# Enhancing Gas Station Customer Experience through Waiting Time Analysis and Innovation

Houghton Queue Monitors

Rahul Teja Bolloju, Mahidhar Manda

## INTRODUCTION

### Two-way ANOVA test

Gas stations play a crucial role in facilitating refueling needs for drivers. One key aspect of customer satisfaction at gas stations is minimizing waiting time. In this project, we propose to investigate the average waiting time for customers at various gas stations and explore the impact of potential innovations aimed at reducing this waiting time.

### Data Description

We have collected data on customer waiting times from multiple gas stations across different locations in Houghton. At the gas station dataset contains, categorical variables include "Gas_station" for the names of stations where waiting times are logged, "Car_type" indicating vehicle types, and "Day" specifying the recording day of the week. Numerical variables consist of "No_pumps" representing total gas pumps, "In_time" denoting vehicle entry time, and "Out_time" indicating vehicle exit time.

#### Independent Variables (Factors):

In a two-way ANOVA, there are two independent variables, each with two or more levels or categories. For example, in our case study two independent variables are type of car and gas station. how this variable are effecting the waiting time of a customer.

- Gas_station
- Car_type

#### Dependent Variable:

This is the variable that is being measured or observed and is expected to change based on the independent variables. It is continuous in nature, such Waiting time.

- Waiting_time

# METHODS

In this study, the population of interest comprises customers visiting gas stations, and the observational unit is each individual customer. This research employed an observational study design rather than a controlled experiment.

The sample collection process involved gathering data on waiting times from various gas stations. To obtain a representative sample, efforts were made to ensure random sampling. Data collection included recording waiting times and categorizing them by gas station, car type, and other relevant factors.

We employed random sampling because we have conducted our study on the gas stations only in Houghton area. This approach aimed to minimize bias and increase the generalizability of findings to the broader population of gas station customers in the area.

Variables measured in this study included: - **Gas_station:** Categorical variable indicating the gas station where the observation was made. - **Car_type:** Categorical variable representing the type of car each customer arrived in. - **Waiting_time:** Continuous variable denoting the time each customer spent waiting.

Data cleaning involved removing outliers, such as extreme waiting times due to unusual circumstances like customers having meals in their cars. This process aimed to ensure the integrity of the dataset for analysis.

The methodology focused on capturing real-world observations of waiting times at gas stations, considering factors such as gas station location and customer car type. This comprehensive approach facilitated an in-depth analysis of waiting time dynamics and their potential influences.

# RESULTS

## Statistical Analysis:

### Two-way ANOVA test hypotheses

A two-way ANOVA test as a comprehensive tool in statistics that delves into the effects of two categorical independent variables on a continuous dependent variable. It's like examining the combined influence of two factors on the average outcome, while also segregation how these factors might interact and influence each other's impact. This analysis provides a clear understanding of how various factors contribute to the variability observed in the dependent variable, offering valuable insights for decision-making and further research.

*Null Hypotheses (H0):*

- H01: There is no significant difference in waiting times across different gas stations.
- H02: There is no significant difference in waiting times among different car types.

- H03: There is no significant interaction effect between gas station and car type on waiting times.

*Alternative Hypotheses (Ha):*

- Ha1: There is a significant difference in waiting times across different gas stations.
- Ha2: There is a significant difference in waiting times among different car types.
- Ha3: There is a significant interaction effect between gas station and car type on waiting times.

### Assumptions of two-way ANOVA test

Two-way ANOVA, similar to other ANOVA tests, relies on the assumption that the data within each group are normally distributed and have equal variances. We'll demonstrate how to verify these assumptions after performing the ANOVA analysis.

