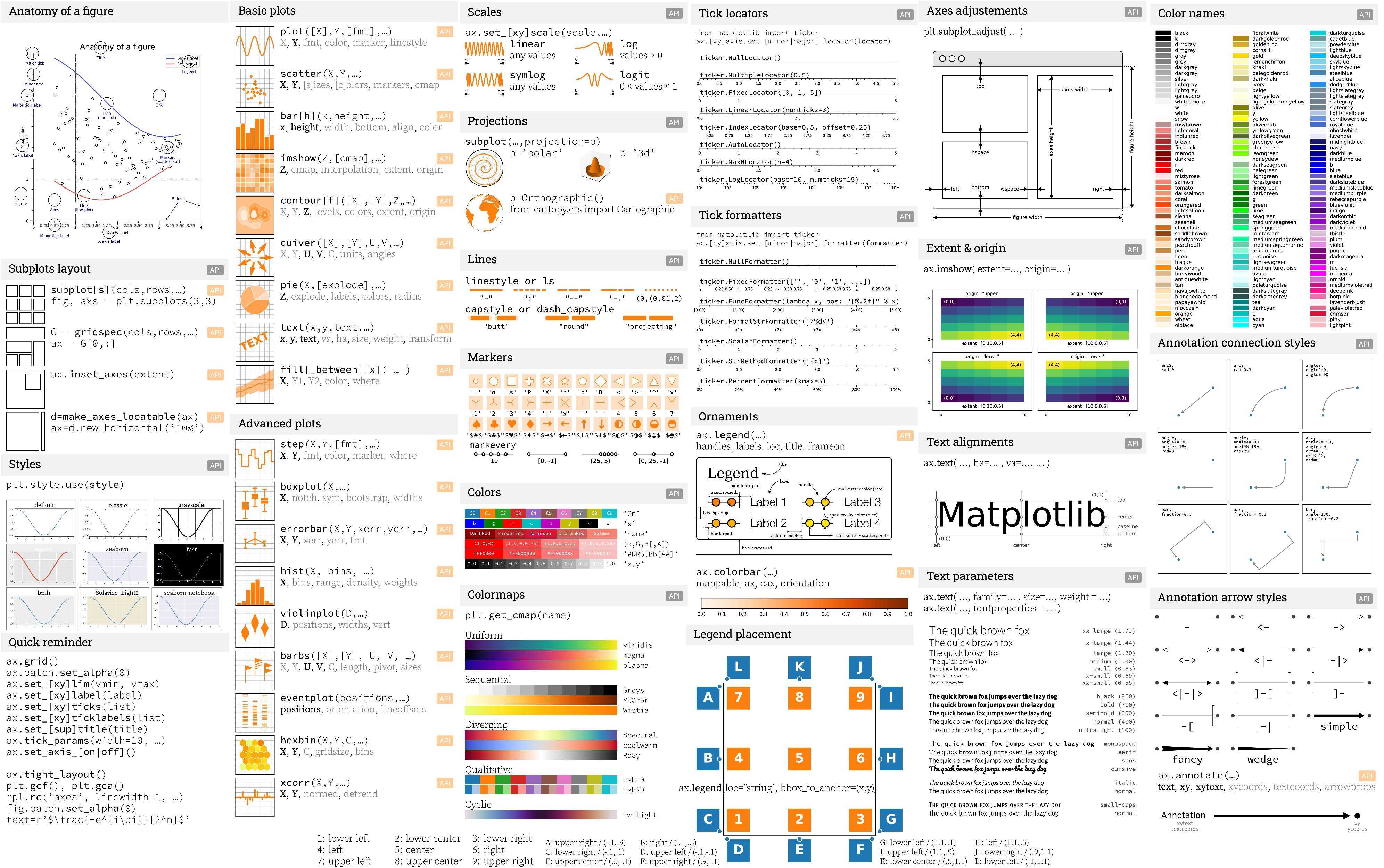
## línea horizontal



Individual Delivery

23/07/2020 to 31/08/2020

**─**

Rosario Marianeschi

Data Science

The Bridge

Table of Contents

General Vision 2

Goals 2

Specifications 2

Software 2

Hardware 2

Requirements 3

Steps 3

I. Research the context 3

II. Get Data 6

III. Data Wrangling 6

IV. Data Mining / Clean Data 7

V. Others. 7

VI. Analysis & Conclusions 7

References 9

# General Vision

My vision is to prove the Hypothesis about ‘Using THC and CBD in some products can offer better effects for health improvement’ and also ‘The importance of cannabidiol as a component in pharmaceutical, beauty and health products’. The aim is to find out the best use of Cannabis Strains & Features, using the tools and processes learnt during the first 7 weeks of DataScience BootCamp at The Bridge.

