Project: Healthcare - Persistency of a drug

Week 8: Deliverables

Name: Krishna Ratna Deepika Haripuram

Email: haripuramdeepika@gmail.com

Country: Canada

Batch Code: LISUM26

Specialization: Data Science

Submission Date: Nov 23rd, 2023

Submitted to: Data Glacier

Table of Contents

1.Problem Description

The challenge we are addressing in this project is a crucial and intricate issue prevalent in the pharmaceutical sector: understanding and predicting the persistency of drug usage as per physician prescriptions. This complex problem extends beyond merely assessing adherence to medications; it delves into exploring a variety of influencing factors such as patient demographics, clinical history, medication characteristics, and socio-economic elements.

In chronic disease management within healthcare, the success of treatment largely hinges on patients consistently following their medication regimens. However, the reality is that non-persistence to medication is a widespread issue that results in less-than-ideal treatment outcomes, elevated healthcare costs, and an increased risk of complications associated with the diseases.

Our project's objective is to confront this challenge head-on by employing data analytics to unveil patterns and contributors to drug persistency. By conducting an exhaustive analysis of patient data, we aim to create predictive models that can accurately determine medication adherence. This initiative is expected to yield crucial insights into patient behavior, providing a foundation for improved healthcare interventions and strategies.

2.Data Understanding

The dataset for this project is a rich amalgamation of patient demographic details and clinical factors. It includes various attributes like age, gender, race, and specific medical history details, along with treatment adherence indicators. The data is comprehensive, covering aspects ranging from patient demographics to intricate clinical parameters and provider attributes.

3.Type of Data for Analysis

The dataset is a diverse mix of:

- **Categorical Data**: This includes variables like gender, race, ethnicity, and treatment types, providing a qualitative assessment of patient profiles and clinical scenarios.

- **Numerical Data**: It encompasses quantitative variables such as the frequency of DEXA scans and the count of risk factors, offering measurable insights into patient treatment and health risks.

4.Problems in the Data

The dataset presents several challenges:

- **Missing Values**: The dataset showed no significant missing values, indicating completeness in data recording.

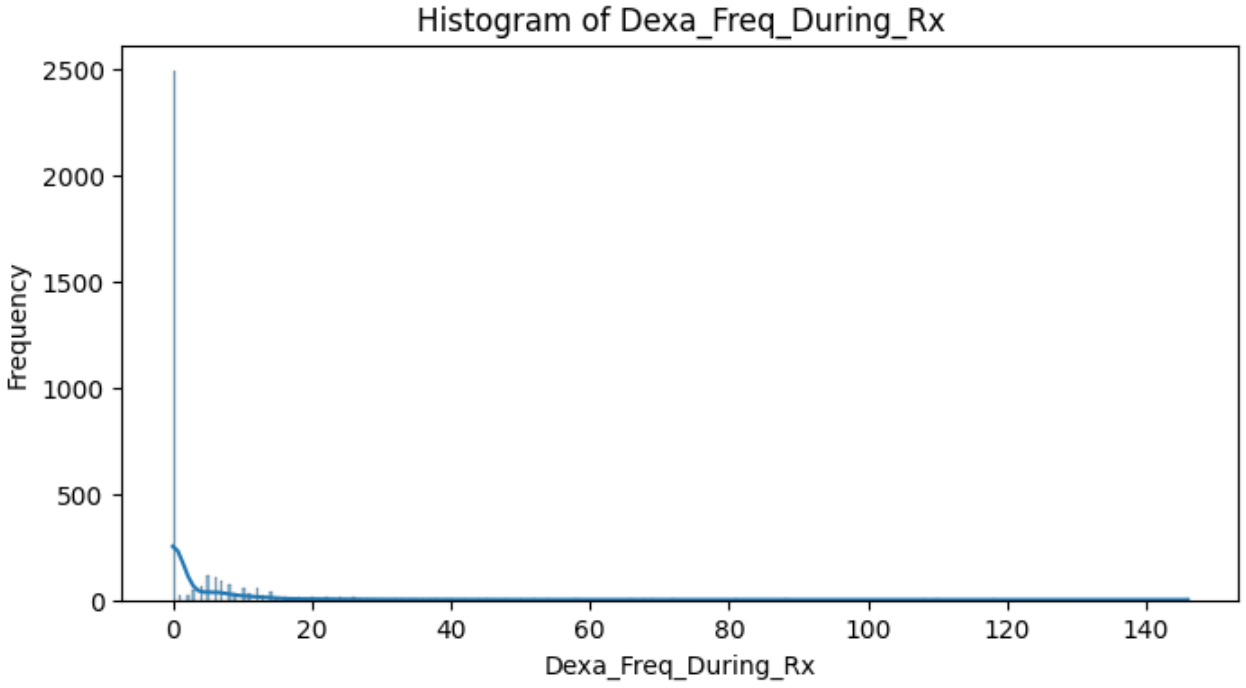
- **Outliers and Skewness:** Notable outliers and skewness were observed in key numerical columns like **`Dexa_Freq_During_Rx`** and **`Count_of_Risks`**, which could potentially skew statistical analyses.

