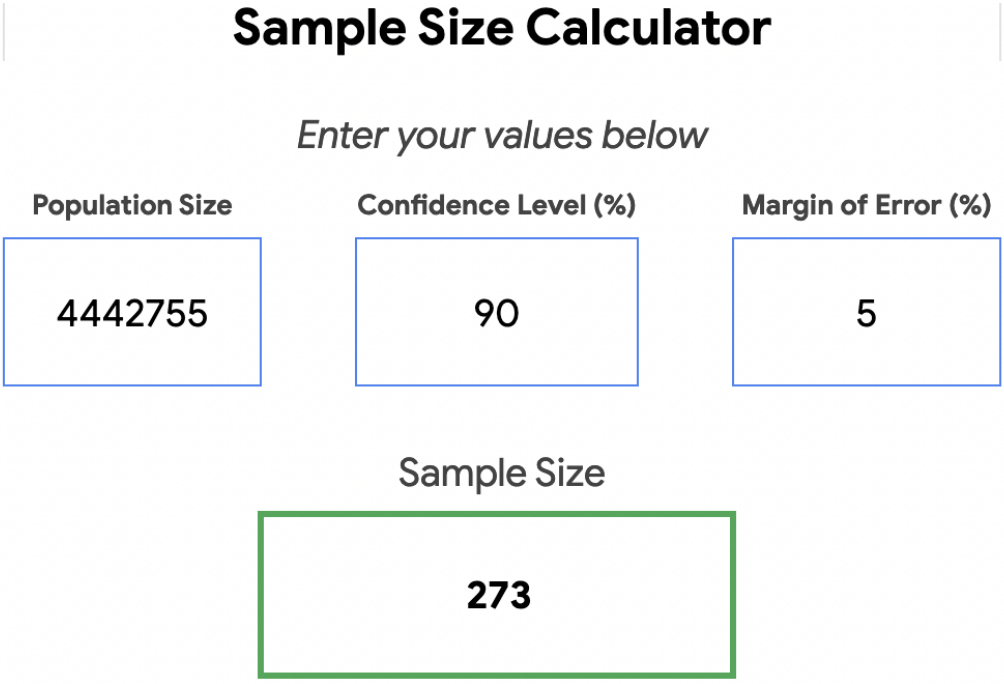
CBD Market Share Report

The main objective of this analysis is to know what are the best products that contain CBD extract that are worth developing on the ecuadorian market

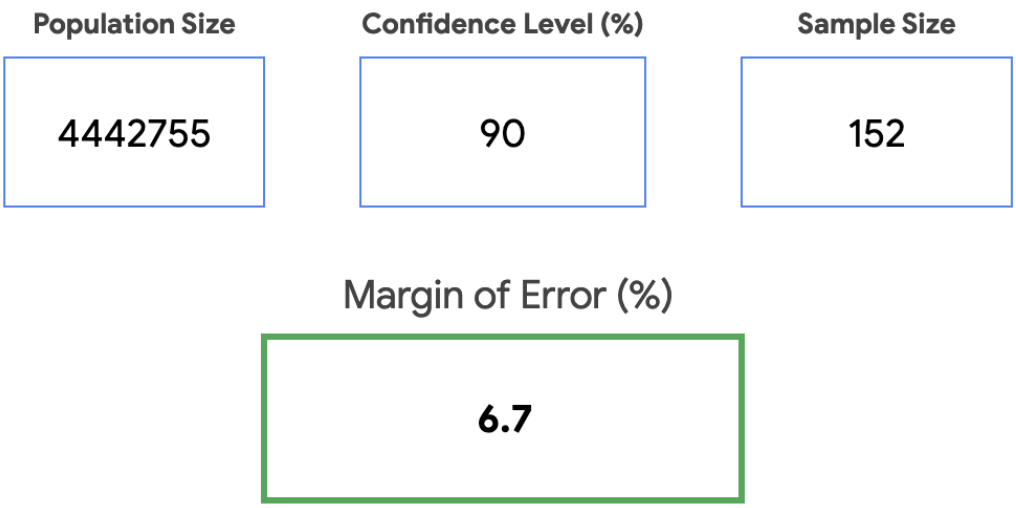
As a secondary objective it is important to determine the age groups of our potential customers in order to define marketing strategies after this analysis.

