LONDON BIKES DATA ANALYSIS REPORT

About Dataset:

Link- https://www.kaggle.com/datasets/hmavrodiev/london-bike-sharing-dataset

Context- The purpose is to try predict the future bike shares.

Content- The data is acquired from 3 sources:

- Https://cycling.data.tfl.gov.uk/ 'Contains OS data © Crown copyright and database rights 2016' and Geomni UK Map data © and database rights [2019] 'Powered by TfL Open Data'
- freemeteo.com - weather data
- https://www.gov.uk/bank-holidays From 1/1/2015 to 31/12/2016]

The data from cycling dataset is grouped by "Start time", this represent the count of new bike shares grouped by hour. The long duration shares are not taken in the count. Metadata:

"timestamp" - timestamp field for grouping the data
"cnt" - the count of a new bike shares
"t1" - real temperature in C
"t2" - temperature in C "feels like"
"hum" - humidity in percentage
"wind_speed" - wind speed in km/h
"weather_code" - category of the weather
"is_holiday" - boolean field - 1 holiday / 0 non holiday
"is_weekend" - boolean field - 1 if the day is weekend
"season" - category field meteorological seasons: 0-spring ; 1-summer; 2-fall; 3-winter.

"Weather_code" category description: 1 = Clear ; mostly clear but have some values with haze/fog/patches of fog/ fog in vicinity 2 = scattered clouds / few clouds 3 = Broken clouds 4 = Cloudy 7 = Rain/ light Rain shower/ Light rain 10 = rain with thunderstorm 26 = snowfall 94 = Freezing Fog.

RAW DATA

Shape- (17414 Rows, 10 Columns)

	A	B	C	D	E	F	G	H	I	J	K
1	timestamp	cnt	t1	t2	hum	wind_speed	weather_co	is_holiday	is_weekend	season	
2	1/4/2015 0:00	182	3	2	93	6	3	0	1	3	
3	1/4/2015 1:00	138	3	2.5	93	5	1	0	1	3	
4	1/4/2015 2:00	134	2.5	2.5	96.5	0	1	0	1	3	
5	1/4/2015 3:00	72	2	2	100	0	1	0	1	3	
6	1/4/2015 4:00	47	2	0	93	6.5	1	0	1	3	
7	1/4/2015 5:00	46	2	2	93	4	1	0	1	3	
8	1/4/2015 6:00	51	1	-1	100	7	4	0	1	3	
9	1/4/2015 7:00	75	1	-1	100	7	4	0	1	3	
10	1/4/2015 8:00	131	1.5	-1	96.5	8	4	0	1	3	
11	1/4/2015 9:00	301	2	-0.5	100	9	3	0	1	3	
12	1/4/2015 10:00	528	3	-0.5	93	12	3	0	1	3	
13	1/4/2015 11:00	727	2	-1.5	100	12	3	0	1	3	
14	1/4/2015 12:00	862	2	-1.5	96.5	13	4	0	1	3	
15	1/4/2015 13:00	916	3	-0.5	87	15	3	0	1	3	
16	1/4/2015 14:00	1039	2.5	0	90	8	3	0	1	3	
17	1/4/2015 15:00	869	2	-1.5	93	11	3	0	1	3	
18	1/4/2015 16:00	737	3	0	93	12	3	0	1	3	
19	1/4/2015 17:00	594	3	0	93	11	3	0	1	3	
20	1/4/2015 18:00	522	3	1.5	93	6.5	3	0	1	3	
21	1/4/2015 19:00	379	3	1	93	7	3	0	1	3	
22	1/4/2015 20:00	328	3	3	93	4	3	0	1	3	
23	1/4/2015 21:00	221	3	2.5	93	5	4	0	1	3	
24	1/4/2015 22:00	178	3	2	93	6	4	0	1	3	
25	1/4/2015 23:00	157	4	3.5	87	5	4	0	1	3	
26	1/5/2015 0:00	83	4	3	93	6	4	0	0	3	
27	1/5/2015 1:00	67	4	3.5	93	5	4	0	0	3	
28	1/5/2015 2:00	32	5	4	87	6	4	0	0	3	
29	1/5/2015 3:00	22	6	4.5	84	7.5	4	0	0	3	

CLEANING AND PROCESSING DATA

Link :https://colab.research.google.com/drive/1polbt5RXcVpq7Wc5LGsTg3NBELaGiAFs#scrollTo=8E3Wof0X7bCv

