

# 공유 자전거 대여 예측

**Capital Bikeshare Program in Washington, D.C.**

3조  
임수연, 장현정, 조은서

# 목차 Contents

**01** 주제 선정 배경  
& 분석 방향

**02** EDA & 전처리  
데이터 설명 / EDA & 전처리

**03** 모델링 적용

**04** 결론  
활용방안 / 한계

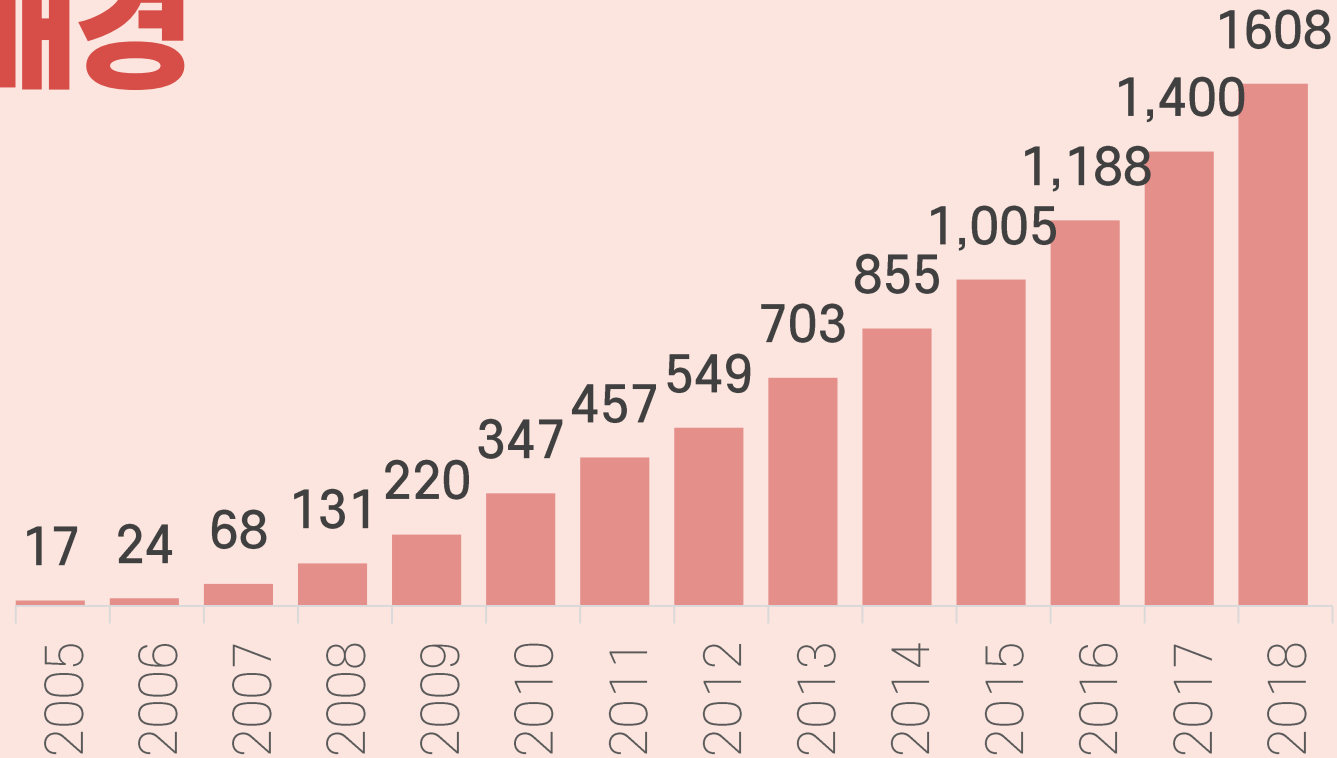
# 01

## 주제 선정 배경 & 분석 방향

# 주제 선정 배경



전 세계 자전거 공유 서비스



출처: Statista

## 공유 자전거에 대한 수요 증가

전 세계적으로 많은 대도시들이 자전거 공유 프로그램을 운영하여 도심 도로 차량 수를 줄여 교통혼잡과 오염을 억제하려고 노력하고 있다. 이에 주목하여 공유 자전거 서비스에 대한 분석을 주제로 하였다.

