I. Context and assumptions

(1) Dataset Import and Overview

This project integrates multiple datasets related to **fuel inventory management**, including:

- **Fuel_Level_Part_1 & Fuel_Level_Part_2** - these two dataset contains information about the continuous records of each tank fuel levels at different timestamps.

- **Invoices** - this dataset includes the purchase orders for each tank with information like transaction dates, quantities, cost, fuel type.

- **Tanks** - this dataset contains information like the basic metadata about each tank, like capacity, fuel type, and station location.
- **Locations** - this dataset contains the basic information about the gas station like the location, address, latitude, longitude of the gas station.

All datasets were imported using the pandas library (pd.read_csv()) to build a dataframe format, followed by an initial inspection with info() and describe() to assess record counts, data types, and missing values. This exploratory step identified structural inconsistencies across datasets and guided the data cleaning process.

(2) Field Standardization and Consistency

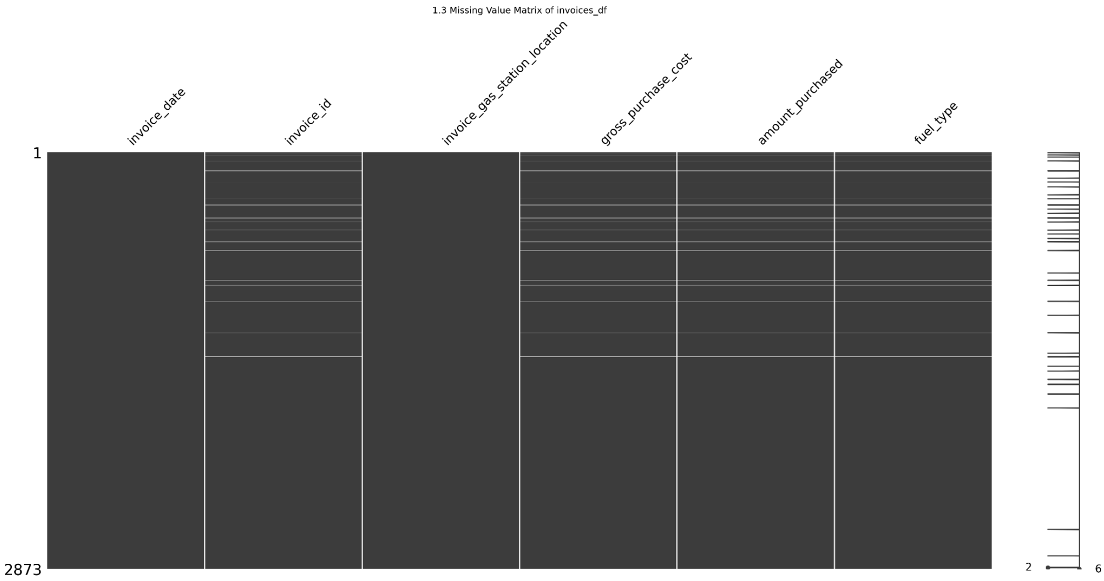
From the inspection of the dataframe using info(), we found the format of the column names were blank, like “Tank ID”. In order to make the subsequent analysis cleaner, we standardized the column names using rename(), and stripped the extra spaces with.str.strip() to avoid mismatched joins (e.g., “T 10” vs. “T10”). This step ensured compatibility during later merges and minimized referential errors.

(3) Data Cleaning and Handling Missing Values

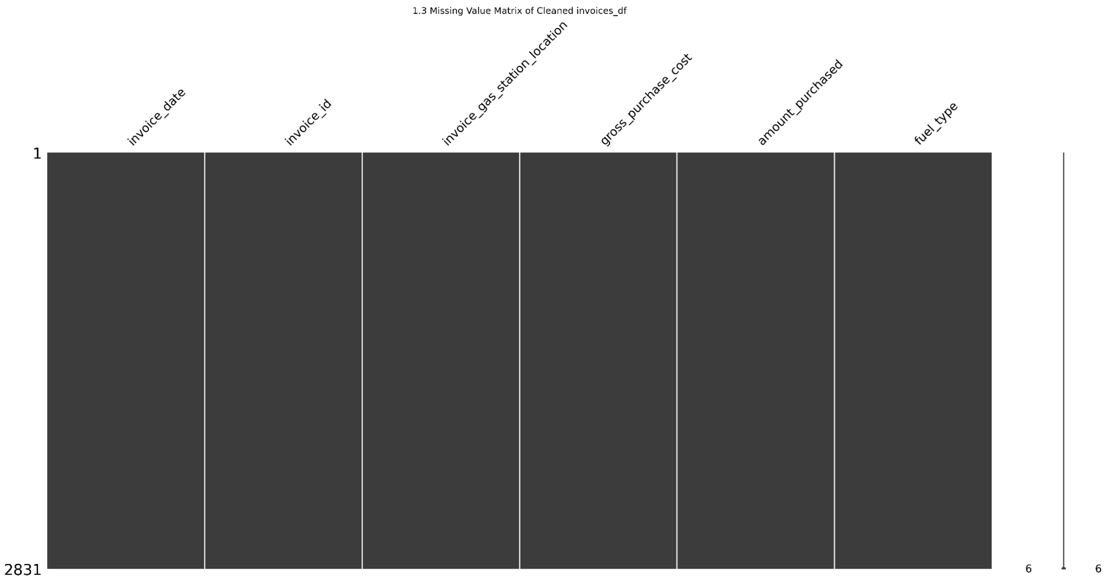
The cleaning process included several major procedures:

1. **Missing Value Treatment**

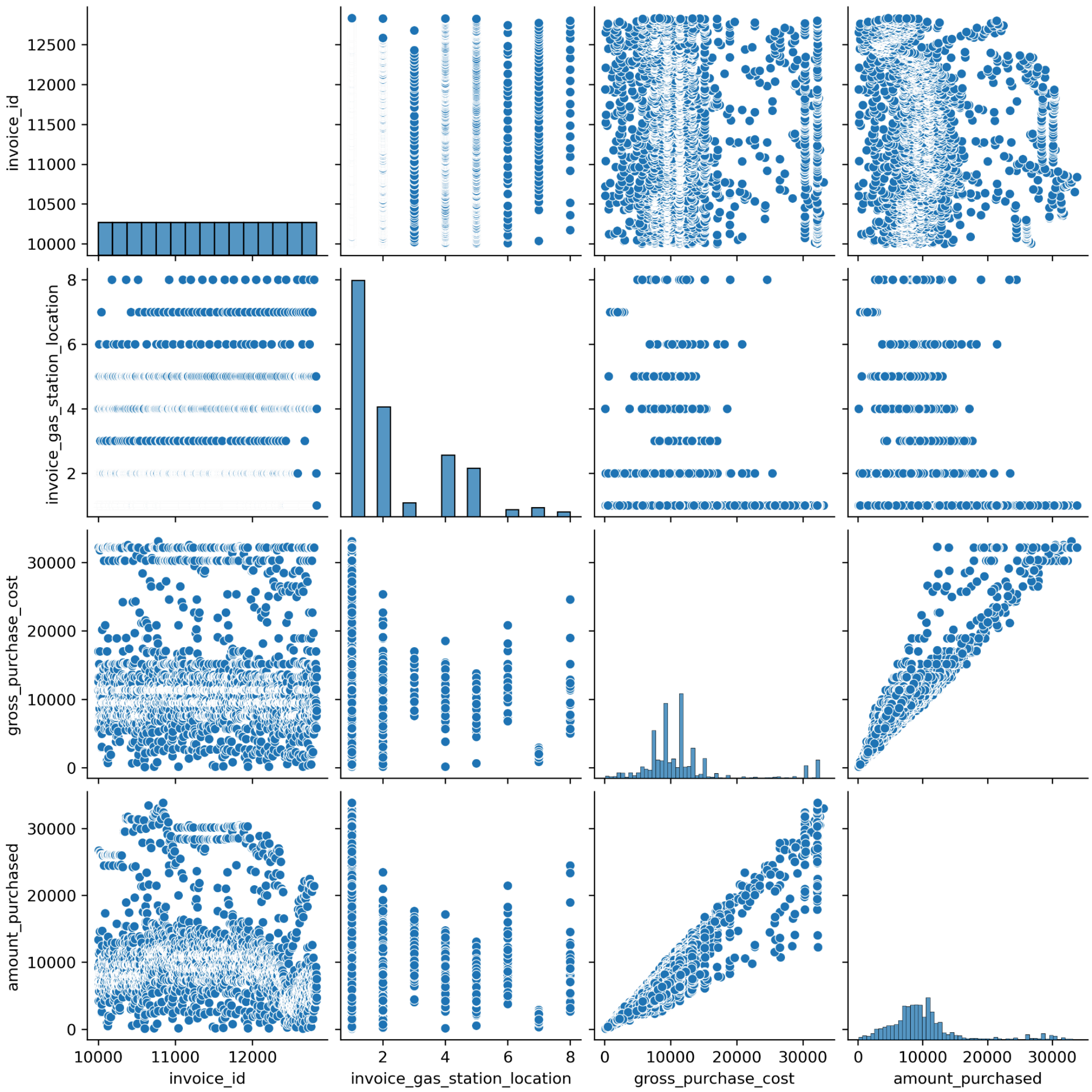
   Missing values in each dataframe were assessed using isnull().sum(). After knowing the missing value distribution, missing values are treated with different methods like dropna() or fillna() to process data into a more analyzable format. For example, records with missing key variables (e.g., fuel level, timestamp, tank ID) were removed using dropna(), while auxiliary variables were filled with zeros or median values using fillna().



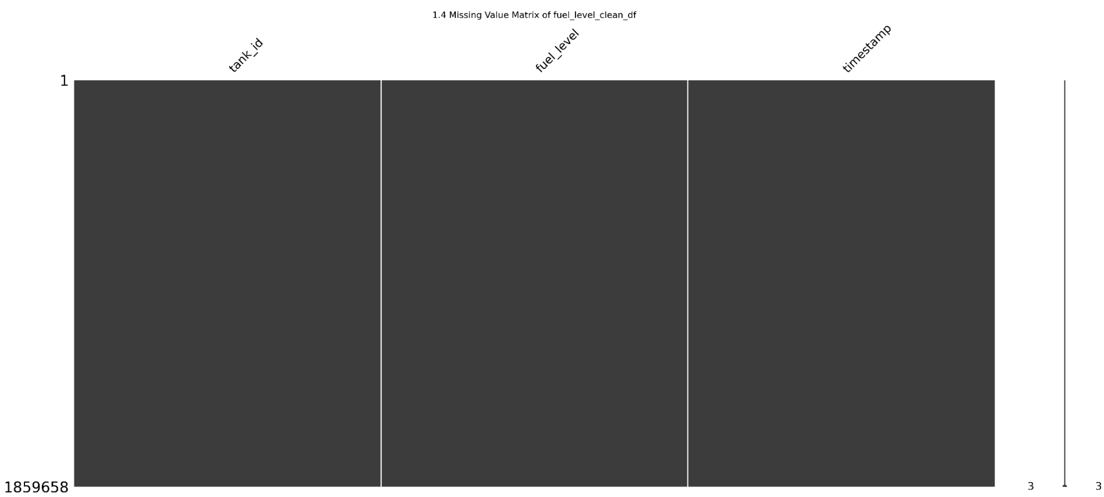
Missing value for the invoices_df



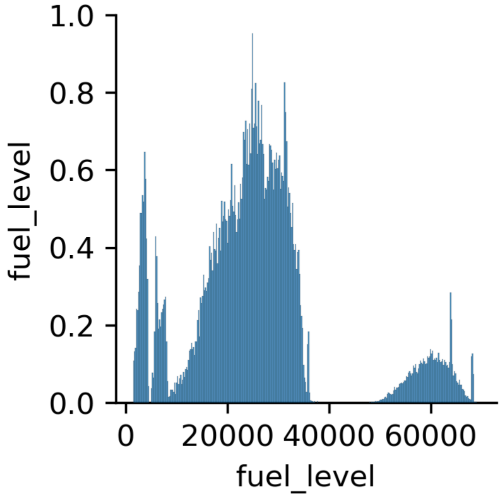
Invoices_df after data cleaning



Pairplot for Invoices_df



fuel_level_df after data cleaning



Pairplot for fuel level

2. **Data Type Conversion**
   All timestamps were standardized with pd.to_datetime() to prepare for the timeseries analysis.

(4) Data Merging and Integration

After cleaning, in order to further analyse the fuel level of each tank, we used .merge() and .concat() to make the dataset more analyse- friendly. For example, we concatenated the 2 parts data of the fuel level into one dataset named fuel_level_df to have a more comprehensive view of the fuel level situation. Also, in the further analysis, we merge tanks_df and fuel_level_df via a left join on tank_id. This merge connected each inventory observation to its respective tank attributes (location, fuel type, capacity). Post - merge validation confirmed no unmatched records remained.

(5) Descriptive Statistics and Preliminary Analysis

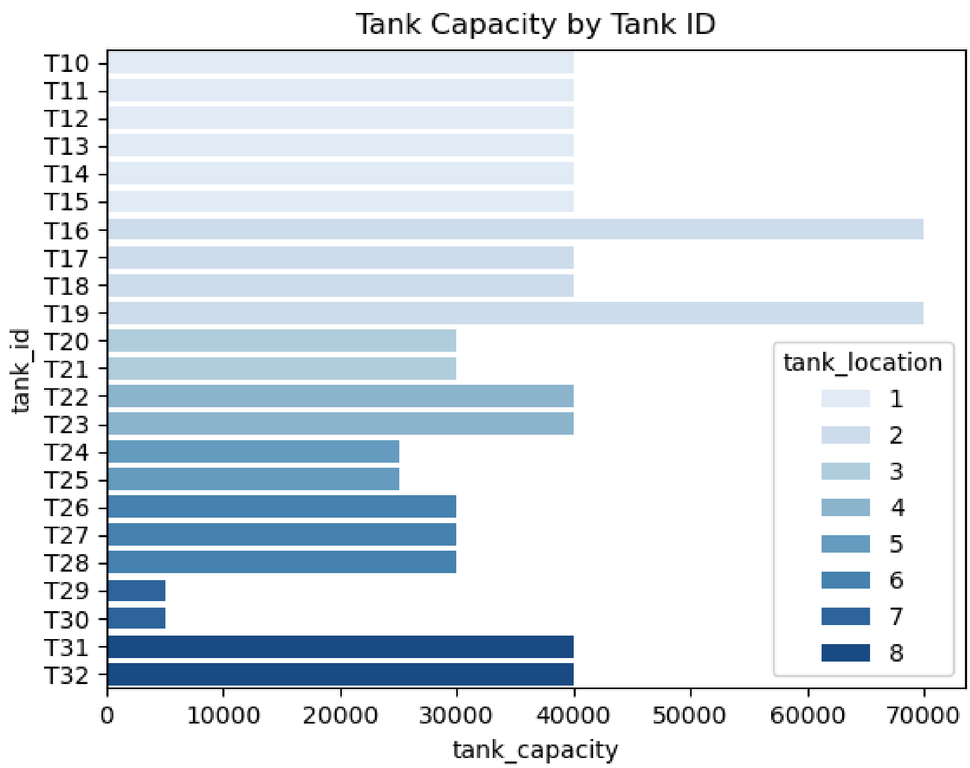
Descriptive analysis summarized inventory dynamics in the invoices, tanks, fuel_level. Using describe(), we knew the average, minimum, and maximum of the dataset.

This summary served as the analytical foundation for subsequent stages, including inventory trajectory visualization, restock optimization, and performance benchmarking.

2. Overall Analysis

2.1 Tank Capacity and Location Overview

To understand the storage structure across all sites, we need to know the tank capacity of each tank by location. In order to achieve that goal, we used the dataset tank_df to have an overview of the tank capacity, and visualized each tank's capacity by location using a horizontal bar chart. The figure below plots Tank Capacity (x-axis) against Tank ID (y-axis), with bar colors indicating the Tank Location. A blue gradient palette was applied to distinguish different sites, providing an intuitive view of capacity distribution.



The visualization reveals significant variability in storage capacity among tanks. While several tanks - such as T16 and T19 - exhibit large capacities exceeding 70000 liters, others like T29 have relatively small capacities under 10,000 liters.

This imbalance suggests that storage allocation is not uniform across sites: certain locations are equipped with higher-capacity tanks which are likely to support stations with greater throughput or regional demand concentration.

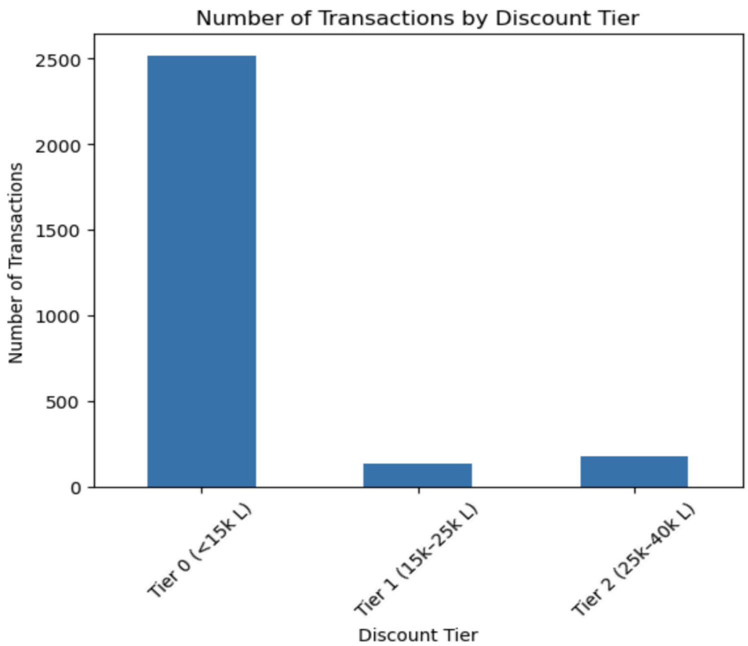
2.2 Discount

To evaluate cost efficiency across fuel purchases, we implemented a tiered discount model based on the volume of fuel ordered per transaction. The discount structure rewards higher-volume purchases with progressively larger per-liter savings. Specifically, purchases below 15,000 liters receive no discount, while those between

15,000 and 25,000 liters earn a 2¢ discount per liter. The discount increases to 3¢ for purchases between 25,000 and 40,000 liters, and reaches 4¢ for orders exceeding 40,000 liters.

Each invoice was analyzed to determine its applicable discount tier. We then calculated the price per liter before and after the discount, allowing us to assess the net cost impact across all transactions. This analysis revealed that a significant portion of purchases fell into the lower tiers, particularly just below the 15,000-liter threshold. This suggests that many stations may be missing opportunities to optimize their order volumes and unlock greater savings.

To visualize this, we generated a bar chart showing the number of transactions in each discount tier. The distribution highlights a concentration of purchases in Tier 0 and Tier 1, with fewer transactions reaching the higher discount thresholds. This pattern indicates potential for strategic adjustments in purchasing behavior to improve cost efficiency.



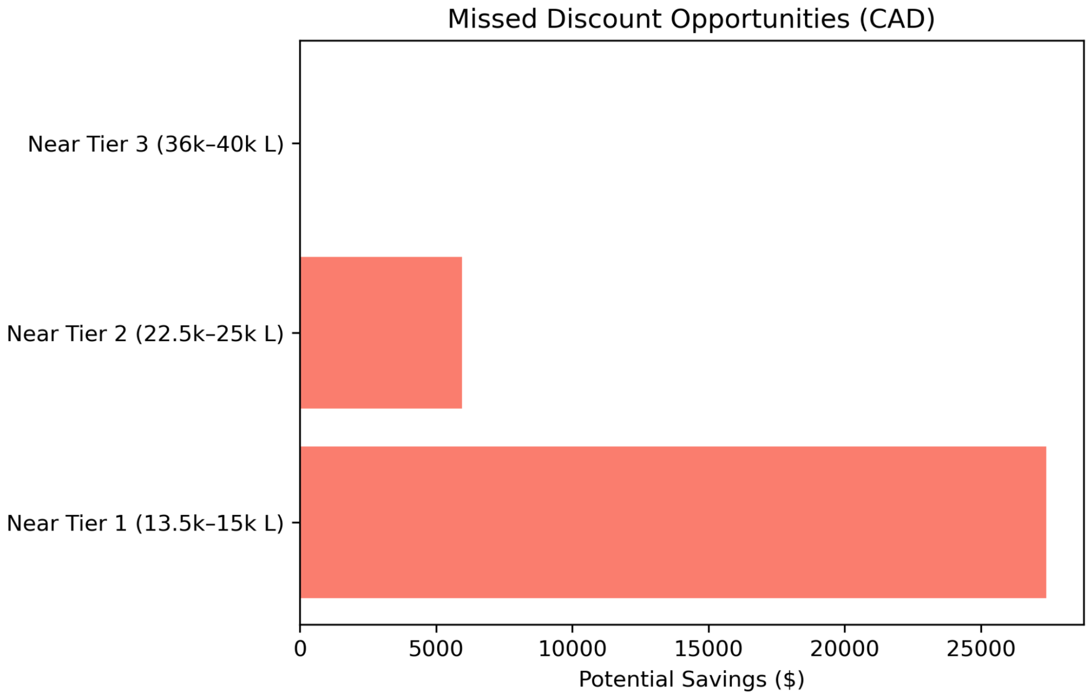
The resulting distribution (as shown in the bar chart) indicates that over 2,500 transactions fall within Tier 0, while only a small number qualify for Tier 1 and Tier 2 discounts. This heavy concentration in the lowest tier suggests that many stations consistently purchase just below the first discount threshold, missing opportunities for volume-based savings.

2.3 Missed Opportunities for Discounts

Building on the tier analysis, we identified transactions that narrowly missed qualifying for a higher discount. Specifically, we focused on purchases that were within 10% of the next tier threshold - for example, orders between 13,500 and 15,000 liters that missed the 2¢ discount by a small margin.

Across all transactions, 96 orders fell within 13.5k-15k liters (Near Tier 1) and 25 orders within 22.5k-25k liters (Near Tier 2), while no orders approached the 40k L threshold for Tier 3. For each of these near-tier transactions, we calculated the additional savings that would have been realized had the purchase volume been increased just enough to qualify for the next discount.

The results were striking: if Near Tier 1 orders had reached 15k L, total savings would have amounted to approximately $27,402.66 CAD. Similarly, raising Near Tier 2 orders to 25k L would have yielded an additional $5,956.20 CAD. In total, these near-miss transactions represent a combined potential savings of about $33,358.85 CAD - a meaningful opportunity that could have been captured with only minor adjustments in order volume.

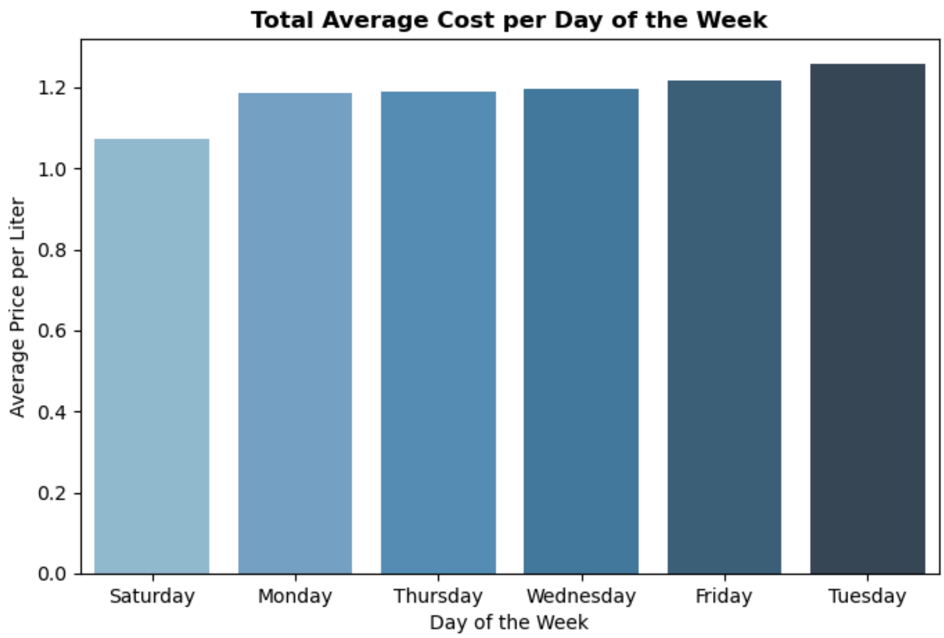


A horizontal bar chart was used to illustrate the total missed savings across the three key thresholds, reinforcing the financial impact of these near misses. The chart is shown below. This analysis underscores the importance of volume-aware purchasing strategies. By rounding up orders that are close to a discount threshold, stations could significantly reduce their fuel costs.

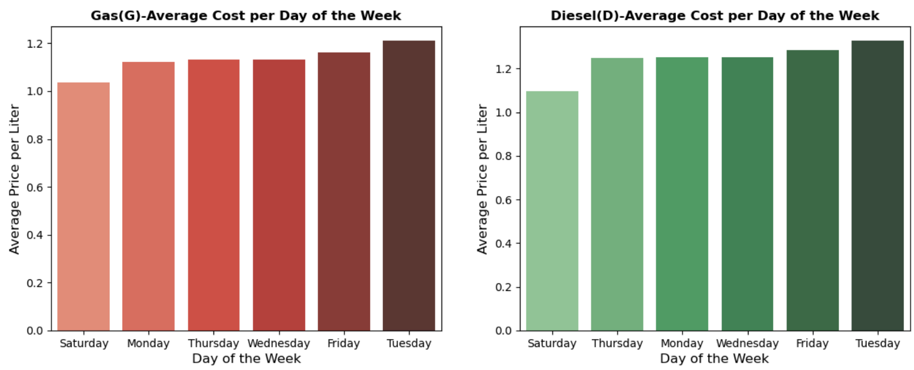
We recommend implementing real-time alerts or purchasing guidelines to help station managers recognize when a small increase in volume could unlock a higher discount tier. Such mechanisms would enable data-driven purchasing decisions and improve overall cost efficiency across stations.

2.4 Total Average Cost per day of the week

To examine potential temporal variations in fuel procurement costs, the analysis calculated the total average price per liter by day of the week. Each transaction was grouped according to its purchase day, and the mean price per liter was computed across all records. The resulting bar chart displays the average adjusted fuel cost from Monday to Saturday, allowing direct comparison of daily price fluctuations.



The results show that the overall variation across the week is relatively small - approximately 0.10 to 0.15 CAD per liter. The lowest average price typically occurs on Saturday, while the highest prices are observed on Tuesday and Friday.



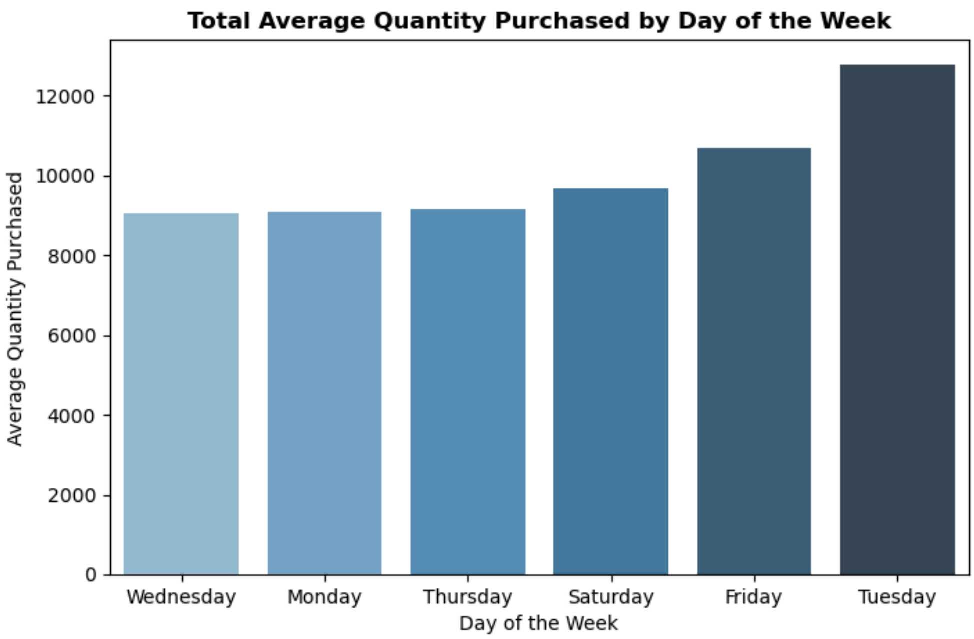
Gasoline prices exhibit a clear weekday trend. The lowest average price (1.03 CAD/L) occurs on Saturday, while Tuesday records the highest price (1.21 CAD/L), the difference of nearly 18 cents per liter. Prices remain relatively stable from Monday to Thursday (≈1.12-1.13 CAD/L).

Diesel prices follow a similar but steeper pattern. The cheapest day is again Saturday

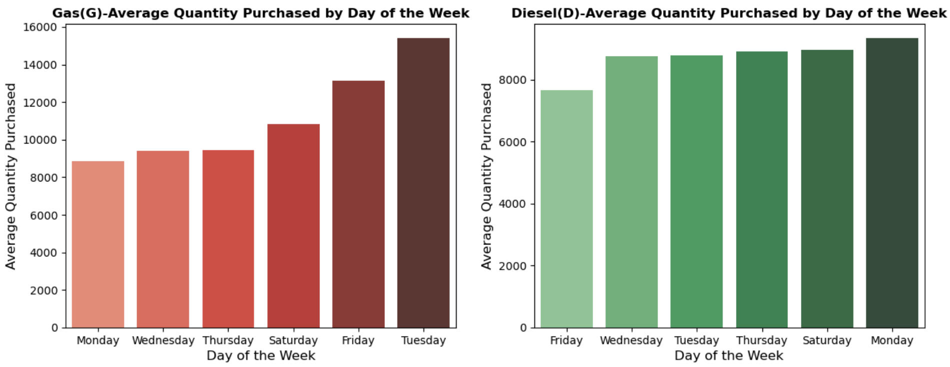
(1.10 CAD/L), while Tuesday sees the highest price (1.33 CAD/L) - a 21-cent gap, larger than for gasoline. Prices are clustered near 1.25 CAD/L during midweek (Monday-Thursday). The overall pattern suggests that both fuels become most expensive early in the week and cheapest during weekends.

2.5 Quantity purchased by day of the week

To complement the price analysis, the dataset was further examined to identify temporal patterns in purchase quantity across different days of the week. By grouping all transactions according to the day of purchase and calculating the mean purchase volume per day, we visualized the average amount of fuel purchased from Monday through Sunday. The bar chart provides a clear view of purchasing behavior trends over the course of a week.



The results show a noticeable variation in procurement volumes by day. Tuesday exhibits the highest average purchase quantity, exceeding 12,000 liters, followed by Friday, while Monday, Wednesday, and Thursday maintain relatively stable purchase volumes of approximately 9,000-10,000 liters. The lowest average quantities appear during the midweek period, suggesting a pattern where refueling activity accelerates toward the end of the week, likely in preparation for weekend demand or to restore inventory after several days of depletion. This pattern suggests that fuel stations tend to accumulate inventory before peak demand periods, demonstrating a practical approach to maintaining operational readiness.