Sample Size

In order to determine the sample size for this survey we have to put some constraints on the population, the survey was only taken on the ecuadorian population, a sample size determined for the two main cities in the country, for Quito a population of 1’399.814 is used and for Guayaquil a population of 3’042.941 is used, giving us a total population of 4’442.755. Using the population we calculated the sample size for our analysis for a 90% confidence level and a 5% margin of error giving a result of 273 responses.



The number of responses could not be met because of time constraints, the next step was to calculate the margin of error of our analysis by using the 152 responses we have collected. The margin of error for the analysis at a 90% confidence level and with the 152 sample size is 6.7%.



The 1.7% more in error margin is tolerable, we could work with this data set.

Data Sources

All of our data has come from surveys developed and recollected using google forms as surveys. The survey was active from July 27th until August 11th 2022. This survey is kept anonymous, no personal contact information is collected. The not collection of location could be a problem in order to asses later the bias of this parameter in the data.

For the original survey you could follow the link

The translated survey is written below:

1. Age
    a. Surveyed Input Required
2. With what gender do you seem identified?
    a. Male
    b. Female
    c. Prefer not to tell
3. Have you ever used marijuana recreationally or medicinally?
    a. Yes
    b. No
4. Do you know the difference between THC and CBD?
    a. Yes
    b. No
5. Do you know any of these medical properties of CBD?
    a. Insomnia Treatment
    b. Natural Anxiolytic
    c. Chronic Pain Treatment
    d. Anti Inflammatory
6. Have you ever consumed any product that contains CBD?

    a. Yes
    b. No
7. If you answered yes to the previous question. What product have you consumed?
    a. Aromatic Flower
    b. Concentrated Oil
    c. Gummies
    d. Cookies/Desserts
    e. Aromatized Beer
    f. Aromatized Beverages
    g. Chocolates
    h. Other (Surveyed Input)
8. Would you be willing to consume any of the following products if they contained CBD?
    a. Aromatic Flower
    b. Concentrated Oil
    c. Gummies
    d. Cookies/Desserts
    e. Aromatized Beer
    f. Aromatized Beverages
    g. None
    h. Other (Surveyed Input)

Cleaning and Data Manipulation

“Consumo de CBD (Responses)” is the title given to the original document, the modified document is named “Consumo de CBD (Manipulado)” this document has all the data cleaning and manipulation described below. This data is located on my personal google drive and also on the github repository created to share this project.

The first step taken to clean this data is to correct the age input, some people entered the age value with the word “años” (translated to “years”) beside. There were two values with these characteristics on row 70 and 135.

To the question “¿Conoce usted alguna de estas propiedades médicas del CBD?” translated as “Do you know any therapeutic properties of CBD?” we have some blank spaces, this would be fill with the “Ninguna” acronym as not available or unknown, filters for blank spaces were applied in order to change fill it. 11 values were completed. Also this column was split into separated columns using the “Split into columns” tool using comma as the separator..

The question “De haber respondido afirmativamente a la pregunta anterior. ¿Qué producto ha consumido usted?” translated as “If the answer to the last question is affirmative. What product have you consumed?”. In the responses we have blank, “na”, “ninguno”, “no he consumido”, this responses have been replaced by the word “Ninguno”, also one response written as “crema para humectar” is changed to “Crema Para Humectar” just for aesthetic reasons but there was two answers where “Ungüentos o cremas corporales”, “Crema dermatologica” and “Cremas” was collected as an input, thes response is replaced by “Cremas Para Humectar” as by definition these are the same.

For the last question “¿Estaría usted dispuesto a consumir alguno de los siguientes productos si tuvieran CBD?” translated as “Would you be willing to consume any of the following products if they contained

CBD?". All the negative questions, 17 responses in total were transformed into the word "Ninguno", also one response has a really long message were the person surveyed put "Flor Aromática, Galletas/Postres, Cerveza Aromatizada, Bebidas Aromatizadas, Consumiría todo lo que se pueda… la planta tal cual con niveles de THC que NO me embriaguen" these answer is replaced by the word "Todo"

One of the last transformations of data was made to the response where a person put "Ninguna" as an option of products he/she will consume but this means "None" so this will be deleted and taken as a typo.

In the same sheet questions 7 and 8 were splitted by using the "Split into columns" tool using the comma separator as the delimiter.

Analysis Summary

For every question we have created a new sheet, the responses for all the questions were counted for every option using the function =COUNTIF() in order to process this in a better way.

The file "Consumo de CBD (Manipulado)" has charts made on Google Spreadsheets which are the charts on this document.

