

# Data Intake Report

Project Name: Healthcare – Persistency of a drug  
Report date: 20.04.2021  
Internship Batch: LISP01 & LICAN01  
Version: 1.0  
Data intake by: DG\_team\_project\_PL-RO-KSA-EGY

## Tabular data details:

### Healthcare\_dataset

Total number of observations	3424
Total number of files	2
Total number of features	69
Base format of the file	.xlsx
Size of the data	898 KB

## File ingestion details:

Programming language used: *Python*

Libraries used for accessing the dataset: *Pandas*

Dataset represented in python: *Data frame*. A Data frame is a two-dimensional data structure, in which data is aligned in a tabular form in rows and columns.

Parameters used for reading: *file\_name, sheet\_name, index\_col=None* (in order not to add another id column as it already exists)

## Dataset in python:

	Ptid	Persistency_Flag	Gender	Race	Ethnicity	Region	Age_Bucket	Ntm_Speciality	Ntm_Specialist_Flag	Ntm_Speciality_Bucket	...	Risk_
0	P1	Persistent	Male	Caucasian	Not Hispanic	West	>75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...	
1	P2	Non-Persistent	Male	Asian	Not Hispanic	West	55-65	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...	
2	P3	Non-Persistent	Female	Other/Unknown	Hispanic	Midwest	65-75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...	
3	P4	Non-Persistent	Female	Caucasian	Not Hispanic	Midwest	>75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...	
4	P5	Non-Persistent	Female	Caucasian	Not Hispanic	Midwest	>75	GENERAL PRACTITIONER	Others	OB/GYN/Others/PCP/Unknown	...	

5 rows × 69 columns