Python for Data Analysis Project: Gas Station Inventory Management Optimization Analysis

Group Members: Bahaa Joudieh, Carly Craffen, Jordan Richardson, Vivian Wu
Johns Hopkins Carey Business School
Nov 05, 2023

1. Introduction

In this project, our team is tasked with analyzing the inventory management practices of 8 gas stations and giving recommendations on how these stations could optimize their cost savings.

1.1 Background and Objective

Gas stations always purchase fuel (gasoline or diesel) in large amounts and sell it to customers. Some gas stations might be tempted to order gas more frequently with smaller amounts each time, but just like wholesale purchases are always associated with some discount - the more a gas station orders fuel each time, the greater the discount rate. With a cost-effective inventory management strategy, however, a gas station could order gas in more significant amounts and less frequently to benefit from discounts without running out of gas.

Our team aims to analyze each gas station's inventory replenishment strategy using the provided dataset and enlist recommendations to maximize savings and optimize inventory management. We first performed a rigorous analysis of the dataset to understand the current purchase strategy that each gas station adopts. As we figured out their inventory trajectory, we calculated how many purchases benefit from discounts and how much potential savings these gas stations could achieve with an optimized inventory management strategy. As a final outcome, we will provide managerial recommendations based on purchase strategy to maximize savings.

2. Data Collection & Preparation

2.1 Overview of the Data

We are given five datasets:

1. The first dataset, called 'Locations,' provides the locations of all gas stations with their respective names.
2. The second dataset, called 'Tanks,' provides information on the tanks at each gas station, including tank capacity, tank ID, tank location, and tank type.
3. The third dataset, 'Invoices,' includes invoice information across all tanks in every gas station. In other words, this dataset consists of all inventory purchase history of all tanks, including each purchase amount, purchase cost, date, and purchase location.
4. The fourth and fifth datasets, called 'Fuel Level 1' and 'Fuel Level 2', contain fuel level information of each tank at frequent and regular time stamps. These datasets allow us to understand the amount of fuel in each tank throughout the timeframe this analysis covers.

2.2 Data Cleansing & Preparation

To prepare all the datasets, we underwent deep data cleaning before creating any data visualizations. We first cleaned duplicated values from all datasets. We reviewed all of the N/A values in each dataset and found only two N/A values, so we decided to remove those rows to ensure the preciseness of the analysis. Next, we renamed all the columns to remove the spaces in the names. For example, in df_fuel_level_1, the column name was 'Tank ID'; we had renamed it 'Tank_ID' instead to make it easier to use those column names later. For data reflecting date/time information, we converted the data type from 'object' to 'datetime' so that we could later use the 'Time_stamp' column for data analysis.

Understanding Fuel Level Trajectory at Each Station

To prepare the data, we first combined 'Fuel Level 1' and 'Fuel Level 2' into one data frame called 'df_total'. Since 'Fuel Level 1' and 'Fuel Level 2' have the same columns, we use .concat() to directly combine these data frames. While our goal is to analyze the inventory management practices of each gas station, we need to be able to see which tank is located at which gas station. Therefore, we modified df_total by adding gas station name information from df locations and corresponding 'station name' to 'tank location' using .merge(). We also added tank capacity to the df_total. We end up with a data frame that includes the fuel level of each tank, the location of the tanks, and the tank capacity.

To understand how often fuel replenishes at each gas location, we grouped the fuel level history in df_total by the gas station location. We end up with 8 data frames, whereas df_FL_1 represents the fuel level change in gas station 1, and so on. We also grouped by all the tank IDs and created new data frames for all tanks; for example, df_T10 contains the data relating to Tank ID T 10. We have 22 data frames total with each tank ID.

Understanding Purchase Cost at Each Station

Next, we will prepare the invoice data frame. The first thing we did with the invoices data frame was create a new column called 'Cost per Liter' using the gross purchase cost divided by the amount purchased in liters to understand how much each gas station paid for the unit cost of gas. To consider the effect of inflation on the purchase cost, we created a data frame called 'df_inflation,' which contains the monthly inflation rate from January 2017 to August 2019 and the corresponding inflation multiplier. Back to the invoices data frame, we adjusted the gross purchase cost by dividing the original purchase cost by the respective inflation multiplier given the transaction date. We end up with a data frame that includes the adjusted costs for all purchases.