For questions 5, 7 and 8 I choose bar graphs as the best way to present this data, I order this in descendent order to put the highest value on top. Question 1 is represented by a histogram, showing the greatest age group to target as a market group, also questions 2, 3, 4 and 6 are being depicted as pie charts to show the percentage for the two options these depict as is easy to see the percentage depict as an area inside the circle.

I also performed a tetrachoric correlation to find if there is any correlation between question 1 and 3, questions 3 and 4, and question 3 and 6.

The tetrachoric tables used for the correlation calculation are depicted below. I used R as a tool to calculate the tetrachoric correlation.

		Gender	
		Male	**Female**
Have You Tried Marihuana in a Recreational Way? (Question 3)	**Yes**	52	26
	No	36	38

The relationship between trying marihuana and gender (either Male or Female) is 0.28, a low correlation which means that there is no effect here.

The other relationship I wanted to clarify is the use relationship between trying marihuana and knowing the difference between THC and CBD.

		Do you know the difference between THC and CBD? (Question 4)	
		Yes	**No**
Have You Tried Marihuana in a Recreational Way? (Question 3)	**Yes**	59	19
	No	28	46

The relationship between trying marihuana and knowing the difference between THC and CBD is 0.57, a moderate correlation that is worth investigating in order to find something.

The last correlation I wanted to explore is the relationship between trying marihuana and the willingness to try CBD products.

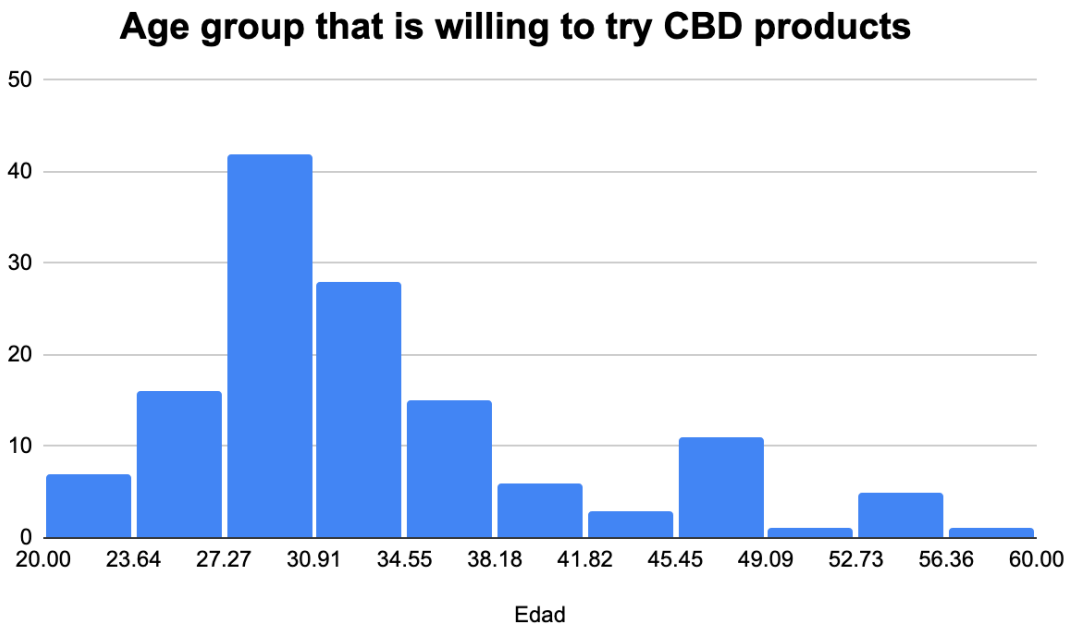
		Have you ever tried any product that contains CBD? (Question 6)	
		Yes	**No**
Have You Tried Marihuana in a Recreational Way? (Question 3)	**Yes**	57	21
	No	14	60

The relationship between trying marihuana and the willingness to try CBD products is 0.76, this correlation value is really high so it means that here there is an effect that is worth studying in depth.