“

자전거 대여 횟수를

예상해보자

”

# 02

## EDA & 전처리

- 1) 데이터 설명
- 2) EDA & 전처리

# 1) 데이터 설명

## 01 데이터 범위

워싱턴 D.C.의

‘Capital Bikeshare program’

2011, 2012년 자전거 대여 횟수를  
매 시간 기록한 데이터

## 02 레이블 설명

Train data:

1일 ~ 19일 (10886, 12)

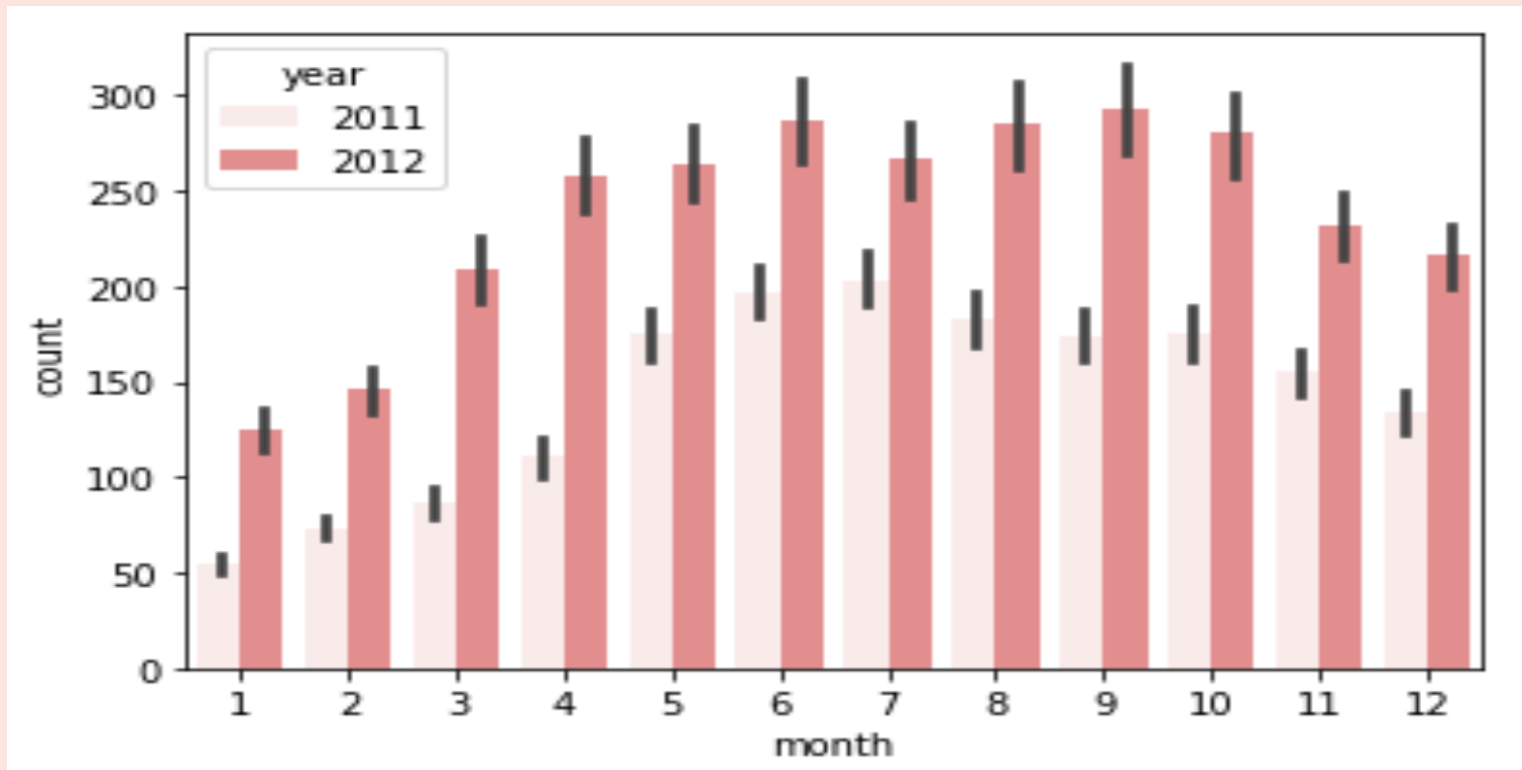
Test data:

