1. Project Introduction

This Business Intelligence project delves into the dataset sourced from a bike rental system, aiming to unravel the intricate demand patterns. By analyzing historical data recorded in 2011 and 2012, the project strives to offer actionable insights to bike rental program managers. The primary focus is on optimizing bike availability on an hourly basis per day, fostering a strategic approach to resource planning.

The database encapsulates essential variables such as the year, day (categorized into holidays or workdays), and the hour of measurement. Key metrics, including the number of bikes rented by registered and non-registered customers, contribute to a comprehensive understanding of rental dynamics.

This project not only provides an in-depth retrospective analysis but also harnesses the power of historical data to make informed predictions. By extrapolating trends and patterns, we seek to anticipate the demand for bikes, facilitating proactive planning for upcoming years. The integration of predictive analytics allows us to streamline operations, ensuring the bike rental program aligns seamlessly with the ebb and flow of demand.

As we embark on this BI journey, the project establishes key performance indicators (KPIs) to gauge success and enhance operational efficiency. By leveraging the insights derived from the dataset, we aim to empower bike rental program managers with the tools necessary for strategic decision-making, ultimately optimizing the overall performance of the bike rental system.

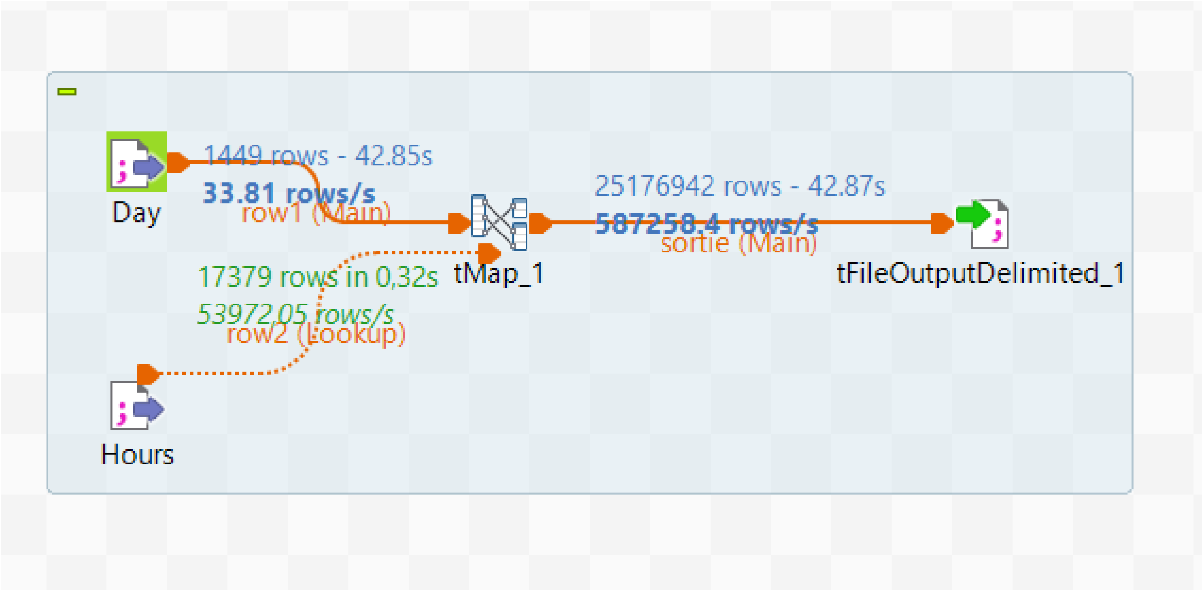
2. Implementation

The database was extracted from the UCI Machine Learning Repository. The dataset is titled "Bike Sharing Dataset" and comprises three files. Two of these files are in CSV format, while the third is a text file providing an explanation of the data contained in both CSV documents.

Here is the link to the dataset : Link

2.2 Data preparation

I began by scrutinizing the data within the raw datasets and identified that both tables shared identical data, except for one column present in one database but not in the other. To address this disparity, I utilized Talend to merge both tables, creating a consolidated database that incorporates both sets of data.



Remarkably, the merged table exhibited no missing values, negating the need for any data removal.

