클라우드기반 데이터 보안 전문가 양성 과정

# 모듈 프로젝트 3

2021년 3월 19일

이름: 박진환

## ■ 프로젝트 분배

- 1. 프로젝트 참가자
  - A. 김승미
  - B. 박진환
  - C. 신선아
  - D. 이희석
- 2. 총괄 책임 → 박진환
- 3. 전체 프로젝트 → 박진환, 이희석, 신선아, 김승미
  - A. 주제 1 / 날씨 → 신선아, 김승미
    - i. 데이터 수집 → 신선아, 김승미
      - 1. Scrapy 작성 → 신선아, 김승미
      - 2. Pipeline 구축 → 박진환, 신선아, 김승미
      - 3. RDS 구축 → 박진환
    - ii. 기능 → 박진환, 신선아, 김승미
      - 1. 5개의 도시의 일-시간 별 온도 비교 그래프 출력 → 박진환, 신선아, 김승미
      - 2. 특정 도시의 일-시간 별 온도와 습도 그래프 출력 → 박진환, 신선아, 김승미
      - 3. 5개의 도시의 여러 정보를 JSON 형식으로 출력 → 박진환, 신선아, 김승미
  - B. 주제 2 / 암호화폐(주식) → 박진환, 이희석
    - i. 데이터 수집 <del>></del> 박진환, 이희석
      - 1. Scrapy 작성 → 이희석
      - 2. Pipeline 구축 → 박진환, 이희석

- 3. RDS 구축 → 박진환
- ii. 기능 → 박진환, 이희석
  - 1. 특정 암호화폐의 일-시간 별 가격 변동 그래프 출력 <del>></del> 박진환, 이희 석
  - 2. 3개의 암호화폐의 변동률 그래프 출력 → 박진환, 이희석
  - 3. 3개의 암호화폐의 여러 정보를 JSON 형식으로 출력 → 박진환, 이희 석
- 4. Python Django RestfulAPI 설계 및 구현 → 이희석, 신선아, 김승미
  - A. 서버 구축 및 API 설계 → 이희석, 신선아, 김승미
  - B. 기능 구현을 위한 코드 작성 → 박진환, 이희석
- 5. 마무리 검수 → 박진환, 이희석

# ■ 모듈 프로젝트 내용

```
1. 날씨
제
                                   2. 암호화폐(주식)
                                   1. 5개의 도시의 일-시간 별 온도 비교 그래프 출력
기
                                   2. 특정 도시의 일-시간 별 온도와 습도 그래프 출력
능
                                   3. 5개의 도시의 여러 정보를 JSON 형식으로 출력
날
씨
                                   1. 특정 암호화폐의 일-시간 별 가격 변동 그래프 출력
기
                                   2. 3개의 암호화폐의 변동률 그래프 출력
능
                                   3. 3개의 암호화폐의 여러 정보를 JSON 형식으로 출력
주
식
                                  1. Scrapy와 apscheduler를 활용하여 날씨 사이트에서 원하는 정보를 5분 간격으
설
계
                                                로 crawling
                                    import scrapy
from weatherbot.items import WeatherbotItem
날
                                    class WeatherbotsSpider(scrapy.Spider):
                                          custom_settings = {\'\]080IR': 'crawl_weatherbots1'}
allowed_domains = ['yahoo.com/news/weather/south-korea/seoul/seoul-1132599/', 'yahoo.com/news/weather/united-states/new-york/new-york-2459115', 'yahoo.co
start_urls = ['https://www.yahoo.com/news/weather/south-korea/seoul/seoul-1132599/', 'https://www.yahoo.com/news/weather/united-states/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new-york/new
씨
                                                  humid = response.xpath('//*/[@id="weather-detail"]/div/div[1]/div/div[2]/div/[2]/div[2]/text()').extract()
high_temp = response.xpath('//*/[@id="Lead-1-WeatherLocationAndTemperature"]/div/section[2]/div/div[2]/span[1]/text()').extract()
low_temp = response.xpath('//*/[@id="Lead-1-WeatherLocationAndTemperature"]/div/section[2]/div/div[2]/span[2]/text()').extract()
title = response.xpath('//*[@id="Lead-1-WeatherLocationAndTemperature"]/div/section[1]/div[1]/div/h1/text()').extract()
                                                   wind = response.xpath('//*[@id="weather-wind"]/div/div[1]/div[2]/div/p/span/text()').extract()
weather = response.xpath('//*[@id="Lead-1-WeatherLocationAndTemperature"]/div/section[2]/div/div[1]/span[2]/text()').extract()
                                                           item['temper'] = row[0]
item['humid'] = row[1]
                                                           item['high temp'] = row[2]
                                                           item['low_temp'] = row[3]
item['title'] = row[4]
item['wind'] = row[5]
                                                           item['weather'] = row[6]
                                        from apscheduler.schedulers.twisted import TwistedScheduler
from weatherbot.spiders.weatherbot import WeatherbotsSpider
                                        process = CrawlerProcess(get_project_settings())
                                        scheduler.start()
                                        process.start(False)
```

