**What is the subject and context of the project?**

The project aims to analyze the availability and usage patterns of the Vélib' bike-sharing system in Paris. The context involves understanding urban mobility, promoting sustainable transportation, and planning for urban infrastructure improvements

**What is the source of the data?**

The data is sourced from the Paris Open Data portal, which provides real-time and historical data on bike availability and station status for the Vélib' system.

**Who collected the data?**

The data is collected by the Vélib' Métropole service, operated by Smovengo under a contract with the Métropole du Grand Paris.

**When was the data collected ?**

**In what context/for what purpose was the data collected?**

The data is collected to monitor and manage the bike-sharing system efficiently, providing users with real-time information on bike and station availability and helping administrators manage the distribution and maintenance of bikes.

**What does the dataset contain? Describe the data, and its type (continuous, categorical...)**

The dataset contains information on each bike station, including identifiers, names, operational status, bike availability (total, mechanical, and electric), docking station availability, and geographical coordinates.

The Vélib' dataset includes various types of data which can be categorized as follows:

- Categorical Data: These are variables that have a finite number of categories or distinct groups.

- Station Name (Nom station): Names of the bike stations.

- Operational Status (Station en fonctionnement): Indicates if the station is operational ('OUI' for yes, 'NON' for no).

- Payment Terminal Availability (Borne de paiement disponible): Indicates if a payment terminal is available at the station ('OUI' for yes, 'NON' for no).

- Return Capability (Retour vélib possible): Indicates if bikes can be returned at this station ('OUI' for yes, 'NON' for no).

- Communes Equipped (Nom communes équipées): Names of the communes where stations are located.

- Numerical Data: These variables are measured on a numeric scale and can be either discrete or continuous.

- Station ID (Identifiant station): A unique identifier for each station.

- Station Capacity (Capacité de la station): The total number of bikes that the station can accommodate.

- Number of Free Docks (Nombre bornettes libres): The number of bike docks available at the station.

- Total Bikes Available (Nombre total vélos disponibles): Total number of bikes available at the station.

- Mechanical Bikes Available (Vélos mécaniques disponibles): Number of mechanical bikes available.

- Electric Bikes Available (Vélos électriques disponibles): Number of electric bikes available.

- Geographical Data: These data types include information about geographical locations.

- Geographical Coordinates\*(Coordonnées géographiques): The latitude and longitude coordinates of the stations.

- Data Update Time (Actualisation de la donnée): Timestamp indicating the last time the data for a station was updated, including date and time.

**What is the dataset's license?**

Typically, data provided by the Paris Open Data portal is under an open license, such as the Open Database License (ODbL), which allows free use and redistribution with attribution.

**Research question**

**How does the availability of different kinds of bikes, terminals, and services availability (return and payment) vary across different areas?**

* The availability of bikes and terminals in the Vélib' Métropole service is influenced by a variety of factors, based on the real-time data provided on their platform:
* Installation Status: Whether a station is actively installed affects availability.
* Renting and Returning Status: Stations must be operational for renting and returning bikes.
* Station Capacity and Usage: The number of docks available and the frequency of bike check-outs and returns can cause fluctuations in availability.
* Time of Day and Day of the Week: Usage patterns during peak hours and on weekends can impact availability.
* Maintenance and Repairs: Regular maintenance or unexpected repairs can temporarily reduce the number of available bikes or terminals.
* Weather Conditions: Adverse weather can affect user demand and thus bike availability.
* Special Events and Strikes: Local events or public transport strikes can lead to higher demand for bikes

1. **sub-questions:**
2. **How does the availability of bikes vary between electric and mechanical bikes at different stations?**

This question seeks to understand the distribution and preference for bike types among users across different areas. This can be answered by making a comparison of availability between electric and mechanical bikes at different stations.

**The top 5 stations with the highest number of available mechanical bikes are**:

Parc André Citroën (55 bikes)

Parvis Corentin Celton (53 bikes)

Grenelle - Dr Finlay (52 bikes)

Regnault - Patay (50 bikes)

François Truffaut - Saint Emilion (49 bikes)

**The same top 5 stations also have electric bikes available but in smaller quantities:**

Parc André Citroën (6 bikes)

Parvis Corentin Celton (2 bikes)

Grenelle - Dr Finlay (4 bikes)

Regnault - Patay (3 bikes)

François Truffaut - Saint Emilion (2 bikes)

From this analysis, it is evident that mechanical bikes are more abundantly available than electric bikes at these top stations. This could suggest a higher demand or a strategic allocation favoring mechanical bikes.

1. **What are the peak times for bike rentals and returns at high-capacity stations?**

This question aims to analyze the flow of bike usage during different times of the day, especially in stations with high capacity, to see if there are patterns or peak usage times.

The primary peak hour for bike activity is at 17:00 (5 PM UTC), showing significantly higher data updates compared to other hours. This suggests that late afternoon may be a busy time for the Vélib' service, possibly due to commuters returning home or tourists moving around the city in the late afternoon

1. **How can predictive analytics be used to optimize bike redistribution across the Vélib' Métropole network to meet user demand more effectively?**

This approach offers a systematic way to harness data and predictive analytics to enhance the management of bike-sharing systems, ultimately leading to more data-driven decision-making. The goal is to use historical bike usage data to predict future demand patterns at various stations throughout the network. This prediction can help in better allocating bikes across stations, ensuring that supply closely matches demand at any given time and location

.