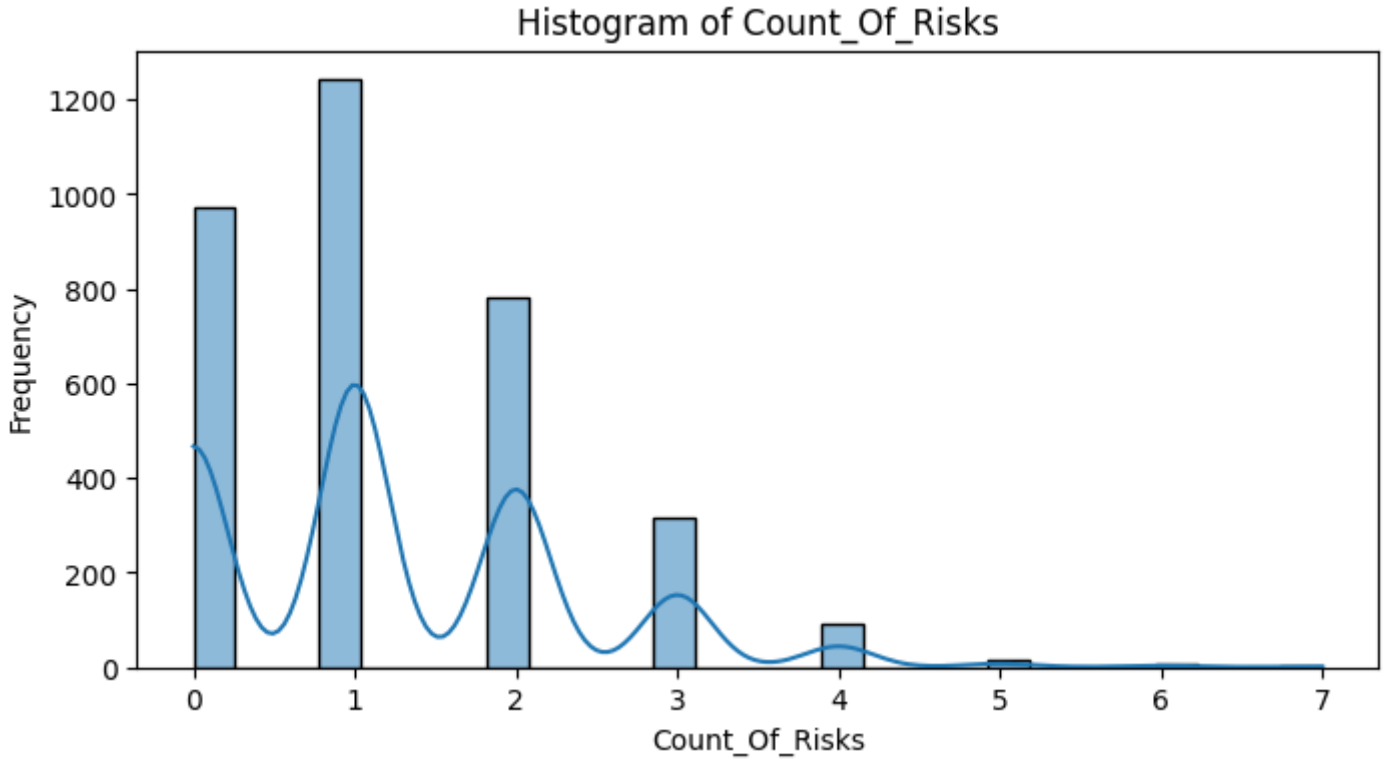
5. Approaches to Overcome Problems

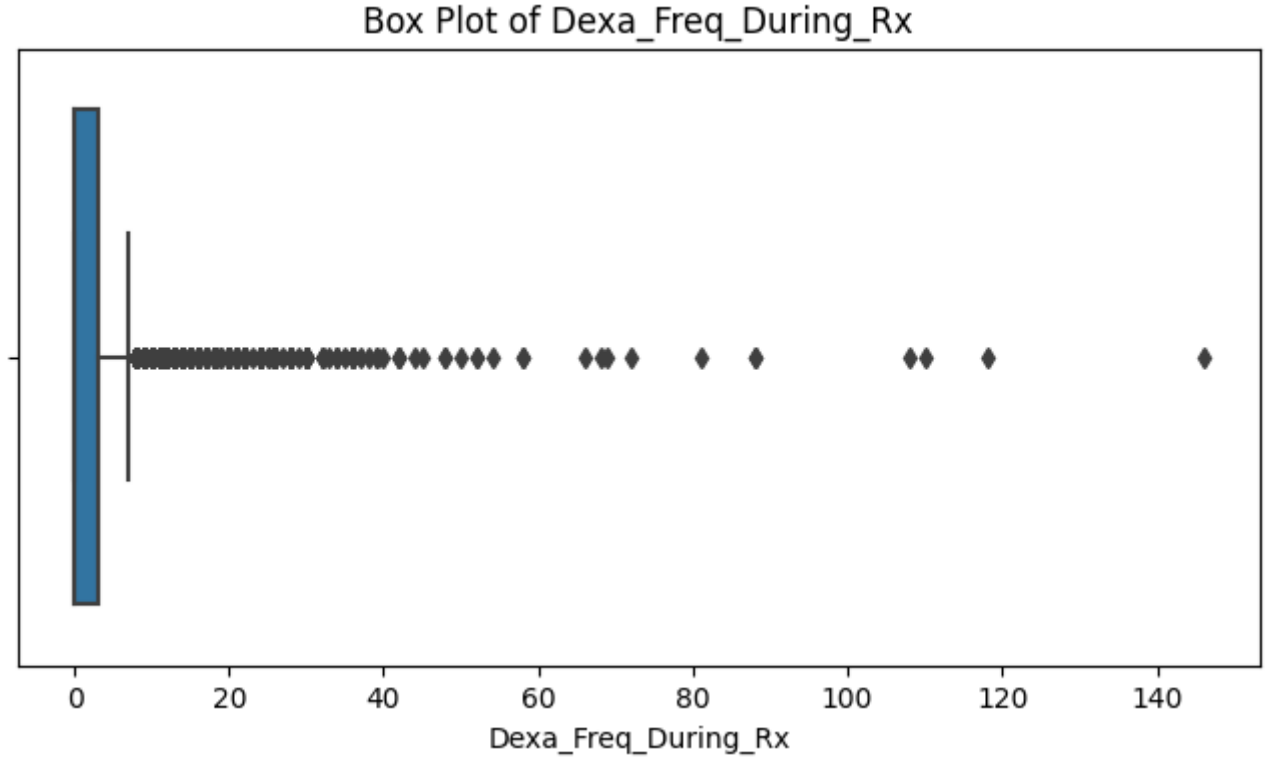
Handling Skewness: Log transformations were applied to skewed numerical columns to normalize their distribution, enhancing their suitability for statistical models and machine learning algorithms.