2. Scrapy의 Pipeline을 활용하여 Scrapy, Kafka, RDS를 연결

```
from kafka import Kafkafroducer
from json import dumps

import time

class WeatherbotPipeline:

def __init__(self):

self.producer = Kafkafroducer(acks=0,

compression_type='gzip',
bootstrap_servers=['localhost:9092'],

value_serializer=lambda x: dumps(x).encode('utf-8'))

def process_item(self, item, spider):

item = dict(item)

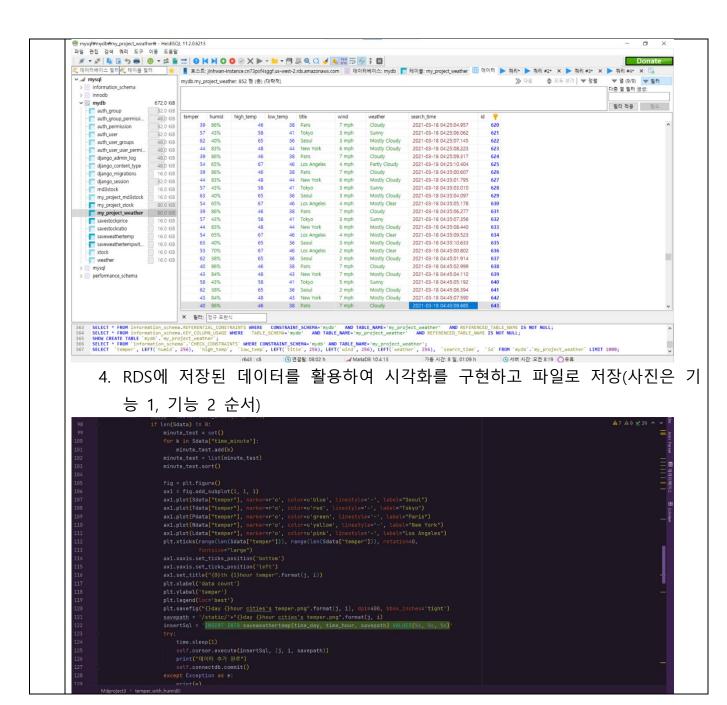
item("temper") = int(item("tiemper"))

item("tomper") = int(item("high_temp"))

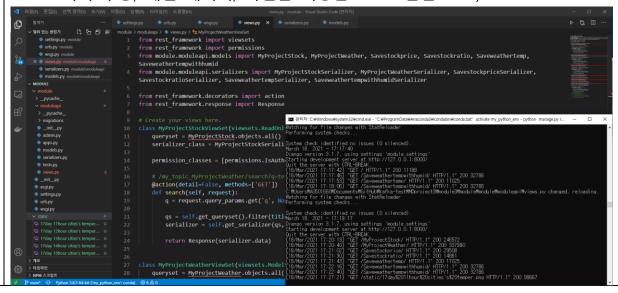
data = {"schema": {"type": "struct", "fields": [{"type": "int32", "optional": "false", "field": "temper"}, {"type": "string", "optiona
```

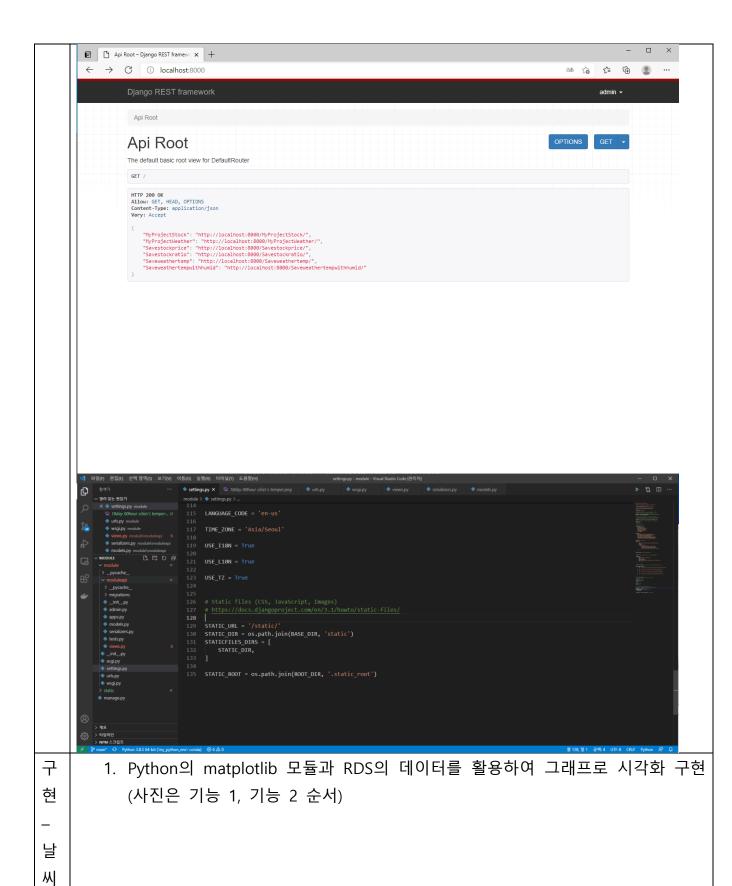
3. 스크래피를 통한 RDS 데이터 삽입을 확인(사진은 스크래피, 토픽 전달 확인, RDS 데이터 사진)

