Analyzing neighborhoods and beer consume by students in the vicinity of University of São Paulo in São Paulo/ Brazil to start a Pub & Beer Store.

Diego Marcelo Inácio de Barros

February 06, 2021

1. INTRODUCTION

Beer is the second most consumed alcoholic beverage in the world and the most consumed in Brazil [1]. In this country the consumption percentage of beer falls in approximately 60% and the Brazilians are the third in consuming this fresh and bubbling drink and presenting growth in this habit [2]. The city of São Paulo is the biggest one in all South America and have 12.33 million habitants and data available about this theme and in that population center inside University of São Paulo, USP, is available on Kaggle[3]. In that city exist a borough called Butantã and there is located the USP, the better college in all South America and one of the biggest campus in all the state. Also the total revenue for Brazil in 2021 was predicted to be 51,160 million dollars and the average per capita for the same year to be 239.07 dollars 4. So the intention in this report is to justify the best time and place to open a Pub & Beer Store using the Foursquare API and create a model to evaluate a tendency in the consumption of beer through time in the same city and what's the best variables to use in that model. The final goal is to obtain the best profitable place in the Butanta's borough in São Paulo and knowing the consumer behavior through a dataset will create an opportunity to ensure bigger sales and better brand exposure.

2. DATA ACQUISITION AND CLEANING

Data about the consume of beer by USP students during the entire year of 2015 was collect from Kaggle, this data as information about:

- Data (year-month-day);
- Average Temperature in degrees Celsius;
- Minimum Temperature in degrees Celsius;
- Maximum Temperature in degrees Celsius;
- Precipitacion in milimeters of rain;
- Weekend – categorical values 0 for no and 1 for yes.
- Beer consumed in liters.

In this study data concerning the geopositional coordinates from the neighborhoods around the college campus was necessary and obtained through the use of Google maps that provided the latitude and longitude of the 6 principal boroughs around Butantã and this neighborhood also. After having obtained this information and specifying the radius and limit of search, the Foursquare API was used and gave us, after some previous use of a function to extract from a json file the necessary info, detailed data concerning the venues as follows:

- Type;
- Latitude;

- Longitude;
- Category.

The cleaning and preparation of the data acquired from Foursquare was necessary to transform the objects (strings) represented in the Venue Category column to several columns and if some venue category was present in some neighborhood the interpretable category type values would show number 1 (yes) and if not 0. This manipulation prepared the info to be used to create a cluster model based on KMeans algorithm.

Dataset across consume of beer as initially 941 rows and 7 columns, but after check for Not a Number, NaN, values and drop the sames from the data frame. This one ended having 365 rows and 7 columns, each one of the rows represent data acquired trough each day of year 2015. After this, select the best features to be used in a linear regression model, for that was necessary using the method .corr () and with that selecting the higher score to be used in the model.

3. EXPLORATORY DATA ANALYSIS

3.1 - Foursquare API

After generated a new data frame using the groupby method the figure 1 could be produced and show that after using the Foursquare API in a radius of 3 Kilometers in each of the 7 boroughs around USP campus the most frequently venues in a decreasing way are restaurant, pizza place, brazilian restaurant, bakery, convenience store, plaza, supermarket, martial arts school and bar. But the only places that could be a menace to the business plan are bar, pizza place, convenience store, so only three out of ten could be considered like that.

