

Python Django

데이터 시각화(Data Visualization)를 위한 Python & Django

Python

<Python Library사용>

[pandas, re(정규 표현식(Regular Expression))활용, 웹 크롤링하기]

pandas, re, bs4(BeautifulSoup) & request 활용

import pandas, import re, from bs4 import BeautifulSoup, import requests 사용
패키지 설치 안 되었을때는 pip install 사용하여 패키지 설치해줘야 합니다.

[웹 크롤링(정적 크롤링, 동적 크롤링)]

<정적 크롤링 vs 동적 크롤링>

정적 크롤링과 동적 크롤링은 웹 크롤링의 종류로, 수집하고자 하는 데이터의 특성에 따라 구분됩니다.

정적 크롤링은 웹 페이지를 처음 방문했을 때 보이는 데이터를 수집하는 방법입니다. 로그인이나 클릭과 같은 추가 조작이 필요하지 않기 때문에 속도가 빠르다는 장점이 있습니다. 예를 들어, 상품 목록이나 뉴스 기사와 같은 데이터를 수집하는 데 적합합니다.

동적 크롤링은 웹 페이지를 방문한 후 추가 조작을 통해 보이는 데이터를 수집하는 방법입니다. 로그인, 클릭, 스크롤, 페이지 이동 등과 같은 조작을 통해 데이터를 노출시켜야 하기 때문에 정적 크롤링에 비해 속도가 느립니다. 예를 들어, 실시간 주식 시세나 날씨 정보와 같은 데이터를 수집하는 데 적합합니다.

정적 크롤링 VS 동적 크롤링

	정적 크롤링	동적 크롤링
연속성	주소를 통해 단발적으로 접근	브라우저를 사용하여 연속적으로 접근
수집 능력	수집 데이터의 한계가 존재	수집 데이터의 한계가 없음
속도	빠름	느림
라이브러리	requests, BeautifulSoup	selenium, chromedriver

[bs4(BeautifulSoup) & request 활용, pandas로 파일 읽기 및 내보내기, matplotlib 활용하기]
웹 크롤링 개념 복습, 영화 평점 가져오기, 날씨 알아내기 및 matplotlib 활용하여 차트 만들기

-> Stable 버전 다운로드 들어가서 Chrome Driver를 설치

-> MySQL을 Python에서 사용에 필요한 라이브러리 패키지

bigdb

Schema

생성,

daum_movie,

forecast

Table

생성

Table Name: daum_movie

Schema: bigdb

CharSet/Collation: Default CharSetDefault Collation

Engine: InnoDB

Comments:

Column Name	Datatype	PK	NN	UQ	B	UN	ZF	AI	G	Default/Expression
num	INT	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
title	VARCHAR(100)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
grade	VARCHAR(45)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
reserve	VARCHAR(45)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL

Column Name:

CharSet/Collation:

Comments:

Data Type:

Default:

Storage:

☐ Virtual☐ Stored

☐ Primary Key☐ Not Null☐ Unique

☐ Binary☐ Unsigned☐ Zero Fill

☐ Auto Increment☐ Generated

Columns

Indexes

Foreign Keys

Triggers

Partitioning

Options

bigdb - Schema

daum_movie - Table

daum_movie

daum_movie

daum_movie - Table

bigdb - Schema

Name: bigdb

Rename References

CharSet/Collation: utf8utf8_bin

Specify the name of the schema here. You can use any combination of ANSI letters, numbers and the underscore character for names that don't require quoting. For more flexibility you can use the entire Unicode Basic Multilingual Plane (BMP), but you will have to quote the name later with backticks.

Refactor model, changing all references found in view, triggers, stored procedures and functions from the old schema name to the new one.

The character set and its collation selected here will be used when no other charset/collation is set for a database object (it uses the DEFAULT value then). Setting DEFAULT here will make the schema to use the server defaults.

Schema

Apply

Revert

Table Name: Schema: **bigdb**

Charset/Collation: Engine:

Comments:

Column Name	Datatype	PK	NN	UQ	B	UN	ZF	AI	G	Default/Expression
num	INT	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
city	VARCHAR(45)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
tmef	VARCHAR(45)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
wf	VARCHAR(45)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
tmin	VARCHAR(45)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
tmx	VARCHAR(45)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL

Column Name: Data Type:

Charset/Collation: Default:

Comments:

