ANALYSING SMART CITY BIKE SHARING DATA USING POWER BI

1. Project Overview and Objective

This project analyses real-time public bike-sharing data to understand station performance, bike availability, and usage patterns across different cities. Using Power BI, the data is cleaned, transformed, and visualized to generate actionable insights for improving operational efficiency and urban mobility planning.

Objectives:

- Analyse bike availability and station utilization
- Identify high-demand and underperforming stations
- Support data-driven decision-making using interactive dashboards

---

2. Data Sources

- **Source Description and Timeline:** Public GitHub repository
- **Domain**: Urban Mobility and Transportation Analytics.

---

3. Problem Statement

- Public bike-sharing systems generate continuous data from hundreds of stations across different cities.
- The challenge is to analyse this real-time bike station data to understand station performance, usage efficiency, and operational patterns.
- By applying Power BI techniques—including data cleaning, data modelling, DAX, and dashboard development.
- The project aims to transform raw data into meaningful insights that support better urban mobility planning and decision-making.

---

4. Attribute

Attribute Name	Data Type	Description
Number	Whole Number	Unique identification number assigned to each bike station
Name	Text	Name of the bike station
Address	Text	Physical address of the bike station
Position	Text	Geographic coordinates (Latitude, Longitude) of the station
Banking	True/False	Indicates whether card payment is supported at the station
Bonus	True/False	Indicates whether the station offers bonus services
Status	Text	Current operational status of the station (Open/Closed)
Contract Name	Text	City or contract area where the station is located
Bike Stand	Whole Number	Total number of docking stands at the station
Available Bike Stand	Whole Number	Number of empty stands available
Available Bikes	Whole Number	Number of bikes currently available
Last Update	Date/Time/Time zone	Timestamp of the last data update

---

5. Tools & Technologies

- **Power BI Desktop** – For data import, transformation, modelling, and visualization
- **Power Query –** Used to clean and preprocess raw data

---

6. Data Pre-Processing

Tasks Performed:

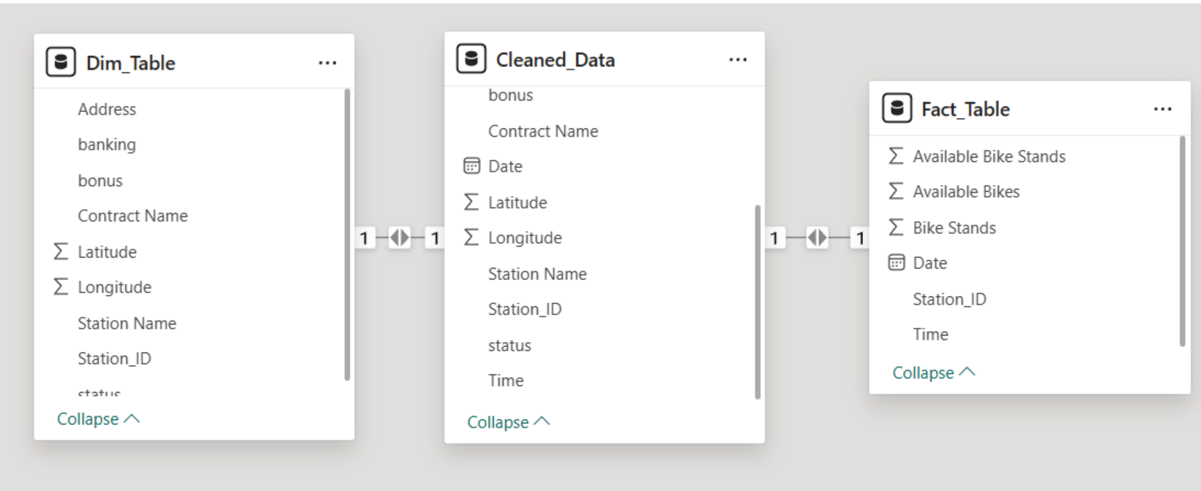
- **Data Cleaning & Transformation:** Used **Power Query in Power BI** to remove duplicates, handle missing values, standardize formats, and split the position column into Latitude and Longitude.
- **Time Zone Handling:** The original dataset included a time zone offset column (+05:30). Since all timestamps are consistently in **Indian Standard Time (UTC +05:30)**, this column was **redundant** and has been removed. All date-time values are considered **local IST time** for the analysis.
- **Filtering & Sorting:** Applied filters and sorting to focus on relevant records and improve data accuracy.
- Convert it into fact and dimension table.

