Coursera Capstone

Capstone Project - The Battle of Neighborhoods

Cannabis shop/services

Problem Statement

Cannabis has been legalized in Canada, Marijuana can be grown at home or bought. Canada is the first G7 country to legalize marijuana for recreational use; on Wednesday, the day after the approval, the companies that produce cannabis have recorded strong earnings on the Toronto stock exchange.

As an entrepreneur that wants to enter into this market in order to exploit the new opportunity, I need to better know the different strains of Cannabis available, in order to fill the right stock and to expand my culture on this subject. For this purpose, I'll use a database of all the existing strains together with their characteristics; with this dataset, besides performing some descriptive statistics analysis, I'll use machine learning techniques in order to try to forecast the strain type in function of their characteristics, this will show if the features are significantly different among the strains and can be used for predictions.

Moreover connecting with Foursquare API I'll download data in order to analyze the Cannabis market in Toronto and see if there is space to open a new activity.

This project can be useful also to cannabis users that want to expand their knowledge on the different strains and have to select a shop near their current place.

Target audience: -
Potential entrepreneurs that want to enter into the Cannabis Market and want to expand their Knowledge of the subject and see market opportunity in a certain geographical area, or Cannabis users looking for new strains to smoke and the nearest shops/services.

Stakeholders: -
1. Cannabis Consumers
2. Potential Cannabis entrepreneurs

Data Section

Following sources are used in order to develop the project.

- Data Item: dataset containing information regarding all the different strains of Cannabis plants.
- Format: CSV
- Description: strain name, type of strain, rating (from 1 to 5), Effects (list of adjectives), Flavour (list of adjectives), general description.
- This dataset was downloaded from: https://github.com/kushyapp/cannabis-dataset/tree/master/Dataset/Strains

- Data item: Foursquare Data regarding all activities resulting from a search in the Toronto area using the words Cannabis or Marijuana.
- Format: json
- Description: name, categories, address, city, country, postal code

Methodology

The cannabis strain dataset contains strains data.

Operation Performed

- Open the file in a Dataframe
- Transform list columns in multiple columns, each containing only one adjective, perfomed for Flavour and Effects
- View descriptive statistics
- Transform Flavour and effects columns into dummy variables
- Partition the dataset in Train and Test subsamples
- Perform classification algorithms, as Classification Tree, Boosting Gradient and Neural Netwok
- Download Data from Foursquare of all the activities resulting from a search in the Toronto area using the words Cannabis or Marijuana.
- Join the 2 dataset for Cannabis and Marijuana in a unique dataset, eliminating duplicates.
- List and evaluate the results