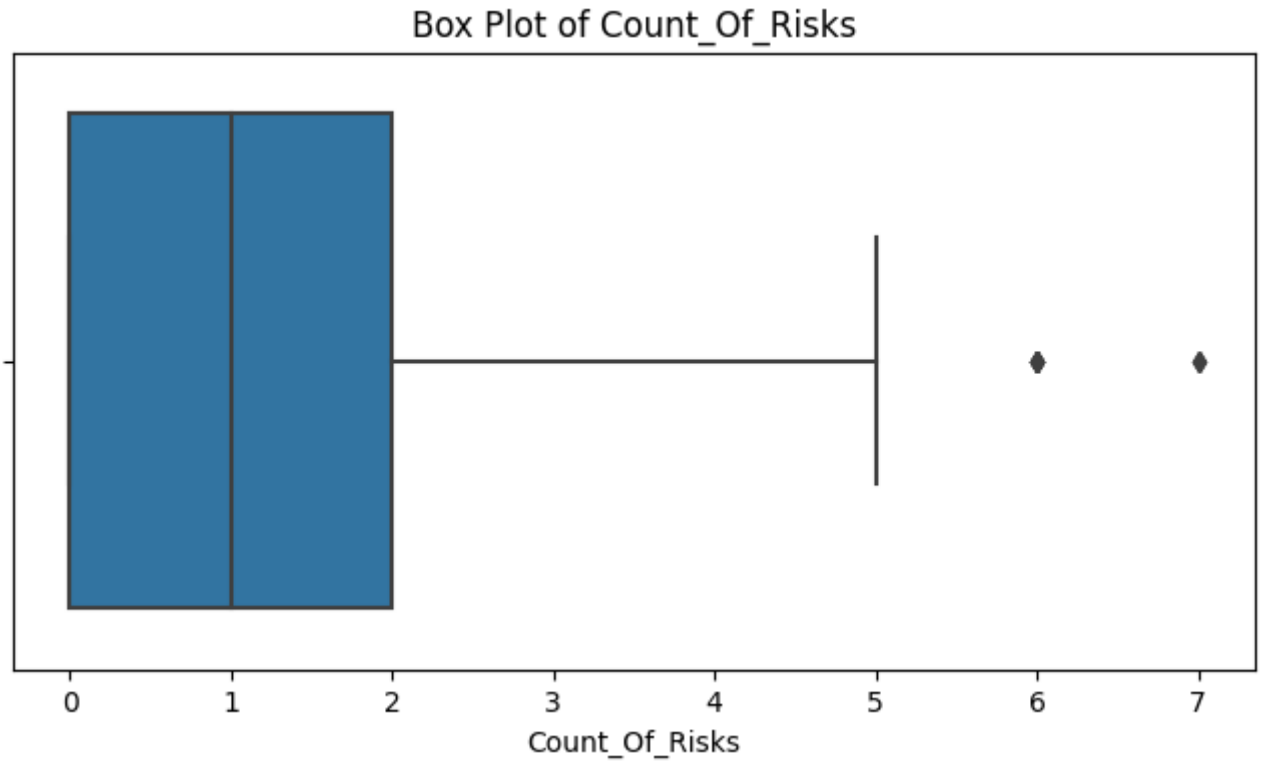
Managing Outliers: Techniques like the Interquartile Range (IQR) method were employed for identifying and treating outliers, thus normalizing the data, and ensuring robustness in analysis.

Data Visualization: Histograms and box plots were utilized to visually assess the distribution of numerical columns, aiding in the identification and interpretation of skewness and outliers.




Histogram of Count_Of_Risks
Frequency
Count_Of_Risks
0
200
400
600
800
1000
1200
0
1
2
3
4
5
6
7


Box Plot of Dexa_Freq_During_Rx
Dexa_Freq_During_Rx
0
20
40
60
80
100
120
140



Github Repo Link
https://github.com/krishnaharipuram/Data-Glacier/upload/main