

# Optimization of Dublin Bikes station-based bike sharing service

Ronan Downes -ID SBA22447

November 25, 2022

# Contents

<b>1</b>	<b>Exploratory Data Analysis</b>	<b>4</b>
1.1	Data Exploration . . . . .	4
1.2	Sampling the data . . . . .	4
1.3	Checking the types of data . . . . .	6
1.4	Data Preparation . . . . .	9
1.4.1	Variables Types and Definitions . . . . .	9
<b>2</b>	<b>Data</b>	<b>10</b>
2.1	Data Loading, Parsing, Storage, and File Formatting . . . . .	10
2.2	Exploration and pre-processing . . . . .	10
2.3	Weather Data . . . . .	11
<b>A</b>	<b>Downloading and consolidating the data-Dublin Bikes - Occupancy Download Data</b>	<b>12</b>

# List of Tables

1.1	Sample of the bikes raw data showing structure and features 1	5
1.2	Bike data types	7
1.3	Sample of the bikes raw data showing structure and features	8
2.1	Your first table.	10
2.2	Your first table.	11
2.3	Bike data	11

# List of Figures

# Chapter 1

## Exploratory Data Analysis

At this part of the data analysis the bikes and weather datasets are described. The meaning of variable is discussed and necessary adjustments are performed to ensure the data is processed as intended at the Machine Learning stage.

### 1.1 Data Exploration

**Variables Types and Definitions¶** The first stage of this analysis is to describe the dataset, understand the meaning of variable and perform the necessary adjustments to ensure that the data will be processed correctly during the Machine Learning process.

### 1.2 Sampling the data

Table 1.1: Sample of the bikes raw data showing structure and features 1

STATION ID		LAST UPDATED		NAME		BIKE STANDS		AVAILABLE BIKE STANDS		AVAILABLE BIKES		STATUS		ADDRESS		LATITUDE		LONGITUDE	
1510468	27	2019-02-16 00:52:48		MOLESWORTH STREET		20	20	0	0	Open		Open		Molesworth Street		53.341290		-6.258117	
2719839	6	2019-09-24 09:38:39		CHRISTCHURCH PLACE		20	18	2	2	Open		Open		Christchurch Place		53.343369		-6.270120	
2531101	32	2019-03-20 04:38:48		PEARSE STREET		30	30	0	0	Open		Open		Pearse Street		53.344303		-6.250427	
360567	23	2019-07-12 08:29:02		CUSTOM HOUSE		30	12	17	17	Open		Open		Custom House		53.348278		-6.254662	
1686586	74	2019-02-21 08:08:09		OLIVER BOND STREET		30	29	1	1	Open		Open		Oliver Bond Street		53.343891		-6.280531	
2305013	109	2019-09-10 13:09:44		BUCKINGHAM STREET LOWER		29	27	2	2	Open		Open		Buckingham Street Lower		53.353333		-6.249319	
58735	95	2019-07-02 22:28:48		ROYAL HOSPITAL		40	0	40	40	Open		Open		Royal Hospital		53.343899		-6.297060	
1806223	9	2019-11-27 21:59:20		EXCHEQUER STREET		24	22	1	1	Open		Open		Exchequer Street		53.343033		-6.263578	

### 1.3 Checking the types of data

Table 1.2: Checking the datatypes because pandas stores variables as strings or objects if data is missing. This needs to be repaired before plotting the data.

Variable	Data Type
STATION ID	int64
TIME	object
LAST UPDATED	object
NAME	object
BIKE STANDS	int64
AVAILABLE BIKE STANDS	int64
AVAILABLE BIKES	int64
STATUS	object
ADDRESS	object
LATITUDE	float64
LONGITUDE	float64
dtype: object	



Table 1.3: Sample of the bikes raw data showing structure and features

	STATION ID	LAST UPDATED	NAME	BIKE STANDS	AVAILABLE BIKE STANDS	AVAILABLE BIKES
1510468	27	2019-02-16 00:52:48	MOLESWORTH STREET	20	20	0
2719839	6	2019-09-24 09:38:39	CHRISTCHURCH PLACE	20	18	2
2531101	32	2019-03-20 04:38:48	PEARSE STREET	30	30	0
360567	23	2019-07-12 08:29:02	CUSTOM HOUSE	30	12	17
1686586	74	2019-02-21 08:08:09	OLIVER BOND STREET	30	29	1
2305013	109	2019-09-10 13:09:44	BUCKINGHAM STREET LOWER	29	27	2
58735	95	2019-07-02 22:28:48	ROYAL HOSPITAL	40	0	40
1806223	9	2019-11-27 21:59:20	EXCHEQUER STREET	24	22	1

## **1.4 Data Preparation**

### **1.4.1 Variables Types and Definitions**

# Chapter 2

## Data

### 2.1 Data Loading, Parsing, Storage, and File Formatting

With portability in mind retrieval functions are defined for each of the four quarters in 2019 and a "data directory" is generated and populated with quarterly csv files. Each is read as a pandas DataFrame and concatenated to csv as dublinbikes\_2019.csv for later analysis.