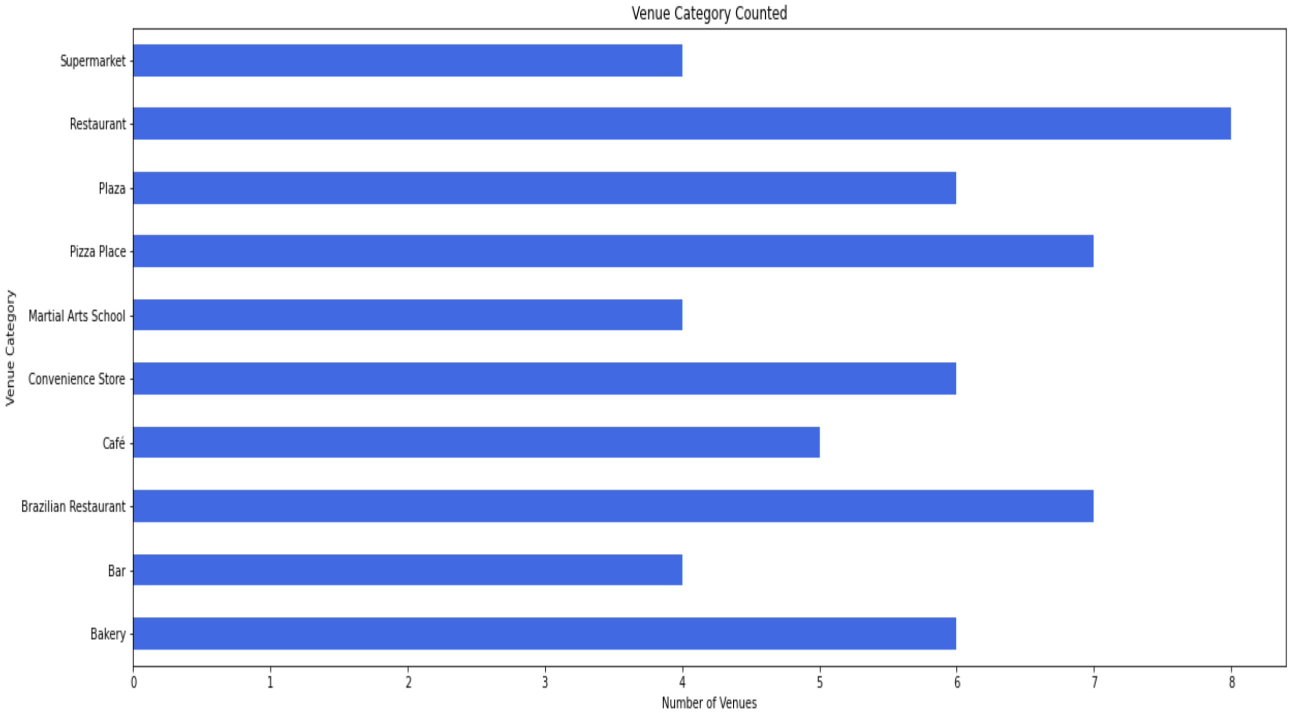

Figure 1: Venue Category Counted

With the same strategy previously utilized and using the Folium library the graph in the figure 2 and 3 provided understandable vision about the geographic location and type of venues that can sell any alcoholic beverages and, with that, fight to conquer possible clients in that neighborhood. And with that could be perceivable that Rio Pequeno would be a better place to start that sort of store (figure 3), when take to mind about competition, but for a matter of contrast the Jardim Everest and Bonfiglioli are located in the neighborhood called as Jardins in the South side of São Paulo. That place has the most expensive square meter in all the city except when compared to the Paulista avenue and the same boroughs accommodate the greater concentration of venues that sell drinks and beer (figure 2).

