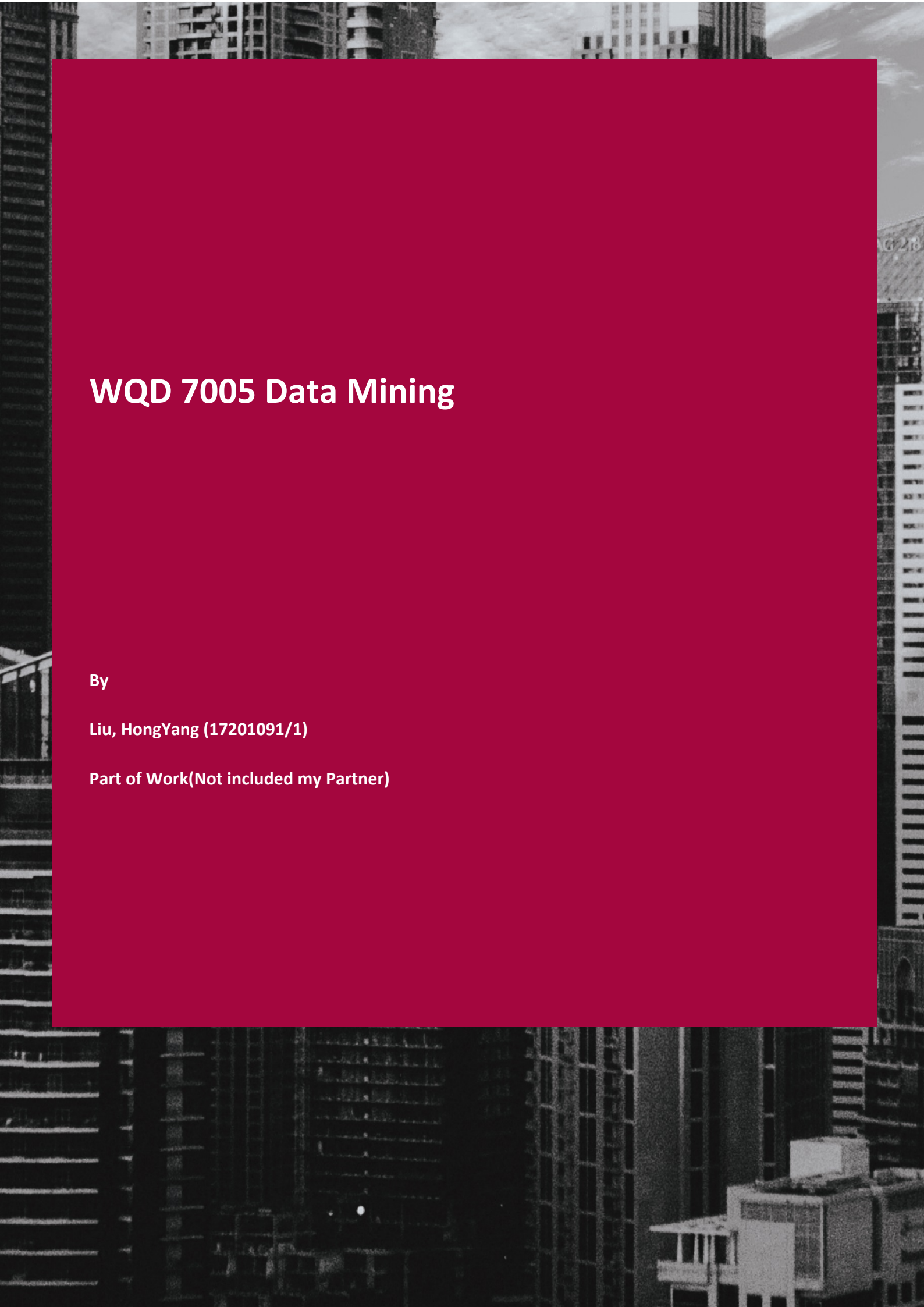

WQD 7005 Data Mining

By

Liu, HongYang (17201091/1)

Part of Work(Not included my Partner)

UNIVERSITY OF MALAYA

KUALA LUMPUR

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

Month	Price	Change
Nov 2019	116.91	-
Dec 2019	119.91	2.57 %
Jan 2020	119.55	-0.30 %
Feb 2020	111.09	-7.08 %
Mar 2020	93.87	-15.50 %
Apr 2020	84.16	-10.34 %

