**TEAM MEMBERS DETAILS**

GROUP NAME:

SPECIALIZATION:

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Email | Country | College/Company |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**PROBLEM DESCRIPTION**

Automate the process of identification of the persistency of the drug as per prescription provided by the physician.

**BUSINESS UNDERSTANDING**

Understanding the persistency is an issue for pharmaceutical companies and so, ABC pharma company has approached XYZ analytics company to provide insights into the same.

The role of XYZ analytics company is to undertake this project and provide a detailed understanding regarding the drug persistency.

Projected Business advantages:

* Adherence to a drug directly impacts the sales dollars a product can generate
* Optimize and improve the efficacy of clinical trials
* Target specific patient populations more effectively
* Better insight into patient behaviour to improve drug effectiveness

**PROJECT LIFECYCLE**

* Sep 05 – Sep 12

Data understanding

Type of data for analysis

Problems in the data (number of NA values, outliers, skewed etc)

* Sep 12 – Sep 19

Data Cleansing and Transformation

* Sep 19 – Sep 26

EDA and Final recommendation

* Sep 26 – Oct 3

EDA presentation for business users

* Oct 3 – Oct 10

Model Selection and Model building

* Oct 10 – Oct 17

Final Report, Code, Ppt submission

**DATA INTAKE REPORT**

Name: Data Science – Healthcare project

Report date: 05/09/2021

Internship Batch: [LISUM02](https://canvas.instructure.com/courses/3110044)

Version:1.0

Data intake by: Group name

Data intake reviewer:<intern who reviewed the report>

Data storage location: github

**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 |

**GITHUB LINK**