Modules used- Pandas ,Zipfile

```
# specifying the column names that I want to use
new_cols_dict ={
    'timestamp':'time',
    'cnt':'count',
    't1':'temp_real_C',
    't2':'temp_feels_like_C',
    'hum':'humidity_percent',
    'wind_speed':'wind_speed_kph',
    'weather_code':'weather',
    'is_holiday':'is_holiday',
    'is_weekend':'is_weekend',
    'season':'season'
}

# Renaming the columns to the specified column names
bikes.rename(new_cols_dict, axis=1, inplace=True)
```

```
# changing the humidity values to percentage (i.e. a value between 0 and 1)
bikes.humidity_percent = bikes.humidity_percent / 100
```

```
# creating a season dictionary so that we can map the integers 0-3 to the actual written values
season_dict = {
    '0.0':'spring',
    '1.0':'summer',
    '2.0':'autumn',
    '3.0':'winter'
}
```

The above provided code snippet outlines a comprehensive set of data cleaning and pre processing tasks for the London Bike Dataset. Let's delve into a more detailed explanation:

Column Renaming: The dataset undergoes a refinement of column names for clarity and expressiveness. The renaming is executed through a dictionary (new_cols_dict), establishing a correspondence between original and desired column names.

Season and Weather Mapping: The script constructs a mapping mechanism for the "season" column, translating numerical representations (0-3) into their corresponding textual equivalents ("spring," "summer," "autumn," "winter") via the season_dict. Similarly, a weather_dict is introduced to encapsulate the translation of numerical codes in the "weather" column to descriptive weather conditions.

```
# creating a weather dictionary so that we can map the integers to the actual written values
weather_dict = {
    '1.0':'Clear',
    '2.0':'Scattered clouds',
    '3.0':'Broken clouds',
    '4.0':'Cloudy',
    '7.0':'Rain',
    '10.0':'Rain with thunderstorm',
    '26.0':'Snowfall'
}

# changing the seasons column data type to string
bikes.season = bikes.season.astype('str')
# mapping the values 0-3 to the actual written seasons
bikes.season = bikes.season.map(season_dict)

# changing the weather column data type to string
bikes.weather = bikes.weather.astype('str')
# mapping the values to the actual written weathers
bikes.weather = bikes.weather.map(weather_dict)
```

Humidity Percentage Conversion: The "humidity_percent" column undergoes a normalization process, with values converted from a percentage scale (0-100) to a fraction scale (0-1) by dividing each value by 100.

Data Type Conversion: The data types of the "season" and "weather" columns are transformed to string, aligning them with their categorical nature and facilitating subsequent mapping operations.

Mapping of Categorical Values: Leveraging the dictionaries created earlier, the script strategically employs the map function to associate numerical representations in the "season" and "weather" columns with their respective human-readable labels. This enhances the interpretability of these categorical variables.

Excel File Export: The culminating dataset, now refined and enriched, is exported to an Excel file named 'london_bikes_final.xlsx.' The 'Data' sheet within this file becomes a repository of the meticulously processed dataset, poised for seamless integration into Tableau for insightful visualizations.

In essence, these pre processing steps transcend mere cosmetic enhancements. They empower the dataset with semantic richness, rendering it more amenable to downstream analysis and visualization endeavors, particularly within the context of Tableau. The code encapsulates a holistic approach to data refinement, ensuring that the dataset not only looks polished but is also imbued with enhanced interpretability and analytical potential.

Transformed Data

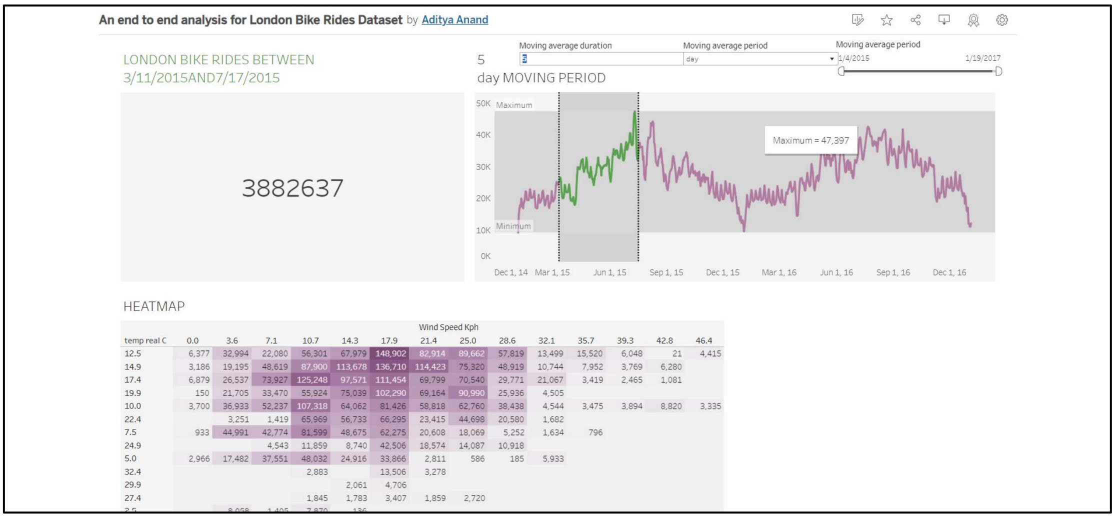
(Link: https://github.com/adiianand03/My-Resume-Projects2/blob/End-to-end-Data-Analysis-with-Interactive-Dashboards-for-London-Bike-Dataset/london_bikes_final.xlsx)

