

기분과 날씨의 상관관계

- 1 프로젝트 주제 선정 동기 및 분석 목적
- 2 데이터셋 소개
- 3 데이터 수집 및 전처리 과정 소개
- 4 데이터 시각화 및 상관관계 분석
- 5 주요 질문 및 결론

HOME > PROJECT > TOP

날씨와 마음의 상관관계

👤 누다심 심리학칼럼니스트 | ⌚ 승인 2016.05.09 16:07 |

[반기성의 날씨와 경제] 美 기상 전문기업들의 날씨 마케팅

경제 | 입력 2021-02-03 21:05:40 | enews2 기자 | 💬 0개



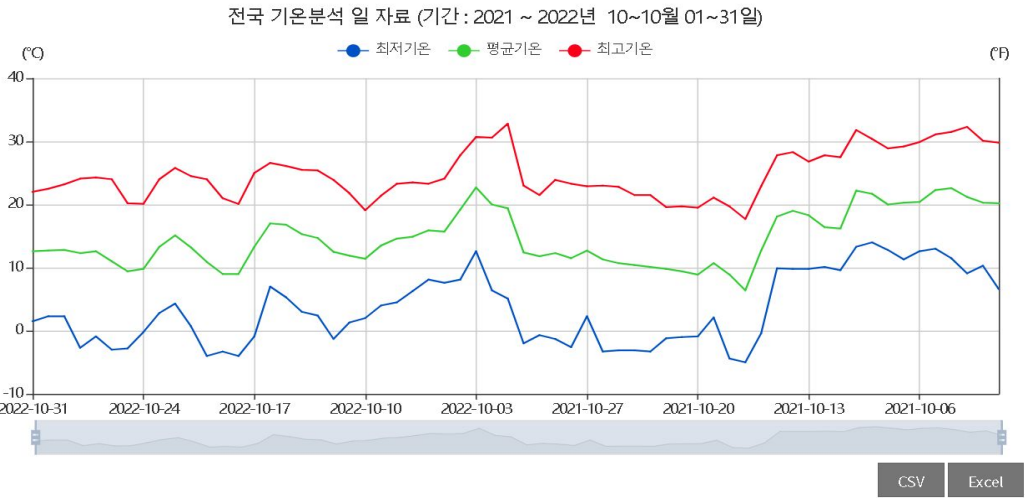
분석 동기와 분석 목적

1. 분석 동기

SNS보급에 따른 인간의 감성을 이용한 감성마케팅 사례
증가
대중의 감성과 다양한 감성 욕구를 파악해 감성 활동
예측

2. 분석 목적

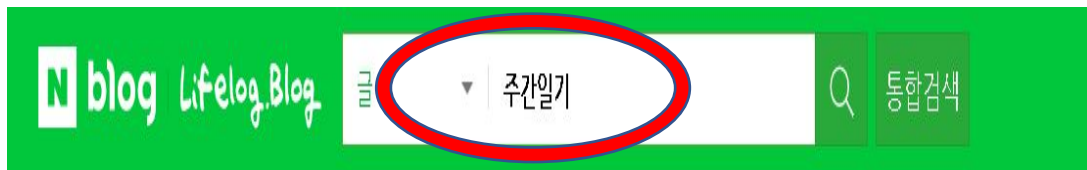
특별한 일이나 기념일 등에 최적화된 일자를 선별한 후
이벤트와 같은 여러 서비스에 활용
감성 욕구를 분석해 일자 별 사업 계획



지역	지역명	일시	평균습도(%rh)	최저습도(%rh)	최저습도지점
0	전국	2022-10-31	74	27	함양군
0	전국	2022-10-30	76	25	함양군
0	전국	2022-10-29	77	22	함양군
0	전국	2022-10-28	77	16	함양군
0	전국	2022-10-27	70	16	함양군
0	전국	2022-10-26	68	21	함양군
0	전국	2022-10-25	66	19	함양군
0	전국	2022-10-24	62	16	함양군
0	전국	2022-10-23	71	21	동두천
0	전국	2022-10-22	69	9	태백
0	전국	2022-10-21	69	11	태백
0	전국	2022-10-20	66	11	태백
0	전국	2022-10-19	63	16	밀양

