WQD7005 Data Mining

Milestone 1: Web Crawling the Real Time Data

Github link: https://github.com/QiaoHui213/data_mining

Video 1: https://www.loom.com/share/f505c53d852e433fa4ab62cddbedd17e

Data is crawled from the website https://www.investing.com/commodities/gold-historical-data

This website showed the gold future historical data and the data is crawled from this website to do further prediction.

```python
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
import time

driver = webdriver.Chrome(executable_path="C:\\Users\\User\\Desktop\\data_mini
ng\\chromedriver.exe")

driver.set_page_load_timeout(30)
driver.get("https://www.investing.com/commodities/gold-historical-data")

for i in range(5):
    dateRangeElements = driver.find_elements_by_xpath("//div[@id='widgetFieldD
ateRange']")

    if len(dateRangeElements) > 0:
        dateRangeElements[0].click()
        break
    else:
        print(str(i + 1) + " trial, can't find element")

    time.sleep(1)
    i = i + 1

elementStartDate = driver.find_element_by_xpath("//input[@id='startDate']")
elementStartDate.clear()
elementStartDate.send_keys("01/01/2019")

driver.find_element_by_xpath("//a[@id='applyBtn']").click()
time.sleep(3)

tableData = driver.find_elements_by_xpath("//table[@id='curr_table']/tbody/tr/
td")
```

```python
resultFile = open("goldprice_" + time.strftime("%d-%b-%Y %H-%M-%S", time.local
time()) + ".csv", 'w')

resultFile.write("Date,Price,Open,High,Low,Volume,Change %\n")

column = 1
for data in tableData:
    if column != 7:
        resultFile.write((data.text).replace(',', '') + ", ")
        column = column + 1
    else:
        resultFile.write(data.text + "\n")
        column = 1

resultFile.close()

driver.close()
```