20일 ~ 말일 (6493, 9)

## 03 변수 설명

Datetime	Weather	Windspeed
Season	Temp	Casual
Holiday	Atemp	Registered
Workingday	Humidity	Count

## 2) EDA&전처리 -연·월에 따른 대여 추이



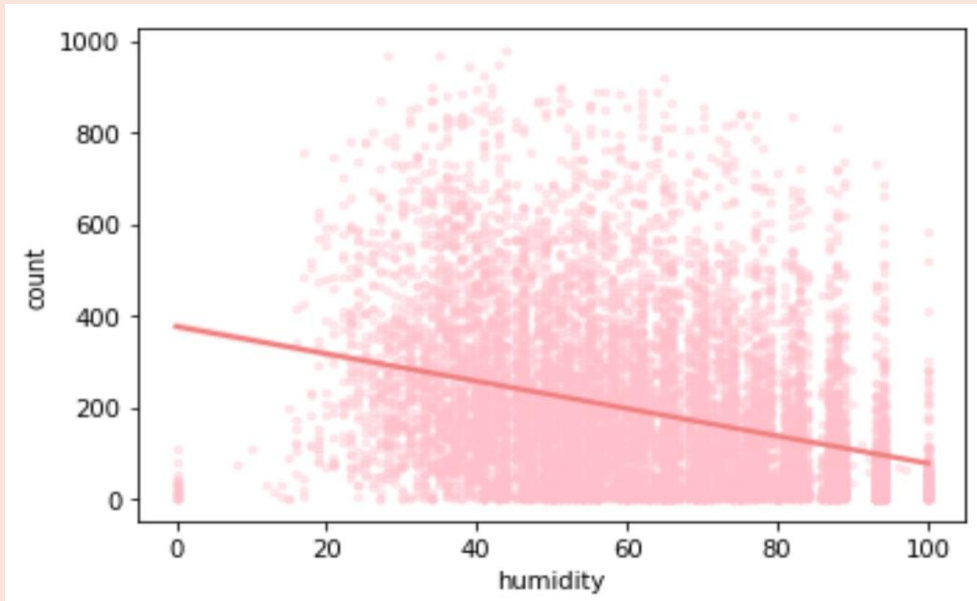
- 2012년의 대여량이 더 많음
- 겨울보다 여름의 대여량이 더 많음

...