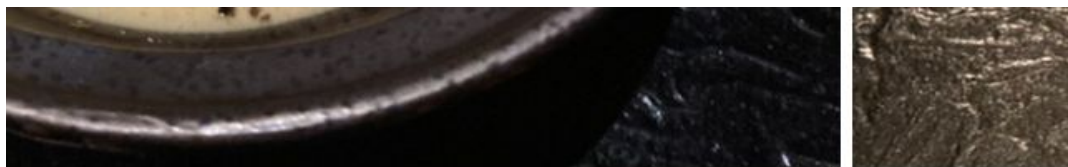
기상자료개방포털 ‘기후통계분석’

1. 전국에서 평균기온과 평균습도로 정의된 일간 기상 데이터
2. 기간 : 2021.10 ~ 2022..10 기간의 데이터 수집
3. 데이터 구성 요소 : 날짜, 평균기온, 평균습도



네이버 블로그 '주간일기'

1. 네이버 블로그 '주간일기'를 활용한 감정 분석
2. 검색어인 '주간일기에 대한 응답값을 기반으로 정의된 월별 검색량 데이터
3. 2021.10 ~ 2022.10 기간의 데이터 수집 (월별 당 1,000 개의 데이터 수집)
4. 데이터 구성 요소: 제목(title), 이름(nickname), 날짜(datetime), 내용(content)



이때까지 다녀온 도쿄등심과는 비교가 안 되는
청담 도쿄등심 시그니처 코스.....!

초대받아서 다녀왔다 ㅎㅎ
장어가 나와서 **너무 행복했던 >**



네이버 블로그 <주간일기> 데이터 크롤링



```
pwd
```

```
'C:\\\\Users\\USER'
```

```
import sys
import os

import pandas as pd
import numpy as np

from bs4 import BeautifulSoup
from selenium import webdriver
import time
from tqdm import tqdm_notebook
```

```
from selenium.webdriver.common.by import By
from selenium import webdriver
from selenium.webdriver.chrome.service import Service
from webdriver_manager.chrome import ChromeDriverManager
```

```
!python --version
```

```
Python 3.9.13
```

```
pd.__version__
```

```
'1.4.4'
```

```
keyword = input("1. 크롤링할 키워드를 입력하세요: ")
```

```
start_date = input("시작 날짜(예:20221123): ")
```

```
end_date = input("종료 날짜(예:20221203): ")
```

```
driver = webdriver.Chrome(r"C:\\chromedriver.exe")
```

```
driver.get("https://section.blog.naver.com/")
time.sleep(2)
```

```
element = driver.find_element('name', 'sectionBlogQuery')
element.send_keys(keyword)
driver.find_element("xpath", '//*[@id="header"]/div[1]/div/div[2]/form/fieldset/a[1]/i').click()
time.sleep(1)
```

네이버 블로그 <주간일기> 데이터 크롤링



```
url_list = []
title_list = []

total_page = 183
for i in tqdm_notebook(range(169, total_page)):
    i = i + 1
    url = 'https://section.blog.naver.com/Search/Post.naver?pageNo='+ str(i+169) + '&rangeType=PERIOD&orderBy=sim&startDate=2022-04-01&endDate=2022-04-30&key'
    driver.get(url)

    titles = "a.desc_inner"
    article_raw = driver.find_elements('css selector', titles)

    for article in article_raw:
        url = article.get_attribute('href')
        url_list.append(url)

    for article in article_raw:
        title = article.text
        title_list.append(title)

    print(title)

print('url갯수: ', len(url_list))
print('url갯수: ', len(title_list))

df = pd.DataFrame({'url':url_list, 'title':title_list})

df.to_excel("blog_url2204(2).xlsx", encoding='utf-8-sig')
```

