Business Analyst Project

By Samuel Wong

Data Command

Libraries

- library(tidyverse)
- library(car)
- library(ggplot2)
- library(gridExtra)

Data Cleaning

- Donna <- read.csv("DonnaGasStation.csv", header = TRUE, sep = ",", quote = '"')
- names(clean_Donna)[1]<- "week_num"
- names(clean_Donna)[2]<- "Weekday"
- names(clean_Donna)[3]<- "D_Price"
- names(clean_Donna)[4]<- "competitor_price"
- names(clean_Donna)[5]<- "D_Gal_sold"
- > colnames(clean_Donna)
  "week_num" "weekday" "D's_Sales" "competitor's_price" "D's_Sales"
- New_clean_Donna <- na.omit(clean_Donna)
- sapply(New_clean_Donna[c('D_Price', 'D_Gal_sold', 'competitor_price')], fivenum)

Data Plots and Graphs

- **#Histogram**

1. ggplot(data = New_clean_Donna) + geom_histogram(mapping=aes(x=competitor_price, fill = weekday), bins = 10)
2. ggplot(data = New_clean_Donna) + geom_histogram(mapping=aes(x=D_Price, fill = weekday), bins = 15)

- **#Boxplot**

1. boxplot(New_clean_Donna$D_Price)
2. boxplot(New_clean_Donna$competitoe_price)

- **#Facet graph - Scatterplot**

1. ggplot(data=New_clean_Donna)+geom_point(mapping=aes(x = D_Price, y = D_Gal_sold,color=weekday))+facet_wrap(~weekday)
2. ggplot(data = New_clean_Donna, mapping = aes(x = D_Price, y = D_Gal_sold)) + geom_point(mapping = aes(color = weekday))+ geom_smooth()
3. ggplot(data = New_clean_Donna, mapping = aes(x = competitor_price, y = D_Gal_sold)) + geom_point(mapping = aes(color = weekday))+ geom_smooth()

Data Regressions

- plot_prices <- ggplot(final_Donna, aes(x = competitor_price, y = D_Price)) + geom_point()
- plot_prices_r <- plot_prices + geom_smooth(method = 'lm')

#Independent variable are pace in hold to predict the dependent variable#

- pred_frame_1 <- data.frame(competitor_price c(= 3.56, 3.66))
- pred_frame_1 <- data.frame(final_Donna$competitor_price, c(3.56, 3.66))

Format: lm(dependent_varible ~ predictor_1, data = df)

- model_1 <- lm(D_Price ~ competitor_price, final_Donna)
- summary(model_1)

 # predict Donna's price by using Competitor's price

- pred_frame_1 <- data.frame(competitor_price = c(3.56, 3.66))
- predict(model_1, newdata = pred_frame_1)

####### MULTIPLE REGRESSION #######

- colnames(final_Donna)

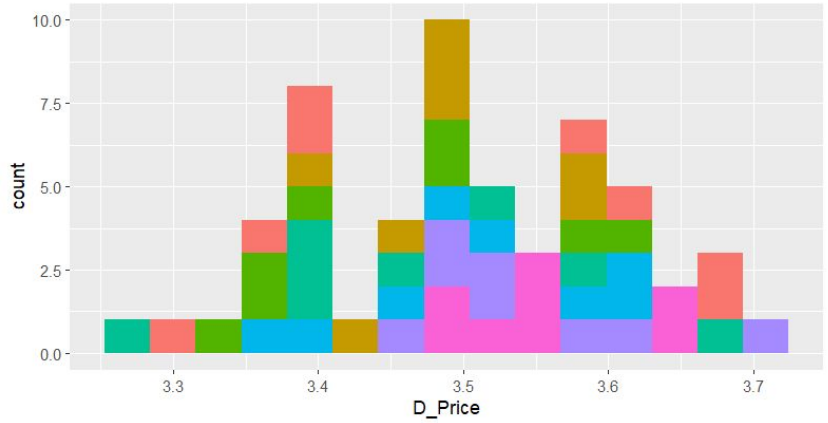
 # predict Donna's price by using Competitor's price and D's gallons sold