# Goals

Accomplish the required individual project. So my specific goals will be:

* Find a good file in .csv to use as raw data.
* Clean and arrange the information in a way that can be useful and readable.
* Find the best functions to rearrange the dataframe and prove the Hypothesis.
* Structure of a real ‘project’, with all the folders and Python files very well developed.
* Answer the questions proposed by the teachers in the assignment.
* Visualize the data obtained in graphs and plots.
* Write conclusions.

# Specifications

To achieve the goals and make the most out of the delivery, below you can see the detailed specifications.

## Software

Minimal Software needed for execution is MacBook Air – iOS Mac HD, OS X, 10.10.5

Platforms:

VS Code has been tested on the following platforms:

* OS X Yosemite
* Windows 7 (with .NET Framework 4.5.2), 8.0, 8.1 and 10 (32-bit and 64-bit)
* Linux (Debian): Ubuntu Desktop 14.04, Debian 7
* Linux (Red Hat): Red Hat Enterprise Linux 7, CentOS 7, Fedora 23

## Hardware

Minimal hardware needed for visualizing this work:

Visual Studio Code is a small download (< 100 MB) and has a disk footprint of 200 MB. VS Code is lightweight and should easily run on today's hardware.

It is recommended:

* 1.6 GHz or faster processor
* 1 GB of RAM

## Requirements

Visual Studio Code

Additional Windows requirements:

Microsoft .NET Framework 4.5.2 is required for VS Code. If you are using Windows 7, please make sure [.NET Framework 4.5.2](https://www.microsoft.com/download/details.aspx?id=42643" \t "_blank) is installed.

Additional Linux requirements:

* GLIBCXX version 3.4.15 or later
* GLIBC version 2.15 or later

Python 3.6 – Libraries to import: matplotlib, pandas, seaborn, plotly.express, plotly.graph\_objects as go, plotly.figure\_factory as ff, from plotly.colors import n\_colors, from plotly.subplots import make\_subplots, nbconvert, nbformat, %matplotlib inline, plot as pltt and pycountry.

# Steps

## Research the context

What is Cannabis? Cannabis is an annual herbaceous plant with two primary classifications — Indica and Sativa. **Marijuana and hemp are members of this plant genus; hemp belongs to the *Cannabis sativa* species only, whereas marijuana can be a member of either *Cannabis indica* or *Cannabis sativa* species.** Because both marijuana and hemp come from *Cannabis sativa*, they share certain traits. **However, marijuana and hemp are not the same things.** Below we highlight several distinct differences between these two plants.

**What is Hemp?**

Hemp is a cannabis plant that is harvested commercially for its seeds, stalks, and flowers. Because it grows sturdy and tall — up to 2 to 4 meters in height — it’s typically cultivated outdoors.

Different parts of the plant are used for different uses:

* Seeds are often used in food and cosmetics.
* Stalks are the source of fiber used in building materials and clothing.
* Flowers, on the other hand, are harvested for its cannabinoid content.
* The cannabinoid content of hemp is where it differs the most from marijuana.

Namely, hemp comes with high concentrations of cannabidiol (CBD), the non-psychoactive cannabinoid, but it carries almost no THC (below 0.3%). It’s the THC content that gives marijuana its psychoactive effects**.**

Unlike marijuana, hemp has been excluded from the Controlled Substances Act with the introduction of the 2018 Farm Bill. According to the new act, hemp can be commercially grown and manufactured into CBD products for sale to the public.

Hemp contains delta-9-tetrahydrocannabinol (THC), which is the psychoactive ingredient found in hashish. THC is present in all hemp varieties to some extent.

In varieties grown for use as a drug, where males are removed in order to prevent fertilization, THC levels can reach as high as 20-30% in the unfertilized females which are given ample room to flower.

In hemp varieties grown for seed or fiber use, the plants are grown very closely together and a very dense biomass product is obtained, rich in oil from the seeds and fiber from the stalks and low in THC content. EU regulations limit THC content to 0.3% in industrial hemp. In Canada, the THC limit is 1%.

The presence of (some) THC in hemp varieties and the fear that THC could be extracted from industrial hemp for illegal purposes has hampered the development of hemp in many countries. Since the early 1990s, however, many countries, including Canada, Australia, the UK, The Netherlands and Germany, allow hemp plantings and commercial scale production. Plant breeders are working on the development of new varieties which are low in THC.