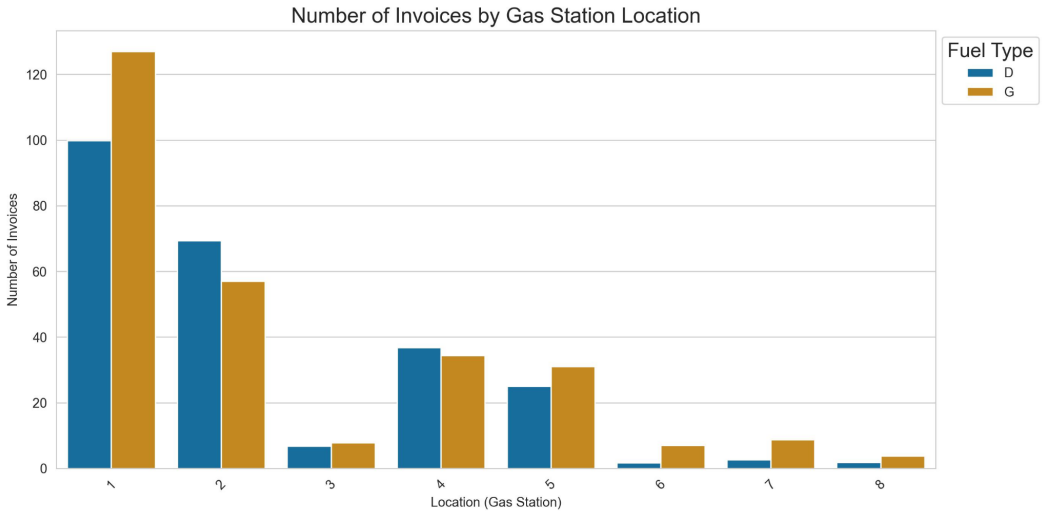
To understand what transactions happened at each gas station, we grouped the purchase history in df_invoices by gas station location. We name these data frames 'df_Inv_{station location ID}'; for example, 'df_Inv_1' includes all purchase data from gas station location 1. We have 8 data frames representing all 8 gas stations with invoice data. Finally, we end up with 8

sets of data frames, in which each of the 8 gas stations has one dataset of its fuel level history and one dataset of its purchase cost history.

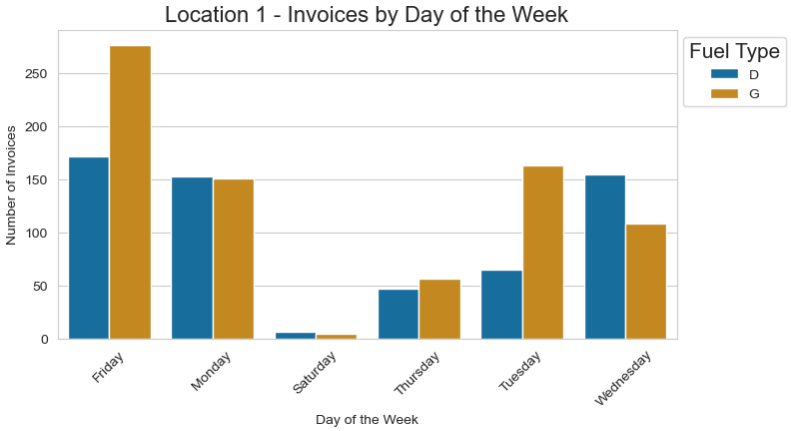
3. The Solution

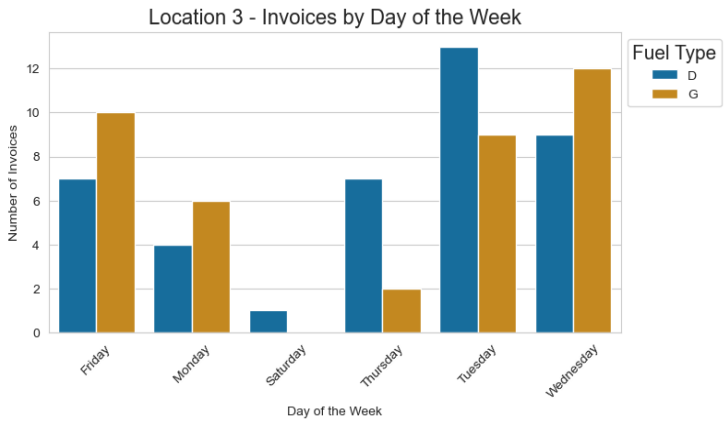
3.1 Adjusted Cost per Day of the Week

Next, we created data visualizations to better understand the data and develop recommendations. We first looked at the frequency of how many invoices we have based on the gas station location number and the fuel type. As you can see below, we have the highest frequency of invoices for gas station location 1. We also have a higher frequency of regular gas in comparison with diesel.

