**

Technical Manual

PDF TimeSheet Reader Tool

AyaHealthcare

CR IT Team

8/18/2022

Content Table

[Change Log 3](#_Toc111791636)

[Description 4](#_Toc111791637)

[Tool WorkFlow 5](#_Toc111791638)

[Pseudo Code 6](#_Toc111791639)

# Change Log

|  |  |  |  |
| --- | --- | --- | --- |
| **Reason** | **Date** | **User** | **Comments** |
| Version 1 | 08/18/2021 | Felipe Gonzalez | Manual creation |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## Description

This is a Python program that scraps data from PDF files, those files are managed by users in San Diego and users from the Data Entry team.

It uses Tabula to read the PDF file, and Regex to scrape the information on the PDF file pages. Pandas for getting the data into Excel. Finally, Tkinter for creating windows .exe application.

## Diagram Description automatically generatedTool WorkFlow

## Pseudo Code

|  |
| --- |
| Start  Receive the report type.  Receive file path.  Read a .json file.  Create an empty dataframe.  Generate audit file.  Read the PDF file.  For each nurse:  Get the nurse name.  Get the GLCODE.  Get the date and time.  Get Comments.  Write the information on the dataframe.  Next nurse.  Write to excel file.  Save Excel file.  Copy the dataframe in a new one.  Empty the old data frame.  If new dataframe is empty:  Return false  Else:  Return true  End |
|  |