Hemp is a viable commercial product that has not yet realized its full potential - despite having been in continuous use by mankind for over 6,000 years.

**THC vs. CBD**

During the selective breeding process for medical marijuana, THC:CBD ratios are accounted for and accommodated to the needs of the client's preference/illness. Due to the large genetic diversity and different geographical climates and environments, a wide range of strains and properties exist.

THC is associated with the psychoactive *high*, while CBD is not psychoactive and is purported to have medicinal properties. There is evidence about the safety & efficacy of cannabinoids in the treatment of epilepsy.

Raw marijuana also boasts heart-healthy “good fats” in the form of omega-3 and omega-6 fatty acids. Omega-rich avocado or salmon offer derive skin benefits, but raw cannabis leaves can offer some of the same perks. Raw cannabis leaves are excellent sources of raw fiber and can aid with digestive issues including constipation.

Further, cannabis fan leaves and sugar leaves are abundant in aromatic terpenes that may have antibacterial, antiviral, and even anti-tumor properties. Researchers have conducted numerous studies on the possible anti-tumor effects of cannabis terpenes, some of which have yielded encouraging findings. For example, myrcene, the most abundant terpene in cannabis, has demonstrated the potential to [kill human breast cancer cells](https://link.springer.com/article/10.1007/s13765-015-0081-3" \l "page-1" \t "_blank) in a 2015 study published in the Journal of the Korean Society for Applied Biological Chemistry. 

Whether your cannabis leaves end up in the compost pile to nourish the earth or in your body to nourish you, they can be nutritional powerhouses.

**What is Marijuana?**

Marijuana is a cannabis plant that is harvested for its euphoric, relaxing, and psychoactive properties. As opposed to hemp, the seeds and stalks of marijuana aren’t used commercially as a food source, or in the textiles industry. Instead, the plant is cultivated for its highly resinous flowers containing an abundance of cannabinoids. The THC content of marijuana is much higher than it is in hemp.

Marijuana can have up to 30% THC per dry weight. This type of cannabis is classified as a Schedule I substance under the Controlled Substances Act of 1970.

Therefore, the US federal government doesn’t recognize any medical uses of marijuana and claims it has a strong potential for abuse. Marijuana remains federally illegal in the USA, although states have begun passing legislation that legalizes either medical or recreational marijuana use.

Given the above differences, those new to cannabis may think CBD from hemp is somehow different than CBD from marijuana. It’s not.

Due to its low THC content, hemp-derived CBD oil has no intoxicating effects on the user. In other words, you won’t get high off of CBD hemp products. Instead, you may feel relaxed, more alert, free of physical and mental discomfort, and notice positive changes in your overall well-being thanks to how CBD works in the endocannabinoid system. CBD oil from marijuana can get the user high, although the psychoactive effects of such products are less intense since CBD counteracts THC’s intoxicating properties. In addition, marijuana-derived CBD oil acts as a strong analgesic, relaxant, anti-depressant, and can tackle a wide range of chronic, often treatment-resistant, conditions.

The use of cannabis for recreational purposes is prohibited in most countries; however, many have adopted a policy of decriminalization to make simple possession a non-criminal offense (often similar to a minor traffic violation). Others have much more severe penalties such as some Asian and Middle Eastern countries where possession of even small amounts is punished by imprisonment for several years.

Countries that have legalized recreational cannabis are [Canada](https://en.wikipedia.org/wiki/Canada), [Georgia](https://en.wikipedia.org/wiki/Georgia_(country)), [South Africa](https://en.wikipedia.org/wiki/South_Africa), and [Uruguay](https://en.wikipedia.org/wiki/Uruguay), plus 11 states, 2 territories, and the [District of Columbia](https://en.wikipedia.org/wiki/District_of_Columbia) in the [United States](https://en.wikipedia.org/wiki/United_States) and the [Australian Capital Territory](https://en.wikipedia.org/wiki/Australian_Capital_Territory) in [Australia](https://en.wikipedia.org/wiki/Australia). Legality varies in these countries and subnational jurisdictions when it comes to commercial sale. A policy of limited enforcement has also been adopted in many countries, in particular the [Netherlands](https://en.wikipedia.org/wiki/Netherlands) where the sale of cannabis is tolerated at [licensed coffeeshops](https://en.wikipedia.org/wiki/Coffeeshop_(Netherlands)).

**Health benefits of cannabis leaves**

Besides adding a flavorful twist to your recipes, cannabis leaves carry many potential health benefits. Cannabis is a plant and as such contains essential nutrients and antioxidant properties as any other leafy green would.