	time	count	temp_real_C	temp_feels_like_C	humidity_percent	wind_speed_kph	weather	is_holiday	is_weekend	season
0	2015-01-04 00:00:00	182	3.0	2.0	0.930	6.0	Broken clouds	0.0	1.0	winter
1	2015-01-04 01:00:00	138	3.0	2.5	0.930	5.0	Clear	0.0	1.0	winter
2	2015-01-04 02:00:00	134	2.5	2.5	0.965	0.0	Clear	0.0	1.0	winter
3	2015-01-04 03:00:00	72	2.0	2.0	1.000	0.0	Clear	0.0	1.0	winter
4	2015-01-04 04:00:00	47	2.0	0.0	0.930	6.5	Clear	0.0	1.0	winter

ISSUES TO BE VISUALIZED

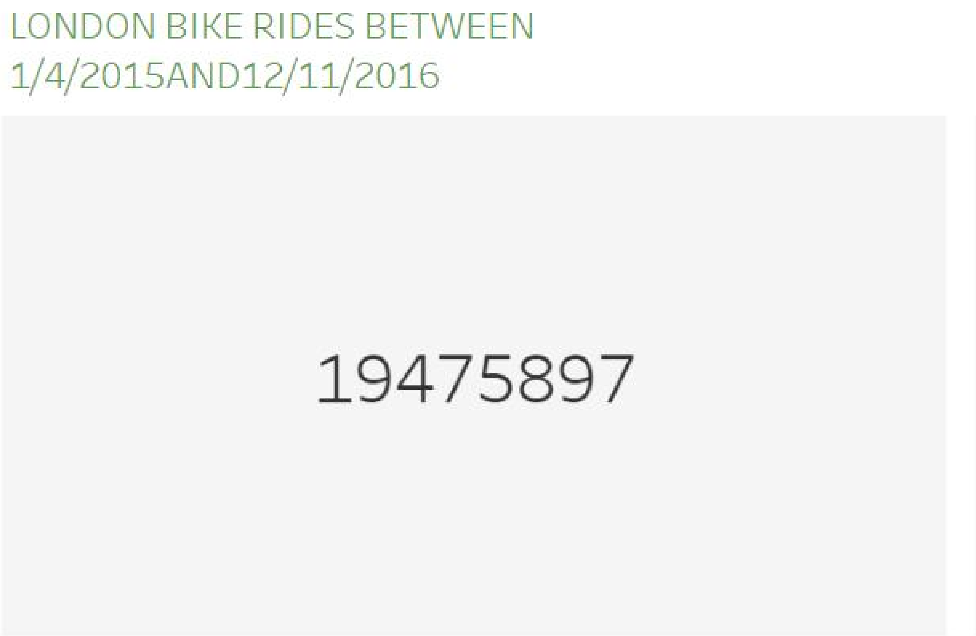
- Total number of bike users between a certain time period
- Effects of factors such as temperature and Wind Speed on number of customers
- To show the number of customers on a particular day with the weather conditions and filtered by the number of hours used

TABLEAU VISUALIZATION ([Tableau Visualization direct link]())