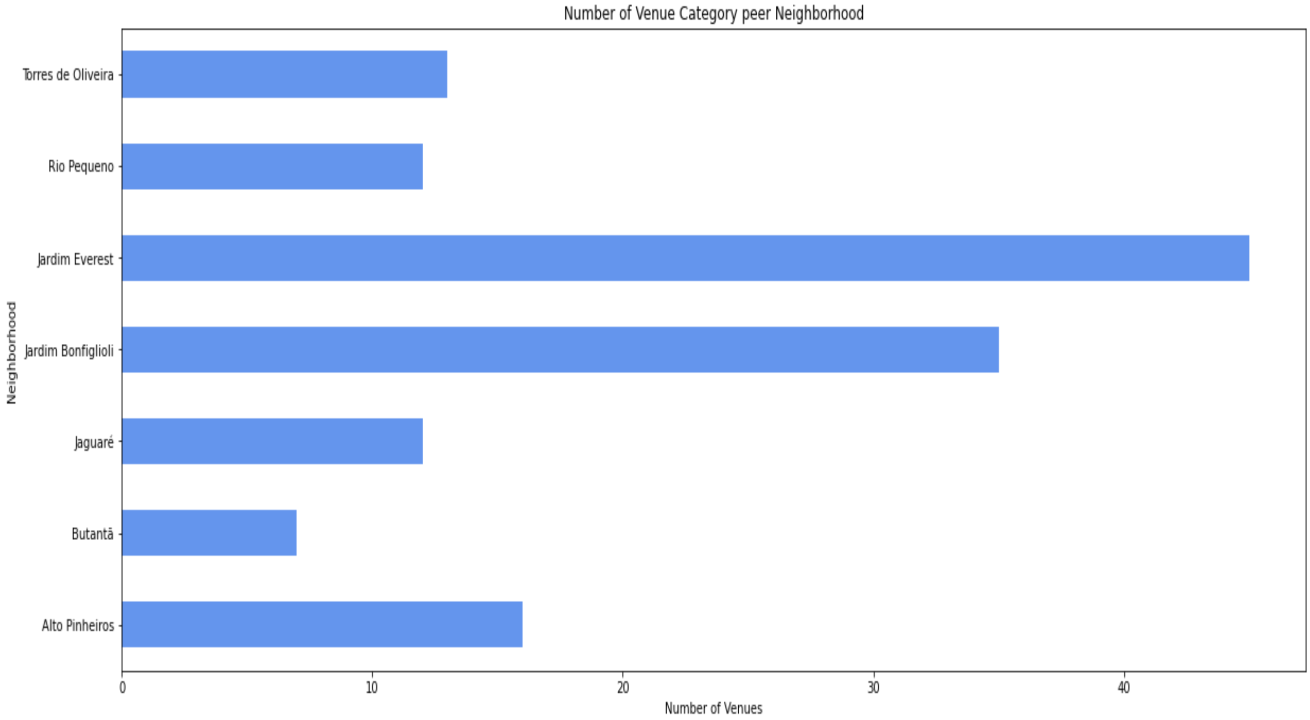

Figure 2: Number of venue category peer Neighborhood

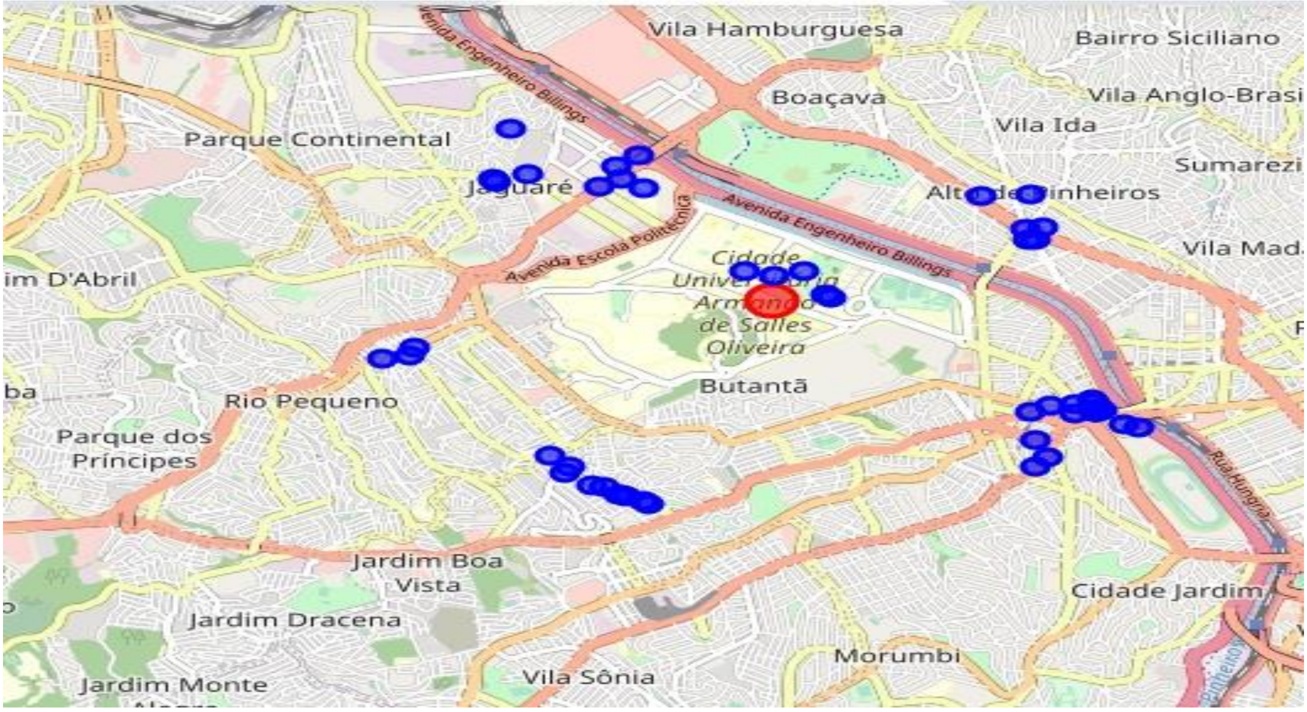

Figure 3: Map with data points in blue showing less concentration of venues in Rio Pequeno and more in Cidade Jardim (Jardim Bonfiglioli and Jardim Everest).

After looping the all the possible numbers of cluster in the KMeans algorithm and test the best silhouette score and the same was decided to be 2 and having and score equal to 0.274. With that info and using Folium library the figure 4 was created.

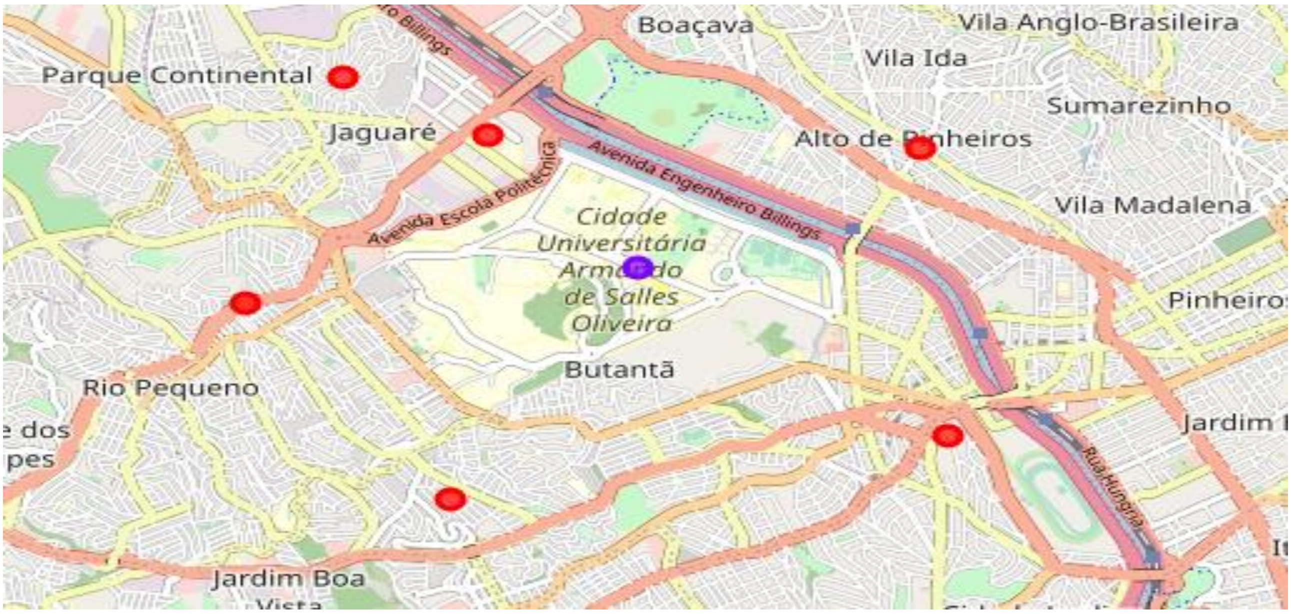

Figure 4: Map with clusters in red and purple.

3.2 - Consuming of beer by University of São Paulo students

Following after cleaning the dataset acquired in Kaggle regarding to consume of beer by USP students, several graphs could be generated and reveal that consume of beer (dependent variable) increase with higher average temperature (figure 5), when it's weekend (figure 6) and tends to decrease when its raining (figure 7). It's necessary to clarify that in figure 6, the blue box plot, representing the days of Week, has several outliers and they are there because in Brazil exists several holidays that not falls in Weekends.