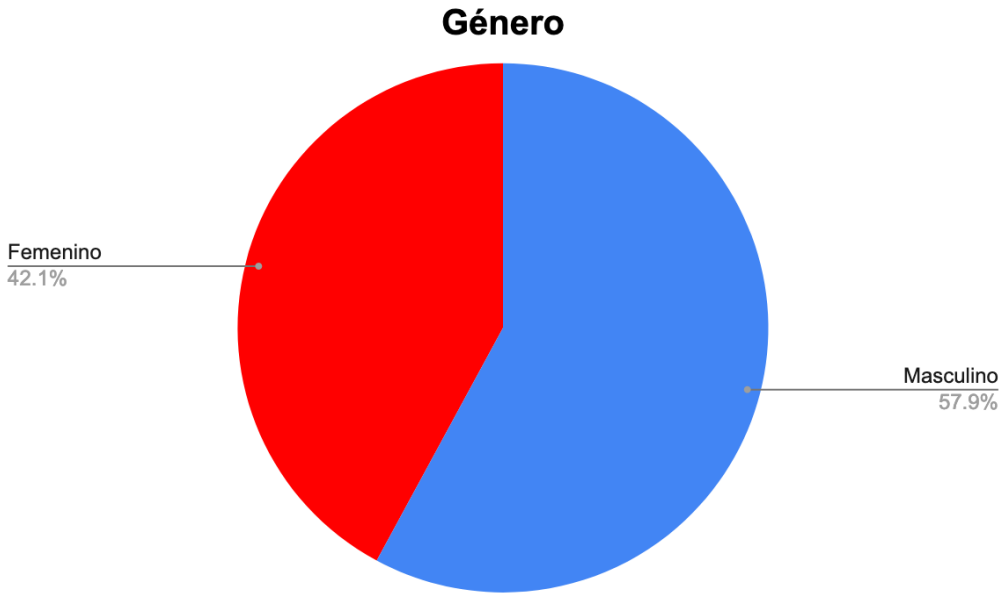
The R code for the calculation of the tetrachoric correlation are shown on the git hub and at the end of this document.

Supporting Visualization and Key Findings

For the first question I graph a histogram in order to find the target age group for our products. Being the 23 to 49 the target age group for our products.

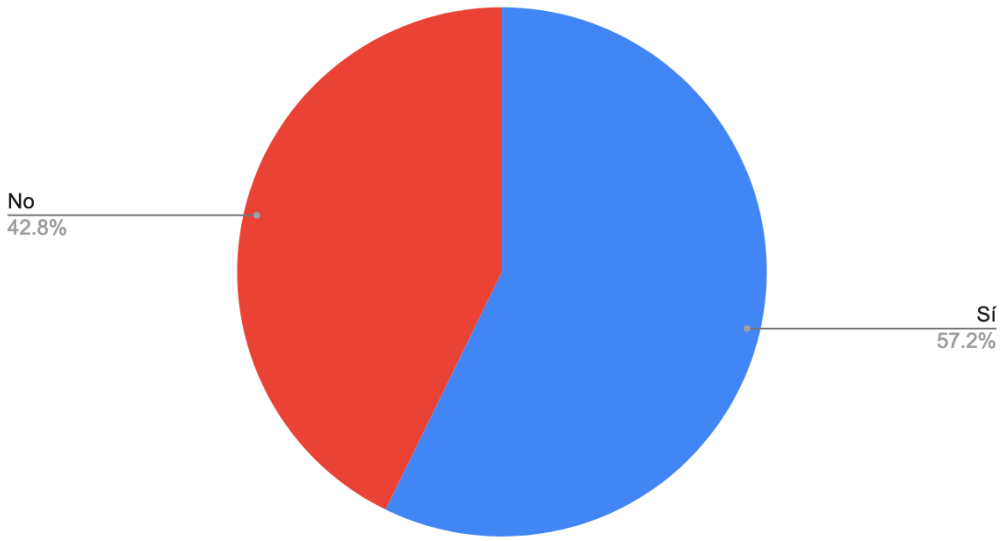


For questions 2, 3 and 5 I graph pie charts as these questions have only two options as answers.

