User Guide

# Code files

1. Web Scraping code
2. Data Cleaning code
3. Data Analysis code
4. Geo-plot code

# Web Scraping Code

## Description

* This code file scrapes the data from the seed data website - [*https://world-crops.com/*](https://world-crops.com/)
* It converts the data into a structured format (pandas data-frame) using python lists, dictionaries and pandas dataframe.

## Input

* This file needs **no input**. The seed data URL is given in the code itself.

## Output

* The output is a csv file “*Web Scraping Output File name here*”. This file is to be used as input for Data Cleaning Code.

# Data Cleaning Code

## Description

* This code file removes any noise, extracts the key words form the textual data.
* It replaces the text data in the data-frame with key words to facilitate further analysis

## Input

* This file needs **“***Web Scraping Output File name here***”** as the input file.

## Output

* The output is a csv file “*Data Cleaning Output File name here*”. This file is to be used as input for EDA.

# Data Analysis Code

## Description

* This code file performs EDA on the clean data set and presents any findings.

## Input

* This file needs **“***Data Cleaning Output File name here***”** as the input file.

## Output

* The output is a RMD file which has been prepared based on the charts created in the code

# Geo-Plot Code

## Description

* This code file performs EDA on crop origin and presents geo-plot of crops based on origin.

## Input

* This file needs **“***countries\_regions.csv” and “crops\_regions\_tables.csv”*as the input files.

## Output

* The output is “*crop\_origins\_plot.html*”. It is a world map that shows the crops originated in that region when you hover over it.