url_list



네이버 블로그 <주간일기> 데이터 크롤링



```
import sys
import os
import pandas as pd
import numpy as np
```

```
url_load = pd.read_excel("blog_url2204(2).xlsx")
```

```
num_list = len(url_load)
url_load
```

```
dict = {} # 전체 크롤링 데이터를 담을 그릇
```

```
number = num_list # 수집할 글 갯수
```

```
# 수집한 url 돌면서 데이터 수집
```

```
for i in tqdm_notebook(range(0, number)):
```

```
    # 글 띄우기
```

```
    url = url_load["url"][i]
```

```
    driver = webdriver.Chrome(r"C:\chromedriver.exe")
```

```
    driver.get(url) # 글 띄우기
```

```
    # 크롤링
```

```
    try :
```

```
        # 글의 iframe 접근
```

```
        driver.switch_to.frame("mainFrame")
```

```
        target_info = {} # 개별 블로그 내용을 담을 딕셔너리 생성
```

```
        # 제목 크롤링
```

```
        overlays = ".se-module.se-module-text.se-title-text"
```

```
        tit = driver.find_element(By.CSS_SELECTOR, overlays)
```

```
        title = tit.text
```

```
        # 글쓴이 크롤링
```

```
        overlays = ".nick"
```

```
        nick = driver.find_element(By.CSS_SELECTOR, overlays)
```

```
        nickname = nick.text
```

```
    # 날짜 크롤링
```

```
    overlays = ".se_publishDate.pco12"
```

```
    date = driver.find_element(By.CSS_SELECTOR, overlays)
```

```
    datetime = date.text
```

```
    # 내용 크롤링
```

```
    overlays = ".se-component.se-text.se-l-default"
```

```
    contents = driver.find_elements(By.CSS_SELECTOR, overlays)
```

```
    content_list = []
```

```
    for content in contents:
```

```
        content_list.append(content.text)
```

```
    content_str = ' '.join(content_list)
```

```
    # 크롤링한 글은 target_info라는 딕셔너리에 담음
```

```
    target_info['title'] = title
```

```
    target_info['nickname'] = nickname
```

```
    target_info['datetime'] = datetime
```

```
    target_info['content'] = content_str
```

```
    # 각각의 글은 dict라는 딕셔너리에 담음
```

```
    dict[i] = target_info
```

```
    time.sleep(0.0000001)
```

```
    # 크롤링 성공하면 글 제목을 출력
```

```
    print(i, title)
```

```
    # 글 하나 크롤링 후 크롬 창 닫기
```

```
    driver.close()
```

```
    # 예러나면 현재 크롬창을 닫고 다음 글(i+1)로 이동
```

```
    except :
```

```
        driver.close()
```

```
        time.sleep(0.00001)
```

```
        continue
```

```
    # 중간 저장 (없어도됨)
```

```
    if i == 30 or 50 or 80:
```

```
        # 판다스로 만들기
```

```
        import pandas as pd
```

```
        result_df = pd.DataFrame.from_dict(dict, 'index')
```

```
        # 저장하기
```

```
        result_df.to_excel("naver_blog_content2204(2).xlsx", encoding='utf-8-sig') # 한글은 오류날 수 있음
```

```
        time.sleep(0.00001)
```

```
    print('수집한 글 갯수: ', len(dict))
```

```
    print(dict)
```

```
    result_df.to_excel("naver_blog_content2204(2).xlsx", encoding='utf-8-sig')
```



