|  |
| --- |
| WQD 7005 Data Mining |
| **By**  **Liu, HongYang (17201091/1)**  **Part of Work(****Not included my Partner)** |



**A close up of a logo

Description automatically generated**

**MASTER OF DATA SCIENCE**

**FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY**

**UNIVERSITY OF MALAYA**

**SEMESTER 2, 2019/2020**

**Assignment Title**

**Monitoring the oil price, correlate with the gold price, and predict what will be the oil price**

|  |  |
| --- | --- |
| **NAME** | **MATRIC NUMBER** |
| **LIU,HONGYANG** | **17201091** |
| **Gunasegarran** | **17043640** |

# Title

Monitoring the oil price, correlate with the gold price, and predict what will be the oil price

# Problem Statement

During the pandemic of Covid-19, we have seen the oil price dropped dramatically, and some other commodities like the gold price have also been impacted. Oil and gold are both important in people’s daily. Especially for Modern people, these people focus on oil prices because most of them driving cars, and they also focus on the gold price, as this price may influence their fortune. Despite we have known the oil price may be influenced by the gold price, while most of us have little knowledge of their relationship. This relationship may help people do financial management.

# Aim

In this project, the main purpose is to find out the correlation between the gold price and oil price. We would utilize python and it is powerful libraries to verify our hypothesis and evaluate their relevance.

# Project Out

The online crude oil price website deployed on Heroku website.

##### Website Address: <https://mighty-ridge-02940.herokuapp.com/>

# Tools

To achieve our goal, we would utilize these tools:

|  |  |
| --- | --- |
| **Tool or package** | **Purpose** |
| Beautiful Soup | Web Scraping |
| Pandas | Data processing |
| Scikit-Learn | Machine Learning |
| Seaborn & Plotly | Data visualization |
| Dash | Build Interactive web app |
| Huroku | Website deployment |

# Datasets:

The data sets comes from the index mundi website.

Each of the data sets has three columns, the month, price and change.

1. The month represents data
2. The price is the quantity of the commodity price.
3. The change is the change rate that the current month value changed compared with las month

A screenshot of a cell phone

Description automatically generated

**Figure 1: Data sets**

# 

# Procedures & Results:

There are five procedures during our project:

1. Milestone 1:

Web crawling the data by using Python

A screenshot of a computer screen

Description automatically generated

**Figure 2: Milestone 1 results**

1. Milestone 2:

Store data into hive data warehouse

A screenshot of a computer

Description automatically generated

**Figure 3: Milestone 2 results**

1. Milestone 3:

Accessing hive data warehouse or data lake using Python or R

A screenshot of a cell phone

Description automatically generated

**Figure 4: Milestone 3 results**

1. Milestone 4:

Interpretation of data & Communication of Insights of data

A picture containing table, water, sitting, kitchen

Description automatically generated

**Figure 5: Milestone 4 results**

1. Milestone 5:

A screenshot of a social media post

Description automatically generated

**Figure 6: Milestone 5 results**

A close up of a map

Description automatically generated

**Figure 7: Milestone 5 results**

# Findings & Conclusion

From the above work, we have found that the oil price had a positive correlation with gold prices from 1992 to 2008. During that time, the oil price increased rapidly and the gold price had also shown similar trends. While during 2008 till 2013, the oil price first drops dramatically and then soared to around 250 dollars. The gold price was still increased during that time. The last phase is from 2013 till now, the oil price has fluctuated, but the gold price has increased to the peak.

Despite we have found out the positive relationship between the oil price and gold price before 2008 using the data sets observations, there is still a deficiency that we still need to detect the impacts on the oil price during 2013 and 2020. In future work, we would propose more relevant variables and verity them to find out their relevance.