In order to avoid the March 2020 lockdown for Covid-19 the 4 quarters of 2019 are considered. The Download Data notebook downloads and consolidates all four 2019 datasets from the Dublin-Bike into a single parquet file. The dataset is freely provided by the Smart Dublin website. It contains the GPS traces of the Dublin buses for January 2013.

### 2.2 Exploration and pre-processing

Supporting end-users travel requirements not only includes the allocation of available bikes at the start of their journey but also requires availability of empty bike stands to allow the user to end their journey at the station of their choice (Mehdizadeh Dastjerdi and Morency 2022).

Executing `df_2019.dtypes` gives the data in table 2.1. TIME and LAST UPDATED can be converted to leaner types. Pandas can easily convert data from one type to another.

Table 2.1: Your first table.

STATION ID	int64
TIME	object
LAST UPDATED	object
NAME	object
BIKE STANDS	int64
AVAILABLE BIKE STANDS	int64
AVAILABLE BIKES	int64
STATUS	object
ADDRESS	object
LATITUDE	float64
LONGITUDE	float64
dtype: object	

`db2019.df.shape` gives (10741877, 11) which means that there are 10741877 rows across 11 independent variables.

```
#remove reduntant rows without update
df_2019 = df_2019.drop(['TIME'], axis = 1)
df_2019.drop_duplicates(keep= 'first', inplace=True)
df_2019.shape
```

Table 2.2: Your first table.		
Value 1	Value 2	Value 3
$\alpha$	$\beta$	$\gamma$
1	1110.1	a
2	10.1	b
3	23.113231	c

Table 2.3: Bike data					
		date	ind	rain	
6RG(6165,245,0.2		6RG(6165,245,0.2	6RG(6165,245,0.2	6RG(6165,245,0.2	6RG(6165,245,0.2
	90881	27-dec-2013 18:00	0	0.4	
	117194	28-dec-2016 03:00	0	0.0	
	48281	16-feb-2009 18:00	0	0.0	
	112220	03-jun-2016 21:00	0	0.0	
	82067	25-dec-2012 12:00	0	0.0	
	126520	20-jan-2018 17:00	0	0.1	
	9537	16-sep-2004 10:00	-1		
	127456	28-feb-2018 17:00	0	0.4	
	42394	16-jun-2008 11:00	0	0.0	
	4023	30-jan-2004 16:00	0	0.1	

2.3 Weather Data

## Appendix A

# Downloading and consolidating the data-Dublin Bikes - Occupancy Download Data

```
import pandas as pd
import numpy as np
import warnings
import time
import datetime as dt
warnings.filterwarnings('ignore')
import matplotlib.pyplot as plt
import urllib
import folium
#import sklearn
import seaborn as sns
import os

#from sklearn.model_selection import train_test_split
from matplotlib import pyplot as plt
#from sklearn.cluster import KMeans
# from matplotlib import inline
from ipywidgets import interact, interact_manual

if not os.path.exists("data"):
    os.makedirs("data")

dataframes = [
    (
        "data/dublinbikes_20190101_20190401.csv",
        "https://data.smartdublin.ie/dataset/33ec9fe2-4957-4e9a-ab55-c5e917c7a9ab/resource/538165d7-
    ),
    (
        "data/dublinbikes_20190401_20190701.csv",
        "https://data.smartdublin.ie/dataset/33ec9fe2-4957-4e9a-ab55-c5e917c7a9ab/resource/76fdda3d-
    ),
    (
        "data/dublinbikes_20190701_20191001.csv",
        "https://data.smartdublin.ie/dataset/33ec9fe2-4957-4e9a-ab55-c5e917c7a9ab/resource/305d39ac-
    ),
    (
        "data/dublinbikes_20191001_20200101.csv",
        "https://data.smartdublin.ie/dataset/33ec9fe2-4957-4e9a-ab55-c5e917c7a9ab/resource/5d23332e-
```

],

# Bibliography

Mehdizadeh Dastjerdi, Aliasghar and Catherine Morency (2022). “Bike-Sharing Demand Prediction at Community Level under COVID-19 Using Deep Learning”. In: *Sensors* 22.3, p. 1060.