Another perceptible thing is that in the summer months in the south hemisphere - December, January and February - consumes tends to be higher even with college vacation (figure 8), this could be explained by the fact that several students are having their families living in the vicinity or in other states, also the public dormitories given by the college to students that don't have financial conditions to rent a place to live are inside the campus and are used by several students by the entire year

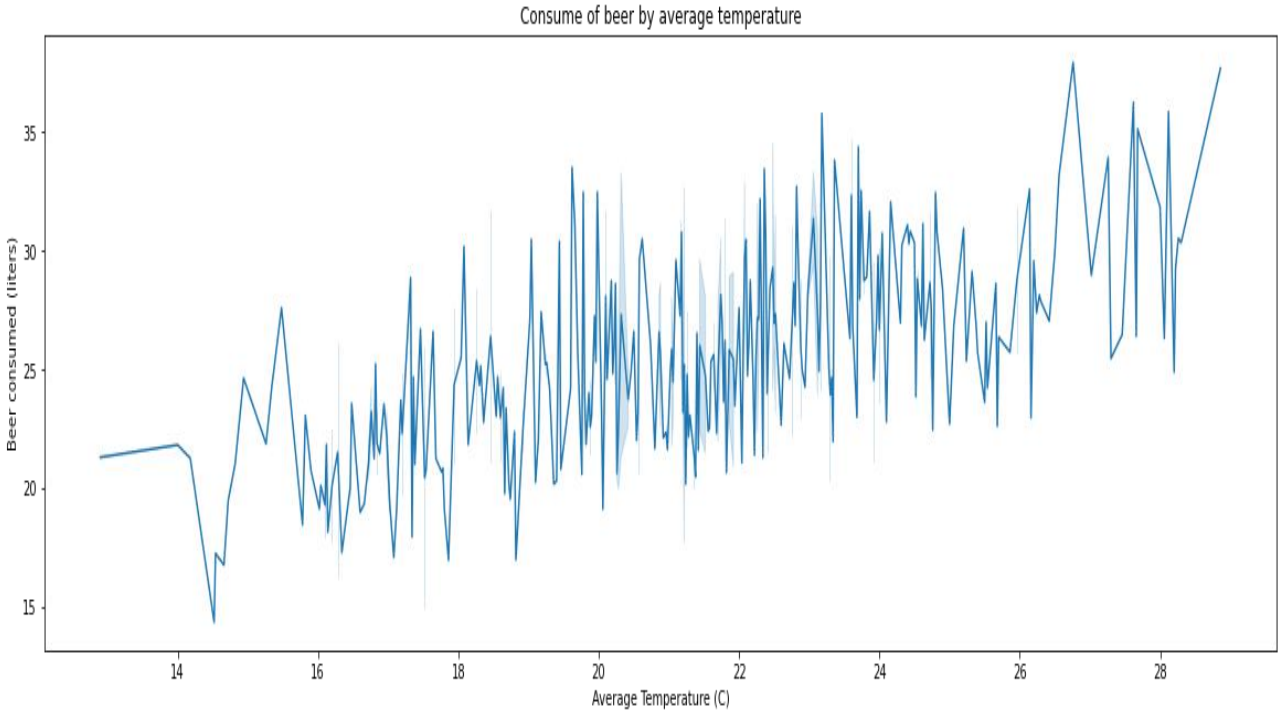

Figure 5: Cosume of beer by average temperature.