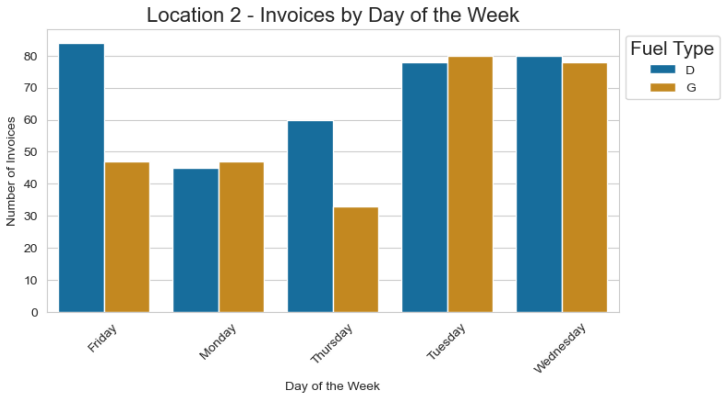


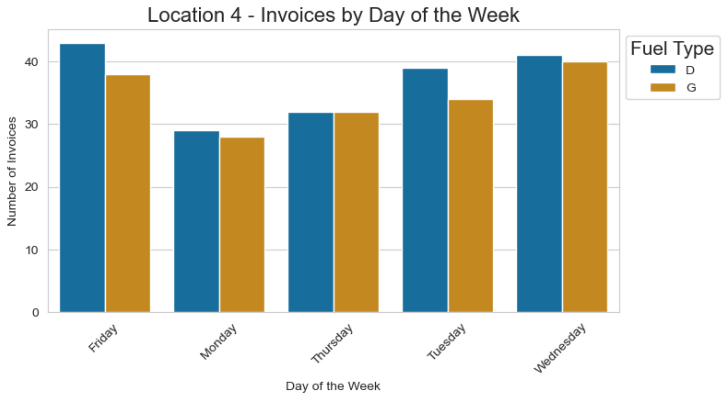
The next thing that we reviewed was how many observations we have of invoices for each day of the week for both diesel and regular gas by gas station location. For gas station location 1, we have the most invoices on Fridays for regular gas, and we have the least amount of invoices on Saturdays for diesel. For gas station location 3, we have the highest number of invoices on Tuesdays for diesel gas and the lowest on Saturdays for diesel.

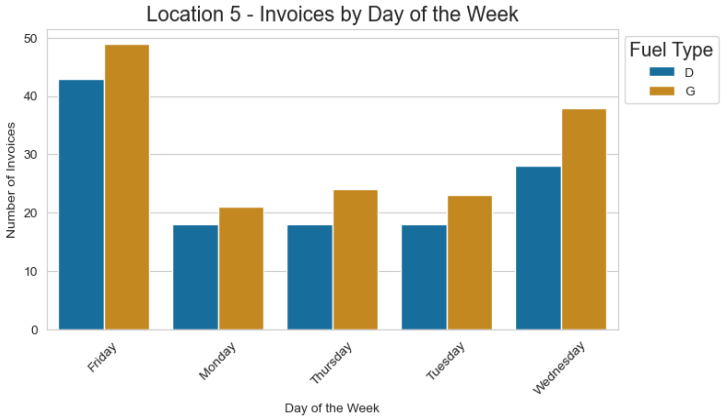


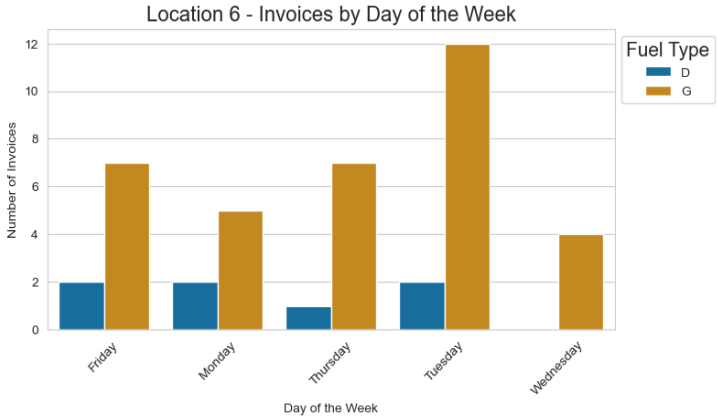


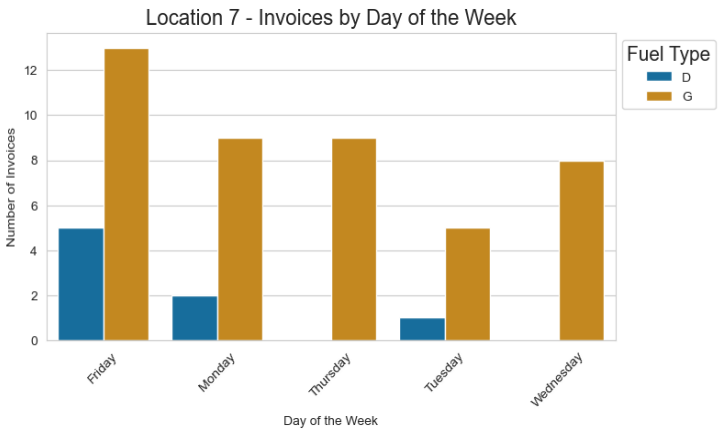
One thing to note about the following few gas station locations is that there are no observations on Saturdays for locations 2, 4, 5, 6, 7, and 8. For gas station locations 2 and 4, the highest number of invoices occurred on Fridays for diesel gas. For gas stations 5 and 7, the highest number of invoices occurred on Fridays for regular gas. Then, for gas station 6, the highest amount of invoices was on Tuesdays for regular, and for gas station 8, it was on Thursdays for regular.

