Seoul Bike Sharing

Midterm Presentation for 2023 Spring Cloud Computing

Kyuri Kim, HaeLee Kim, Nayaeun Kwon

SEOUL

A Complete Guide to Attractions in Seoul | The Official Travel Guide to Seoul (visitseoul.net)

is the capital city of South Korea

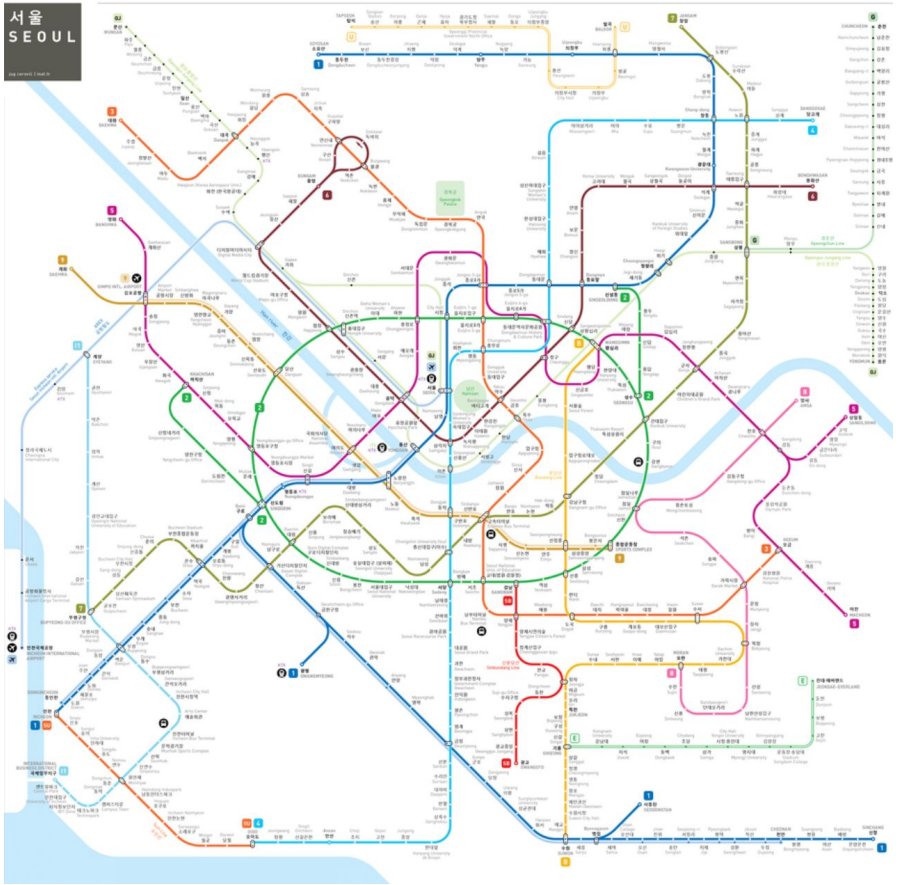
with a population of over 10 million people

having a rich history dating back over 2000 years

but also known for cutting-edge technology

full with fascinating cultural attractions.

You can go anywhere in Seoul with the public transportation system.



Bike Sharing System in Seoul bikeseoul.com

city government system

20,000+ bikes

24/7 operation

stations strategically located

web/mobile registration

GPS tracking

Project Introduction

- Rental bikes are introduced in Seoul for the enhancement of mobility comfort.
- It is important to make the rental bike available and accessible to the public at the right time as it lessens the waiting time.
- Providing the city with a stable supply of rental bikes becomes a major concern.
- The dataset contains weather information & the number of bikes rented per hour and date information.

Dataset Information

Number of Instances : 8760

Number of Features : 14

Types of Features : Integer, Float, and Character

Date Range : 12/1/2017 - 11/30/2018

Date Donated : 3/1/2020

Data Source URL: https://archive.ics.uci.edu/ml/datasets/Seoul+Bike+Sharing+Demand#

Original Datasets URL: http://data.seoul.go.kr/ and publicholidays.go.kr

The Implemented Features

Date : year-month-day

Hour : Hour of the day

Rented Bike count : Count of bikes rented at each hour

Temperature : Temperature in Celsius

Humidity : %

Windspeed : m/s

Visibility : 10m

Dew point temperature : Celsius

Solar radiation : MJ/m2

Rainfall : mm

Snowfall : cm

Seasons : Winter, Spring, Summer, Autumn

Holiday : Holiday / No holiday

Functional Day : NoFunc (Non Functional Hours), Fun (Functional hours)

Expected Outcomes

Goal

Predict the bike count required at each hour and season for the stable supply of rental bikes

Target

Rented bike count

Hypothesis 1

During rush hours, the bike rental count will be higher compared to non-rush hours.

Hypothesis 2

Bike rental count will decrease when the temperature drops below 0°C (32°F) or rises above 30°C (86°F).

ML Models

Regression Analysis, Time Series Analysis, Decision Tree Analysis

Logical architecture of AWS cloud services

AWS cloud services to be used are VPC & EC2.

Logical architecture

1. Create a VPC - Public Subnet
2. Launch EC2
3. Set up Git Repository
4. Collaborate on the code
5. Run the ML models

* Optional steps: Create individual IAM users and S3/EBS

The Data Flow

#	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(°C)	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(°C)	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

To prepare EDA and ML analysis,

Data will be cleaned up / processed by

- dropping nulls and outliers
- converting Seasons into factor

  (Spring : 1, Summer : 2, Fall : 3, Winter : 4)
- & Holiday and Functioning Day into binary

  (Holiday : 0, No Holiday 1)

  (Functional Day : 0. No Functional Day : 1)


Thank you!