Results
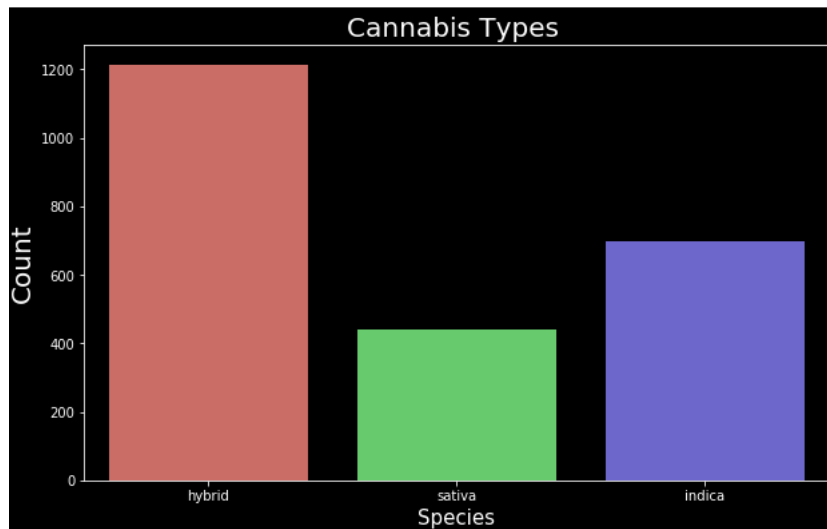

```
Distribuition Type
Type
hybrid     1212
indica      699
sativa      440

Percentage of distribuition Type
Type
hybrid     51.55
indica     29.73
sativa     18.72
```
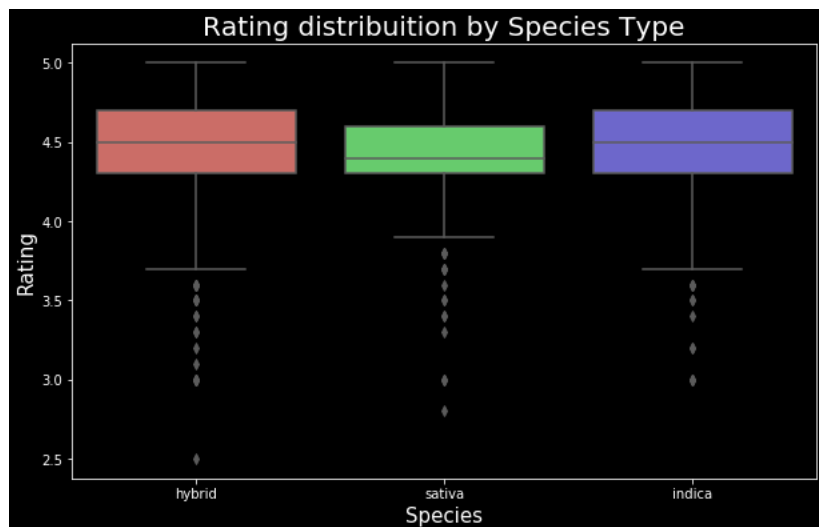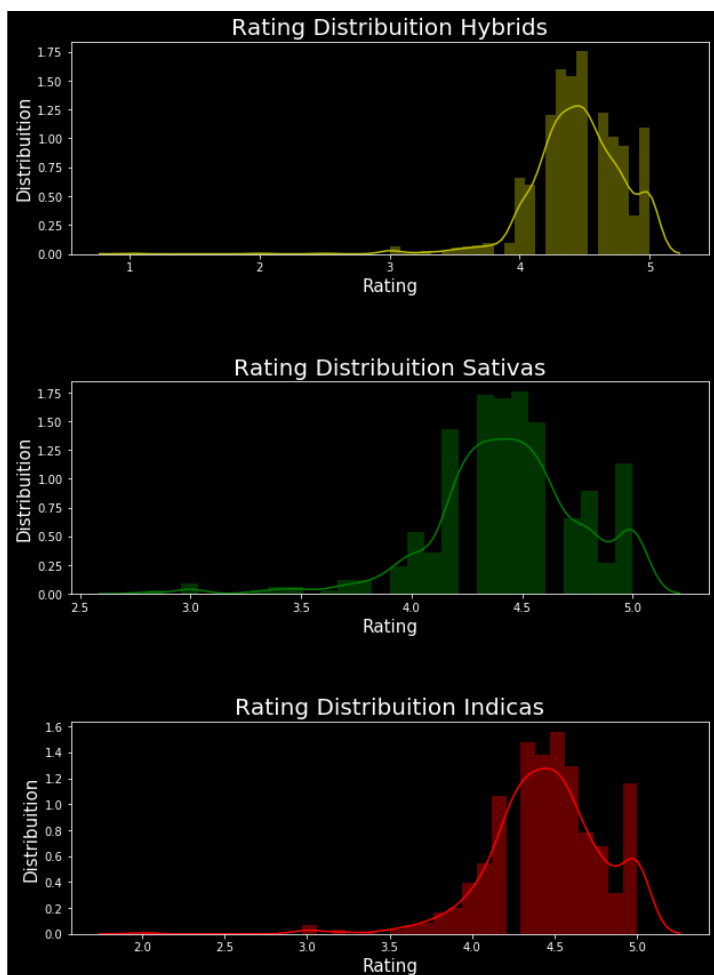
The most diffused Cannabis specie is Hybrid, followed by Indica and Sativa

Looking at the rating distributions we see that Hybrid and Indica are better rated than Sativas.

```
The top 5 First Effects related
Relaxed     825
Happy       476
Euphoric    249
Uplifted    244
Sleepy       89
Name: Effect_1, dtype: int64
```


