

Syed Zahoor Raza, Jayden Hewitt, Andre Bianchi and Eric Smith

DANL 200-02: Beer Market Data Analysis

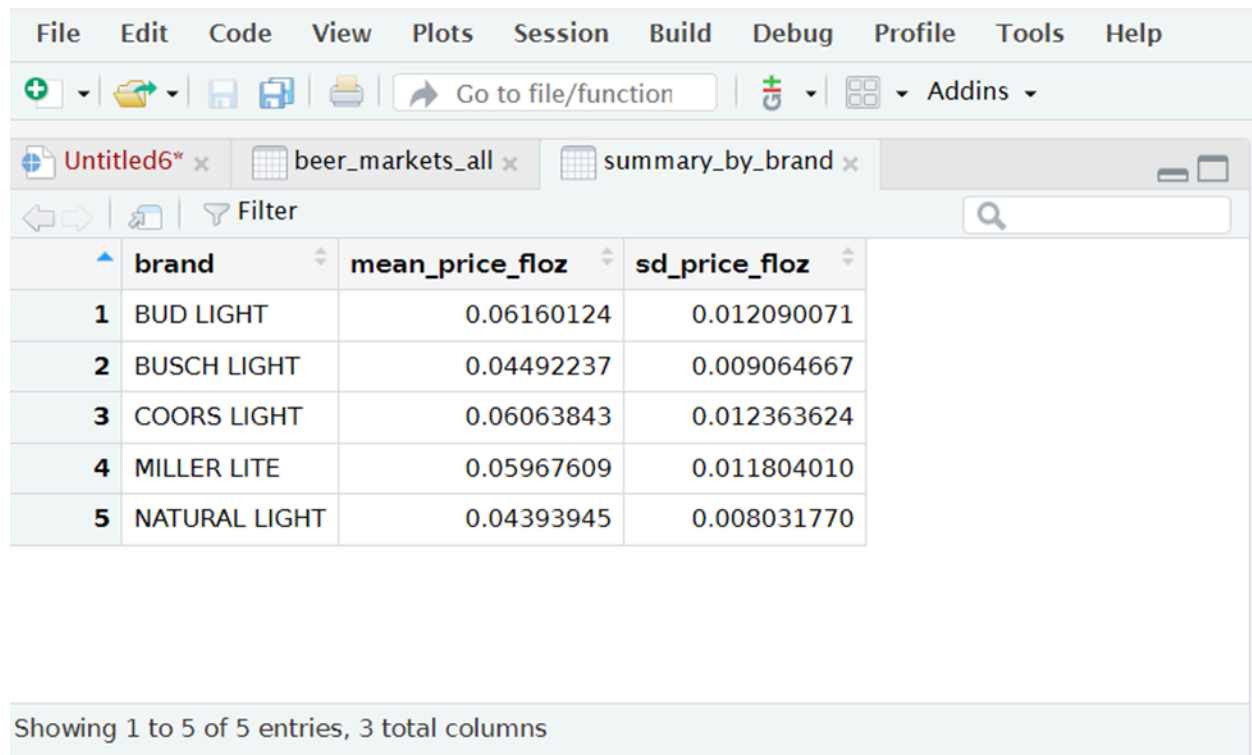
How Do the Prices Per Fluid Ounce Vary from One Brand to Another?

Analyzing price per fluid ounce across brands is crucial for both consumers and businesses within the beer industry. Customers are always looking to get the most out of their purchases, and comparing the cost per fluid ounce of various beer brands helps them identify which brands are the best deals on beer. This is crucial for consumers on a tight budget who might value quantity above brand awareness. Thus, they can make informed purchasing decisions using this information. Moreover, the analysis can also help with market segmentation. By understanding how price impacts various consumer groups, businesses can tailor marketing and promotions. For instance, to draw customers, a retailer may decide to price some goods cheaply or to give premium brands larger margins. If data indicates that a particular beer brand has a higher price point than rivals, a grocery store chain may choose to launch a promotion on it to draw in more consumers or boost sales volume. In public policy, regulatory authorities may also find this data useful to evaluate possible price gouging or to ensure fair competition.