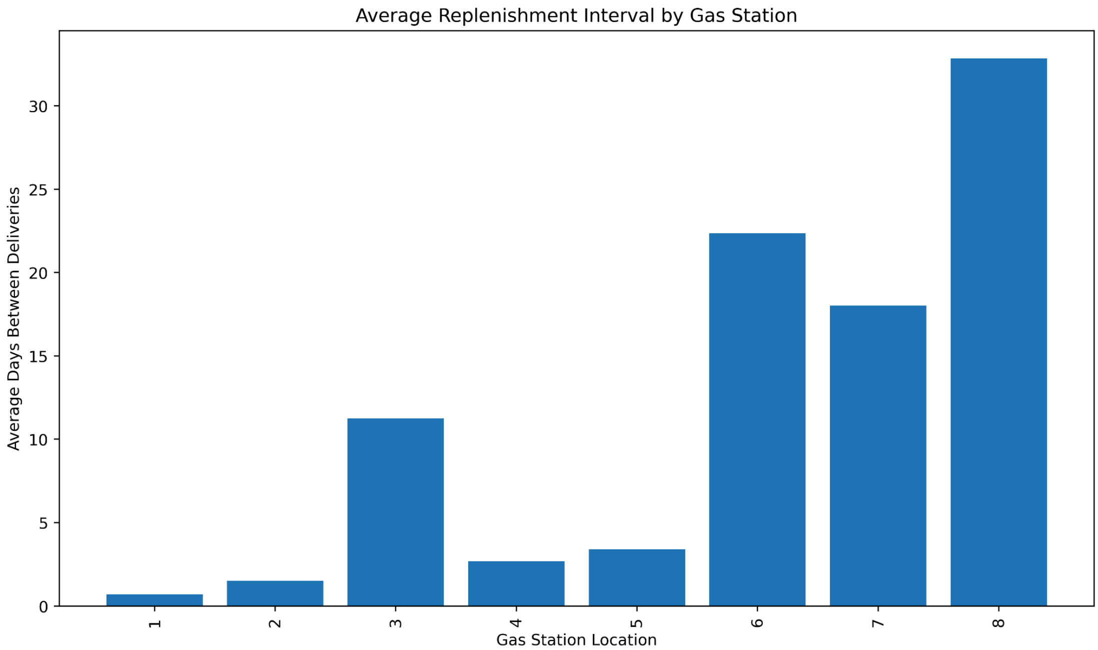
Gasoline purchases show a strong weekday pattern, peaking on Tuesday (15,395 L) and Friday (13,156 L), while reaching the lowest levels on Monday (8,841 L). Diesel purchases are relatively stable throughout the week, fluctuating within a narrower range (7,670-9,336 L). The highest purchase occurs on Monday (9,336 L), while Friday (7,670 L) marks the lowest. This suggests that diesel, often used for logistics or fleet vehicles, follows a steady weekly refueling pattern rather than concentrated bulk orders.

3. Station Analysis

3.1 Average Replenishment Interval & Average Fill Ratio by Gas Station

3.1.1 Average Replenishment Interval by Gas Station

In this part, we will focus on the analysis of replenishment behavior of the eight gas stations which reveals significant disparities in inventory management practices. For the first metric, we used the average replenishment interval on gas stations which can be calculated from the invoices data, calculating by the number of days between the two replenishment dates and finding the average of the replenishment interval. After the calculation, we also visualize the average by the bar chart below.



Interpret and business insights

As we can see from the chart, the average replenishment interval varies widely. From the station with the most frequent replenishment of less than 1 day to the less frequent replenishment of over 30 days, the difference indicates the absence of a unified refueling policy.

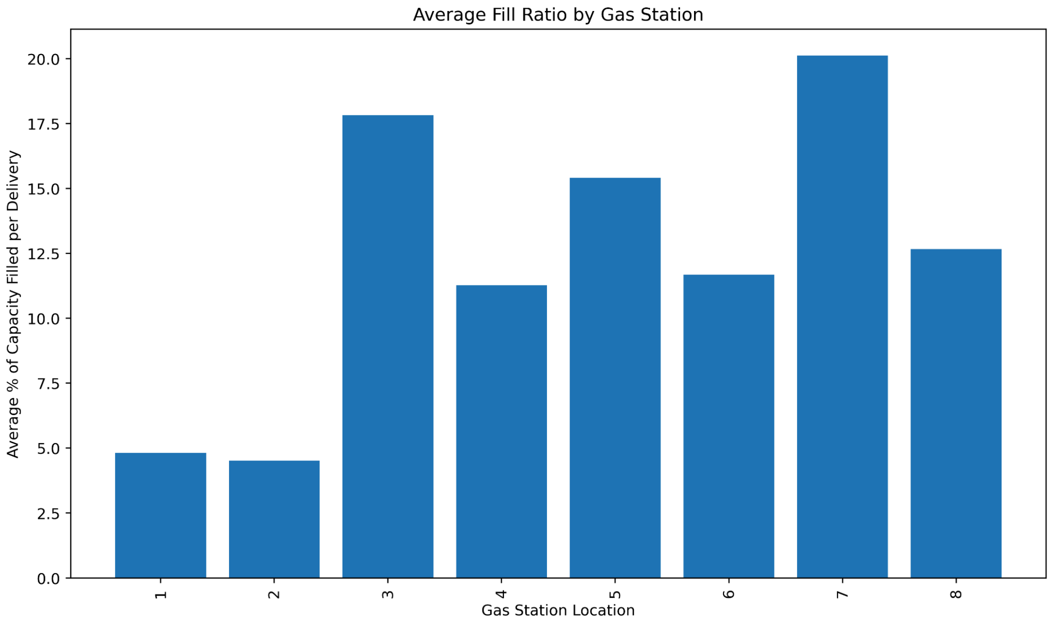
For station 1 (EastMount) and station 2 (Eastgate) exhibit extremely short intervals of 0.7 and 1.5 days, suggesting that these stations frequently reorder fuel in small quantities, which can result in unnecessary transportation costs, missed purchase discounts, and inefficient use of storage capacity, more importantly, the inefficient use of the discount.

In contrast, station 6 (Oakville), station 7 (Circle), and station 8 (Chappel) display long replenishment intervals ranging from 18 to 33 days, which means they have a relatively larger amount of purchases each replenishment. This indeed increases the possibility that they efficiently use the discount to lower the cost. However, the long replenishment increases the likelihood of low-stock risks and potential service interruptions.

In general, we consider the replenishment interval for station 3 (Central), station 4 (Chedoke) and station 5 (Mountain View) to have more moderate intervals ranging 2 to 11 days, reflecting relatively balanced refueling patterns that align better with efficient inventory management principles.

3.1.2 Average Fill Ratio by Gas Station

In this part, we will use a more quantitative metric to have a view about the replenishment situation on station level. From calculating the average fill ratio, we can have further insight into the efficiency of these replenishment behaviors. The chart showing the average fill ratio by station is shown below.



Interpret and business insights

Station 1 (EastMount) and station 2 (Eastgate) have low fill ratios, around 4.5-4.8% of total capacity per delivery, confirming that their frequent deliveries are indeed inefficient. Conversely, station 3 (Central), station 5 (Mountain View), and station 7 (Circle) achieve higher fill ratios between 15% and 20%, which shows that those 3 stations have more effective order planning that reduces fixed transportation costs per liter and minimizes operational redundancy. Station 4 (Chedoke) and station 6 (Oakville) maintain moderate fill ratios of around 11-12%, suggesting stable but suboptimal performance that could benefit from minor adjustments.

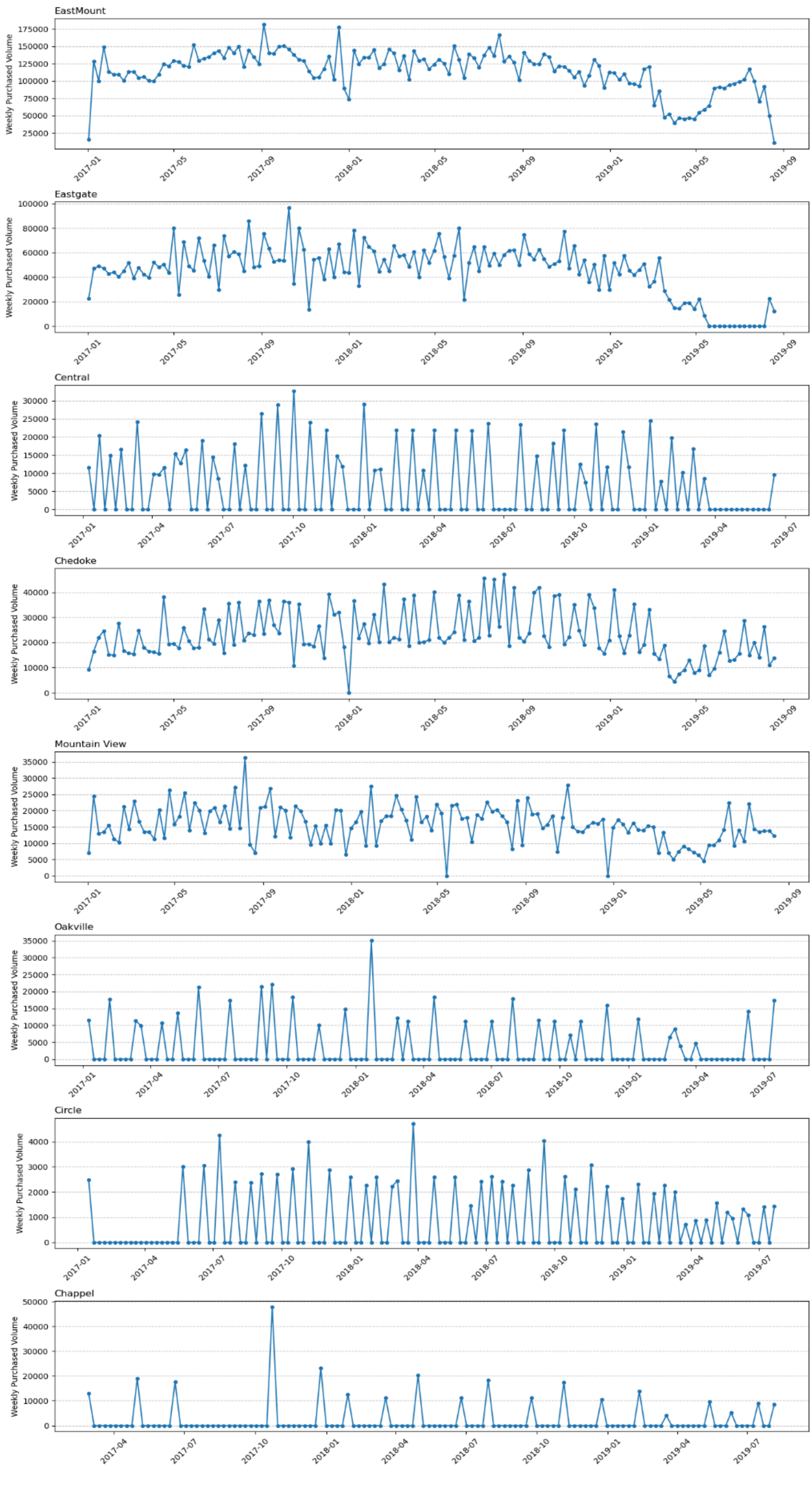
Summary

Taken together, the results demonstrate that for the general stations, current replenishment and ordering patterns lack standardization, leading to the high possibility of both over-supply and under-supply inefficiencies across the network.

3.2 Weekly & Monthly Trends

In this part, we analysed the data over time to identify temporal patterns and cross-station differences on weekly and monthly levels. Both weekly and monthly fuel purchase volumes were plotted in the next page for each of the eight gas stations - EastMount, Eastgate, Central, Chedoke, Mountain View, Oakville, Circle, and Chappel.

Weekly Fuel Purchase Trend by Gas Station



Monthly Fuel Purchase Trend by Gas Station

Interpret and Business Insights

Weekly Trend

From the chart we know the weekly trend captures short-term fluctuations in purchasing behavior, which highlights differences in operational rhythm - some sites follow disciplined replenishment cycles, while others appear reactive and irregular. EastMount and Eastgate consistently record the highest volumes, averaging 140,000 and 60,000 liters per week, with several noticeable peaks in mid-2018 that reflect periods of elevated demand or stock replenishment.

However, a visible decline emerges across both stations beginning in early 2019, indicating a potential slowdown in sales or reduced operational throughput. Chedoke and Mountain View maintain moderate but regular purchase patterns of 20,000-40,000 liters per week, showing periodic restocking cycles consistent with stable demand.

In contrast, smaller stations such as Central, Oakville, Circle, and Chappel demonstrate sporadic purchasing activities with many weeks registering zero transactions, suggesting lower fuel turnover rates or intermittent operation.

Monthly Trend

At the monthly level, aggregate trends reveal broader shifts in fuel demand and purchasing efficiency. EastMount remained the largest consumers, maintaining high monthly volumes of 400,000-600,000 liters throughout 2017 and 2018 before declining sharply in 2019. Chedoke exhibits a steady upward trajectory over 2017-2018, peaking at around 130,000 liters per month, before contracting slightly thereafter, mirroring the network-wide decline. Mountain View displays consistent monthly volumes ranging from 60,000 to 90,000 liters, implying balanced inventory control and regular refilling schedules. Meanwhile, smaller stations such as Circle maintain extremely low purchase levels, below 10,000 liters per month, often with long intervals between orders, which is typical for low-traffic or backup fueling sites.

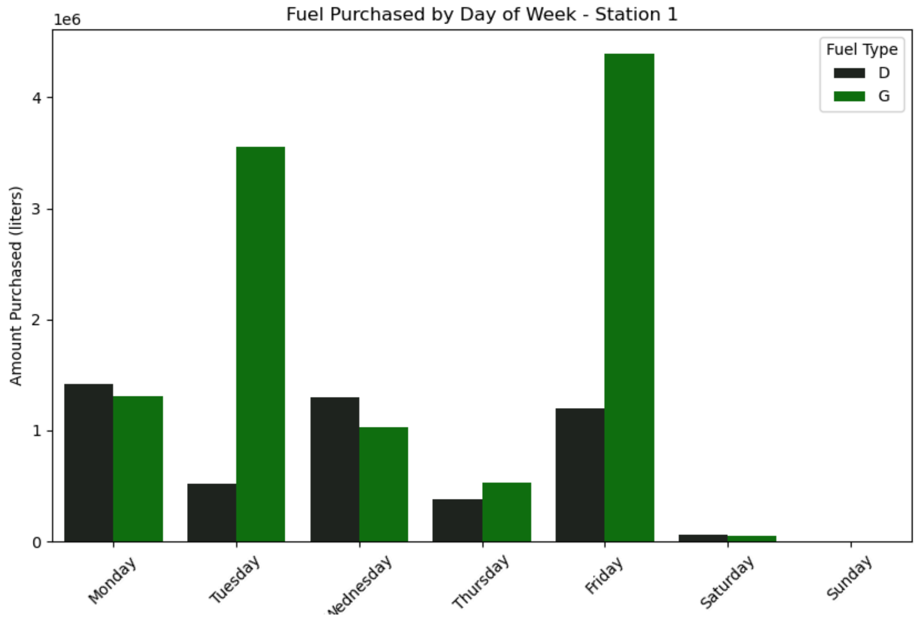
Summary

Overall, the combined weekly and monthly visualizations provide strong insight into the temporal structure and efficiency of fuel purchasing across the network. The synchronization between EastMount and Eastgate suggests shared supply dependencies or similar demand dynamics, while the overall reduction in 2019 across nearly all stations points to a system-wide contraction - possibly due to macroeconomic shifts, pricing changes. Moreover, the variation in purchase regularity underscores differing levels of inventory management discipline across stations which shows high-volume locations maintain predictable cycles, smaller locations rely on ad hoc replenishment.

3.3 Quantity purchased by day of the week for each gas station

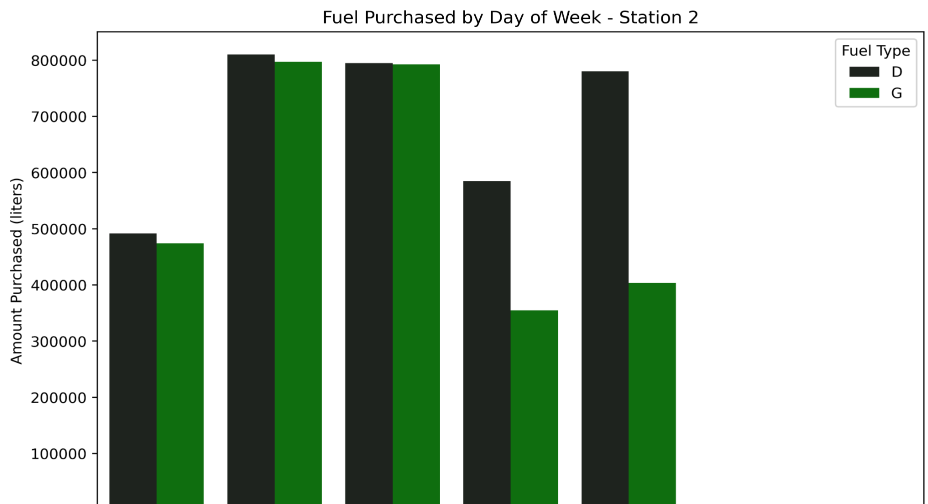
In this stage, we examined how the gross purchase cost varied for each station on the weekday level to identify potential cost-saving patterns based on different days of the week. The dataset was grouped by the variable "day_of_week," and the mean gross purchase cost per day was calculated. Bar charts are created to further see the situation of fuel purchased by day of week categorized by different fuel types like G stands for gasoline, D stands for diesel.

Station 1 (EastMount)



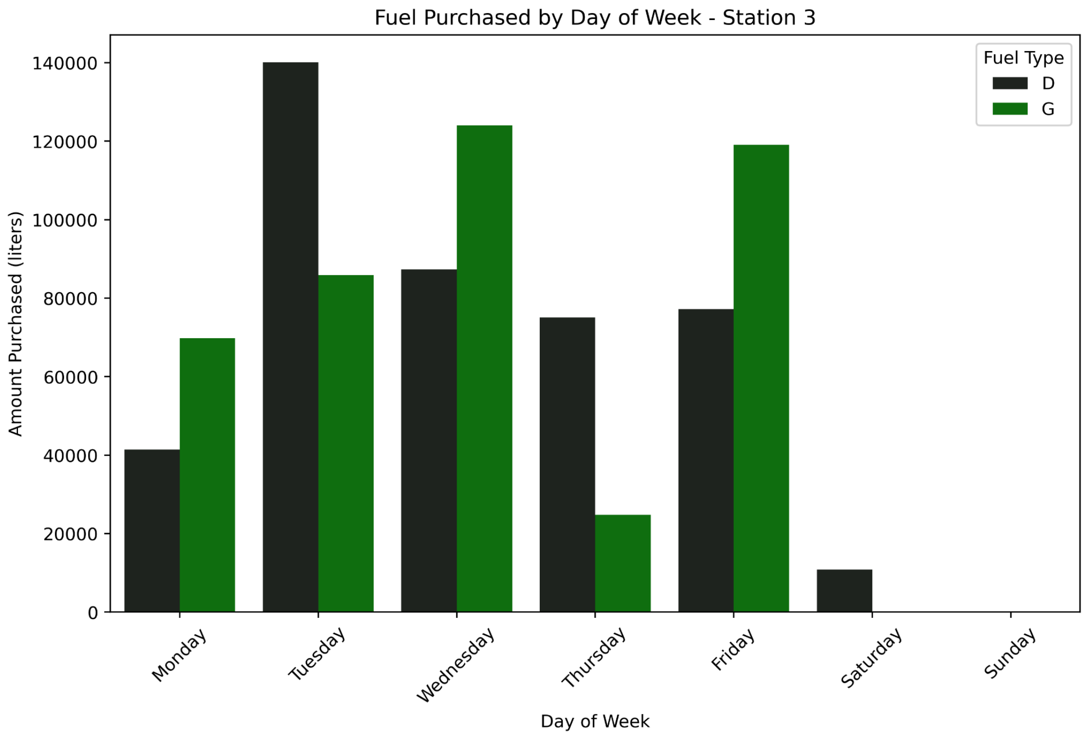
From the chart, we know that station 1 shows strong variability in weekly purchasing behavior, across most days. The top 2 highest gasoline volumes occurred on Friday (≈4.3 million L) and Tuesday (≈3.5 million L), suggesting concentrated bulk purchases twice per week in the operation cycle. For the diesel, the volumes remain relatively stable (≈1.2-1.4 million L) on weekdays, with minimal weekend activity. This dual-peak structure implies coordinated refueling cycles aligned with high turnover days.

Station 2 (Eastgate)



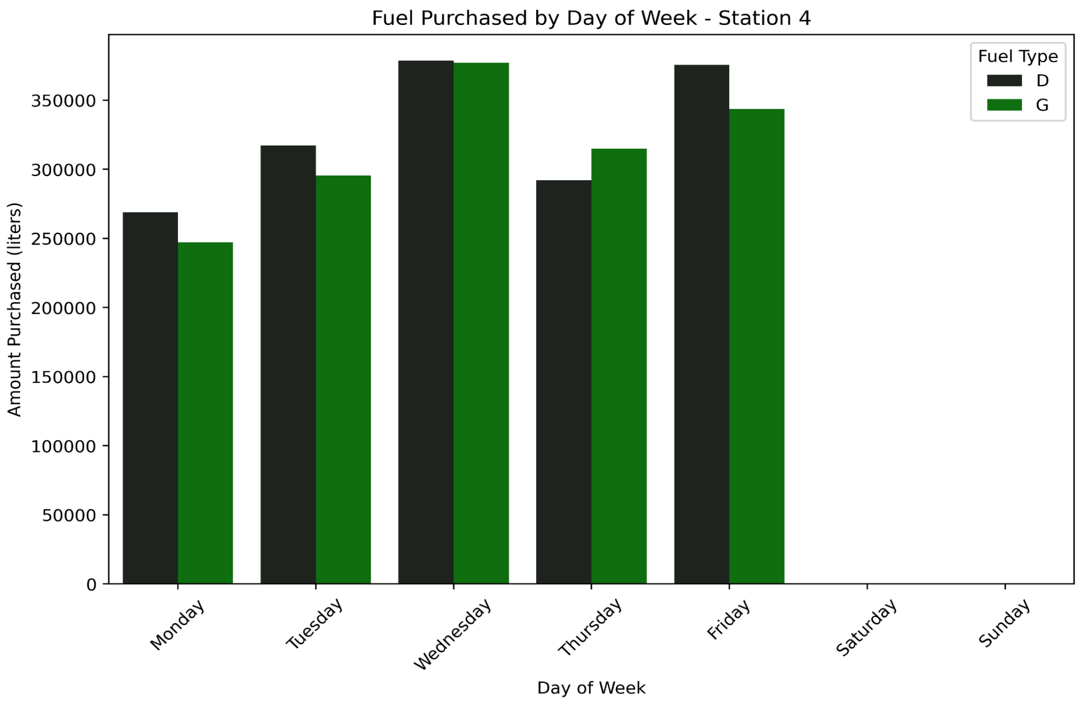
From the chart, we can see that station 2 demonstrates a balanced purchasing pattern between gasoline and diesel peaking on Tuesday and Wednesday (≈0.8 million L each). Purchases decline toward Thursday and Friday, and nearly cease on weekends. The consistent weekday demand reflects steady throughput and possibly synchronized delivery scheduling for both fuel types.

Station 3 (Central)



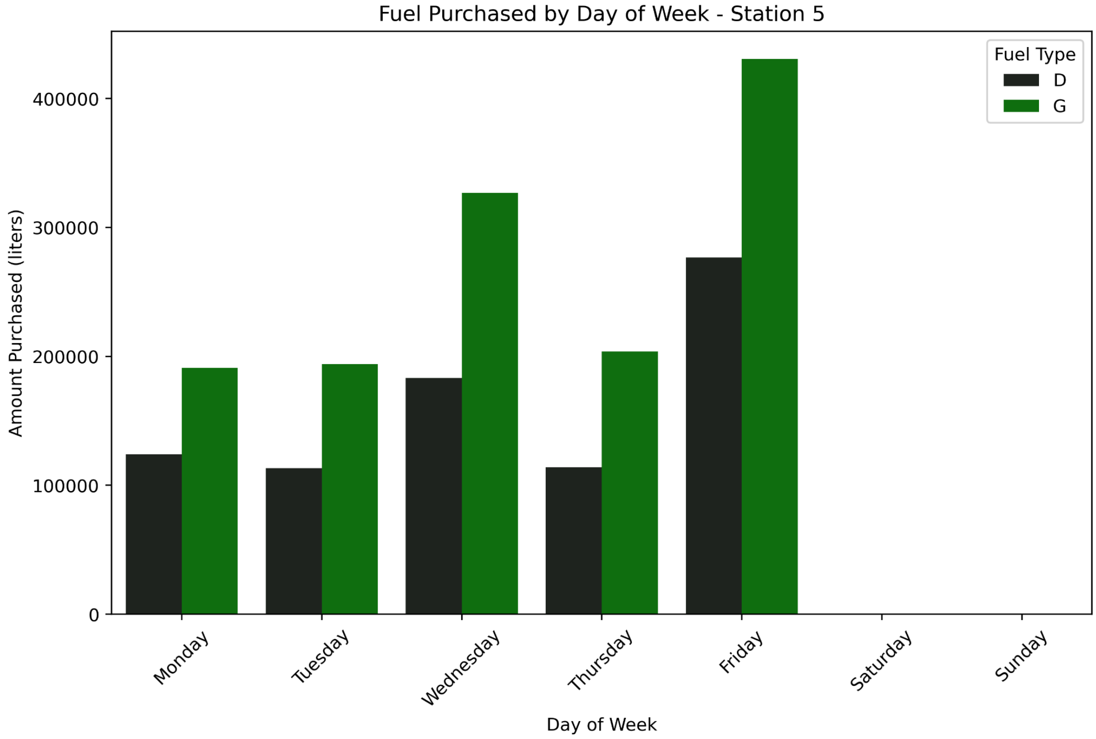
At Station 3, diesel purchases dominate early in the week, peaking on Tuesday (≈140 k L), while gasoline peaks mid-to-late week (Wednesday-Friday, 120-130 k L). The alternating pattern suggests operational segmentation of the two types of fuel.

Station 4 (Chedoke)



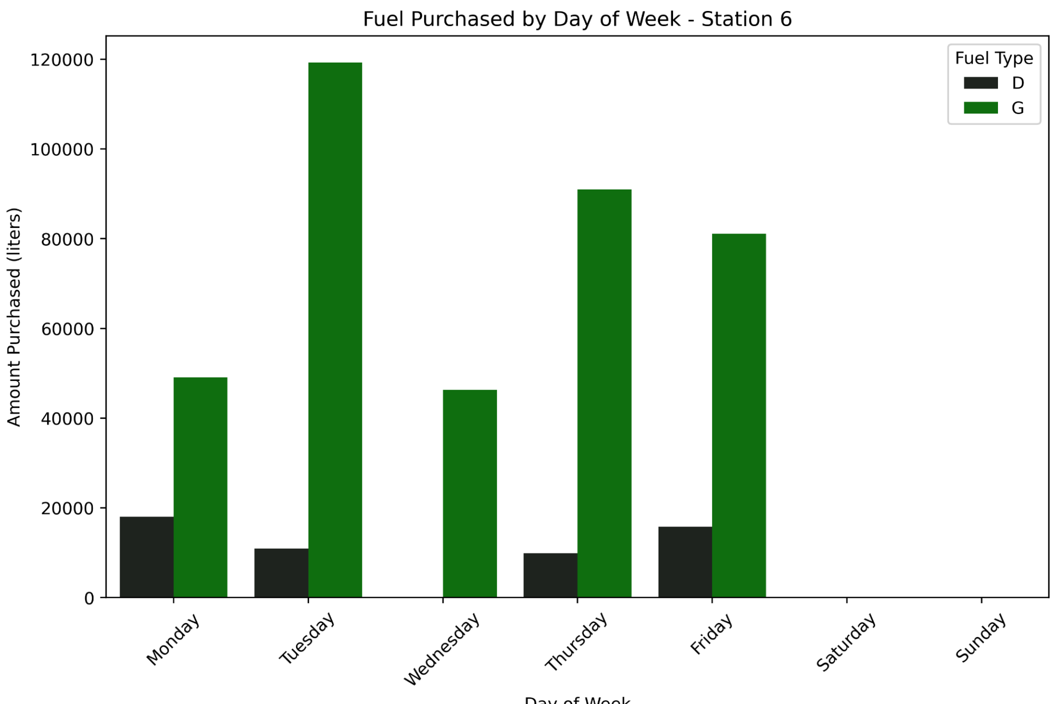
During the weekday, station 4 exhibits even distribution, with both fuel types maintaining 250-370k purchases per day. Slightly higher activity on Tuesday and Friday indicates a possible twice a week refueling rhythm. The station’s balanced pattern and absence of strong spikes suggest disciplined procurement and stable demand.

Station 5 (Mountain View)



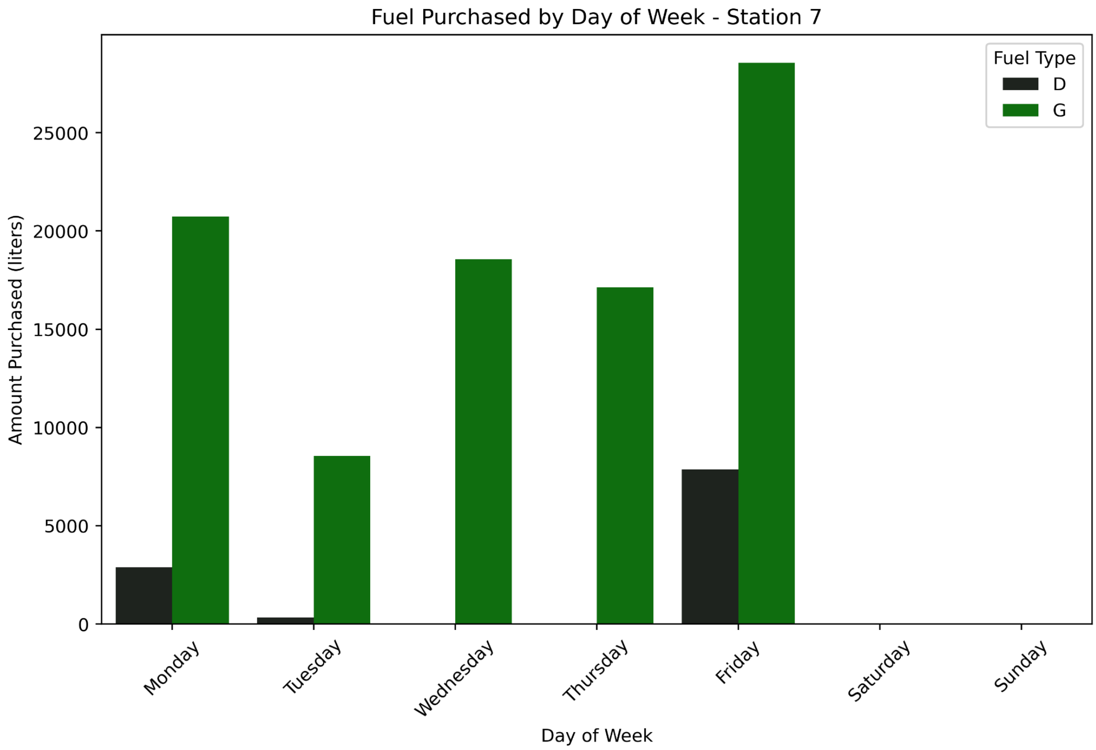
Station 5 shows clear gasoline predominance, particularly on Friday, when gasoline purchases exceed 400 k L, compared to around 300 k L diesel. Monday through Thursday volumes gradually rise, indicating progressive stock depletion and end-of-week replenishment.

