Data Intake Report

Name: Toy dataset
Report date: 04.07.2021
Internship Batch: LISUM01
Version:<1.0>
Data intake by: Askhat Maratov
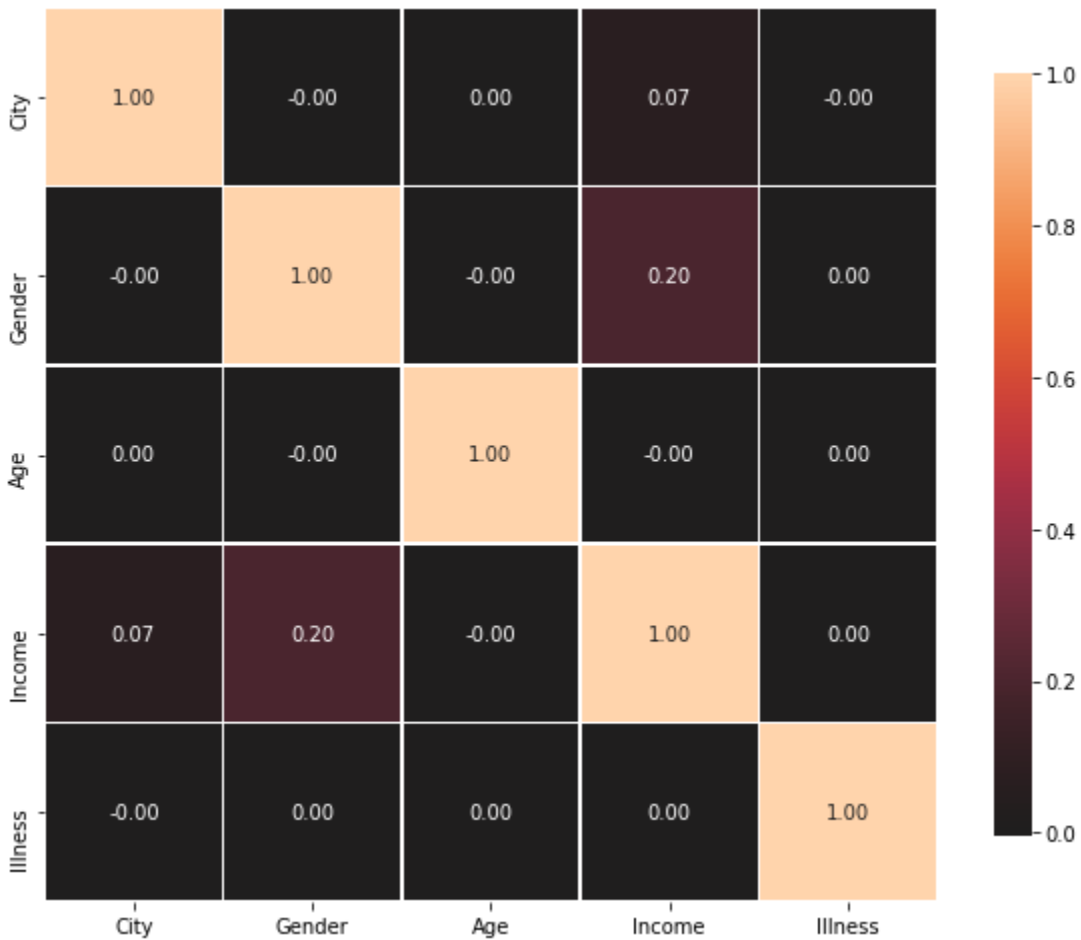Data intake reviewer: DataGlacier
Data storage location: https://github.com/TheKing-hub/Week-4

Tabular data details:

Total number of observations	150000
Total number of files	2
Total number of features	5
Base format of the file	csv, ipynb
Size of the data	5.7 MB

My goal was to identify how all features affect illness.
I made a heatmap correlation to identify the dependence among all features

The data here is clean, so I used logistic regression with multiple inputs, no NA values, no columns with a strong correlation. So I made a prediction using Logistic regression model with 4 inputs except column "number"