### Compute two-way ANOVA test in R:

*Import your data into R*

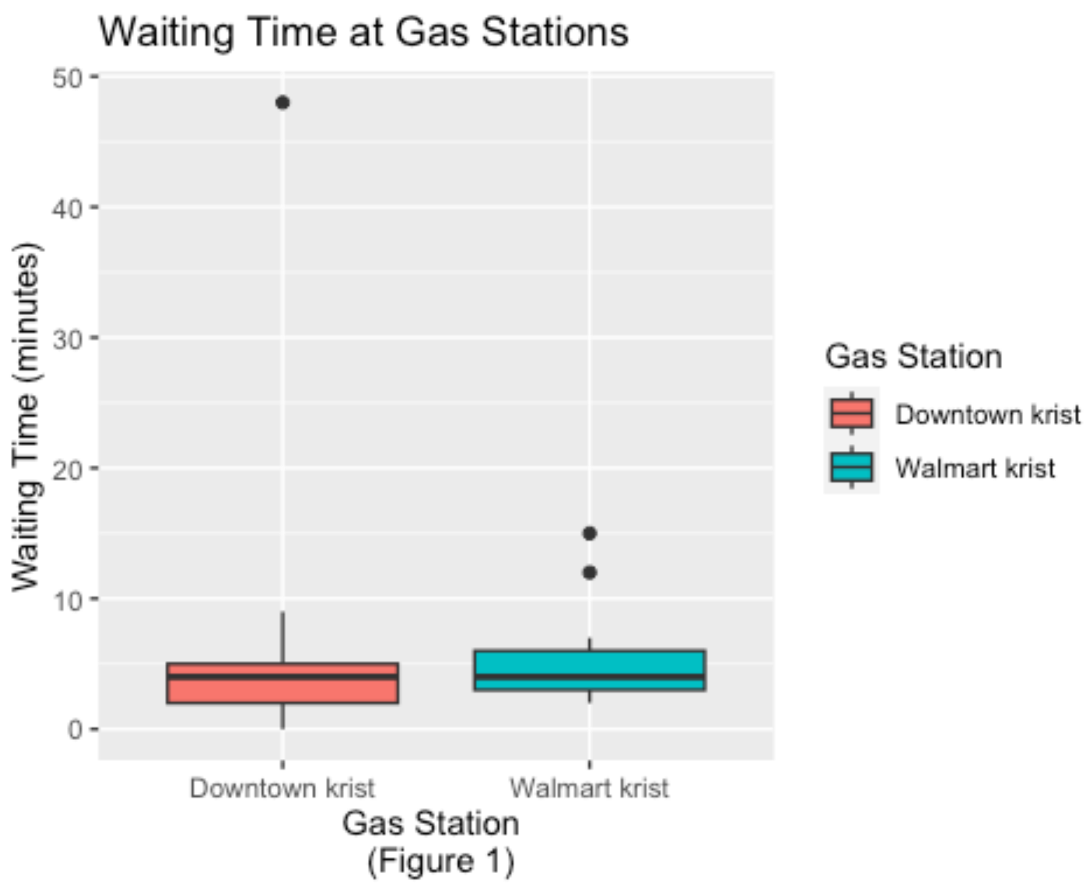
```
# data <- read.csv("//homes.mtu.edu/home/Downloads/Gas_station_data.csv")
data <- read.csv("/Users/rahulteja/Library/Mobile Documents/com~apple~CloudDo
cs/Desktop/COURSE/sem 4/MA 5701 Stastical Methods/project/Gas_station_data.cs
v")
data$Car_type <- tolower(data$Car_type)
head(data)

##     Gas_station No_pumps     Car_type         Date       Day In_time Out_t
ime
## 1 Walmart krist        8 pickup truck Mar 15, 2024    Friday   16:06    16
:13
## 2 Walmart krist        8        sedan Mar 15, 2024    Friday   16:09    16
:15
## 3 Walmart krist        8 pickup truck Mar 15, 2024    Friday   16:01    16
:13
## 4 Walmart krist        8          suv Mar 15, 2024    Friday   16:01    16
:16
## 5 Walmart krist        8 pickup truck Mar 15, 2024    Friday   16:11    16
:16
## 6 Walmart krist        8        sedan Apr 10, 2024 Wednesday   16:18    16
:21
##   Waiting_time
## 1            7
## 2            6
## 3           12
## 4           15
## 5            5
## 6            3
```

*Visualizing the data*

Boxplot

```
library(ggplot2)
ggplot(data, aes(x = Gas_station, y = Waiting_time, fill = Gas_station)) +
  geom_boxplot() +
  labs(title = "Waiting Time at Gas Stations",
       x = "Gas Station \n (Figure 1)",
       y = "Waiting Time (minutes)",
       fill = "Gas Station")
```




We can observe in the Figure 1 there is one outlier in the downtown gas station. While collecting the data we have observe that this outlier has happened due to a customer was having his lunch and waiting in the car.