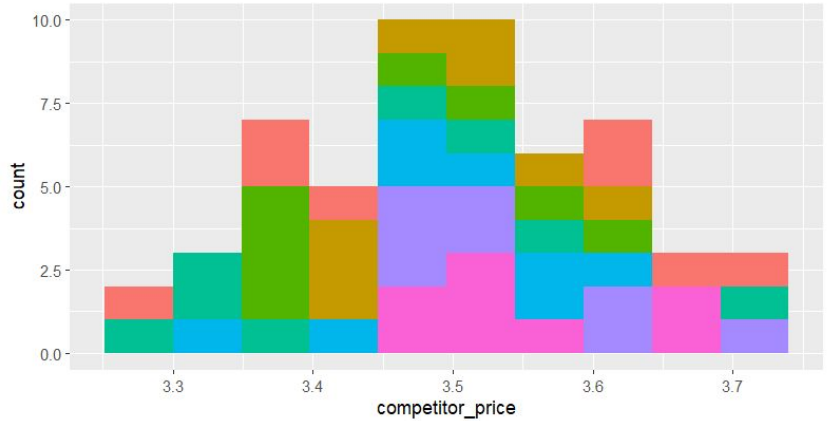
- model_2 <- lm(D_Price ~ competitor_price + D_Gal_sold, data = final_Donna)
- summary(model_2)

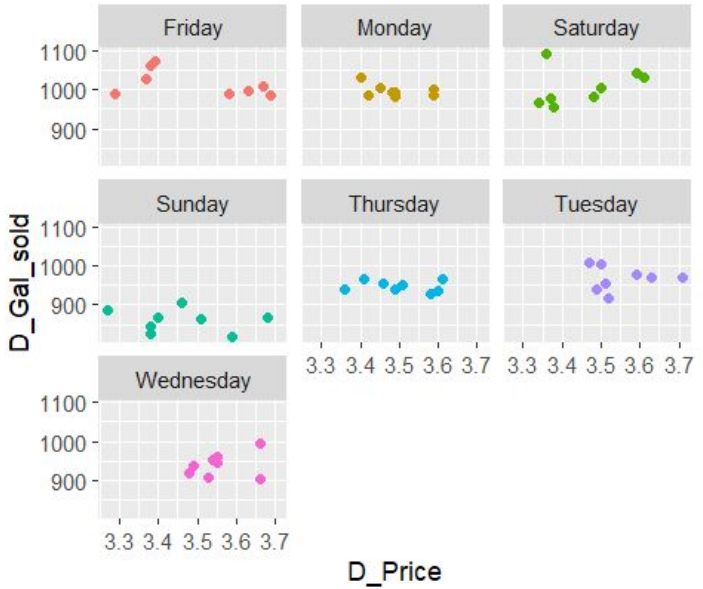
 # predict

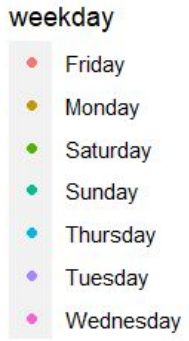
- pred_frame_2 <- data.frame(competitor_price = c(3.55, 3.66, 3.71), D_Gal_sold = c(968.5, 968.5, 968.5))
- predict(model_2, newdata = pred_frame_2)