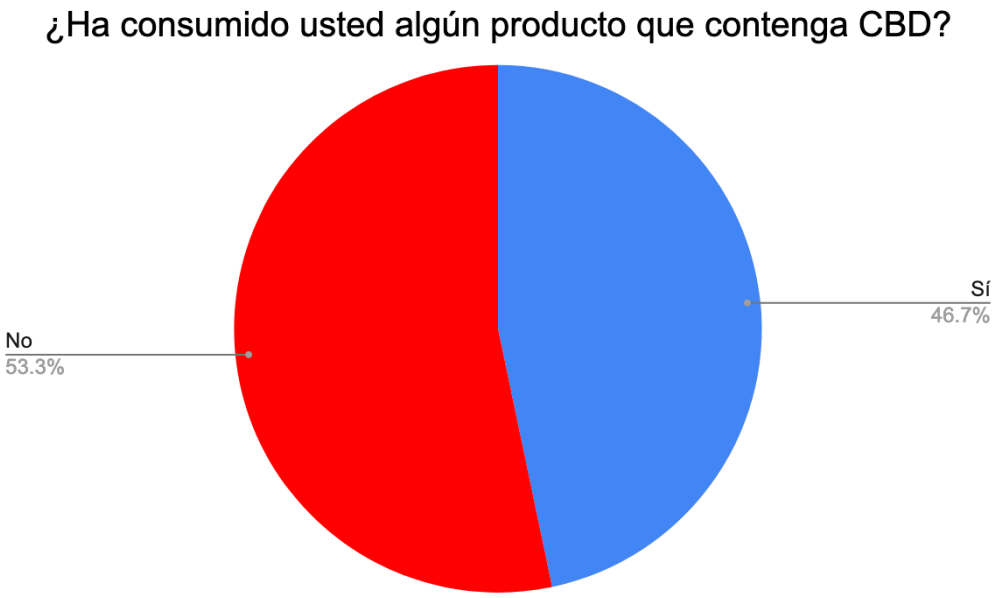


Uso de Marihuana de Manera Recreativa o Medicinal

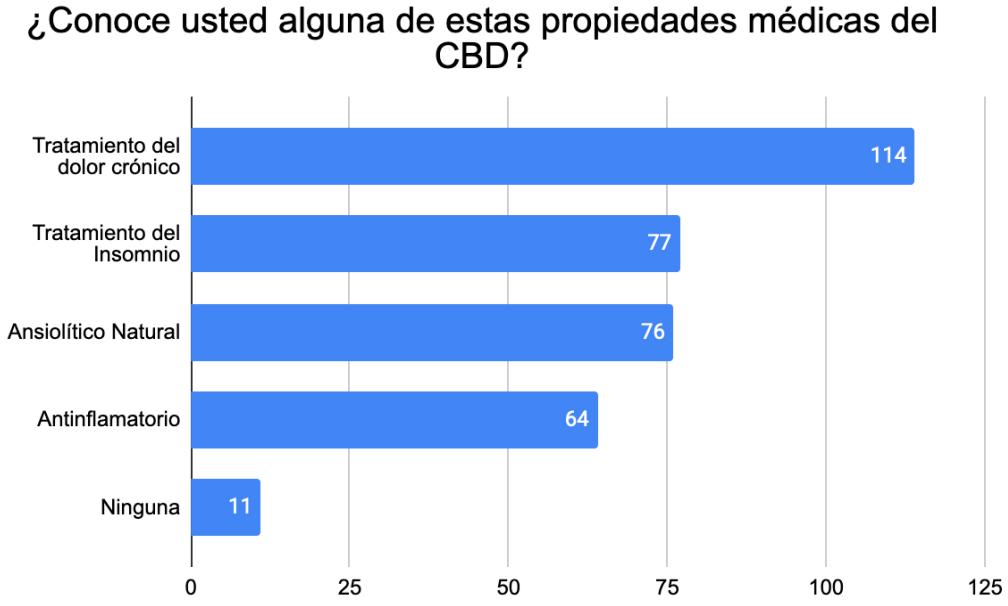
¿Conoce la Diferencia entre THC y CBD?



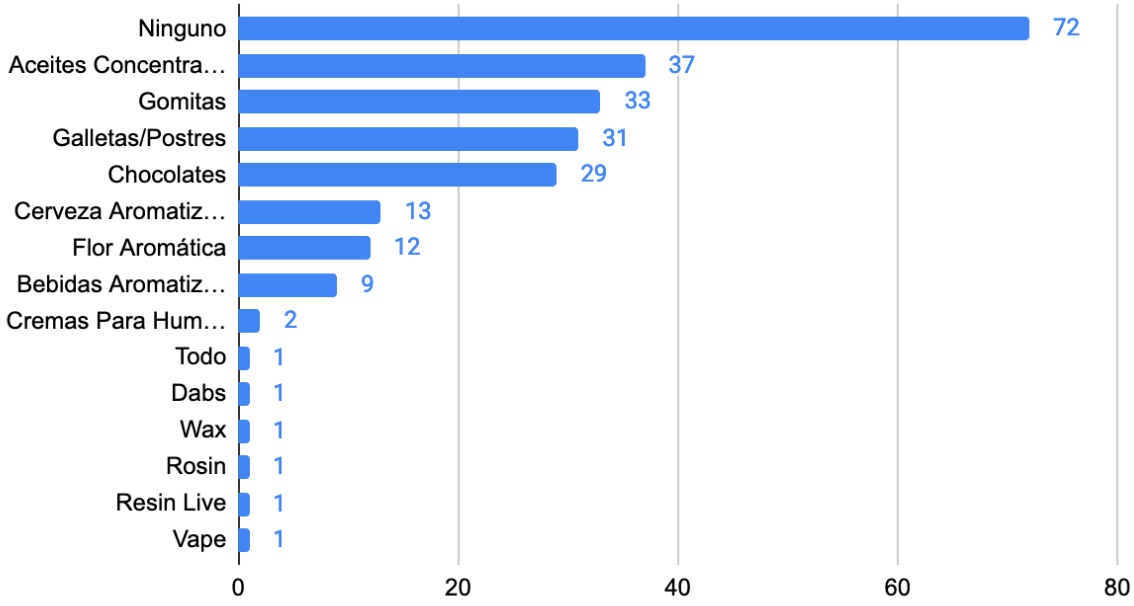
Also for question 6 I choose to graph a pie chart, for representing these the “Yes” or “No” answer.



