

# **Data Science Project**

## **Healthcare – Persistency of a Drug**

### **Week 7 works**

Team member's details:

|                  |  |
|------------------|--|
| Group Name:      | DS_SS  |
| Name:            | SONIYA SUNNY   |
| Email:           | <a href="mailto:soniyasunny1210@gmail.com">soniyasunny1210@gmail.com</a> |
| Country:         | Canada   |
| College/Company: | Data Glacier   |
| Specialization:  | Data Science   |

# Problem Description

To identify the persistency of a drug, a pharmaceutical company approached to develop a model based on data analysis. Factors that affect the persistence of drugs should be identified, along with data insights with predictive analytics, to help the company for their smooth and efficient functioning, with the help of dataset provided by the company.

# Business Understanding

Data-driven analysis and findings can help health care industries to function better, especially for pharmaceutical companies. By carefully analyzing historical data, it would be helpful to identify some knowledge on persistency of a drug. With the aid of physician descriptions, these companies can perform more, if they are getting proper predictive and descriptive analytics.

# Project Lifecycle along with deadline

|                  |  |
|------------------|--|
| July 19, 2022    | Problem description<br>Business understanding<br>Data collection<br>Data Intake report<br>Github Setup |
| July 26, 2022    | Data understanding<br>Initial Data Analysis<br>Identifying approaches                                  |
| August 2, 2022   | Data cleaning<br>Data transformation   |
| August 9, 2022   | EDA<br>Final Recommendation  |
| August 16, 2022  | EDA presentation for business users and technical users  |
| August 23, 2022  | Model Selection and Model Building   |
| August 30, 2022: | Final presentation<br>Project Report and Code  |

# Data Intake Report

Name: Data Science Final Project – ‘Healthcare – Persistency of a Drug’

Report date: July 16, 2022

Internship Batch: LISUM10: 30

Version:<1.0>

Data intake by: Soniya Sunny

Data intake reviewer:<intern who reviewed the report>

Data storage location: [Healthcare dataset.xlsx - Google Drive](#)

## Tabular data details:

|                                     |        |
|-------------------------------------|--------|
| <b>Total number of observations</b> | 3425   |
| <b>Total number of files</b>        | 1      |
| <b>Total number of features</b>     | 69     |
| <b>Base format of the file</b>      | .xlsx  |
| <b>Size of the data</b>             | 899 KB |

# Github Repo Link

[Soniyasunny1/Final Project DS SS \(github.com\)](https://github.com/Soniyasunny1/Final_Project_DS_SS)