## Get Data

The dataset used for the assignment is from the Cannabis Strains Marijuana Dataset from LiamLarsen in Kaggle. It contains 2,350 strains of marijuana. Each strain has a combination of effects and flavors (among other features) that are used in the app. There are 16 unique effects and 50 unique flavors:

<https://raw.githubusercontent.com/JimKing100/strains-live/master/data/cannabis.csv>

Acknowledgements, where the info was taken: www.strainofweed.com <https://cannasos.com/strain>

## Data Wrangling

The marijuana strains data came from the Cannabis Strains Marijuana Dataset from LiamLarsen in Kaggle. The data is loaded into a dataframe, the strains with nans are dropped (a small number <70), and the Effects and Flavor features combined into a Criteria feature.

I selected the file that is in .csv format and always up-to-date, to import it in Python and turn it into a dataframe.

Strains: describe the lineage of a variety, a major aromatic component, or a possible effect. As a result, strains are often marketed based on the characteristics that their names advertise. There are either pure or hybrid varieties of the plant genus Cannabis, which encompasses the species C. sativa, C. indica, and C. ruderalis. Varieties are developed to intensify specific characteristics of the plant, or to differentiate the strain for the purposes of marketing and make it more effective as a drug.

Type: indica, sativa, hybrid

Rating: user ratings averaged

Effects: 16 unique different effects

Taste: taste of smoke

Flavor: 50 unique flavors.

## Data Mining / Clean Data

The delivery is divided in folders: documentation, resources and rsc.

In the documentation, I designed a PowerPoint presentation and the guidelines for this work.

In the ‘resources’ folder, I’ve included the Jupyter file where I started working and trying to clean the data.

In the ‘rsc’ folder, I kept the ‘api’ folder and ‘utils’ along with the jupyter file called main. As part of the activity, I created a flask server file, to build an API to enable access from another device and get the necessary information. All that coding is in the ‘api’ folder, and that’s how my .json file can be available and accessible.

## Others.

Hence, I converted the file containing the rating of cannabis into a .json. I made the dictionary (.json) available for anybody on the screen of the API. It can be accessed through my IP, endpoint and enter the token code.

Creating a clean dictionary, the API and the Flask server where the core of the activity.

## Analysis & Conclusions

After reviewing the dataframe and graphs, I can answer some of the questions and come down to some conclusions.

a. Was it possible to demonstrate the hypothesis? Why?

Demonstrating the Hypotesis was not posible because I lacked of information. I cended up centering the analysis in the type of Cannabis leaf, based on the well organized data provided by LiamLarsen on Kaggle.

b. What can you conclude about your data study?

I can conclude that the most used type of cannabis leaf is Sativa. The Zombie-Kush Strain is the best rated by users and the majority of the effects include “Uplifting”.

c. What would you change if you need to do another EDA project?

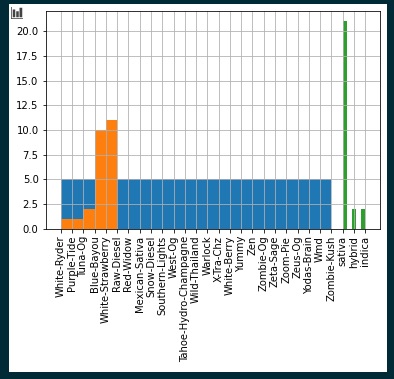
I would use more than one dataframe, gather more numeric data in order to show better graphs and would prepare a better planning (timeline/estimate hours of work) for the stages of the project

d. What did you learn doing this project?

I learnt a lot about the differences between CBD, THC, hemp and Marijuana. The raw cannabis leaf found in nature, the effects and different uses of the product, what is forbidden and what is legal, varieties and breeding processes.

Show the histogram of each column of your dataset with bins =5. How are the ranges painted?

Ranges show the frequency of the data that is present in the dataframe. The ranking is from 0 to 5 and it’s represented in blue. The Strains that are mostly found are White-Strawberry and Blue-Bayou (in orange). At last, the 3 types of leaf are represented in Green.



# References

<https://www.hempbasics.com/shop/general-hemp-information>

<https://www.kaggle.com/nvisagan/cannabis-strains-features>

<https://jimking100.github.io/2020-03-09-Post-9/>

<https://weedmaps.com/learn/the-plant/indica-sativa-hybrid-read-cannabis-leaf/>

<https://royalcbd.com/cbd-oil-hemp-vs-marijuana/>

<https://en.wikipedia.org/wiki/Legality_of_cannabis#:~:text=Countries%20that%20have%20legalized%20recreational,Australian%20Capital%20Territory%20in%20Australia>.