Storage: ☐ Virtual ☐ Stored

☐ Primary Key ☐ Not Null ☐ Unique

☐ Binary ☐ Unsigned ☐ Zero Fill

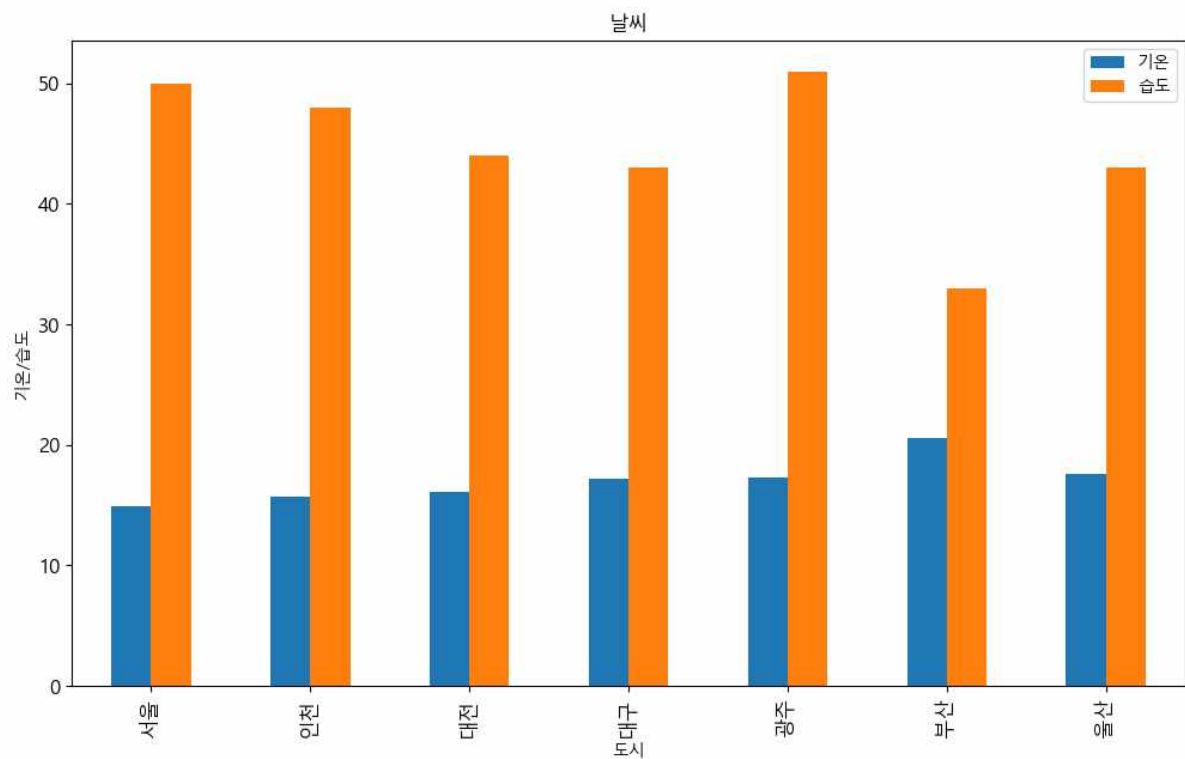
☐ Auto Increment ☐ Generated