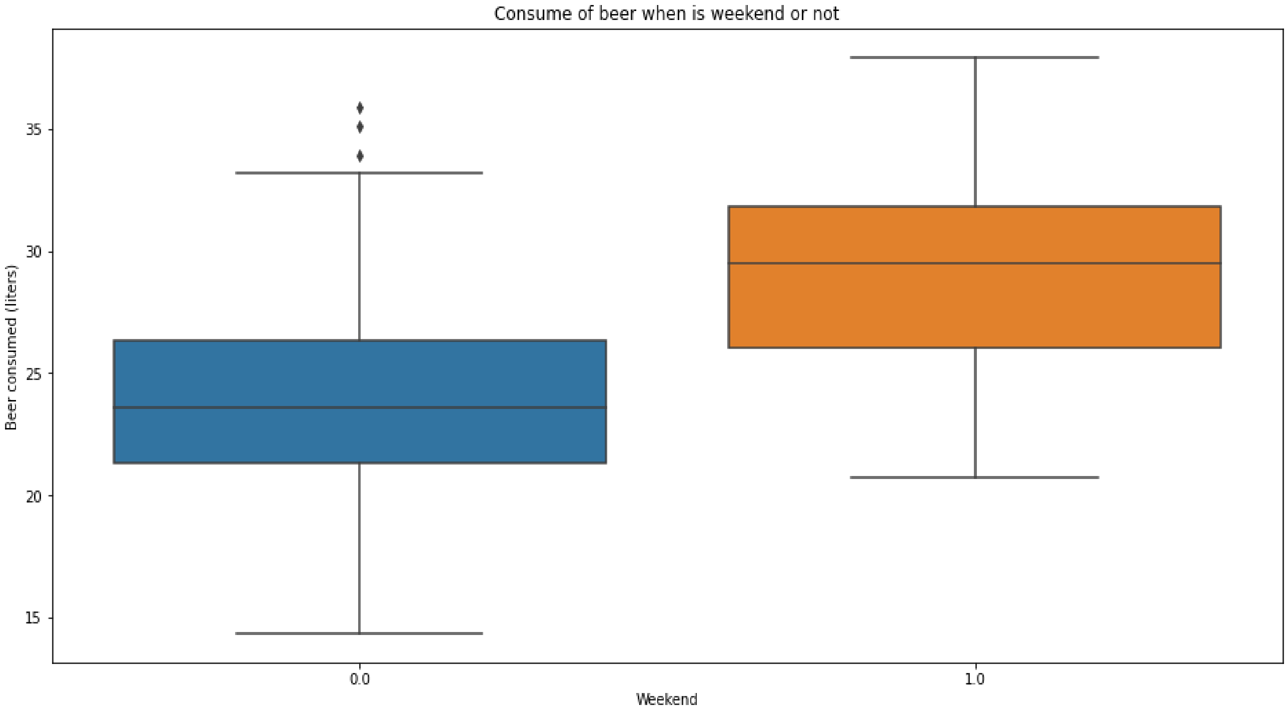

Figure 6: Consume of beer when is Weekend or not.

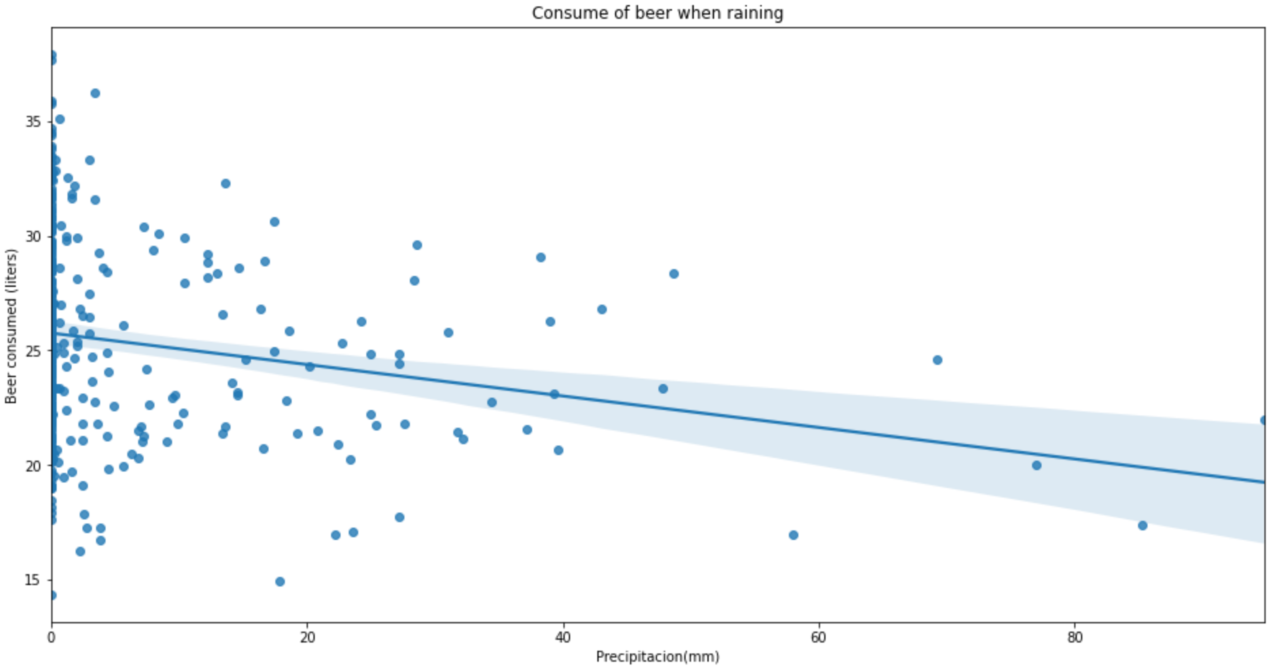

Figure 7: Consume of beer peer volume of rain.