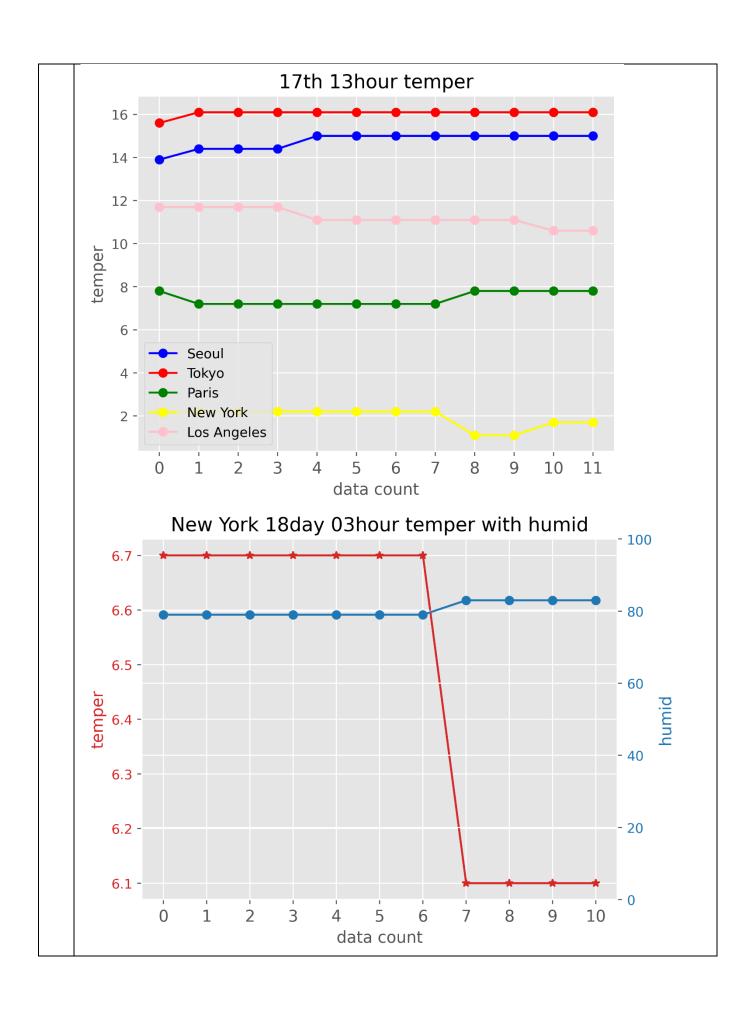
```
| THE **TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN**TOWN
```



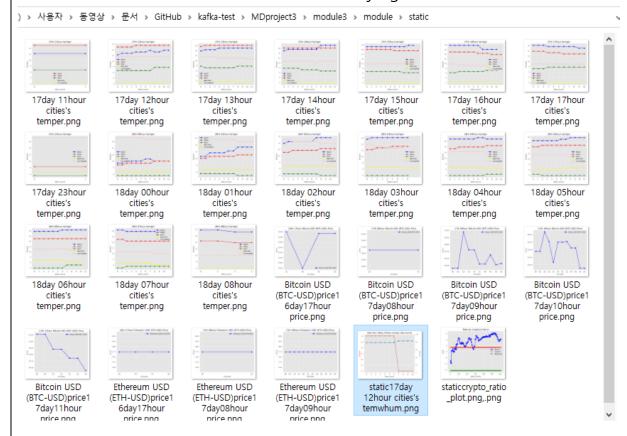
5. Python Django를 활용하여 RestAPI 구축하고 기능과의 연동 준비 완료(사진은 서버 구동, 메인 페이지, 사진을 저장한 Static 관련 코드)