---

7. Data Modelling and DAX

Data Model:

Data modelling organizes your dataset into **Fact and Dimension tables** to make analysis easier, faster, and more accurate in Power BI.



Star Schema

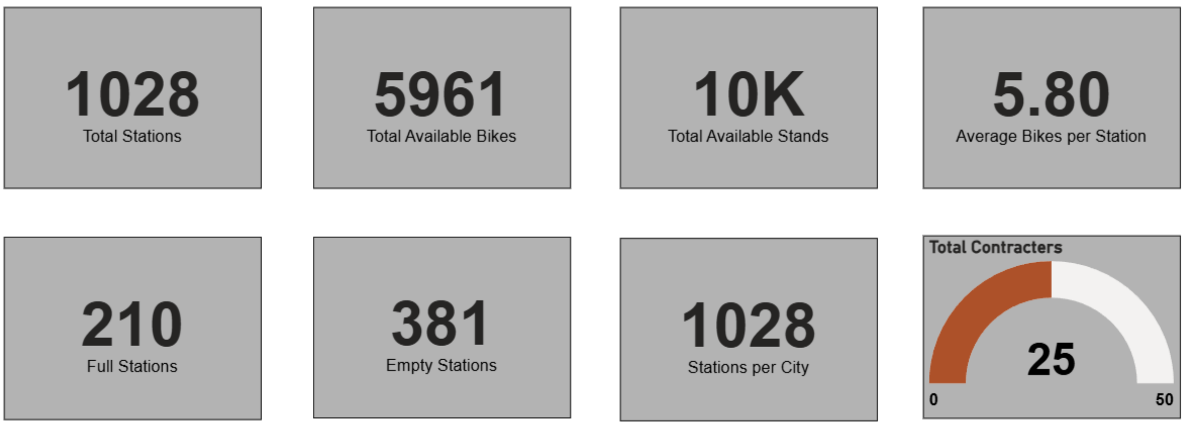
Calculated Columns & DAX Measures:

Implemented DAX formulas for key metrics, such as

1. Average Bikes per Station = AVERAGE (Fact_ Table [Available Bikes])
2. Empty Stations =
   CALCULATE (COUNTROWS (Fact _Table), Fact _Table [Available Bikes] = 0)
3. Full Stations =
   CALCULATE (COUNTROWS (Fact_ Table), Fact_ Table [Available Bike Stands] = 0)
4. Stations per City = DISTINCTCOUNT (Fact_ Table [Station_ ID])
5. Total Available Bikes = SUM (Fact_ Table [Available Bikes])
6. Total Available Stands = SUM (Fact_ Table [Available Bike Stands])
7. Total Stations = DISTINCTCOUNT (Fact_ Table [Station_ ID])

8.Analysis and Visualizations

Card: Total Stations, Contract , Available Bikes, Available Stands, Average Bikes per Station, Full Stations, Empty Stations, Stations per City.