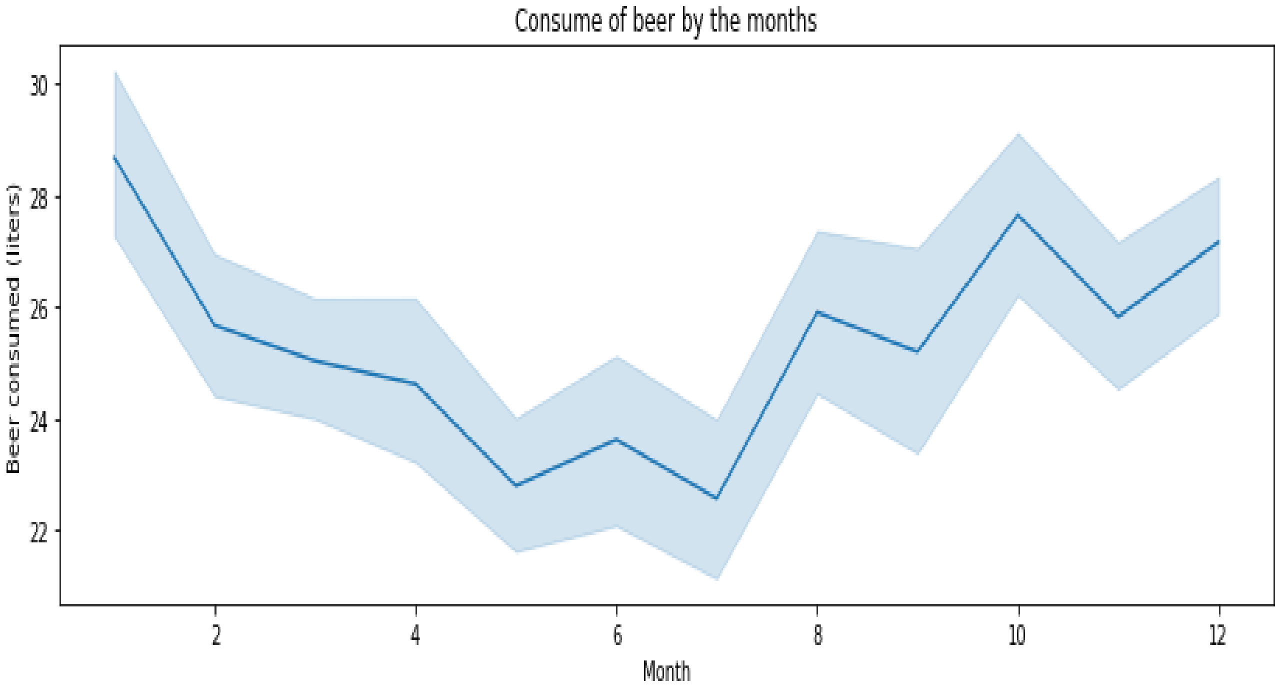

Figure 8: Consume of beer trough the 2015 year.

Conclusion

We analysed several types of data and could determine the better place surrounding the University of São Paulo, and also the best months, time conditions and We analysed several types of data and could determine the better place surrounding the University of São Paulo, and also the best months e time conditions to sell our product. The stakeholders should take to mind costs about buying or renting a local, legal and sanitary questions when the time to take the plan to action. Maybe further investigation on that fields should be advided.

References

1. https://ourworldindata.org/alcohol-consumption#:~:text=consumption%20per%20person-,Alcohol%20consumption%20across%20the%20world%20today,of%20pure%20alcohol%20per%20year.
2. https://www.kirinholdings.co.jp/english/news/2019/1224_01.html
3. https://www.kaggle.com/dongeorge/beer-consumption-sao-paulo
4. https://www.statista.com/outlook/10010000/115/beer/brazil#market-arpu