

Data Cleansing and Transformation

Missing Values

- Technique 1: Mean Imputation
 - Missing numerical values were replaced with the mean. This ensures consistency in the dataset while minimizing data loss.
- Technique 2: Mode Imputation
 - Missing categorical values were filled with the most frequent category (mode), maintaining the integrity of categorical data.

Outliers

- Technique 1: IQR Method
 - Outliers were identified using the Interquartile Range (IQR) and removed. This reduced the impact of extreme values on the model.
 - Technique 2: Capping
 - Extreme values were adjusted to the 5th and 95th percentiles, ensuring no critical data was lost while managing skewed distributions.
-

Final Dataset Summary

Metric	Original Dataset	Cleaned Dataset
Rows	3,424	2,956
Columns	69	69

Review Comments

- Since this project was completed individually, the notebook was self-reviewed. Code outputs were validated step-by-step to ensure accuracy and consistency.