Variables Table

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
instant	ID	Integer		record index		no
dteday	Feature	Date		date		no
season	Feature	Categorical		1:winter, 2:spring, 3:summer, 4:fall		no
yr	Feature	Categorical		year (0: 2011, 1: 2012)		no
mnth	Feature	Categorical		month (1 to 12)		no
hr	Feature	Categorical		hour (0 to 23)		no
holiday	Feature	Binary		weather day is holiday or not (extracted from http://dchr.dc.gov/page/holiday-schedule)		no
weekday	Feature	Categorical		day of the week		no
workingday	Feature	Binary		if day is neither weekend nor holiday is 1, otherwise is 0		no
weathersit	Feature	Categorical		- 1: Clear, Few clouds, Partly cloudy, Partly cloudy		no

Rows per page 10 0 to 10 of 17

Variables Table

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
temp	Feature	Continuous		Normalized temperature in Celsius. The values are derived via (t-t_min)/(t_max-t_min), t_min=-8, t_max=+39 (only in hourly scale)	C	no
atemp	Feature	Continuous		Normalized feeling temperature in Celsius. The values are derived via (t-t_min)/(t_max-t_min), t_min=-16, t_max=+50 (only in hourly scale)	C	no
hum	Feature	Continuous		Normalized humidity. The values are divided to 100 (max)		no
windspeed	Feature	Continuous		Normalized wind speed. The values are divided to 67 (max)		no
casual	Other	Integer		count of casual users		no
registered	Other	Integer		count of registered users		no
cnt	Target	Integer		count of total rental bikes including both casual and registered		no

Rows per page 10 10 to 17 of 17

Throughout the mapping and merging process, I deliberately retained only the pertinent columns. **Weather-related data was deliberately omitted to streamline subsequent data analysis.**

The forthcoming analysis will adopt a comparative approach, focusing on years and dates as key parameters for evaluation.

After executing the job in Talend, I got a CSV that contains the merged data of the two data sources. which served as input later on in the Data visualization.

In the given dataset, we can identify the fact and dimension tables as follows:

- Fact Table:

- `cnt` (count of total rental bikes, including both casual and registered users) is the primary fact in the dataset. **It represents the central numeric measure around which the analysis revolves.**

- Dimension Tables:

1. Time Dimension:
   - `instant` (record index) and `dteday` (date) can be part of the time dimension. These attributes provide a temporal context to the bike rental data.

2. Season Dimension:
   - `season` (season index) can be considered a dimension representing the different seasons: 1 for spring, 2 for summer, 3 for fall, and 4 for winter.

3. Year Dimension:
   - `yr` (year) can be treated as a dimension, distinguishing between 2011 (0) and 2012 (1).

4. Month Dimension:
   - `mnth` (month) represents another dimension capturing the month of the year, ranging from 1 to 12.

5. Hour Dimension:
   - `hr` (hour) is part of the time dimension, representing the hour of the day (0 to 23). However, it's worth noting that this attribute is not available in the `day.csv` dataset.

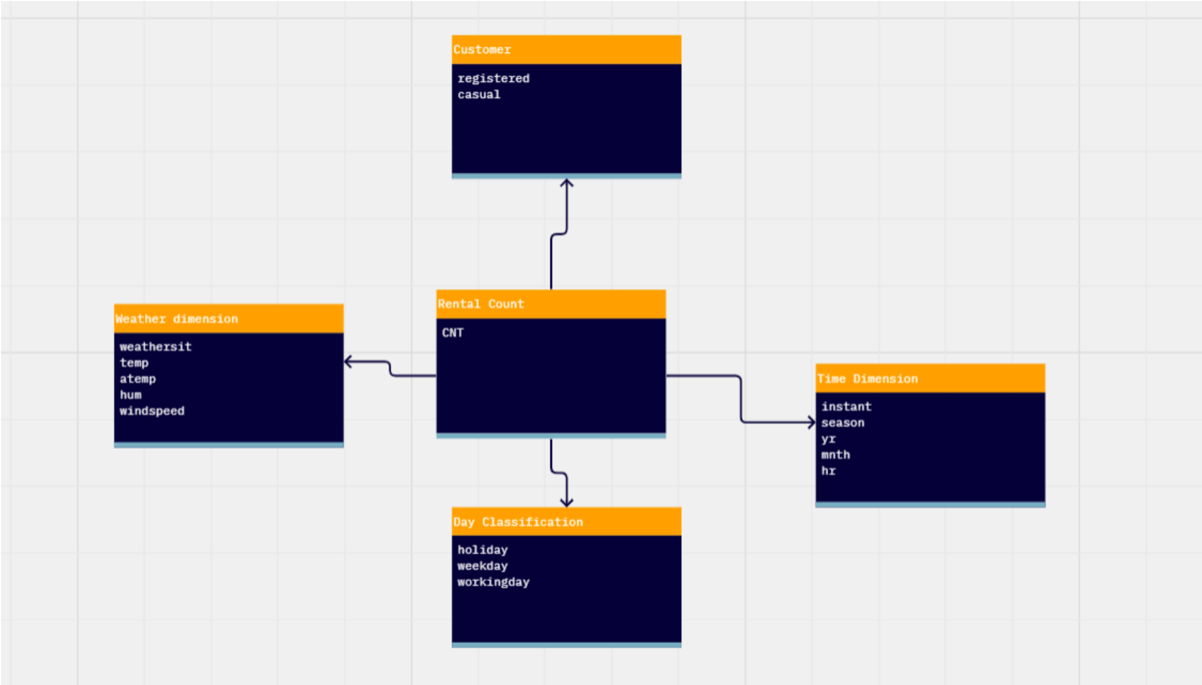
6. Holiday Dimension:
   - `holiday` (indicator of whether the day is a holiday or not) can be considered a binary dimension representing the holiday status.