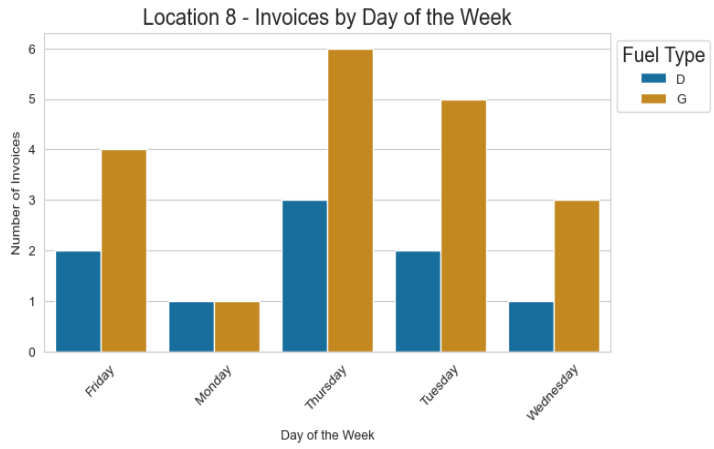




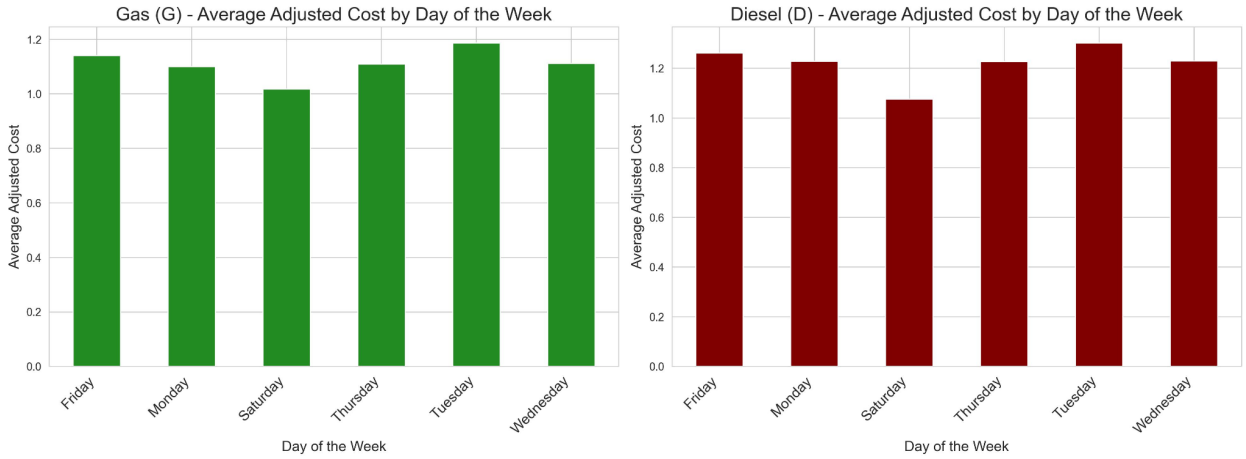






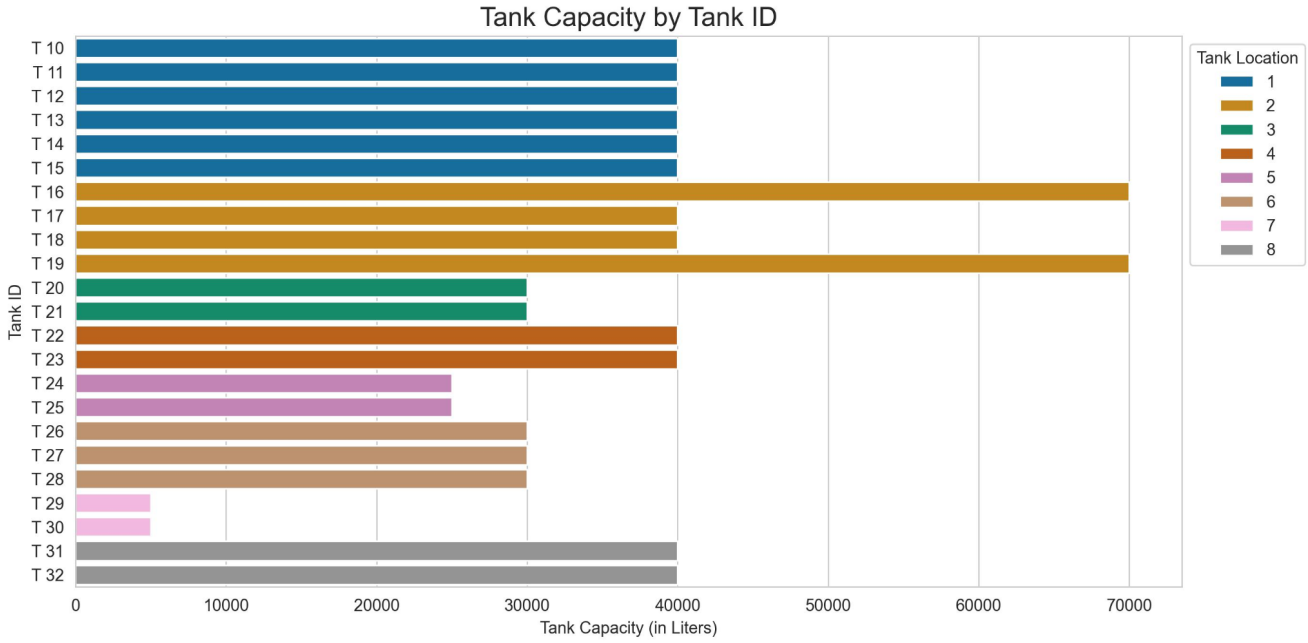


These analyses are essential to understand since we will be reviewing what day is the cheapest to buy fuel. Since we have no observations on Saturdays for gas station locations 2, 4, 5, 6, 7, and 8, this may affect purchasing decisions. The figure below shows which day is the cheapest to buy fuel based on the adjusted cost, which we calculated by dividing the gross purchase cost by the original purchase cost by the inflation multiplier. This graph shows that the cheapest day to purchase fuel is Saturdays since the adjusted price is $1.05, while the most expensive day is Tuesdays, with the adjusted cost being $1.23. We also calculated this by