Visual Assets - Histogram, Facet Graph



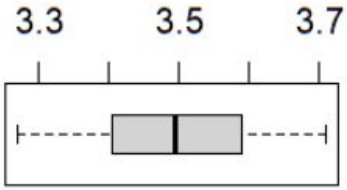




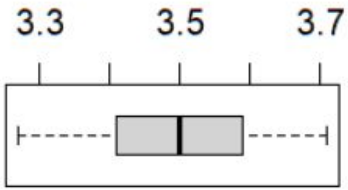


Visual Assets – Scatterplot, Boxplot

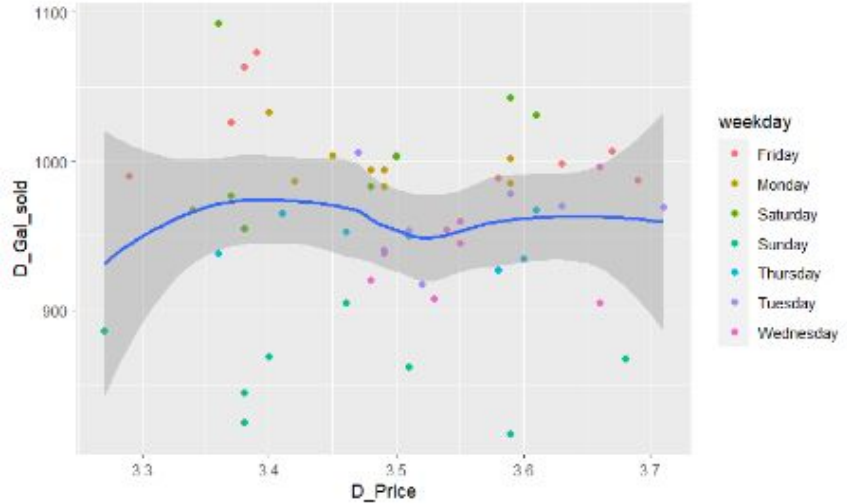
Boxplot of Donna's price



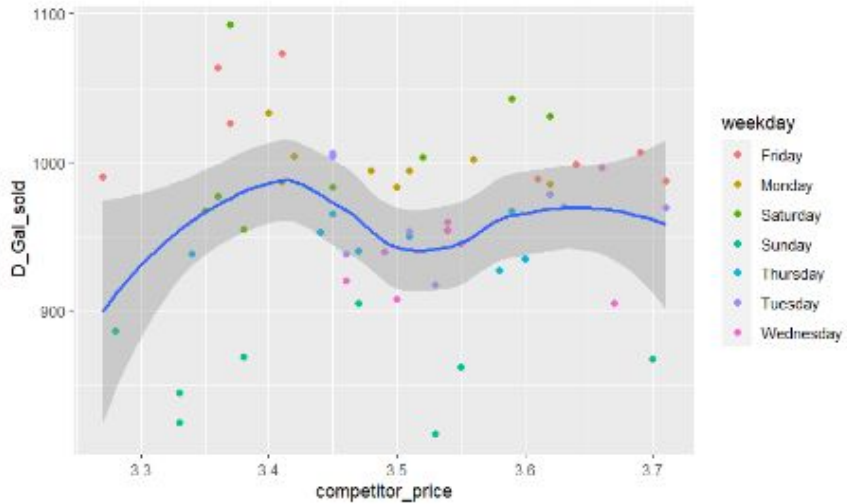
Boxplot of Competitor's price



Donna's Price correlating to Donna's Gallons Sold



Competitor's Price correlating to Donna's Gallons Sold

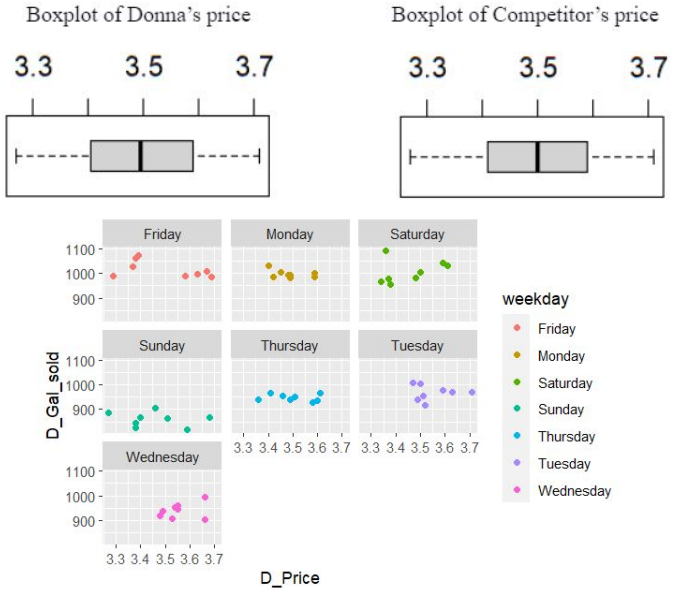


Descriptive Analysis - Explain the Current-State

- **Any similarities? If so, why?**
  - The boxplots and the scatterplots are nearly identical.
- **Any outliers? If so, why?**
  - The facet graph, an outlier exists on all Sundays of the eight weeks period which has the least amount of gallons sold out of all the other days between the range of 800 to 900 gallons.
- **Any patterns? If so, why?**
  - In the scatterplots that offer trend lines the line is shaped like an letter M, as both Donna and her competitor's prices rise up ten dip down between the range of $3.40 to $3.50. However, as prices rise Donna's total gallons sold rise too, albeit never exceeding the maximums sold from when the price was below.