Below, I use the R code to calculate the average price per fluid ounce for each brand of beer and standard deviation to see how much the price per fluid ounce varies within each brand. The R code to calculate these is as follows:

```
summary_by_brand <- beer_data %>%  
group_by(brand) %>%  
summarize(mean_price_floz = mean(price_per_floz),
```

```
sd_price_floz = sd(price_per_floz))
```



	brand	mean_price_floz	sd_price_floz
1	BUD LIGHT	0.06160124	0.012090071
2	BUSCH LIGHT	0.04492237	0.009064667
3	COORS LIGHT	0.06063843	0.012363624
4	MILLER LITE	0.05967609	0.011804010
5	NATURAL LIGHT	0.04393945	0.008031770

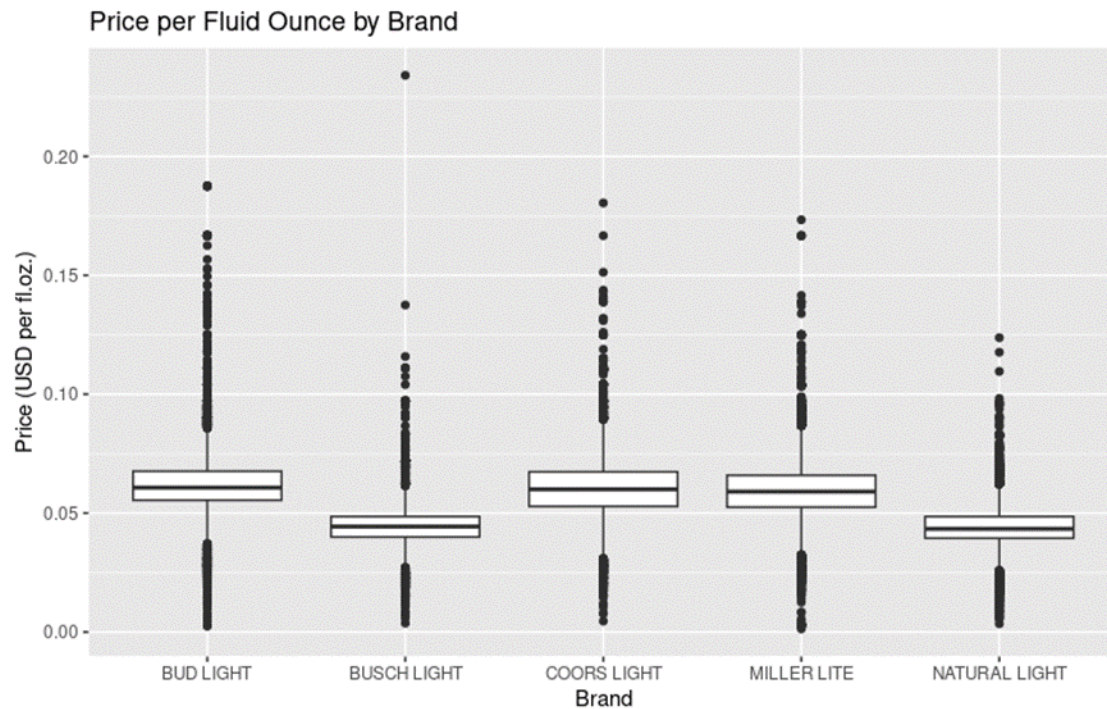
Showing 1 to 5 of 5 entries, 3 total columns

The box plot shows the distribution of price per fluid ounce for each brand of beer. This allows us to see the median (middle value of a dataset), upper (75th percentile), lower quartiles (25th percentile), minimum and maximum, and outliers for each brand (data points that significantly differ from the rest of the data). In addition to this, this boxplot also helps us to see if there are any outliers (data points that significantly differ from the rest of the data).

The R code to create the boxplot is as follows:

```
ggplot(beer_data, aes(x = brand, y = price_per_floz)) +  
  geom_boxplot() +  
  labs(title = "Price per Fluid Ounce by Brand",  
        x = "Brand",  
        y = "Price (USD per fl.oz.)")
```

For data transformation, the outliers can be removed from the dataset as extreme values can skew the results and impact measures like mean and standard deviation.



How does Total Sales vary across different brands?

In this section, we will calculate the total amount spent on each brand and visualize the data using a bar plot. Additionally, we will provide summary statistics such as mean and standard deviation of the total amount spent per brand.

