

Data Dictionary & Project Description

This supplement describes the data provided for your group project. These are real-world datasets. To protect the data provider's proprietary information, the structure of these datasets, the locations of the tanks therein, and the invoices have been obfuscated so as *not* to reflect the real information of the data provider.

The datasets chronicle over a year's fuel purchases (by the gas station owners) and sales at all city gas stations.

Locations.csv

This dataset lists all the gas station locations and contains the following columns:

- Gas Station Location: The unique ID of the gas station
- Gas Station Name: The gas station name
- Gas Station Address: The gas station address
- Gas Station Latitude: The gas station latitude
- Gas Station Longitude: The gas station longitude

Tanks.csv

Each gas station location may have more than one tank. This dataset contains information about these tanks and their attributes

- Tank ID: A unique ID of each tank in the system
- Tank Location: Gas station this tank is located at
- Tank Number: ID of each tank in a specific location
- Tank Type: The type of fuel this tank is used for: U for regular gas, D for Diesel, and P for premium
- Tank Capacity: Capacity of the tank in liters

Invoices.csv

Each gas station purchases different fuel types from its supplier(s). Every delivery of each fuel type to all tanks of a location generates one invoice. The Invoices.csv dataset contains information about these invoices over time and has the following columns:

- Invoice Date: Date of the purchase
- Invoice ID: Unique ID of the invoice
- Invoice Gas Station Location: Gas station location
- Gross Purchase Cost: Total Canadian Dollar (CAD) paid for the purchase
- Amount Purchased: Total number of fuel liters purchased
- Fuel Type: Purchased fuel type

Fuel_Level_Part_1.csv and Fuel_Level_Part_2.csv

These two datasets contain fuel level information in each tank at frequent and mostly regular time stamps. These two datasets contain the following columns:

- **Tank ID:** ID of the tank
- **Fuel Level:** The amount of remaining fuel (inventory in liters)
- **Time Stamp:** The time of inventory reporting

Problem Description:

A gas station purchases fuel in bulk (thousands of liters) and sells it to customers like you and me. A typical gas station (location) may offer different types of fuel (regular gas, premium gas, diesel) and each type of gas may be stored in one or more than one tanks in that particular gas station. These tanks are usually underground, out of safety and limited space considerations. The number of tanks and capacity of each tank is driven by many factors such as available space, city regulations, closeness to the suppliers' reservoirs, and demand, among other factors. It is common for a gas station to carry tens of thousands of liters of fuel at any time. At this relatively large scale, the following decisions may have significant consequences on the survival and profitability of the gas station:

- Fuel replenishment frequency
- Fuel replenishment quantity

This is an exciting managerial question for a business school student, with analytical skills, like you! In the one hand, frequent replenishment in small quantities is attractive as it has less cash tied up to the fuel inventory. On the other hand, larger and less frequent deliveries may qualify the gas station for the quantity discount at every fuel replenishment offered by its supplier(s).

Independent of the fuel type, the supplier offers the following quantity discounts:

| Purchase quantity (liters) | Discount per liter |
|----------------------------|--------------------|
| 0-15000 | 0 |
| 15000-25000 | 2 cents |
| 25000-40000 | 3 center |
| 40000+ | 4 cents |

Business Question

Ultimately, your team is responsible for thoroughly exploring the provided dataset, providing descriptive statistics, inspecting each gas station's inventory replenishment pattern, visualizing it, and suggesting a better inventory policy that may save these gas stations a significant amount of money. Your decisions must be based on the provided data processed using python and its data analysis packages. You can ignore the gas delivery cost and focus on making the correct inventory replenishment decision that may reduce total purchasing cost while maintaining an excellent customer service level (by not running out of gas).

What questions should you answer in your report?

When embarking on a data-driven decision-making process, it is crucial to determine your analysis's direction. Typically, hypotheses are formed during the initial exploration of the datasets. In this project, we aim to analyze the fuel price and purchasing order data to evaluate how well we manage our fuel tanks' inventory and order fuel. By visualizing the inventory evolution trajectory, we can gain insights into our inventory management practices and identify areas for improvement. We can also determine which locations manage inventory effectively and save money and which locations have riskier inventory management practices (maintain lower safety inventory). To quantify performance, we can compare the amount of money saved to the maximum potential savings possible if we optimize our purchasing strategy. It is important to consider inflation in our calculations, as the purchasing power of money changes over time. To do this, we need to find Canada's monthly inflation rates, create a small new dataset with these rates, and join it with our existing data.

Based on your analysis, you can develop recommendations for improving the inventory management policy of each location and estimate potential cost savings. Additionally, we can evaluate whether increasing the capacity of existing tanks would be beneficial and identify which fuel stations would benefit most. We can also explore whether a particular day of the week is best for ordering fuel. Keep in mind that answering these questions requires several rounds of data cleaning, merging, transforming, and visualization. While these directions are important, they are not exhaustive. Your analysis should explore significantly outside the scope of these directions to achieve a thorough understanding of our inventory management practices, cost structure, and overall efficiency.

Group Report Details

Each team is responsible for organizing its report. Each team will submit:

- One report in pdf format
- One notebook containing your code that reads the provided csv files and performs the analysis. Please do not change the provided file names. However, you need to change column names (using pandas) in each file.

Evaluation:

We will evaluate your work for:

- Data processing: cleaning, merging, ...
- Clarity of your code. Do not forget to leave useful comments in your code
- Exploring the dataset and providing an overview of these datasets
- Asking and answering the right business questions
- A thorough and a well-formatted report
- Nicely formatted graphs
- Academic integrity of your work
- Following sound logic in answering the business questions

We will run your code and check your code results with the submitted report during the grading.