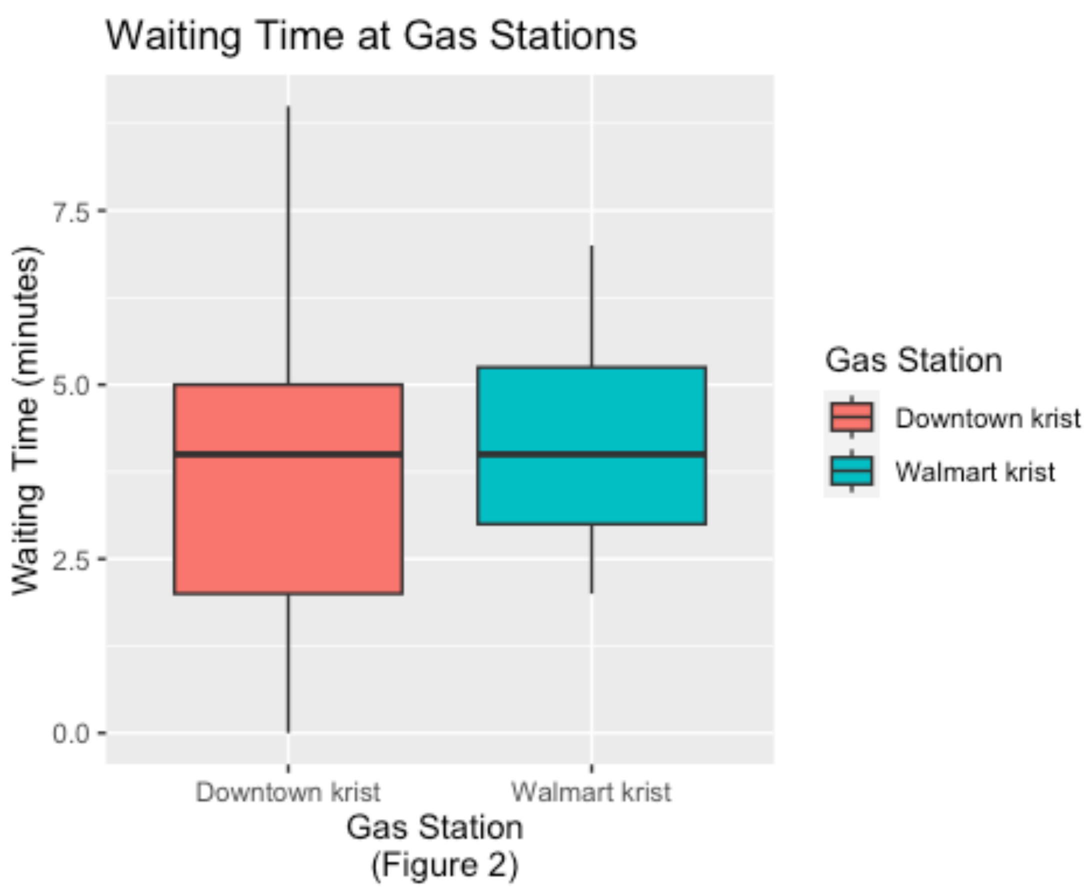
We are also removing the outliers in the Walmart krist

```
df <- data[data$Waiting_time <= 10, ]
head(df)

##      Gas_station No_pumps     Car_type          Date       Day In_time Out
_time
## 1 Walmart krist        8 pickup truck   Mar 15, 2024    Friday   16:06
16:13
## 2 Walmart krist        8        sedan   Mar 15, 2024    Friday   16:09
16:15
## 5 Walmart krist        8 pickup truck   Mar 15, 2024    Friday   16:11
```

```
16:16
## 6 Walmart krist        8        sedan   Apr 10, 2024 Wednesday   16:18
16:21
## 7 Walmart krist        8 pickup truck April 10, 2024 Wednesday   16:18
16:23
## 8 Walmart krist        8        sedan April 10, 2024 Wednesday   16:19
16:25
##   Waiting_time
## 1            7
## 2            6
## 5            5
## 6            3
## 7            5
## 8            6

ggplot(df, aes(x = Gas_station, y = Waiting_time, fill = Gas_station)) +
  geom_boxplot() +
  labs(title = "Waiting Time at Gas Stations",
       x = "Gas Station \n (Figure 2)",
       y = "Waiting Time (minutes)",
       fill = "Gas Station")
```