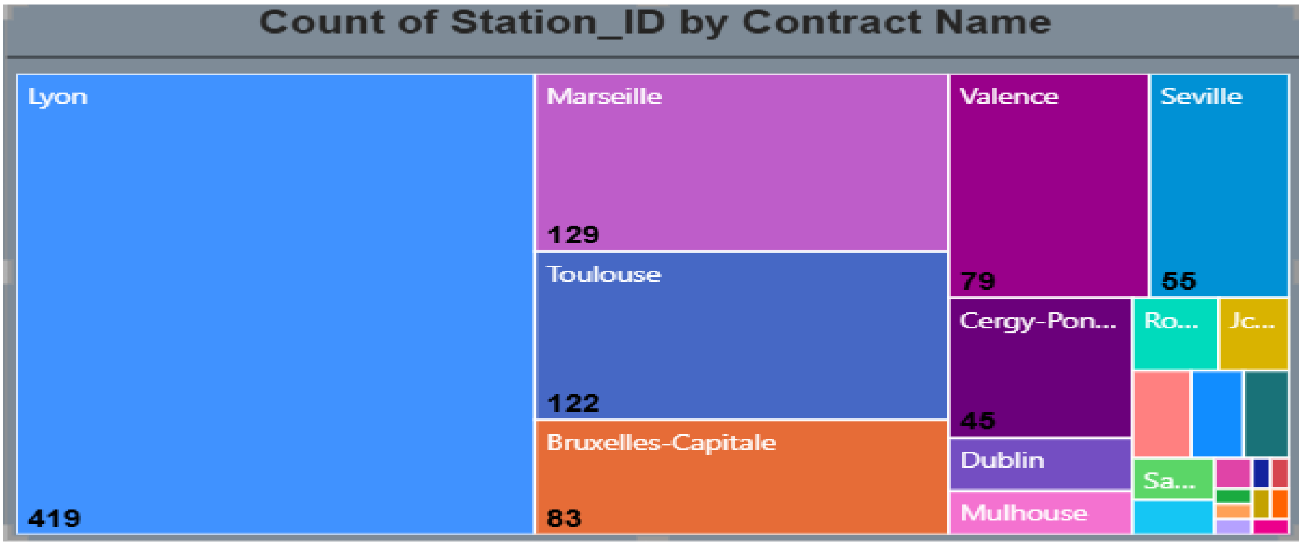
Highest Available Bikes Across Station ID:

Highest Available Bikes	Station_ID	Station Name
54	10002	Insa
40	112	Manuel Broseta I Pont - Naranjos
39	117	Hanover Quay East
39	7002	Universités Lyon Iii / Lyon Ii
38	188	Hospital Nueva Fe (Consultas Externas)
38	7035	Marseille / Université
36	100	Gare De Luxembourg / Luxemburgstation
35	10056	11 Novembre / Gaston Berger
34	10005	11 Novembre
33	10101	Einstein / 11 Novembre 1918
32	6043	Cité Internationale / Cinéma
31	3010	Partdieu / Cuirassiers (Far)
31	3019	Partdieu / Deruelle
30	110	Upv Trinquet
30	2021	Rambaud / Ecole Du Cinéma
30	7019	Croix Barret / Bourdeix
30	7033	Saintluc / Saintjoseph

Insights:

- Certain stations consistently record the highest number of available bikes, indicating lower demand or overcapacity at those locations
- **Station ID 10002** recorded the highest number of available bikes (54), followed by **Station ID 112** with 40 bikes, and **Station IDs 117 and 7002** with 39 bikes each.

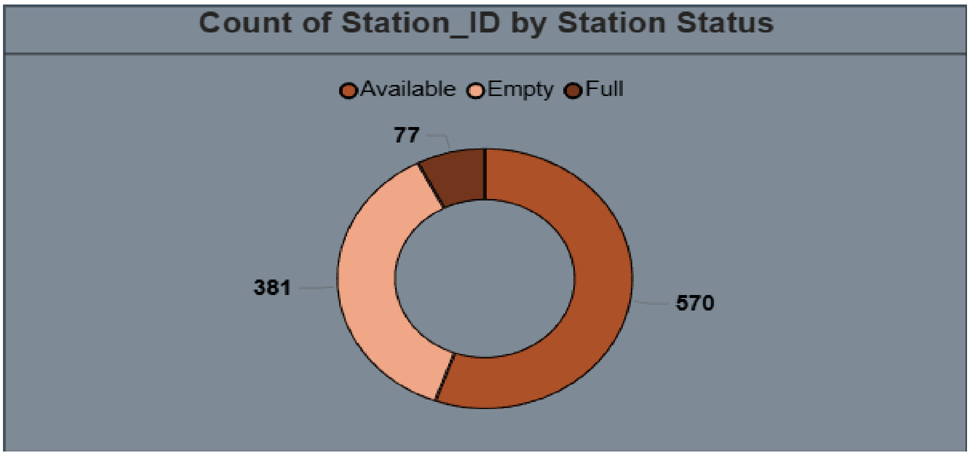
Contract Name–wise Station Distribution:



Insights:

- The number of stations varies significantly across different contract cities, indicating uneven infrastructure deployment.
- Cities with a **higher station count** demonstrate stronger bike-sharing network coverage and higher service accessibility. Contracts with **fewer stations** may indicate limited operational scale or lower demand.
- **Lyon** has the highest number of stations with 419, followed **by Marseille** with 129 stations and **Toulouse** with 122 stations.

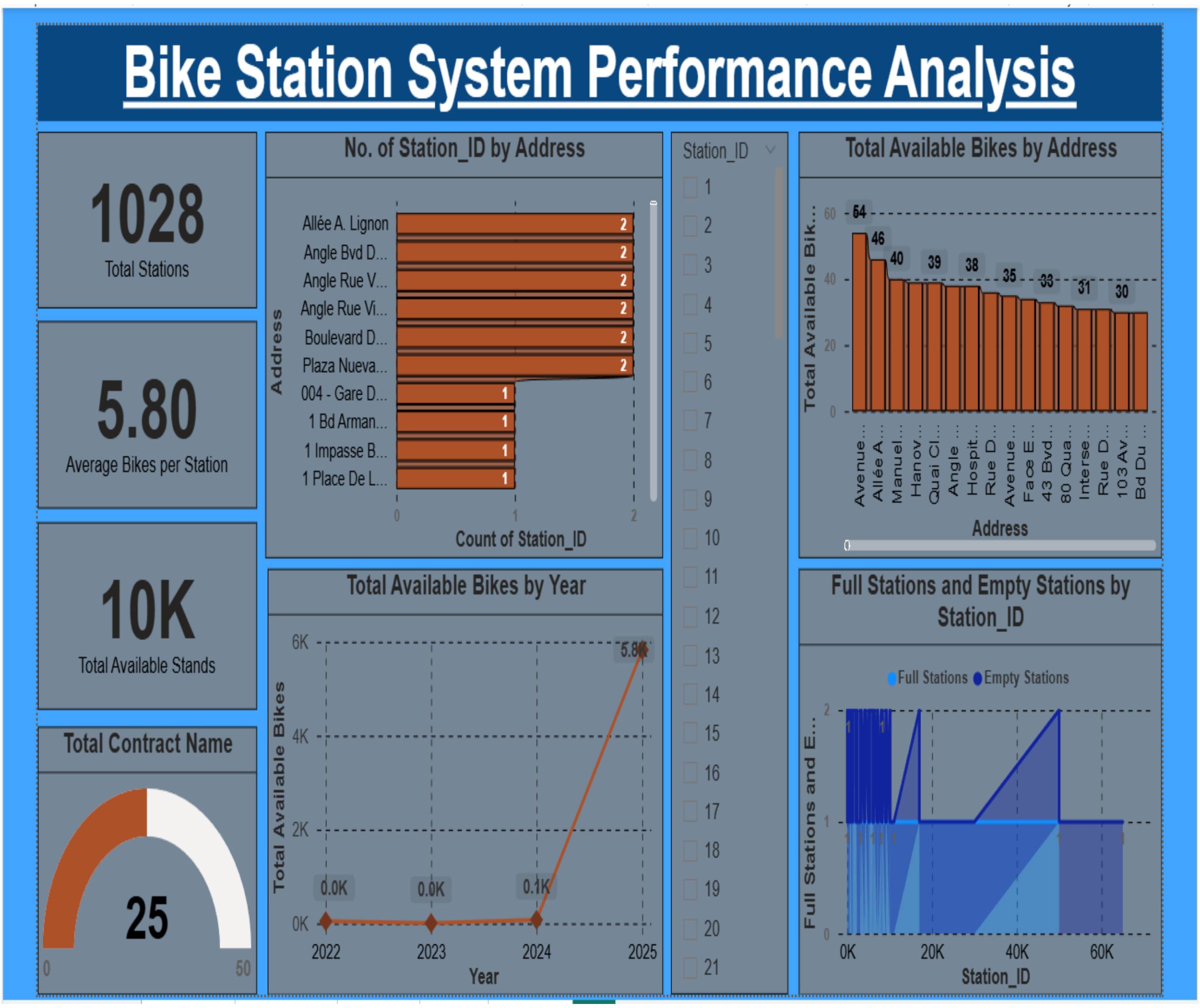
Count of Station ID by Status:

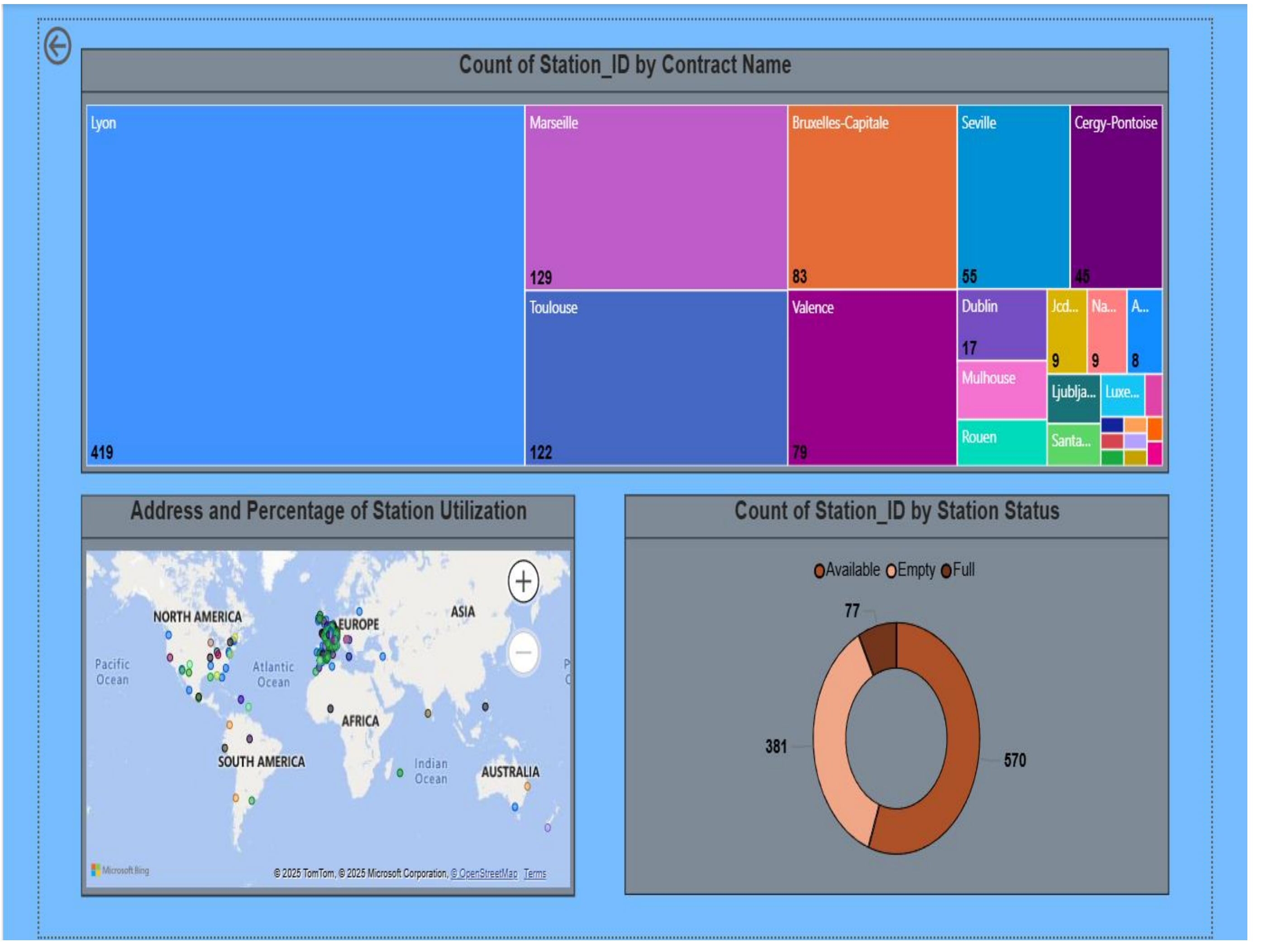


Insights:

- A majority of stations are **available** (570), indicating good overall bike availability across the network.
- 381 stations are **empty**, suggesting periods or locations with high demand where bikes are frequently unavailable.
- Only 77 stations are **full**, showing that fewer stations experience complete saturation of bike stands.
- This distribution highlights a **demand–supply imbalance**, where bike availability varies significantly across locations, emphasizing the need for better redistribution strategies.

Report:





9. Insights

Key Findings:

- **Total Station:** 1028 | **Total Contract Name:** 25
- **Total Available Bike:** 5961 | **Total Available Stand:** 10k
- **Available Bikes per Station:** 5.80
- **Full Station:** 210 | **Empty Station:** 381

Analysis Insights:

Descriptive

- The dataset includes **1,028 bike stations** operating across **25 contract regions**.
- A total of **5,961 bikes** are currently available for use.
- The system also has approximately **10,000 available bike stands**.
- On average, each station holds **5.8 bikes**, indicating moderate bike availability.
- **381 stations are empty**, suggesting high demand or frequent usage in those locations.
- **210 stations are completely full**, indicating lower usage or excess supply in some areas.
- Overall, the analysis highlights the need for **better bike redistribution** to improve availability and operational efficiency.

Diagnosis

- The bike-sharing system shows **uneven distribution of bikes** across stations, leading to imbalances in availability.
- A significant number of **empty stations (381)** indicate high demand in certain areas, potentially causing user inconvenience.
- The presence of **210 full stations** suggests underutilization or excess supply in some locations.
