**Project 5 - Build Data Warehouse  
for Cyclist Data**

MGMTMSA 405: Data Management - MSBA 2024 Final Project

Your goal for this project is to build a small data warehouse based on *Cyclist* dataset.  
The dataset information is located here: <https://divvybikes.com/system-data>

The full dataset can be downloaded from [here](https://www.dropbox.com/scl/fi/392la2w5bzr5cvh4hqs7d/Cyclist-data.zip?rlkey=b70ddt83we8e3wp48v32jrb44&dl=0).

The scope of this project includes analyzing the dataset, loading the data into Snowflake, and producing fact and dimension tables to allow you to generate SQL query to produce the KPI.

The KPIs will then be represented in Tableau dashboards.

The Tableau dashboard should be customizable, allowing to select different dimensions to observe data.

The final deliverable for this project will be a project report and a Tableau dashboard.

To complete this project, students should demonstrate experience in data analytics, database management, and data visualization. They should use strong communication skills and be able to work collaboratively together to ensure that the final deliverable meets all requirements.

Upon completion of this project, the student team will have gained valuable experience in data management and analytics.

Additionally, they will be able to shared their work on social media like LinkedIn or during interviews.

**Recommended Steps**

1. Load the dataset into Snowflake (you can either copy the file to your S3 bucket or try to load directly from Snowflake from your desktop). Describe which method you used and why.
2. Build a Data Dictionary of the dataset (look for the Metadata Source on the crime data dataset page on the data.gov website, link above.
3. Dimensional Modeling
   * Define the business or event process.
   * Define the grain.
   * Define your Fact and Dimensions to produce the KPIs listed below using the 7Ws (who does what, when …)
4. Build the Fact and Dimension tables
   * Provide an ERD Diagram of your fact and dimension tables (a star schema showing the relationship between the tables and the column names).   
     I recommend DBSchema (<https://dbschema.com/>) but you’re free to use any other tools. (Another one is Navicat)
5. Generate your Fact and Dimension in your Snowflake account using SQL.
6. Populate your fact and dimension tables using SQL (date\_dim, timeOfday\_dim if necessary), and other necessary dimensions for your analytics.
7. Write the SQL query to generate the KPIs below.
8. Connect Tableau to your Snowflake account, and build 5 dashboards to represent the different KPIs below.
9. Use the Tableau dashboards you have built to provide analytical observation as if you were presenting this data visualization to the Chicago Department of Transportation.
   * Try to extract some important insights ; to tell a story.

**KPIs**

1. **Ride duration**: This KPI measures the duration of the ride, based on the time it started and the time it ended.
2. **Most common rideable type**: This KPI measures the type of rideable bike (eg. Classic bike, electric bike, etc.) that was most common during a given period of time and/or location.
3. **Most frequented route**: This KPI measures the route (Start station – End station) with the most trips done in a period of time (e.g. day of week, morning/afternoon, etc.).
4. **Max bike usage**: This KPI measures the total number of bikes being used in the entire network (all stations), during a period of time. The aim is to understand what moments of the day have the peak usage of bikes. Students should look at the data and determine the time intervals that make more sense to split the day in.
5. **Casual-member rate**: This KPI measures the percentage rate between casual rides and member rides. It is calculated as (# casual / # member) \* 100 for a given location/route/destination and/or time period.

***Important Note About Finals Submission****:*

* *For the Tableau dashboard, please use Tableau Data Extracts instead of a Live Connection, and submit a .twbx file.  
  Compress the .twbx file.*
* Store all your files into **1** **Folder**.
* 1 folder per group !
* Compress the Folder.
* Include the project number. Include your Finals Group Number in the folder name too.
* Example : Group8\_Project5.zip

**Project Report**

The project report should include the following:

* **Executive Summary**:

A brief and concise summary of the project, highlighting the most critical and relevant insights and findings.

* **Project Statement**:

Project introduction showing a clear understanding of the dataset and the project goals.

* **Data Dictionary**:

Structured description of the data elements, attributes, and entities present in the data warehouse (tables and relationships, attributes, data types, lengths, domains, and constraints)

* **Dimensional Modeling + ERD**

Use the 4 critical steps learnt in class to define your fact and dimension tables.  
Represent your fact and dimension tables in a star schema. Explain why you choose these dimensions. What are the measures? Paste your star schema in the document.

* **Data Transformation**:

Explain how you populated your dimensions and fact table. Don’t paste long SQL into the document. Store your SQL command in a SQL file and add to the folder.

* **Produce KPIs using Fact and Dim**:

Explain which dimension and fact you use to generate each KPIs. If SQL is few lines you may copy in the document.   
Include your SQL also in a separate file and store in the folder.

* **Data Visualization**:

Show your dashboards (screenshot) and provide analytical observation. Try to extract some important insights, to tell a story. Maybe make a list of recommendations or actions.

* **Project Challenges**

Brief summary of the project limitations, challenges faced, and how you solved them.

**Tableau Help: Creating Visualizations and Building a Dashboard in Tableau**

1. **Download and Install Tableau Desktop**

Tableau offers free one-year Tableau license to students (access to Tableau Desktop and other Tableau learning resources) through the Tableau for Students Program

To get your free one-year Tableau license, open the ‘*Tableau for Students’* webpage:

<https://www.tableau.com/academic/students>

Click on “Get Tableau for Free”

Graphical user interface, application, Teams

Description automatically generated

Enter your contact information along with your “UCLA or Anderson” email

Graphical user interface, text

Description automatically generated

For ‘*How will you be using your Tableau license?*’, enter “Learning on own”, and click on “Verify Student Status”

Graphical user interface, text, application

Description automatically generated

Upload your documentation (e.g., Admission Letter, BruinBill, etc.) and click on “Submit”

Graphical user interface, text

Description automatically generated

You will receive an email from SheerID Verification with your “Product Key” and “Access Key” (retain it for future use)

Graphical user interface, text, application

Description automatically generated

Download and Install Tableau Desktop using this link:

<https://www.tableau.com/support/releases>

Graphical user interface, application, Teams

Description automatically generated

Once the installation is complete, open Tableau Desktop, click on “Activate Tableau” (you don’t need to fill the 14-day free trial form) and enter your information (same as that entered on the ‘*Tableau for Students*’ webpage

Enter your “Product Key” (received on email) and click “Activate”

You are all set to use Tableau now!

1. **Connect to Snowflake**

Connect Tableau to Snowflake using the *Tableau Snowflake Connector* and build your dashboard

Please note the following:

* You may need to download drivers to use the Tableau connectors
* Use ‘*Tableau Data Extracts*’ instead of ‘*Live Connection*’ to improve performance
* Don’t use ‘*Tables*’ as data source, instead write ‘*Customer SQL Queries*’ to pull the required data
* Try to create only a few queries to get the required data for all your KPIs
* Save your Tableau workbook as a ‘*.twbx*’ file
* You can share the final data extracts and Tableau workbook within your group, to work on different visualizations separately