Team Name: Data Minders

Team Members:
Jacob Farrington; jtsfarrington@gmail.com , USA, University of North Texas, Data Science
Xiaoyan Zhang; xiaoyanhouston@gmail.com , USA, University of Texas at Dallas, Data Science
Marina Tsvetkova; marinatsv07@gmail.com , Poland, N/A , Data Science
Pamela S. D. Martey; pmlmartey@gmail.com, USA, Clark Atlanta University, Data Science

Problem description:

ABC is a pharmaceutical company keen on understanding persistency of a drug based on physician prescriptions for patients. To tackle this, ABC has engaged an analytics company to automate the identification of persistency. This analytics firm has tasked Team Data Minders with creating an automated solution to evaluate and enhance the persistency of a drug for ABC.

Business Understanding:

ABC Pharmaceutical is focused on optimizing drug persistency to enhance patient health outcomes and improve market competitiveness. Effective persistency is essential for reducing healthcare costs and ensuring treatments are effective, which also aids in regulatory compliance and strengthens the company's reputation. ABC has partnered with an analytics firm to develop a system that automates the monitoring of drug persistency, providing crucial insights to support strategic decisions and boost operational efficiency.

Project lifecycle along with deadline:

Week 8: Deliverables (Due date 26 June 2024)
  Group zoom to discuss progress and submission (25 June)

Week 9: Deliverables (Due date 2 July 2024)
  Group zoom to discuss progress and submission (1 July)

Week 10: Deliverables (Due date 9 July 2024)
  Group zoom to discuss progress and submission (8 July)

Week 11: EDA Presentation and proposed modeling technique (Due date 16 July 2024)
  Group zoom to discuss progress and submission (15 July)

Week 12: Model Selection and Model Building/Dashboard (Due date 23 July 2024)
  Group zoom to discuss progress and submission (22 July)

Week 13: Final Project Report and Code (Due date 30 July 2024)
  Group zoom to discuss progress and submission (29 J

Data Intake Report

Name: ABC Pharmaceutical - DataMinders
Report date: 6/19/24
Internship Batch: LISUM33
Version: 1.0
Data intake by: Data Minders
Data intake reviewer: Data Minders
Data storage location: https://github.com/S0n0f1saac/LISUM33-GroupProject

Tabular data details:

Total number of observations	3424
Total number of files	1
Total number of features	69
Base format of the file	.xlsx
Size of the data	899 KB

Proposed Approach:
- Identification of duplicates:
  - We will conduct exact matching for features where we expect duplicates and clustering to identify potential duplicates for all other features.

Github Repo Link:

https://github.com/S0n0f1saac/LISUM33-GroupProject