

02 PJT

파이썬과 Pandas를 사용한 데이터 처리 (데이터 사이언스 기초)

챕터의 포인트

- [도전] 넷플릭스 주가 데이터 분석
- 제출

[도전]
넷플릭스 주가 데이터 분석

| 공통 요구사항

- 캐글을 활용하여 데이터를 다운로드 받아 활용합니다.
- 데이터셋: “[Netfilx Stock Price Prediction\(넷플릭스 주식 가격 데이터\)](#)”
- 데이터셋 요약: 2018-02-05 ~ 2022-02-04 까지의 일별 데이터
- 명시된 요구사항 이외에는 자유롭게 작성해도 무관합니다.

| 세부 요구사항

- A. 데이터 전처리 - 데이터 읽어오기
 - B. 데이터 전처리 - 2021년 이후의 종가 데이터 출력하기
 - C. 데이터 분석 - 2021년 이후 최고, 최저가 출력하기
 - D. 데이터 분석 - 2021년 이후 월 별 평균 종가 출력하기
 - E. 데이터 시각화 - 2022년 1월 이후 월 별 최고, 최저, 종가 시각화
- [참고] 출력 결과는 데이터를 다운로드 받는 시기에 따라 다르게 나올 수 있습니다.

A. 데이터 전처리 - 데이터 읽어오기

- Pandas 를 사용하여 csv 파일(NLFX.csv)을 DataFrame 으로 읽어옵니다.
- 이 때, ['Date', 'Open', 'High', 'Low', 'Close'] 필드만 읽어오도록 구성합니다.
- 출력 결과 예시

```
In [3]: # DataFrame 출력  
df
```

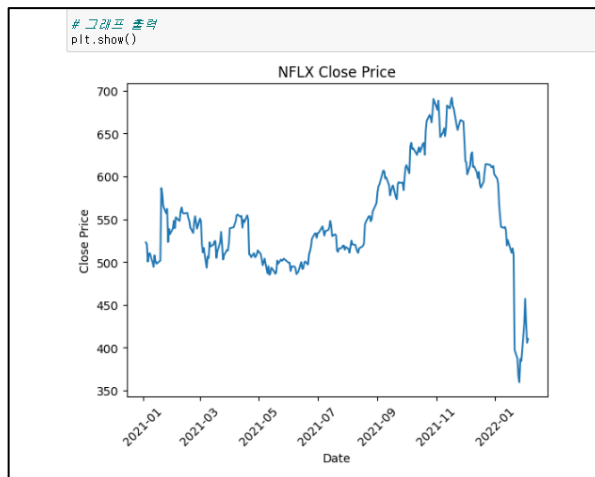
```
Out[3]:
```

	Date	Open	High	Low	Close
0	2018-02-05	262.000000	267.899994	250.029999	254.259995
1	2018-02-06	247.699997	266.700012	245.000000	265.720001
2	2018-02-07	266.579987	272.450012	264.329987	264.559998
3	2018-02-08	267.079987	267.619995	250.000000	250.100006
4	2018-02-09	253.850006	255.800003	236.110001	249.470001
...
1004	2022-01-31	401.970001	427.700012	398.200012	427.140015
1005	2022-02-01	432.959991	458.480011	425.540009	457.130005
1006	2022-02-02	448.250000	451.980011	426.480011	429.480011
1007	2022-02-03	421.440002	429.260010	404.279999	405.600006
1008	2022-02-04	407.309998	412.769989	396.640015	410.170013

1009 rows × 5 columns

B. 데이터 전처리 - 2021년 이후의 종가 데이터 출력하기

- csv 파일을 DataFrame 으로 읽어와 2021년 이후의 데이터만 필터링합니다.
 - [힌트] 필터링이 가능한 형식으로 데이터 타입을 변경한 후 필터링을 진행합니다.
 - Pandas 의 to_datetime() 을 활용합니다.
- 필터링이 완료된 DataFrame 의 종가 데이터를 Matplotlib 를 사용하여 시각화 합니다.
- 출력 결과 예시



| C. 데이터 분석 - 2021년 이후 최고, 최저 종가 출력하기

- csv 파일을 DataFrame 으로 읽어와 2021년 이후의 데이터만 필터링합니다.
- 종가(Close) 필드를 활용하여, 2021년 이후 가장 높은 종가와 가장 낮은 종가를 출력합니다.
- Pandas 의 내장 함수를 사용합니다.
- 출력 결과 예시