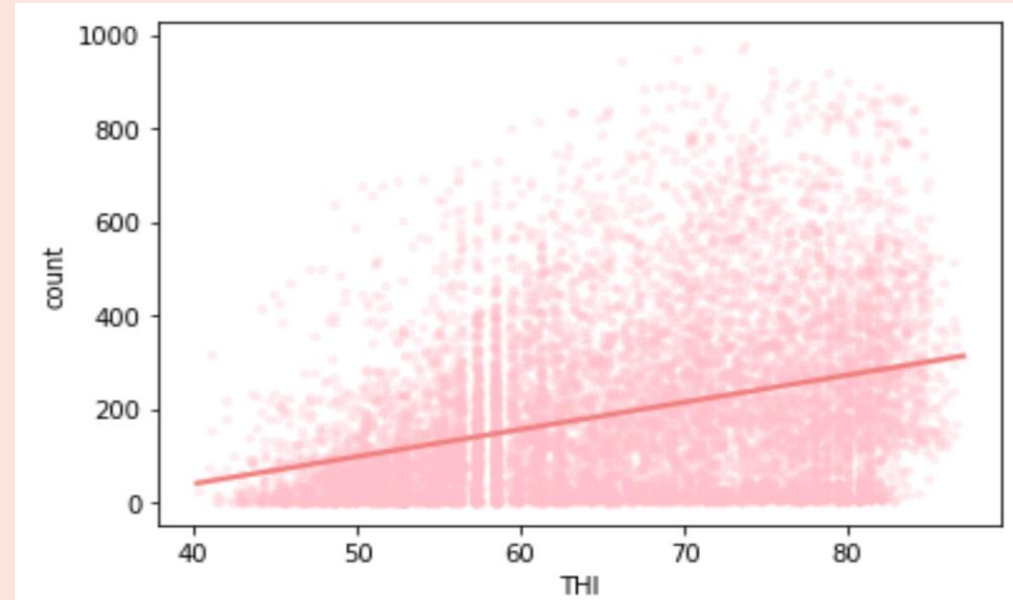


## 2) EDA&전처리

### 습도에 따른 대여 차이



### 온도에 따른 대여 차이

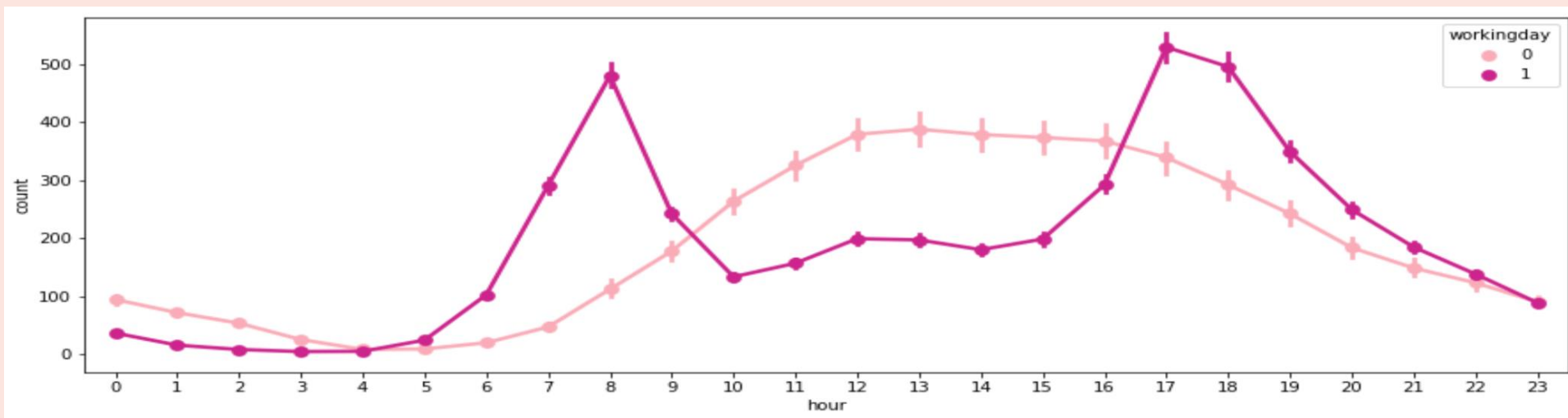


...