fuel type with two separate data frames for regular gas and diesel, but they depicted roughly the same numbers as shown below.



3.2 Tank ID, Tank Capacity, and Fuel Level Replenishments

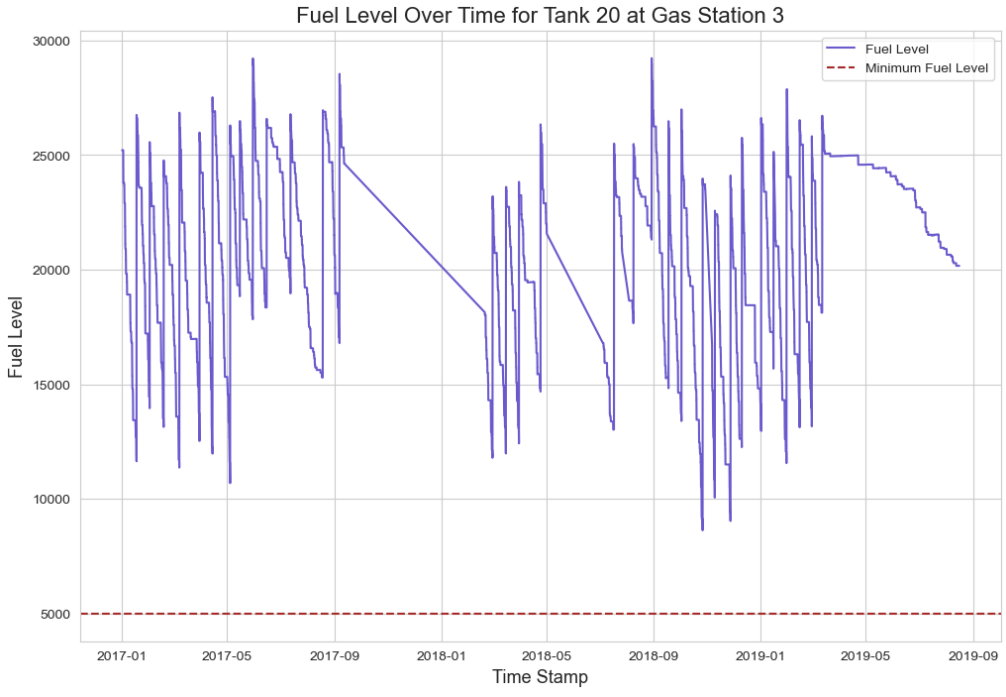
Next, we looked at the tank ID and the tank capacity. Tanks T16 and T19 (located at gas station location 2) have the highest tank capacity at 70,000 liters, which could mean that the gas station doesn't have to fill them up as often as compared to tanks T 29 and T30 located at gas station 7, which has the lowest capacity at 5,000 liters. Tanks T 20, T 21, T 26, and T 27 have a capacity of 30,000 liters. Tanks T 24 and T 25, located at gas station 5, have a capacity of 25,000 liters. However, the rest of the tanks have a capacity of 40,000 liters.



