

# Final Project

## 중간 발표

---

---

2021-2

---

Final Project

---

4조

---

# 목차

---

1. 데이터 소개 & 전처리

---

2. EDA

---

3. 모델링 문제

---

# 목차

---

## 1. 데이터 소개 & 전처리

---

## 2. EDA

---

## 3. 모델링 문제

---

# 서울시 공유 자전거 데이터

- 2017.12 ~ 2018.12월 서울시 공유 자전거(따릉이) 대여 수치
- 대기 관련 변수, 날씨 관련 변수 (공휴일 여부, 기능 중지 여부)
- 시간대별 데이터

	Date	Rented Bike Count	Hour	Temperature(* C)	Humidity(%)	Wind speed (m/s)	Visibility (10m)	Dew point temperature(* C)	Solar Radiation (MJ/m2)	Rainfall(mm)	Snowfall (cm)	Seasons	Holiday	Functioning Day
0	01/12/2017	254	0	-5.2	37	2.2	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
1	01/12/2017	204	1	-5.5	38	0.8	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
2	01/12/2017	173	2	-6.0	39	1.0	2000	-17.7	0.0	0.0	0.0	Winter	No Holiday	Yes
3	01/12/2017	107	3	-6.2	40	0.9	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
4	01/12/2017	78	4	-6.0	36	2.3	2000	-18.6	0.0	0.0	0.0	Winter	No Holiday	Yes

# 데이터 전처리

# 결측치 없음

bike.isnull().sum()

```
Date 0
Rented Bike Count 0
Hour 0
Temperature( 0
Humidity(%) 0
Wind speed (m/s) 0
Visibility (10m) 0
Dew point temperature( 0
Solar Radiation (MJ/m2) 0
Rainfall(mm) 0
Snowfall (cm) 0
Seasons 0
Holiday 0
Functioning Day 0
dtype: int64
```

bike.info() #dtype 통일 필요. Date는 시계열로 변환 필요.

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 8760 entries, 0 to 8759
```

```
Data columns (total 14 columns):
```

#	Column	Non-Null Count	Dtype
0	Date	8760 non-null	object
1	Rented Bike Count	8760 non-null	int64
2	Hour	8760 non-null	int64
3	Temperature(켈)	8760 non-null	float64
4	Humidity(%)	8760 non-null	int64
5	Wind speed (m/s)	8760 non-null	float64
6	Visibility (10m)	8760 non-null	int64
7	Dew point temperature(켈)	8760 non-null	float64
8	Solar Radiation (MJ/m2)	8760 non-null	float64
9	Rainfall(mm)	8760 non-null	float64
10	Snowfall (cm)	8760 non-null	float64
11	Seasons	8760 non-null	object
12	Holiday	8760 non-null	object
13	Functioning Day	8760 non-null	object

```
dtypes: float64(6), int64(4), object(4)
```

```
memory usage: 958.2+ KB
```

# 데이터 전처리

# 결측치 없음

bike.isnull().sum()

```
Date 0
Rented Bike Count 0
Hour 0
Temperature( 0
Humidity(%) 0
Wind speed (m/s) 0
Visibility (10m) 0
Dew point temperature( 0
Solar Radiation (MJ/m2) 0
Rainfall(mm) 0
Snowfall (cm) 0
Seasons 0
Holiday 0
Functioning Day 0
dtype: int64
```

bike.info() #dtype 통일 필요. Date는 시계열로 변환 필요.

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 8760 entries, 0 to 8759
```

```
Data columns (total 14 columns):
```

#	Column	Non-Null Count	Dtype
0	Date	8760 non-null	object
1	Rented Bike Count	8760 non-null	int64
2	Hour	8760 non-null	int64
3	Temperature(켈)	8760 non-null	float64
4	Humidity(%)	8760 non-null	int64
5	Wind speed (m/s)	8760 non-null	float64
6	Visibility (10m)	8760 non-null	int64
7	Dew point temperature(켈)	8760 non-null	float64
8	Solar Radiation (MJ/m2)	8760 non-null	float64
9	Rainfall(mm)	8760 non-null	float64
10	Snowfall (cm)	8760 non-null	float64
11	Seasons	8760 non-null	object
12	Holiday	8760 non-null	object
13	Functioning Day	8760 non-null	object

```
dtypes: float64(6), int64(4), object(4)
```

```
memory usage: 958.2+ KB
```

# 데이터 전처리

```
# Holiday열에서 Holiday이면 1, No Holiday이면 0으로 더미변수화
bike.loc[bike['Holiday'] == 'Holiday', 'Holiday'] = 1
bike.loc[bike['Holiday'] == 'No Holiday', 'Holiday'] = 0
```

```
# Functioning Day열에서 Yes이면 1, No이면 0으로 더미변수화
bike.loc[bike['Functioning Day'] == 'Yes', 'Functioning Day'] = 1
bike.loc[bike['Functioning Day'] == 'No', 'Functioning Day'] = 0
```

[illegible]

# 목차

---

1. 데이터 소개 & 전처리

---

2. EDA

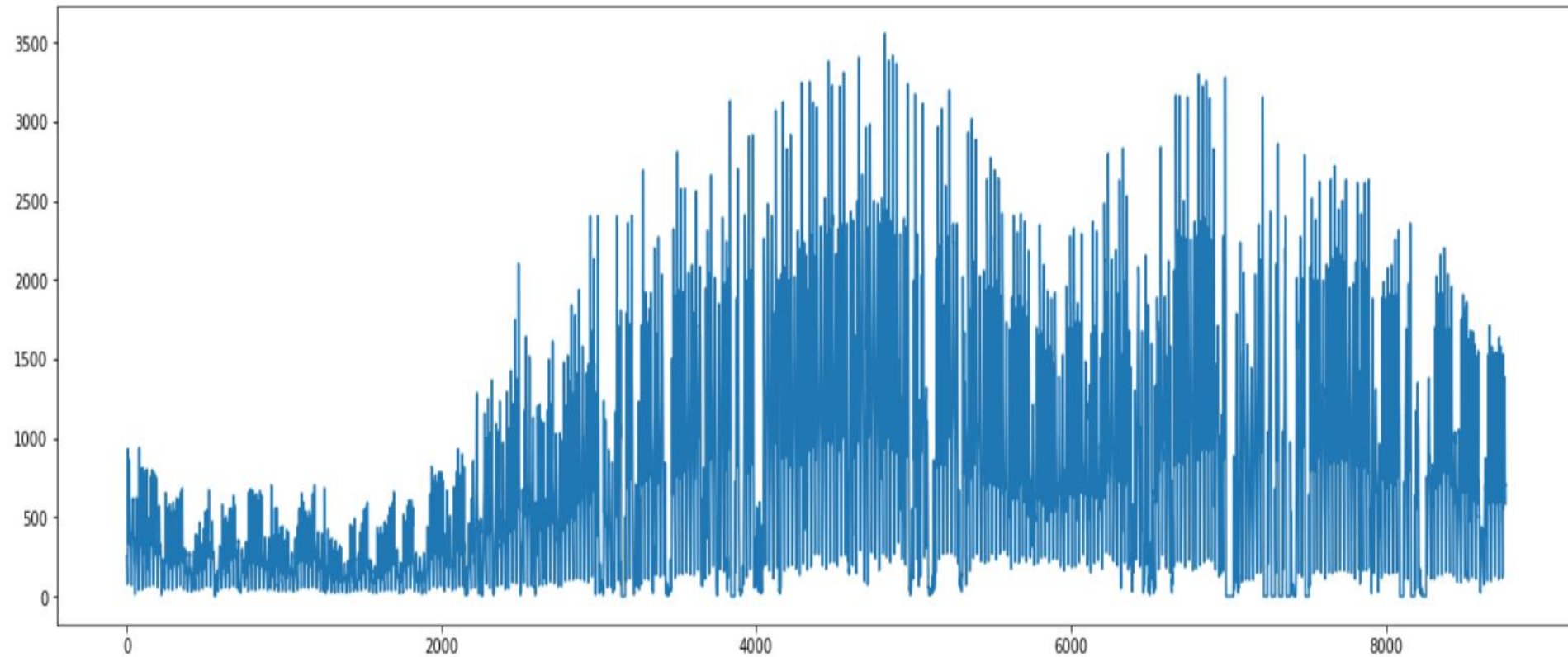
---

3. 모델링 문제

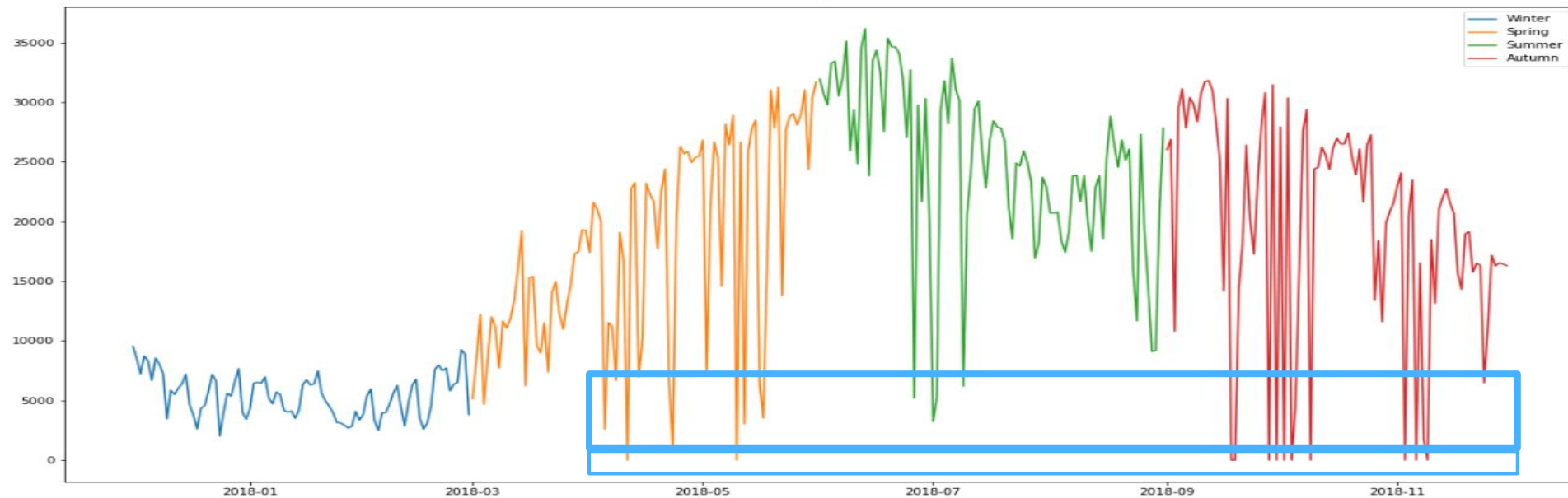
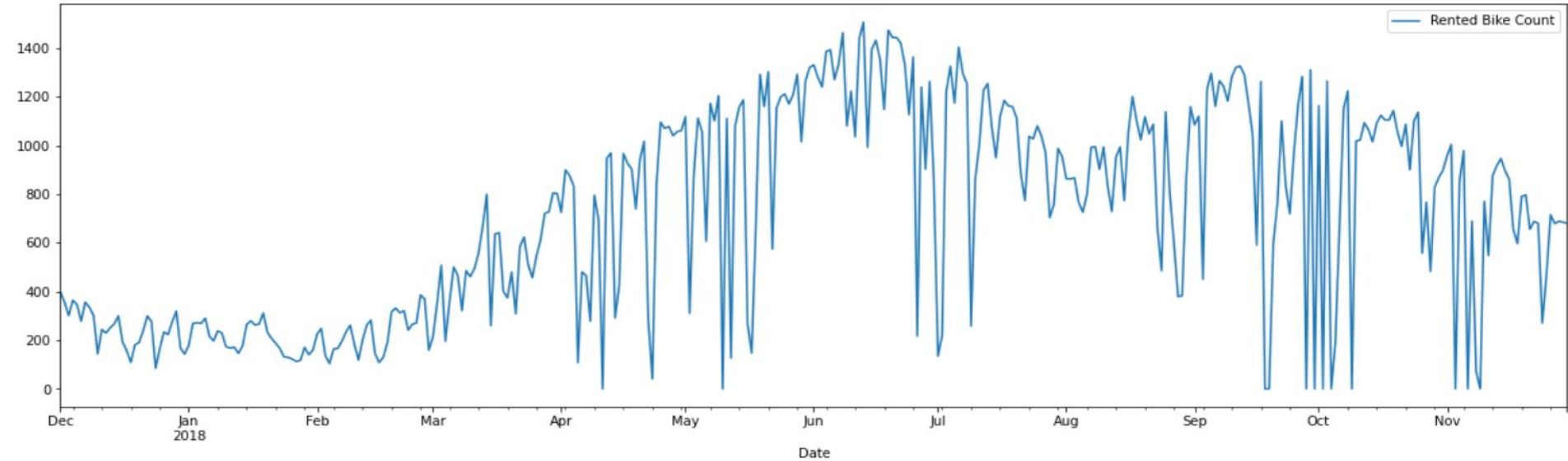
---