- 습도와 대여량이 음의 상관관계를 가짐
- 온도와 대여량이 양의 상관관계를 가짐

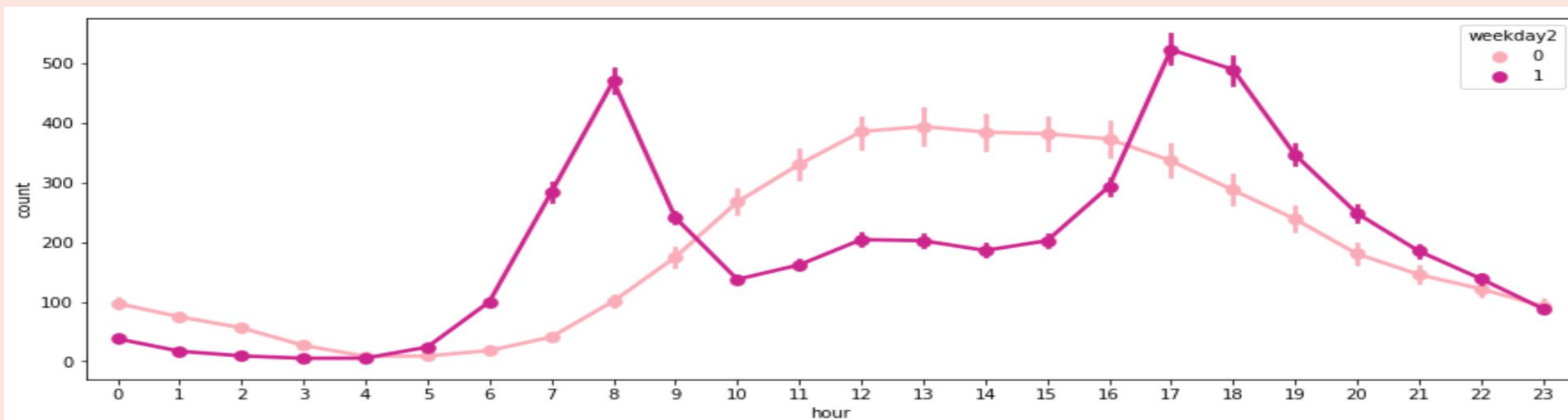
## 2) EDA&전처리

### 근무일 여부에 따른 시간별 대여 차이

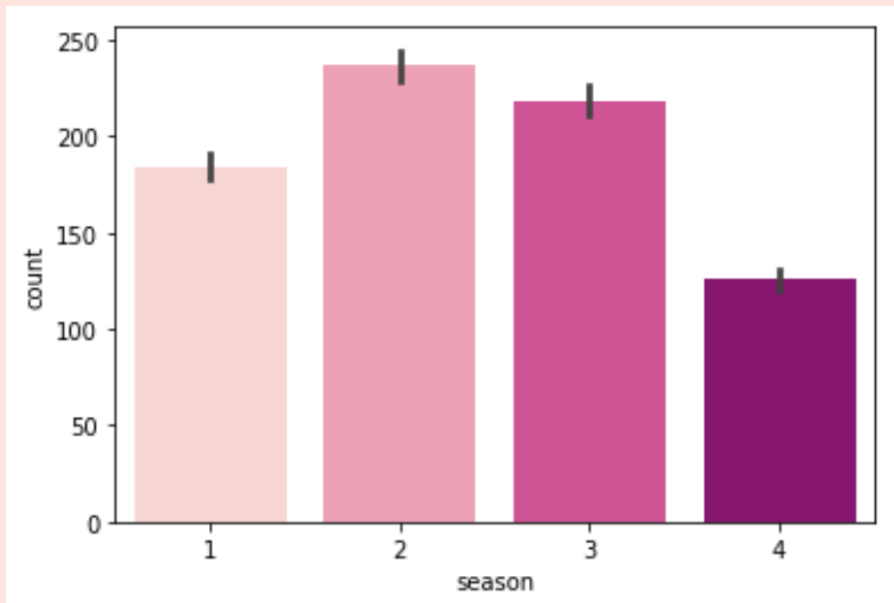


### 평일/주말 여부에 따른 시간별 대여 차이

- 평일과 주말을 나눠 보기 위해 요일변수 추가



## 2) EDA&전처리



...