Figure 1: Data sets

Procedures & Results:

There are five procedures during our project:

1. Milestone 1:

Web crawling the data by using Python

Data TABLE

Year	OilPrice	Gold Price
Jan 1992	40.23	369.05
Feb 1992	41.09	352.33
Mar 1992	41.17	362.53
Apr 1992	44.09	394.73
May 1992	45.57	389.32
Jun 1992	47.92	380.74

Figure 2: Milestone 1 results

2. Milestone 2:

Store data into hive data warehouse

```
hive> LOAD data inpath '/usr/oilCrawl' into table oilCrawl;
Loading data to table default.oilcrawl
[Warning] could not update stats.
OK
Time taken: 25.2 seconds
hive> select * from oilCrawl;
OK
        Year    AverageClosing Price    Year Open       Year High       Year LowYear Close      Annual% Change
0       2020    $51.56  $61.18  $63.27  $31.04  $31.04  -49.16%
1       2019    $57.05  $46.54  $66.30  $46.54  $61.06  34.46%
2       2018    $64.90  $60.37  $76.41  $42.53  $45.41  -24.84%
3       2017    $50.84  $52.33  $60.42  $42.53  $60.42  12.47%
4       2016    $43.58  $36.76  $54.06  $26.21  $53.72  45.03%
5       2015    $48.72  $52.72  $61.43  $34.73  $37.04  -30.70%
6       2014    $93.17  $95.14  $107.95 $53.45  $53.45  -45.55%
7       2013    $97.98  $93.14  $110.62 $86.65  $98.17  6.90%
8       2012    $94.05  $102.96 $109.39 $77.72  $91.83  -7.08%
9       2011    $94.88  $91.59  $113.39 $75.40  $98.83  8.15%
10      2010    $79.48  $81.52  $91.48  $64.78  $91.38  15.10%
11      2009    $61.95  $46.17  $81.03  $34.03  $79.39  78.00%
12      2008    $99.67  $99.64  $145.31 $30.28  $44.60  -53.52%
13      2007    $72.34  $60.77  $99.16  $50.51  $95.95  57.68%
14      2006    $66.05  $63.11  $77.05  $55.90  $60.85  -0.34%
15      2005    $56.64  $42.16  $69.91  $42.16  $61.06  40.82%
16      2004    $41.51  $33.71  $56.37  $32.49  $43.36  33.37%
17      2003    $31.08  $31.97  $37.96  $25.25  $32.51  4.17%
18      2002    $26.19  $21.13  $32.68  $18.02  $31.21  56.36%
19      2001    $25.98  $27.29  $32.21  $17.50  $19.96  -25.30%
20      2000    $30.38  $25.56  $37.22  $23.91  $26.72  3.73%
21      1999    $19.35  $12.42  $28.03  $11.38  $25.76  112.19%
22      1998    $14.42  $17.41  $17.93  $10.82  $12.14  -31.22%
23      1997    $20.61  $25.55  $26.55  $17.60  $17.65  -31.85%
24      1996    $22.12  $19.83  $26.55  $17.33  $25.90  32.55%
25      1995    $18.43  $17.45  $20.53  $16.86  $19.54  9.96%
26      1994    $17.20  $14.52  $20.72  $13.89  $17.77  25.23%
27      1993    $18.43  $19.03  $21.05  $13.98  $14.19  -27.19%
28      1992    $20.58  $19.43  $23.03  $17.89  $19.49  1.78%
29      1991    $21.54  $26.53  $32.25  $17.43  $19.15  -32.76%
30      1990    $24.53  $22.88  $41.07  $15.43  $28.48  30.40%
31      1989    $19.64  $17.38  $24.62  $16.99  $21.84  27.57%
32      1988    $15.97  $17.77  $18.54  $12.58  $17.12  2.27%
33      1987    $19.20  $18.13  $22.44  $15.12  $16.74  -6.64%
Time taken: 3.485 seconds, Fetched: 35 row(s)
```

