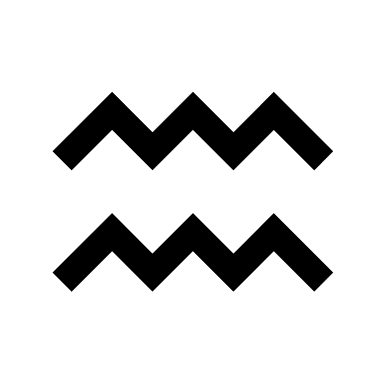
****

|  |  |
| --- | --- |
| An image of a part of a periodic table and test tubes  Quantic  *an analytics company* | Week 8 - deliverables  This project tackles a critical organizational challenge centered around data understanding and quality improvement. The team has embarked on a comprehensive data analysis journey, examining the dataset's characteristics, including the identification of issues such as missing values, outliers, and skewness. To rectify these data anomalies, a range of approaches are being applied, and the rationale behind these methods is elucidated.  Ansel Vallejo | Data Scientist  LISUM25 |

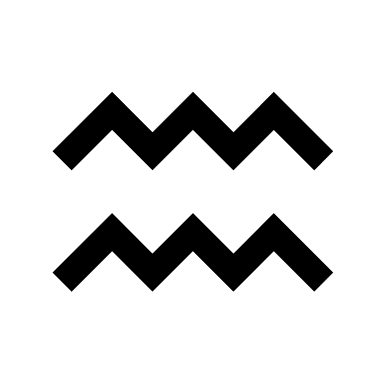
QUANTIC

Table of Contents

**About Us1**

Quantic Team2

**Overview3**

Business Scope3

Data Intake Report4

**Problem Description5**

**Data Understanding6**

Data7

**References8**

Github Repo Link8

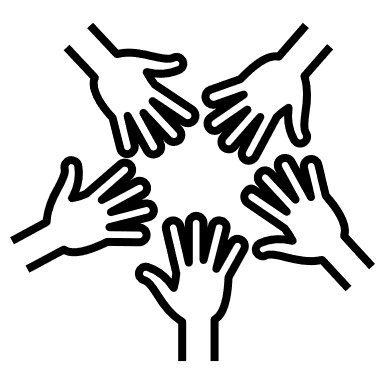
**i**

About Us

**Quantic** is an analytics company that places a strong emphasis on healthcare. We are dedicated to the idea that data can be a catalyst for positive change in the healthcare industry. With a talented team of data scientists and analysts, our primary objective is to tackle complex healthcare challenges and enhance patient outcomes. Our distinctive approach combines state-of-the-art data analytics with deep healthcare sector knowledge to deliver actionable insights, fostering informed decisions and meaningful advancements in healthcare provision. At Quantic, we are committed to a future where healthcare is not only data-driven but also healthier and more efficient.

**1**

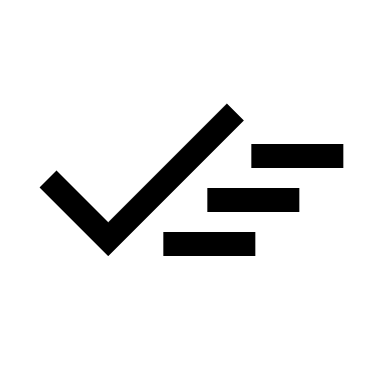
Our Team



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Email** | **Country** | **Institution** | **Specialization** |
| Ansel Vallejo | msavg@hotmail.com | Japan | Flatiron School | Data Science |

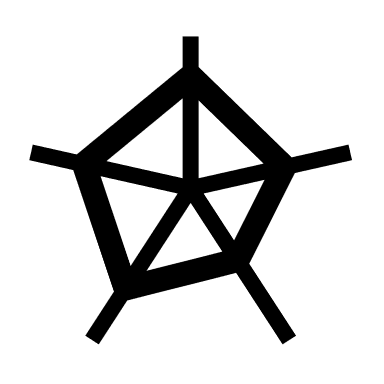
**2**

Overview

****

One of the persistent challenges faced by pharmaceutical companies lies in comprehending the duration of drug persistence as per physician prescriptions. To solve this problem, ABC Pharma Company recognized this issue and engaged Quantic to streamline and automate the identification process. By leveraging data analytics, the pharmaceutical company aimed to gain valuable insights into drug persistency patterns, ultimately enhancing their decision-making and ensuring better patient care. The collaboration between ABC Pharma and Quantic demonstrates a commitment to harnessing data-driven solutions to address critical industry challenges. Through this initiative, they strive to advance pharmaceutical practices and optimize patient outcomes.

Business Scope

****

The project scope entails the development of an automated system in collaboration with Quantic to analyze and identify drug persistency patterns within the pharmaceutical domain. This data-driven solution will enhance decision-making for ABC Pharma Company, ultimately leading to improved patient care and the advancement of pharmaceutical practices.

**3**

Data Intake Report

Name: *Healthcare\_dataset*

Report date: *October 26th 2023*

Internship Batch: LISUM25

Version:1.0

Data intake by: Ansel Vallejo

Data intake reviewer: N/A

Data storage location: N/A

**Tabular data details:**

**City**

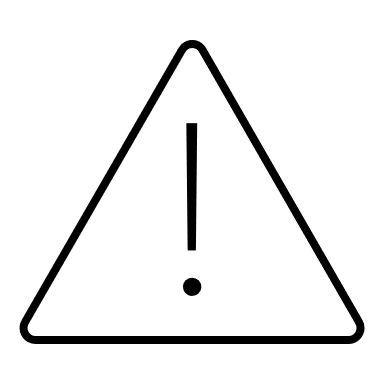
|  |  |
| --- | --- |
| **Total number of observations** | 3424 |
| **Total number of files** | 1 |
| **Total number of features** | 69 |
| **Base format of the file** | .CSV |
| **Size of the data** | 899 KB |

**Proposed Approach:**

* Check data for any missing values.
* Check for outliers.
* Check for skewed data.