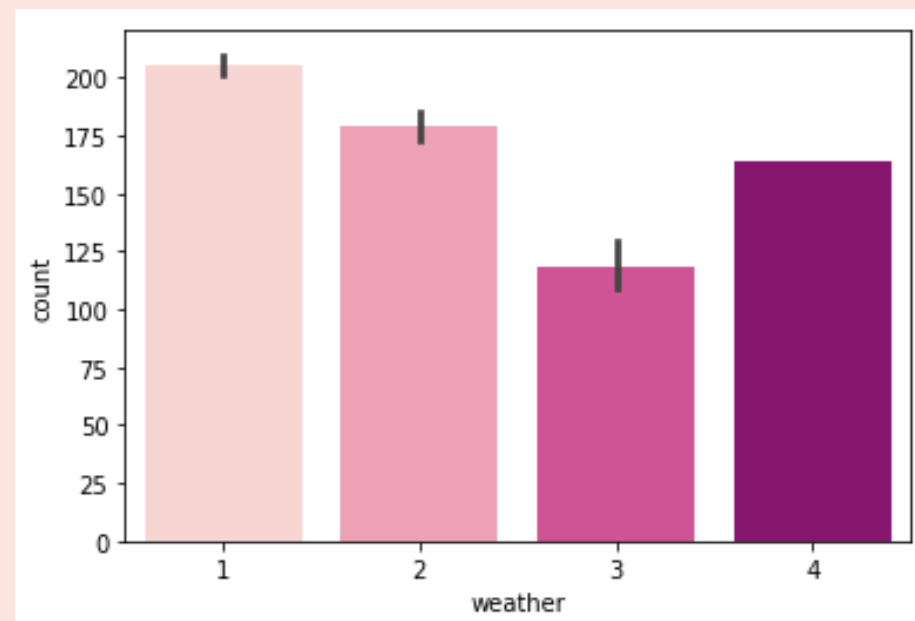
### 계절 변수 확인

1 ~ 3월	봄
4 ~ 6월	여름
7 ~ 9월	가을
10 ~ 12월	겨울

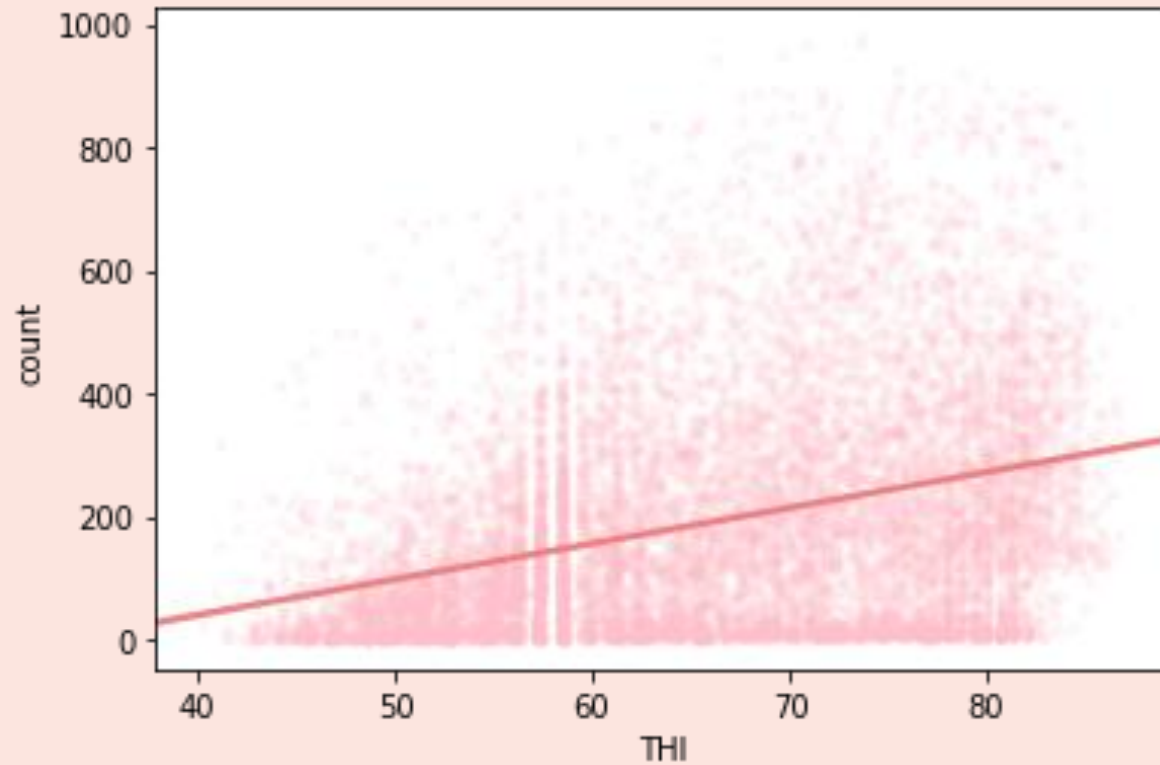
...