네이버 블로그 <주간일기> 데이터 크롤링



	title	nickname	datetime	content
0	2021. 11월	울스	2021. 11.	주말이 지나고 찾아오는 월요일이안식의 날처럼 느껴져 참 좋다.가족들을 모두 출근 혹은 등교시키고 나면시험 친 다음날의 후련함과마음의 여
1	#밥상일기	하지	2021. 11.	2021. 10 마지막 주신혼주간밥상일기 그동안 정말 잘 먹고 살아와서이번주는 오빠랑 독소빼는 주간!!!으로 정하고큰 맘 먹고 오랜만에 #사과디톡
2	[주간 일기	위시	2021. 11.	※ 긴 글 주의. 오글 주의. TMI 주의.월요일.모두가 지치고 힘든 월요일은 양심상 조용히 찌그러져 있었다. 화요일.못 참고 나대기. 가볍게 언니에.
3	주간밥상/	옥이	2021. 11.	올해는 시간이 빨리지나간다..♡월요일콩이 심장초음파추적검사 하는날. 수면마취를 해야해서 혼자 가는데 긴장했는데 역시나 대병은 혼자 힘들
4	주간일기 :	엘진자	2021. 11.	역시나 읽어주는 사람이 있으면 신나서 쓰게되는 일기(나 혼자 보려고 쓴다면서 이런다 ☺)그럼 10월 마지막주 일기 시작! 그렇다 10월 25일 미
5	10월 주간	민지짱	2021. 11.	두둥2주전 주간일기출장때메 아예 블로그 들어와보지도 못했더니블로그 쓰고 심은 병에 걸렸지 모야.....바야흐로..2주전아직도 10월10/18(월) 급
6	11월 주간	차샤	2021. 11.	11월의 첫 주간일기 시작은 10월의 가장 큰 소비였던 아이폰13미니! 11번가 4차 사전예약으로 구매! 11/1 도착 한다는 예정과 달리 10/26 도착하
7	한동안 이	RIM	2021. 11.	지난주 월요일엔 어떤 하루를 보냈길래한 장의 사진도 안 남은 것인지,시작은 화요일, 골골대는 몸으로헤정 언니를 만나기 위해 성수동으로 갔다
8	10월 첫째	로디	2021. 11.	10월 1일 새로운 집으로 이사 와서또 열심히 시작해 보는 주부송의 집밥 요리일기 개천절 대체공휴일에 먹었던 점심 같다! 트레이더스에서 사
9	10월 4주차	MDina 엠	2021. 11.	8시에 필라테스 수업 갈 수 있는 마지막 주차....왜냐하면 11월부터 워드코로나로출근시간이 9시로 바뀌었기 때문이다... 망할 ㅈㅈ 매머드 커피
10	[안스일상:	안스	2021. 11.	가장 아쉬웠던 점 1: 주방 before 사진을 못찍음 ^^...입주 첫날 아버지가 오셔서 벅벅벅 닦아주신덕분에정말... 깨끗하게 사용중! 올라빠처럼 딸(
11	2021년 11	kelly	2021. 11.	가을인지 겨울인지 모를 뒤죽박죽 날씨가다시 정신 차리나 심더니 갑자기 분위기 미세먼지.....또 왔구나 미세먼지 너,,,?올해 11월 틈틈이 끼워 놓
12	2021. 11월	울스	2021. 11.	일주일에 한 번 컴퓨터를 마주하는소동한 시간.ㅋㅋ주간 일기를 기록하는 지금!소름 돋을 정도로 천천히,증멜 매주 손톱만큼 나아지고 있는 듯한 1
13	주간일기 :	개굴킴	2021. 11.	주간일기를 그 다음주 목요일이 돼서야써오는 사람이 있다?그 사람이 바로 저예요☺...지난주부터 묘한 우울감 피로감이 겹치더니이번주는 몸
14	#밥상일기	하지	2021. 11.	2021년 11월 첫째 주신혼주간밥상일기 12월이 한 달도 채 남지 않았다니항상 11월쯤이 되면 시간이, 세월이 무색함을 느낀다.어쩔 이렇게 빠르
15	[11월 셋째	rawbeing	2021. 11.	"곤경에 빠지는 건 뭔가를 몰라서가 아니다. 뭔가를 확실히 안다는 착각 때문이다"마크 트웨인저도 가끔 뜨끔합니다...알고 있는 것과 실천은 다른