# 공유 자전거 대여량



# 공유 자전거 대여량



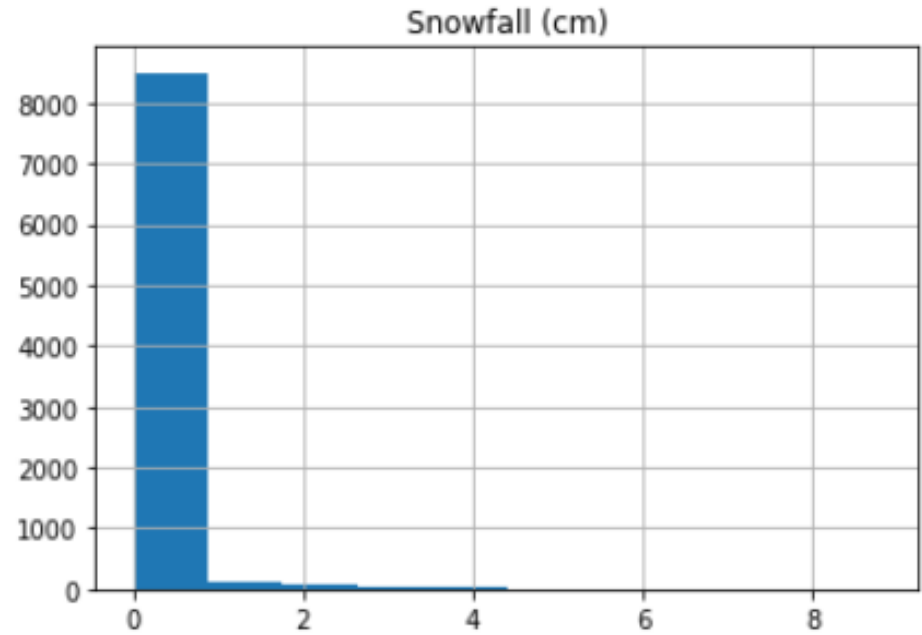
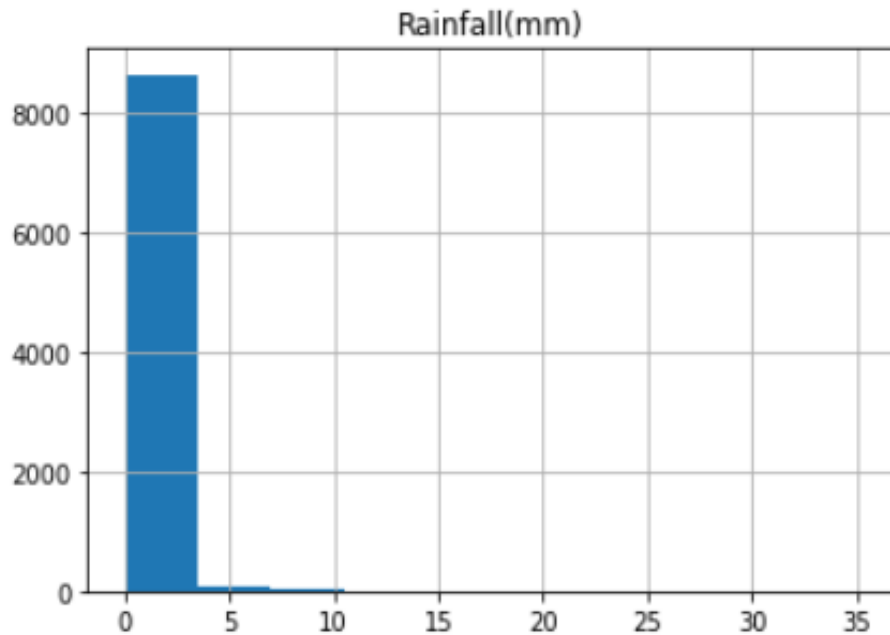
	Date	Functioning Day
3144	2018-04-11	No
3145	2018-04-11	No
3146	2018-04-11	No
3147	2018-04-11	No
3148	2018-04-11	No
...	...	...
8251	2018-11-09	No
8252	2018-11-09	No
8253	2018-11-09	No
8254	2018-11-09	No
8255	2018-11-09	No

295 rows × 16 columns

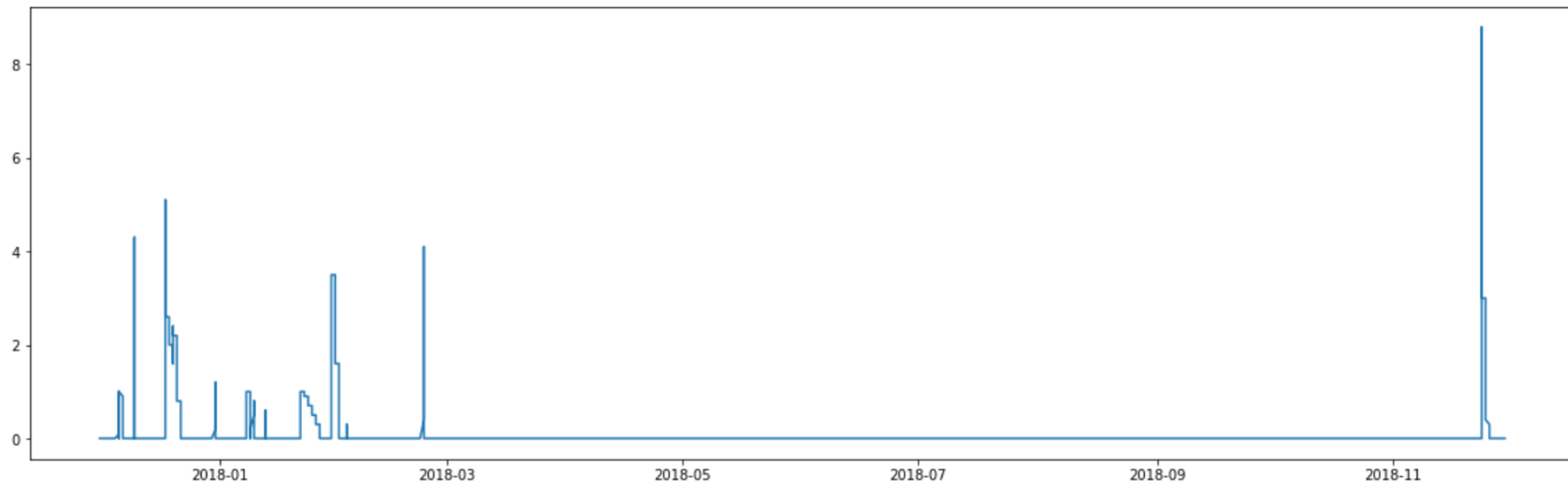
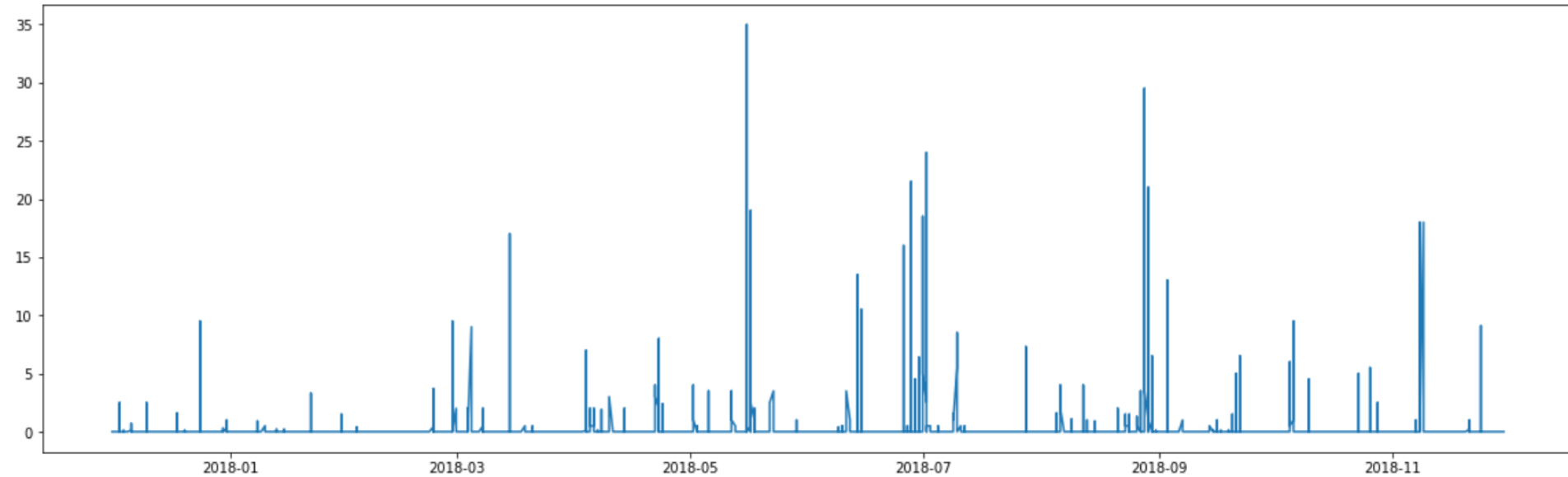
```
zerodata[zerodata['Functioning Day']=='Yes']
```