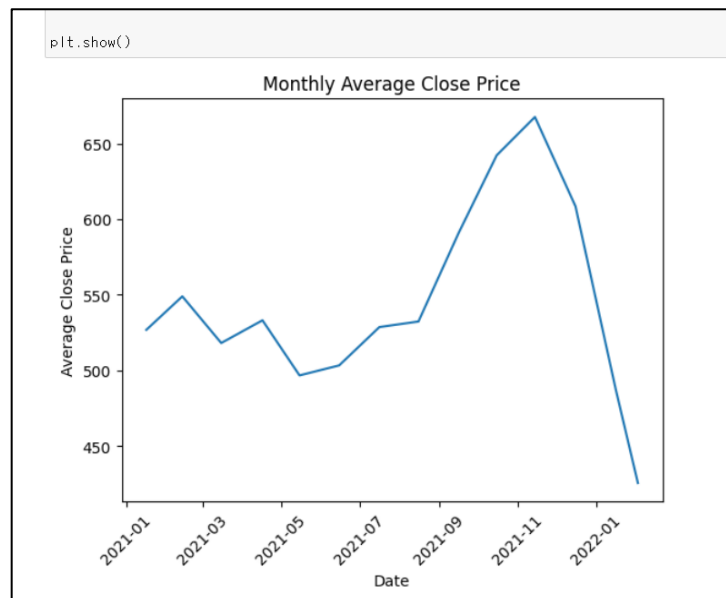
```
In [10]: print("최고 종가:", max_price)
          print("최저 종가:", min_price)
```

최고 종가: 691.690002

최저 종가: 359.700012

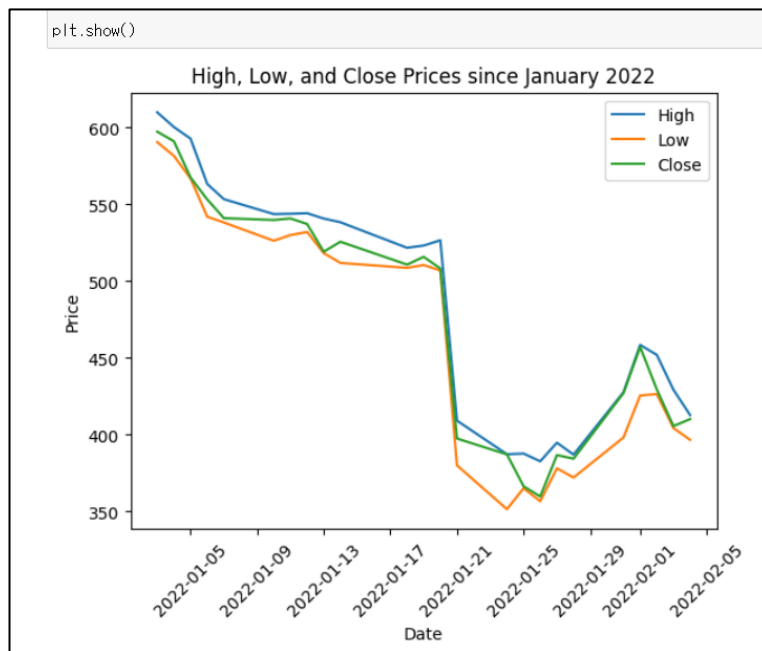
D. 데이터 분석 - 2021년 이후 월 별 평균 종가 출력하기

- csv 파일을 DataFrame 으로 읽어와 2021년 이후의 데이터만 필터링합니다.
- 월 별로 그룹화하여 평균 종가를 계산한 새로운 DataFrame 을 만들어 그래프로 시각화 합니다.
- 출력 결과 예시



E. 데이터 시각화 - 2022년 이후 최고, 최저, 종가 시각화하기

- csv 파일을 DataFrame 으로 읽어와 2022년 이후의 데이터만 필터링합니다.
- Matplotlib 를 활용하여 3가지 필드를 한 번에 분석할 수 있도록 아래와 같이 시각화 합니다.
- 출력 결과 예시



제출

| 제출 시 주의사항

- 제출기한은 금일 18시까지입니다. 제출기한을 지켜 주시기 바랍니다.
- 반드시 README.md 파일에 단계별로 구현 과정 중 학습한 내용, 어려웠던 부분, 새로 배운 것들 및 느낀 점 등을 상세히 기록하여 제출합니다.
 - 단순히 완성된 코드만을 나열하지 않습니다.
- 위에 명시된 요구사항은 최소 조건이며, 추가 개발을 자유롭게 진행할 수 있습니다.
- <https://lab.ssafy.com/> 에 프로젝트를 생성하고 제출합니다.
 - 프로젝트 이름은 '프로젝트 번호 + pjt' 로 지정합니다. (ex. 02_pjt)
- 반드시 각 반 담당 교수님을 Maintainer 로 설정해야 합니다.