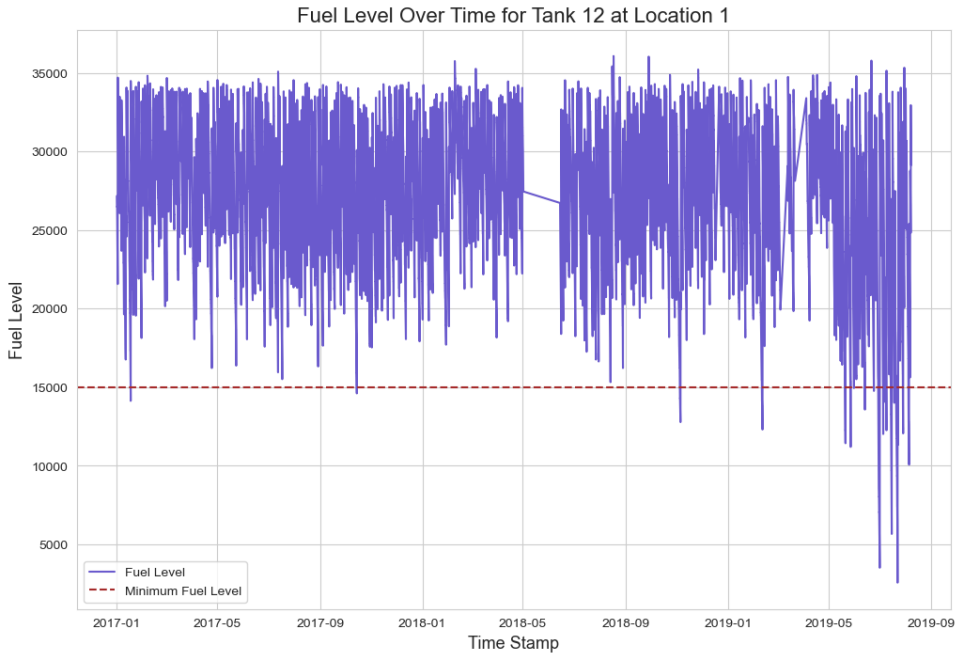
We used the above analyses of tank capacity to review how each gas station location utilizes the supplier discounts offered. Before doing that, below is the table of how the supplier discounts the amount of fuel purchased. For our remaining analysis, we will assume that gas station locations can maximize their fuel capacity at any given time. This assumption will enable us to calculate the maximum discount achievable.

Purchase Quantity (liters)	**Discount per Liter**
0 - 15,000	$0.00
15,000 - 25,000	$0.02
25,000 - 40,000	$0.03
40,000+	$0.04

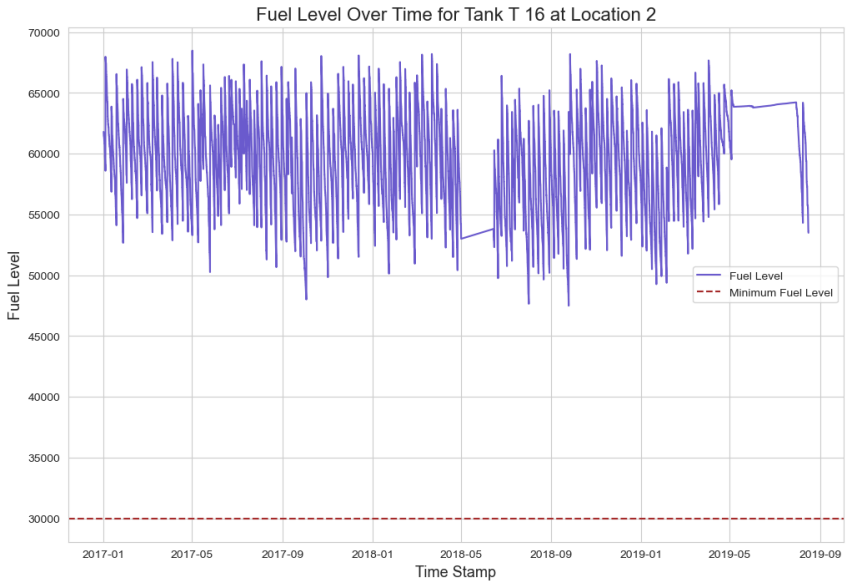
First, we visualized the times series of fuel levels for three tanks to better understand when the gas station purchases fuel. The first tank we reviewed is Tank T 20 at gas station location 3, which has a capacity of 30,000 liters. For the graph below, we calculated how Gas Station 3 could utilize the maximum discount. They would need to fill a minimum of 25,000 liters based on the table above, so we subtracted the capacity by the minimum to get 5,000 liters, which is depicted by the fixed red line. If the fuel level falls below that red line, it means that gas station 3 would be maximizing the discount that the supplier offers. When looking at the graph below, we noticed that the fuel level never falls below that red line, which indicates that the gas station is not utilizing the maximum supplier discount.



Next, we looked at the time series of fuel levels for Tank T 12 at gas station location 1, which has a tank capacity of 40,000 liters. To maximize the supplier discount, Gas Station 1 would need the fuel level to fall below 15,000 liters, as depicted with the fixed red line. Gas station 1 didn't utilize that benefit up until late 2018.



Lastly, we reviewed the time series of fuel levels for Tank T 16 at gas station 2, which has a capacity of 70,000 liters. To get the maximum discount, they need to fill a minimum of 40,000 liters, which leaves only 30,000 liters in the tank. As shown below, Gas Station 2 has yet to utilize the supplier discount since the fuel level has consistently stayed above the minimum.