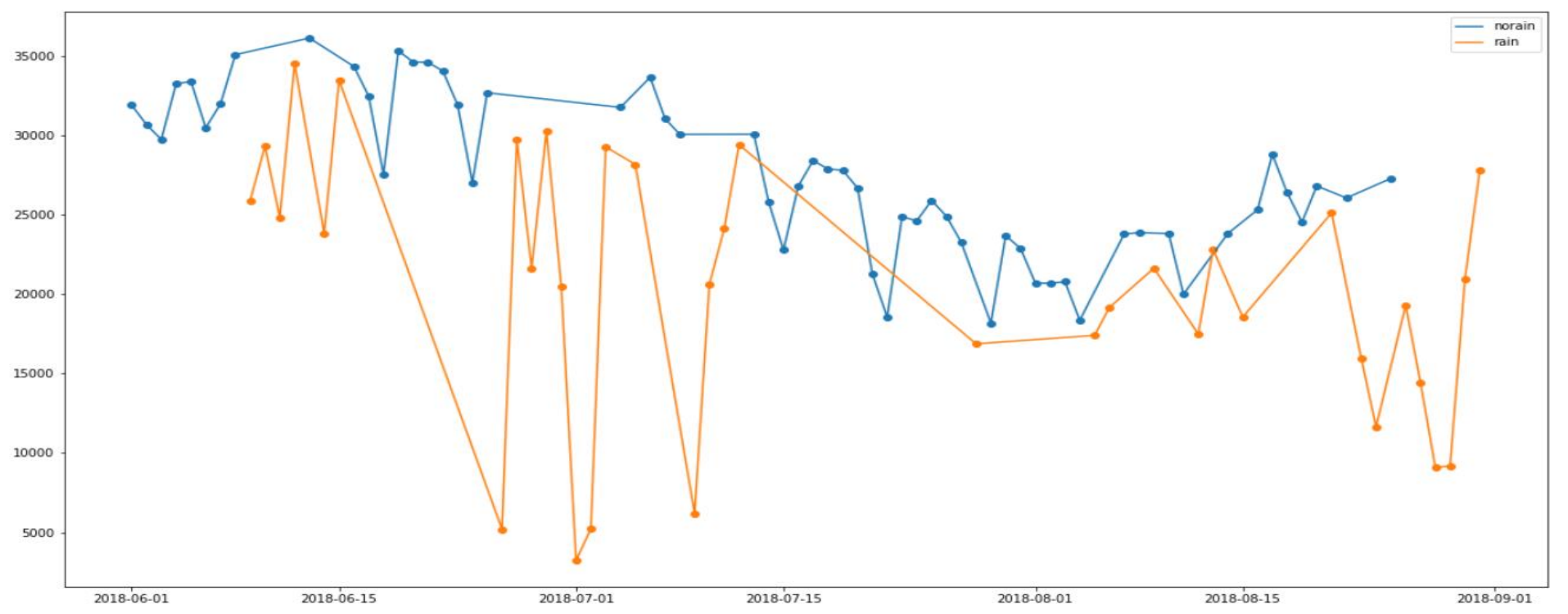
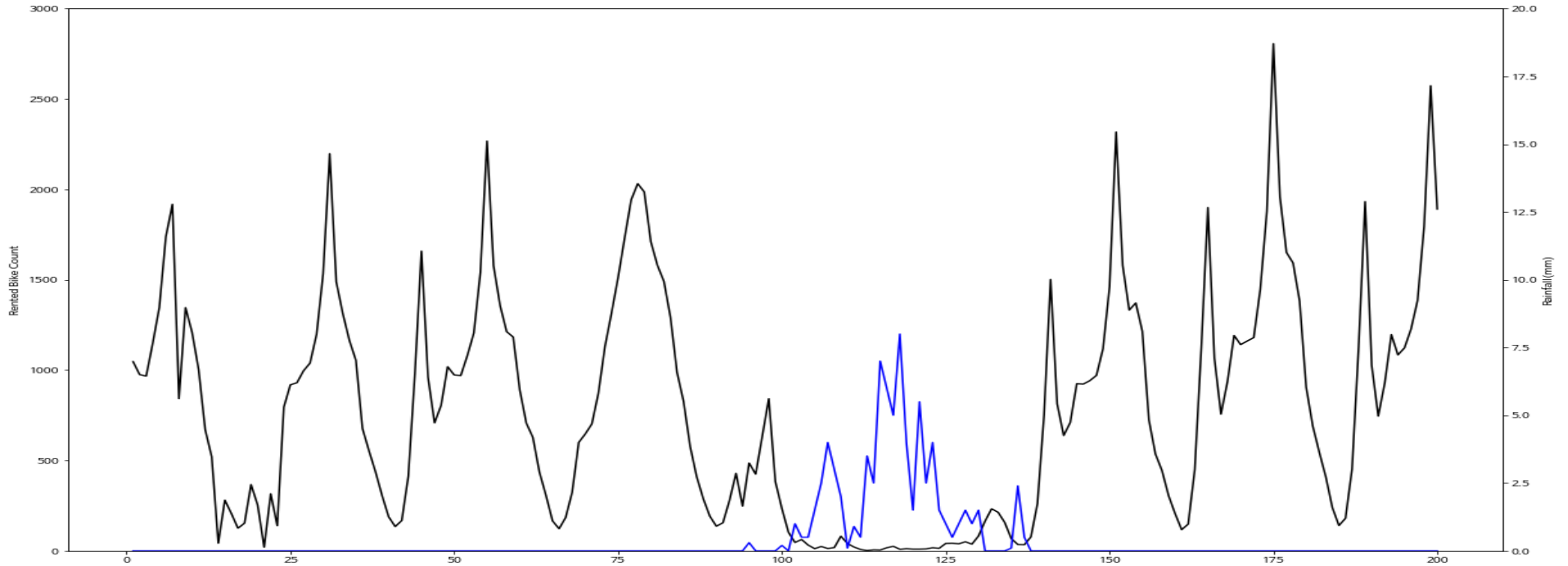
	Date	Rented Bike Count	Hour	Temperature(° C)	Humidity(%)
--	------	-------------------	------	------------------	-------------