datetime	feeling	평균기온(°C)	평균습도(%rh)
2021-10-01	1	20.2	82
2021-10-01	0	20.2	82
2021-10-01	1	20.2	82
2021-10-01	0	20.2	82
2021-10-01	0	20.2	82
2021-10-01	1	20.2	82
2021-10-01	1	20.2	82
2021-10-01	1	20.2	82
2021-10-01	0	20.2	82
2021-10-01	1	20.2	82
2021-10-02	0	20.3	80
2021-10-02	1	20.3	80
2021-10-02	1	20.3	80
2021-10-02	1	20.3	80
2021-10-02	1	20.3	80
2021-10-02	1	20.3	80
2021-10-02	1	20.3	80

데이터 전처리

방향

1. 엑셀파일을 활용해 감성분석 분류 결과로 나뉜 월별 표
2. 감성분석단어 분류 (크게 2가지의 분류)
 - a. 긍정 (행복하다, 기쁘다, 상쾌하다, 즐겁다)
 - b. 부정(슬프다, 우울하다, 서운하다, 화나다)
3. 데이터 구성 요소: 날짜(datetime), 감정(feeling), 평균기온, 평균습도
4. Matplotlib과 Scattare Plot 사용

데이터 시각화



```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
result2 = pd.read_excel('2110_2204 시각화.xlsx')
result2.dtypes
```

```
datetime          datetime64[ns]
feeling            float64
temperature        float64
humidity           float64
dtype: object
```

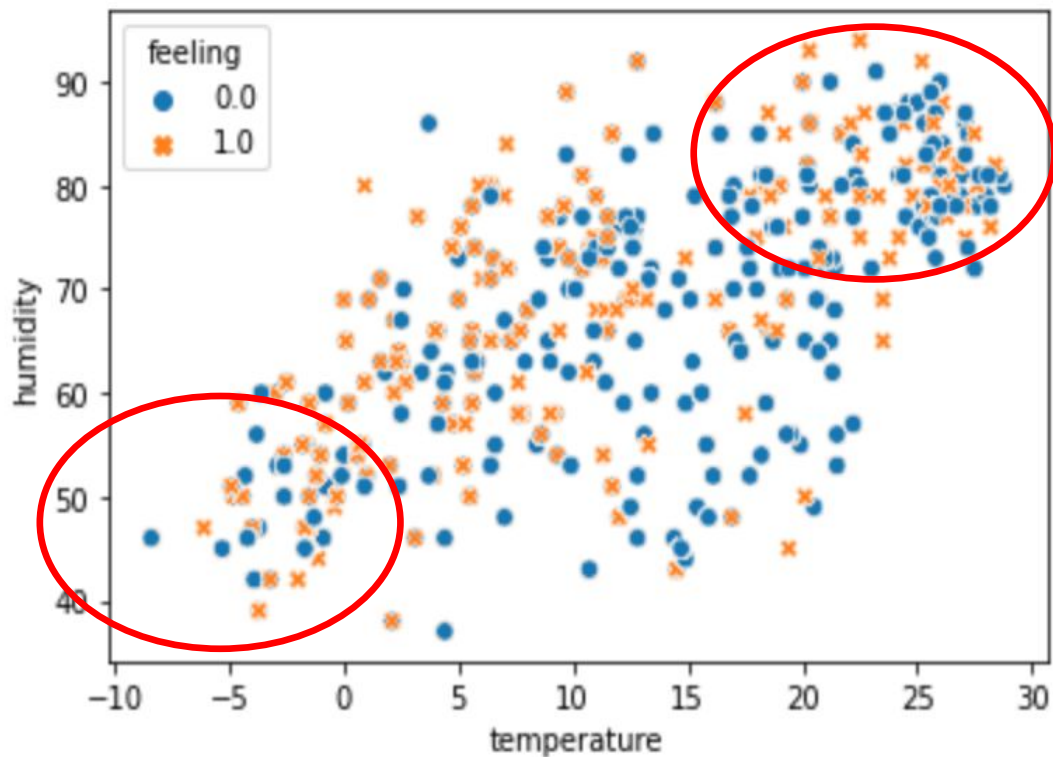
```
a = pd.read_excel('202205_202210_전처리.xlsx')
```

```
final = pd.concat([result2,a])
final.to_excel('최종결과.xlsx')
final
```

```
sns.scatterplot(x='temperature',
                y='humidity',
                hue='feeling', # different colors by group
                style='feeling', # different shapes by group
                s=40, # marker size
                data=final)

plt.show()
```

