

# Release Notes

## V1.0

### ETL Foreclosure Data

Muhammad Kashif Irshad

[Kashif.ir@gmail.com](mailto:Kashif.ir@gmail.com)

## Contents

Introduction .....	3
Software Configuration.....	3
• Python: .....	3
• PIP .....	3
• Python APIs .....	3
• Chrome Extension and Script.....	3
Project Directory Structure .....	4
Process of transforming Sherrif's PDF to required EXCEL.....	4

## Introduction

This project is initiated in-order to automate the data collection of Sherrif's provided property list. Sherrif's property lists are availed in pdf format which will be used as input and an output excel file will be generated to summarize the stats available on two of the websites:

- <https://zillow.com>
- <http://www3.nccde.org/parcel/search/default.aspx>

## Software Configuration

- Python:
- Install python from the following url:
  - <https://realpython.com/installing-python/>
- Make sure to check the checkbox to add python path to system path variable so that python command may be execute from path in the system. There is also a checkbox to install python for all of the users on machine, better to select that as well.

Note: For following installations, setup.bat is created in project folder, which can be executed to install all rest of the following dependencies.

In case bat file fails, detail of installation is provided below

- PIP
- Install PIP and below modules by using following commands or by running setup.bat file in project directory.
- --user can be added at the end of the command if python is installed for current user only
  - `python get-pip.py --user`
- Python APIs
  - `pip install xlrd --user`
  - `pip install xlwt --user`
  - `pip install progressbar --user`
- Chrome Extension and Script
- Add following extension to chrome:

<https://chrome.google.com/webstore/detail/custom-javascript-for-web/poakhIngfciodnhlhgnaaelnpjijja?hl=en>

- Open the url <http://www3.nccde.org/parcel/search/default.aspx> in chrome
- Click on cjs icon to open setting
- Paste the code from "chrome cjs script.js" text file from the project directory in the text area and make sure the highlighted fields shown in image below and click save.



## Project Directory Structure

Project Structure	Type	Description
chrome cjs script.js	File	Script file to add to CJS Chrome Extension
db	Directory	SQLITE Database Folder
docs	Directory	Constains Documents related to project
output	Directory	Generated excel files are saved to this directory
processed	Directory	Original files are saved to this directory after processing
program.py	File	Main Python Program File
setup.bat	File	Batch File to Install dependencies of Python for this project
Start-Program.bat	File	Batch file to start execution of program
to_be_processed	Directory	Files to be processed will be copied to this directory

## Process of transforming Sherrif's PDF to required EXCEL

- Go to the following web page:
  - <https://smallpdf.com/pdf-to-excel>
- Upload the PDF file to the above link and it will convert it to an excel file which can be downloaded. This excel file generated from above website from the pdf will be used as input to this project
- Copy the converted excel file to the project directory "\\to\_be\_processed"
- Execute the batch file to start the process "Start-Program.bat" present in the project directory

- Open the Google Chrome and open the following website which will be used by python program to fetch information, automatically.
  - <http://www3.nccde.org/parcel/search/default.aspx>
- Progress of the program will be shown on console as status bar
- Any alert message will also be shown on console
- After progress is meet to 100% the resultant file will be present in the project directory “\output”
- Source excel file will be copied to project directory “processed”