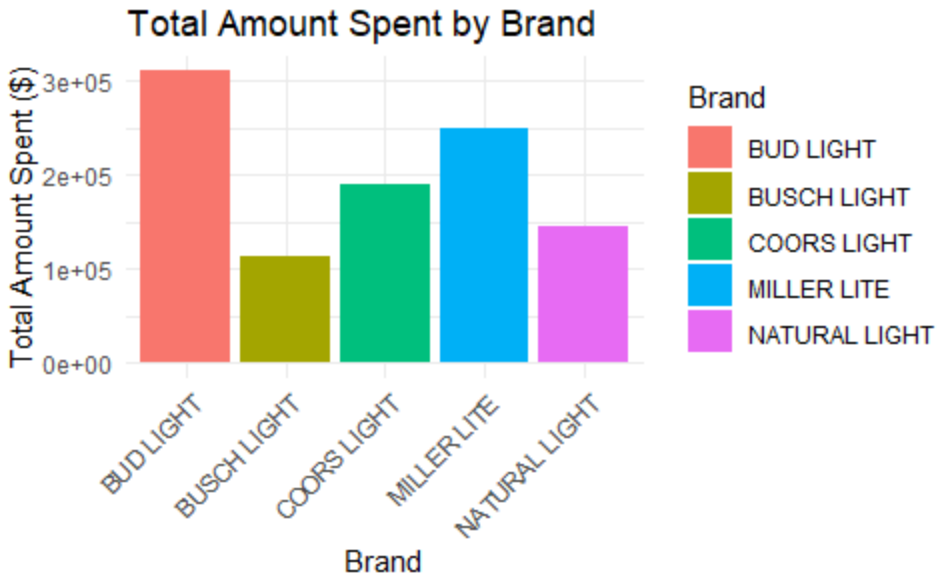
Calculate the total amount spent per brand:

```
summary_by_brand <- beer_markets %>%  
  group_by(brand) %>%  
  summarize(total_spent = sum(dollar_spent, na.rm = TRUE),  
            mean_spent = mean(dollar_spent, na.rm = TRUE),  
            sd_spent = sd(dollar_spent, na.rm = TRUE))  
  
print(summary_by_brand)
```

brand	total_spent	mean_spent	sd_spent
<chr>	<dbl>	<dbl>	<dbl>
BUD LIGHT	309653.2	14.34111	9.025775
BUSCH LIGHT	113953.4	13.13735	8.127524
COORS LIGHT	189540.5	14.49752	8.348152
MILLER LITE	248962.6	14.50915	9.038281
NATURAL LIGHT	145245.7	11.51282	8.099684

The bar plot will show the total amount spent on each brand.

```
total_spent_by_brand <- beer_markets %>%
  group_by(brand) %>%
  summarize(total_spent = sum(dollar_spent, na.rm = TRUE))
ggplot(total_spent_by_brand, aes(x = brand, y = total_spent, fill =
brand)) +
  geom_bar(stat = "identity") +
  labs(title = "Total Amount Spent by Brand",
    x = "Brand",
    y = "Total Amount Spent ($)",
    fill = "Brand") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



We can see by analyzing this data, that price is not a major driving factor for beer consumers. We know this because even though Busch light and Natural light have the cheapest beer in terms of price per fluid ounce, they come up short being the two lowest brands in terms of total sales.

What is the favorite beer of each state?

The following is the R code used to answer this question.

```
> top_brands <- beer_markets |>
> group_by(state, brand) |>
> summarise(total_sales = sum(dollar_spent)) |>
> top_n(1, total_sales) |>
> arrange(state)
```

region	brand	total_sales
<chr>	<chr>	<dbl>
1 alabama	MILLER LITE	5476.
2 arizona	BUD LIGHT	15742.
3 arkansas	BUSCH LIGHT	3449.
4 california	BUD LIGHT	27733.
5 colorado	BUD LIGHT	5162.

6 connecticut	BUD LIGHT	2957.
7 delaware	COORS LIGHT	972.
8 district of columbia	MILLER LITE	135.
9 florida	MILLER LITE	29102.
10 georgia	BUD LIGHT	6826.

#This is the first 10 rows of the data set I created to show the favorite brand by state.

```
> install.packages("maps")
> install.packages("mapdata")

> State <- map_data("state")

> top_brands$state <- tolower(top_brands$state)
> top_brands <- top_brands |>
>   rename(region = state)

