



 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



Table of Contents

About Us	1
Quantic Team	2
Overview	3
Business Scope	3
Data Intake Report.....	4
Problem Description.....	5
Data Understanding	6
Data	7
References	8
Github Repo Link.....	8

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.

Our Team



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

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.

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:

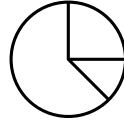
- Check data for any missing values.
- Check for outliers.
- Check for skewed data.

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.

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

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
<p><i>Data :</i></p> <p>Values that contain “N” and “Y” values will be converted into numerical categorical values, such as 0 and 1, respectively.</p> <p>Values that contain missing values will be eliminated according to missing data threshold.</p>	

References



Github



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

End of Documentation



QUANTIC