Figure 3: Milestone 2 results

3. Milestone 3:

Accessing hive data warehouse or data lake using Python or R

```
> dbGetQuery(conn, "select * from oilcrawldata")
        index year  averageclosingprice  yearopen  yearhigh  yearlow
1   test file <NA>                <NA>      <NA>      <NA>     <NA>
2                Year AverageClosing Price Year Open Year High Year Low
3             0 2020               $51.56    $61.18    $63.27   $31.04
4             1 2019               $57.05    $46.54    $66.30   $46.54
5             2 2018               $64.90    $60.37    $76.41   $42.53
6             3 2017               $50.84    $52.33    $60.42   $42.53
7             4 2016               $43.58    $36.76    $54.06   $26.21
8             5 2015               $48.72    $52.72    $61.43   $34.73
9             6 2014               $93.17    $95.14   $107.95   $53.45
10            7 2013               $97.98    $93.14   $110.62   $86.65
11            8 2012               $94.05   $102.96   $109.39   $77.72
12            9 2011               $94.88    $91.59   $113.39   $75.40
13           10 2010               $79.48    $81.52    $91.48   $64.78
14           11 2009               $61.95    $46.17    $81.03   $34.03
15           12 2008               $99.67    $99.64   $145.31   $30.28
16           13 2007               $72.34    $60.77    $99.16   $50.51
17           14 2006               $66.05    $63.11    $77.05   $55.90
18           15 2005               $56.64    $42.16    $69.91   $42.16
19           16 2004               $41.51    $33.71    $56.37   $32.49
20           17 2003               $31.08    $31.97    $37.96   $25.25
21           18 2002               $26.19    $21.13    $32.68   $18.02
22           19 2001               $25.98    $27.29    $32.21   $17.50
23           20 2000               $30.38    $25.56    $37.22   $23.91
24           21 1999               $19.35    $12.42    $28.03   $11.38
25           22 1998               $14.42    $17.41    $17.93   $10.82
26           23 1997               $20.61    $25.55    $26.55   $17.60
27           24 1996               $22.12    $19.83    $26.55   $17.33
28           25 1995               $18.43    $17.45    $20.53   $16.86
29           26 1994               $17.20    $14.52    $20.72   $13.89
30           27 1993               $18.43    $19.03    $21.05   $13.98
31           28 1992               $20.58    $19.43    $23.03   $17.89
32           29 1991               $21.54    $26.53    $32.25   $17.43
33           30 1990               $24.53    $22.88    $41.07   $15.43
34           31 1989               $19.64    $17.38    $24.62   $16.99
35           32 1988               $15.97    $17.77    $18.54   $12.58
36           33 1987               $19.20    $18.13    $22.44   $15.12
```