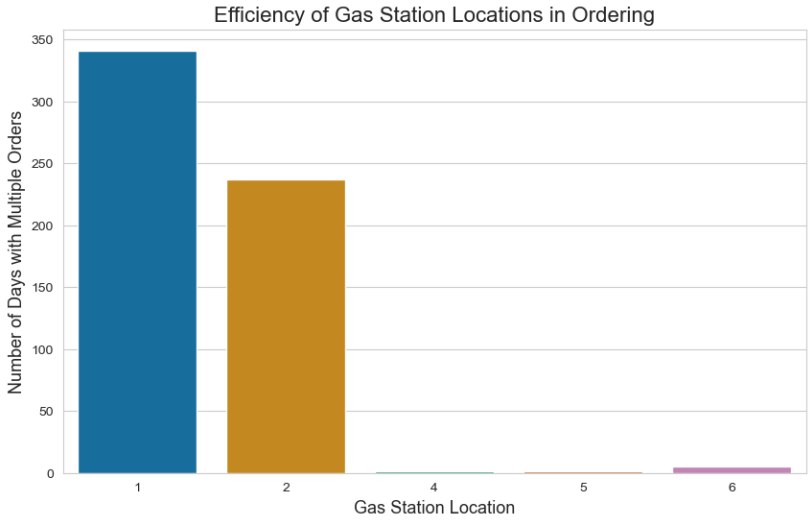


Next, we will review the percentages of how often each gas station benefits from the maximum discount offered based on tank capacity. Gas station location 1 benefitted the most from the discounts provided by the supplier, specifically for tank T 13, in which 65.89% of the replenishments received a discount (capacity for this tank is 40,000 liters). Also, from gas station location 1, 36.35% of the replenishments benefitted from the discount for tank T 10 (capacity for this tank is 40,000 liters). Gas station 8 came in third for tank T 32, with 12.30% of the replenishments benefitting from the discount of $0.04 per liter (capacity for this tank is 40,000 liters). Gas station locations 3 and 6 did not take advantage of the supplier discounts at all (the capacity for these tanks is 30,000 liters). Unfortunately, gas station 7 could not take advantage of any supplier discount due to both of their tanks (T 29 and T 30) only having a capacity of 5,000 liters each. Gas station location 2 has two tanks, T 16 and T 19, with the highest capacity of 70,000 liters out of all the other tanks, and they did not take advantage of the supplier discounts. Instead, gas station location 2 took advantage of the supplier discounts for tank T 17 (0.27%) and T 18 (0.12%).

Gas Station Location	Tank ID	Number of replenishments that benefited from the maximum discount.
1	T10	36.35%
	T11	2.26%
	T12	1.25%
	T13	65.89%
	T14	0.3%
	T15	0.33%
2	T16	0.00%
	T17	0.27%
	T18	0.12%
	T19	0.00%
3	T20	0.00%
	T21	0.00%
4	T22	0.00%
	T23	0.32%

5	T24	0.75%
	T25	5.34%
6	T26	0.00%
	T27	0.00%
	T28	0.00%
8	T31	1.52%
	T32	12.3%

One of the last things that we looked at in terms of fuel level replenishments is how many gas station locations ordered the same type of fuel more than twice a day. As shown below, gas station locations 1, 2, 6, 5, and 4 have ordered fuel more than twice a day. Specifically, gas station 1 has ordered more than twice a day 341 times out of 463 orders in total, which is roughly 73.65%.



3.3 Price Calculations for Gas Station Location 1

For our subsequent analyses, we will review the price calculations for only gas station location 1. As shown in a previous visualization, gas station location 1 has the following tanks: T 10, T 11, T 12, T 13, T 14, and T 15, all with a tank capacity of 40,000 liters. As a reminder, we created multiple data frames with these different tank IDs. In these data frames, we added a new column called 'Fuel Level Difference,' which calculates how much fuel was used between each observation. We then used that calculation to find the average consumption of fuel used per day for each tank ID for gas station location 1 by summing the Fuel Level Difference for days and dividing by the number of days in our data. The tank with the highest average consumption was T 12 at 4,786.56 liters daily. The tank with the lowest average consumption is T 13, which is 1627.83 liters daily.

```
The average consumption per day for T10 is 3,039.42 liters per day
The average consumption per day for T11 is 4,728.46 liters per day
The average consumption per day for T12 is 4,786.56 liters per day
The average consumption per day for T13 is 1,627.83 liters per day
The average consumption per day for T14 is 3,315.47 liters per day
The average consumption per day for T15 is 2,407.06 liters per day
```

Then, we calculated the weekly average consumption for each tank at gas station location 1 by simply multiplying the averages per day by 7. As you can see below, the tank with the highest weekly consumption is T 12 at 33,505.9 liters per week. The lowest weekly consumption is tank T 13 at 11,394.78 liters per week.

```
The average weekly comsumption for T10 is 21,275.96 liters
The average weekly comsumption for T11 is 33,099.23 liters
The average weekly comsumption for T12 is 33,505.9 liters
The average weekly comsumption for T13 is 11,394.78 liters
The average weekly comsumption for T14 is 23,208.27 liters
The average weekly comsumption for T15 is 16,849.41 liters
```

Since the weekly consumption for all tanks is below the capacity of 40,000 liters, they can all benefit from ordering once a week when the amount is between 25,000 and 40,000 liters to benefit from the maximum discount, which is $0.03 per liter.