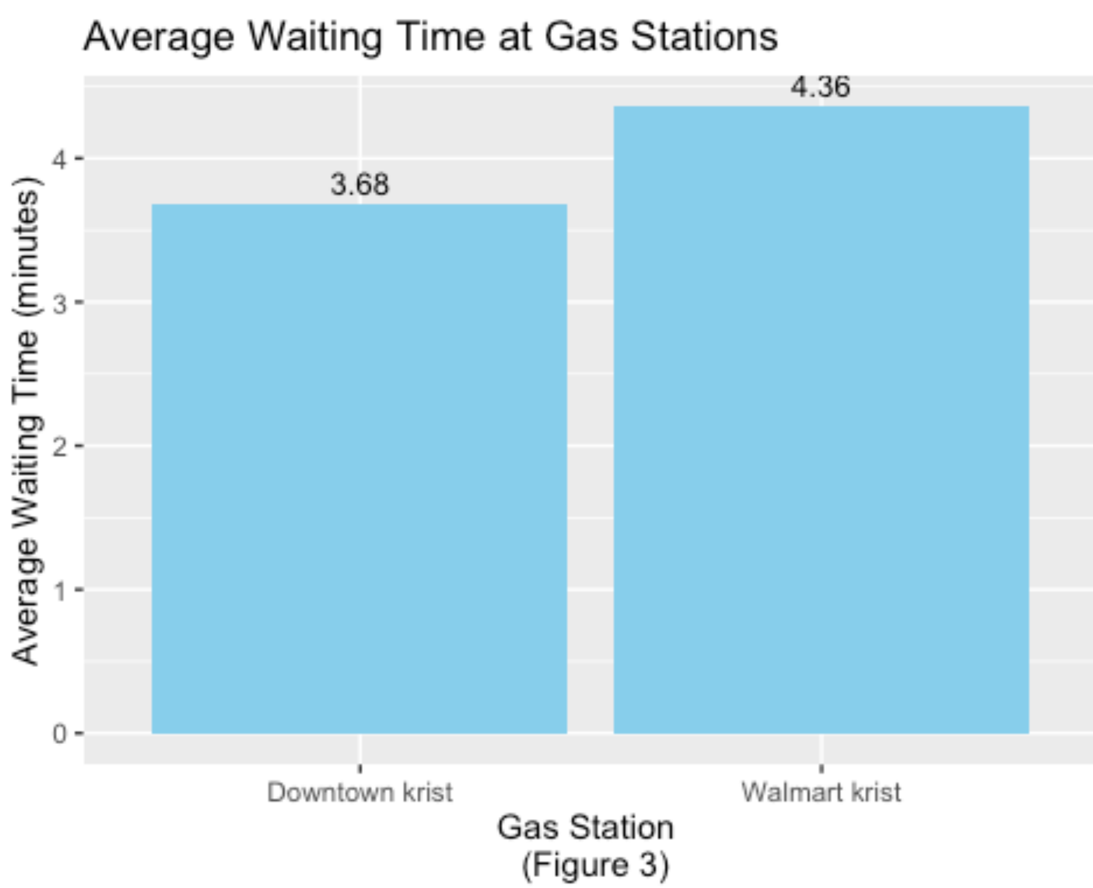
From the Figure 2 box plot, we can observe that both the box plots have almost same mean values.

Bar plot observe the distribution between the variables.

```
avg_waiting <- tapply(df$Waiting_time, df$Gas_station, mean)

avg_waiting_df <- data.frame(
  gas_station = names(avg_waiting),
  avg_waiting_time = avg_waiting
)

ggplot(avg_waiting_df, aes(x = gas_station, y = avg_waiting_time)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Average Waiting Time at Gas Stations",
       x = "Gas Station \n (Figure 3)",
       y = "Average Waiting Time (minutes)") +
  geom_text(aes(label = round(avg_waiting_time, 2)), vjust = -0.5, size = 3.5
)
```




Looking at Figure 3, we see that the average price at the Walmart gas station is higher than at the Downtown gas station, but not by much.

car types in the dataframe

```
unique(df$Car_type)

## [1] "pickup truck" "sedan"        "suv"
```

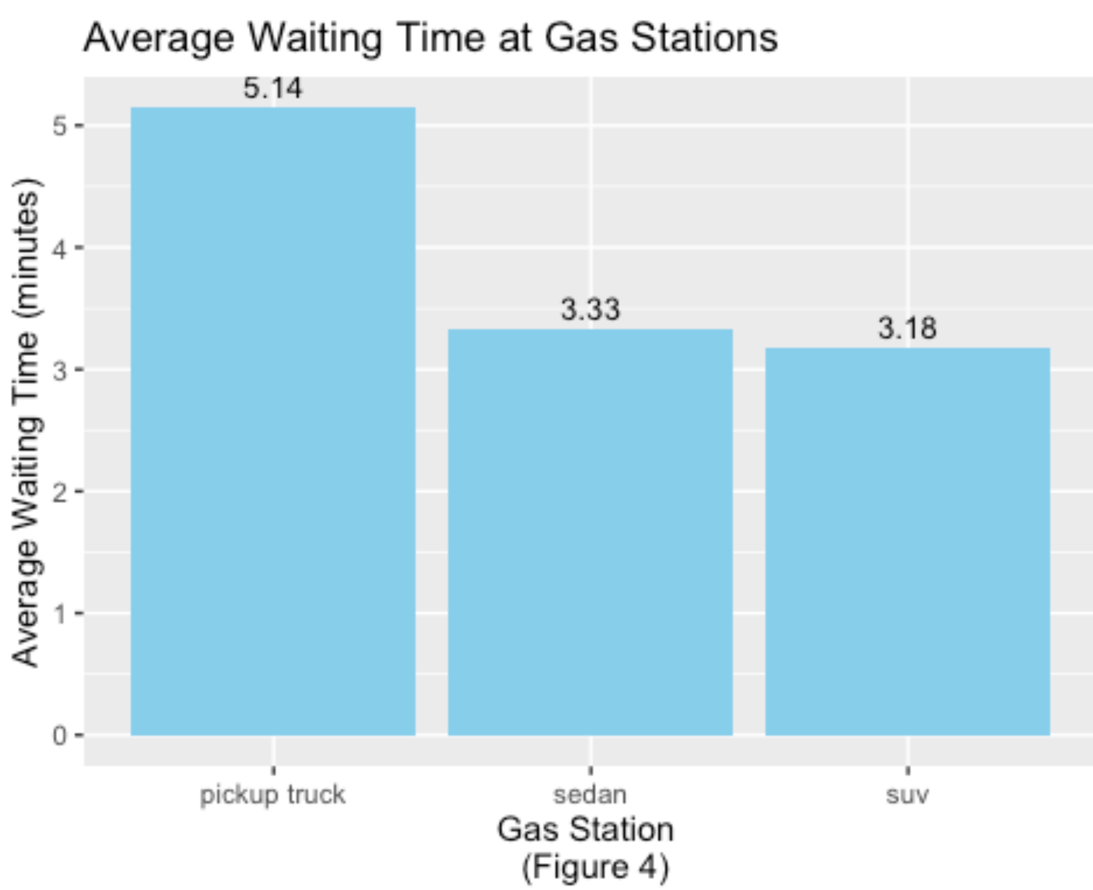
Based on the collected data we have got "pickup truck", "sedan", and "suv".