Station 6 (Oakville)

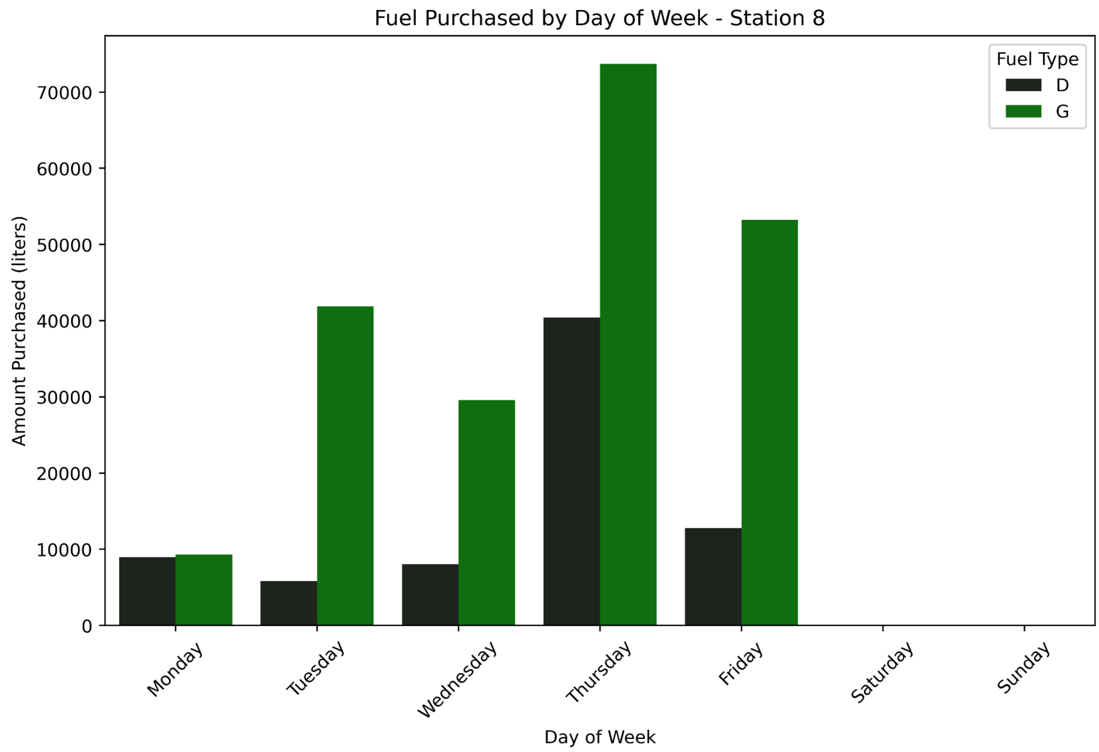


Station 6 record shows significant gasoline bias, with volumes between 50k L (Monday) and 120k L (Tuesday), while diesel remains consistently low (<20k L). Tuesday and Friday stand out as primary purchasing days, showing a weekly rhythm for gasoline replenishment.

Station 7 (Circle)



Similar to Station 6, the purchase pattern for Station 7 is also gasoline-based, reaching the peak around 26k L on Friday. Diesel purchases are minimal, occurring sporadically.



Station 8 (Chappel)

Station 8 also follows a similar gasoline-dominant pattern, with Thursday being the top amount purchased day followed by smaller orders on Tuesday and Friday. Diesel consumption remains moderate (≈40 k L on Thursday), pointing to occasional replenishment rather than continuous demand.

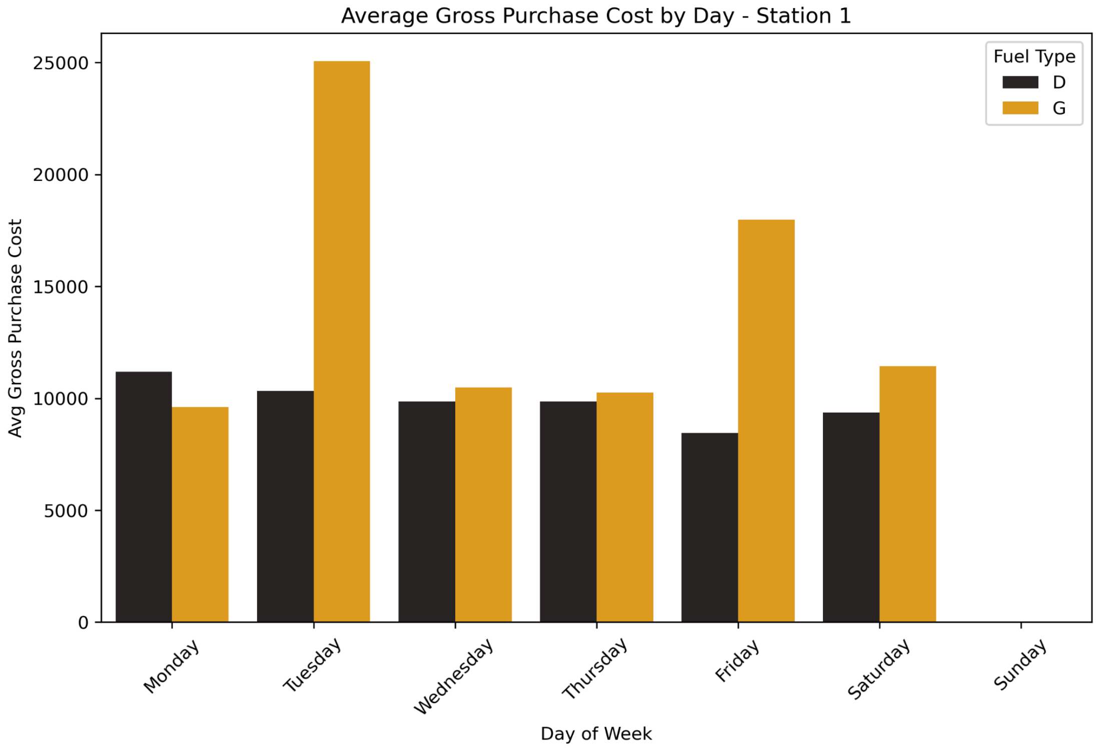
Summary

Across all eight stations, purchasing activity is heavily concentrated on weekdays, with negligible weekend transactions. Gasoline purchases tend to peak between Tuesday and Friday, while diesel demand is more evenly spread but smaller. This temporal clustering highlights a weekday operational cycle. Aligning large-volume gasoline orders with lower-cost days identified in earlier analyses could further enhance cost efficiency.

3.4 Average Gross Purchase Cost by Day Analysis

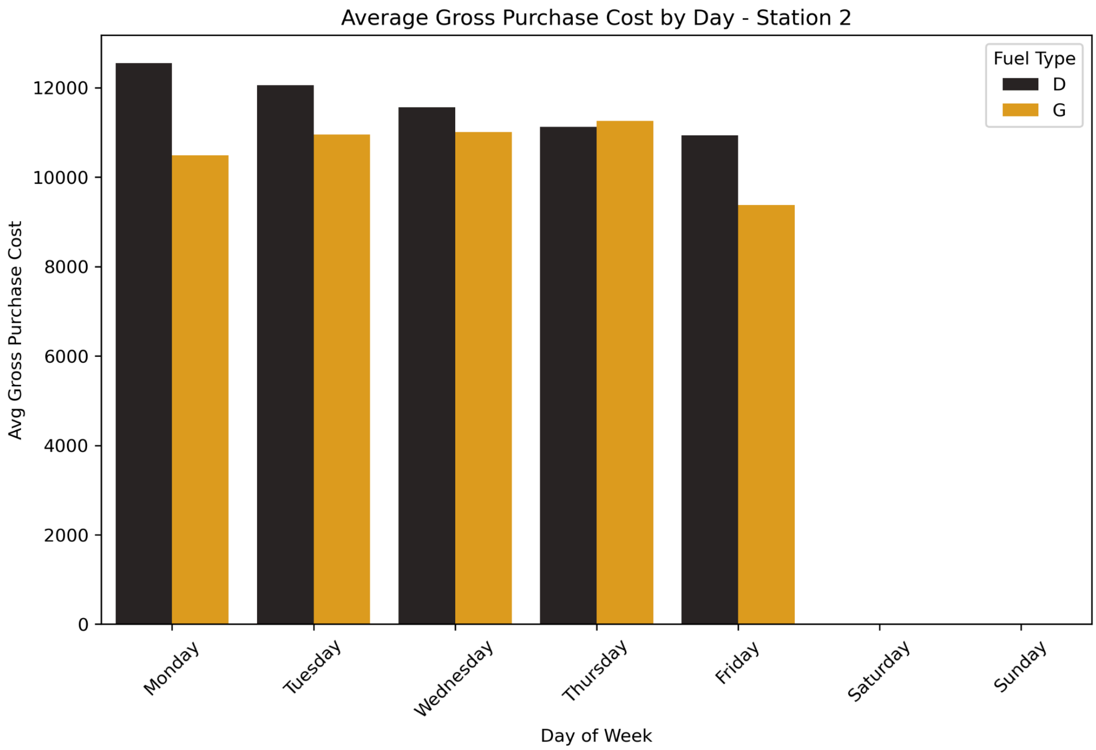
To further understand the purchasing intensity, we analyzed the average quantity of fuel purchased per weekday across all stations. For the analysis, we grouped the dataset by the variable “day_of_week,” and the mean purchase volume was computed for each day based on the fuel type. Bar charts were generated to visualize weekly purchasing patterns categorized by different fuel type like G stands for gasoline, D stands for diesel, and have comparison between days

Station 1 (EastMount)



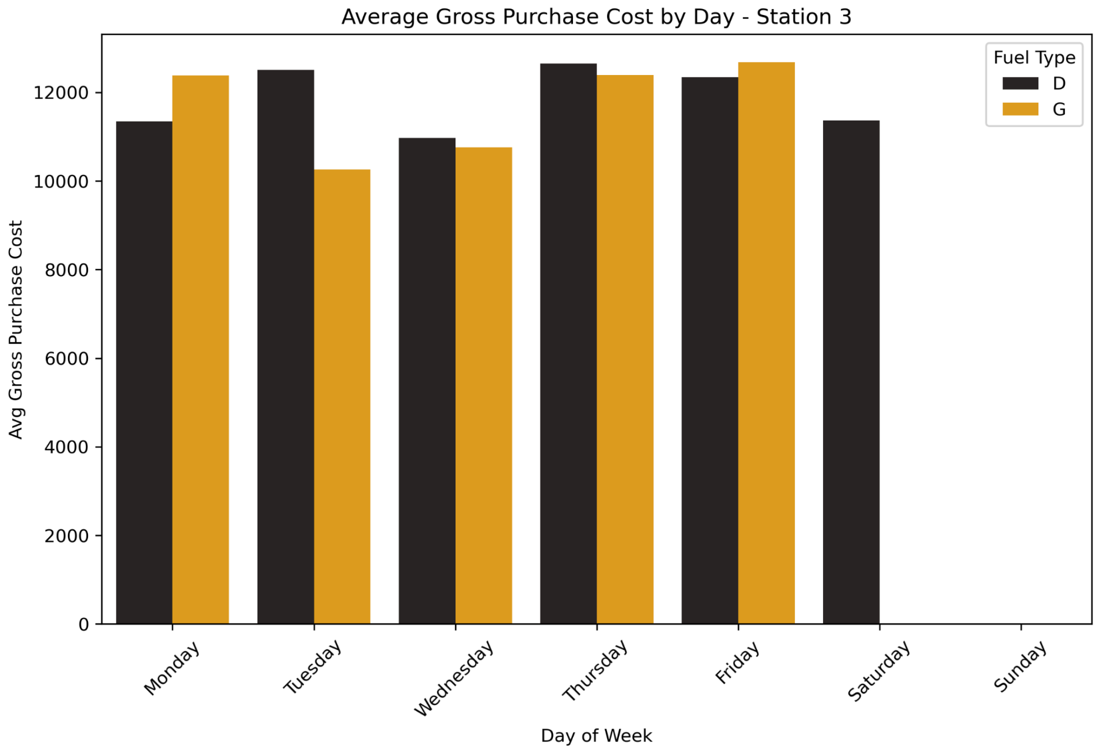
For station 1, we can see there are significant weekly routines, with gasoline dominating total purchase cost. The highest gasoline cost occurs on Tuesday of ~25,000 CAD and the second highest gasoline occurs on Friday of ~18,000 CAD, besides those two days, the rest of weekdays generally remain around 10,000-11,000 CAD. Meanwhile, costs for diesel are steadier, fluctuating between 9,000-11,000 CAD and reaching a maximum on Monday.

Station 2 (Eastgate)



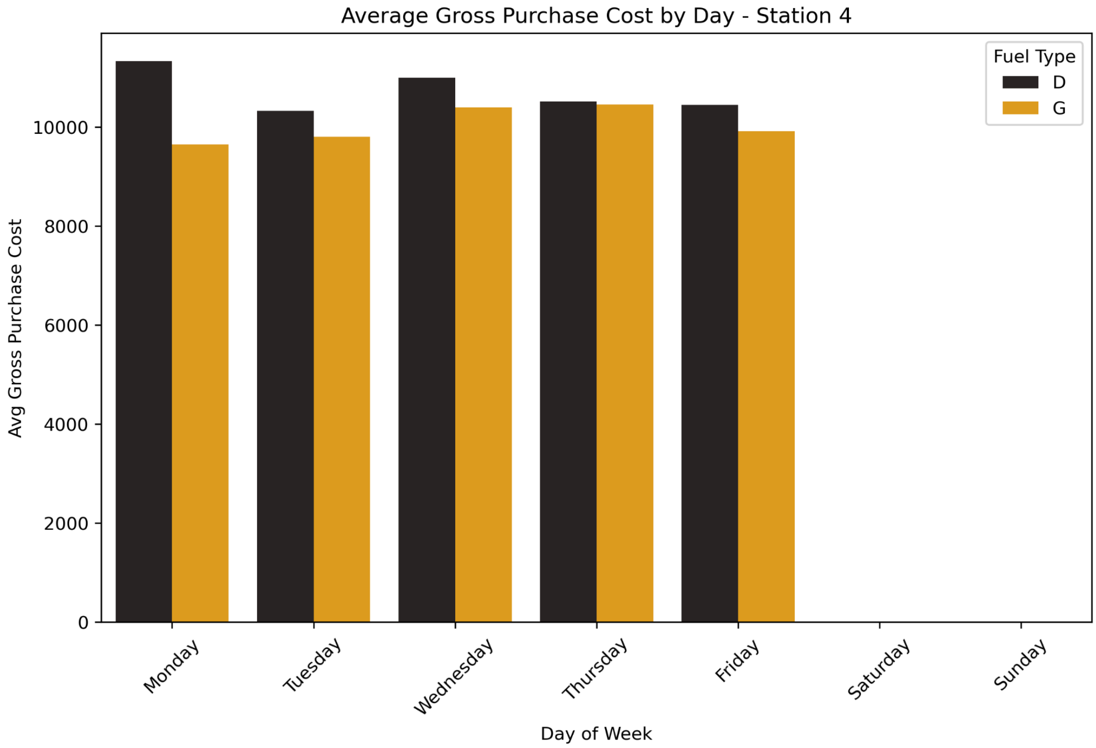
Unlike station 1, the pattern for station 2 shows more balance, with Monday being the largest gross purchase for gasoline and Thursday being the largest gross purchase for diesel. For both types of fuel, the cost remains within an overall narrow range, especially for the diesel cost. before tapering off toward Friday. The pattern indicates mid-week procurement concentration and consistent weekday demand.

Station 3 (Central)



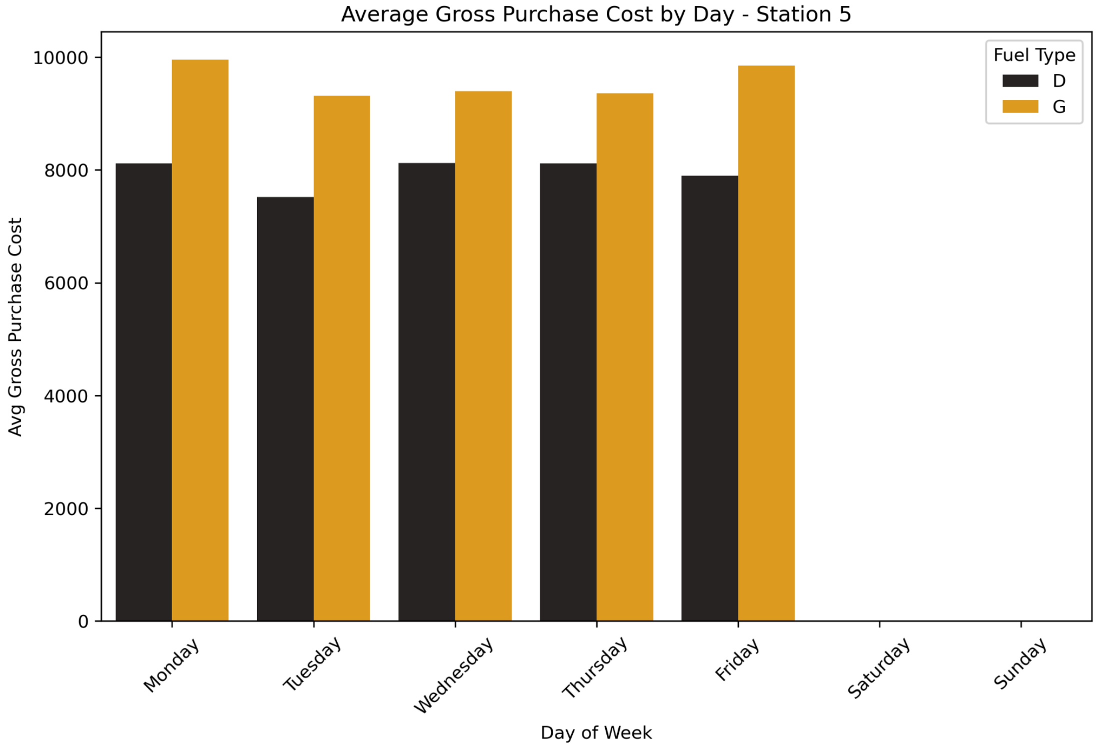
For station 3, the purchase costs for diesel peaks on Tuesday of 12.5 k CAD, while the purchase costs for gasoline increase from Wednesday to Friday. This pattern likely reflects operational segmentation.

Station 4 (Chedoke)



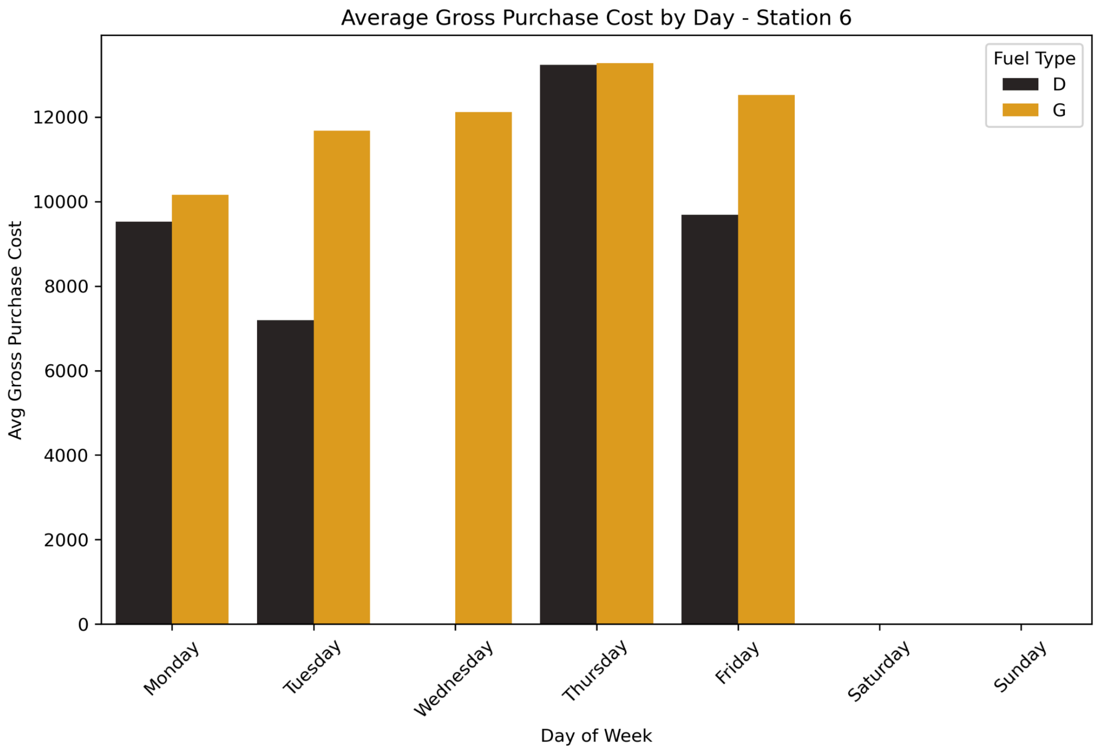
For station 4, the patterns for both diesel and gasoline are quite stable, with both diesel and gasoline averaging between 9.8 k and 11.3 k throughout the week. Diesel costs show slight increases on Monday and Wednesday, while the cost for gasoline becomes higher mid-week.

Station 5 (Mountain View)



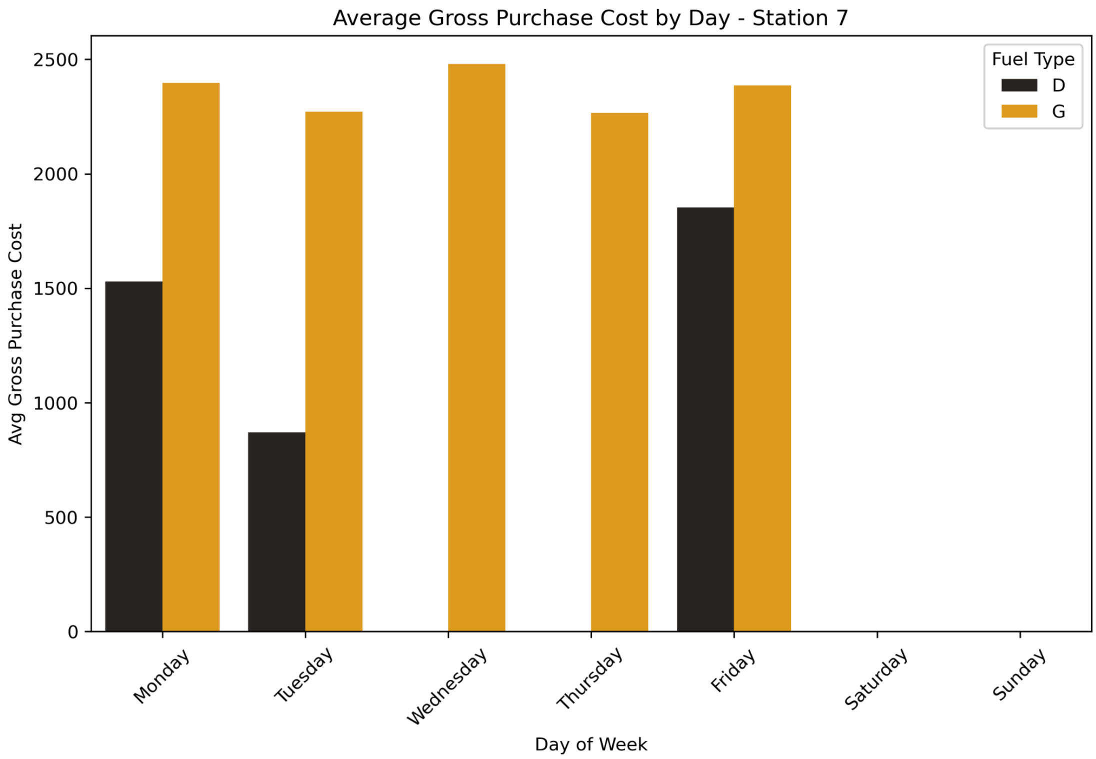
Station 5 shows a gasoline-dominant pattern, with the highest costs on Friday, followed by Monday, and lower levels mid-week. For the diesel gross purchase cost, the pattern is quite stable, and shows less difference the entire week.

Station 6 (Oakville)



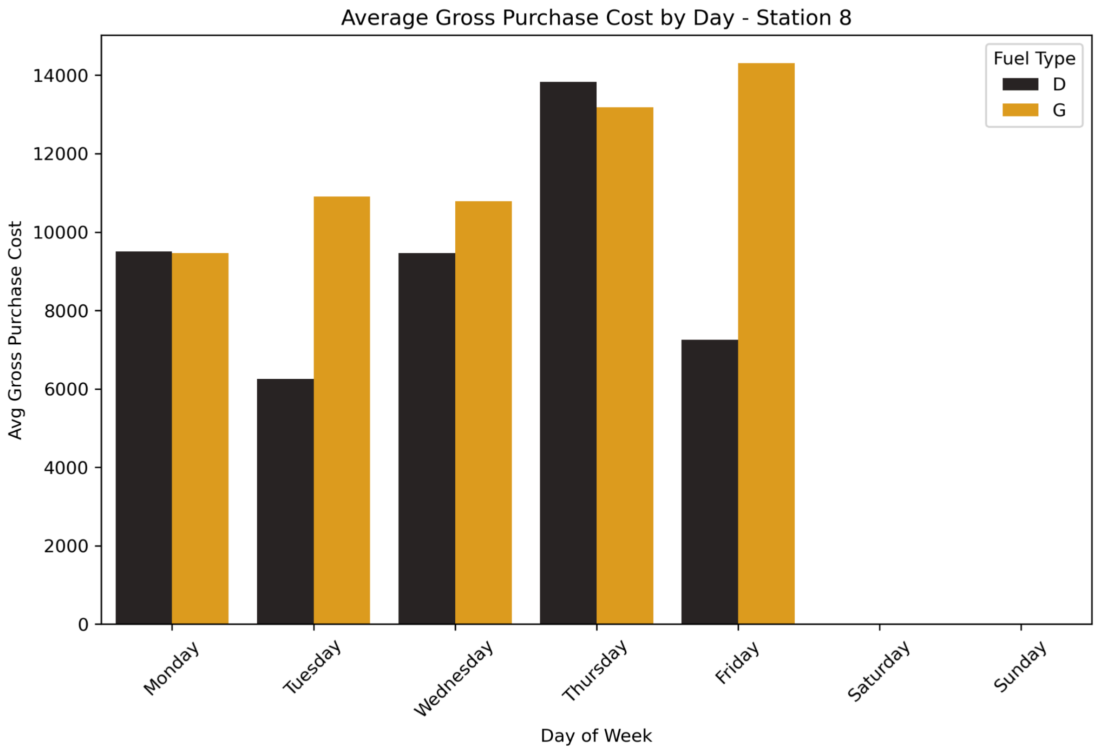
For station 6, we can see a very gasoline-dominated pattern. On Wednesday, there is no purchase on diesel, while the purchase for gasoline remains high. The average gross purchase reaches the peak on Thursday. For both gasoline and diesel, they both reach the peak on Thursday (≈13 k CAD) and drop thereafter. This suggests that most orders are placed late in the workweek.

Station 7 (Circle)



The pattern for station 7 shows small purchase volumes in general and largely gasoline-based (~2.3-2.5 k CAD across weekdays). Diesel purchases are minimal and sporadic, appearing mainly on Monday, Tuesday, and Friday. Given its low throughput, the station's costs are more sensitive to timing.

Station 8 (Chappel)

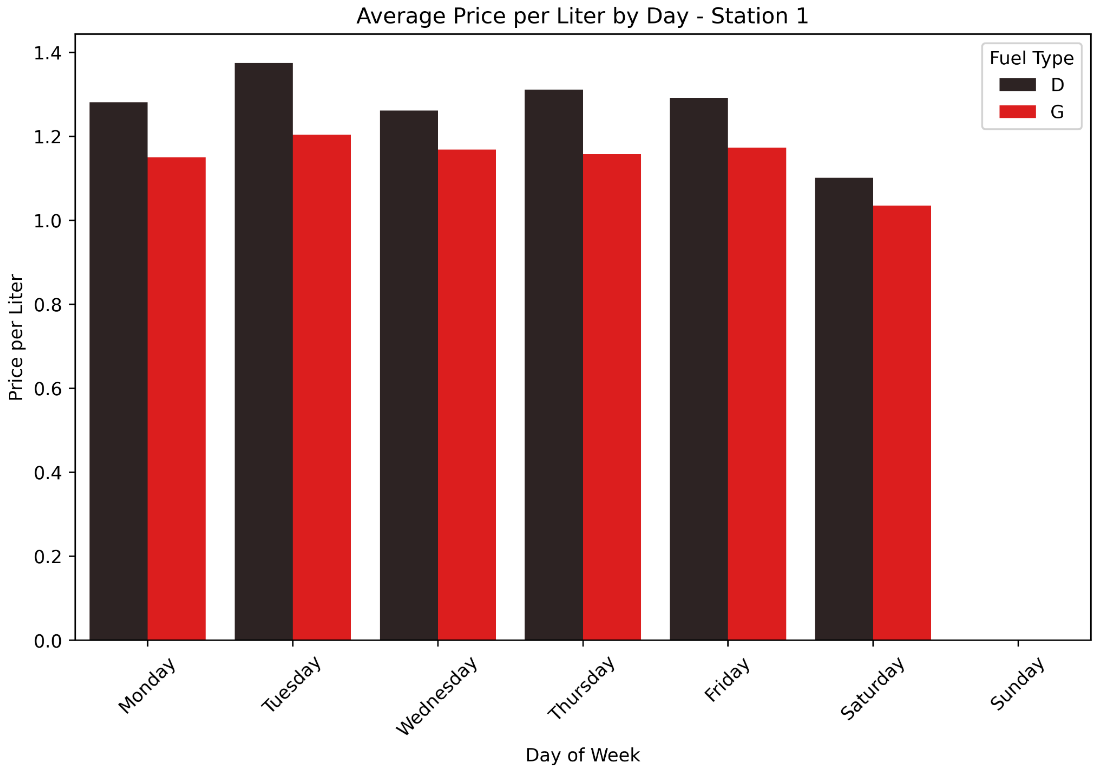


Station 8 displays late-week peaks - gasoline reaches its maximum on Friday, while diesel peaks on Thursday (~13.8 k CAD). Moderate costs appear mid-week (Tuesday-Wednesday). This indicates synchronized end-week replenishment for both fuel types.