- Tank T 10: gas station 1 could order, on average, 25,000 liters on Saturdays four times a month.
- Tank T 11: they could order, on average, 35,000 liters on Saturdays four times a month.
- Tank T 12: they could order, on average, 35,000 liters on Saturdays four times a month.
- Tank T 13: they could order, on average, 25,000 liters once every two weeks on Saturdays.
- Tank T 14: they could order, on average, 25,000 liters on Saturdays four times a month.
- Tank T 15: they could order, on average, 35,000 liters on Saturdays twice a month.

Based on these analyses, gas station location 1 could have had only 588 replenishments these past two years and a half (assuming an average of 30 days per month) instead of 1,361 replenishments.

Using these calculations and results, we created what the average monthly cost could have been for the replenishments of each tank ID. We used Saturday's cost per day and subtracted the maximum discount price of $0.03 to get the discounted price of $1.023. Using that

discounted price, we calculated the monthly cost of each tank based on the tank replenishment recommendations above. For example, for Tank T 10, we multiplied 25,000 liters by 4 by the discounted price to get $102,300, as shown below.

```
The average monthly estimated cost of replenishment for Tank T10 is $102,300.00
The average monthly estimated cost of replenishment for Tank T11 is $143,220.00
The average monthly estimated cost of replenishment for Tank T12 is $143,220.00
The average monthly estimated cost of replenishment for Tank T13 is $51,150.00
The average monthly estimated cost of replenishment for Tank T14 is $102,300.00
The average monthly estimated cost of replenishment for Tank T15 is $71,610.00
```

Therefore, the highest average monthly estimated cost would be for tanks T 11 and T 12, which is $143,220, and the lowest for tank T 13, which is $51,150. Based on these calculations, we determined the total estimated cost of replenishments for each tank for the entire length of the dataset, which is roughly two and a half years.

```
The total estimated cost of replenishment for tank T10 for the duration provided is $3,007,620.00
The total estimated cost of replenishment for tank T11 for the duration provided is $4,210,668.00
The total estimated cost of replenishment for tank T12 for the duration provided is $4,210,668.00
The total estimated cost of replenishment for tank T13 for the duration provided is $1,503,810.00
The total estimated cost of replenishment for tank T14 for the duration provided is $3,004,210.00
The total estimated cost of replenishment for tank T15 for the duration provided is $2,105,334.00
```

The highest estimated replenishment cost would be for tanks T 11 and T 12 at $4,210,668, and the lowest for tank T 13 at $1,503,810.

Therefore, Gas Station Location 1 could have paid $18,042,310 instead of $18,198,952.88 for the past two and a half years, meaning that Gas Station Location 1 could have saved $156,642.88 if they had followed the tank replenishment schedule above.

4. Summary & Recommendations

Throughout our analysis of these eight gas stations, we looked at their current operating procedures before implementing ours. The significant problems that we came across could be summarized in the following points:

- The gas stations' strategy for ordering fuel was entirely arbitrary.
- The gas stations were not benefitting from the discounts provided by the supplier.
- Some gas stations were ordering the same fuel type multiple times a day.

Based on our findings and our in-depth cost analysis for a sample (Location 1), we were able to address every single problem and come up with recommendations that apply either to all or some of the gas stations. First, Saturday proved to have the cheapest fuel price per liter. Only two gas stations were ordering fuel on Saturdays in small quantities. We recommend all gas stations adopt that approach, especially Location 7. Gas Station 7 has two small tanks and cannot

benefit from the supplier discounts; hence, the only way to save money is to have most fuel orders placed on Saturdays. Second, the supplier offers excellent discounts of 2, 3, and 4 cents per liter if the quantity ordered is between 15,000-25,000 liters, 25,000-40,000 liters, and above 40,000 liters, respectively. As per our analysis, gas station 1 has been benefiting the most from supplier discounts compared to the other locations, but only on tanks T10 and T13.

To further back our analysis, we performed an in-depth cost and tank analysis study for gas station location 1. We proposed an inventory management system that will allow gas station 1 to reduce the number of replenishments and benefit from the maximum supplier discount. Not only did our system allow location 1 to continue normal operations, but it also could have helped them save $156,642.88 during the past two years and a half.

Adopting a cost-effective inventory management strategy is a must for these gas stations. Following a frequent replenishment system in small quantities might appear attractive to these gas stations as it has less cash tied up to the fuel inventory. However, our team recommends that they follow a different approach. We recommend these gas stations shift their replenishment schedule to having the bulk of orders on Saturdays and opt for reducing the number of replenishments by ordering in large quantities and benefitting the supplier discount.