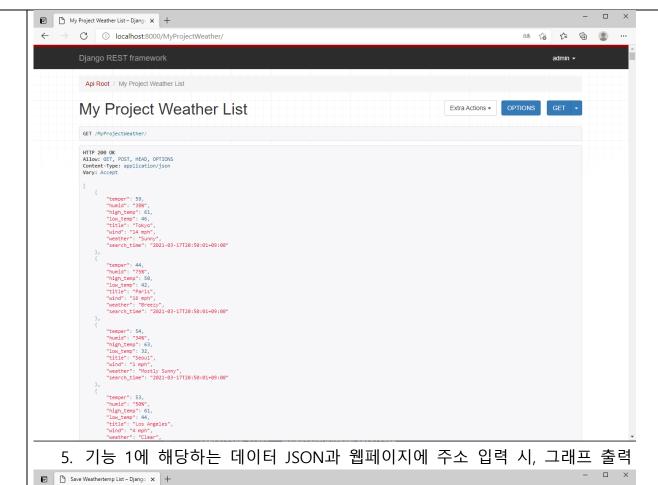


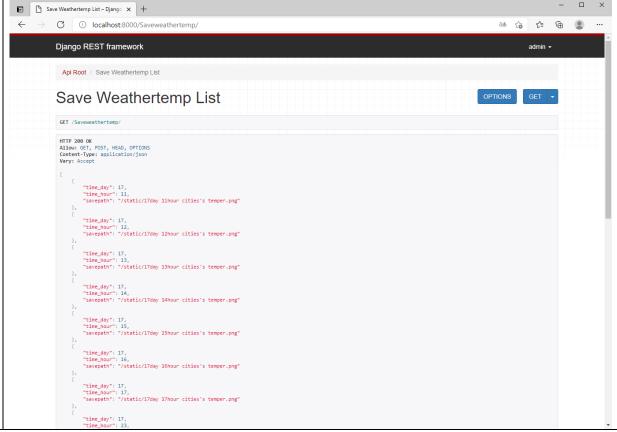


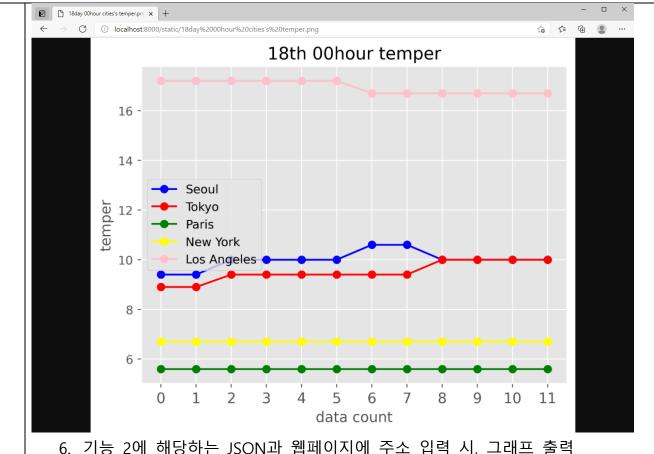
- 2. Python Django를 활용하여 RestfulAPI에 데이터 JSON파일과 경로 입력 시, 그 래프 출력 기능 구현
- 3. 그래프 출력을 위한 사진이 모여 있는 Django의 Static 폴더