# 강수량 / 적설량 분포

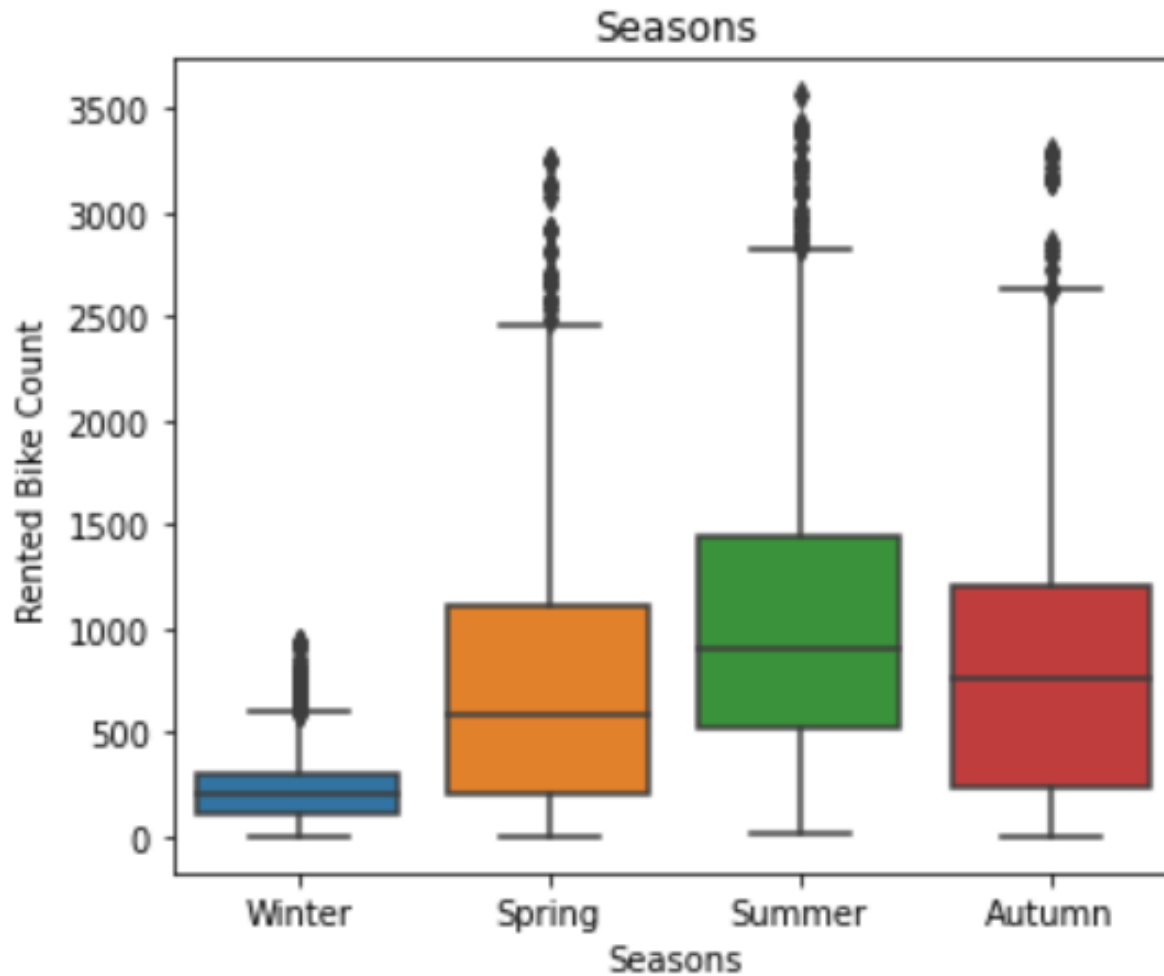


# 강수량 / 적설량 분포

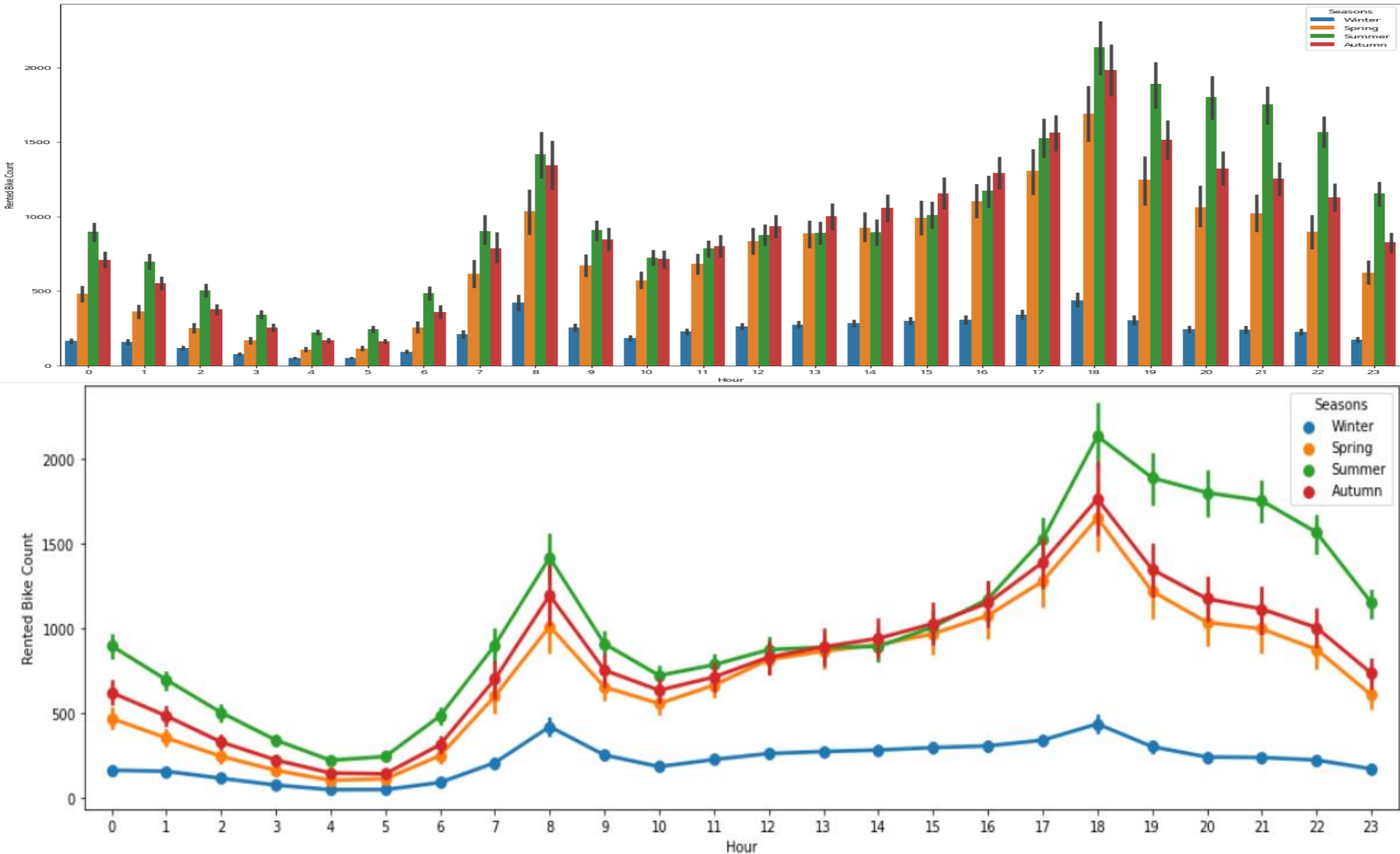




# 공유 자전거 계절별 대여량

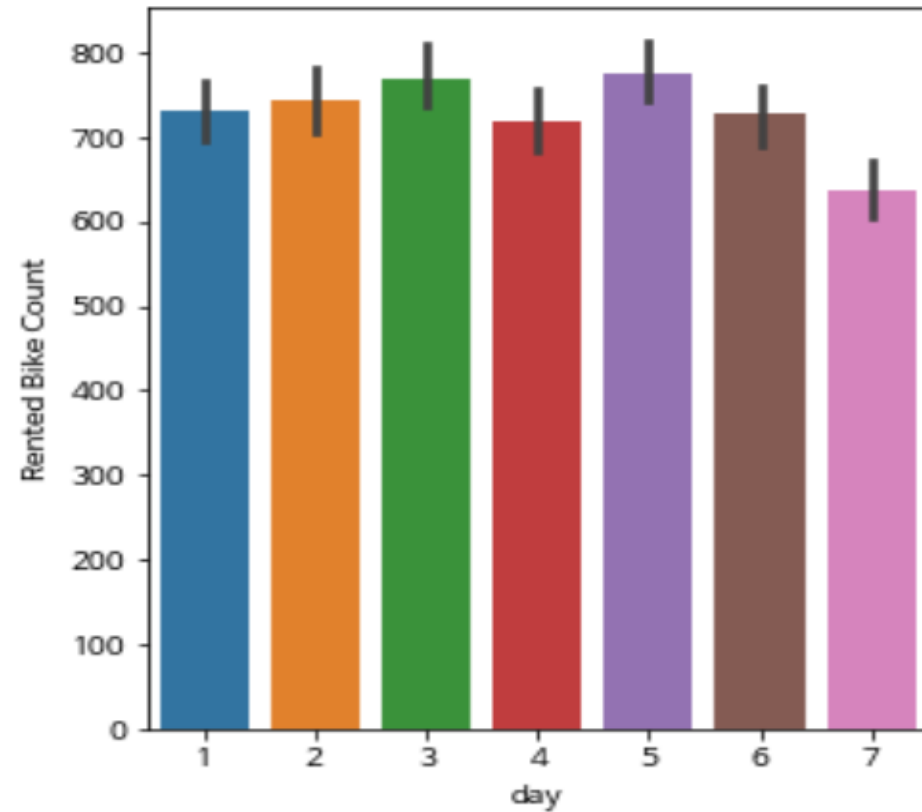
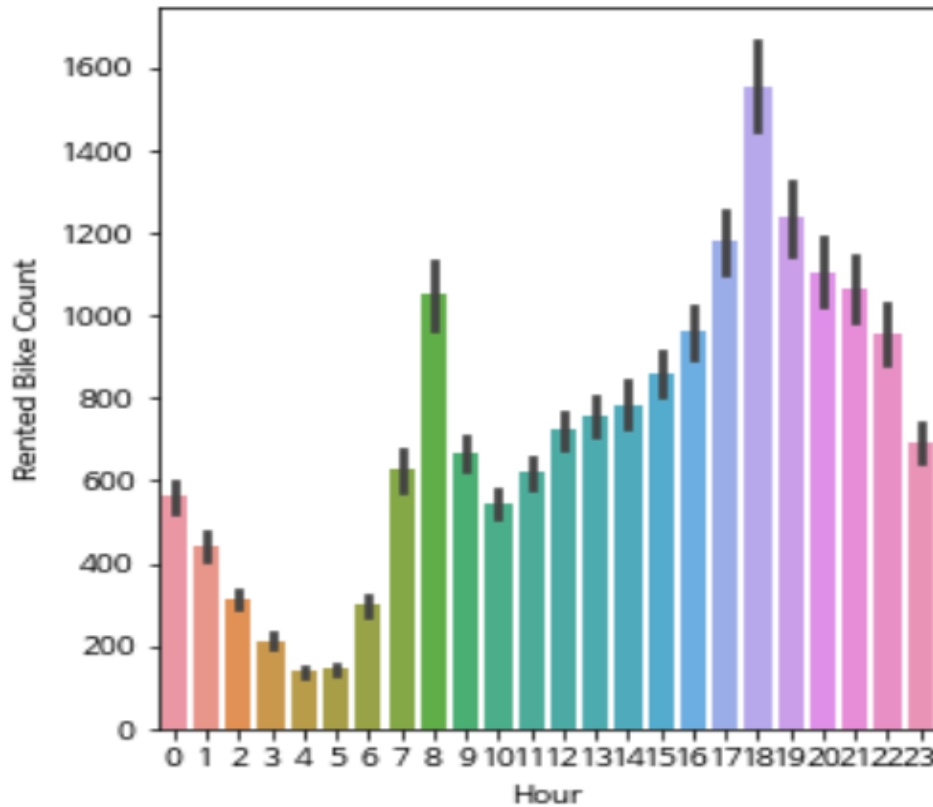