7. Weekday Dimension:
   - `weekday` (day of the week) is another dimension representing the days of the week.

8. Working Day Dimension:
   - `workingday` (indicator of whether the day is a working day or not) can be treated as a binary dimension, indicating whether the day is neither a weekend nor a holiday.

.9. User Type Dimension:
   - `casual` and `registered` can be considered dimensions representing user types. These attributes capture the count of casual and registered users, respectively.

These identified dimensions and the central fact (`cnt`) collectively form the foundation for a multi-dimensional analysis of the bike-sharing dataset.



OLAP (Online Analytical Processing) Analysis

1. OLAP Cube Structure

Our OLAP analysis centers around a well-defined cube structure, comprising a central fact table and several dimension tables. The fact table, `cnt` (count of total rental bikes), is connected to key dimensions, enabling multi-dimensional analysis.

2. Key Dimensions

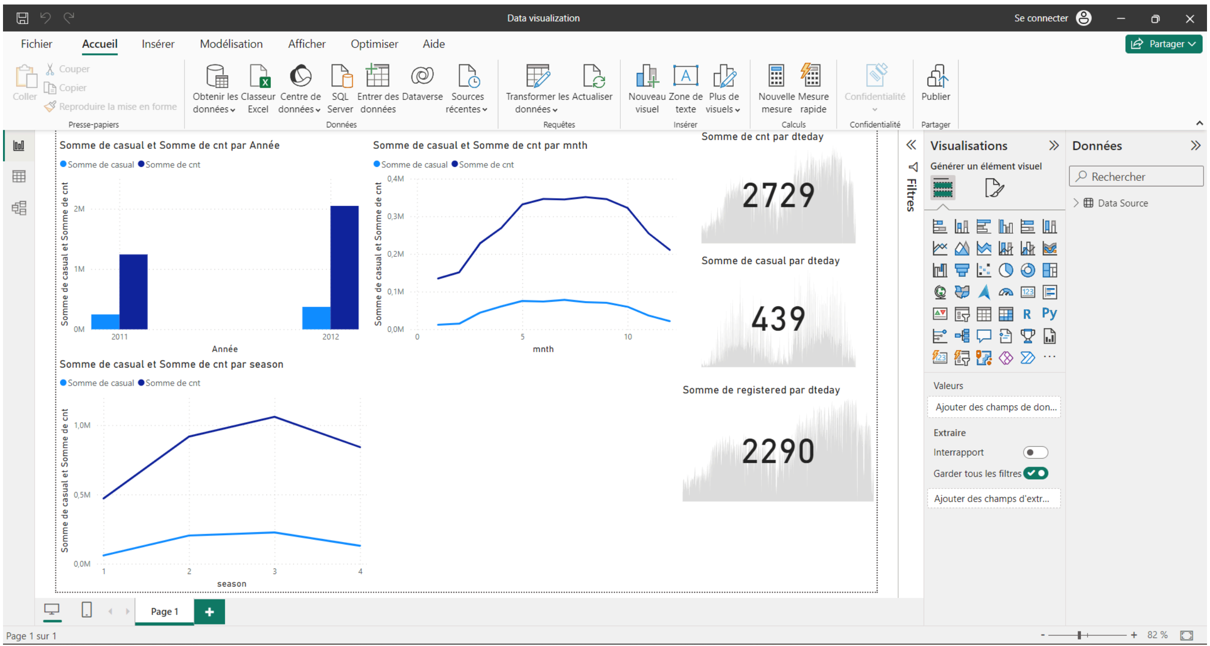
We have identified and defined key dimensions that form the foundation of our OLAP cube. These dimensions include:

-Weather
-Customer
-Time
-Day Classification

3. Measures

The primary measure in our OLAP cube is the `cnt` (count of total rental bikes), representing the central numeric value around which our analysis revolves. Additionally, we have measures related to casual and registered users.