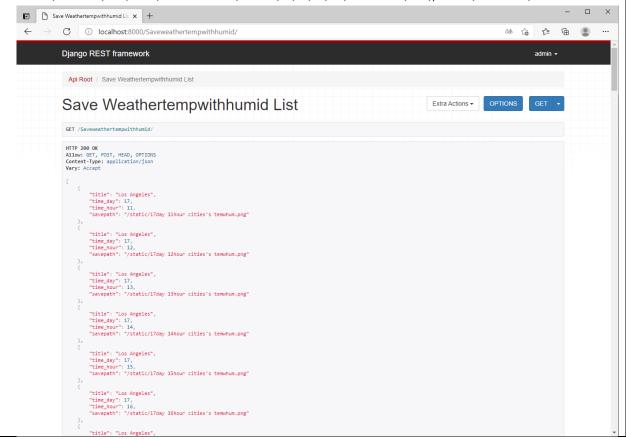
4. 기능 3에 해당하는 날씨 데이터 JSON 출력

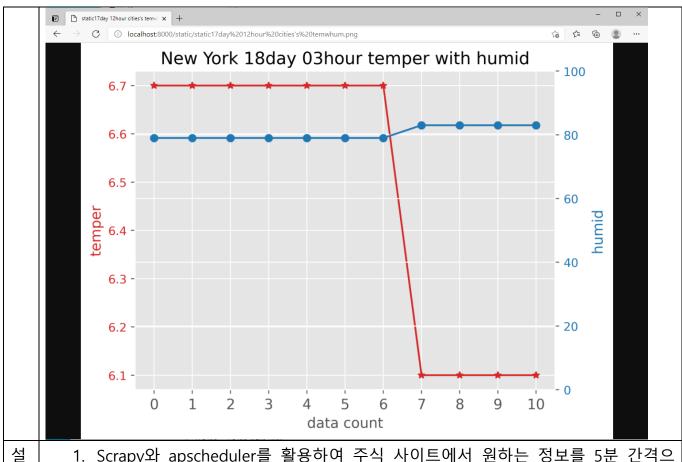






#### 6. 기능 2에 해당하는 JSON과 웹페이지에 주소 입력 시, 그래프 출력





1. Scrapy와 apscheduler를 활용하여 주식 사이트에서 원하는 정보를 5분 간격으로 crawling

계

암

호

화

폐

(주

식)

```
from scrapy.crawler import CrawlerProcess

from scrapy.utils.project import get_project_settings

from apscheduler.schedulers.twisted import TwistedScheduler

from stockscraper.spiders.stockbots import StockbotsSpider

process = CrawlerProcess(get_project_settings())

scheduler = TwistedScheduler()

scheduler.add_job(process.crawl, 'cron', args=[StockbotsSpider], minute='0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55')

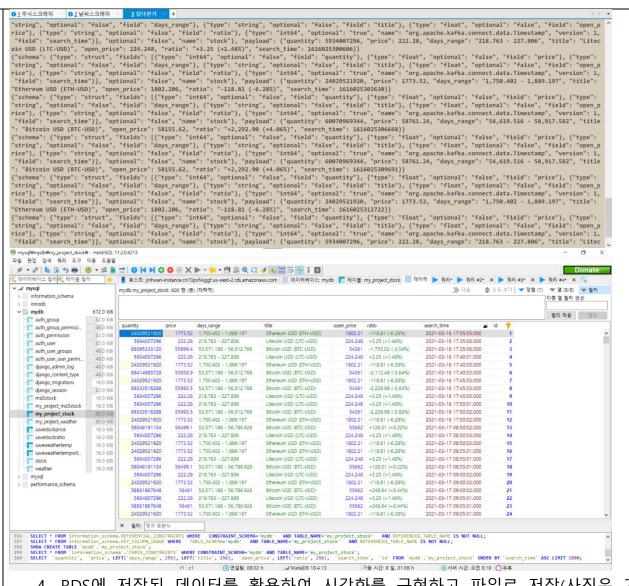
scheduler.start()

process.start(False)
```

2. Scrapy의 Pipeline을 활용하여 Scrapy, Kafka, RDS를 연결

3. 스크래피를 통한 RDS 데이터 삽입을 확인(사진은 스크래피, 토픽 전달 확인, RDS 데이터 저장 사진)

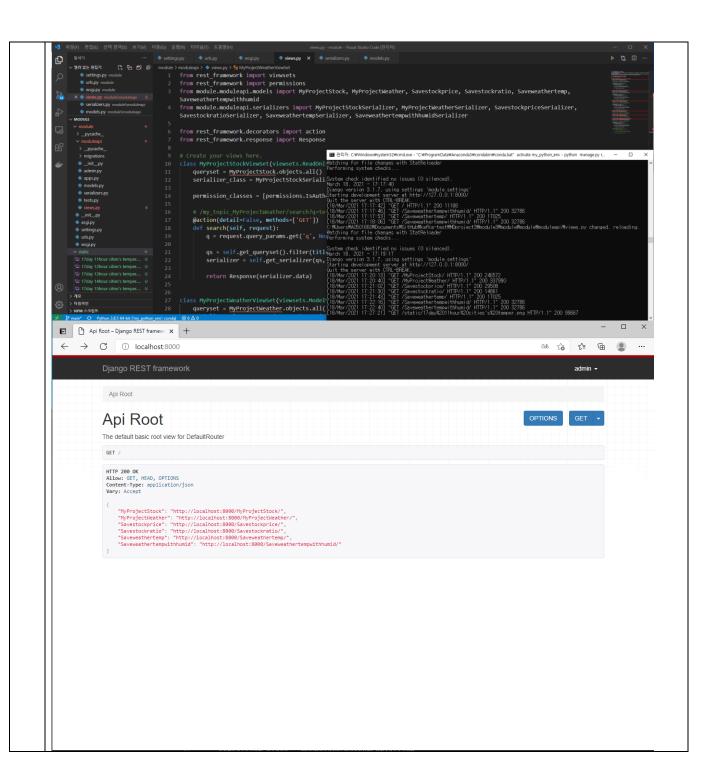
```
uest 6: MetadataRequest_V1(topics=['my_project_stock'])
2021-03-18 08:05:08 [kafka.protocol.parser] DEBUG: Received correlation id: 6
2021-03-18 08:05:08 [kafka.protocol.parser] DEBUG: Processing response MetadataResponse_v1
2021-03-18 08:05:08 [kafka.protocol.parser] DEBUG: Processing response MetadataResponse_v1
2021-03-18 08:05:08 [kafka.protocol.parser] DEBUG: More (Inode_id=0, host=ip-172-31-47-219.us-west-2.compute.internal:) port=0902, rack=None)], controller_id=0, to pics=[(error_code=0, topic='my_project_stock', is_internal=false, partitions=[(error_code=0, partition=0, leader=0, replicas=[0], isr=[0])]))
2021-03-18 08:05:08 [kafka.cluster] DEBUG: More duster metadata to ClusterMetadata(posets: 1, topics: 1, groups: 0)
2021-03-18 08:05:08 [kafka.protocol.parser] DEBUG: Sending neguest MetadataRequest_V1(topics=['my_project_stock']) to node 0
2021-03-18 08:05:08 [kafka.protocol.parser] DEBUG: Sending neguest MetadataRequest_V1(topics=['my_project_stock'])
2021-03-18 08:05:08 [kafka.protocol.parser] DEBUG: Received correlation id: 63
2021-03-18 08:05:08 [kafka.protocol.parser] DEBUG: Received correlation id: 63
2021-03-18 08:05:08 [kafka.protocol.parser] DEBUG: Processing response MetadataResponse_v1
2021-03-18 08:05:08 [kafka.protocol.parser] DEBUG: Received correlation id: 63
2021-03-18 08:05:08 [kafka.complotes] [more content of the more content of the m
```