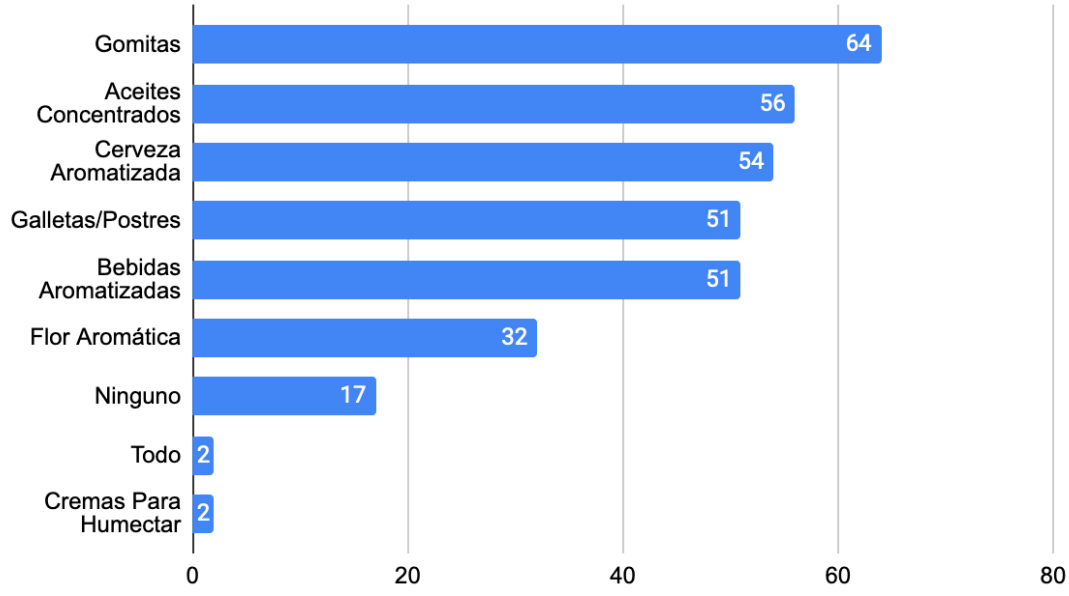
For questions 5, 7 and 8 I choose to graph these as horizontal bar charts in descending order showing the most voted option on top.



Productos que han sido consumidos



Productos dispuestos a consumir



Also these charts have been uploaded into a dashboard on Tableau.
Key Findings

After I have finished my analysis my key findings are these:

- The age group we have to target for our products is between 23 to 49 years old
- The most known medical effect for CBD products is the treatment of chronic pain
- Also almost half our surveyed have not tried CBD products, but the ones that have tried CBD products mentioned that concentrated oils, gummies and cookies/desserts are the most demanded.
- The products that the people are more willing to try are gummies, concentrated oils and aromatized beer.
- The tetrachoric correlation that has the most strong positive correlation is when we compare the results from question 3 related to the consumption of marihuana as a recreational matter and question 6 related to the consumption of CBD products with a 0.76 as a correlation coefficient.

Append

R code used for Tetrachoric Correlation

```
library(psych)

#Creating a 2x2 matrix

data1 = matrix(c(59,28,19,46), nrow = 2)

#View Table
data1

#Tetrachoric Correlation of
tetrachoric(data1)

#Creating a 2x2 matrix

data2 = matrix(c(57,14,21,60), nrow = 2)
```