3.5 Price per Liter Average by day of the week

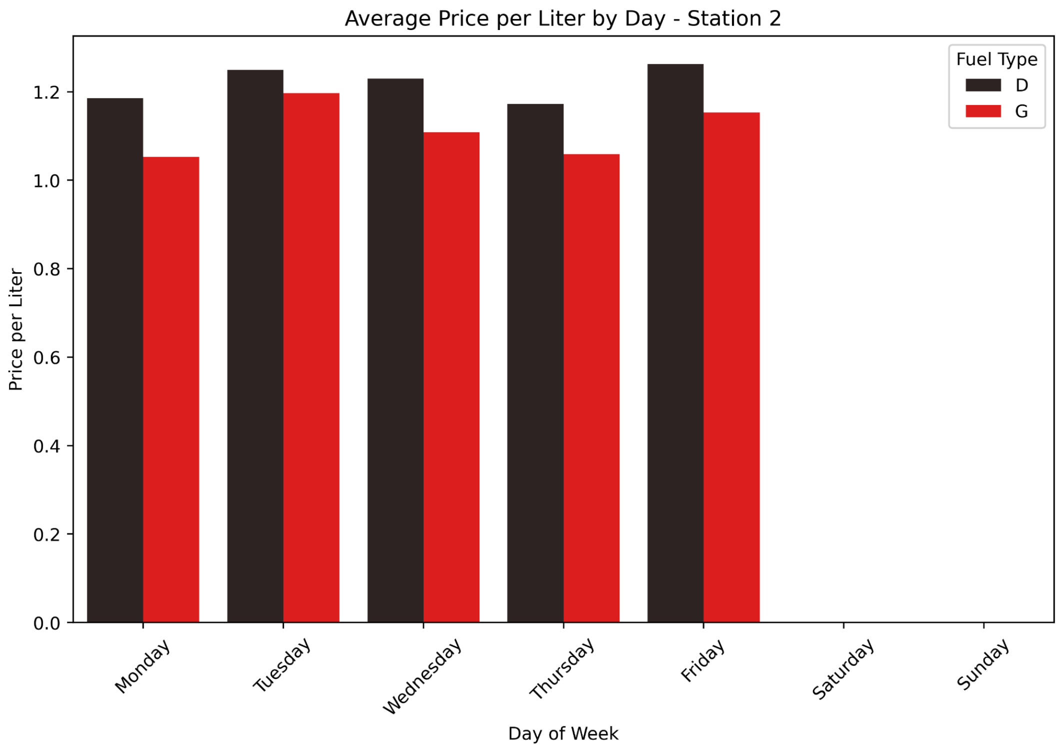
To explore station-level pricing dynamics, we further grouped the dataset by station, weekday, and fuel type. For each combination, the mean price per liter was computed and plotted using grouped bar charts. This approach allowed us to compare diesel and gasoline prices across days within each station.

Station 1 (EastMount)



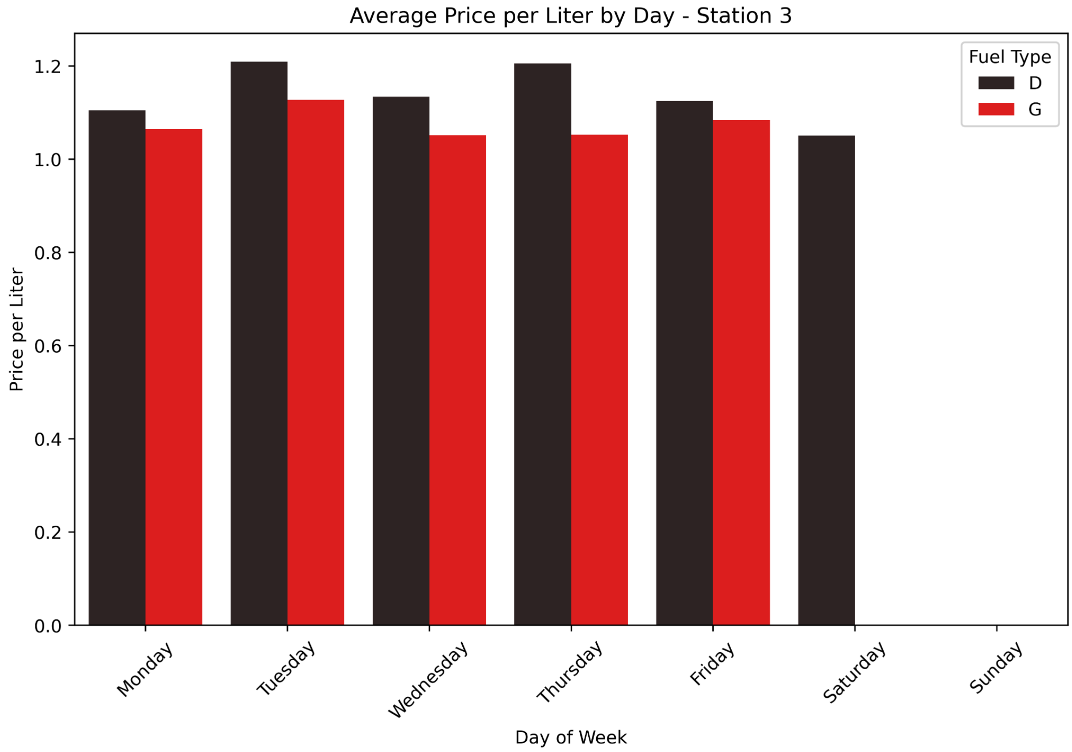
Fuel prices at Station 1 remain relatively stable throughout the week, averaging between 1.0-1.4 CAD/L. Diesel is priced slightly higher than gasoline, with a peak on Tuesday of 1.38 CAD/L and a minor decline toward the weekend.

Station 2 (Eastgate)



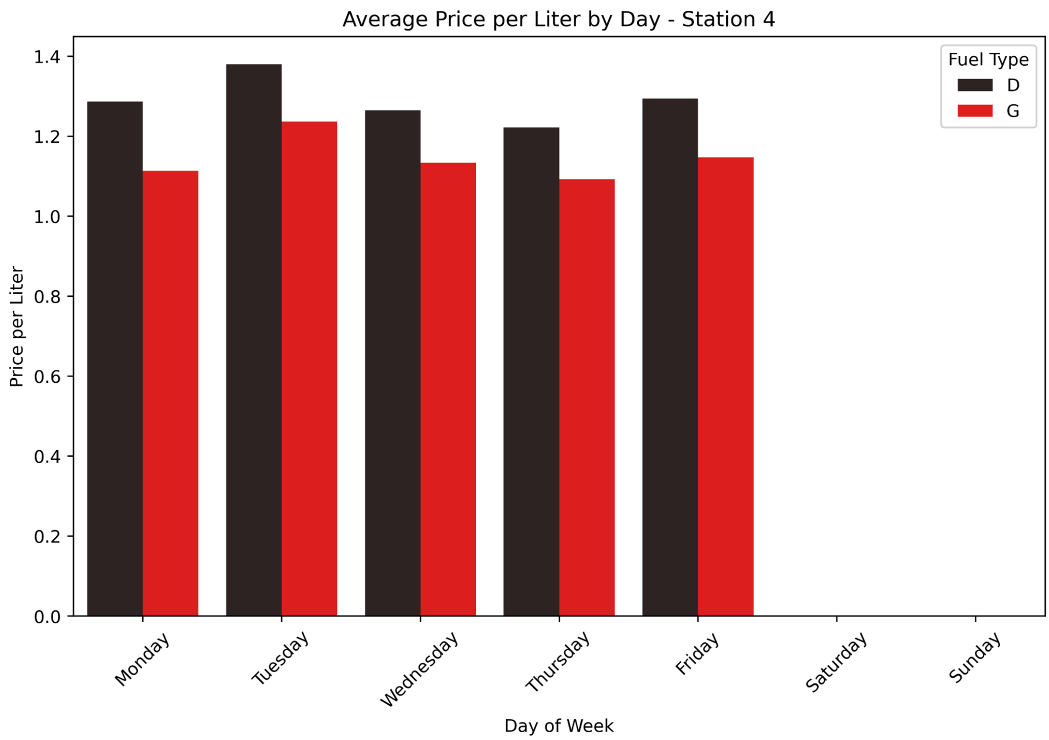
Similar to station 1, both fuel types show moderate fluctuations with diesel being is slightly more expensive, reaching its peak point on Friday, while gasoline stays relatively consistent around 1.1-1.2 CAD/L.

Station 3 (Central)



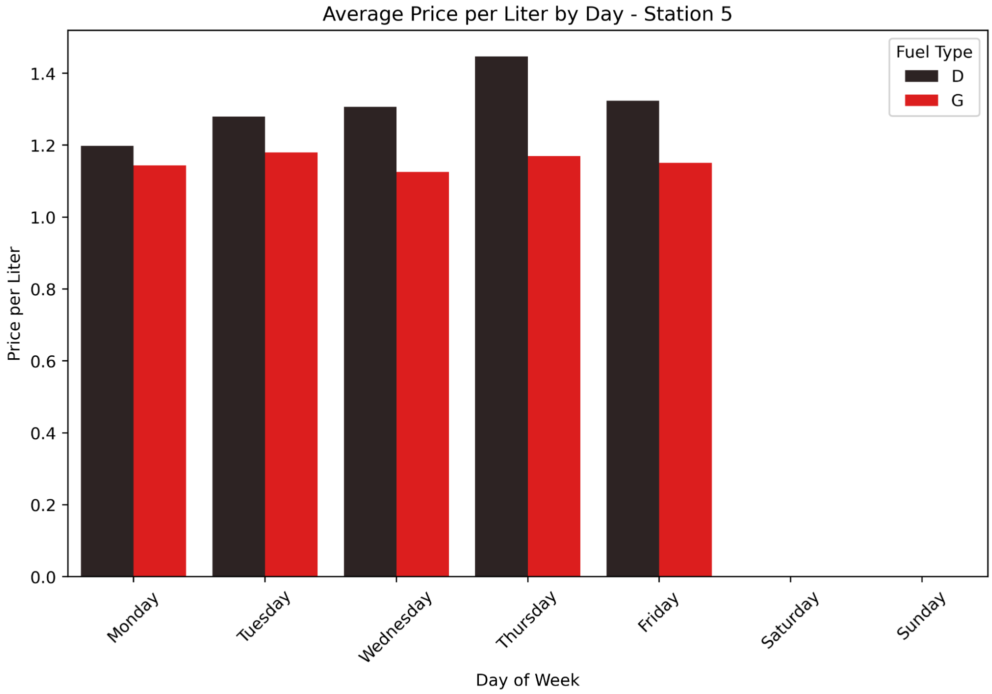
Station 3 shows limited daily variation, with both fuels ranging between 1.05-1.22 CAD/L. Diesel peaks slightly on Tuesday, while gasoline remains steady or slightly lower midweek.

Station 4 (Chedoke)



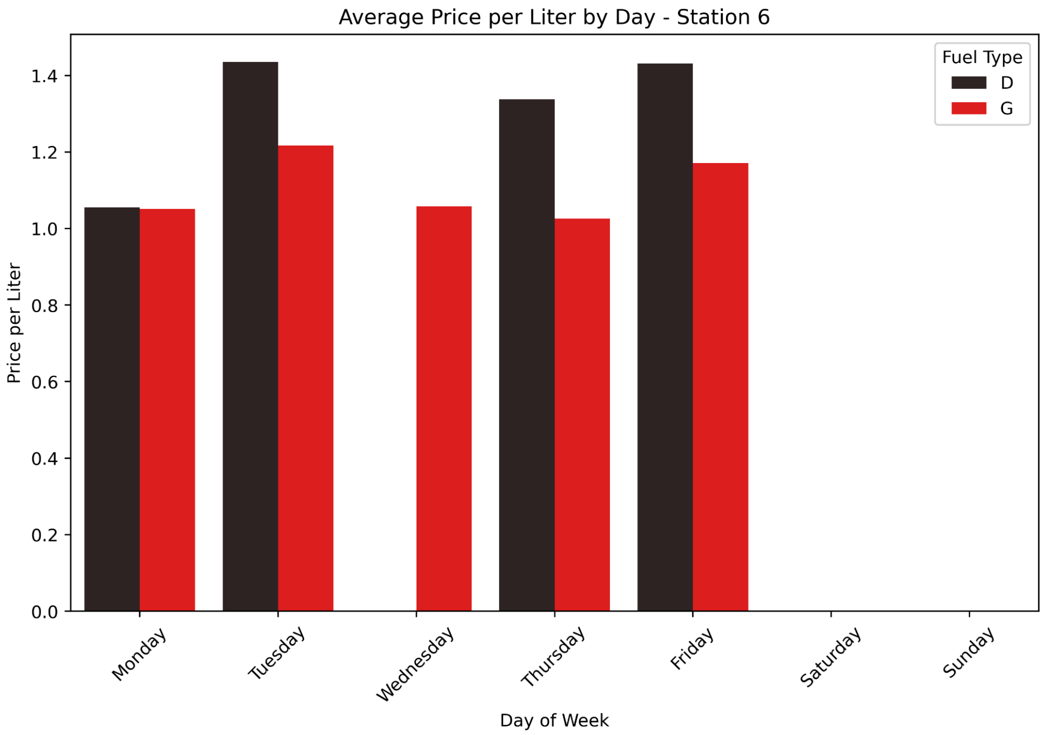
Fuel prices at Station 4 are slightly higher than the rest stations, ranging between 1.1-1.4 CAD/L. The average price for diesel remains higher than gasoline, with Tuesday being the day with the highest price, while Wednesday and Thursday show relatively lower prices.

Station 5 (Mountain View)



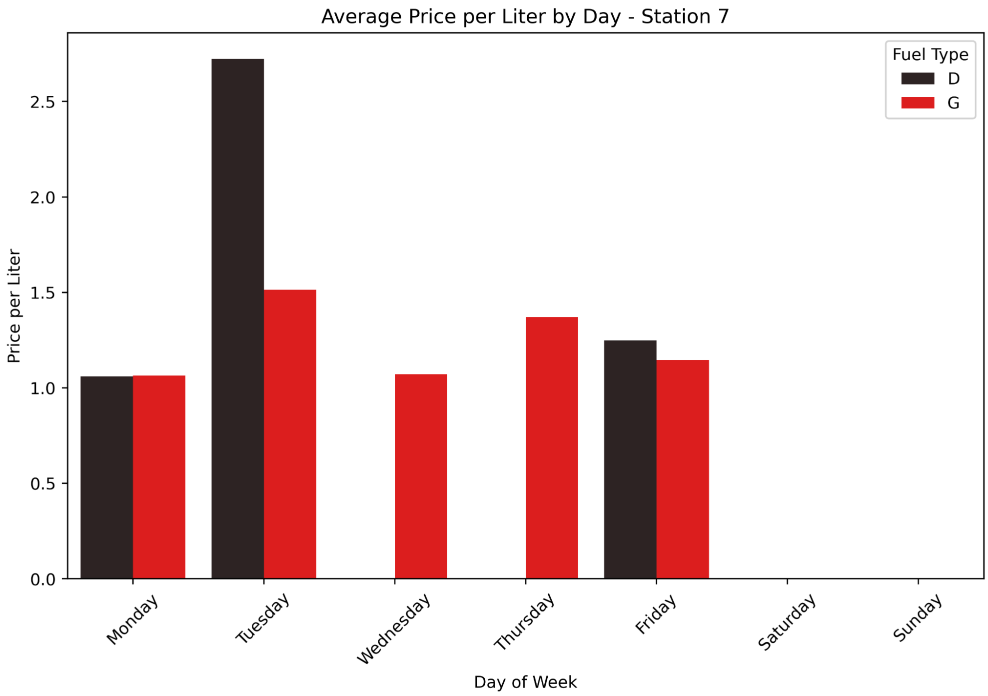
The price for diesel in station 5 reaches its highest level on Thursday (~1.45 CAD/L), while gasoline stays in the range of 1.15-1.18 CAD/L. The difference between the two fuel types is more obvious than at most other stations.

Station 6 (Oakville)



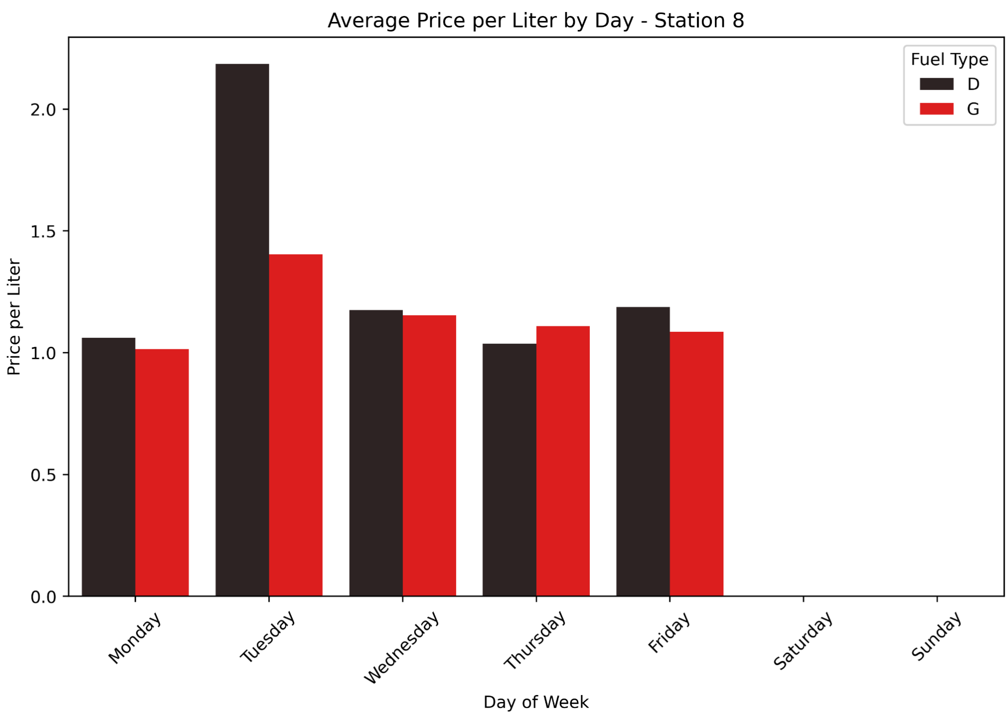
From the chart, we can see the diesel price on Wednesday remains 0, which aligns with the previous analysis. The highest average price is on Tuesday and Friday around 1.44 CAD/L, and the highest average gasoline price happens on Tuesday around 1.22 CAD/L.

Station 7 (Circle)



Station 7 shows the largest volatility among all stations. To start with, the diesel prices for both Wednesday and Thursday are missing, but the diesel prices spike sharply on Tuesday around 2.7 CAD/L, while the price for gasoline also rises around 1.5 CAD/L. Such sharp changes suggest small transaction volumes and unstable supply conditions.

Station 8 (Chappel)

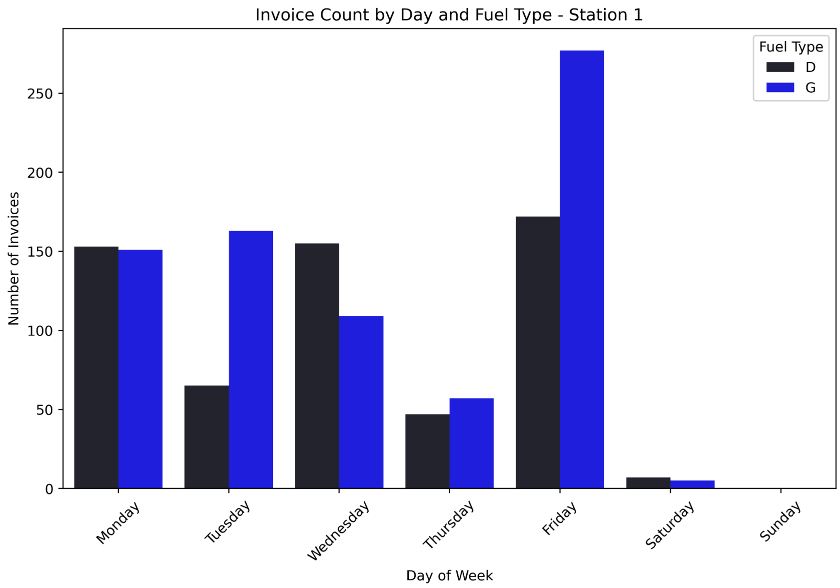


Station 8 displays a noticeable price spike for diesel on Tuesday. Prices for both fuels remain stable for the rest of the week, averaging between 1.0-1.2 CAD/L.

3.6 Number of invoices by day of the week

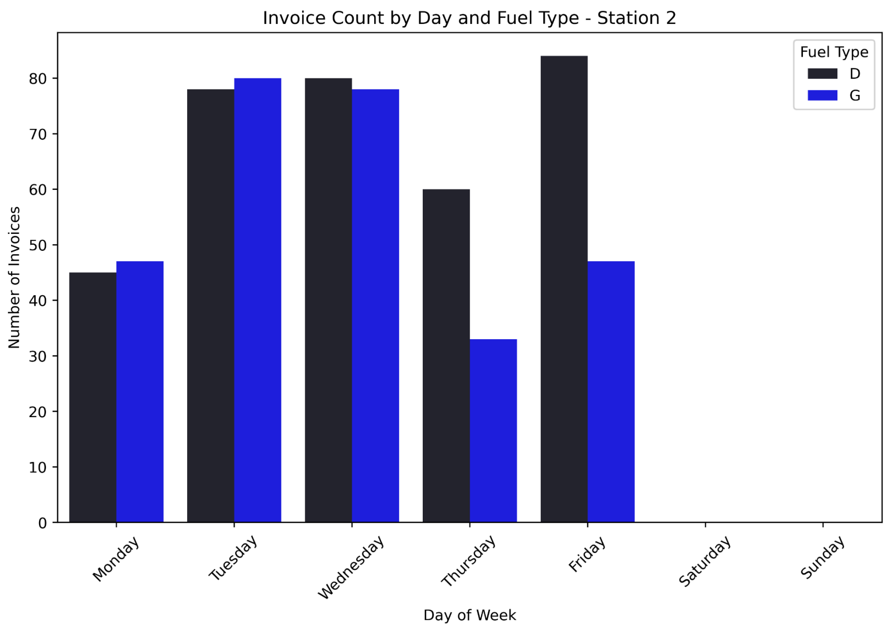
In this part, we analysed transaction frequency by counting the number of invoices issued per station, fuel type, and day of the week. The data were grouped accordingly, and the counts were visualized through bar charts for each station with the fuel types divided into D for diesel and G for gasoline.

Station 1 (Eastmount)



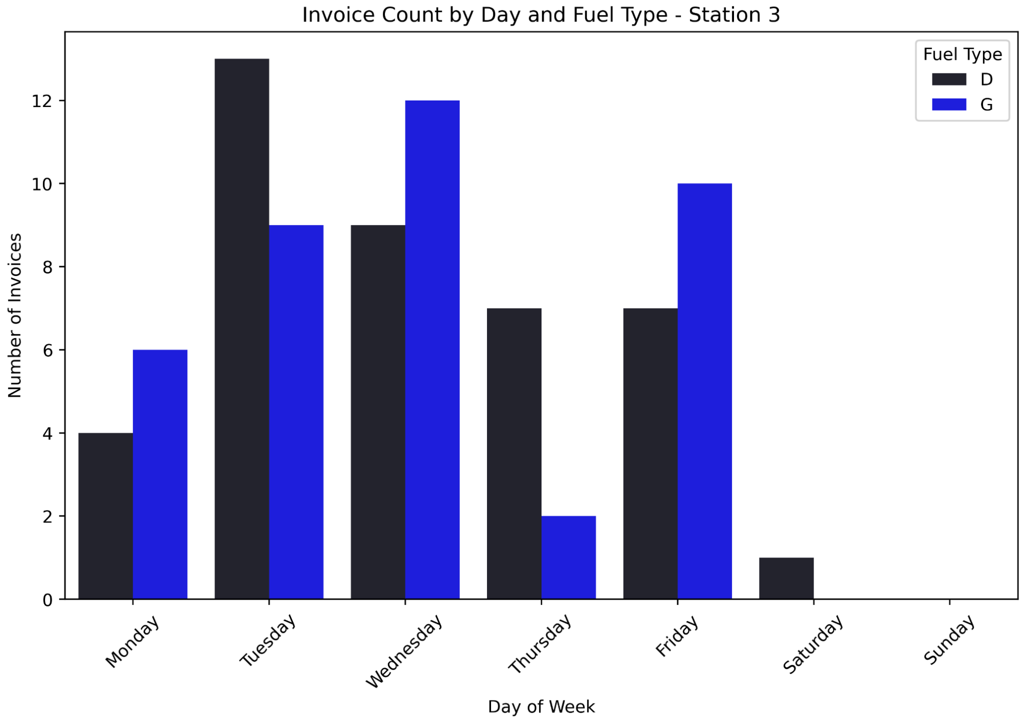
Station 1 exhibits a high overall invoice volume, particularly for gasoline. The number of invoices for gasoline on Friday is around 270. The invoice number for diesel reaches the top on Friday around 170. Monday also records considerable activity for both fuels, while Saturday shows minimal transactions.

Station 2 (Eastgate)



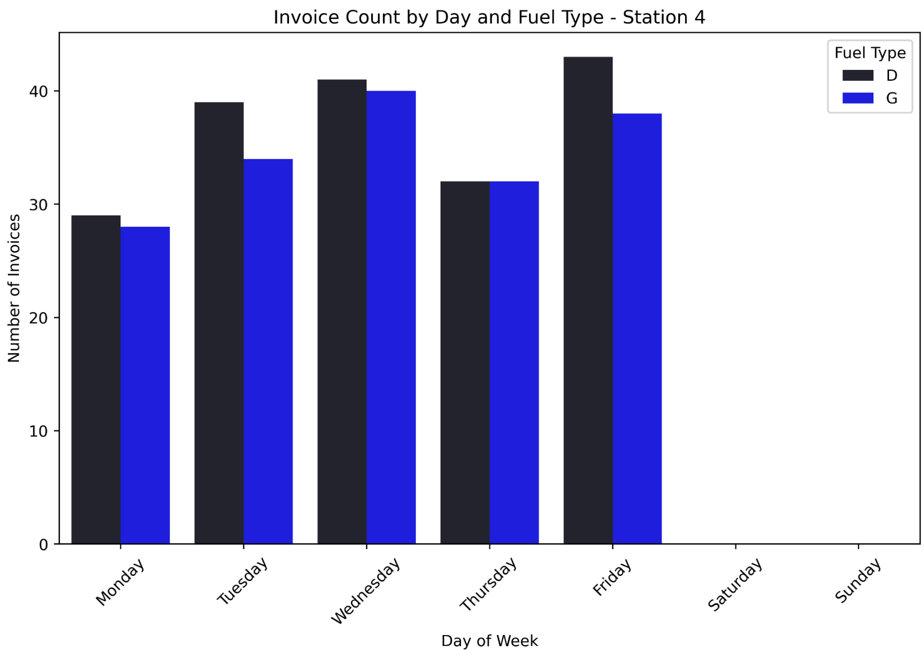
At Station 2, the biggest invoice number for diesel happens on Friday, the peak invoice number for gasoline happens on Tuesday. The invoice number for diesel maintains a slightly higher count than gasoline on most days. Both fuels show limited activity on weekends.

Station 3 (Central)



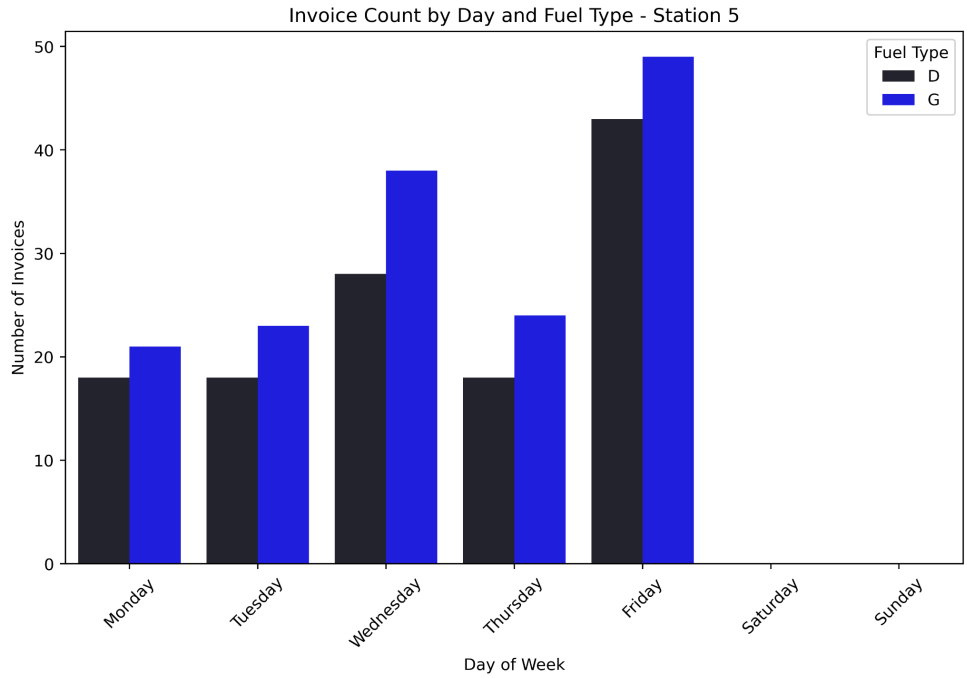
From the chart, the invoice number for diesel reached a peak on Tuesday, and the invoice number for gasoline reached the peak on Wednesday. There are still some diesel activities on Saturday. For both Saturday and Sunday, there is no sign of gasoline activity.

Station 4 (Chedoke)



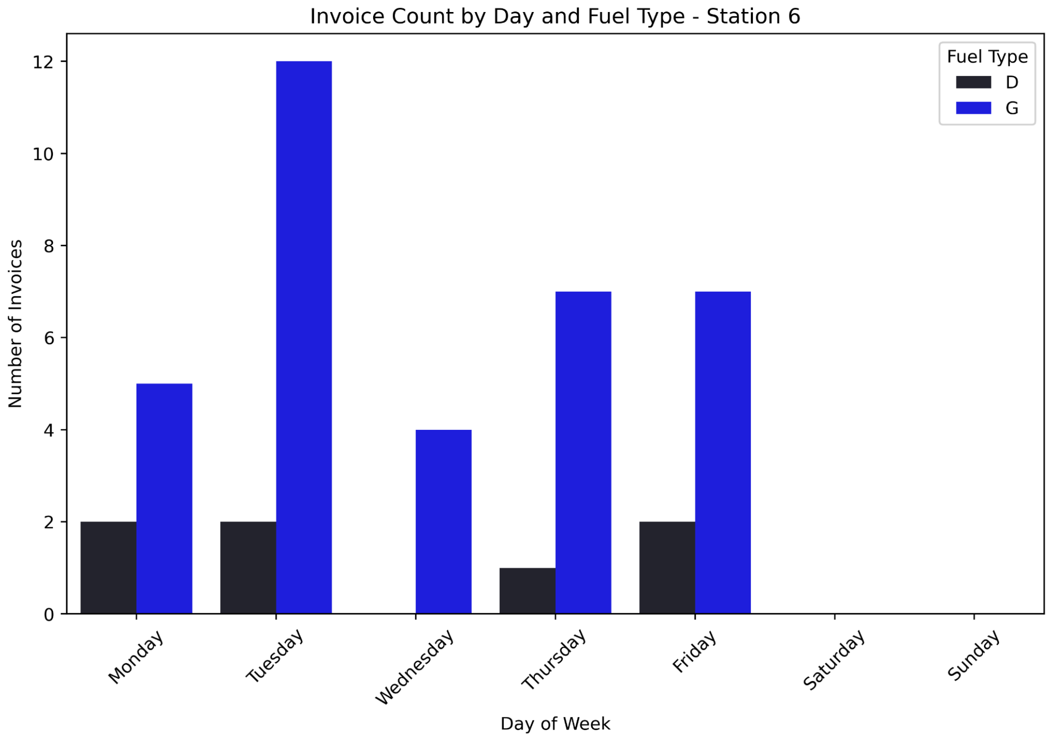
Station 4 demonstrates consistently high invoice counts throughout the week, ranging between 30-45 invoices per day. Surprisingly, the invoice number for both diesel and gasoline volumes are nearly identical. This even distribution reflects regular purchase scheduling.