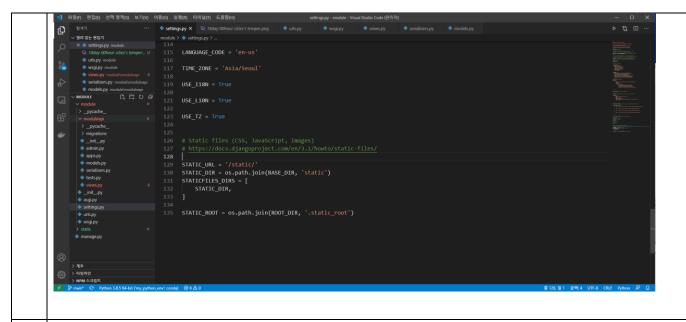


4. RDS에 저장된 데이터를 활용하여 시각화를 구현하고 파일로 저장(사진은 기능 1, 기능 2 순서)

```
plt.xticks(range(len(ndata["price"])), range(len(ndata["price"])), rotation=0, fontsize="large")
                         plt.xlabel('data check')
                        plt.ylabel('price')
plt.legend(loc='best')
                        plt.savefig(k+'price'+j+'day'+i+'hour price.png', dpi=400, bbox_inches='tight')
savepath = '/static/'+k+'price'+j+'day'+i+'hour price.png'
                         insertSql = 'INSERT INTO savestockprice(stock_title, time_day, time_hour, savepath) VALUES(%s, %s, %s, %s)
                              time.sleep(1)
       elif "Litecoin" in i:
   Ldf["ratio"] = df["ratio"][df["title"] == i]
Bdf["ratio"] = Bdf["ratio"].map(lambda x: self.ratio_parse(x))
Ldf["ratio"] = Ldf["ratio"].map(lambda x: self.ratio_parse(x))
Edf["ratio"] = Edf["ratio"].map(lambda x: self.ratio_parse(x))
ax1.plot(Bdf["ratio"], marker=r'o', color=u'blue', linestyle='-', label='Bitcoin')
ax1.plot(Ldf["ratio"], marker=r'+', color=u'red', linestyle='--', label='Litecoin')
ax1.plot(Edf["ratio"], marker=r'*', color=u'green', linestyle='--', label='Ethereum')
ax1.xaxis.set_ticks_position('bottom')
ax1.yaxis.set_ticks_position('left')
insertSql = """INSERT INTO savestockratio(datacount, savepath) VALUES(%s, %s)"""
```