Figure 4: Milestone 3 results

4. Milestone 4:

Interpretation of data & Communication of Insights of data

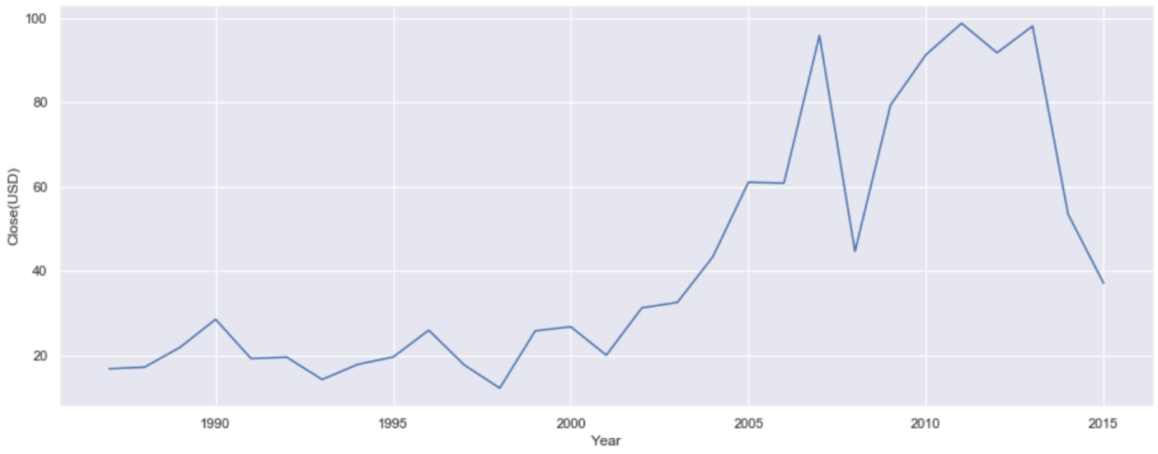

Figure 5: Milestone 4 results

5. Milestone 5:

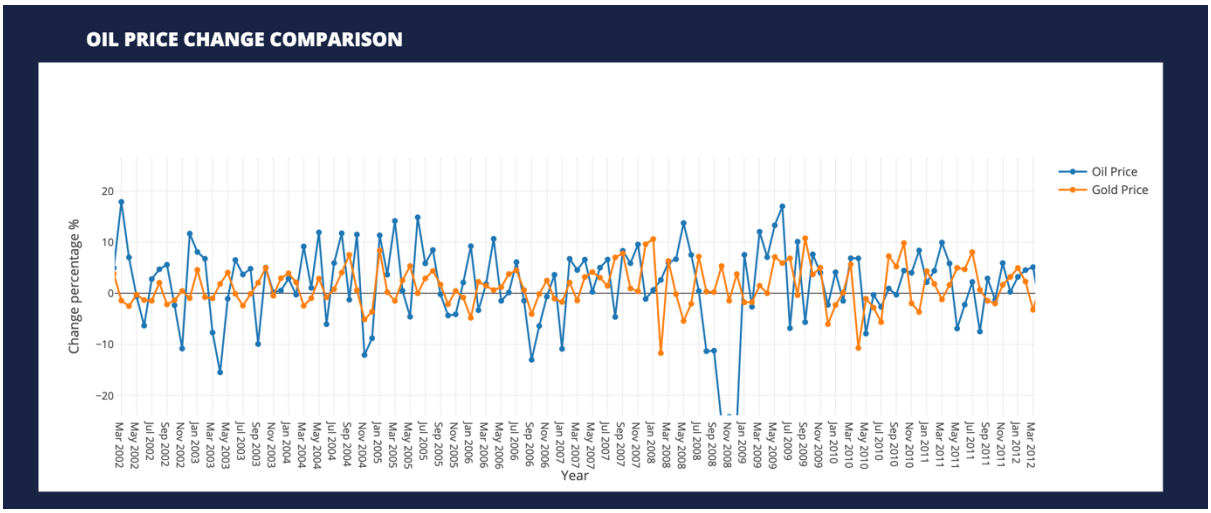

Figure 6: Milestone 5 results

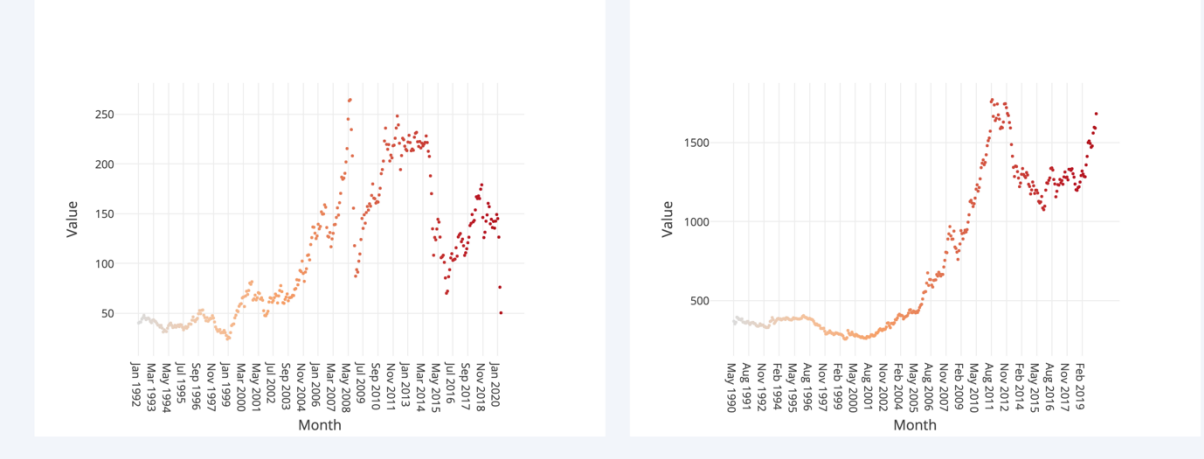

Figure 7: Milestone 5 results

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.