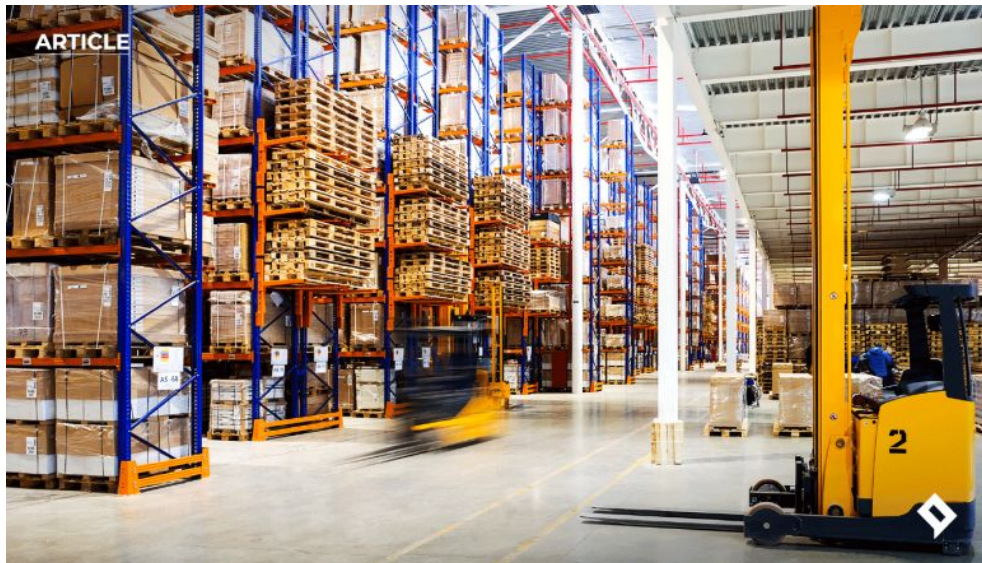




데이터 상관관계

분석

1. 전체 기간의 최종적인 그래프 시각화 (x축: 기온, y축: 습도)
2. 기온과 습도가 낮을 때보다 높을 때 감정 단어의 사용 빈도가 높음
3. 기온이 높고 습도가 낮을 시 부정 단어의 사용 빈도가 더 높음 (ex. 여름)
4. 기온이 낮고 습도가 높을 시 긍정 단어의 사용 빈도가 더 높음 (ex. 겨울)
5. 상대적으로 습도보다 기온의 영향을 더 많이 받음



주요

질문

1. 날씨별 나타나는 평균적인 감정상태는 무엇인가
2. 기분과 날씨의 상관관계를 활용한 효과적인 마케팅 방안은 무엇인가
 - 마케팅고객이 느끼는 감정에 따라 제품을 살지가 결정되는데 가장 큰 영향
 - 현재의 날씨 조건으로 날씨에 따라 고객들이 선호하거나 선호하지 않는 제품을 매일 분석해 판매자에게 알려 재고 관리 및 상품의 배치 등을 대비
 - 음료 마케팅 , 명품 마케팅

감사합니다 ☀