### 계절값 변경

3 ~ 5월	봄
6 ~ 8월	여름
9 ~ 11월	가을
12 ~ 2월	겨울



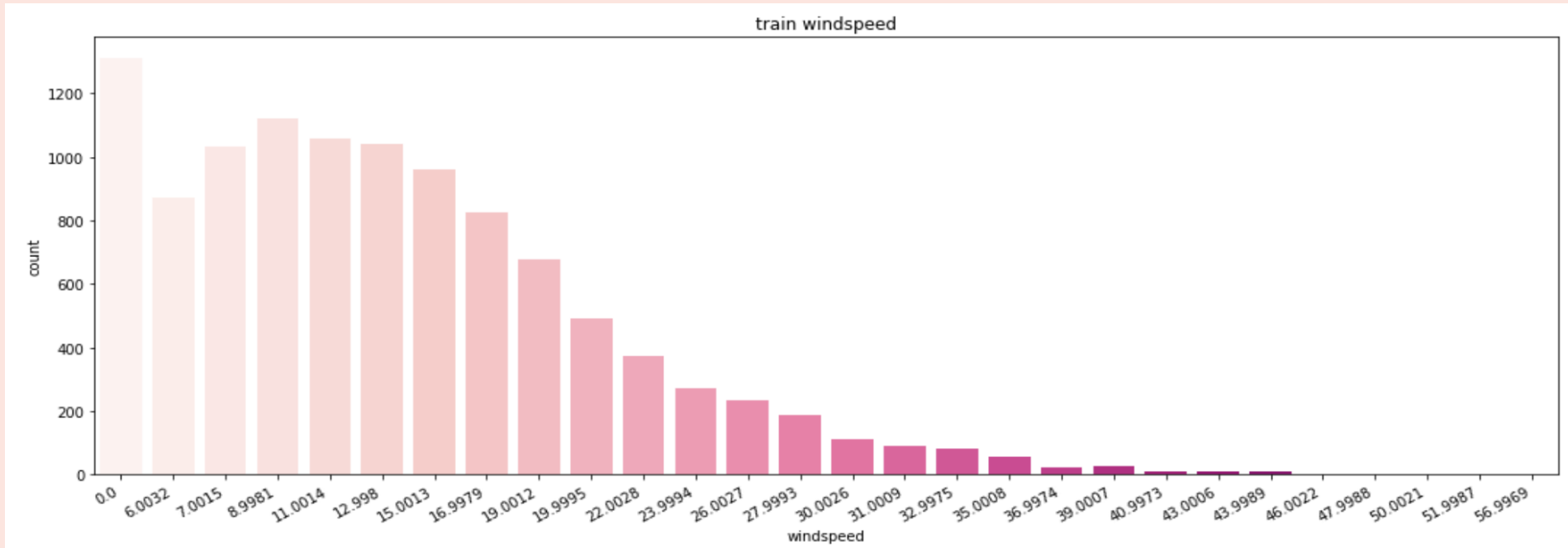
# 필요 변수 추가 -불쾌지수



- Temp와 Humidity를 매개변수로 하는 불쾌지수 변수 생성

...