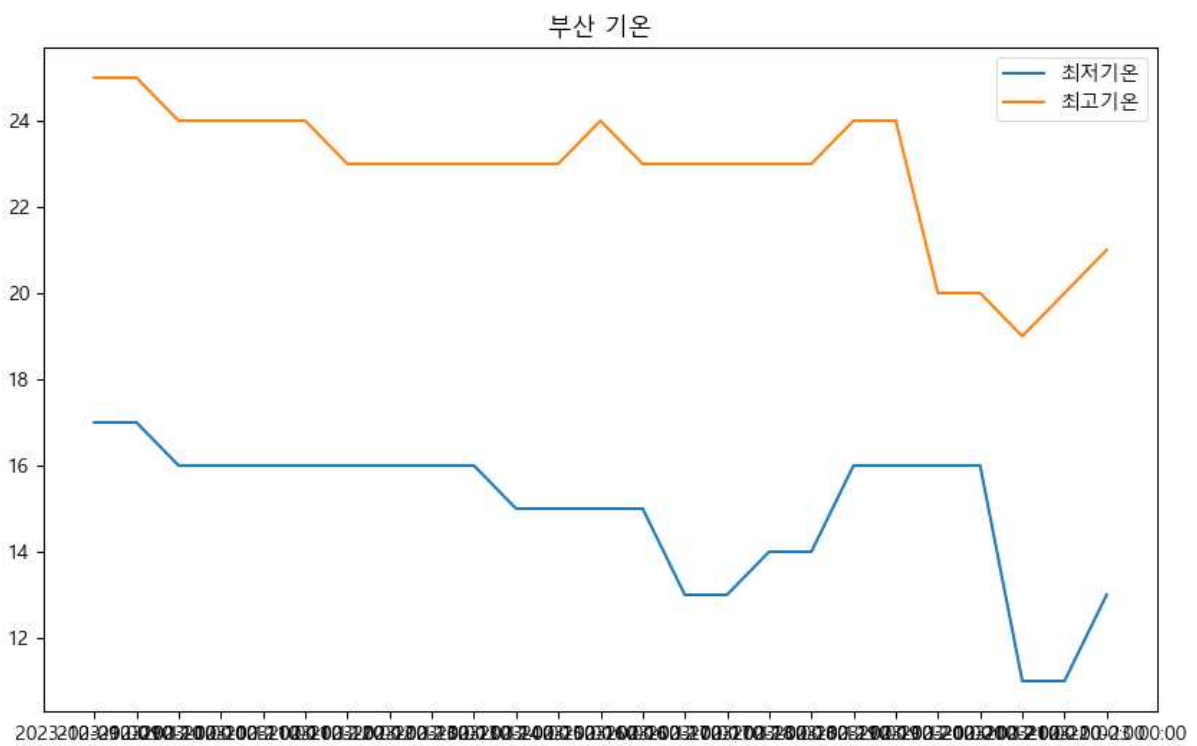
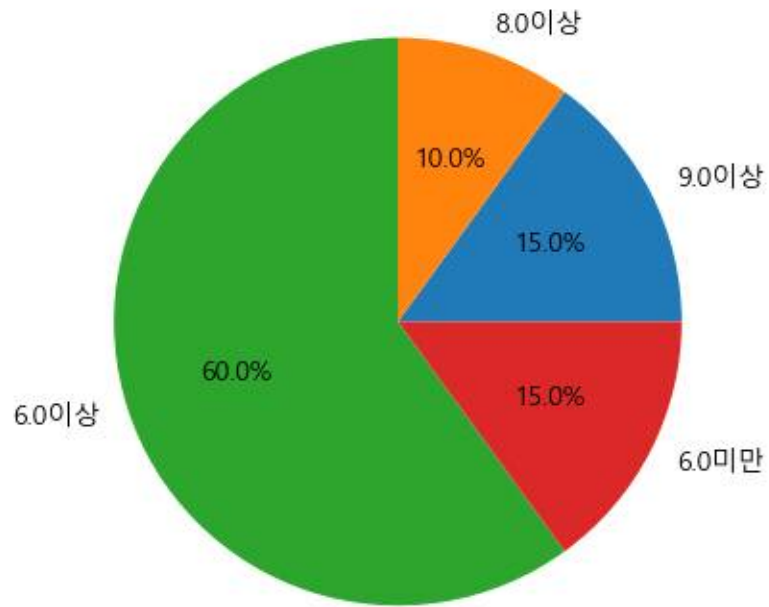
Columns | Indexes | Foreign Keys | Triggers | Partitioning | Options