5. Python Django를 활용하여 RestAPI 구축하고 기능과의 연동 준비 완료(사진은 서버 구동, 메인 페이지, 사진을 저장한 Static 관련 코드)

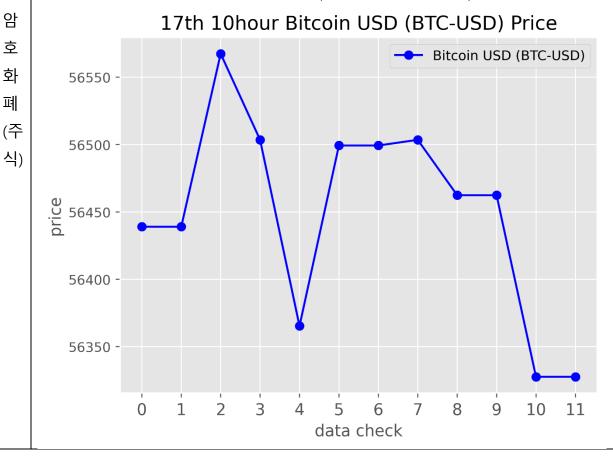


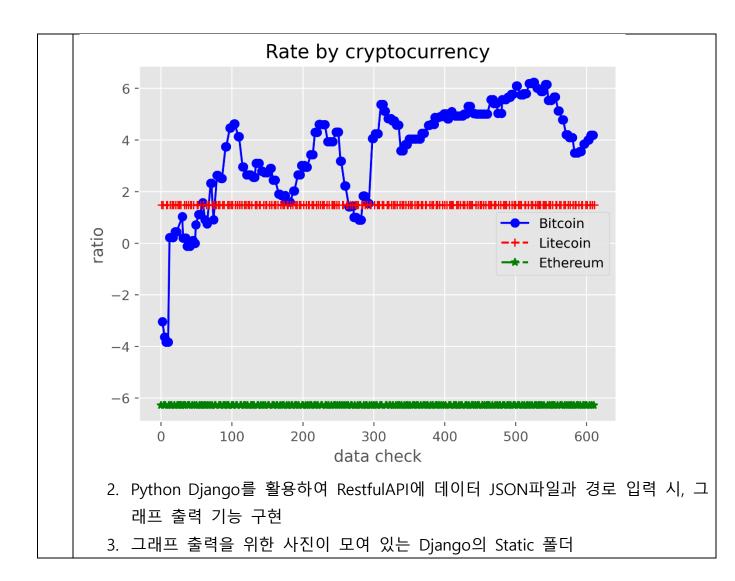


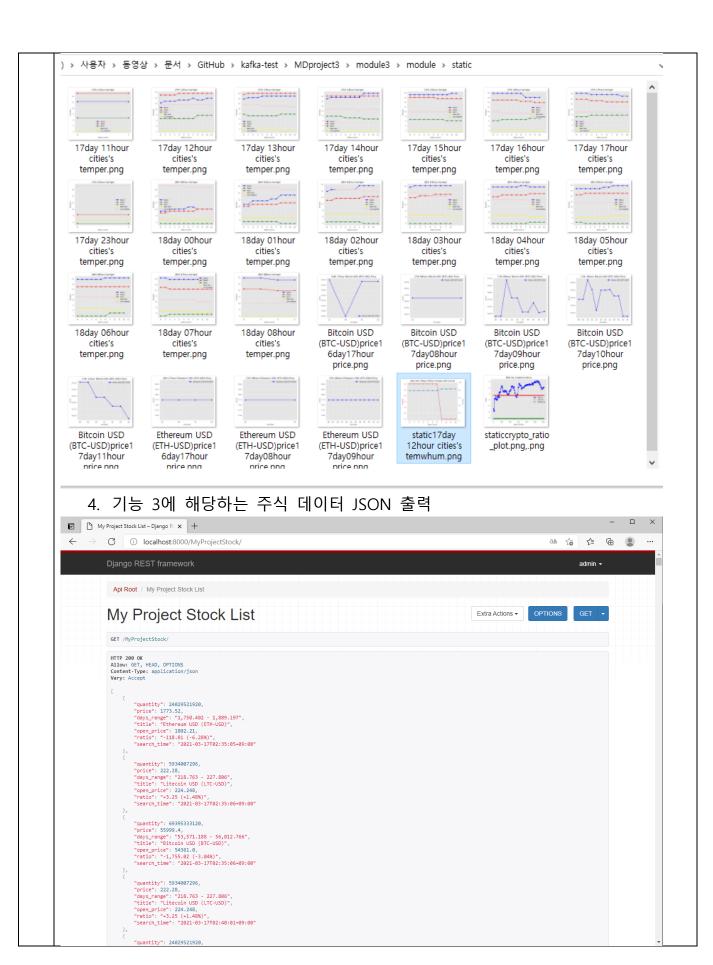
구

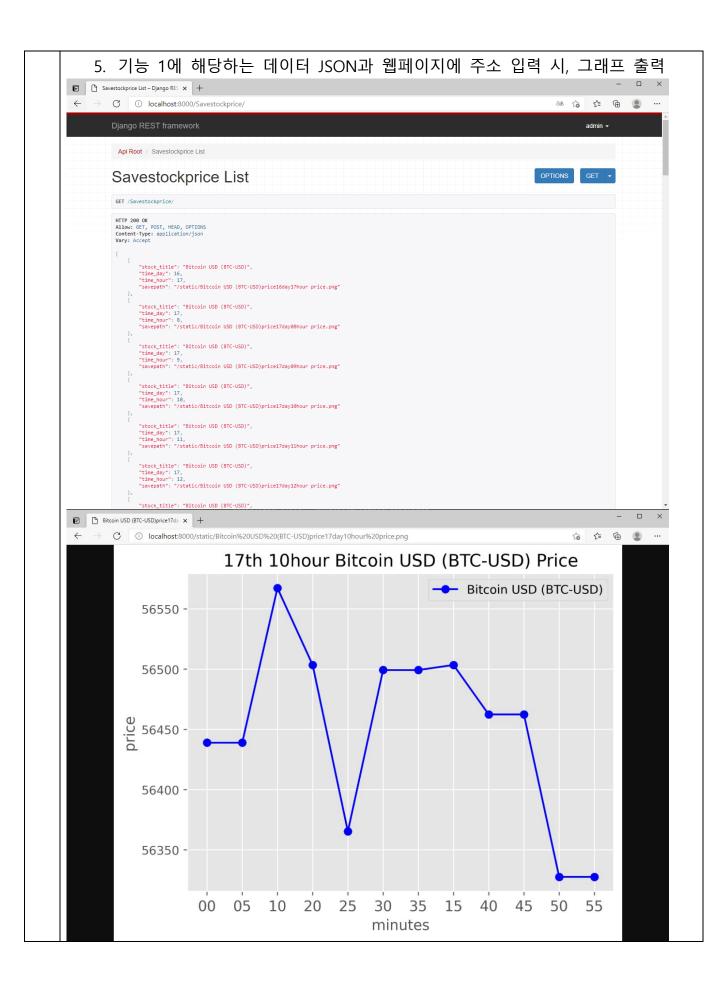
혀

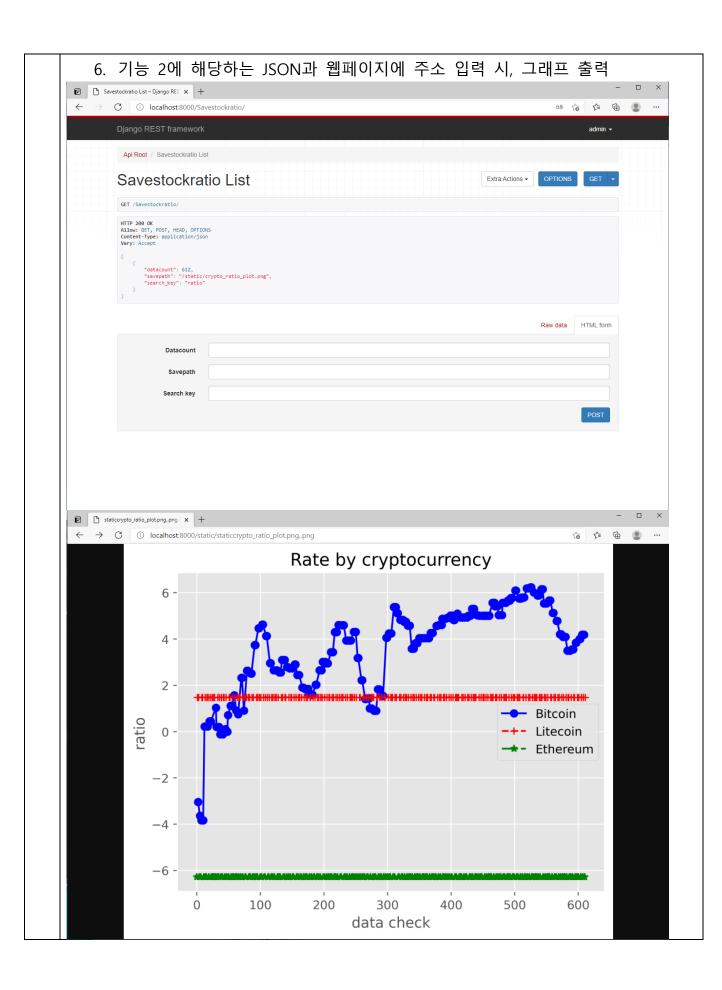
1. Python의 matplotlib 모듈과 RDS의 데이터를 활용하여 그래프로 시각화 구현 (사진은 기능 1, 기능 2 순서)(★기능 2의 경우 2개의 데이터가 동일한 데이터 를 가져오는 것을 확인하였으며, 수정해보려 했으나, 결국 실패했습니다.)











RDS	1. 주소 jinhwan-instance.cn73pxf4sggf.us-west-2.rds.amazonaws.com
정보	2. 계정 admin
	3. 비밀번호 1q2w3e4r5t

### ■ 프로젝트 후기

금번 프로젝트는 전체 과정에서 첫번째 팀 프로젝트였습니다. 이전에도 2인 규모의 소규모 과제를 진행한적이 있었지만, 프로젝트는 4명과 함께 이번 과정에서 배운 모든 것을 활용한 프로젝트였고, 제 인생에 있어서는 처음으로 총괄 책임을 맡아 진행해본 프로젝트였습니다.