we are plotting Bar plot to observe the waiting time of different type of cars in the car type.

```
Car_avg_waiting <- tapply(df$Waiting_time, df$Car_type, mean)

Car_avg_waiting_df <- data.frame(
  car_type = names(Car_avg_waiting),
  avg_waiting_time = Car_avg_waiting)


ggplot(Car_avg_waiting_df, aes(x = car_type, y = avg_waiting_time)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Average Waiting Time at Gas Stations",
       x = "Gas Station \n (Figure 4)",
       y = "Average Waiting Time (minutes)") +
  geom_text(aes(label = round(avg_waiting_time, 2)), vjust = -0.5, size = 3.5
)
```



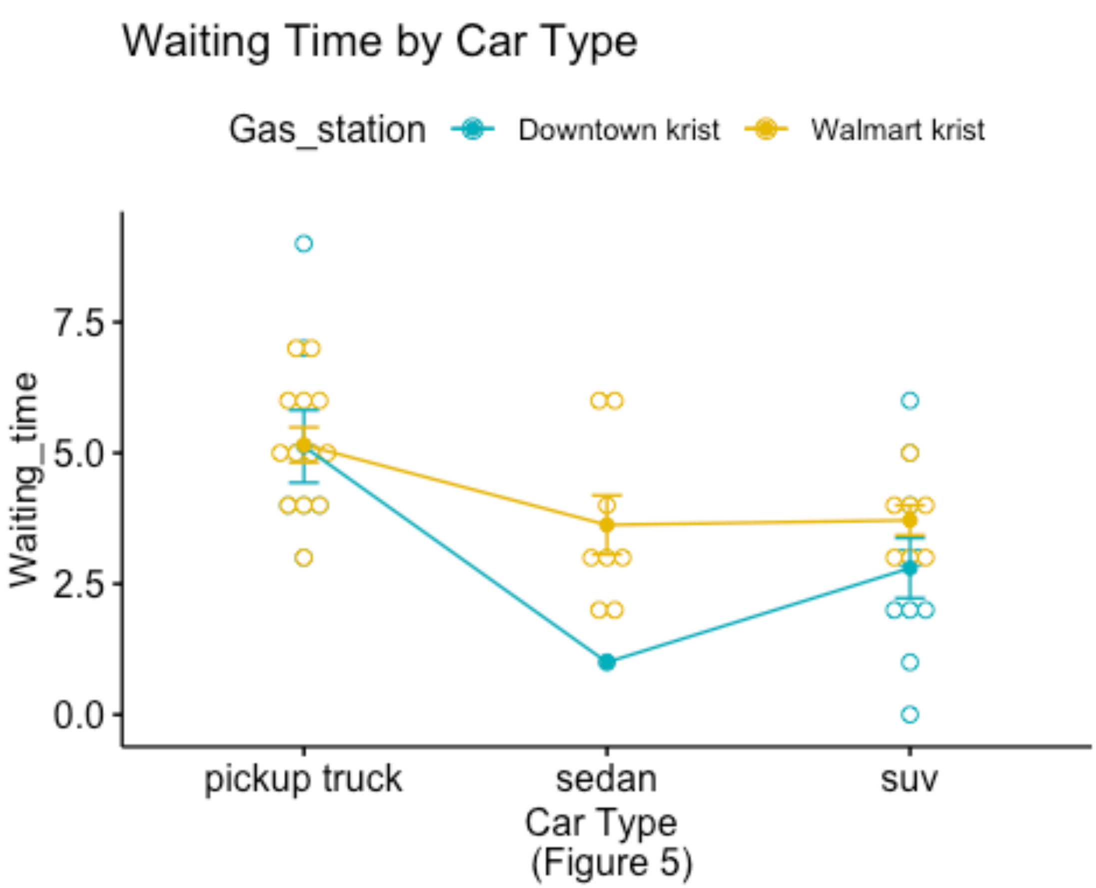
From the Figure 4, we can observe that the average waiting time for the Pickup truck is higher than the Sedan and SUV.