Station 5 (Mountain View)



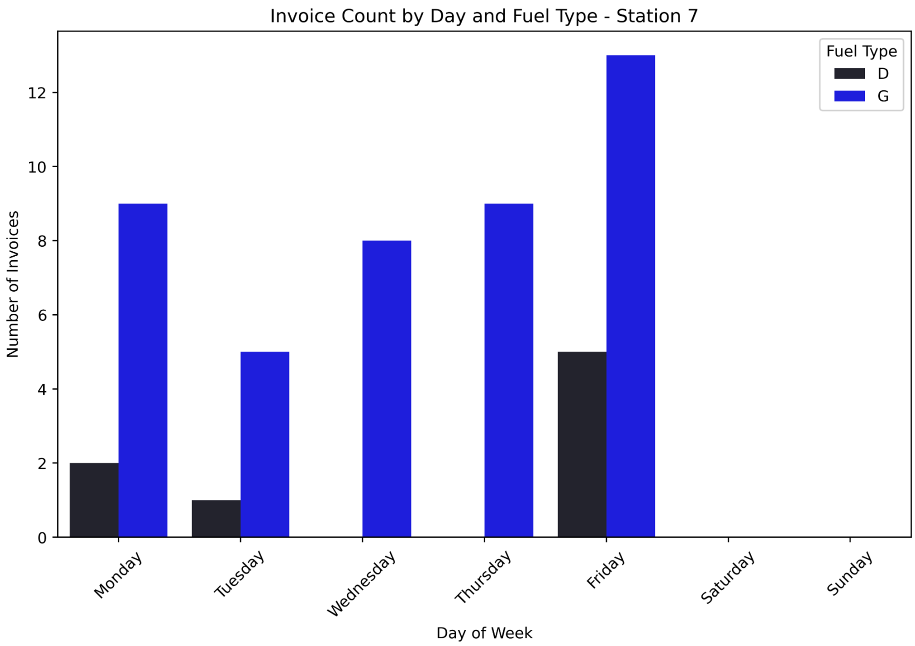
Station 5 records aFriday increase, with both fuel types reaching the peak number of invoices on Friday (~45-50 invoices). Gasoline consistently exceeds diesel in count, suggesting higher transaction frequency for gasoline deliveries.

Station 6 (Oakville)



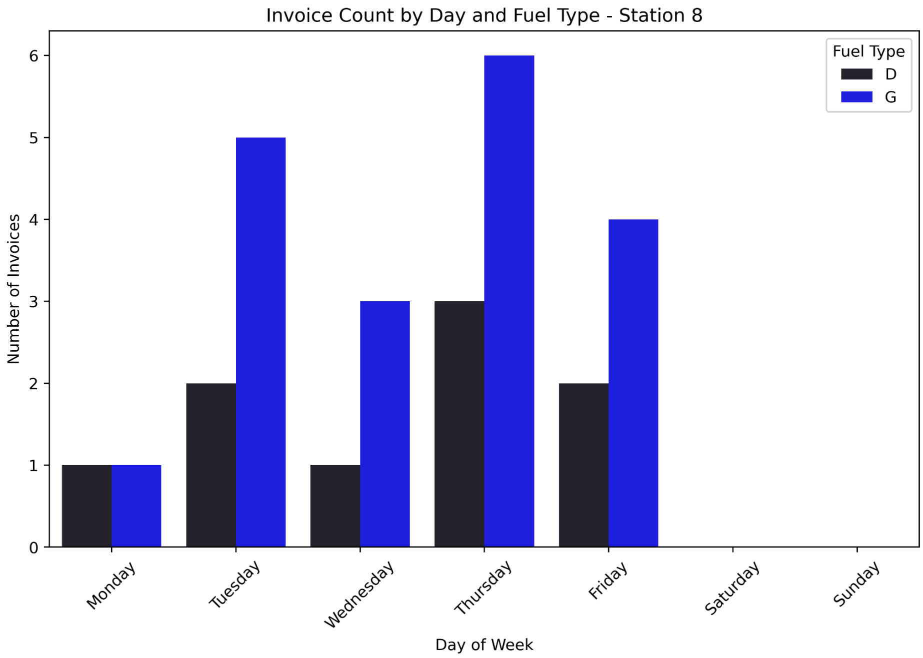
For station 6, we can see there’s no invoice count for diesel. From the previous analysis, we know that station 6 has a more gasoline-dominated pattern. The chart also confirms this point, we can see there's a great difference between the invoice number of those two types of fuel. The invoice count of gasoline peaks on Tuesday (12 invoices), while diesel transactions remain minimal.

Station 7 (Circle)



Similar to station 6, the invoice for station 7 shows strongly gasoline dominated. The pattern for diesel is irregular, and shows orders sporadically, mainly on Monday and Friday. Invoice activity is modest but consistent for gasoline, averaging 8-13 invoices on weekdays, and peaking on Friday. Diesel orders appear sporadically, mainly on Monday and Friday.

Station 8 (Chappel)



For station 8, the invoice numbers for both diesel and fuel are very low, fewer than 6 invoices per day. Gasoline accounts for most invoices, with peaks on Tuesday. Diesel activity is limited to small early-week purchases. On Monday, the invoice numbers for both types are identical.

3.7 Evaluating Gas Station Fuel Inventory Management and Optimization

To conduct a precise and accurate analysis on tank level and station level, the fuel inventory management is a critical and inevitable component of operational efficiency and financial control in the industry. Managing how much fuel is stored at each gas station directly impacts three key aspects of business performance. First, it’s the capital utilization. Excess inventory ties up working capital and increases holding costs. Second, it’s the operational risk. Insufficient inventory levels increase the risk of stockouts, disrupting customer service, which in turn damages the profit. Last, it’s the supply chain efficiency. Optimized deliveries improve resource use and reduce logistics costs.

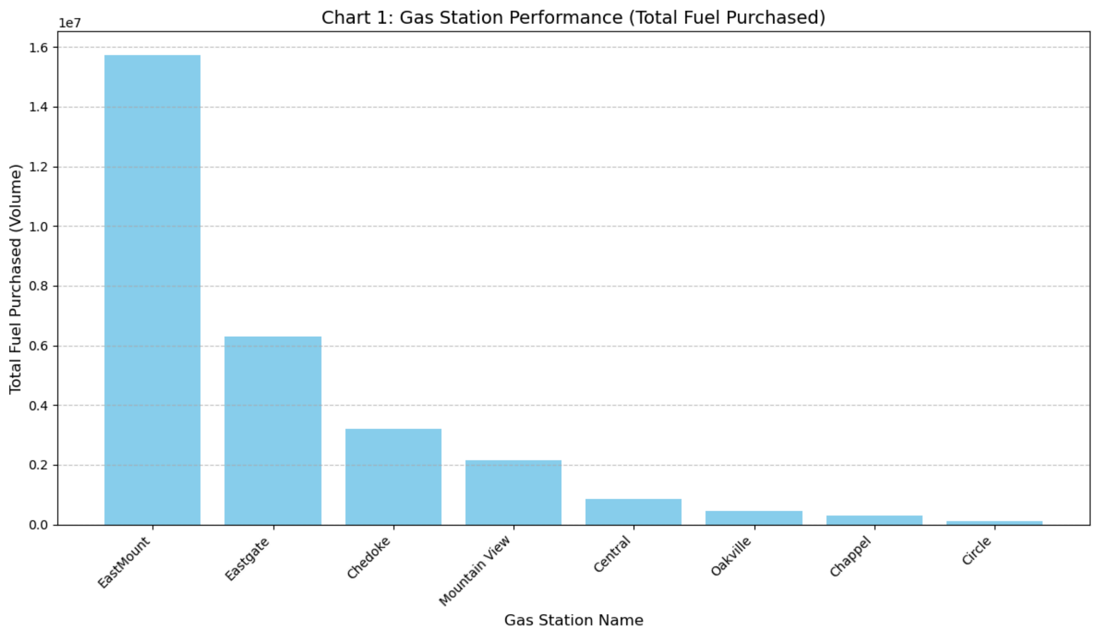
In this part, we will focus on evaluating how effectively a network of gas stations manages its fuel inventory and ordering strategies. Using detailed datasets of tank capacities, fuel level readings, and invoice histories, we will analyze both efficiency, which means how well stations control inventory levels and risk which means how often stations operate near critical low levels. The analysis also visualizes station-level performance, identifies which stations perform well or poorly, and provides strategic recommendations for improvement.

3.7.1 Gas Station Performance Ranking Based on Total Fuel Purchased

In this part, we evaluate the performance on the station level to see the performance using the total fuel purchased by the station, which can assess overall station performance and ranking over the entire dataset period.

In this part, we merged the dataset invoices_df with the 'gas_station_location' and the 'gas_station_name' columns in the locations_df. Using the merged dataset, we grouped the station name with .groupby(), then sum total fuel purchased. In order to make the code more readable, we visualized the data using barplot.

After the analysis, we found that in this metric calculation, the station with the best performance is EastMount with the total fuel purchased of 15,738,669. The station with least satisfying performance is the station Circle with the total fuel purchased of 104,645.

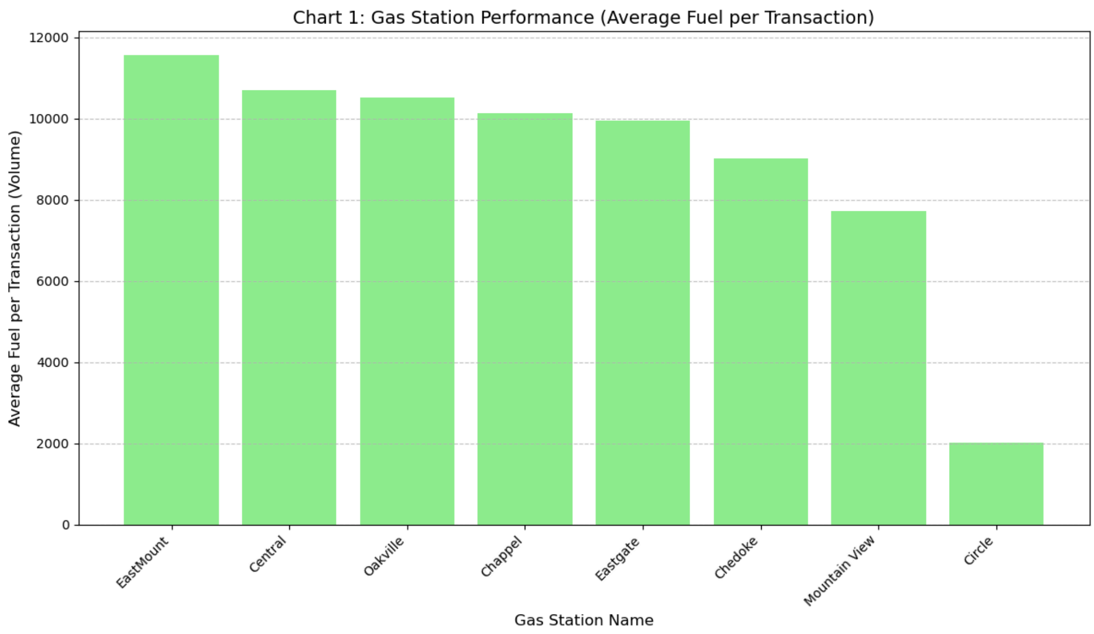


3.7.2 Gas Station Performance Ranking Based on Average Fuel per Transaction

In this part, we used the metric of average fuel per transaction to show the efficiency of the station, and also be the supplement for the total fuel purchased analysis that can help identify the stations that need improvement.

Using the previously merged dataset, we grouped the station name with .groupby(), then calculated the mean of total fuel purchased. In order to make the code more readable, we also visualized the data using barplot.

After the analysis, we found similar results as the analysis using total fuel purchased. The station with the best performance is EastMount with the average fuel per transaction of 11,564.05. The station with least satisfying performance is the station Circle with the average fuel per transaction of 2,012.41. However, unlike the total fuel purchased results, there is not much difference between each station.



For those stations with high efficiency like Eastgate, EastMount, the analysis suggests that these stations have high average transaction volume, suggesting they attract fleet customers or high-volume purchasers. Therefore, their improvement should focus on maintaining strong customer relationships, offering volume discounts, or providing fast-lane service to retain the current transaction size.

For those stations with low efficiency stations like Circle, Chappel, Central, the results show that they have a low average transaction volume. This may indicate they primarily serve small private vehicles or customers making small, intermittent refills. Therefore, their strategic suggestions should focus on implementing measures to incentivize customers to purchase more fuel at once, such as: loyalty point rewards, minimum purchase discounts, or bundle offers with car washes/convenience store items to increase the total customer spend.

3.8 Use More Complicated Metrics to Evaluate the Performance of the Station

In this part, we enter the core analytical part to evaluate how effectively and safely each gas station manages its fuel inventory. To achieve this, two primary performance metrics were defined:

Metric 1: Turnover Rate (%)

Turnover Rate (%) = Total Fuel Purchased / Total Tank Capacity

This metric can help measure the efficiency and capital utilization of each station.

Metric 2: Low Safety Stock Frequency (%)

Low Safety Stock Frequency (%) = the Number of Stations Perform under the

Low_Safety_Stock_Threshold / the Total Number of Station

The metric measures operational risk and stability.

With those two metrics, we will be able to have a more accurate view about the performance of each station. Additionally, we visualized both metrics to compare stations and interpret trends.

3.8.1 Using turnover rates to see which station would benefit most from increased tank capacities

This section aims to evaluate the capacity utilization efficiency of each gas station by analyzing the relationship between fuel sales (demand) and tank capacity (supply constraint).

A higher turnover rate indicates that a station's fuel storage capacity is being used more intensively relative to its sales volume — meaning it operates under greater capacity strain and may benefit most from additional tank capacity or more frequent refueling cycles.

Metric 1: Turnover Rate (%)

The turnover rate was calculated for each gas station as follows:

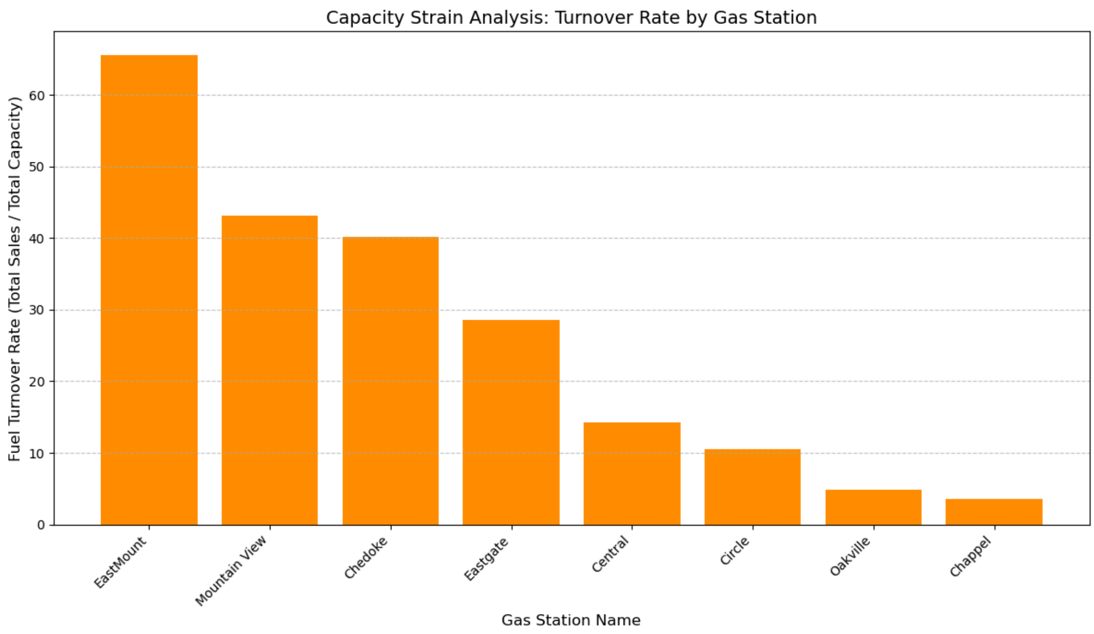
$$\text{Turnover Rate}_i = \frac{\text{Total Fuel Purchased}_i}{\text{Total Tank Capacity}_i}$$

From the results of the code, we can see that EastMount station has the highest turnover rate and benefits the most. The turnover rate directly measures how efficiently each station uses its available capacity:

A high turnover rate indicates the station sells fuel quickly relative to its capacity – a sign of high demand or limited storage. A low turnover rate suggests under-utilized capacity or weaker demand. The analysis also ranked stations by turnover rate to identify those that would benefit most from capacity expansion.

Visualization of the turnover rate

In the visualization part, we used a bar chart comparing turnover rates across all gas stations, where taller bars indicate greater utilization pressure shown in the figure, highlighting capacity strain differences among stations. This bar chart clearly illustrates which stations are operating near or beyond optimal capacity levels and which have spare capacity available.



Interpretation and Insights

From the ranking and visualization, we come up with the following interpretation. The turnover rate stands for the relationship between fuel level and total tank capacity. The definition suggests that stations at the top of the ranking exhibit higher turnover rates and tend to high customer demand and frequent refueling cycles. The potential operational risks are about supply interruptions. To mitigate the potential operational risks, these stations can increase tank capacity or optimize delivery frequency. As for the stations with lower turnover rates, the stations might be over-capitalized in storage infrastructure relative to their actual sales volume. The potential policies for those types of stations with lower turnover rate could potentially reduce tank size or adjust logistics to optimize utilization.

```
--- Capacity Strain Analysis: Which Fuel Stations Would Benefit Most? ---
Benefit is determined by the Turnover Rate (Total Sales / Total Capacity).
Stations with the HIGHEST Turnover Rate are under the MOST strain and benefit most.
--------------------------------------------------------------------------
gas_station_name total_fuel_purchased total_tank_capacity turnover_rate
       EastMount           15,738,669             240,000         65.58
   Mountain View            2,158,447              50,000         43.17
         Chedoke            3,211,086              80,000         40.14
        Eastgate            6,283,414             220,000         28.56
         Central              855,334              60,000         14.26
          Circle              104,645              10,000         10.46
        Oakville              441,310              90,000          4.90
         Chappel              283,793              80,000          3.55
--------------------------------------------------------------------------
```

Conclusion

This analysis quantifies how effectively each gas station utilizes its fuel storage capacity. Stations with high turnover rates are the most efficient but also face greater operational strain and risk of stockouts. By identifying these high-strain locations, management can make data-driven decisions to expand capacity, optimize delivery schedules, and enhance supply chain resilience.

3.8.2 Using low safety stick frequency to see which station would have higher inventory risk

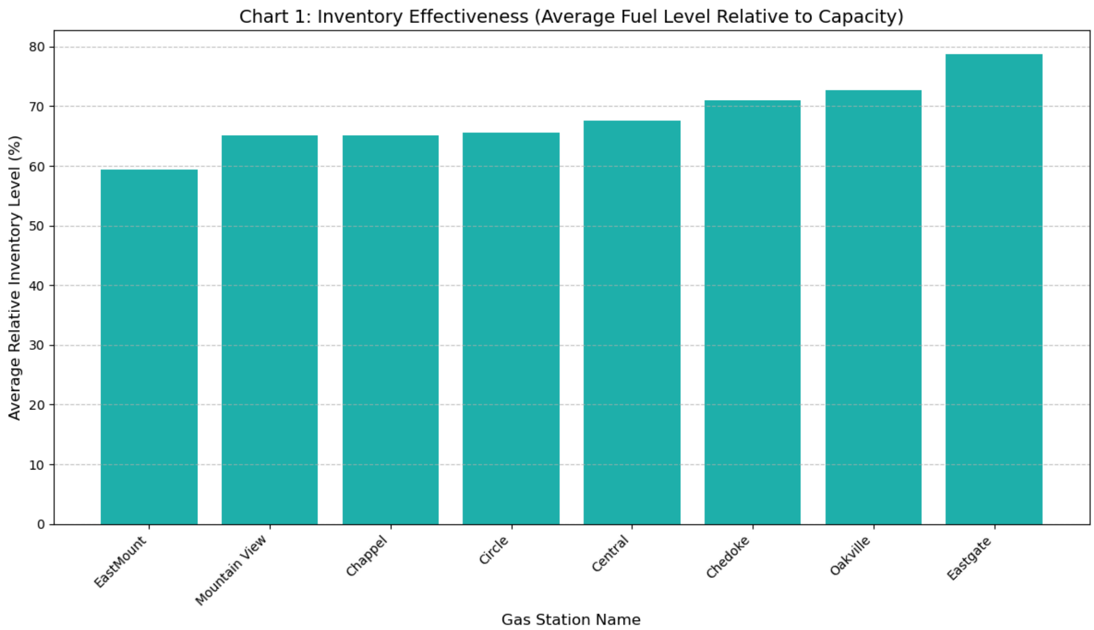
In this part, we will use two metrics to define which station would have higher inventory risk.

Metric 2 – Relative Fuel Level (%)

$$\text{Relative Fuel Level } (\%) = \frac{\text{Fuel Level}}{\text{Tank Capacity}} \times 100$$

Visualization

A bar chart showing the relative fuel level on gas station level to represent the relative safety level of each tank.



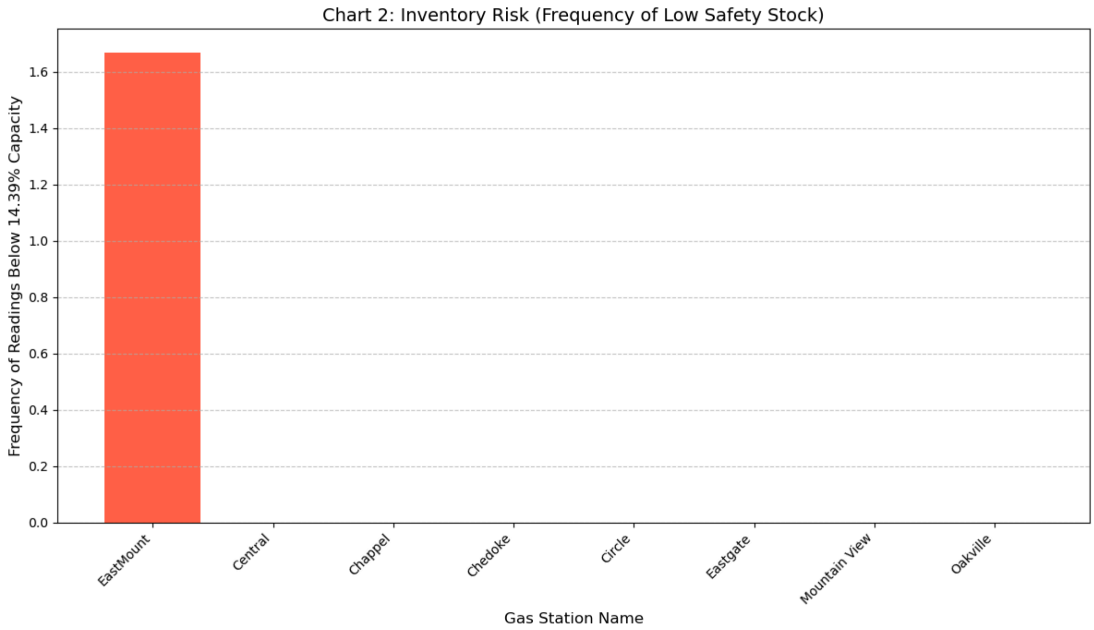
From the chart, we can see the bars with shorter bars are capital-efficient.

Metric 3 – Low Safety Stock Frequency (%)

$$\text{Low Safety Stock Frequency } (\%) = \frac{\text{Count of Readings Below Threshold}}{\text{Total Readings}} \times 100$$

Visualization

The second bar chart plots the frequency of low-stock events per station. In this chart, stations are sorted from highest risk to lowest. The chart has some blanks on several stations, but it is normal since from the statistics of these stations the frequency of low safety stick frequency is 0, which means those stations have a more moderate inventory policy.



Low Safety Stock Level Definition

In order to evaluate operational risk, we introduced the concept of low safety stock threshold, set to 14.39% of tank capacity. From the tanks_df, we can see that the tank capacity has great differences. For T29, T30, they have relatively low tank capacity of 5000. For T16, T19, they have relatively high tank capacity of 70000. In order to have a more accurate amount of LOW_SAFETY_STOCK_THRESHOLD, we dropped those tank capacities to calculate the average of the ratio of previously defined LOW_SAFETY_STOCK of 5000. This threshold was determined from historical consumption data — approximately the average level at which emergency deliveries or low-level alarms occur. For the fuel level below this threshold represents a risk event, such as potential stockout exposure.

Interpretation and Insights

Higher frequency means higher risk, which means the station frequently runs close to depletion, risking stockouts and lost sales, therefore has negative influence for the profit of the station. On the contrary, for the stations with low frequency, the risk tends to be lower, which means the station maintains a comfortable buffer to absorb demand variability. Also the stations with lower frequency show more effective management which balance the capital efficiency and risk control.

Conclusion

This analysis provides quantitative proof of the possible influence of the possible strategies change, like increasing the tank capacity. Those actionable insights can help gas stations manage their fuel inventories to the fullest. By comparing efficiency (average stock levels) and risk (low-stock frequency), management can make informed decisions to optimize fuel logistics, and maintain more reliable operations.

The findings emphasize the importance of striking a balance between saving money and ensuring reliability. From the analysis, we know cost-saving focus stations like EastMount can face the problem of stockouts. However, without careful oversight, some stations could undermine operational stability.

By applying these insights, we can provide more practical strategies for different types of stations which in turn strengthen their fuel distribution system and sustain high service levels.

4. Tank Analysis

4.1 Assumptions for Safety Line

For this part, we will discuss the assumption for the safety line on each tank. Mentioned in Part 3.8.2, we assume the ideal safety line quantity for a standard-sized capacity tank is 5000. However, from the tank capacity result, we found that there are several tanks having larger capacity, while some tanks have smaller tank capacity. Therefore, using 5000 to be the exact number of the safety line is impractical. Therefore, we decided to use ratio to be the safety line. In this way can we standardize safety stock thresholds across fuel tanks with varying storage capacities, and establish a fair and scalable rule for safety stock, rather than using a single arbitrary number for all tanks, eventually we can ensure that smaller tanks maintain proportionally higher safety reserves (to buffer limited storage) and larger tanks hold appropriate safety levels aligned with their size. The following methodology was implemented to calculate, normalize, and apply customized safety levels across different tank sizes using real operational data.

Methodology

To start with, we created a temporary DataFrame tank_df_temp to filter tanks within a defined operational range, which are tanks with capacities greater than 5,000 and less than 70,000 were considered “standard-sized” tanks. The capacity of these tanks represent the majority of operational assets, forming the basis for calculating a standardized safety rate.

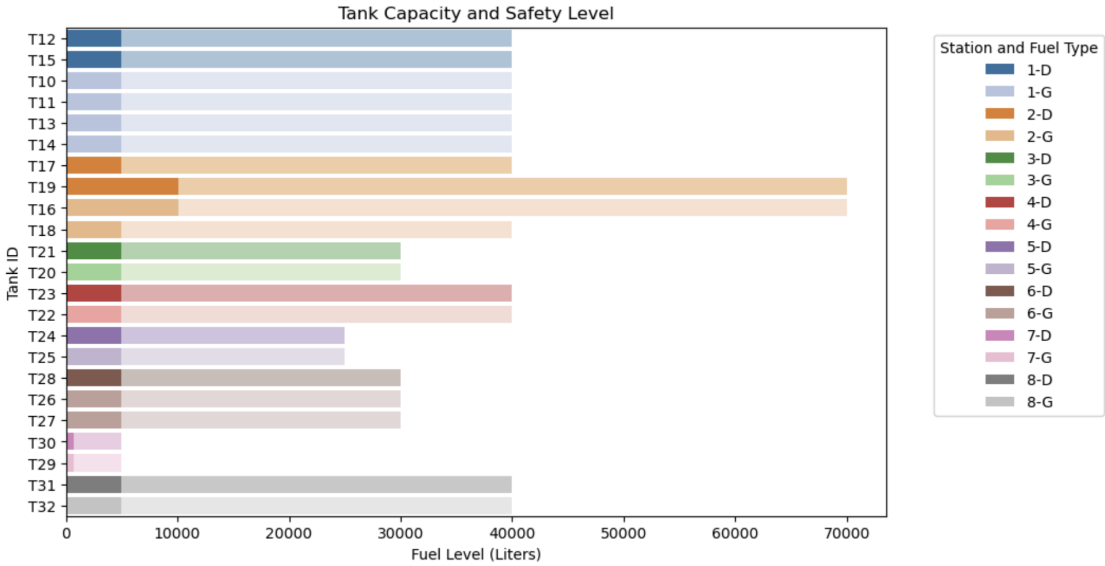
Under the dataframe of tank_df_temp, we can further calculate the low safety stock rate, under the methodology below:

$$\text{Low Safety Stock Rate} = \frac{5000}{\text{Tank Capacity}}$$

This ratio quantifies the proportion of total capacity considered “safe” for each tank within the target range. Based on the low safety stock rate, we can compute the average safety level ratio as the average safety level ratio, under the methodology of average safety level rate equals to the mean of all low safety stock rates. This average ratio represents the typical safety margin used across the system and serves as a scaling factor for non-standard tanks (those smaller than 5,000 or larger than 70,000). After calculation, we used the function of safety_level(row) to create a new column that shows the safety level and safety level rate of different tanks. After applying the function, the data shows the tanks between 5,000 and 70,000 capacity were assigned a flat safety level of 5,000 units. For the tanks outside this range were assigned a scaled safety level equivalent to avg_safety_level * tank_capacity. Therefore, we can make sure that standard-sized tanks retain a fixed safety stock of 5,000 units, and as for the non-standard tanks (very small or very large), those tanks receive a scaled safety level proportional to their capacity, maintaining consistency

	tank_id	tank_location	tank_number	tank_type	tank_capacity	safety_level	safey_level_rate	station_fuel_type
0	T12	1	3	D	40000	5000.0	0.12	1-D
1	T15	1	6	D	40000	5000.0	0.12	1-D
2	T10	1	1	G	40000	5000.0	0.12	1-G
3	T11	1	2	G	40000	5000.0	0.12	1-G
4	T13	1	4	G	40000	5000.0	0.12	1-G
5	T14	1	5	G	40000	5000.0	0.12	1-G
6	T17	2	2	D	40000	5000.0	0.12	2-D
7	T19	2	4	D	70000	10073.0	0.14	2-D
8	T16	2	1	G	70000	10073.0	0.14	2-G
9	T18	2	3	G	40000	5000.0	0.12	2-G
10	T21	3	2	D	30000	5000.0	0.17	3-D
11	T20	3	1	G	30000	5000.0	0.17	3-G
12	T23	4	2	D	40000	5000.0	0.12	4-D
13	T22	4	1	G	40000	5000.0	0.12	4-G
14	T24	5	1	D	25000	5000.0	0.20	5-D
15	T25	5	2	G	25000	5000.0	0.20	5-G
16	T28	6	3	D	30000	5000.0	0.17	6-D
17	T26	6	1	G	30000	5000.0	0.17	6-G
18	T27	6	2	G	30000	5000.0	0.17	6-G
19	T30	7	2	D	5000	719.5	0.14	7-D

with the average safety behavior of the fleet.