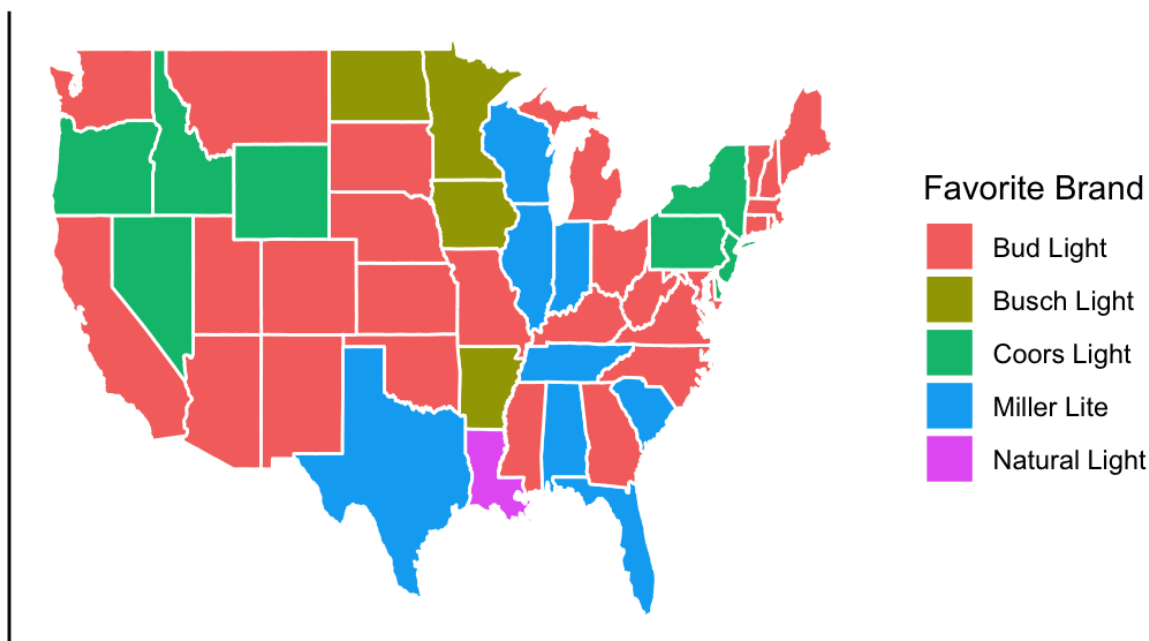
mappedbrand <- merge(top_brands, State, by = "region")
```

#I merged the beer data set to a data set which includes the latitude and longitude of every state in America, so that I could add the top brand data onto a map.

```
> mappedbrand$brand <- tools::toTitleCase(tolower(mappedbrand$brand))

> ggplot(data = mappedbrand, aes(x=long, y=lat, fill=brand, group=group)) +
>   geom_polygon(color = "white") +
>   scale_fill_discrete(name = "Favorite Brand") +
>   labs(title = "The Lower 48's Favorite Beer") +
>   theme_classic() +
>   theme(axis.title.x=element_blank(), axis.text.x=element_blank(),
>         axis.ticks.x=element_blank(), axis.title.y=element_blank(),
>         axis.text.y=element_blank(), axis.ticks.y=element_blank())
```

The Lower 48's Favorite Beer



This is the map of the United States. Each state's color corresponds to their favorite brand of beer.

How much does each can cost, per pack size?

The following is the R code used to answer this question.

```
> beer_markets$packsize <- as.numeric(ifelse(grepl("\\d", beer_markets$`_purchase_desc`),  
>                                     gsub("\\D","", beer_markets$`_purchase_desc`),  
>                                     1))
```

#First I converted the “_purchase_desc” variable into a numeric variable by taking the numeric value (which indicated the size of the pack bought) and removing the rest of the characters in each value. This created a new variable called “packsize”

```
> cans_only <- beer_markets |>  
> filter(container == "CAN")  
  
> pricebyquant <- cans_only |>  
> select(packsize, quantity, dollar_spent) |>  
> mutate(packprice = dollar_spent / quantity)  
  