Interaction plot

```
library("ggpubr")
ggline(df, x = "Car_type", y = "Waiting_time", color = "Gas_station",
       add = c("mean_se", "dotplot"),
       palette = c("#00AFBB", "#E7B800"),
       main = "Waiting Time by Car Type",
       xlab = "Car Type \n (Figure 5)")

## Warning in stats::qt(ci/2 + 0.5, data_sum$length - 1): NaNs produced

## Bin width defaults to 1/30 of the range of the data. Pick better value wit
h
## `binwidth`.
```




In the above interaction plot (Figure 5) we can observe that the waiting time of pickup truck is almost same around 5 minutes in the both the gas stations. Where as the average waiting time of sedan in Downtown krist is very less, and from above plot we can clearly say that there is no interaction. Coming to SUV type the mean waiting for both the gas station there is a slight variation.

## Compute two-way ANOVA test

The R function aov() can be used to answer this question. The function summary.aov() is used to summarize the analysis of variance model.

```
anova_result <- aov(Waiting_time ~ Gas_station + Car_type, data = df)
summary(anova_result)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## Gas_station  1   5.13   5.126   2.179 0.147154    
## Car_type     2  41.40  20.701   8.802 0.000625 ***
## Residuals   43 101.13   2.352                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the ANOVA table we can conclude that Car_type is statistically significant. These results would lead us to believe that Car_type has a significant effect on Waiting_time of a Gas_station.

### Two-way ANOVA with interaction effect

```
# anova_result_interaction <- aov(Waiting_time ~ Gas_station * Car_type, data
= df)
anova_result_interaction <- aov(Waiting_time ~ Gas_station + Car_type + Gas_s
tation:Car_type, data = df)
summary(anova_result_interaction)
```

```
##                      Df Sum Sq Mean Sq F value   Pr(>F)    
## Gas_station           1   5.13   5.126   2.201 0.145546    
## Car_type              2  41.40  20.701   8.890 0.000621 ***
## Gas_station:Car_type  2   5.66   2.830   1.215 0.307057    
## Residuals            41  95.47   2.329                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value for the interaction (Gas_station:Car_type) is 0.307057 which is greater than significant level (0.05). We can conclude there is no significant interaction effect between gas station and car type on waiting times. Hence we failed to reject Null Hypothesis.

### Interaction Effect:

This is the combined effect of the two independent variables on the dependent variable. It tells you whether the effect of one independent variable depends on the level of the other independent variable. In other words, it examines whether the effect of one factor varies across different levels of the other factor.

## Pair Wise Analysus

### Multiple pairwise-comparison between the means of groups

In an ANOVA test, a low p-value means that there's a difference in averages between some groups, but we're not sure which ones. To find out, we can compare pairs of groups to see if their average differences are actually meaningful.