First Effect Related x Rating by Species Type

We see that Uplifted, Euphoric and Creative receive are associated with best grades in all 3 types.
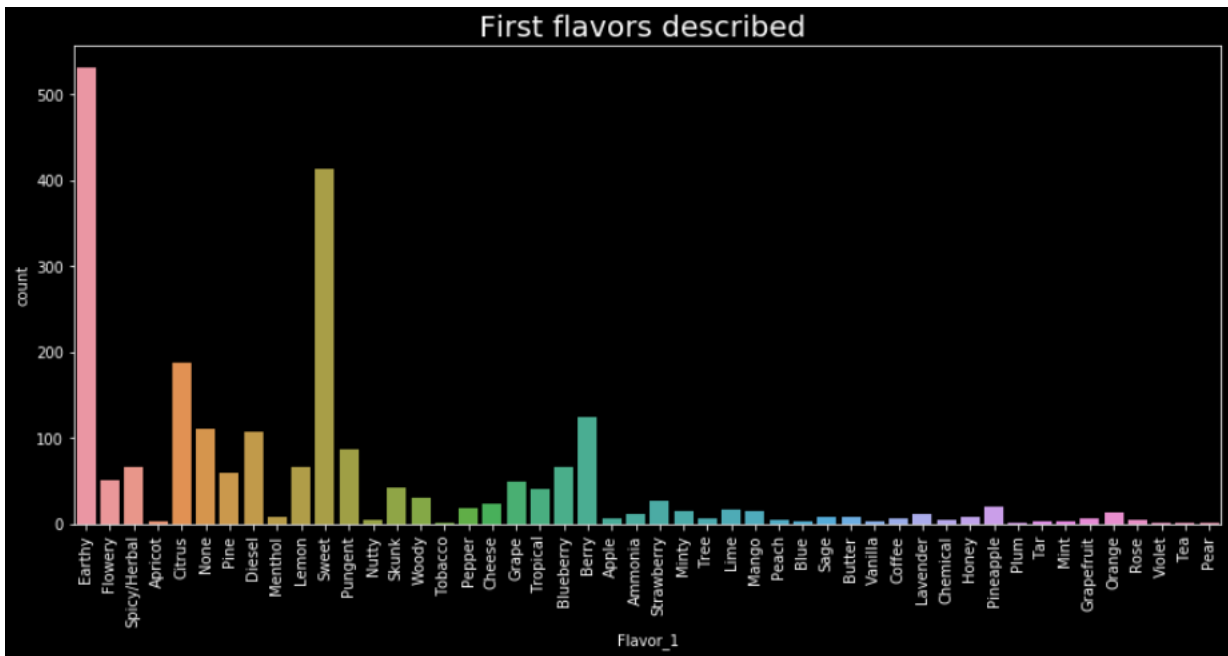
Effects Distribution by Cannabis type:

Sativa

- 'Happy': 342
- 'Uplifted': 328
- 'Euphoric': 276
- 'Energetic': 268

Indicas

- 'Relaxed': 628
- 'Happy': 562
- 'Euphoric': 516

Hybrids

- 'Happy': 967
- 'Relaxed': 896
- 'Uplifted': 848
- 'Euphoric': 843

Earthy and sweet the most selected in all the types.

We try to predict the Cannabis types using the feature
Rating, effects and flavor.
We do this with Decision Tree, Gradient boosting and Neural Network.

Classifier	Accuracy score
Decision Tree	0.51
Gradient boosting	0.63
Neural Network	0.56

Gradient boosting best classifier.

Looking at the activities Cannabis related already opened in Toronto retrieved from Foursquare, we have:

	name	categories
0	Toronto Medical Cannabis Prescriptions	Business Service
1	Cannabis & Coffee	Coffee Shop
2	Apollo Cannabis Clinic	Doctor's Office
3	Cannabis Culture	Drugstore
4	Canuck Cannabis Club	Drugstore
5	SESH Cannabis	Health & Beauty Service
6	Cannabis Lawyers Toronto Canada - Harrison Jor...	Lawyer
7	Allevi8	Marijuana Dispensary
8	Apollo Medical Marijuana Clinic	Medical Center
9	Bodystream Medical Marijuana	Medical Center
10	Bodystream Medical Marijuana Services	Medical Center
11	Grow Legally Marijuana Clinic and Consulting	Medical Center
12	GrowWise Health Medical Cannabis Education	Medical Center
13	Marijuana Access Canada	Medical Center
14	White Cedar Medical Cannabis Doctors	Medical Center
15	Friendly Stranger - Cannabis Culture Shop	Smoke Shop
16	Cannabis Magazine	Tech Startup

We have mainly:

Lawyer firms: 1
Startup Magazine: 1
Cannabis Shops: 6
Medical : 9

Discussion

Looking at the strains dataset we have seen that the most diffused Cannabis specie is Hybrid, followed by Indica and Sativa and the rating distributions indicate that Hybrid and Indica are better rated than Sativas. We have seen that Uplifted, Euphoric and Creative receive are associated with best grades in all 3 types.

Trying to predict the Cannabis types using the feature Rating, effects and flavor we see that the best classifier is Gradient boosting.

In Toronto's Area, in few months, some business activities already were started up following the new Cannabis market, mainly medical or consumer shops but also a Lawyer firm and a startup magazine.

Conclusion

The Cannabis Market is still at the initial stage in the Toronto area, there are many spaces to open a classical reseller activity or also something more naïve following the new market opportunities.
The Variety of strains permits to diversify and satisfy nearly all the customer tastes or desire to experiment something new.