> avg_pricing <- pricebyquant |>
```

```
> group_by(packsize) |>
> summarise(avg_price = mean(packprice)) |>
> arrange(packsize)
```

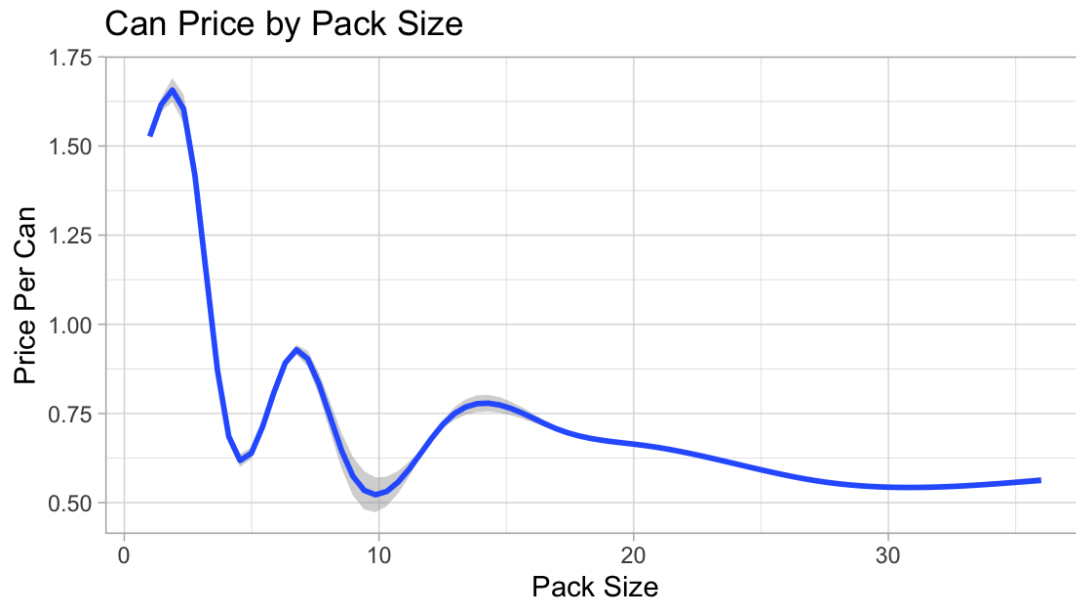
```
> avg_pricing <- avg_pricing |>
> mutate(can_price = avg_price / packsize)
```

	packsize	avg_price	can_price
	<dbl>	<dbl>	<dbl>
1	1	1.53	1.53
2	2	2.90	1.45
3	3	4.86	1.62
4	4	2.75	0.686
5	6	5.02	0.836
6	8	3.57	0.447
7	9	7.80	0.866
8	12	8.09	0.674
9	15	9.71	0.648
10	18	12.3	0.685
11	20	13.7	0.686
12	24	14.6	0.609
13	30	16.3	0.543
14	36	20.2	0.562

#This shows the average price of each sized pack or beers, and the average price per can for each pack size

```
> all_canprice <- pricebyquant |>
> mutate(cans = packprice / packsize)
```

```
> ggplot(all_canprice, aes(x = packsize, y = cans)) +
>   geom_smooth() +
>   labs(title = "Can Price by Pack Size",
>         y = "Price Per Can",
>         x = "Pack Size") +
>   theme_light()
```

This is a visualization of the price per can, as the packsize goes up. Generally the more beers you buy, the cheaper each can tends to be.

Favorite Beer Brands by Buyer Type

What are Men's Favorite Beer Brands?

Below is the R code used to answer what brand of beer men bought the most.

```
df_sum <- beer_markets |>
  filter(buyertype == 'male') |>
  count(brand)
```

Below is a chart showing the results:

1	BUD LIGHT	3945
---	-----------	------

2	BUSCH LIGHT	1807
3	COORS LIGHT	2168
4	MILLER LITE	2797
5	NATURAL LIGHT	2507

Bud Light is the most popular, Miller Lite second, Natural Light third, Coors Light fourth, then Busch Light is last for men.

What are Married Couples' Favorite Beer Brands?

Below is the R code used to answer what brand of beer married couples bought the most.

```
df_married <- beer_markets |>
  filter(buyertype == 'married') |>
  count(brand)
```

Below is a chart showing the results:

1	BUD LIGHT	14706
2	BUSCH LIGHT	6003
3	COORS LIGHT	9575

4	MILLER LITE	12169
5	NATURAL LIGHT	7846

Bud Light is the most popular, Miller Lite second, Coors Light third, Natural Light fourth, then Busch Light is last for married couples.

What are Female's Favorite Beer Brands?

Below is the R code used to answer what brand of beer females bought the most.

```
df_sum <- beer_markets |>
  filter(buyertype == 'female') |>
  count(brand)
```

Below is a chart showing the results:

1	BUD LIGHT	2941
2	BUSCH LIGHT	864
3	COORS LIGHT	1331
4	MILLER LITE	2193
5	NATURAL LIGHT	2263

Bud Light is the most popular, Natural Lite second, Miller Light third, Coors Light fourth, then Busch Light is last for females.

Using this information, we deduce that Bud Light is the favorite beer brand for men, women, and married couples.

Works Cited

1. <https://umu.diva-portal.org/smash/get/diva2:1833378/FULLTEXT03.pdf>
2. <https://www.geeksforgeeks.org/boxplots-in-r-language/>
3. <https://www.statology.org/summary-statistics-in-r-dplyr/>