4. Data Visualization
Power BI Desktop was employed for data visualization. I imported the merged database as the input and initiated the creation of graphs. These visualizations are seamlessly integrated into an interactive dashboard provided by the tool, enhancing the accessibility and usability of the data representations.



Each Key Performance Indicator (KPI) serves a specific purpose in providing valuable insights into different aspects of the bike-sharing system. Here's how these KPIs might be useful for users of the analysis system:

1. Number of Bike Rentals per Year (2011 / 2012) - Segmented by Category:
   - Purpose: This KPI helps users understand the overall trend of bike rentals across two consecutive years, distinguishing between registered members and casual members.
   - Usefulness: It allows stakeholders to identify growth or decline in overall bike usage and assess the contribution of registered versus casual members to these trends. This insight aids in strategic planning, marketing, and resource allocation.

2. Number of Bike Rentals per Season - Segmented by Category:
   - Purpose: This KPI provides a seasonal breakdown of bike rentals, allowing users to identify patterns influenced by weather conditions and member categories.
   - Usefulness: Stakeholders can understand the seasonality of bike usage and tailor marketing or operational strategies accordingly. For example, adjustments in bike availability or promotional campaigns can be optimized based on seasonal demand.

3. Average Total Rentals per Day:
   - Purpose: This KPI offers a daily average of total bike rentals, providing an overview of the system's overall utilization.
   - *Usefulness:* Users can gauge the system's general popularity and establish benchmarks for daily performance. Fluctuations in this metric may prompt further investigation into specific days that deviate from the average.

4. Average Rentals by Registered Members:
   - Purpose: Focuses on the average number of rentals made by registered members, providing insights into their consistent engagement.
   - Usefulness: This KPI helps in understanding the loyalty and regularity of registered users. It's crucial for customer relationship management and can inform loyalty programs or incentives tailored for this segment.

5. Average Rentals by Casual Members:
   - Purpose: Similar to the previous KPI, this focuses on the average rentals but for casual members.
   - Usefulness: Allows users to understand the behavior of occasional or one-time users. Tailoring marketing strategies or promotions for this segment can be based on these insights.

6. Total Rentals per Day:
   - Purpose: Provides a straightforward count of total bike rentals each day, emphasizing the daily demand.
   - Usefulness: Users can identify peak days of activity, which is crucial for operational planning, staffing adjustments, and ensuring the availability of sufficient bikes during high-demand periods.

In summary, these KPIs collectively offer a holistic view of the bike-sharing system, enabling stakeholders to make informed decisions related to operational efficiency, marketing strategies, and overall system optimization.

5. Conclusion

In conclusion, this Business Intelligence project has successfully navigated the intricate landscape of bike-sharing system data, shedding light on patterns and insights that hold immense significance for the effective management and strategic planning of bike rental programs. By amalgamating datasets from 2011 and 2012 and employing the powerful capabilities of Talend and Power BI, we have crafted a comprehensive analysis that addresses key performance indicators (KPIs) crucial to the understanding of user behavior, seasonal variations, and overall system dynamics.

The systematic examination of KPIs, such as the number of bike rentals per year, segmented by member category, and the seasonal distribution of rentals, has empowered stakeholders to make data-driven decisions. Through the lens of these KPIs, we gained nuanced insights into yearly trends, seasonal fluctuations, and the distinct preferences of registered versus casual members. This knowledge equips decision-makers with the tools to optimize resource allocation, enhance user experiences, and tailor marketing strategies based on user behaviors and demand patterns.

The adoption of Power BI as our data visualization tool proved instrumental in transforming complex datasets into interactive dashboards, providing an intuitive platform for stakeholders to explore and derive actionable insights. The visual representations facilitated a deeper understanding of the data, enabling quick and informed decision-making.

As we look to the future, the predictive power embedded in our analysis offers a strategic advantage for anticipating demand, planning resource allocation, and optimizing operational efficiency in the coming years. The project's success in meeting its objectives underscores the value of Business Intelligence in translating raw data into actionable intelligence for the enhancement of bike-sharing programs.

In essence, this project not only contributes to the academic exploration of data analysis but also holds practical implications for the effective and strategic management of bike rental services. The insights gained through this endeavor serve as a foundation for continuous improvement, emphasizing the ongoing need for data-driven decision-making in the dynamic landscape of urban mobility solutions.