Tosca Voogd Manual Kickstarter Project

This manual describes the steps that have to be taken to be able to run the tool. The scripts include a scraper, clean file, merge file, and analysis file. The scraper gathers data from the Kickstarter website.

Install Python and Selenium
Python has to be installed in order to follow the steps → https://www.python.org/downloads/. Selenium is used for web scraping, please install the Selenium Chromedriver (using Google Chrome): https://chromedriver.chromium.org/downloads → important here is to remember where you save it, you need the path later.

PIP install
pip install selenium
pip install numpy
pip install pandas
pip install request
pip install nltk
pip install spacy

Kickstarter_projectinfo.py
Make sure that you are in the '1 Scraper' folder. Here, you can run the command 'python kickstarter_projectinfo.py' that runs the script. In the terminal, it will ask you: *'Enter Number of pages to scrape (1 page = 12 entries):'*, here, you fill in the number of pages you want to scrape. This will create a file named kickstarter_data.xlsx with the project details.

Kickstarter_description_comments.ipynb
In this file, some adaptions have to be made in order to run it on your own device.
First, change the path of kick to the file kickstarter_data.xlsx.

```
kick = pd.read_excel('/Users/toscavoogd/Git_thesis_kick/Kickstarter_thesis/1 Scraper/kickstarter_data.xlsx')
```

Then, change the PATH of the chromedriver to where you located the webdriver. Also, you can install a proxy network in the chromedriver, but this is not obliged (else: leave it out).

```
PATH = "/Users/toscavoogd/chromedriver"

options = {
    'proxy':
    {
        'http': 'http://lum-customer-hvanl-zone-kickstarter-country-nl:kw0acksxm1rl@zproxy.lum-superproxy.io:22225',
        'https': 'https://lum-customer-hvanl-zone-kickstarter-country-nl:kw0acksxm1rl@zproxy.lum-superproxy.io:22225'
    },
}
```

In addition, be aware to change all paths to the ones of your device.

Finally, it will give you kickstarter_description_comments.xlsx (with the project details, appended with the comments and description) and comment_count.xlsx.

Clean.ipynb
The dataset that is used here is kickstarter_description_comments.xlsx. In this file, the textual data is cleaned. Be aware of changing all paths to your own device. This file will give you data_commentsdescription.xlsx and cleaned_data.xlsx.

Sentiment.ipynb
This script makes the comments ready for sentiment analysis by splitting all the comments. The file test_comments1 is the output.

Install Alteryx
For the next steps, Alteryx is used, which can be installed here: https://www.alteryx.com/sparked#. Be aware: these steps have to be performed on a windows device.

- Sentiment_analysis_tosca.yxmd
- Innovativeness_tosca.yxmd
- Sustainability_comments.yxmd
- Sustainability_description.yxmd

Merge.ipynb
The steps conducted in Alteryx gave the following datasets: sustainability_data, innovation_data, and sentiment_data. In the merge file, these are merged into one dataframe together with the Kickstarter data. Be aware of changing the paths to your own. This script will return merged_data.xlsx.

Install R
For the final step, R (https://cran.r-project.org) and R Studio (http://www.rstudio.com/ide) need to be installed. In the correlation_regression_kickstarter.R file, the correlations are checked and regressions are made.