처음에는 프로젝트를 진행하기 앞서 진행 방식을 정하고자(메인 & 서브 프로그래머 방식 / 개인 파트별로 나누어 이후 합쳐서 조율)하였으며, 개개인의 역량 향상을 위하여 개인 파트별로 나누어 이후 합쳐서 조율하는 방식을 택했습니다.

하지만 개개인의 역량 파악이 생각보다 덜 되었고, 프로젝트 구조 자체를 처음에 잘못 이해하여 파트별로 나누는 것에 있어서 문제가 발생하였고, 서로가 비슷한 속도로 진행되어 파트를 합쳐서 조율하는 것이 힘들 것으로 예상되었고, 앞서 정했던 진행 방식을 변경하여 메 인 & 서브 프로그래밍 방식과 파트별로 나누어 코딩하는 방식을 결 합하여, 제가 총괄 책임을 맡고, 조원 분들을 각자 메인 프로그래머로 지정하여 파트별로 작성을 부탁한 뒤, 필요한 곳마다 작성을 도우는 방식으로 진행을 하였습니다.

결과적으로, 좋은 점도 있었고 안 좋은 점도 있었습니다. 일단 프로 젝트를 진행하는 속도를 향상시킬 수 있었습니다. 각자 맡은 부분을 책임자로써 필요한 부분마다 도와주어서 지연 시간을 최소화하였습니다. 하지만 아쉬우면서도 팀원들에게 미안했던 점은 너무 제 위주로 진행되어 개인의 역량 향상을 방해하고 제 방식으로 프로그램이 작성되었다는 점이였습니다.

앞으로는 진행 전에 프로젝트를 확실하게 이해하고 역할 분담을 확실하게 하여, 최대한 간섭을 덜 하여 개인의 역량 향상을 보장하고, 모두가 만족할 수 있는 프로젝트를 만들 수 있도록 노력해가겠다는 생각이 들었으며, 첫 책임자의 귀중한 경험을 잊지 않고 앞의로의 팀 프로젝트에서 책임자이든 팀원이든 좀 더 체계적인 활동에 도움이 될 경험이었습니다.