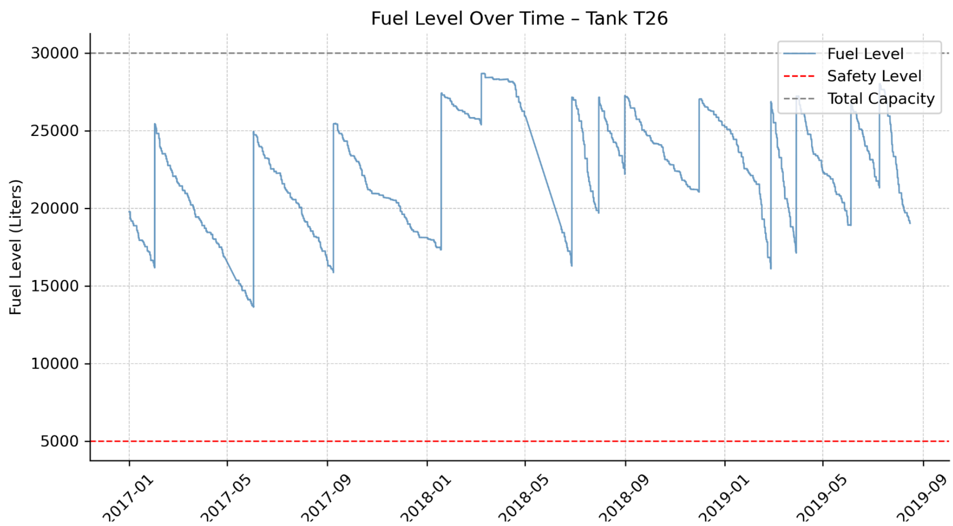


4.2 Fuel Level Over Time for Each Tank

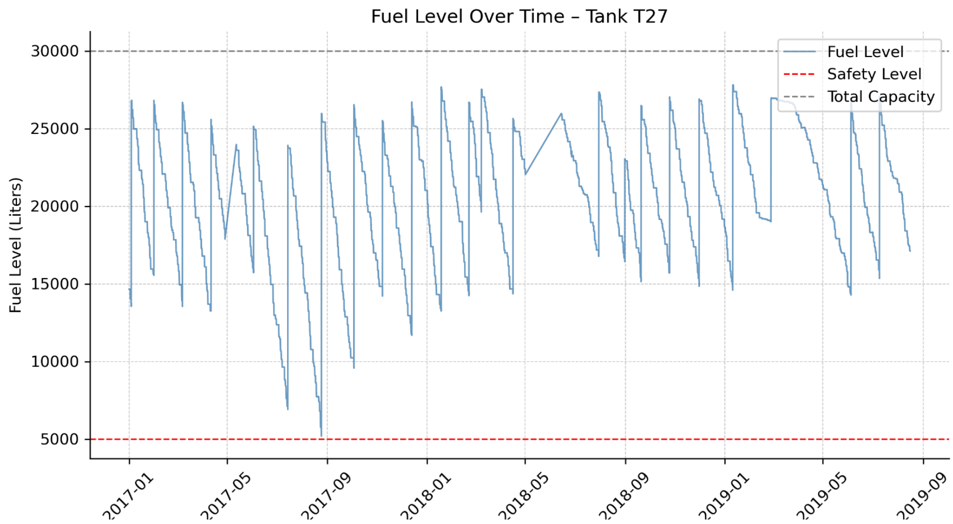
In this part, we looked into the fuel level over time on tank level, and visualized fuel level dynamics over time for each storage tank across the fuel station network.
By monitoring how fuel levels fluctuate relative to total tank capacity and defined safety thresholds, we can identify tanks operating under risky or inefficient inventory conditions. The code creates individual time-series plots for each tank, enabling detailed observation of fuel utilization patterns, refill cycles, and potential operational issues such as excessive depletion or irregular refueling frequency. From this part, we can track the fuel level trends over time for every tank, and visually compare the actual fuel levels with the tank's total capacity and safety level, and identify potential stockout risks or overfilled conditions.

Class 1 - Stable and Predictable Tanks

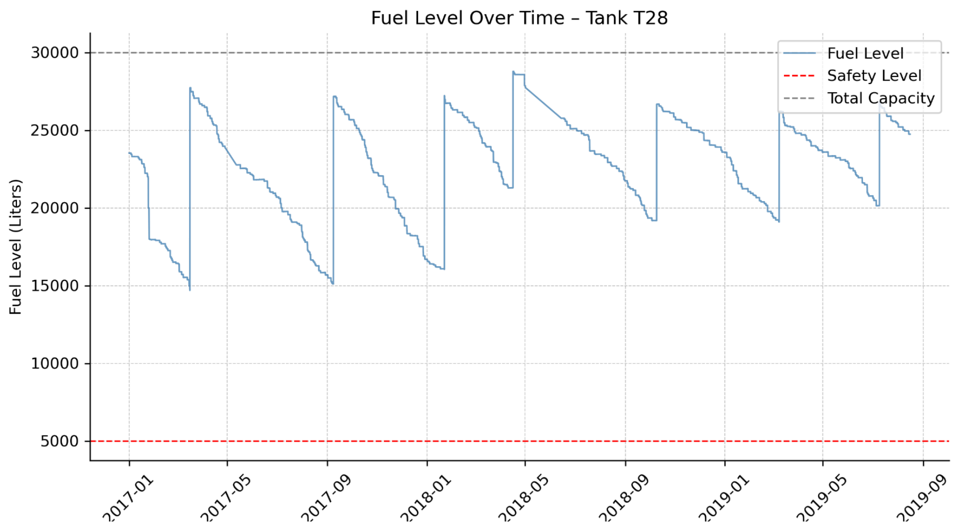
Examples: T26, T27, T28, T29, T30, T31, T32



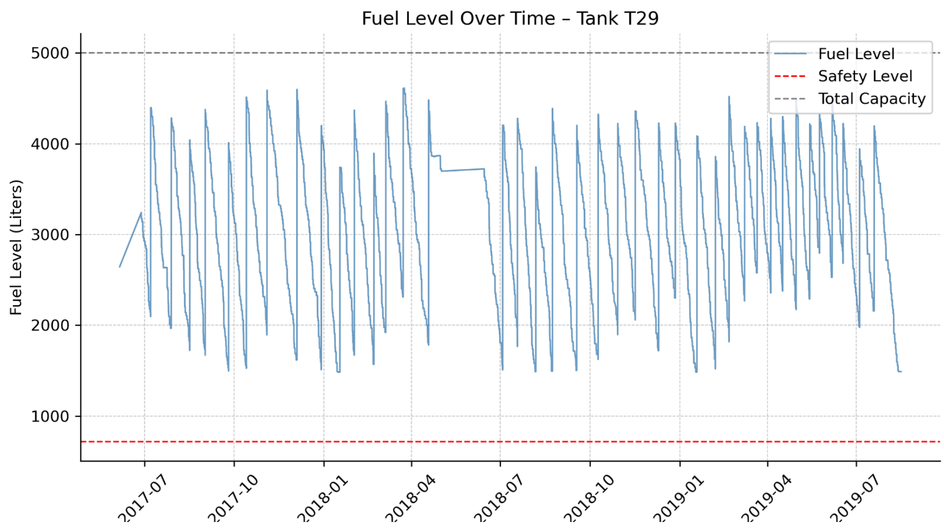
Fuel Level Over Time – Tank T26
Fuel Level (Liters)
Fuel Level
Safety Level
Total Capacity
30000
25000
20000
15000
10000
5000
2017-01
2017-05
2017-09
2018-01
2018-05
2018-09
2019-01
2019-05
2019-09

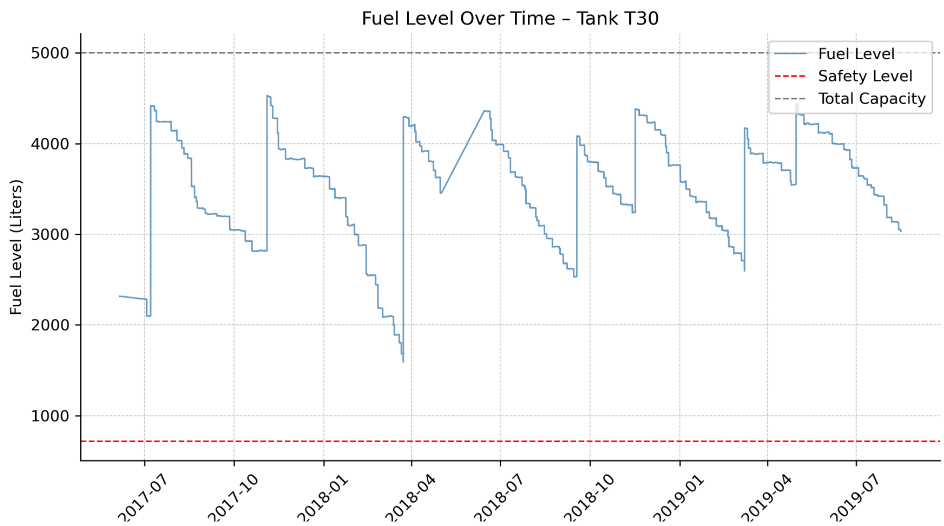


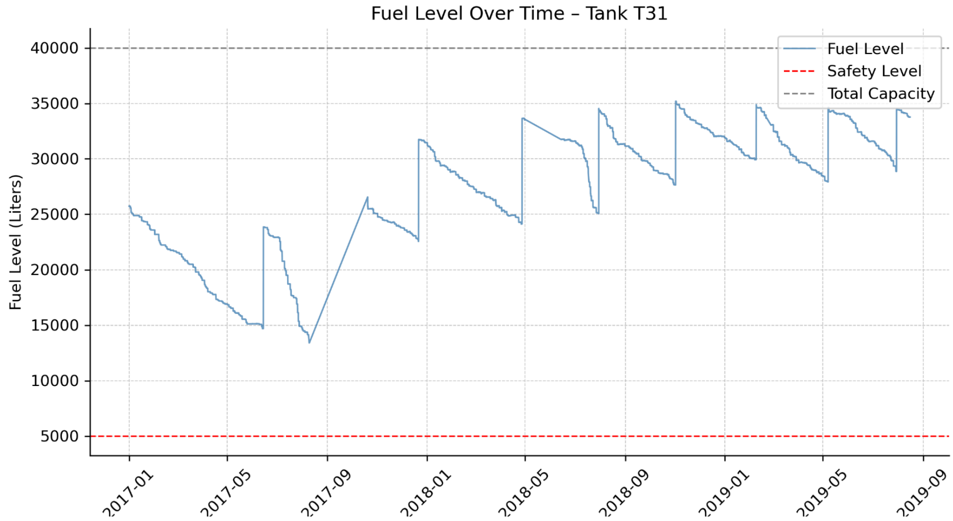
Fuel Level Over Time – Tank T27
Fuel Level (Liters)
Fuel Level
Safety Level
Total Capacity
30000
25000
20000
15000
10000
5000
2017-01
2017-05
2017-09
2018-01
2018-05
2018-09
2019-01
2019-05
2019-09

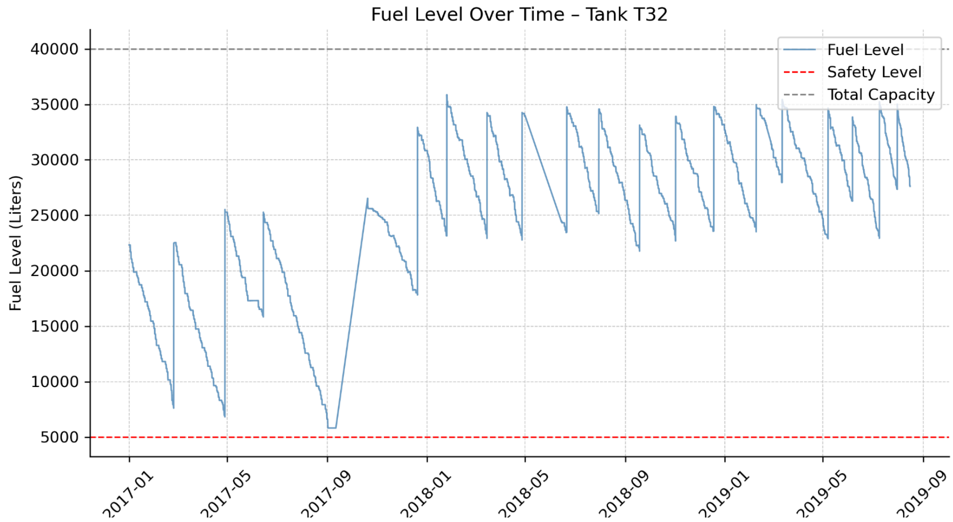


Fuel Level Over Time – Tank T28
Fuel Level (Liters)
Fuel Level
Safety Level
Total Capacity
30000
25000
20000
15000
10000
5000
2017-01
2017-05
2017-09
2018-01
2018-05
2018-09
2019-01
2019-05
2019-09


Fuel Level Over Time – Tank T29
Fuel Level (Liters)
5000
4000
3000
2000
1000
2017-07
2017-10
2018-01
2018-04
2018-07
2018-10
2019-01
2019-04
2019-07
Fuel Level
Safety Level
Total Capacity


Fuel Level Over Time – Tank T30
Fuel Level (Liters)
5000
4000
3000
2000
1000
2017-07
2017-10
2018-01
2018-04
2018-07
2018-10
2019-01
2019-04
2019-07
Fuel Level
Safety Level
Total Capacity





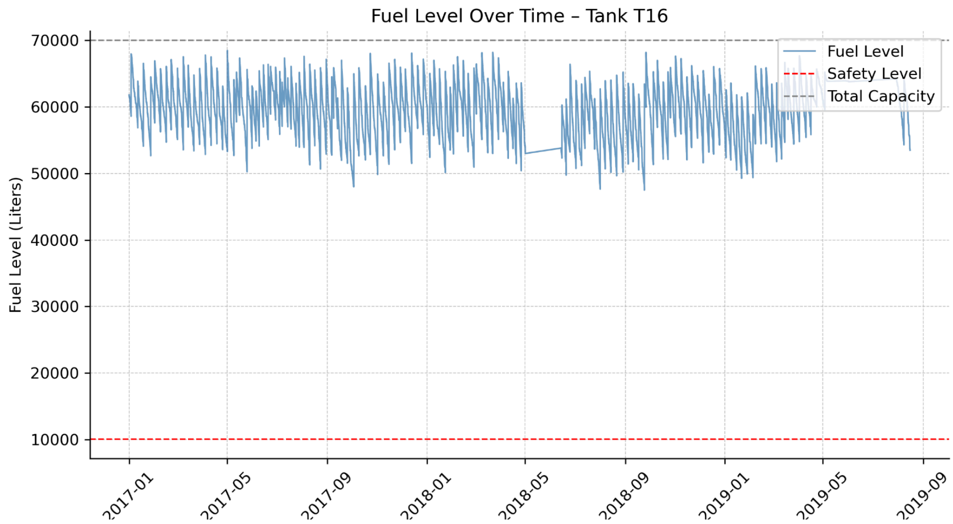
These tanks demonstrate the most ideal inventory behavior. Their fuel levels follow a consistent, repeating saw-tooth pattern, reflecting steady consumption followed by well-timed replenishment. Refueling events occur at regular intervals, typically every three to five weeks, and the replenishment volume is sufficient to restore the tank close to its upper capacity without significant overfilling. Such curves indicate that both demand forecasting and logistics scheduling are functioning efficiently, ensuring a balance between fuel availability and cost control.

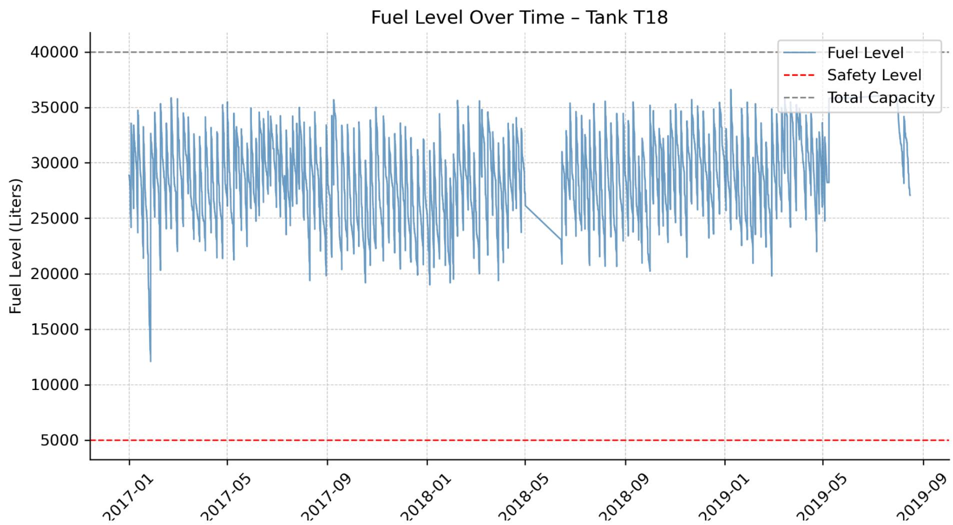
The main advantage of this group is operational predictability: these tanks seldom fall below the low-stock threshold, minimizing the risk of service interruptions. Their replenishment rhythm also suggests mature coordination between local stations and central supply operations.

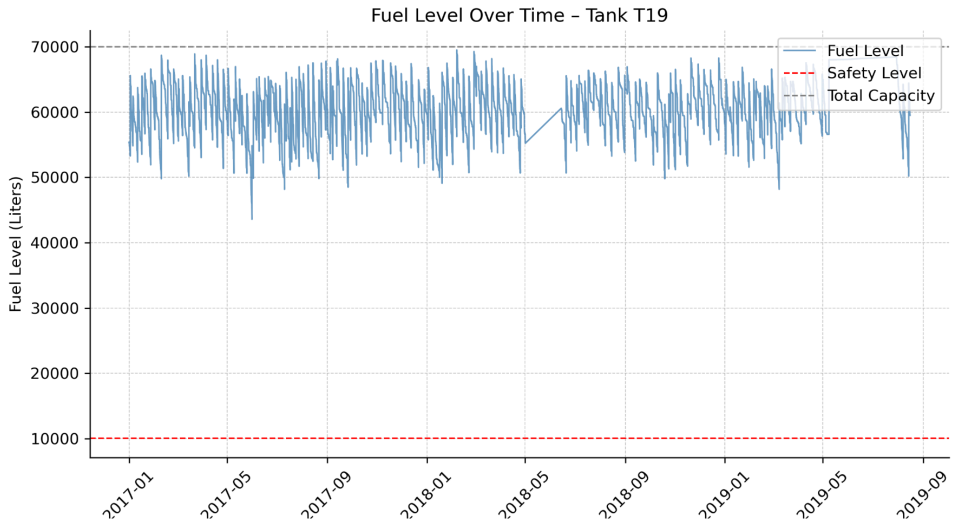
However, Tank 28 and Tank 31’s fuel are replenished as soon as the tank reaches roughly 50% of its total capacity. This reflects disciplined operational planning - refueling occurs at fixed intervals before any stockout risk arises. This strategy, while stable, sacrifices economic efficiency. By initiating refills when half the tank remains, these stations lose the opportunity to consolidate orders or capture bulk-purchase discounts.

Class 2 - Moderately Stable but Conservative Tanks

Examples: T16, T18, T19





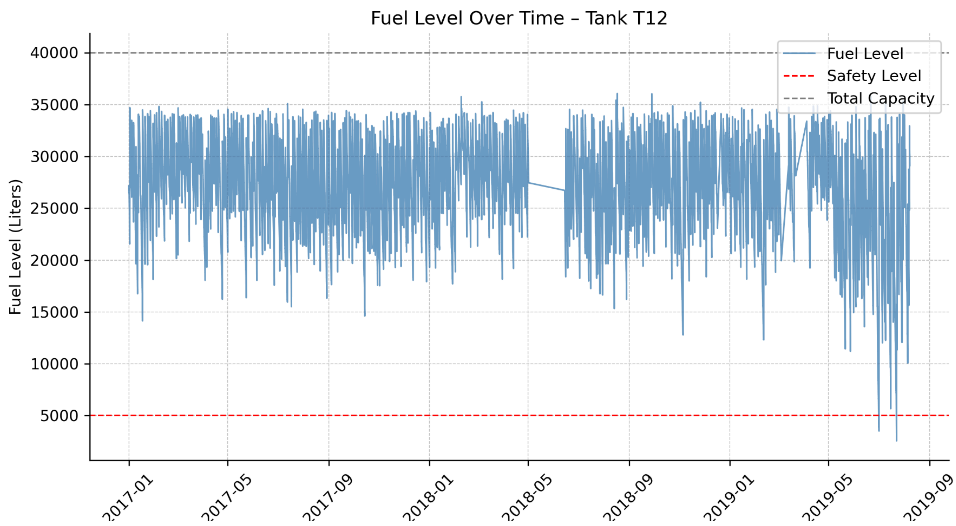


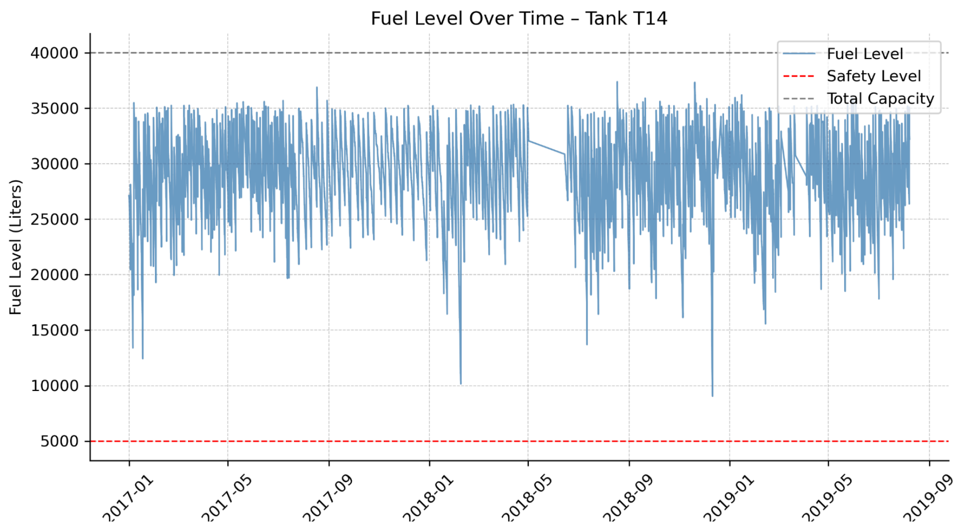
This group of tanks maintains generally high inventory levels, with curves fluctuating within a narrow range between 60% and 90% of total capacity. Refueling tends to occur well before the low-stock threshold, and the intervals between deliveries vary moderately. This pattern reflects a risk-averse inventory strategy, where managers prioritize fuel availability over cost efficiency.

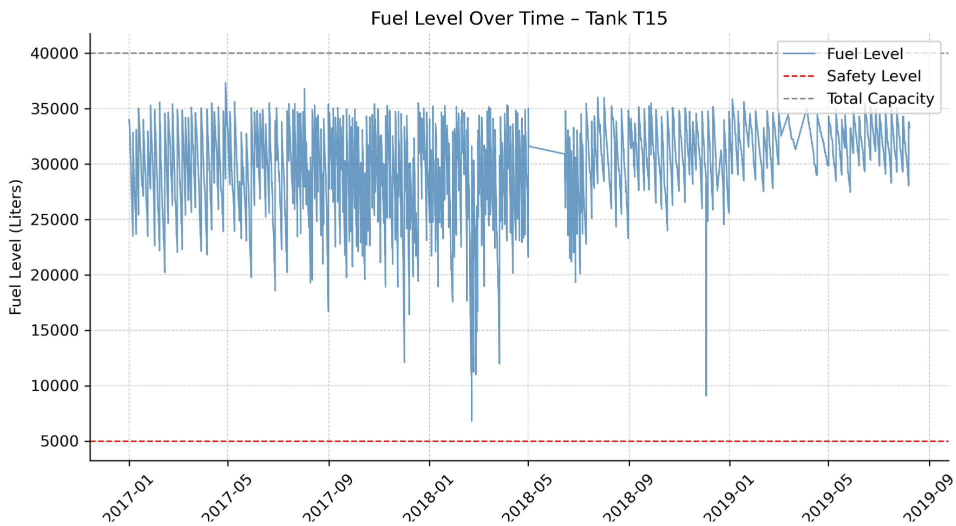
While such conservatism ensures uninterrupted service, it leads to two inefficiencies. First, maintaining high inventory for long periods ties up working capital that could be better utilized elsewhere. Second, high tank levels increase the likelihood of fuel aging and evaporation losses, particularly during warmer months.

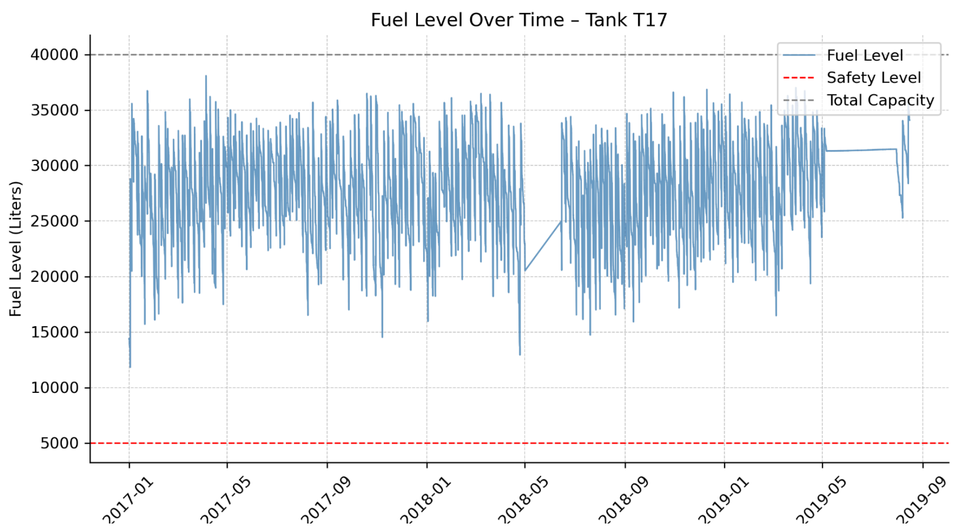
Class 3 - High-Frequency Small-Batch Tanks

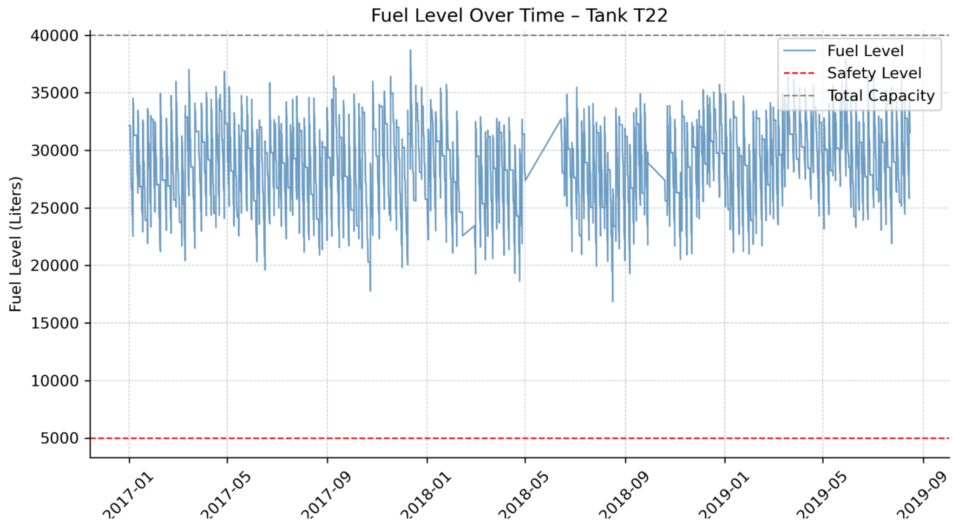
Examples: T12, T14, T15, T17, T22, T23

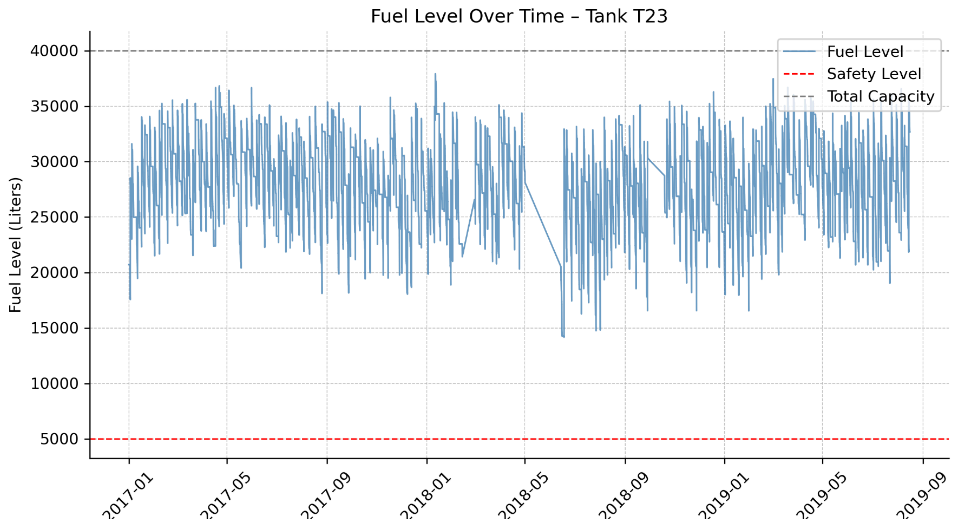



Fuel Level Over Time – Tank T14
Fuel Level
Safety Level
Total Capacity
Fuel Level (Liters)
40000
35000
30000
25000
20000
15000
10000
5000
2017-01
2017-05
2017-09
2018-01
2018-05
2018-09
2019-01
2019-05
2019-09


Fuel Level Over Time – Tank T15
Fuel Level
Safety Level
Total Capacity
Fuel Level (Liters)
40000
35000
30000
25000
20000
15000
10000
5000
2017-01
2017-05
2017-09
2018-01
2018-05
2018-09
2019-01
2019-05
2019-09


Fuel Level Over Time – Tank T17
Fuel Level
Safety Level
Total Capacity
Fuel Level (Liters)
40000
35000
30000
25000
20000
15000
10000
5000
2017-01
2017-05
2017-09
2018-01
2018-05
2018-09
2019-01
2019-05
2019-09



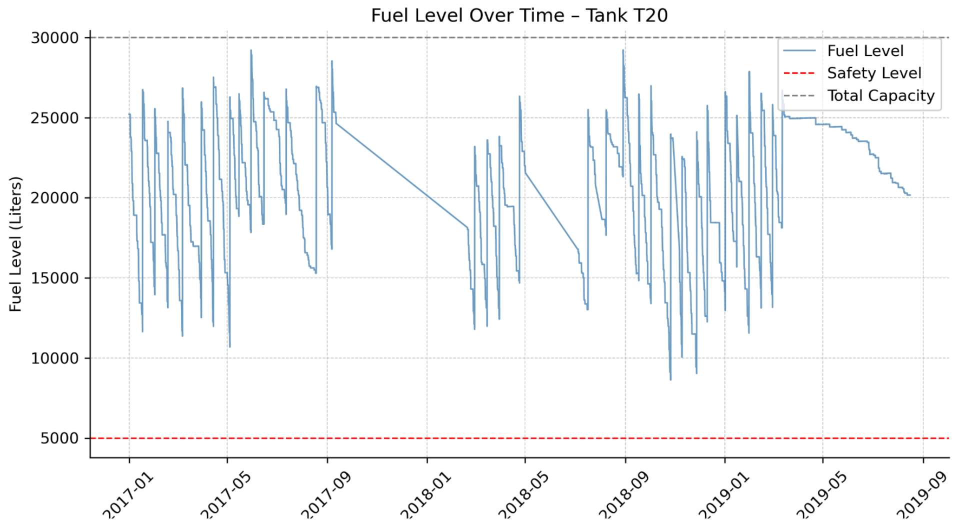


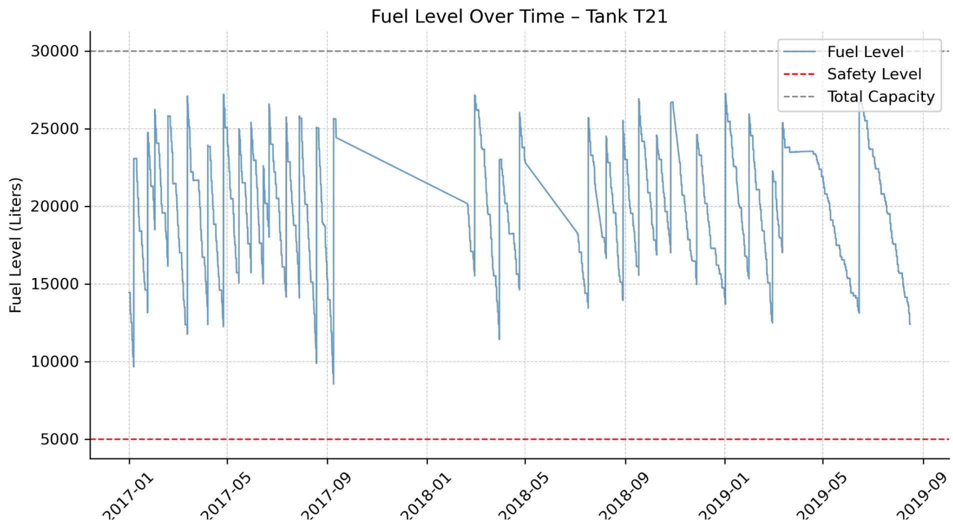
The fuel level trajectories of these tanks fluctuate rapidly and irregularly, indicating multiple small replenishments within short intervals. Each refill increases the level only slightly before another delivery occurs. This pattern implies that these stations are refilling reactively, likely triggered by manual decisions rather than predictive scheduling.