# 공유 자전거 계절별 대여량

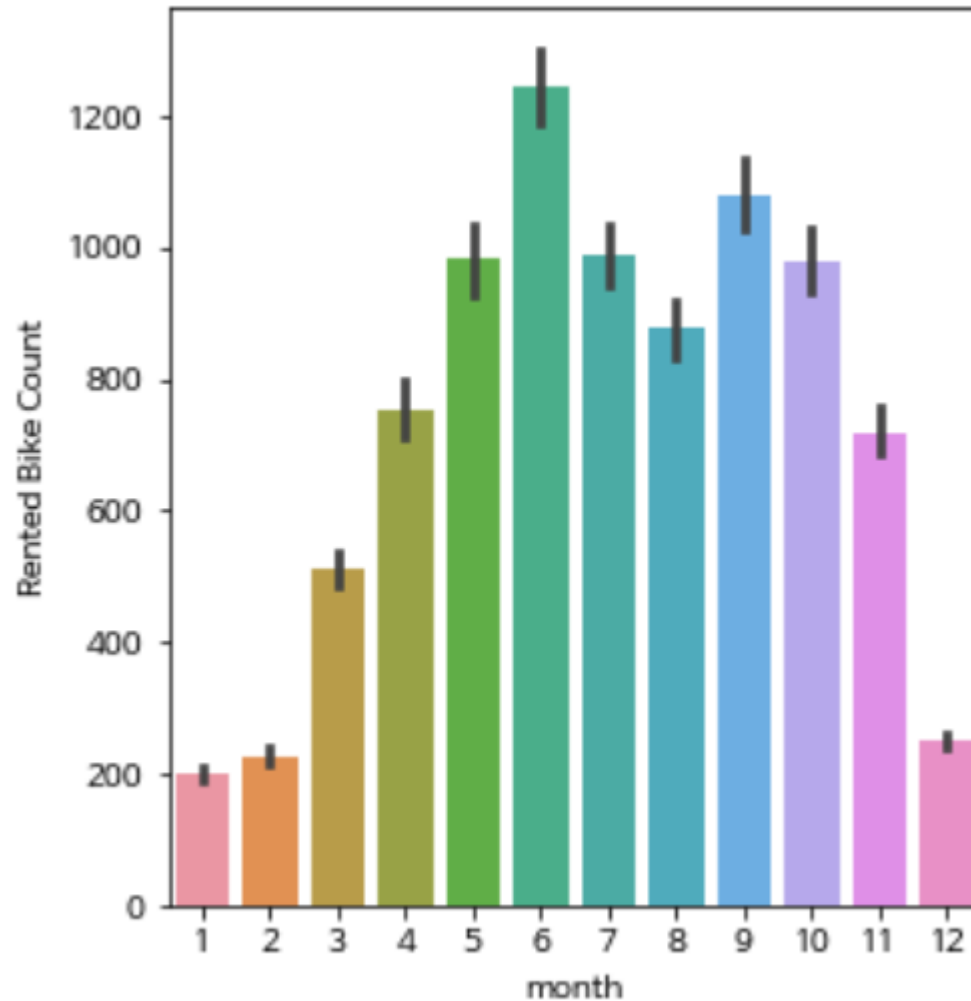




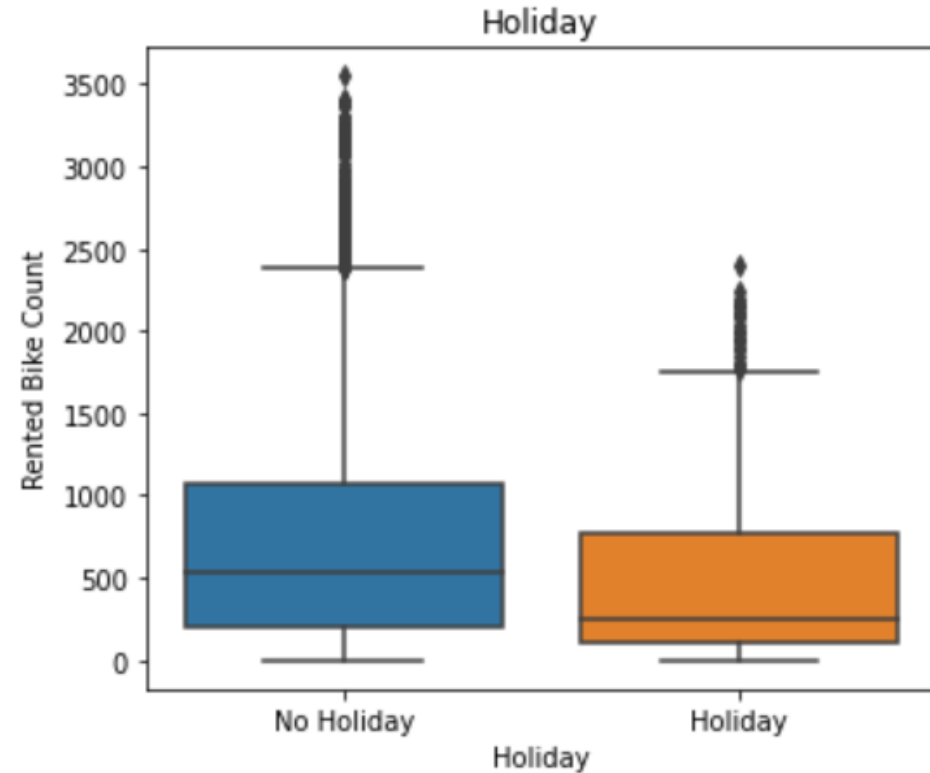
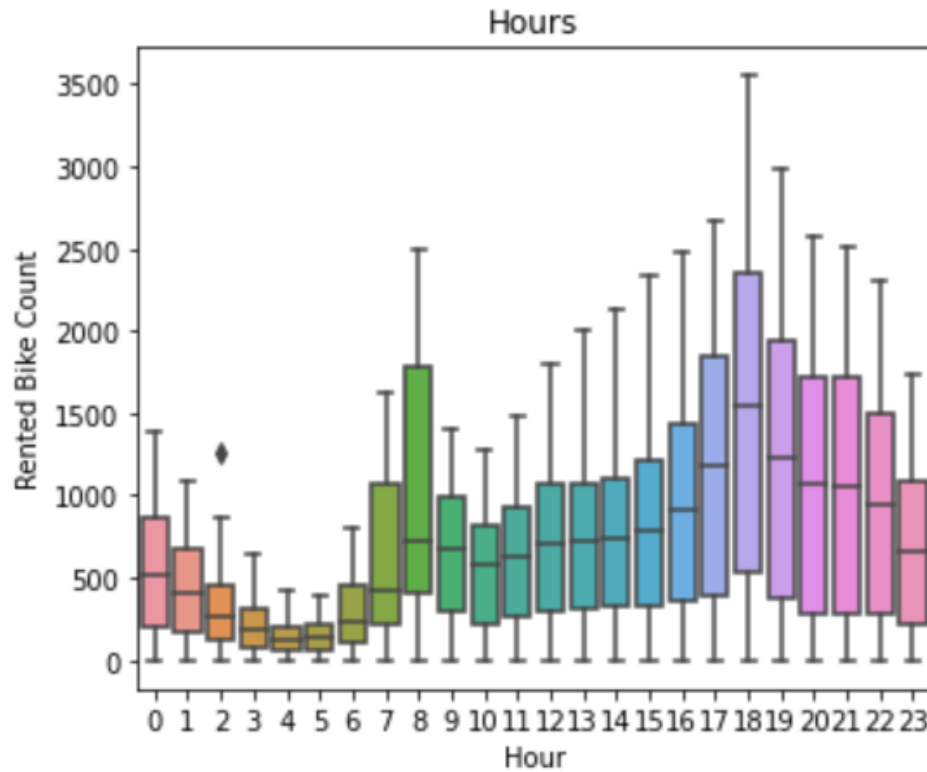
# 공유 자전거 시간대별 대여량

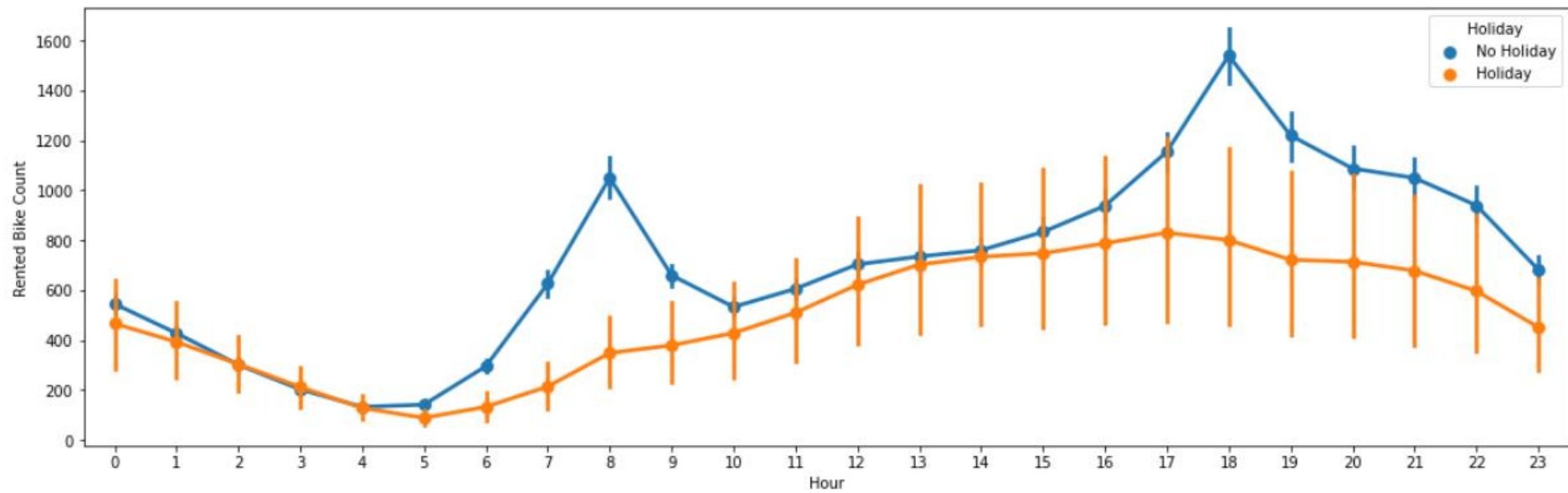
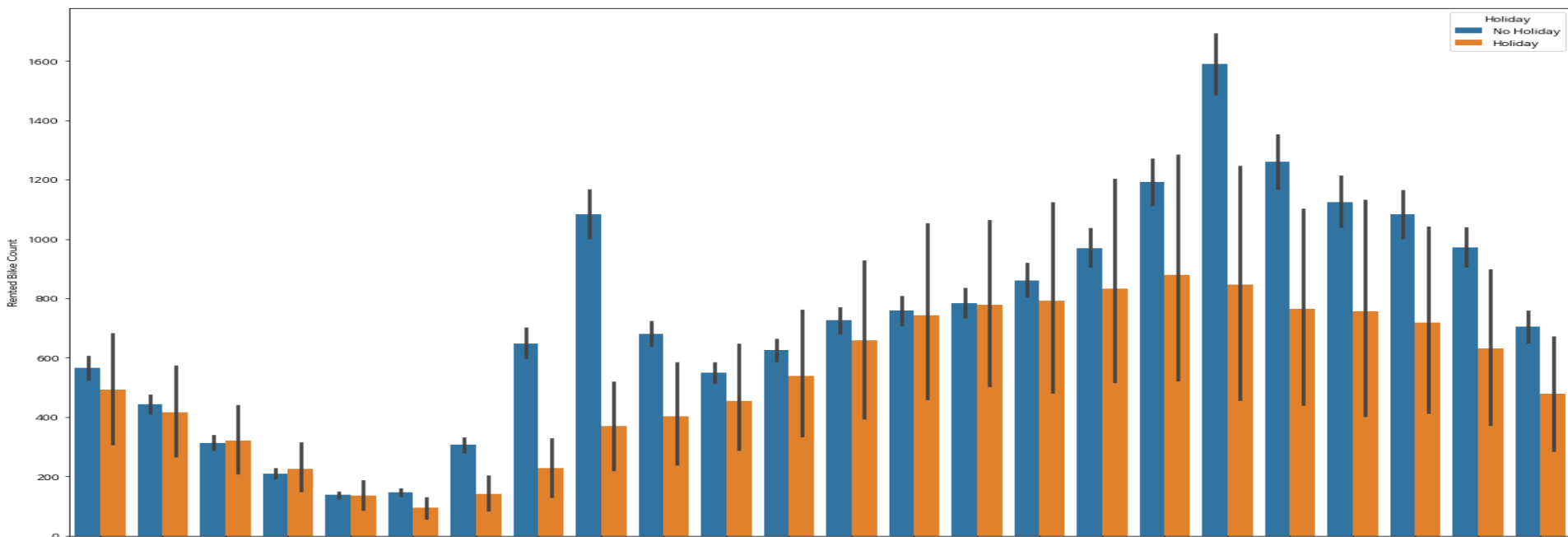


