CA 1

CSE353

EDA Project

Submitted By : Amardeep Singh Gujraal

Reg No. : 12006933

Section: K20CH

**Introduction to dataset.**

This dataset is about how alcohol consumption is related to happiness of people around the world.

* **Basic** Structure

Text

Description automatically generated

Dataset contains data for 122 countries and as 9 columns namely:

1. Country
2. Region
3. Hemisphere
4. HappinessScore
5. HDI
6. GDP\_PerCapita
7. Beer\_PerCapita (in liters)
8. Spirit\_PerCapita (in liters)
9. Wine\_PerCapita (in liters)

Luckily, we have zero null values so we can easily work on this dataset without worrying about missing or improper values.

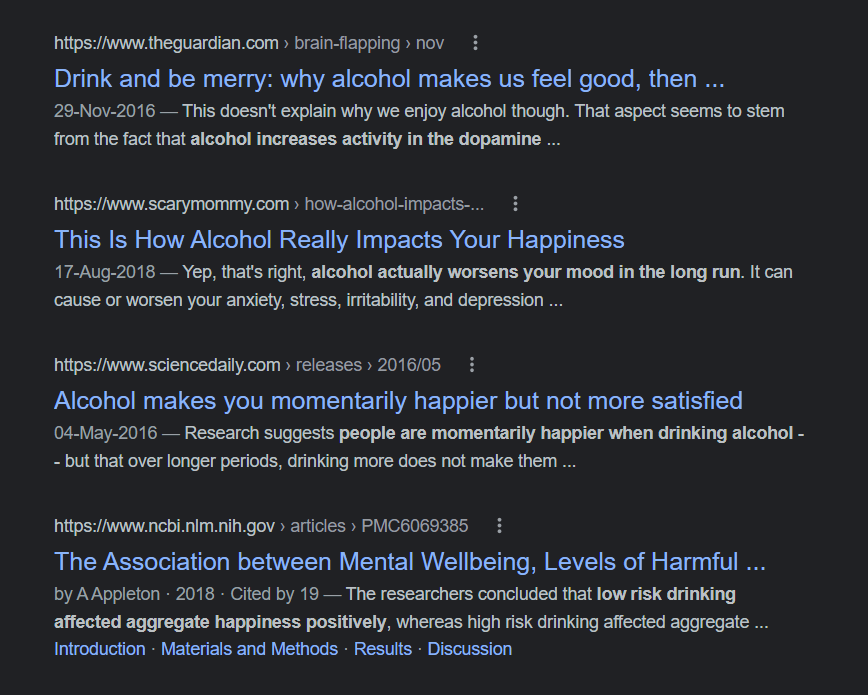
**Why this dataset**

All of us have heard this that “*happiness and alcohol consumption go hand in hand*”

so, I choose this dataset to find if that’s even true? Or its just a myth companies have been taking advantage of and a billion-dollar market is running on.

Using data analysis tools like Pandas, NumPy, matplotlib and seaborn I can find many relations between happiness and alcohol consumption and in depth about which beverage in particular has that happiness potion in it.

Why this Topic?



Here are top 4 google results on same topic, you can see they all have different opinions. So how can we know which one is true?

Let’s find out ourselves.

For basic instance here’s how top 5 rows look like:

Graphical user interface

Description automatically generated with medium confidence

This will give you enough idea about what we will be working on, So lets get started.

**What are the insights we will be**

**finding here**

1. We will be looking at which regions around the world are happiest and which aren’t
2. We will also be looking at which countries are on top, and bottom of happiness score provided by United Nations.
3. Then we will be looking at human development index of all the countries.
4. Then we will check GDP of these countries.
5. Next, we will be looking at which countries consume the most amount of alcohol.
6. Finally, we will be looking at the correlations of this data and conclude our study.

We will be using powerful libraries like pandas, NumPy, seaborn, matplotlib etc to demonstrate this data on graphs and charts.