Such over-frequent replenishment creates significant inefficiencies. The logistics cost per liter rises due to repeated truck dispatches with partial loads. Moreover, frequent human intervention increases operational complexity and reduces coordination between supply and station demand. While this approach avoids stockouts, it also prevents the system from capturing bulk purchase discounts or optimizing delivery routes.

Class 4 - Long-Gap Replenishment Tanks

Examples: T20, T21



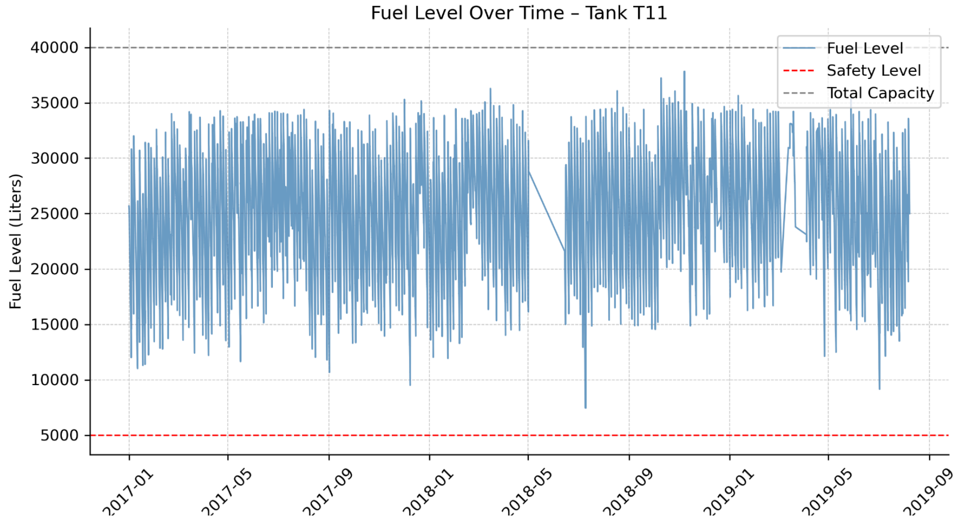


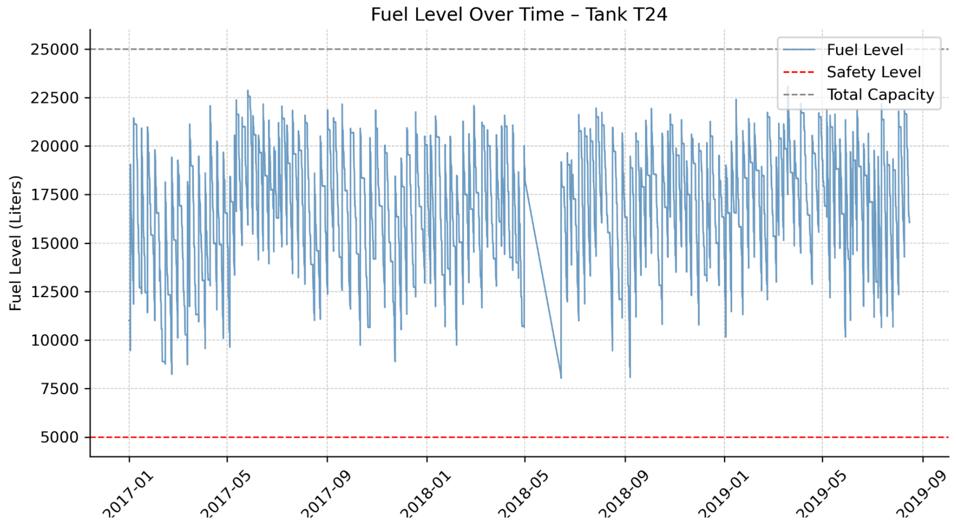
These tanks exhibit long downward slopes with sharp upward jumps, showing infrequent but large replenishments. Refueling events are separated by extended gaps, sometimes lasting months. During these gaps, the inventory often drops near or below the low-stock threshold, exposing the station to high supply risk.

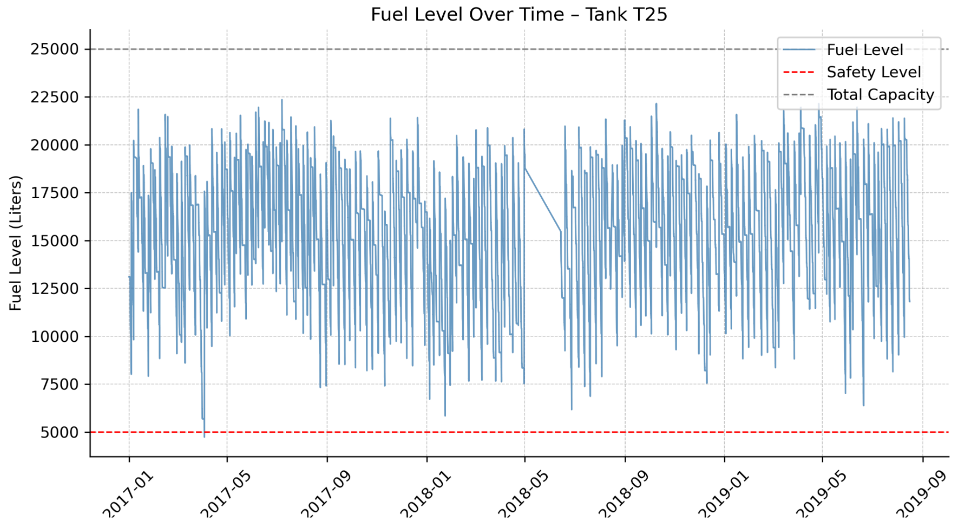
This pattern typically arises from logistical constraints or inconsistent demand. Remote stations or those with lower sales may delay refills until transport is cost-effective. However, this “wait until empty” approach increases the chance of service interruptions and emergency deliveries.

Class 5 - Moderate-Interval, High-Volume Refills

Examples: T11, T24, T25



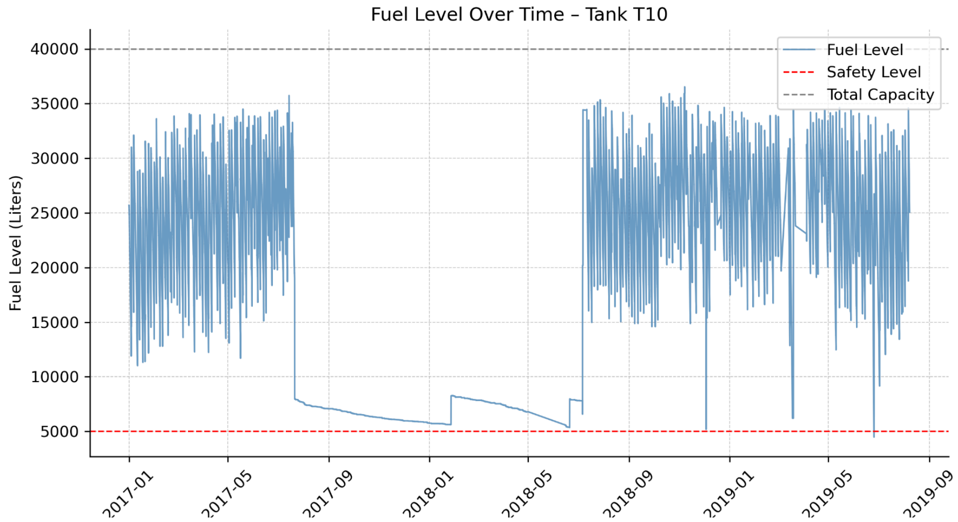


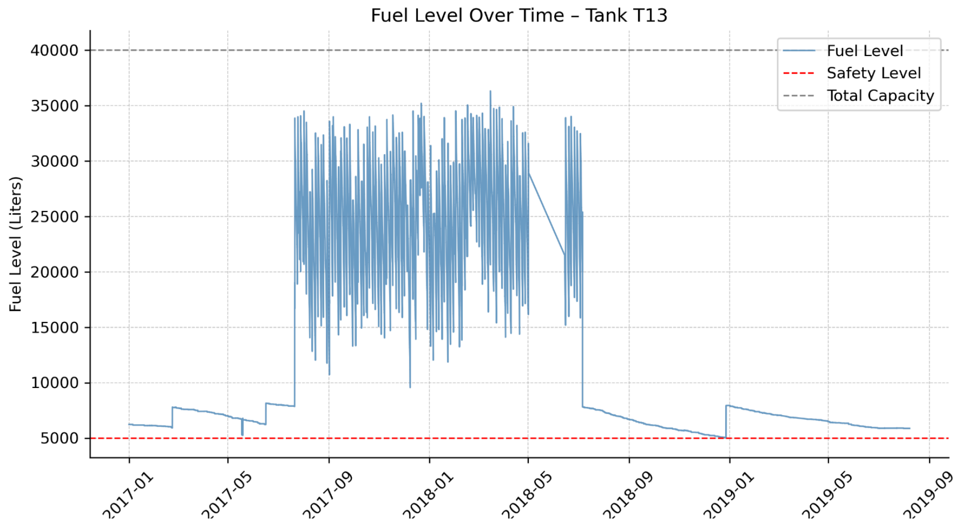


Tanks T11, T24 and T25 display a moderate-interval and high-volume refills pattern that distinguishes them from other long-gap tanks. Their inventory curves show consistent cycles roughly every three to four weeks, with refills triggered when fuel levels drop to around 40-50 percent of capacity. This strategy strikes a balance between cost efficiency and supply safety: the stations avoid frequent deliveries while still preventing stockouts. However, the slightly early refill timing limits capacity utilization and reduces the potential to capture volume-based discounts.

Class 6 - Understocked or Idle Tanks

Examples: T10, T13





These tanks remain near zero inventory for extended periods, with only a few isolated refilling spikes. The pattern implies that these tanks are either inactive, serving as backup capacity, or associated with stations that have temporarily halted operations.

The key issue here is asset underutilization. Maintaining tanks that seldom operate still incurs maintenance, inspection, and safety costs without corresponding revenue.

If these tanks belong to low-volume sites, continuing to allocate logistics resources to them represents inefficiency.

Conclusion

From the chart, we have a better understanding of the monitor consumption cycles showing the fuel level line typically declines steadily over time as customers refuel, then spikes upward upon replenishment showing the delivery cycle frequency. For operational safety, for each graph, if the blue “fuel level” line frequently dips below the red “safety level” line, it indicates risk of stockout and potential service disruption, which means the station needs better inventory strategies for safer stock. On the contrary, if the fuel level consistently remains near the gray “total capacity” line, it suggests excess inventory holding, tying up capital unnecessarily. Besides, we also need to look into irregular behavior. Sudden or erratic spikes in graphs may reveal data recording issues, irregular supply deliveries, or equipment malfunctions.

5 New Strategy

5.1 Suggested Strategies and Fuel Level Over Time After Changing Strategies

Since the invoice data only contains the station number and fuel type, our strategy aggregates the tank information at the station–fuel type level. Specifically, we divided the eight stations into Diesel and Gasoline categories, resulting in sixteen distinct station–fuel-type combinations (e.g., 1-D, 1-G, 2-D, 2-G, …).

This means that for each station, the refill decision is not made at the individual tank level, but rather based on the overall inventory status of each fuel type. When a refill is triggered, all tanks storing the same fuel type at that station are refilled simultaneously. Additionally, the discount applies collectively to all tanks when a station places a combined order for the same fuel type.

We then provided detailed recommendations for each of these sixteen groups.

In this section, we also simulated and visualized the evolution of fuel inventory levels after implementing the optimized refill strategy for each station-fuel type combination. The purpose was to verify whether the proposed strategy can maintain safe fuel levels while reducing refill frequency and total purchasing costs.

The analysis was implemented in Python using a time-series simulation approach.

First, the cleaned fuel-level dataset (tank_fuel_df) was grouped by station-fuel type and timestamp, summing up the total fuel across all tanks belonging to each type.

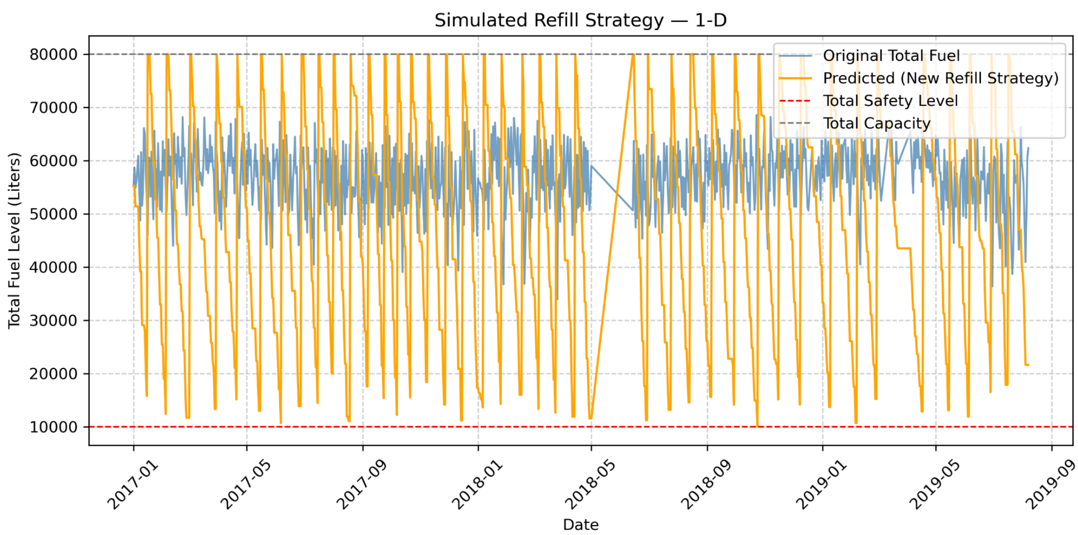
Second, the timestamps were converted into daily observations to remove noise and make the refill pattern clearer.

Third, each station-fuel type, we retrieved its total tank capacity and safety stock level from the summary table (tanks_result_df).

Fourth, the algorithm iterated through each day in chronological order, computing the change in fuel level (diff) between consecutive timestamps. When diff was negative, it indicated consumption; When the simulated fuel level dropped below the safety threshold, the code triggered an automatic refill event, instantly bringing the tank level up to full capacity.

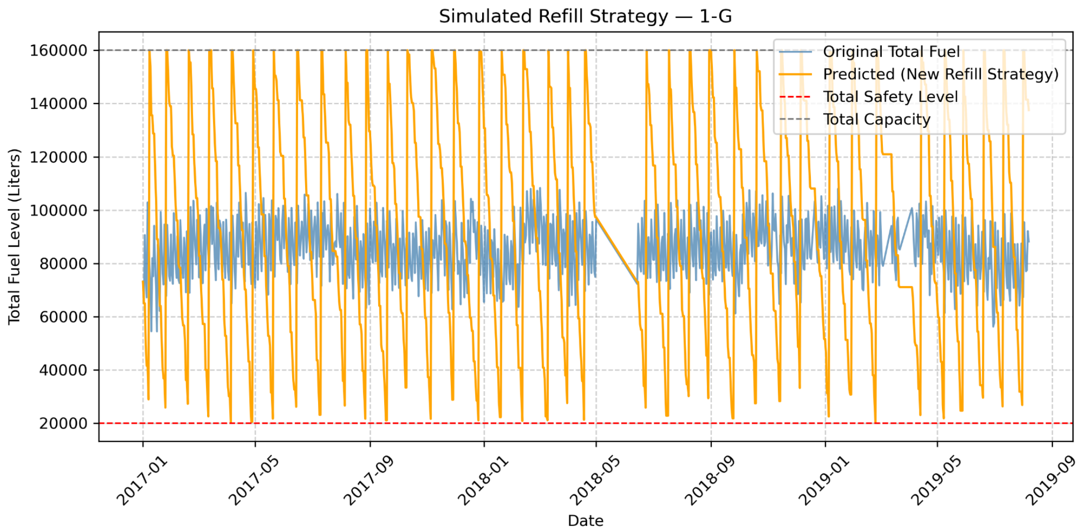
Last, both the original observed fuel level (blue line) and the simulated refill level under the new strategy (golden line) were plotted, along with the safety line (red dashed) and capacity line (gray dashed).

This simulation framework effectively models the operational behavior of each station under the new refill policy. By comparing the trajectories before and after optimization, we can assess whether the revised strategy successfully stabilizes inventory levels, reduces excessive refills, and minimizes low-stock occurrences - thereby confirming the strategy's practical viability.



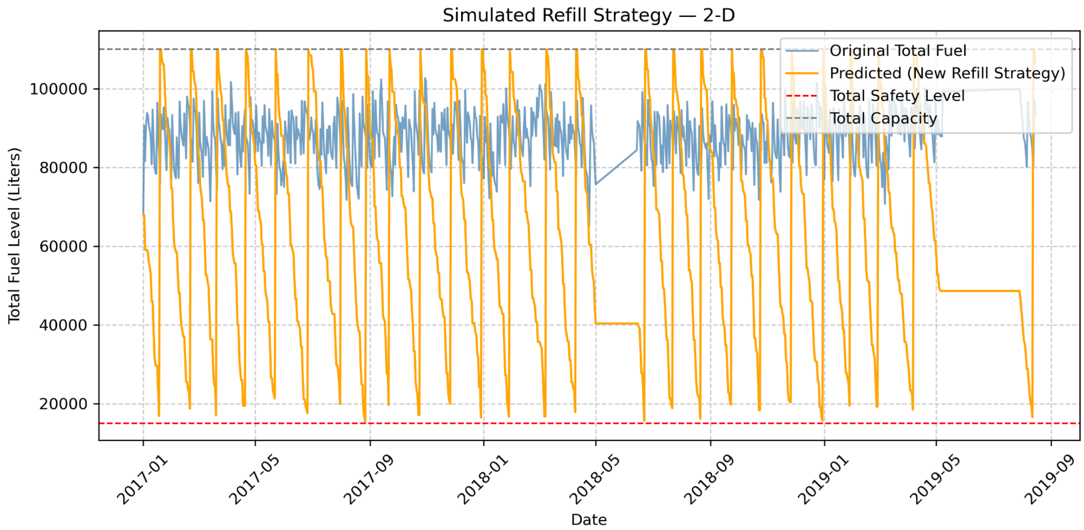
Station 1' diesel tanks include tank 10 and tank 15. Their total capacity level is 80000L, and their total safety level is 10000L. We suggest a strategy to refill the 2 tanks up to their total capacity of 80000L every time when their total fuel_level meets their total safety line of 10000L.

For Station 1's diesel tanks, the simulated refill strategy (gold line) successfully maintains stable inventory levels between the safety threshold of 10,000 L and full capacity of 80,000 L. Compared with the original trajectory (blue line), the optimized plan prevents excessive depletion and eliminates erratic deep drops. Refill cycles become more regular and aligned with predictable consumption, ensuring operational reliability and improved inventory turnover.



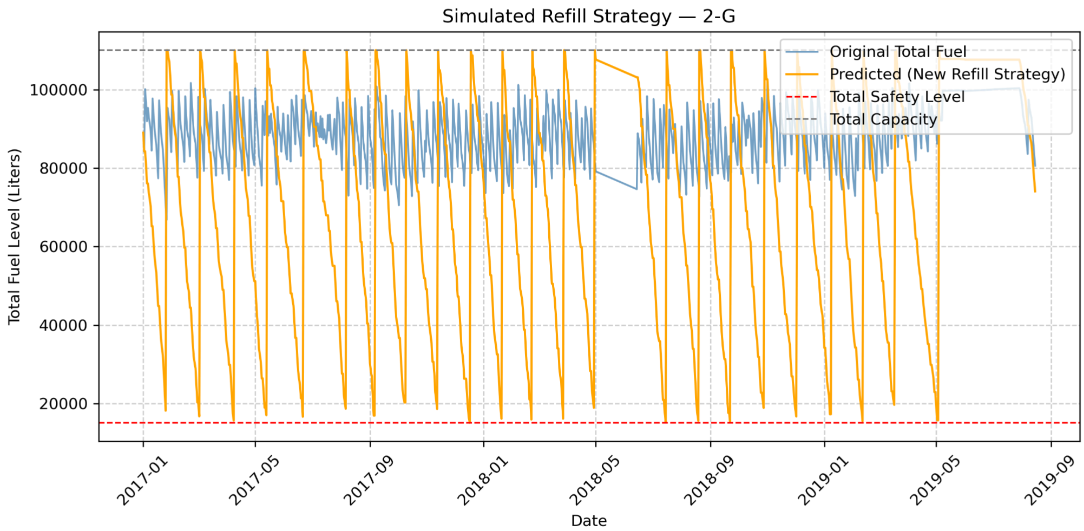
Station 1' gasoline tanks include tank 10, tank 11, tank 13, and tank 14. Their total capacity level is 160000L, and their total safety level is 20000L. We suggest a strategy to refill the 4 tanks up to their total capacity 160000L every time when their total fuel_level meets their total safety line of 20000L.

Station 1's gasoline inventory demonstrates the most substantial improvement. The optimized trajectory shows a consistent saw-tooth pattern oscillating between the 20,000 L safety level and 160,000 L capacity. The blue line previously indicated irregular replenishment, while the gold line now reflects more disciplined refills at full-tank levels. This results in smoother inventory cycles, reduced risk of shortage, and the largest overall cost saving across all stations (≈ $246 K CAD).



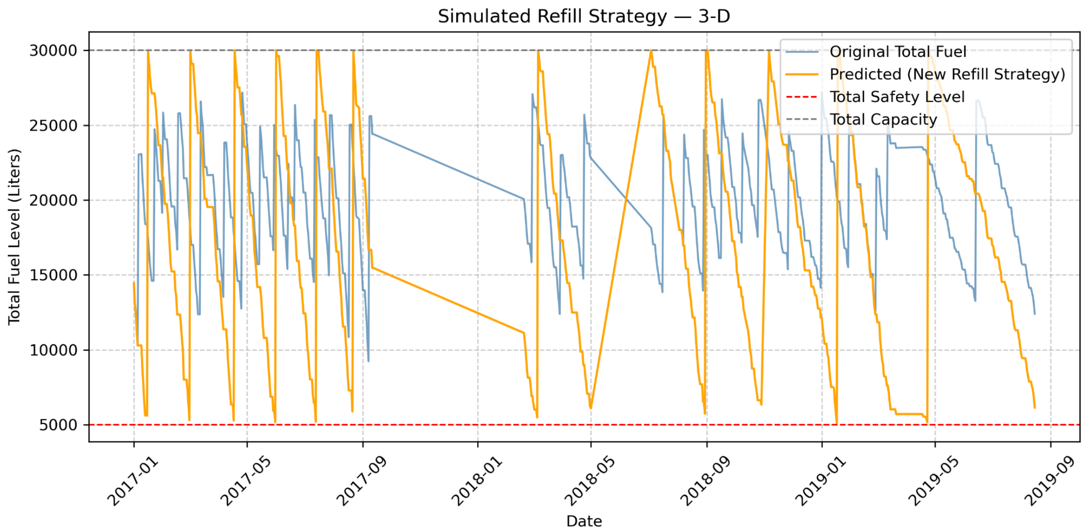
Station 2's diesel tanks include tank 17 and tank 19. Their total capacity level is 110000L, and their total safety level is 15073L. We suggest a strategy to refill the 2 tanks up to their total capacity of 110000L every time when their total fuel_level meets their total safety line of 15073L.

Under the new strategy, Station 2's diesel tanks maintain fuel levels well above the 15,000 L safety limit, fluctuating mainly between 50,000 L and 110,000 L. The optimized model performs fewer but higher-volume refills, improving cost efficiency while preserving supply security. The station achieves a saving ratio of 3.28 %, confirming the effectiveness of volume-based replenishment.



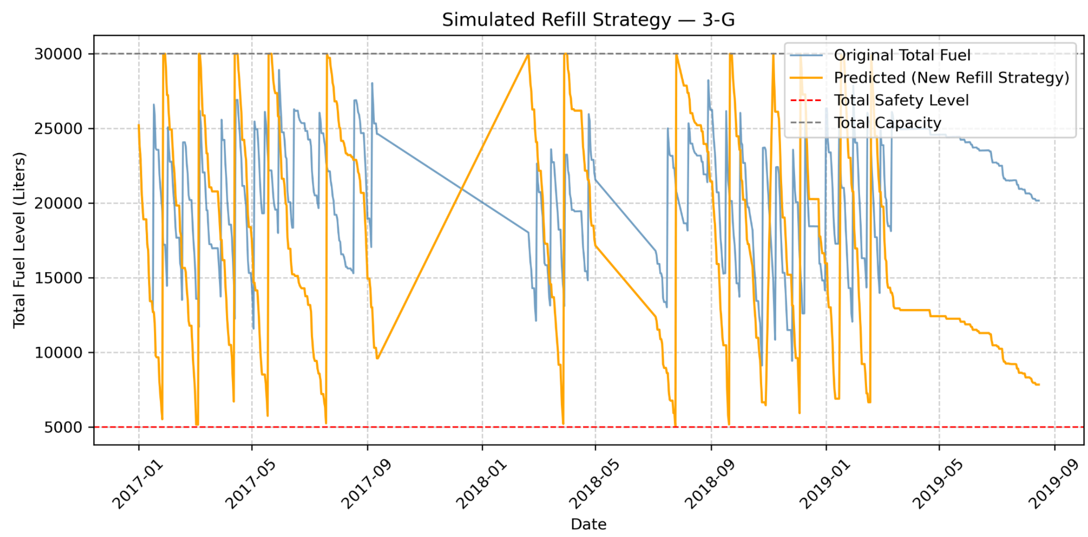
Station 2's gasoline tanks include tank 16 and tank 18. Their total capacity level is 110000L, and their total safety level is 15073L. We suggest a strategy to refill the 2 tanks up to their total capacity of 110000L every time when their total fuel_level meets their total safety line of 15073L.

The gasoline tanks at Station 2 show clear stabilization effects. The new trajectory ensures consistent refills to full capacity, avoiding the sharp downward spikes evident in the historical pattern. The improved timing aligns with supplier discount thresholds, yielding the highest saving ratio among all units (3.6 %). This demonstrates that combining capacity-based triggers with discount-tier awareness maximizes both operational efficiency and financial return.



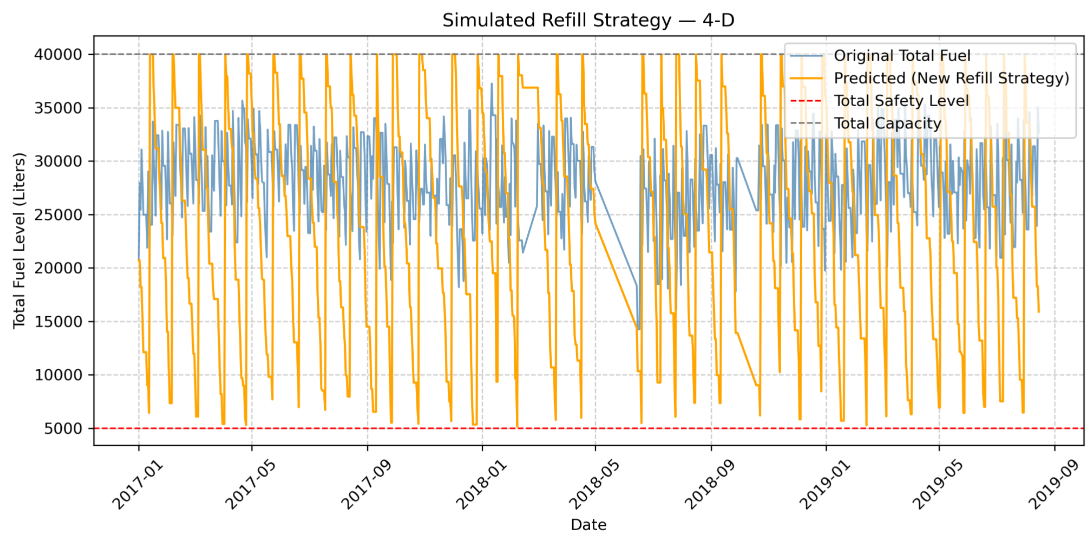
Station 3's diesel tanks only include tank 21. The capacity level is 30000L, and the safety level is 5000L. We suggest a strategy to refill the tank up to a total capacity of 50000L every time when fuel level meets the safety line of 5000L.

Station 3's diesel tank, with a total capacity of **30,000 L**, shifts from frequent small refills to fewer, more efficient full-capacity replenishments. The predicted curve indicates deliberate consumption down to the **5,000 L safety mark**, followed by instant restoration to capacity. This pattern smooths out volatility and achieves better discount utilization, reducing the overall purchasing frequency without increasing stock-out risk.



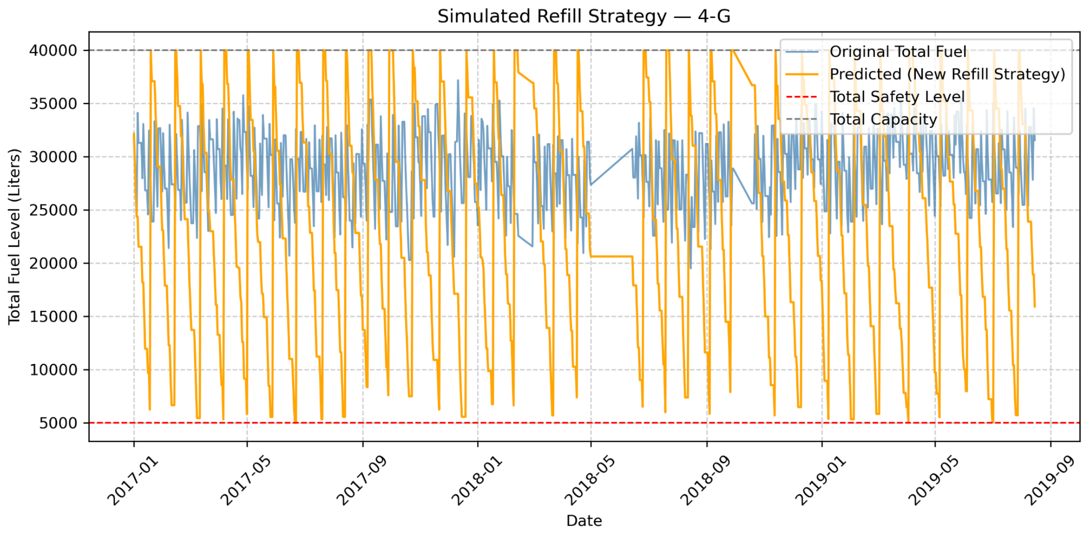
Station 3's gasoline tanks only include tank 20. The capacity level is 30000L, and the safety level is 5000L. We suggest a strategy to refill the tank up to a total capacity of 30000L every time when fuel level meets the safety line of 5000L.