# 공유 자전거 시간대별 대여량

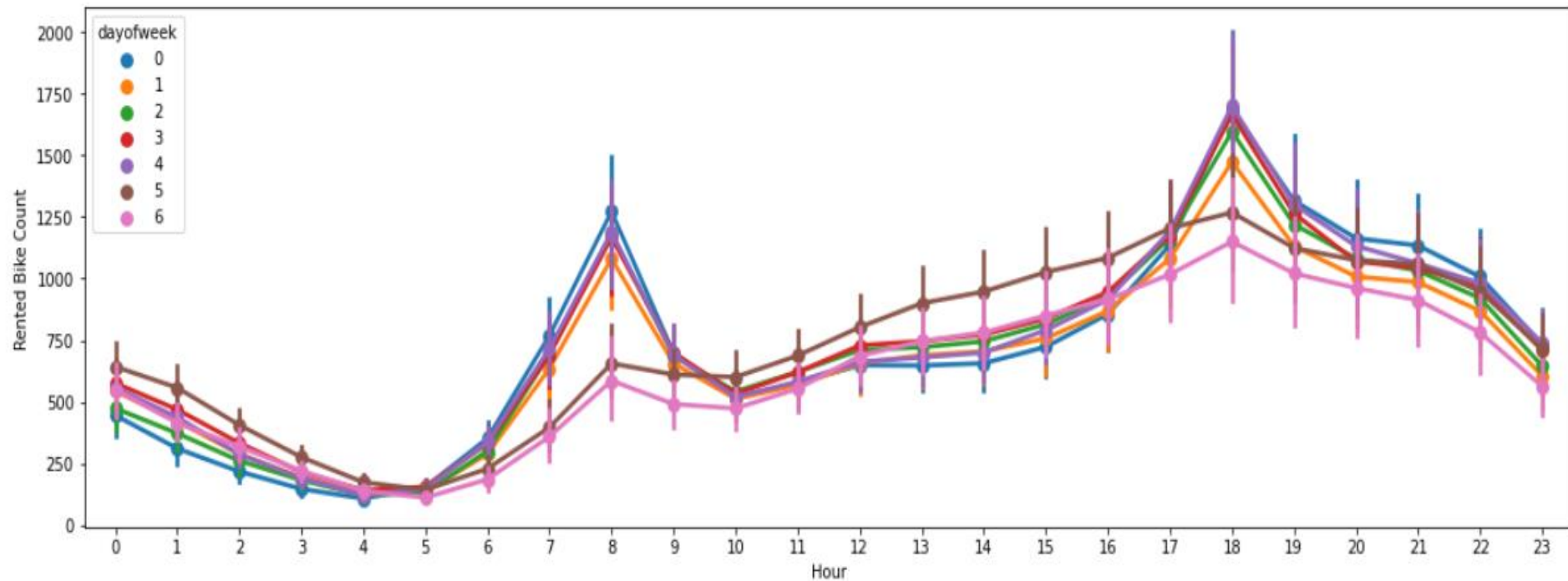


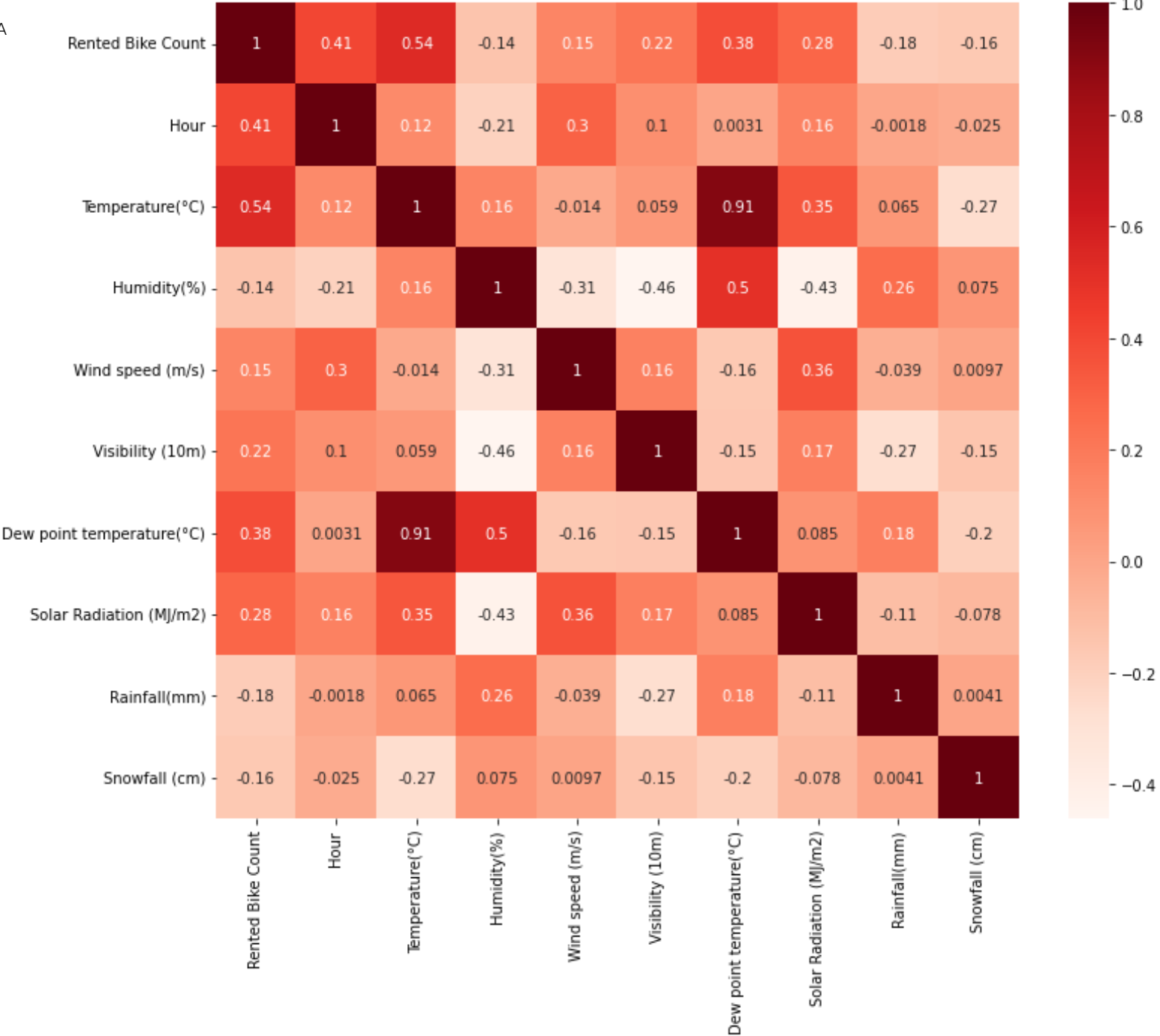
# 공유 자전거 시간대별 대여량





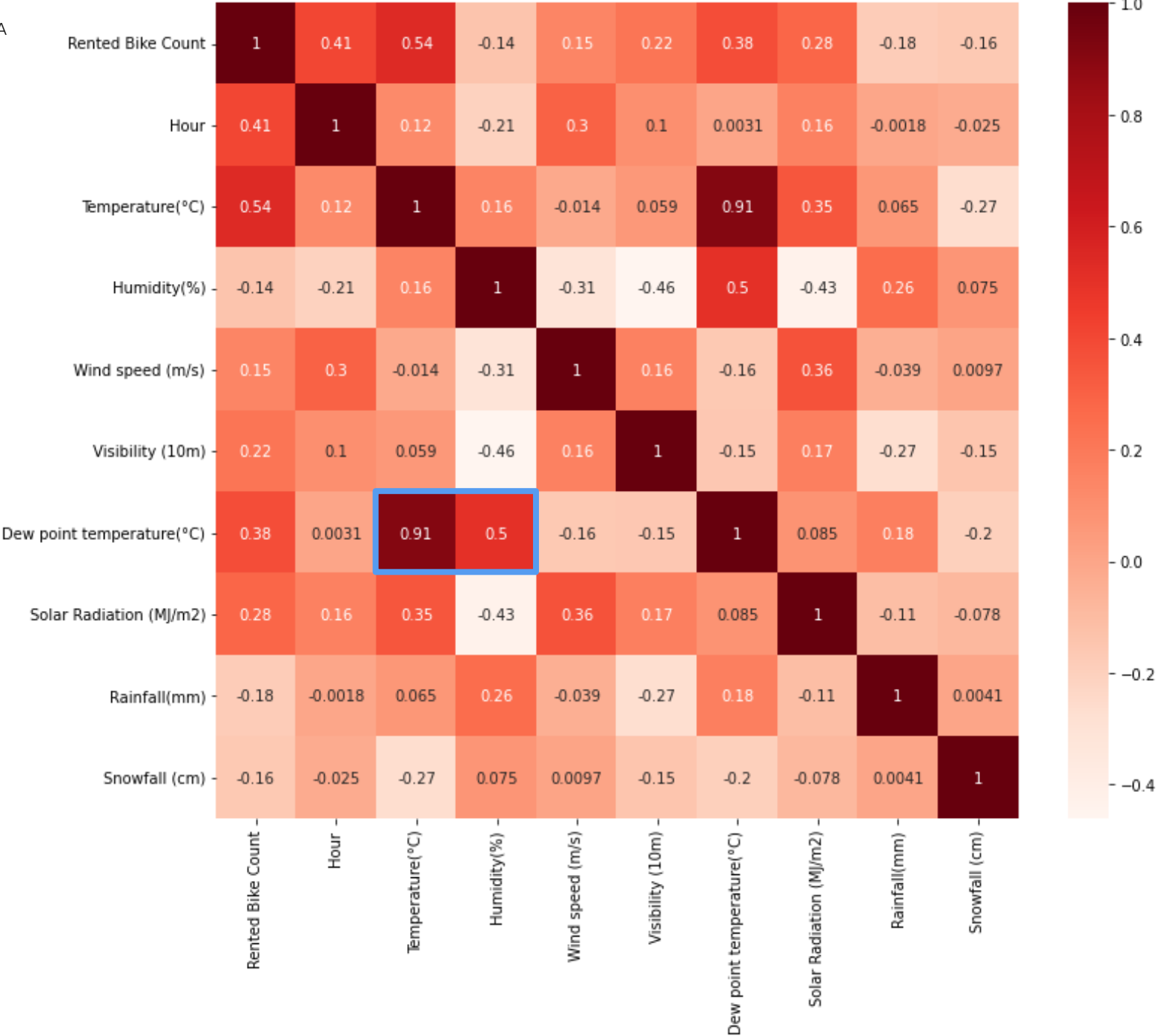
# 공유 자전거 계절별 대여량





# VIF Factor

	VIF Factor	features
0	1.057800	Holiday
1	1.082729	Rainfall(mm)
2	1.121269	Snowfall (cm)
3	2.883130	Solar Radiation (MJ/m2)
4	4.421340	Hour
5	4.806280	Wind speed (m/s)
6	9.121531	Visibility (10m)
7	17.993491	Humidity(%)
8	25.511824	Dew point temperature
9	27.817104	Functioning Day
10	44.879238	Temperature

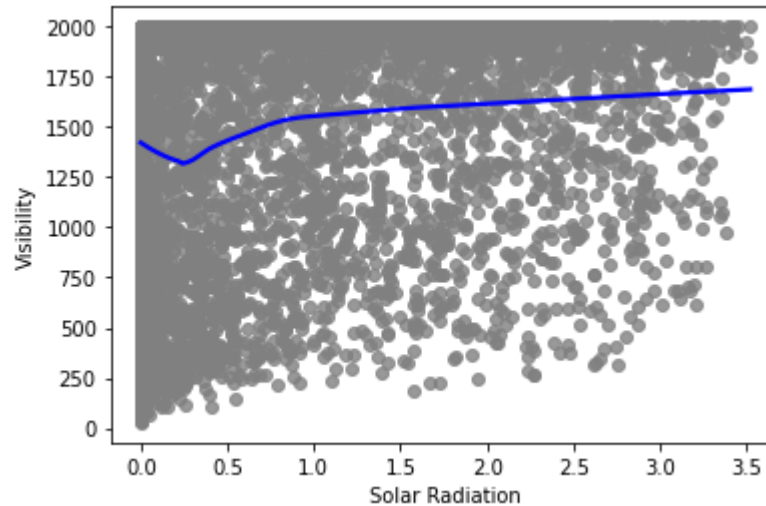
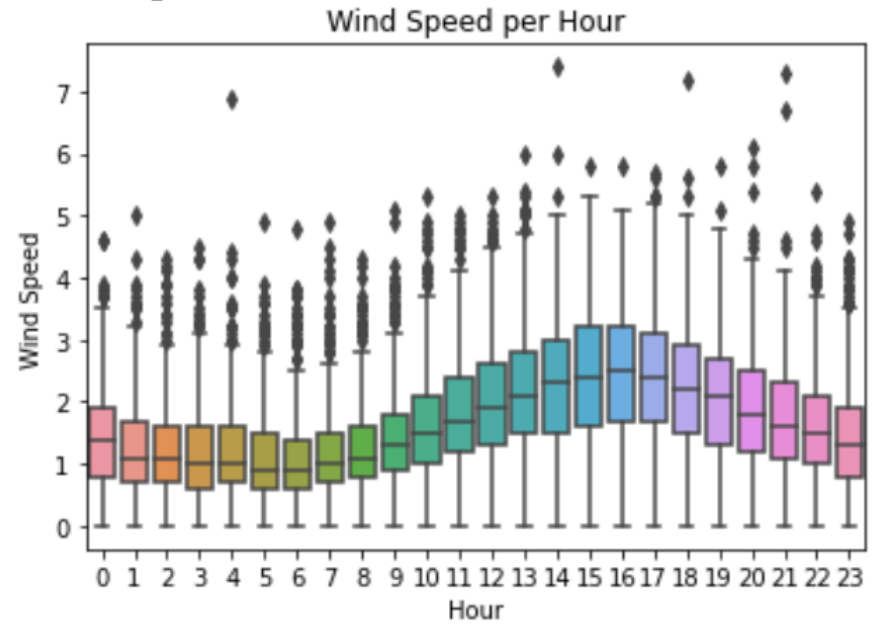
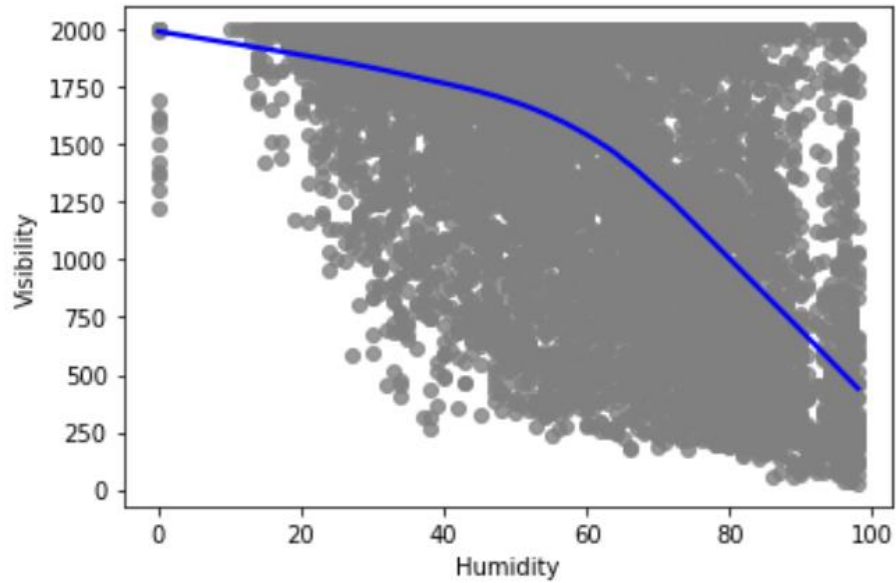