- sapply(New_clean_Donna[c('D_Price', 'D_Gal_sold', 'competitor_price')], fivenum)

	D_Price	D_Gal_sold	competitor_price
[1,]	3.270	817.03	3.27
[2,]	3.405	936.14	3.41
[3,]	3.495	968.45	3.50
[4,]	3.590	997.52	3.59
[5,]	3.710	1092.33	3.71



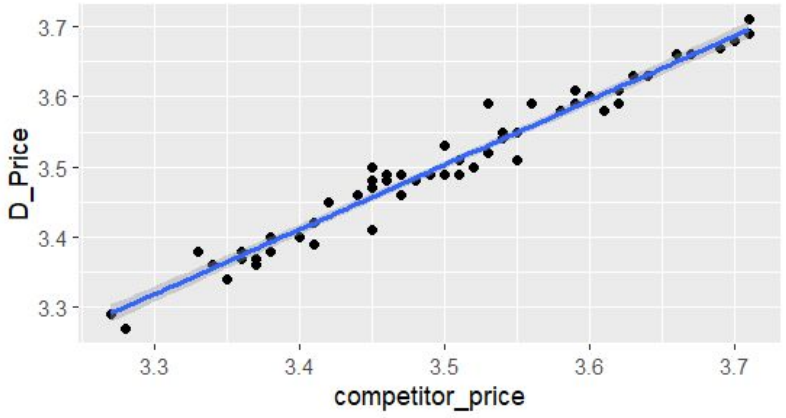
Regression Analysis

1st model prediction

- The model_1, this predict Donna's Price with the competitor's price.
  - Independent variable is her competitor's prices
  - Dependent variable is Donna's prices
- From this we know there is a one cent difference for almost every price, however we can't be entirely sure until checking with the
- **Results**: The competitor is setting price by one cent below Donna's price

2nd model prediction

- The model_2, this predict Donna's Price with the competitor's price and the total gallons Donna sold.
  - 1st independent variable = competitor's prices
  - 2nd independent variable = Donna's gallons sold
  - Dependent variable is Donna's prices
- The same result occur, only the 3rd variable predict a 2 cent difference.
- **Results**:The competitor is setting price by one to two cent below Donna's price



Data Regressions

#Independent variable are pace in hold to predict the dependent variable#

- pred_frame_1 <- data.frame(competitor_price c(= 3.56, 3.66))
- pred_frame_1 <- data.frame(final_Donna$competitor_price, c(3.56, 3.66))

Format: lm(dependent_varible ~ predictor_1, data = df)

- model_1 <- lm(D_Price ~ competitor_price, final_Donna)
- summary(model_1)

 # predict Donna's price by using Competitor's price

- pred_frame_1 <- data.frame(competitor_price = c(3.56, 3.66))
- predict(model_1, newdata = pred_frame_1)

1	2
3.558429	3.650348

####### MULTIPLE REGRESSION #######

- colnames(final_Donna)

 # predict Donna's price by using Competitor's price and D's gallons sold

- model_2 <- lm(D_Price ~ competitor_price + D_Gal_sold, data = final_Donna)
- summary(model_2)

 # predict

- pred_frame_2 <- data.frame(competitor_price = c(3.55, 3.66, 3.71), D_Gal_sold = c(968.5, 968.5, 968.5))
- predict(model_2, newdata = pred_frame_2)

1	2	3
3.548675	3.649978	3.696025

Discussion of business risk

What's the Risk

- **Recap**: Donna's business problem is her competitor is setting their prices below hers.
  - The competitor has the advantage in the long-run, if nothing is done.
  - Donna will lose more profits, and if mismanage, her business to her competitors.
- This is strategic risk on the long-run where Donna must make long-term plans to succeed in protecting her gas station and keep most of the profits in her local area.

Business Solutions and Strategies

- **Short-term**: The product, gasoline, the same at all gas station and the demand for it is mostly inelastic so Donna must appeal to the customers or risk long-term sustainability.
  - **Price Match Policy, prioritize the current customer (*to improve relationships between producer and her customers*), and include small discounts in-store purchases with gas purchase.**
- **Long-term**: To slowly but surely improve her business model as more data comes in.