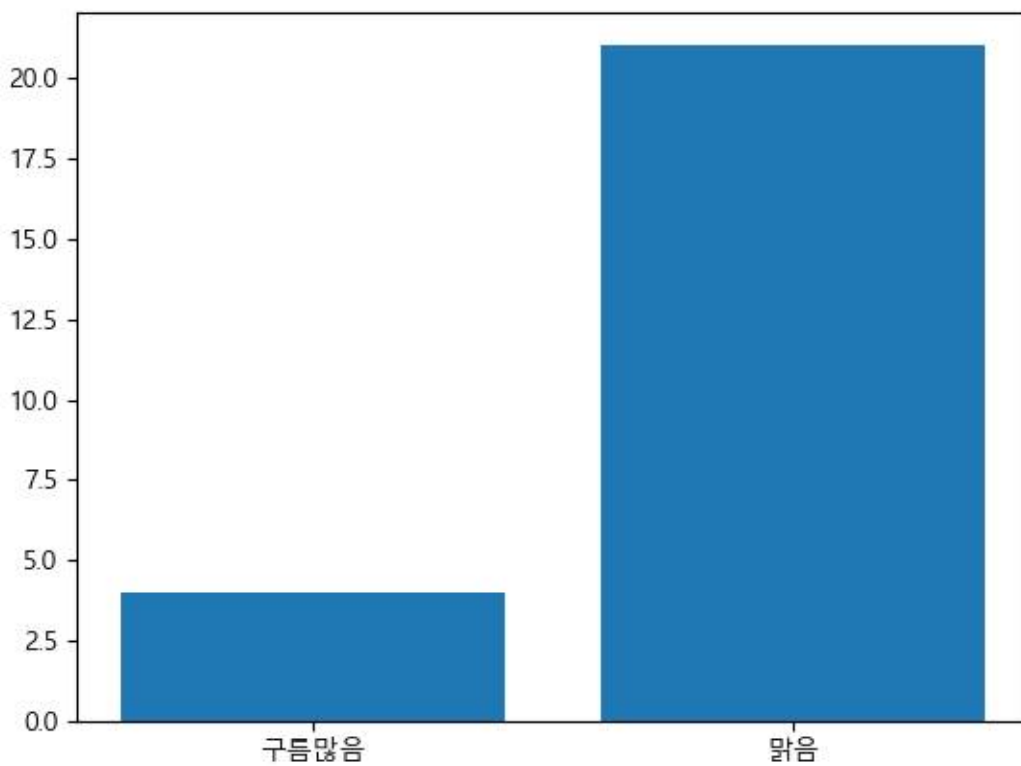
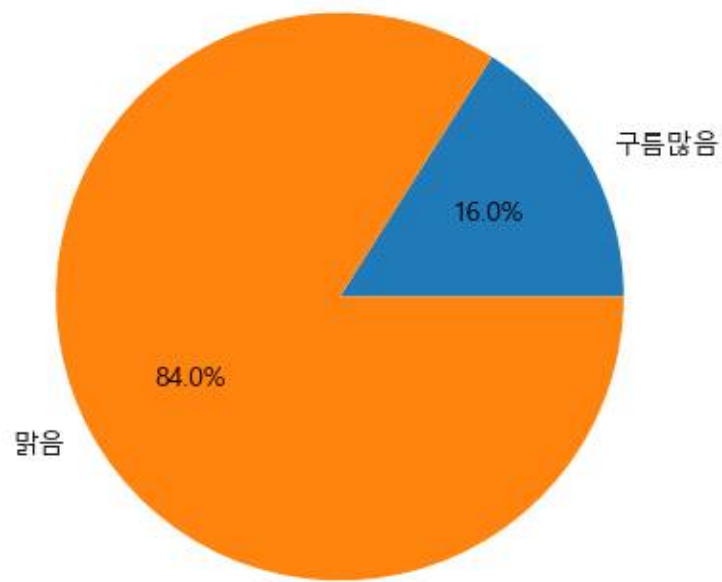
Apply Revert

DB 연결 및 DB 삽입, 선택 작업

PieChart 그리기(matplotlib 활용), 다른 여러 가지 그래프들 (Matplotlib 활용)



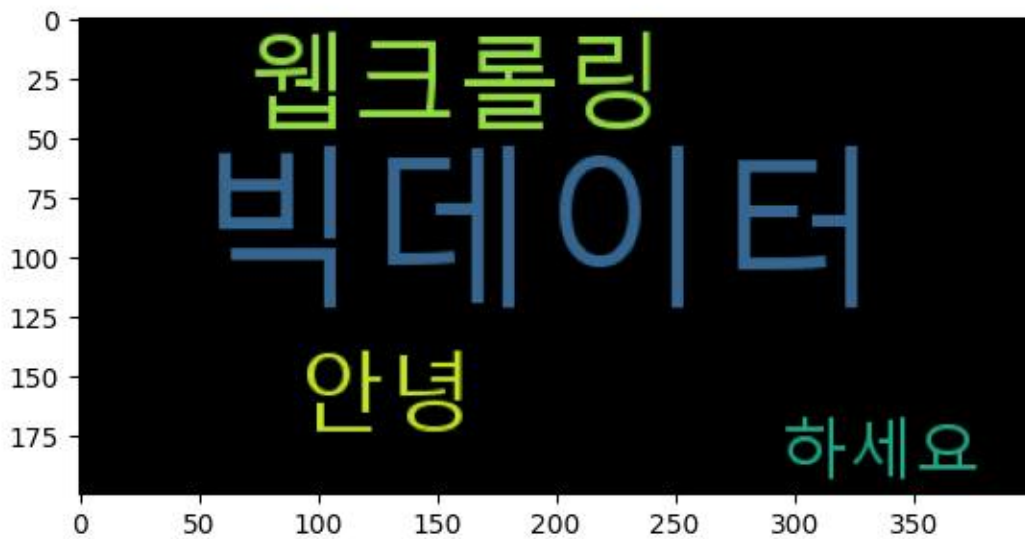




`pip install lxml` (xml 파싱 라이브러리)

[웹 크롤링 DB 연동하기]

- 날씨 정보 추출, 날짜별 최고기온, 최저기온 출력
- WordCloud 활용



WordCloud 활용을 하게 되면 위와 같은 이미지의 형태가 나옵니다.