- The average of **5.8 bikes per station** reflects moderate availability but highlights the need for better redistribution.
- Overall, the system requires **improved balancing strategies** to ensure optimal bike availability and efficient utilization across all contract areas.

Predictive

- Based on current usage patterns, stations with consistently **high empty rates** are likely to continue facing bike shortages during peak hours.
- Stations with **frequent full capacity** are expected to experience low turnover unless bike redistribution strategies are improved.
- As demand grows in high-traffic areas, **bike shortages may increase**, especially during peak commuting times.

- Predictive trends suggest that **rebalancing bikes from low-demand to high-demand stations** can significantly improve availability.
- Without operational adjustments, the gap between high-demand and low-demand stations is expected to widen over time.

Prescriptive

- Implement **dynamic bike redistribution strategies** to move bikes from low-demand stations to high-demand stations during peak hours.
- Increase **bike capacity or station size** in areas that consistently show high usage and frequent empty conditions.
- Reduce excess bikes or relocate docking stations from **low-utilization areas** to improve overall system efficiency.
- Introduce **real-time monitoring and alerts** to identify empty or full stations early and take corrective action.
- Use historical demand patterns to **optimize redistribution schedules**, especially during peak commuting hours.
- Plan future station expansions in areas showing **consistent high demand and growth potential**.

---

10. Conclusions

The analysis highlights key gaps in bike availability and station utilization across cities. Addressing these imbalances through better redistribution and planning can improve system efficiency and user satisfaction.