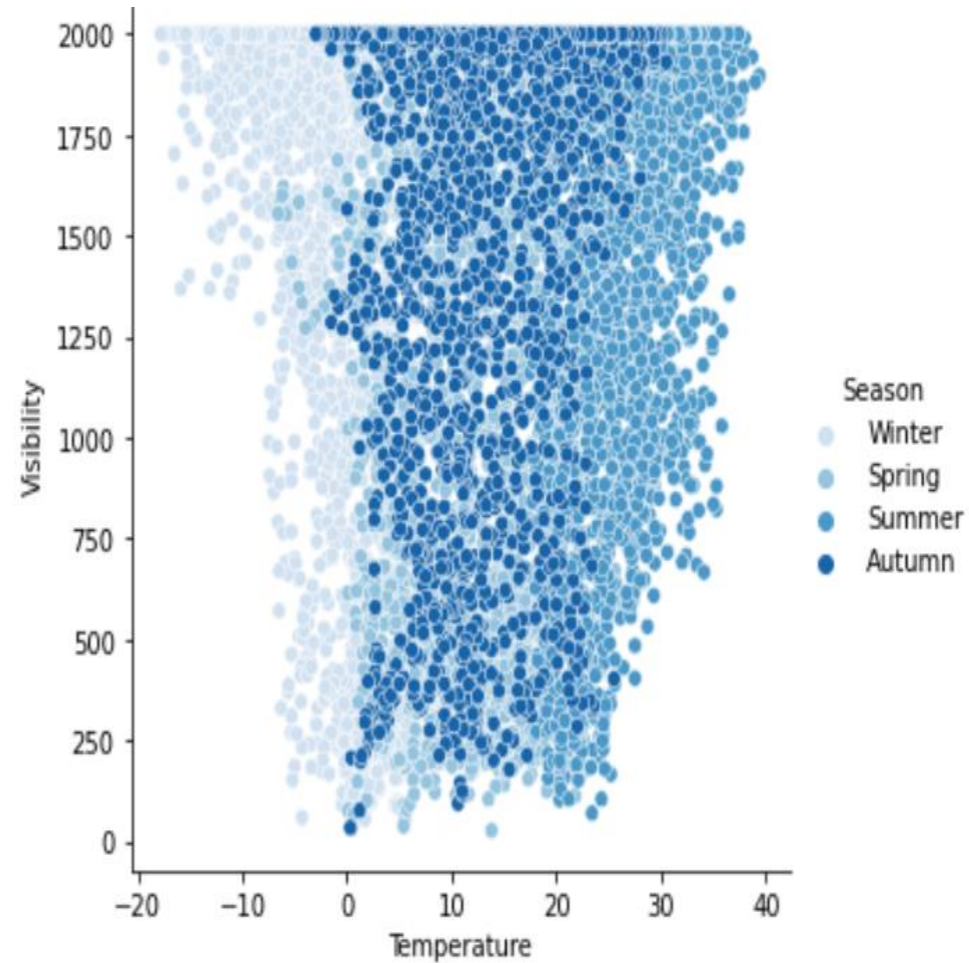
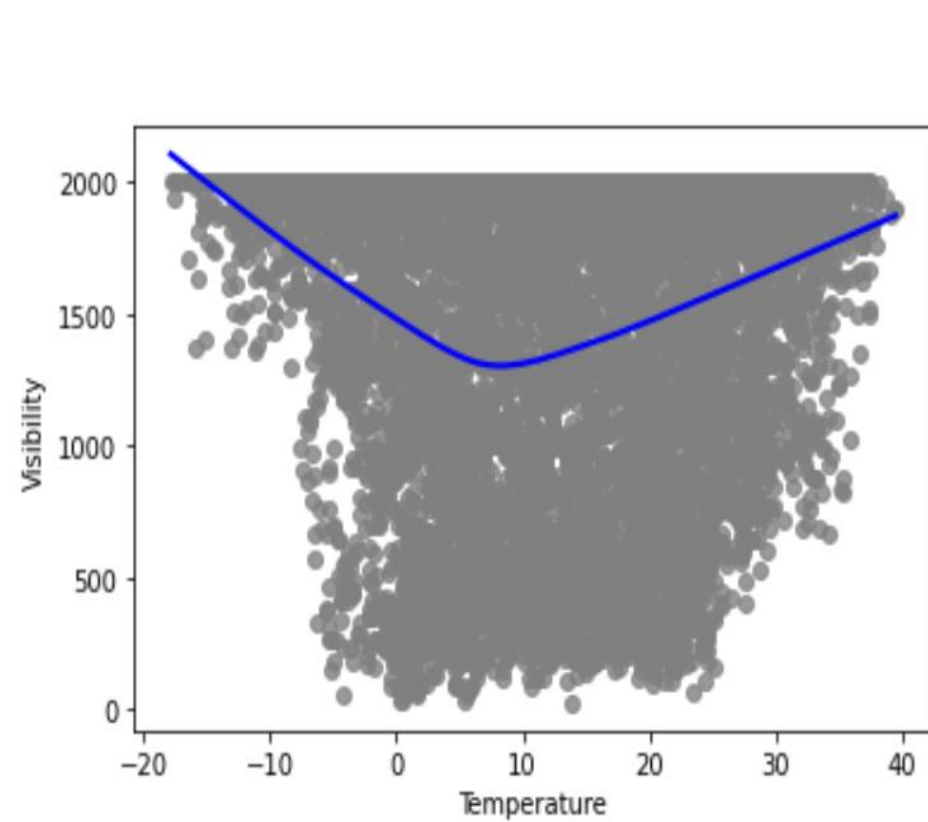
# 이슬점 ~ 습도 + 온도

상대습도 (%)	대기온도(℃)									
	-5	0	5	10	15	20	25	30	35	40
90	-6.5	-1.3	3.5	8.2	13.3	18.3	23.2	28.0	33.0	38.2
85	-7.2	-2.0	2.3	7.3	12.5	17.4	22.1	27.0	32.0	37.1
80	-7.7	-2.8	1.9	6.5	11.6	16.5	21.0	25.9	31.0	36.2
75	-8.4	-3.6	0.9	5.6	10.4	15.4	19.9	24.7	29.6	34.5
70	-9.2	-4.5	0.2	4.5	9.1	14.2	18.6	23.3	28.1	33.5
65	-10.0	-5.4	-1.0	3.3	8.0	13.0	17.4	22.0	26.8	32.0
60	-10.8	-6.5	-2.1	2.3	6.7	11.9	16.2	20.6	25.3	30.5
55	-11.6	-7.4	-3.2	1.0	5.6	10.4	14.8	19.1	23.9	28.9
50	-12.8	-8.4	-4.4	-0.3	4.1	8.6	13.3	17.5	22.2	27.1
45	-14.3	-9.6	-5.7	-1.5	2.6	7.0	11.7	16.0	20.2	25.2
40	-15.9	-10.8	-7.3	-3.1	0.9	5.4	9.5	14.0	18.2	23.0
35	-17.5	-12.1	-8.6	-4.7	-0.8	3.4	7.4	12.0	16.1	20.6
30	-19.0	-14.3	-10.2	-6.9	-2.9	1.3	5.2	9.2	13.7	18.0

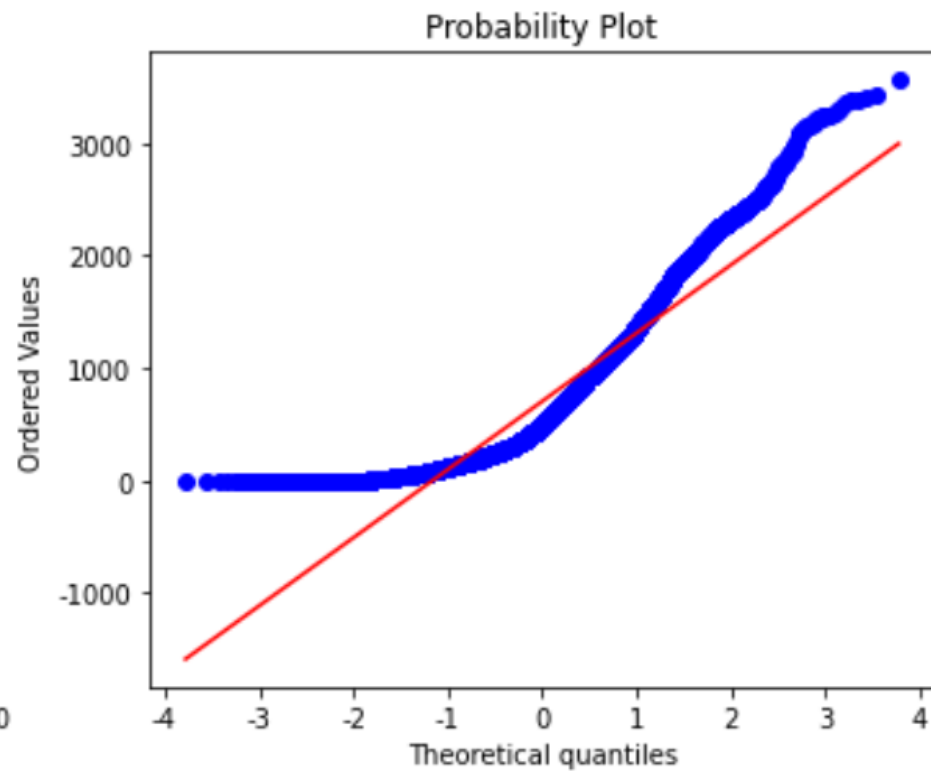
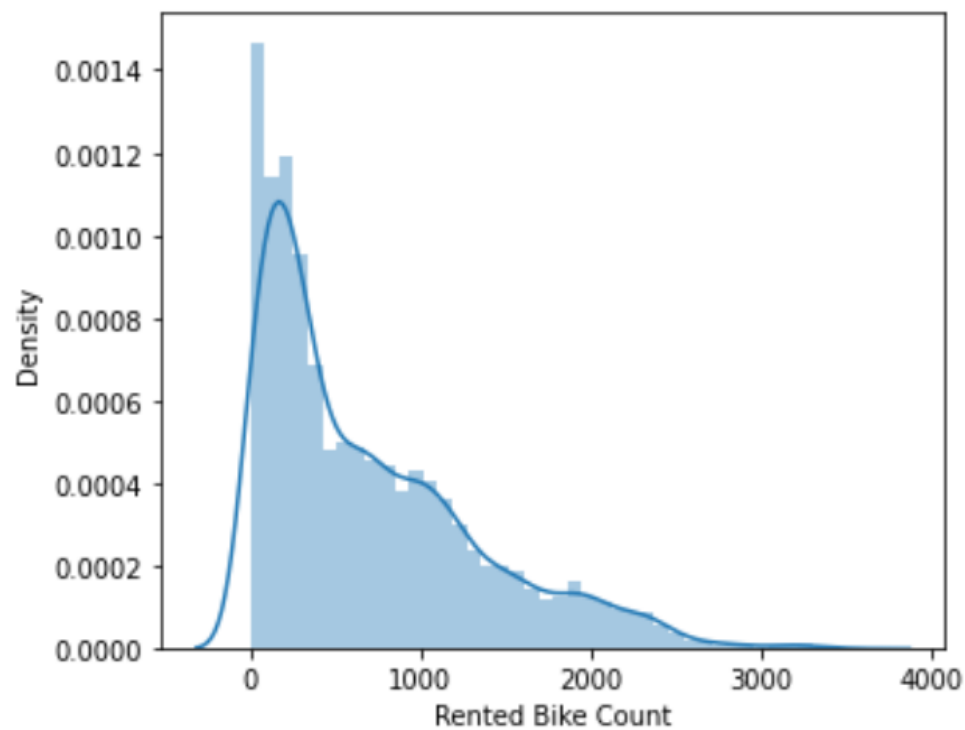
# 대기 변수