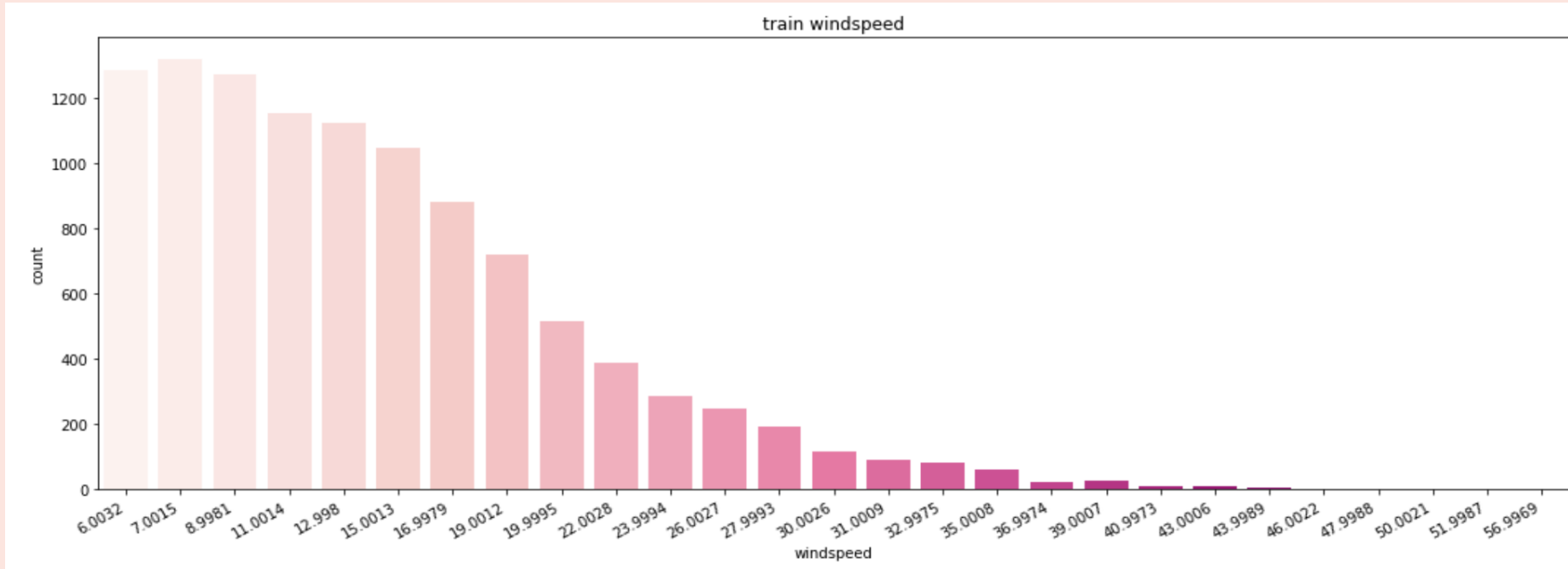
# Windspeed 변수 결측치 채우기



- 0이 과도하게 많이 관찰됨
- 값을 확인해보면 급격히 0으로 바뀌는 경우가 많음

→ 0인 값을 결측치로 판단

# Windspeed 변수 결측치 채우기



...

Random Forest를 이용해 결측치 채우기

# 03

## 모델링 적용

- 1) Decision Tree
- 2) Random Forest

# 모델링 적용 결과

Kaggle의 RMSLE 평가방식

Decision Tree	0.52623
Random Forest1	0.45602
Random Forest2	0.45488
Random Foresst3	0.46151
Xgboost	0.43746

## Random Forest1

- 전체 변수

## Random Forest2

- 전체 변수
- parameter값 조정

## Random Forest3

- 변수 수정
- parameter값 조정



# 04

## 결론

- 1) 활용방안
- 2) 한계

# 1) 활용방안



- **EDA 과정**을 통해
  - 날짜, 시간, 날씨 등의 영향에 따른 워싱턴 D.C 시민들의 이동 양상 파악 가능
- **Prediction**을 통해
  - 대여량이 높을 것으로 예측될 때 자전거 거치 수를 늘려 이용객 확보
  - 대여량이 낮을 것으로 예측될 때 자전거를 수거하여 수리 및 점검

## 2) 한계



- **대여소 변수**가 없어 더 자세한 분석이 불가능해 아쉬움
- **windspeed**가 0으로 나오는 값이 많아 예측의 정확도가 떨어져서 아쉬움
- **windspeed, temp, humidity** 등의 **연속변수**에서 **discrete**하게 나타난 것으로 보아 세부적으로 측정이 이루어지지 않은 것 같아 아쉬움

발표를 들어주셔서

**감사합니다**

...