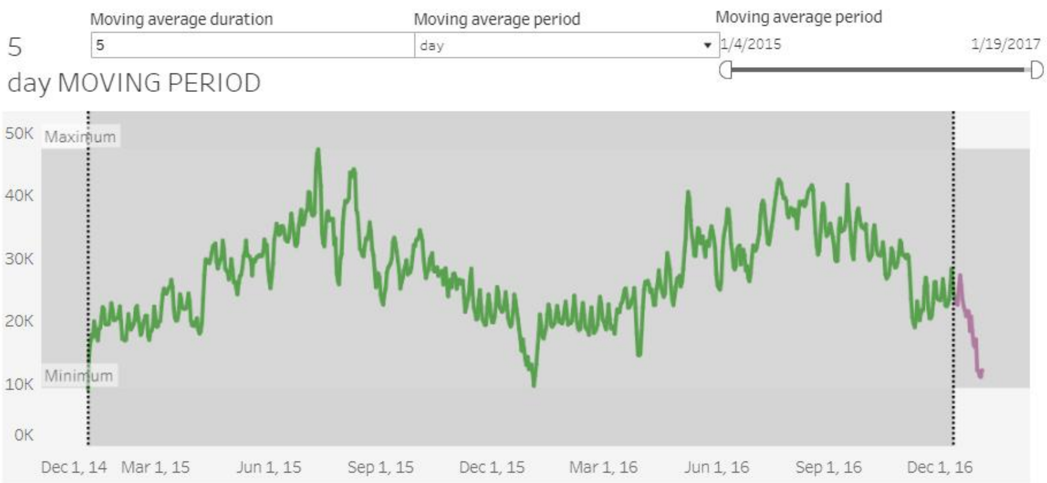


Dashboard

DASHBOARD COMPONENTS



Total Users between a certain period



Interactive slider component to adjust the time period with an option to choose the average duration and average period as day, week or month

HEATMAP

	Wind Speed Kph														
temp real C	0.0	3.6	7.1	10.7	14.3	17.9	21.4	25.0	28.6	32.1	35.7	39.3	42.8	46.4	50.0	53.6
12.5	38,329	243,483	316,093	484,275	431,665	569,516	242,265	217,725	164,249	72,636	45,625	17,407	3,782	7,453		
19.9	7,388	129,062	214,658	344,471	309,814	522,831	316,058	265,246	79,149	15,239	6,525					
17.4	25,121	152,956	315,759	507,545	405,156	499,875	317,845	227,048	82,936	32,888	9,285	2,465	1,081			
10.0	46,608	298,385	289,148	485,291	378,838	391,221	238,427	229,930	143,285	57,150	42,182	12,988	13,244	3,960		
7.5	33,943	264,216	190,398	315,003	250,060	355,676	173,669	164,153	82,539	21,711	12,227	3,289	8,210	1,794	1,180	556
22.4	4,664	76,132	65,674	190,170	177,028	276,143	105,367	144,232	59,643	6,022						
5.0	29,640	168,184	216,760	311,854	239,244	221,395	134,281	83,865	56,857	26,342	12,642		1,874			65
24.9		15,761	44,068	85,219	87,122	121,557	67,528	38,495	20,600	10,289						
2.5	23,380	98,933	107,158	149,321	51,956	89,782	23,377	18,016	9,866	208	2,427					
29.9		5,471		6,605	15,259	28,016	10,530	5,046								
27.4	2,036	17,863	21,722	28,057	16,840	25,198	15,443	5,895								
32.4		2,727	4,984	2,883		21,703	3,278									
0.0	18,238	61,694	35,012	48,408	22,671	7,259	3,453									
-2.5	2,072	12,755	6,174	1,083	77											

Heatmap of total customers as a comparison of Wind Speed and Temperature as weather factors influencing the use.

CONCLUSION

In transforming the London Bike Dataset for impactful visualization, a meticulous data preparation journey unfolded. Initially, columns were intuitively renamed for clarity. Categorical variables, like "season" and "weather," were artfully mapped to meaningful labels. Humidity percentages were normalized, and data types were re-defined for seamless interpretation. Following this, a sophisticated Tableau dashboard emerged, encapsulating vital insights. Total users within a specified time-frame became instantly visible and accessible. An interactive slider facilitated dynamic period adjustments, offering users the flexibility to choose average duration and periods—be it day, week, or month. The crowning achievement materialized in a dynamic heat map, expertly showcasing the total customers against the factors of wind speed and temperature as variables. This visionary depiction brilliantly illuminated the relationship between weather factors and bike usage, providing a user-centric narrative that goes beyond mere data representation.
This dynamic visualization, at the heart of the dashboard, transforms raw data into a compelling story, empowering stakeholders with actionable insights and fostering a deeper comprehension of the intricate dynamics at play.