**TEAM MEMBERS DETAILS**

GROUP NAME: PACHACUTEAM

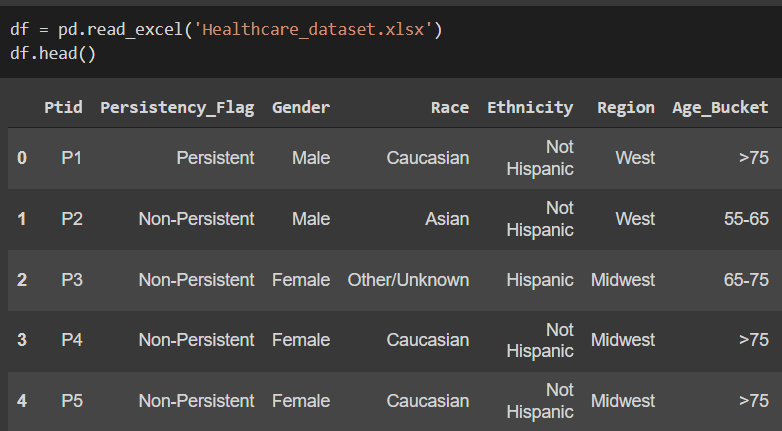
SPECIALIZATION: Data Science

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Email | Country | College/Company |
| Andersson Romero | [andromdez@gmail.com](mailto:andromdez@gmail.com) | Perú | Freelancer |

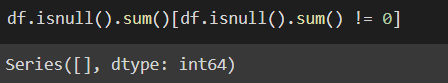
**Data Understanding**

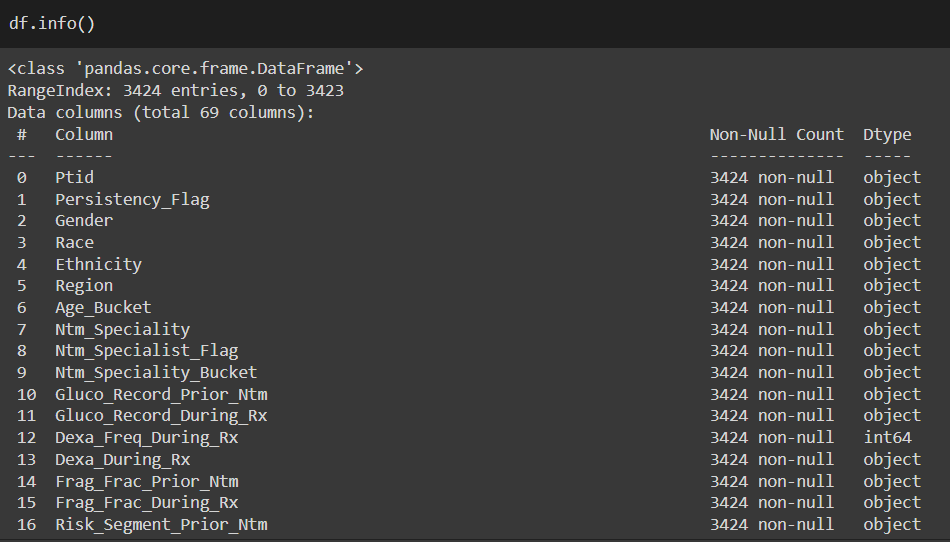
As a first step, we load the dataset using pandas.

We look at the data.