For Station 3's gasoline tank, the optimized strategy produces a similarly disciplined pattern. The gold line follows a clear downward-and-refill sequence, keeping fuel levels consistently within the safe operating range. In contrast, the original series shows irregularities and occasional understocking. The new plan ensures stable operation, validating the model's adaptability to smaller-scale stations.



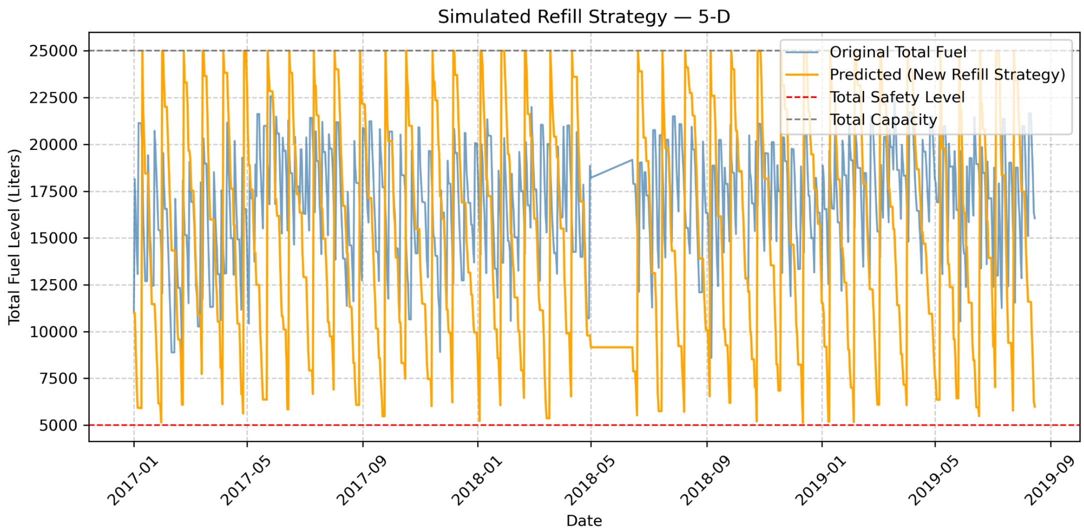
Station 4's diesel tanks only include tank 23. The capacity level is 40000L, and the safety level is 5000L. We suggest a strategy to refill the tank up to a total capacity of 40000L every time when fuel level meets the safety line of 5000L.

Station 4's diesel tank's optimized line oscillates predictably between 5,000 L and 40,000 L. The simulation confirms that the revised timing achieves a steady balance between consumption and replenishment, preventing both overstocking and low-fuel emergencies.



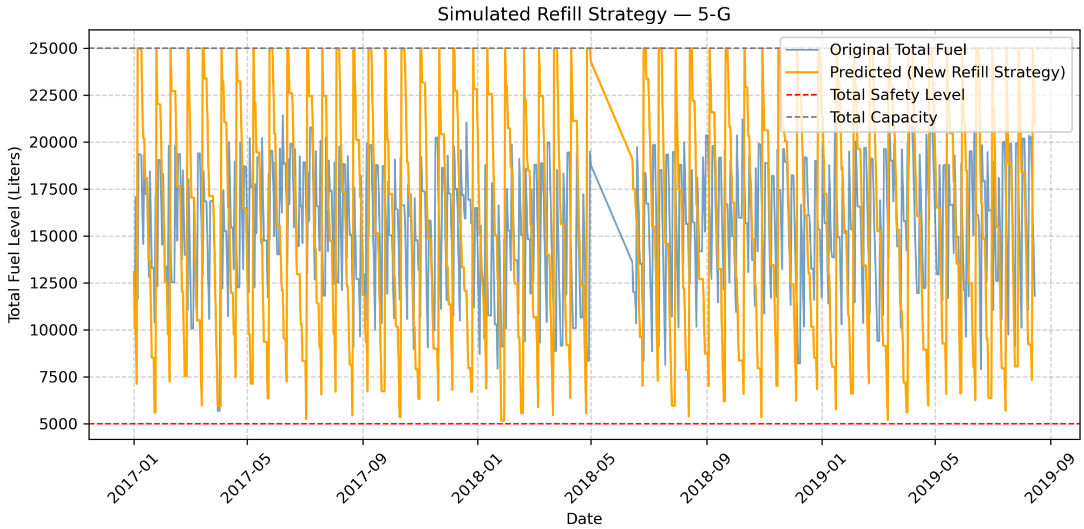
Station 4’s gasoline tanks only include tank 22. The capacity level is 40000L, and the safety level is 5000L. We suggest a strategy to refill the tank up to a total capacity of 40000L every time when fuel level meets the safety line of 5000L.

The gasoline inventory at Station 4 displays a nearly ideal saw-tooth pattern under the new plan. Refills are triggered precisely when levels approach the 5,000 L safety line, reaching full capacity immediately thereafter. Compared to the historical curve, the improved schedule enhances stock predictability and shortens idle tank periods, resulting in a steady 2.74 % cost saving.



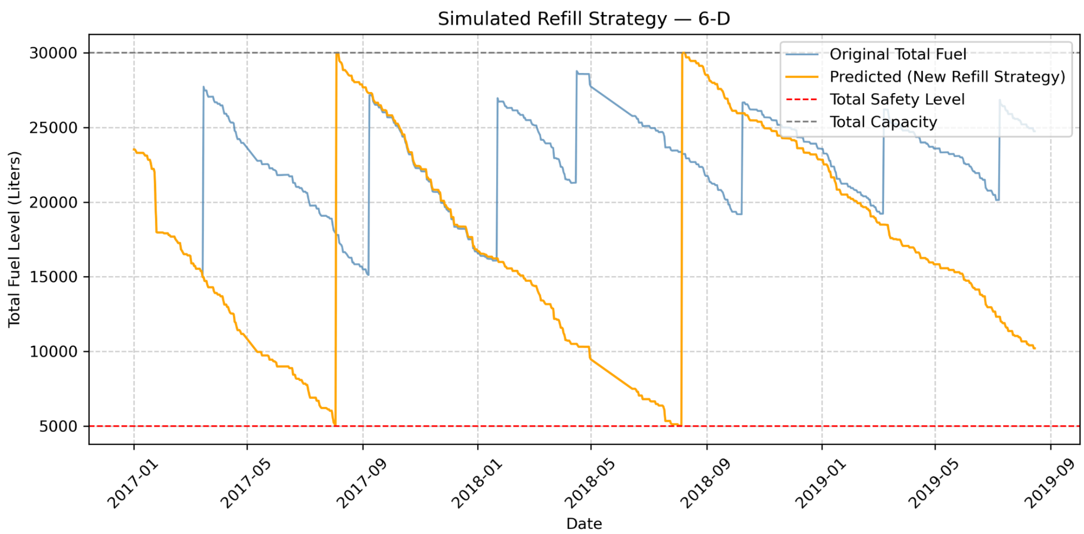
Station 5’s diesel tanks only include tank 24. The capacity level is 25000L, and the safety level is 5000L. We suggest a strategy to refill the tank up to a total capacity of 25000L every time when fuel level meets the safety line of 5000L.

For Station 5's diesel tank (capacity = 25,000 L), the optimized simulation shows a consistent oscillation between the 5,000 L safety level and full capacity, without any safety violations. The blue line indicates more frequent minor refills, while the golden trajectory reflects efficient, consolidated replenishment cycles. This streamlined behavior confirms that even smaller stations can gain operational benefits from a structured refill rule.



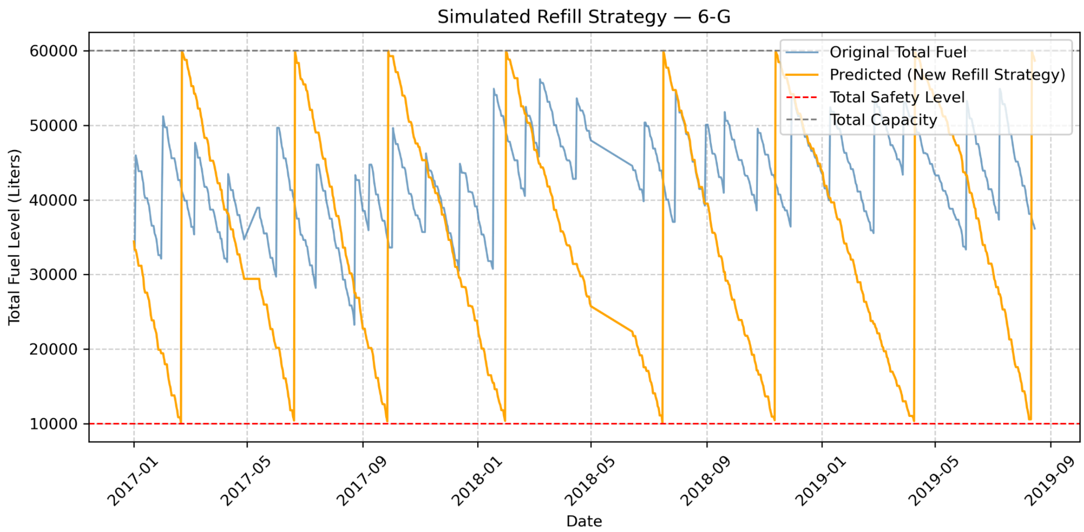
Station 5's gasoline tanks only include tank 25. The capacity level is 25000L, and the safety level is 5000L. We suggest a strategy to refill the tank up to a total capacity of 25000L every time when fuel level meets the safety line of 5000L.

The gasoline tank at Station 5 follows a similar improvement trend. The new trajectory demonstrates stable periodic replenishment cycles, maintaining safe levels while reducing unnecessary refills. Although absolute savings are moderate (≈ $27 K CAD), the optimization enhances reliability and ensures consistent service availability.



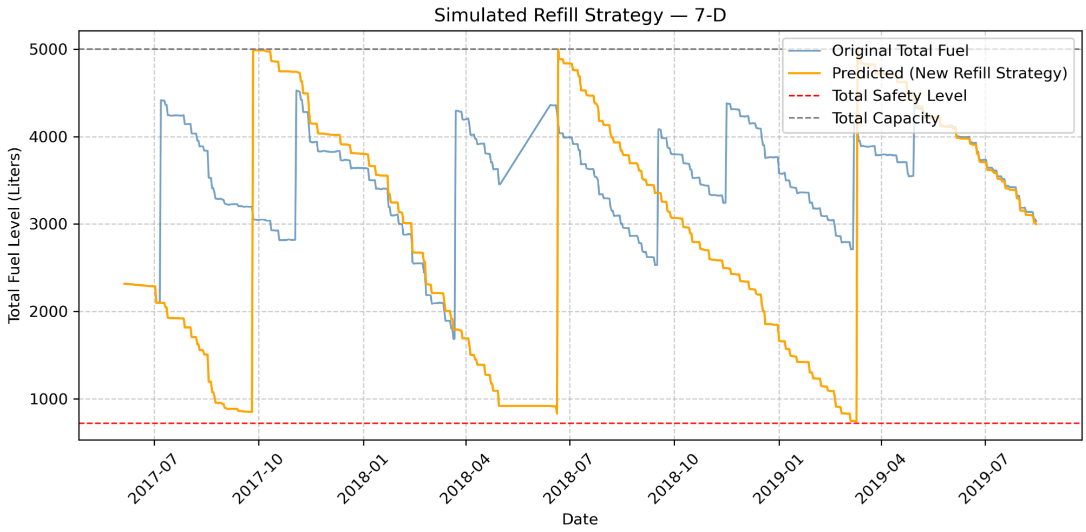
Station 6's diesel tanks only include tank 28. The capacity level is 30000L, and the safety level is 5000L. We suggest a strategy to refill the tank up to a total capacity of 30000L every time when fuel level meets the safety line of 5000L.

For Station 6's diesel tank, the optimized refill pattern (gold line) shows infrequent but well-timed replenishments, maintaining inventory safely between 5,000 L and 30,000 L. The original trajectory (blue line) reflects inconsistent refill intervals. The simulated strategy now adopts a clearer rule - refill only when fuel approaches the threshold - resulting in fewer but fuller replenishments. This behavior minimizes logistics cost and stockouts, delivering a 2.48% saving ratio while maintaining operational security.



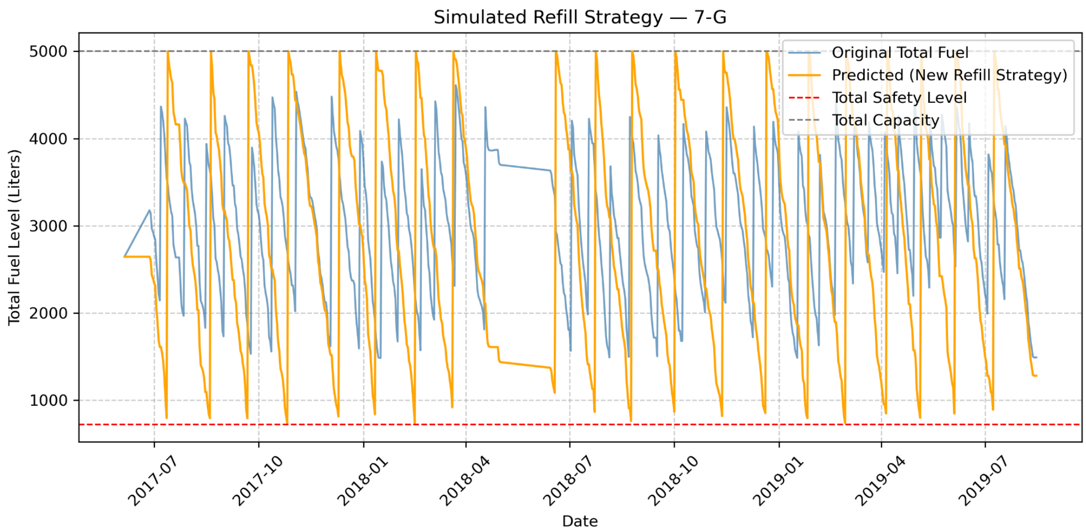
Station 6's gasoline tanks include tank 26 and tank 27. Their total capacity level is 60000L, and their total safety level is 10000L. We suggest a strategy to refill the 2 tanks up to their total capacity of 60000L every time when their total fuel_level meets their total safety line of 10000L.

The gasoline tanks at Station 6 show one of the clearest improvements among smaller stations. The gold curve follows a disciplined cycle - fuel levels decline steadily to the 10,000 L safety line before being fully replenished to 60,000 L. The pattern eliminates minor fluctuations visible in the original series, suggesting more efficient refill coordination. Station 6-G achieved a 3.34% saving ratio, the highest within its size category, confirming strong benefits from capacity-based triggers.



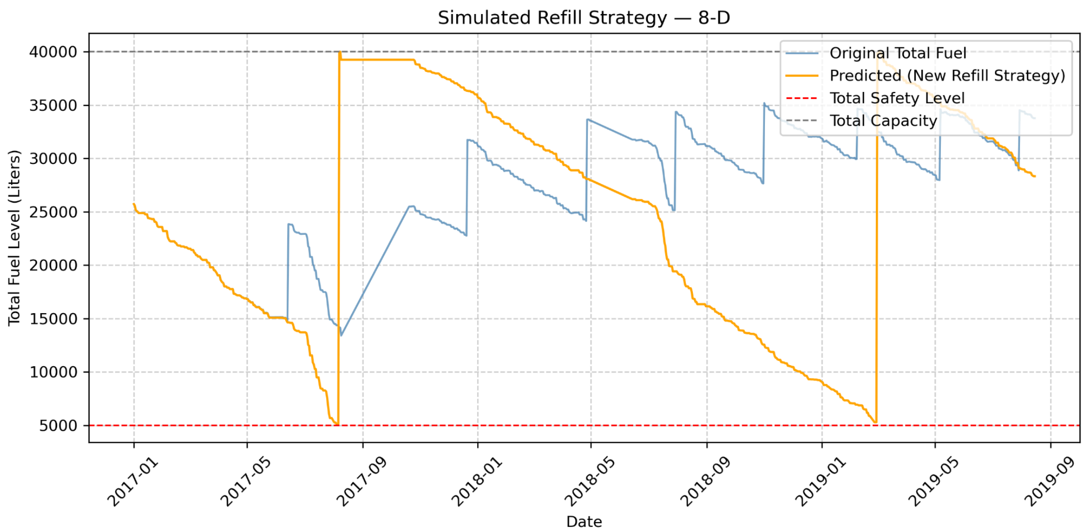
Station 7's diesel tanks only include tank 30. The capacity level is 5000L, and the safety level is 719.5L. We suggest a strategy to refill the tank up to a total capacity of 5000L every time when fuel level meets the safety line of 719.5L.

For Station 7's diesel tank (capacity 5,000 L), the predicted line keeps inventory oscillating safely above the 700 L threshold, refilling to full whenever levels approach it. Compared with the sparse and delayed replenishment of the original trend, the simulated strategy guarantees continuous availability and eliminates stockouts, even for low-volume tanks.



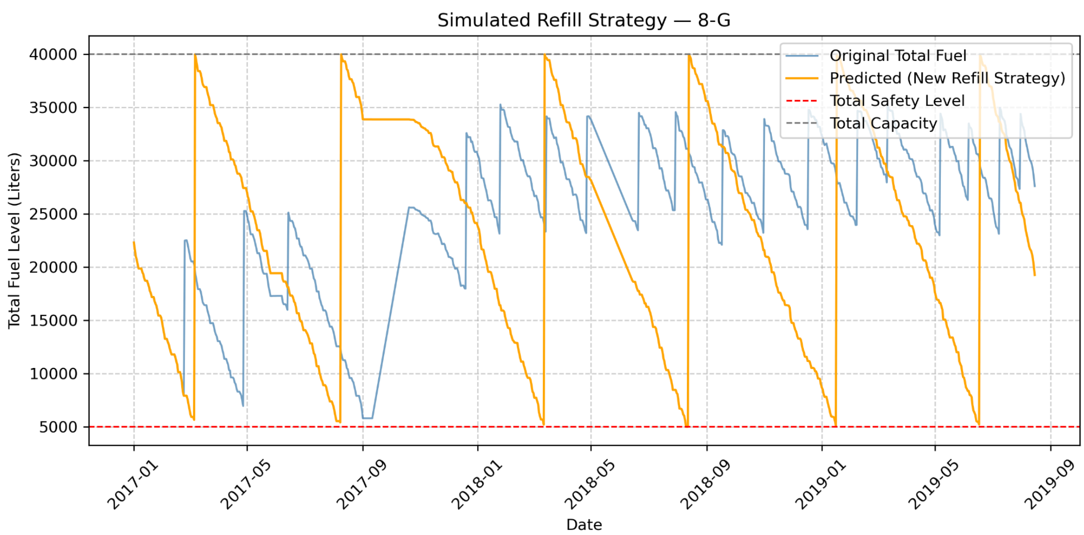
Station 7's gasoline tanks only include tank 29. The capacity level is 5000L, and the safety level is 719.5L. We suggest a strategy to refill the tank up to a total capacity of 5000L every time when fuel level meets the safety line of 719.5L.

The gasoline tank at Station 7 displays a highly responsive and frequent refill pattern. Both original and simulated curves show rapid consumption cycles, but the optimized plan maintains more consistent refill timing and avoids dips below the 700 L safety mark. The high turnover frequency matches the small tank capacity, ensuring reliability despite limited storage.



Station 8’s diesel tanks only include tank 31. The capacity level is 40000L, and the safety level is 5000L. We suggest a strategy to refill the tank up to a total capacity of 40000L every time when fuel level meets the safety line of 5000L.

For Station 8’s diesel storage, the optimized trajectory significantly stabilizes inventory variation. The fuel level now declines gradually from 40,000 L to the 5,000 L safety line, followed by instant full replenishment. The smooth, consistent cycle contrasts sharply with the uneven original trend, indicating tighter operational control and improved scheduling.



Station 8’s gasoline tanks only include tank 32. The capacity level is 40000L, and the safety level is 5000L. We suggest a strategy to refill the tank up to a total capacity of

40000L every time when fuel level meets the safety line of 5000L.

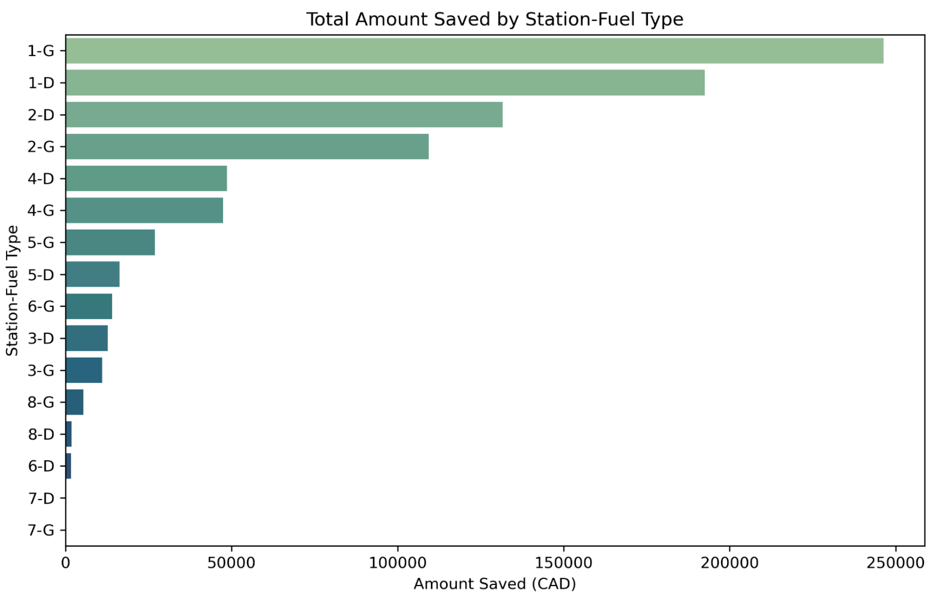
The gasoline tank at Station 8 follows a disciplined pattern under the simulated strategy, oscillating between 5,000 L and 40,000 L. The optimized refills are well-timed and consistent, ensuring that the tank never drops below the safety limit. Compared with the irregular replenishments of the past, the improvement enhances both safety and predictability, supporting stable long-term operations.

5.2 Calculate Amount Saved Based on New Strategy

In this stage, we developed and simulated an optimized fuel refill strategy aimed at reducing purchasing costs while maintaining safe inventory levels across all stations and fuel types.

First, we calculated each station-fuel combination's baseline price per liter before discounts by reverse-engineering the tiered pricing structure. Based on the discount tiers ("Tier 0-3"), we added the appropriate margin to reconstruct the undiscounted price. We then computed the total historical purchase cost under previous strategies and the expected cost under new refill strategies, which adjust the purchasing frequency according to each tank's total capacity and safety level.

Next, we derived the new optimized price per liter by simulating different capacity utilization thresholds - larger tanks or higher refill volumes received deeper discounts. The resulting "new purchase cost" was compared against the previous cost to obtain both total amount saved (CAD) and saving ratio (%).



	station_fuel_type	amount_saved	saving_ratio(%)
0	1-D	1.93e+05	3.27
1	1-G	2.46e+05	2.00
2	2-D	1.32e+05	3.28
3	2-G	1.09e+05	3.60
4	3-D	1.26e+04	2.56
5	3-G	1.11e+04	2.48
6	4-D	4.86e+04	2.47
7	4-G	4.74e+04	2.74
8	5-D	1.62e+04	1.63
9	5-G	2.69e+04	1.81
10	6-D	1.64e+03	2.48
11	6-G	1.40e+04	3.34
12	7-D	1.82e-12	0.00
13	7-G	0.00e+00	0.00
14	8-D	1.79e+03	2.05
15	8-G	5.39e+03	2.32

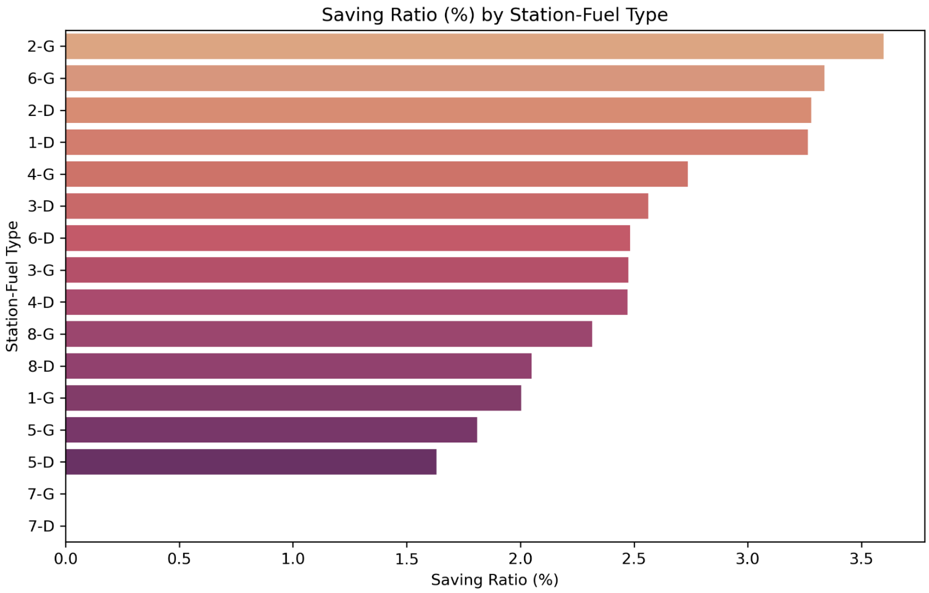
This *“Total Amount Saved by Station-Fuel Type”* chart, illustrates the absolute monetary savings (in CAD) for each station-fuel combination.

Station 1 demonstrates the most substantial savings for both fuel types, achieving approximately $246,000 CAD for gasoline (G) and $193,000 CAD for diesel (D). Together, Station 1 contributes nearly $440,000 CAD, which accounts for over one-third of total savings across all stations. This is mainly attributed to its larger tank capacities (160,000 L for G and 80,000 L for D) and correspondingly high purchase volumes (10.9 M L and 4.9 M L, respectively).

Station 2 follows closely, with $132,000 CAD (D) and $109,000 CAD (G) saved. Despite slightly smaller total capacities (110,000 L each), both fuel types benefit from a higher per-liter discount (reduction of $0.04 per L), producing saving ratios above 3%.

Station 4 ranks third, saving roughly $48,600 CAD (D) and $47,400 CAD (G).

Moderate savings appear in Stations 3, 5, 6, where total savings range between $10K-30K CAD, while Stations 7 and 8 contribute minimal gains due to smaller storage volumes (≤ 5,000-40,000 L).



The chart *"Saving Ratio (%) by Station-Fuel Type,"* emphasizes relative efficiency rather than absolute value.

The highest saving ratios are achieved by Station 2-G (3.6%) and Station 6-G (3.3%), indicating the optimized strategy yields particularly strong benefits for gasoline storage in medium-sized tanks. Diesel at Stations 1 and 2 also performs efficiently (3.27% and 3.28%), showing consistent cost sensitivity to capacity-based discounting.

In contrast, Station 7 (D & G) records 0% savings, as its extremely limited capacity (only 5,000 L) fails to reach any higher discount tier. Stations 5 and 8 yield modest savings (1.6-2.3%), primarily constrained by smaller purchase quantities and limited discount leverage.

Overall, the optimized refill plan provides both significant total cost reductions and notable proportional improvements, especially for stations with high-capacity tanks and frequent purchases. This confirms that adjusting refill timing to maximize tier discounts can generate measurable financial benefits without increasing operational risks.

6. Summary

This comprehensive analysis provides a data-driven evaluation of fuel inventory management across eight gas stations, integrating transactional, operational, and temporal perspectives. By merging datasets on fuel levels, invoices, tanks, and locations, the study established a unified analytical framework that connects purchasing behavior with storage capacity, pricing structure, and operational efficiency.

At the descriptive level, the results reveal substantial heterogeneity in both tank capacity and purchasing practices. Large stations such as EastMount and Eastgate consistently handle high transaction volumes but operate with short replenishment intervals and low fill ratios, indicating frequent small-batch refills that raise logistics and opportunity costs. Conversely, smaller stations, including Circle and Chappel, show infrequent deliveries and low utilization rates, exposing them to higher risks of stockouts and idle capacity.

The cost structure analysis demonstrates that the tiered discount mechanism exerts a measurable impact on overall procurement efficiency. Over 2,500 transactions fell within the no-discount tier, while near-threshold purchases (96 orders for Tier 1 and 25 for Tier 2) collectively missed potential savings of ≈ $33,358 CAD. This finding underscores the need for volume-aware ordering policies and automated alerts that encourage minor volume adjustments to unlock higher discount tiers.

Temporal analysis further uncovered weekly pricing and quantity cycles. Average adjusted fuel costs vary modestly, about 0.10-0.15 CAD per liter, but show consistent peaks on Tuesdays and Fridays. Purchase quantities, meanwhile, concentrate on Tuesdays and Fridays as well, revealing a partial timing mismatch between price advantage and purchase intensity. Aligning high-volume orders with low-price days could therefore generate additional cost optimization without disrupting operations.

Station-level investigations of replenishment intervals and fill ratios confirmed the absence of a standardized refueling policy. Stations 1 and 2 replenish too frequently (every 0.7-1.5 days, ≈ 5 % fill ratio), while stations 6-8 replenish too infrequently (18-33 days). Optimal performance lies in the moderate range - stations 3-5 achieve balanced cycles of 2-11 days and 15-20 % fill ratios. This differentiation highlights both over-supply and under-supply inefficiencies across the network.

Performance metrics at the tank level, including turnover rate and low-safety-stock frequency, provided a quantitative basis for operational risk assessment. High turnover stations exhibit strong demand but greater exposure to stockout risk, while low turnover stations maintain excess inventory, tying up working capital. Striking a balance between capital efficiency and reliability is therefore essential.

Finally, the simulation of an optimized refill strategy demonstrated clear financial and operational benefits. By synchronizing refill timing with tank capacity and discount thresholds, the model achieved both significant total cost savings - led by Station 1 (≈ $440 K CAD) - and smoother, more predictable inventory cycles. Across all stations, the optimized approach reduced refill frequency, minimized low-stock occurrences, and increased per-transaction efficiency without compromising service continuity.

Overall, this project illustrates how integrated data analytics can transform fuel-supply decision-making. By unifying cost, capacity, and temporal analyses, the study identifies actionable levers for improving profitability and resilience.

Implementing predictive scheduling, discount-tier optimization, and real-time monitoring systems can enable each station to achieve a sustainable balance between operational stability, cost control, and capital utilization, establishing a scalable model for data-driven fuel inventory management.