**4**

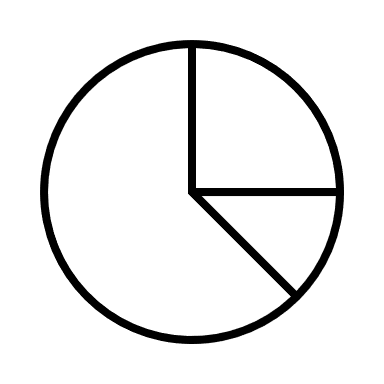
Problem Description



The problem at hand pertains to the pharmaceutical industry's struggle to gain insights into drug persistency duration as influenced by physician prescriptions. It is crucial to comprehend the duration of patient adherence to prescribed drug regimens, as this knowledge is vital for enhancing treatment effectiveness, patient well-being, and overall pharmaceutical strategies. The key issue in this context lies in the absence of an efficient and automated system for identifying persistency patterns, hampered by data anomalies such as missing values (NA), outliers, and skewed distributions. Addressing these data quality concerns is paramount to improving the understanding of patient drug persistency and ultimately enhancing pharmaceutical practices.

**5**

Data Understanding



**Scope:**

The scope of data understanding involves addressing critical data issues such as missing values, outliers, and skewed distributions. By identifying and mitigating these problems, the analysis aims to ensure the dataset's integrity and improve the accuracy of insights into drug persistency patterns influenced by physician prescriptions.

**Observation:**The dataset presented contains *3,424* datapoints and *69* variables.

For better understanding, the data is grouped in buckets:

* Demographics
* Provider Attributes
* Clinical Factors
* Disease and Treatment Factors

**Target:**

***Persistency Flag***

**6**

**Data**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Bucket** | **Variable** | **Datatype** | **Missing Data** | **Unknown** | **Outliers** | **Skewed** |
| Demographics | Age\_bucket | Object | **✖** | **✖** | **✖** | **✖** |
| Race | Object | **✖** | ✔ | **✖** | **✖** |
| Region | Object | **✖** | ✔ | **✖** | **✖** |
| Ethnicity | Object | **✖** | ✔ | **✖** | **✖** |
| Gender | Object | **✖** | **✖** | **✖** | **✖** |
| Idn\_Indicator | Object | **✖** | **✖** | **✖** | **✖** |
| Provider Attributes | Ntm\_Specialty | Object | **✖** | ✔ | **✖** | **✖** |
| Ntm\_Specialist\_Flag | Object | **✖** | ✔ | **✖** | **✖** |
| Ntm\_Specialty\_Bucket | Object | **✖** | **✖** | **✖** | **✖** |
| Clinical Factors | Gluco\_Record\_Prior | Object | **✖** | **✖** | **✖** | **✖** |
| Gluco\_Record\_During\_Rx | Object | **✖** | **✖** | **✖** | **✖** |
| Dexa\_Freq\_During\_Rx | Integer | **✖** | **✖** | **✖** | **✖** |
| Dexa\_During\_Rx | Object | **✖** | **✖** | **✖** | **✖** |
| Frag\_Frac\_Prior\_Ntm | Object | **✖** | **✖** | **✖** | **✖** |
| Frag\_Frac\_During\_Rx | Object | **✖** | **✖** | **✖** | **✖** |
| Risk\_Segment\_Prior\_Ntm | Object | **✖** | **✖** | **✖** | **✖** |
| Tscore\_Bucket\_Prior\_Ntm | Object | **✖** | **✖** | **✖** | **✖** |
| Risk\_Segment\_During\_Rx | Object | **✖** | **✖** | **✖** | **✖** |
| Tscore\_Bucket\_During\_Ntm | Object | **✖** | **✖** | **✖** | **✖** |
| Change\_T\_Score | Object | **✖** | **✖** | **✖** | **✖** |
| Change\_Risk\_Segment | Object | **✖** | **✖** | **✖** | **✖** |
| Disease and Treatment Factors | NTM\_Injectable Experience | Object | **✖** | **✖** | **✖** | **✖** |
| NTM\_Risk\_Factors | Object | **✖** | **✖** | **✖** | **✖** |
| NTM - Comorbidity | Object | **✖** | **✖** | **✖** | **✖** |
| NTM – Concomitancy | Object | **✖** | **✖** | **✖** | **✖** |
| Adherence | Integer | **✖** | **✖** | **✖** | **✖** |

**Approach to overcome Data Errors:**

|  |  |
| --- | --- |
| **Unknown** | *Use the mode as an imputer* |
| *Data :*  Values that contain “N” and “Y” values will be converted into numerical categorical values, such as 0 and 1, respectively.  Values that contain missing values will be eliminated according to missing data threshold. | |

**7**

References



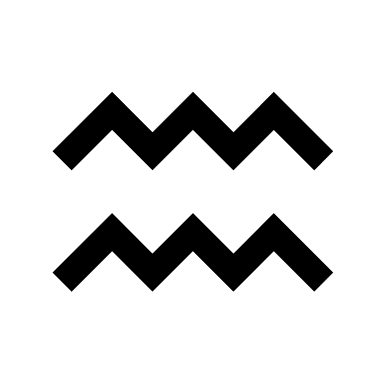
Github



https://github.com/anvadev/Healthcare-Drug\_Persistency/tree/main/Week%208%20-%20Data%20Understanding

**8**

**End of Documentation**

****

QUANTIC