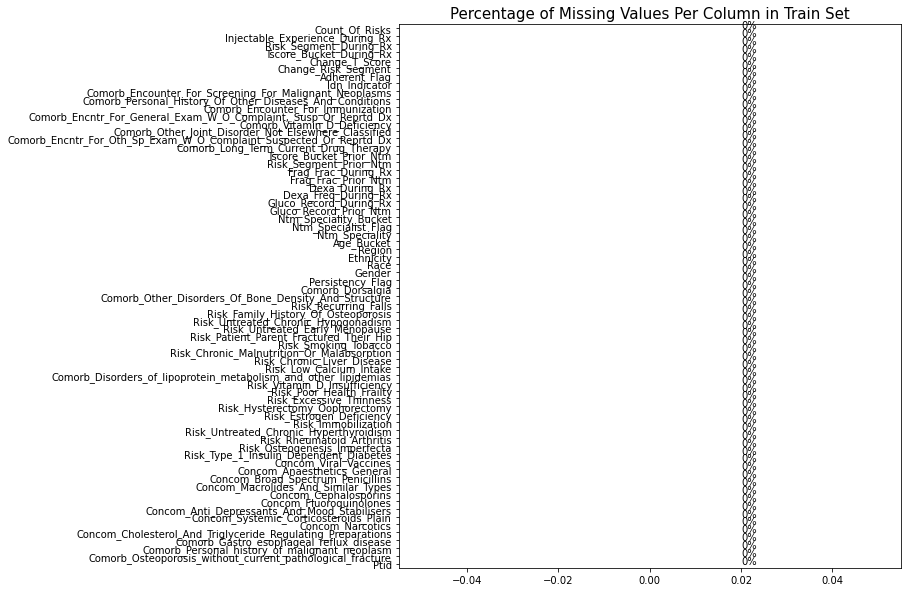


We can quickly calculate that there are no null elements.

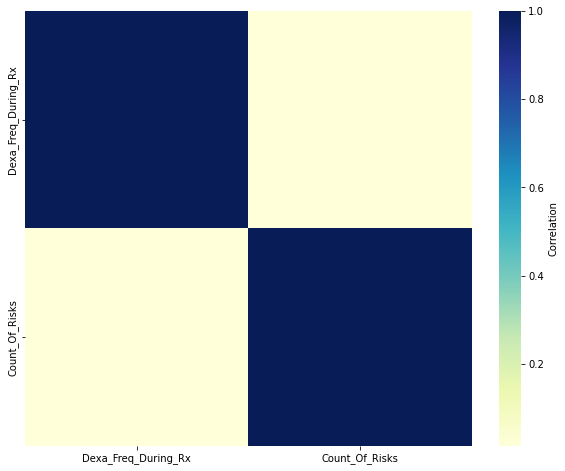


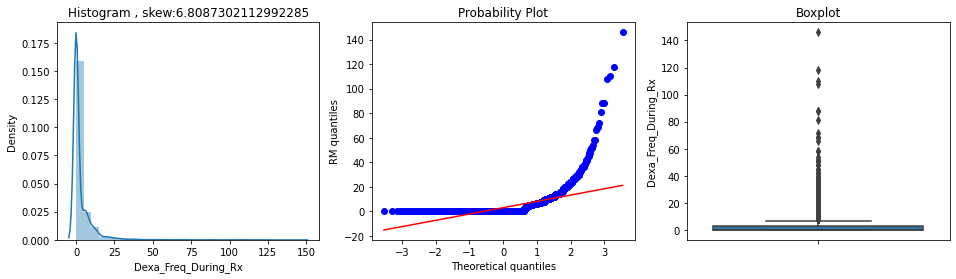
For more detail we use the info() function, where we see that almost all variables are objects.

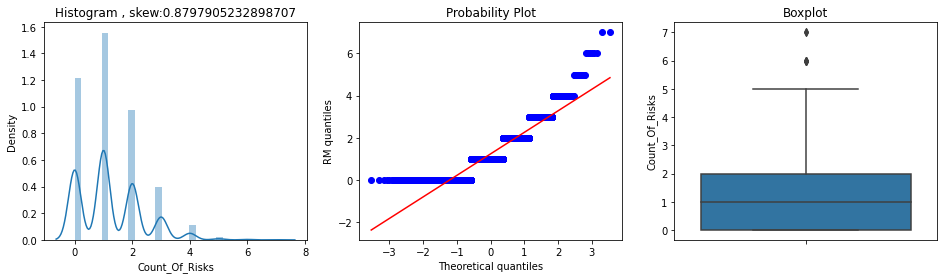
We observe graphically that there are no null or empty elements.



