

DATA 583 project proposal: Analysis of Seoul Bike Sharing Demand

Jade Yu, Yahan Cong

Introduction:

Bike-sharing system in Korea serves as a vital component of urban transportation which offers widespread accessibility, convenience, and energy conservation. For operators of bike-sharing services, maintaining a stable supply and strategically managing bike resources pose significant challenges. This research aims to explore how the atmosphere and holiday conditions affect bike renting. And build machine-learning models for rented bike prediction.

Dataset Description:

The dataset “Seoul Bike Sharing Demand” records the number of public bicycles rented per hour in the Seoul Bike Sharing System from January 2017 to November 2018. It contains 8760 samples of hourly rented bike numbers with corresponding weather data and holiday information.

This dataset has 2 dependent variables: Rented Bike Count, and Functioning Day. Rented Bike Count denotes the number of rented sharing bikes in Seoul per hour. And Functioning day is a binary variable, it indicates whether the Seoul Bike System is in operation or not.

There are 12 independent parameters which might affect the rent bike amount:

Time parameters;

Date: The dates on which the shared bikes were rented.

Hour: The hour on which the shared bikes were rented.

Parameters of weather (continuous numeric variable):

Temperature: The temperature in Celsius($^{\circ}\text{C}$) for the hour

Humidity: The humidity (%) for the hour

Wind Speed: The wind speed in meters per second (m/s)

Visibility: The distance at which objects are clearly visible under atmospheric conditions. (Unit: 10m)

Dew point temperature: The Celsius temperature($^{\circ}\text{C}$) at which air becomes saturated with water vapour and begins to condense.

Solar radiation: The solar radiation amount (MJ/m²)

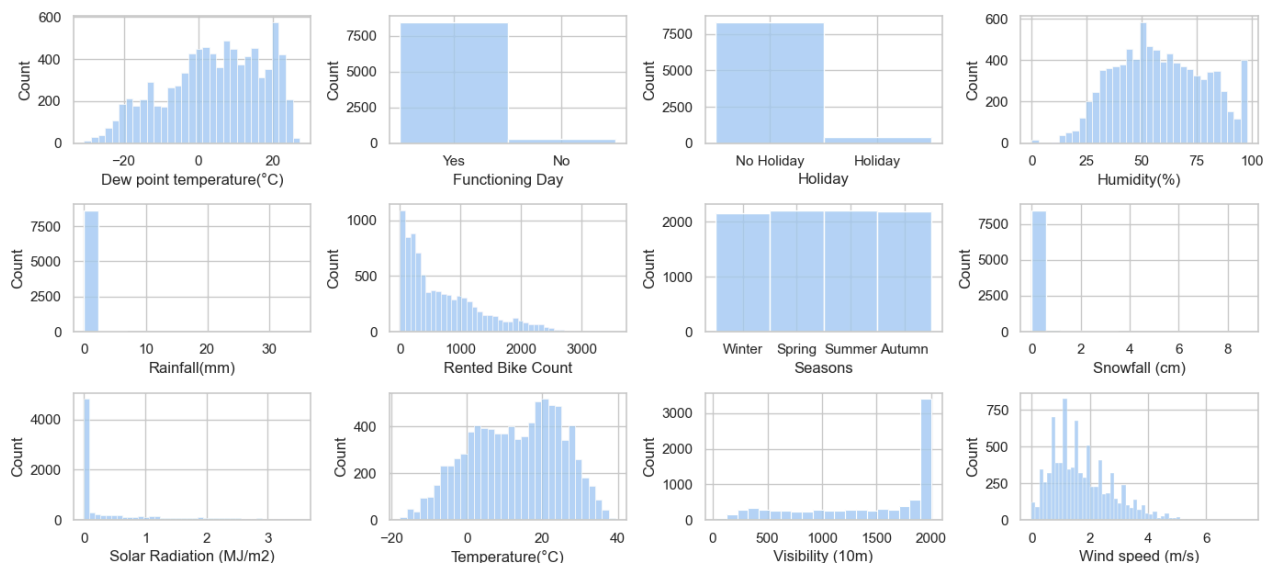
Rainfall: The rainfall amount (mm)

Snowfall: The snowfall amount (cm)

Categorical and binary variables:

Seasons: Categorical Spring: Summer, Autumn, Winter

Holiday: Binary variable: whether it is a holiday or not



Data Background:

This dataset is a business data set. It was collected by Dr. Sathishkumar VE from Sunway University, Malaysia in 2020. Data used include Seoul Bike and Capital Bikeshare Program data from 2017 to 2018. Raw data related to bike rent counts is downloaded from the official open data website of the City of Seoul. Datetime information, season, holiday and functional days are added through feature engineering. The weather information is downloaded from the Korea Meteorological Administration.

Since the original data is from the official website of the South Korean government, scientific processes that may affect the data could include the inaccurate measurement of weather data resulting from instrument calibration, data processing algorithms or natural reasons like atmospheric dynamics, solar influence and so on.

Scientific Questions:

We are interested in three questions regarding this dataset as follows:

1. Is it possible to perform clustering techniques to distinguish non-functional days and identify the weather conditions that will cause the cessation of the Seoul bike system?
2. Is it possible that the rented bike count numbers follow a certain probability distribution? How to examine our hypothesis?
3. How to construct machine learning models about how the weather conditions impact the number of bikes rented?

For question 1, since the sample of “non-functioning days” is much less than the sample of “functioning days”, we are considering ensemble classification methods such as boosting trees and random forests. For question 2, since the bike rented number can be considered an event number occurring within a fixed time, we suspect that it might fit a Poisson distribution.

In terms of question 3, we are expecting to use GLM or neural network to depict the relationship between rented bike counts and the weather and holiday information. However, further investigation awaits.

Appendix :

seoulBike.sample(10)														
	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
1423	29/01/2018	231	7	-11.1	37	1.6	2000	-22.9	0.00	0.0	0.0	Winter	No Holiday	Yes
5307	10/07/2018	136	3	20.6	98	0.8	1986	20.2	0.00	0.5	0.0	Summer	No Holiday	Yes
1382	27/01/2018	193	14	-4.1	25	2.0	1730	-21.2	1.20	0.0	0.2	Winter	No Holiday	Yes
2270	05/03/2018	550	14	10.0	44	1.9	2000	-1.7	2.46	0.0	0.0	Spring	No Holiday	Yes
8359	14/11/2018	1047	7	5.2	78	0.3	1182	1.6	0.00	0.0	0.0	Autumn	No Holiday	Yes
7201	27/09/2018	488	1	16.1	53	1.4	2000	6.5	0.00	0.0	0.0	Autumn	No Holiday	Yes
4082	20/05/2018	473	2	13.9	43	0.9	2000	1.5	0.00	0.0	0.0	Spring	No Holiday	Yes
3702	04/05/2018	521	6	9.0	57	1.7	1903	0.9	0.00	0.0	0.0	Spring	No Holiday	Yes
8284	11/11/2018	189	4	6.9	82	0.5	337	4.0	0.00	0.0	0.0	Autumn	No Holiday	Yes
5058	29/06/2018	3113	18	28.1	60	1.9	959	19.5	1.05	0.0	0.0	Summer	No Holiday	Yes