### Tukey multiple pairwise-comparisons

The ANOVA test shows significance, we can use Tukey HSD (Tukey Honest Significant Differences, which is done with the R function: TukeyHSD()) to compare the means of different groups in multiple pairwise comparisons. To do this, we pass the fitted ANOVA as an argument to the TukeyHSD() function.

```
TukeyHSD(anova_result_interaction, which = "Gas_station:Car_type")

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Waiting_time ~ Gas_station + Car_type + Gas_station:Car
_type, data = df)
##
## $`Gas_station:Car_type`
##                                                             diff       l
wr
## Walmart krist:pickup truck-Downtown krist:pickup truck  0.02884615 -2.0204
82
## Downtown krist:sedan-Downtown krist:pickup truck       -4.12500000 -8.9622
07
## Walmart krist:sedan-Downtown krist:pickup truck        -1.50000000 -3.7802
81
## Downtown krist:suv-Downtown krist:pickup truck         -2.32500000 -4.4882
65
## Walmart krist:suv-Downtown krist:pickup truck          -1.41071429 -3.7710
30
## Downtown krist:sedan-Walmart krist:pickup truck        -4.15384615 -8.8865
66
## Walmart krist:sedan-Walmart krist:pickup truck         -1.52884615 -3.5781
74
## Downtown krist:suv-Walmart krist:pickup truck          -2.35384615 -4.2721
20
## Walmart krist:suv-Walmart krist:pickup truck           -1.43956044 -3.5775
85
## Walmart krist:sedan-Downtown krist:sedan                2.62500000 -2.2122
07
## Downtown krist:suv-Downtown krist:sedan                 1.80000000 -2.9831
59
## Walmart krist:suv-Downtown krist:sedan                  2.71428571 -2.1611
61
```

```
## Downtown krist:suv-Walmart krist:sedan                    -0.82500000 -2.9882
65
## Walmart krist:suv-Walmart krist:sedan                      0.08928571 -2.2710
30
## Walmart krist:suv-Downtown krist:suv                       0.91428571 -1.3331
84
##                                                                 upr     p ad
j
## Walmart krist:pickup truck-Downtown krist:pickup truck  2.0781739 1.000000
0
## Downtown krist:sedan-Downtown krist:pickup truck        0.7122073 0.133891
9
## Walmart krist:sedan-Downtown krist:pickup truck         0.7802814 0.378779
0
## Downtown krist:suv-Downtown krist:pickup truck         -0.1617351 0.028761
8
## Walmart krist:suv-Downtown krist:pickup truck           0.9496012 0.485481
4
## Downtown krist:sedan-Walmart krist:pickup truck         0.5788735 0.114558
6
## Walmart krist:sedan-Walmart krist:pickup truck          0.5204816 0.246750
0
## Downtown krist:suv-Walmart krist:pickup truck          -0.4355720 0.008533
8
## Walmart krist:suv-Walmart krist:pickup truck            0.6984644 0.353206
9
## Walmart krist:sedan-Downtown krist:sedan                7.4622073 0.589328
3
## Downtown krist:suv-Downtown krist:sedan                 6.5831586 0.868284
1
## Walmart krist:suv-Downtown krist:sedan                  7.5897324 0.562582
7
## Downtown krist:suv-Walmart krist:sedan                  1.3382649 0.861800
9
## Walmart krist:suv-Walmart krist:sedan                   2.4496012 0.999997
2
## Walmart krist:suv-Downtown krist:suv                    3.1617556 0.826655
6
```

By observing the above Tukey multiple comparisons of means we can clearly observe that there is no significant difference between the mean values of the groups because all groups p-values are grater than significant level (0.05), which states that there is no statistical significance between the groups. Hence we clearly say that all pairs comes under a single group.

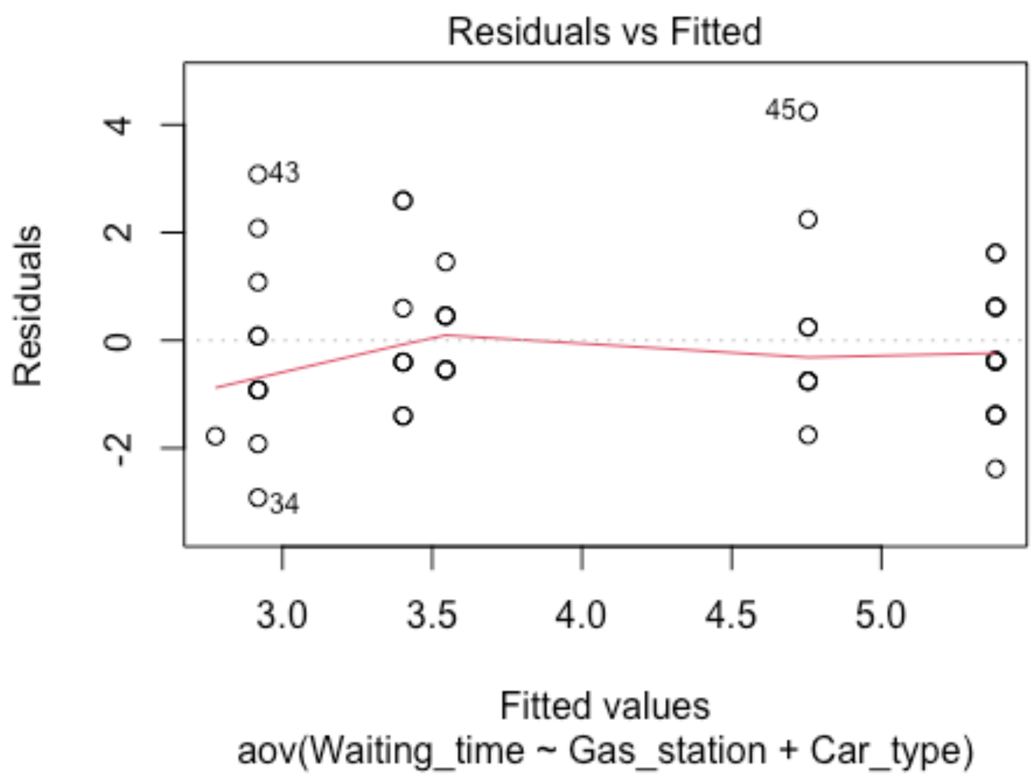
### Check ANOVA assumptions: test validity?

