Logo, company name

Description automatically generated

**Analysis of Factors Influencing**

**Sapporo Beer Sales**

December 19th, 2022

Text

Description automatically generated

**Executive Summary:**

In this project, we investigate the factors that influence Sapporo’s beer sales in the US market. We analyzed various data such as U.S. Census Bureau’s demographic and socio-economic data, restaurant and stores’ Google ratings and Business patterns by zip codes  to understand the relationship between these and Sapporo’s sales. We find evidence that diversity in areas, greater density of retail & food stores & localities with higher median income all drive sales for Sapporo beer. Moreover, our results also show that the ratings of the stores at which sapporo beer is sold does not have a positive relation with sales of the beer.

**Business Problem:**

Sapporo Beer is a Japanese brand that has been in the US market since 1964. Since then, Sapporo has grown as a brand throughout the United States. With this growth begets the obvious question of what factors are driving this growth or sales of Sapporo beer. One such set of factors are the demographic and socioeconomic characteristics of areas in which Sapporo beer is sold. Sapporo wants to understand factors that influence their beer sales and identify new avenues in which they can focus their marketing efforts.

**Empirical Framework:**

*Hypotheses*

* Sapporo consumption varies across NY based on local socio-demographic factors.
* Higher Google ratings may influence a positive trend in beer sales.
* Areas with higher income individuals - have a higher consumption of beer​.​

*Statistical Methods*

We use regression analysis to investigate the relationship between Sapporo consumption and a series of socio-demographic, economic, and restaurant-related factors in a zip code. This allows us to test our hypotheses regarding the strength of impact of various potential drivers of Sapporo beer consumption. We rely on a random-effects model, a standard approach for estimating statistical relationships with data from multiple zip codes over time. We also use classic ordinary least square models and estimate models by year for additional support. We provide more information in the Appendix regarding our approach.

*Data*

For this project, data was collected from Sapporo Beer's CSV files and external sources such as the US Census Bureau's ACS website to investigate the relationship between demographic and socio-economic factors and Sapporo's beer sales in the US market. The data was aggregated at the ZCTA, county, and tract levels and transformed into the zip code level using crosswalk files, which allowed for a simpler merge with the response variable of Sapporo's sales data. The number of stores that sell beer in an area, such as convenience stores, restaurants, and liquor stores, was also analyzed in relation to Sapporo's sales. The collected data was comprehensive and allowed for an investigation of the factors influencing Sapporo's beer sales and the creation of recommendations. Google ratings of restaurants and stores were also analyzed in the study.

*Results*

In Table 1, we observe that a 1% increase in median income results in a statistically-significant 1.08% increase in Sapporo sales. Other results indicate that a 1% increase in the number of food and retail stores results in a 1.39% increase in beer sales. Overall, our results showed that variables indicating diversification have an overall positive impact on Sapporo sales.

Another OLS regression was conducted to test the difference of effect on sales in above and below average rated stores. Since the mean of the ratings of all the stores in the dataset was 3.93 (4), we separated stores with a dummy variable called *above\_average\_rating*. The results of this regression indicate that there is a 17% increase in the average Sapporo beer sales when we switch from a below 4 rated stores to a 4 or above rated stores.

**Discussion**

*Suggestions*

Based on the regression results, Sapporo should focus their marketing efforts in areas with higher median incomes, as there is a significant increase in sales in these areas as median salaries increase. Finally, Sapporo should consider the student population as a factor in locating potential stores, as there is a positive relationship between the student population and beer sales. Moreover, Sapporo should definitely focus on trying to enter different areas through stores with higher Google ratings as we see a significant improvement in average sales.

*Limitations*

The lack of seasonal data is a limitation to the model as it masks the potential of independent variables during different times of the year. Since our dataset is a snapshot of the whole year's sales, the model generated cannot reveal the independent variables' effect on sales at different points of time. In addition, the model could be missing important variables for which data could not be gathered for. There are many factors that influence the sales of Sapporo beer, and it is not accurate to associate sales only to the variables included in the model. Another limitation is that the data used in the model was restricted to one state. It is important to note that these same independent variables could have different effects on the sales in different states. The data used would obviously change and could lead to a different outcome.

**Appendix:**

Table

Description automatically generated

*Table 1*: Regression results of the random effects & yearly OLS regression models

Table

Description automatically generated

*Figure 1:* OLS Regression - log transformed total sales v. above average rating (dummy variable)

**Code Files Used:**

1. **R Code File** : This file contains the specifications of the models used to derive the results in the table above. Models include : Random Effects & OLS Regression.

* File: *Sapporo\_models.R*

1. **Python File:** Multiple files that consists of data scraping from the internet, census data API and census data websites. These python files also include the cleaning, merging and arranging of data that was in the end used to run regressions on.

* Files: *data\_scraping.ipynb* & *Sapporo\_analysis.ipynb*

**The random effects model:**

A random effects model is a statistical model that is used to analyze data from multiple groups or clusters. The groups are assumed to have different underlying population distributions. In the context of Sapporo beer, a random effects model could be used to analyze data from different geographical areas or zip codes, where each area or zip code is considered a separate group or cluster.

To fit a random effects model for Sapporo beer, we first collected data on factors that may influence sales, such as the Asian population, income, age groups,  and stores in different zip codes. We then used statistical software to fit a regression model that includes one or more predictor variables (such as the Asian population) and one response variable (Sapporo’s beer sales). The model takes into account the fact that there may be different underlying population distributions in different geographical areas or zip codes, and estimates the relationship between the predictor variables and the outcome for each group.

After the model has been fitted, we analyzed the results to see if the relationship between predictor variables and our response variable were in line with our hypothesis. This would help Sapporo understand the factors that influence their beer sales and identify ways to increase sales in the future. For example, if the Asian population in a given area increases, the model would predict how much Sapporo’s beer sales would increase as a result. Additionally, the model could test whether the relationship between the predictor variables and sales is linear or if there are certain thresholds where sales start to increase more quickly.