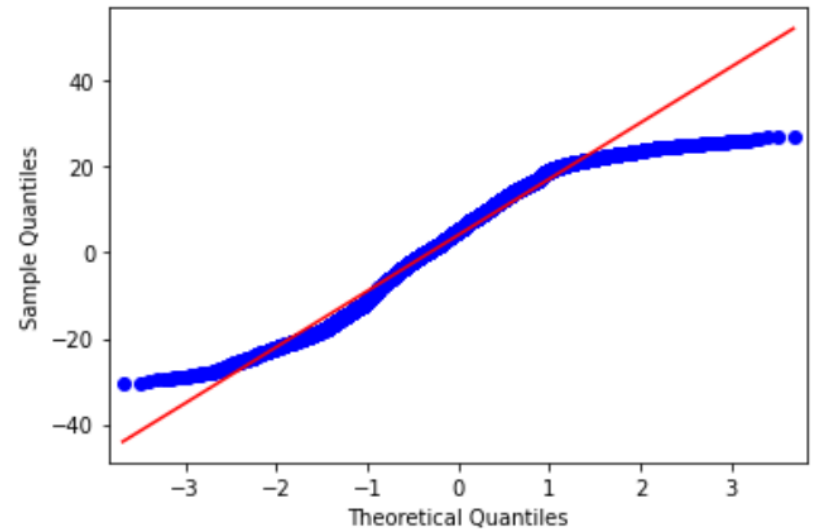
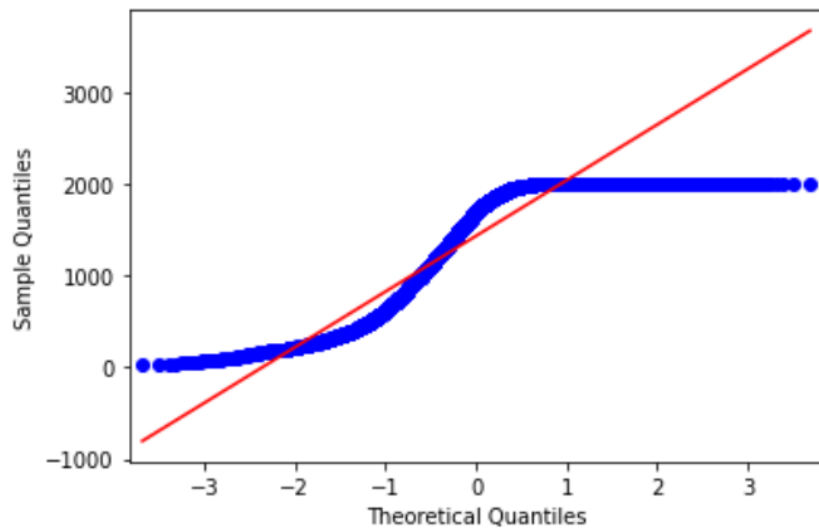
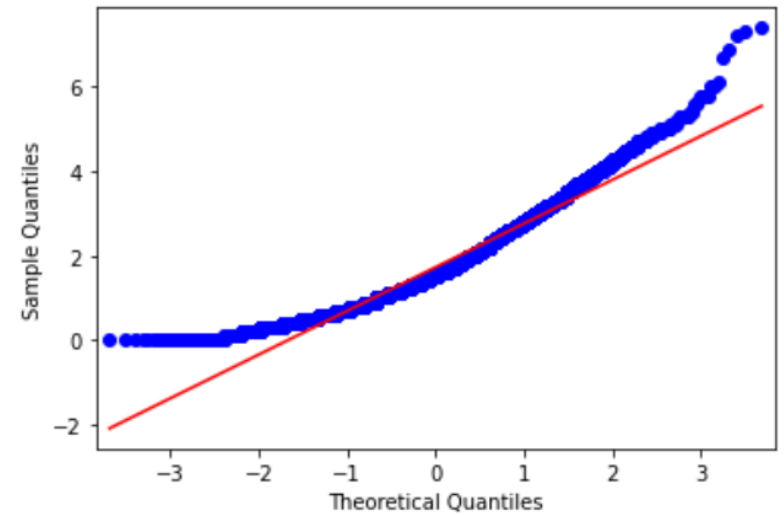
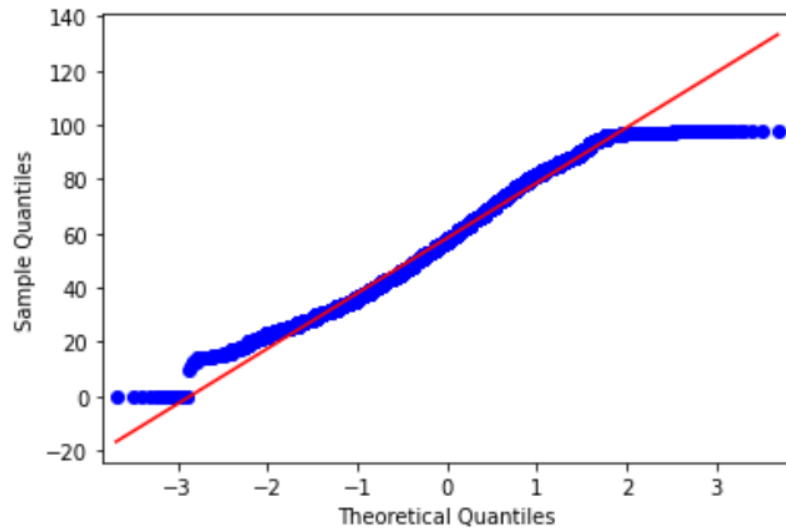
# 온도 ~ 시정(가시거리)



# 각 변수 정규성



# 각 변수 정규성



# 목차

---

1. 데이터 소개 & 전처리

---

2. EDA

---

3. 모델링 문제

---

# 1. 문제 정의

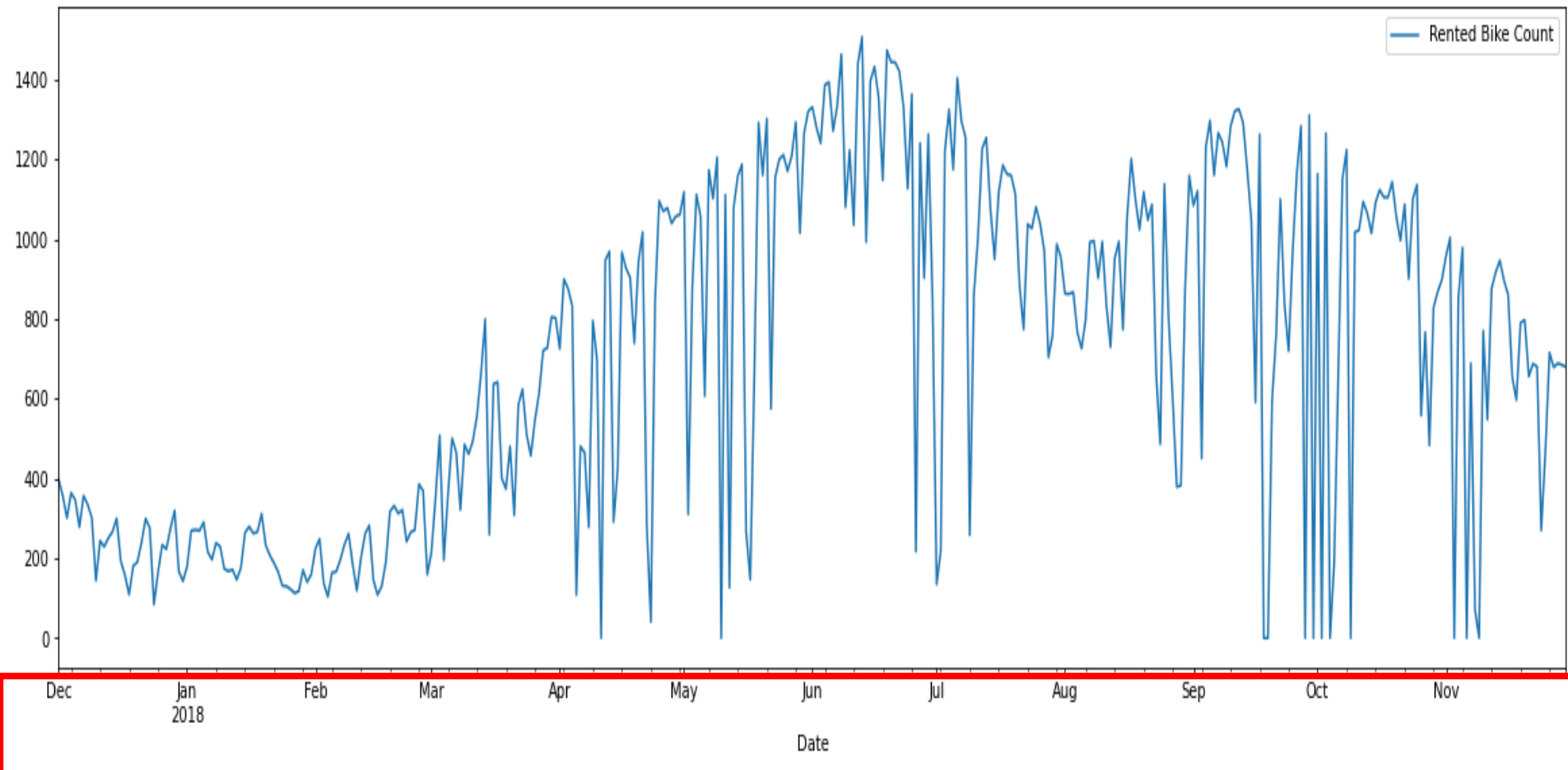
## 시간대별 데이터

- 시간대별 예측?
- 일자별 예측?

## 1년치 데이터

- 데이터 추가 및 예측 모델링
- 마지막 1달을 validation set

# 데이터 문제





감사합니다

---