ANOVA assumes that the data are normally distributed and the variance across groups are homogeneous. We can check that with some diagnostic plots.

### Check the homogeneity of variance assumption

Residuals vs Fitted

```r
plot(anova_result, 1)
```



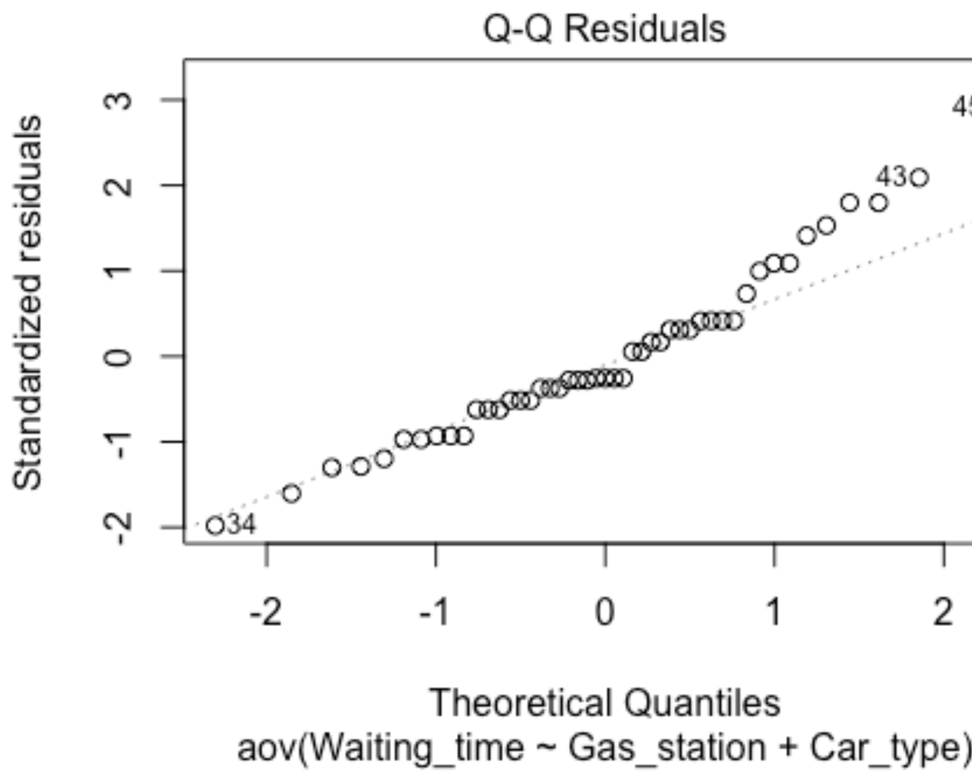

(Figure 6)

In Figure 6 (Residuals vs Fitted), R identified points 43, 34, and 45 as outliers. Apart from those, we observe a straight line with data points scattered randomly around it, indicating that the assumption of equal variance (homoscedasticity) is met here.

### Check the normality assumpttion

Q-Q Residuals:

```r
plot(anova_result, 2)
```




(Figure 7)

In Figure 7 (Q-Q Residuals), R automatically marked the three data points with significant residuals (observations 34, 43, and 45). Besides these points, the observations mostly align closely with the 45-degree line in the QQ-plot, suggesting that normality is likely maintained.

## CONCLUSION

In our analysis of gas station waiting times, we conducted a thorough observation into the influences of both location and vehicle type on customer waiting time. Through statistical methodologies like two-way ANOVA tests and pairwise comparisons, we uncovered significant trends. Notably, we found that the type of vehicle had a significant impact, with pickup trucks often facing longer waits compared to sedans and SUVs. This discovery emphasizes the importance of tailoring management strategies to efficiently accommodate larger vehicles, particularly during peak periods of demand.

While our analysis revealed variations in average waiting times across different gas station locations, we did not identify any significant interaction effects between station location and vehicle type on waiting times. This suggests a consistent influence of vehicle type on waiting duration irrespective of station location. However, it also underscores the need for further exploration into other potential factors affecting waiting times, such as staffing levels or operational procedures. Our study's observational approach provided valuable real-world insights from gas stations in Houghton, but future research could benefit from investigating additional variables and expanding the study's geographic reach for more comprehensive insights applicable to gas station management practices in diverse regions.

## References


http://www.sthda.com/english/wiki/two-way-anova-test-in-r#what-is-two-way-anova-test

https://www.investopedia.com/terms/a/anova.asp

https://ademos.people.uic.edu/Chapter12.html

https://www.scribbr.com/statistics/anova-in-r/