예제1] part03/ex01_exploratory_analysis.py

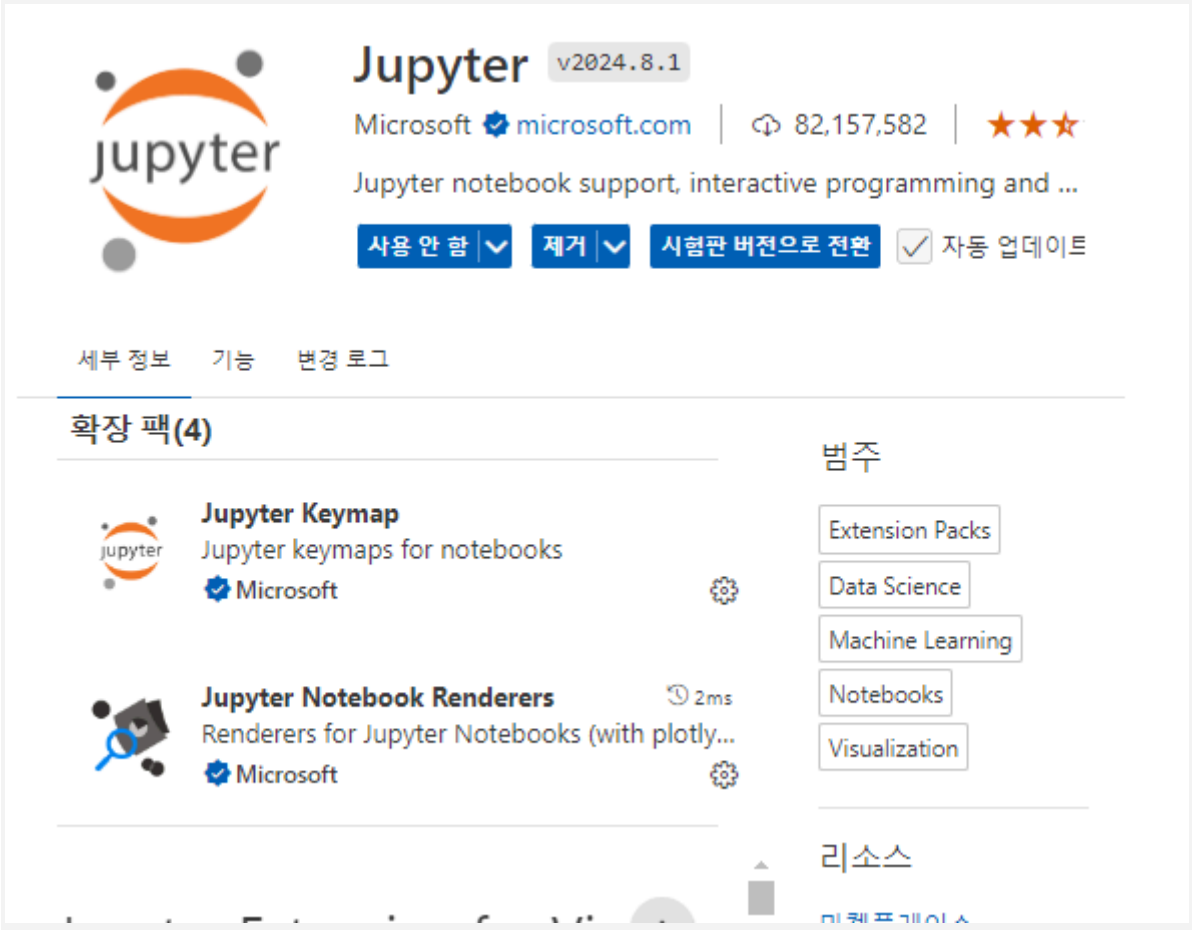
```python
# -*- coding: utf-8 -*-
import pandas as pd

df = pd.read_csv('./data/auto-mpg.csv', header=None)
df.columns = ['mpg','cylinders','displacement','horsepower','weight',
              'acceleration','model year','origin','name']

# 처음 5개의 행
print(df.head())
# 마지막 5개의 행
print(df.tail())

# df의 모양과 크기 확인
print(df.shape)
#내용 확인
print(df.info())
#자료형 확인
print(df.dtypes)
print(df.mpg.dtypes)
print(df.cylinders.dtypes)

#데이터 출력
print(df.describe())
print(df.describe(include='all'))

#유효한 원소의 갯수
print(df.count())
print(type(df.count()))
#특정 열의 고유값
unique_values = df['origin'].value_counts()
print(unique_values)

# 평균값
print(df.mean())
print(df['mpg'].mean())
print(df.mpg.mean())
print(df[['mpg','weight']].mean())

# 중간값
print(df.median())
print(df['mpg'].median())
print(df['origin'].median())

# 최대값
print(df.max())
print(df['mpg'].max())
print(df['horsepower'].max())

```

```python
# 최소값
print(df.min())
print(df['mpg'].min())

# 표준편차
print(df.std())
print(df['mpg'].std())

# 상관계수
print(df.corr())
print(df[['mpg','weight']].corr())

```

예제2] part03/ex02_df_plot.py

pip install openpyxl

pip install matplotlib



설치할 것

```python
# -*- coding: utf-8 -*-
import pandas as pd

df = pd.read_excel('./data/남북한발전전력량.xlsx', engine='openpyxl')


df_ns = df.iloc[[0, 5], 3:]
df_ns.index = ['South','North']
df_ns.columns = df_ns.columns.map(int)
print(df_ns.head())

#선 그래프1
df_ns.plot()

#선 그래프2
tdf_ns = df_ns.T
print(tdf_ns.head())
tdf_ns.plot()

#막대 그래프
df_ns.plot(kind='bar')
tdf_ns.plot(kind='bar')

```

예제3] part03/ex03_df_plot_scatter.py

```python
# -*- coding: utf-8 -*-

import pandas as pd

df = pd.read_csv('./data/auto-mpg.csv', header=None)

df.columns = ['mpg','cylinders','displacement','horsepower','weight',
              'acceleration','model year','origin','name']

df.plot(x='weight',y='mpg', kind='scatter')

df[['mpg','cylinders']].plot(kind='box')

```