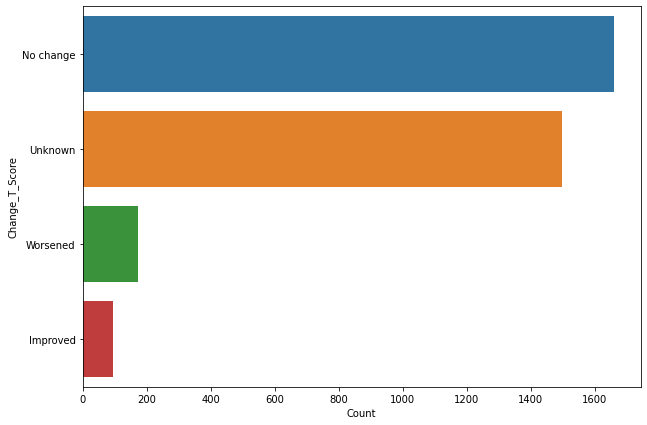
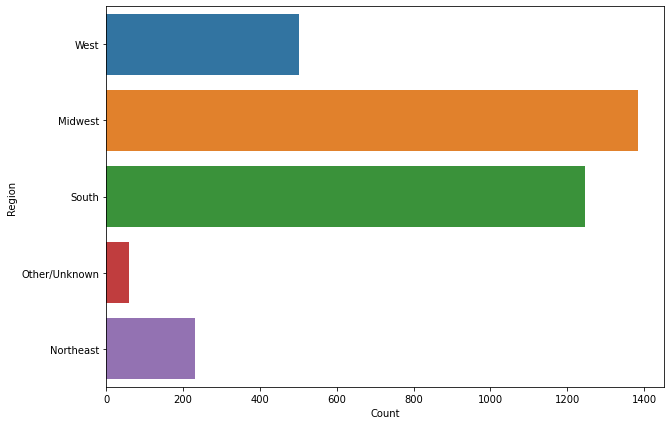
We can see that there is a high correlation between the numerical variables.

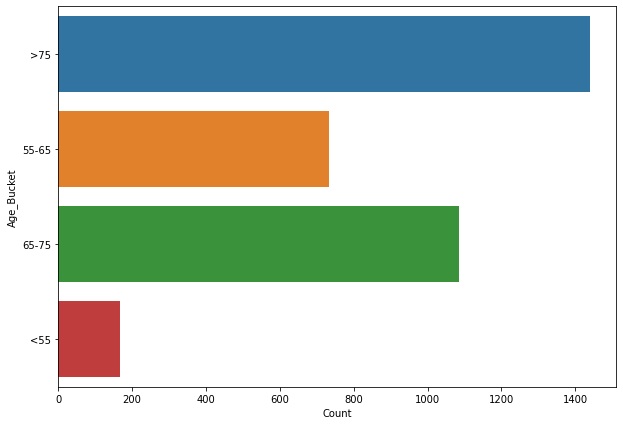


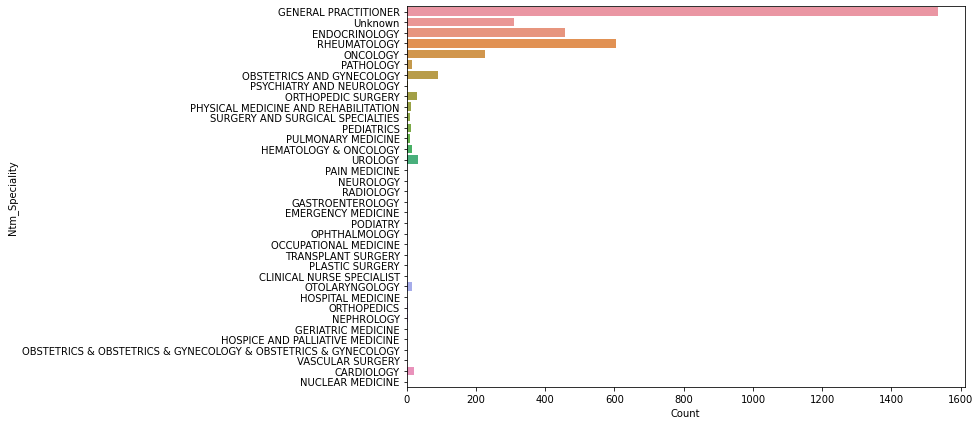
Distribution analysis, probability and outliers of ‘Dexa\_Freq\_During\_Rx’

Analysis of distribution, probability and outliers of ‘Count\_Of\_Risks’

Count of unique variables in each item. (sample)



**DATA INTAKE REPORT**

Name: Healthcare project

Report date: 08/04/2022

Internship Batch: [LISUM07](https://canvas.instructure.com/courses/3110044)

Version: 1.0

Data intake by: PACHACUTEAM

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:** https://github.com/And2300/Healthcare\_PersistencyDrug