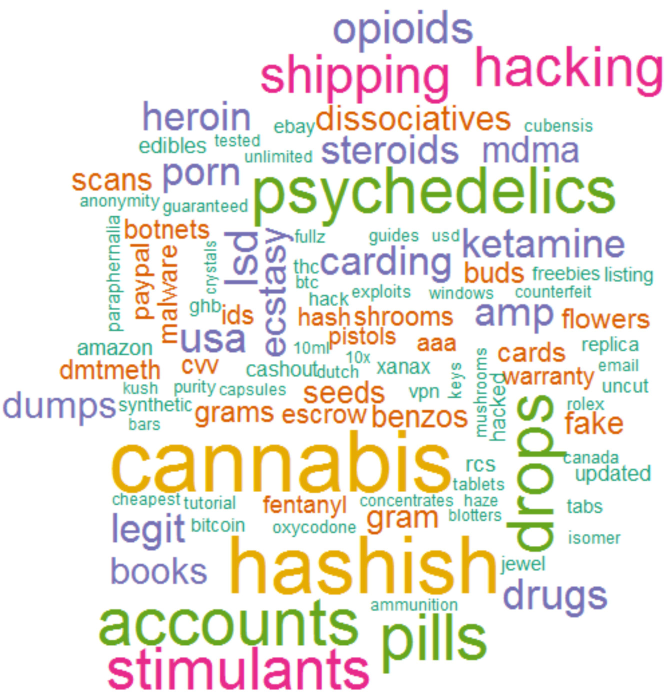
Code ▾

Data Mining - Dark Web Market

Simon Delecourt & Edouard Donze



1 - Introduction

For the past few years, a lot of *web market* has been developed and you can now easily buy items of any kinds. However, plenty of those marketplaces are growing up on *Dark Web*. Even if government attempts to fight against this kind of illegal market, new ones resurface or re-migrate frequently. On top of that, currently, we don't really know in detail how these websites operate.For more information on *Deep web* and *Dark Web marketplaces* please read the article: "Mining the Dark Web" [1]. You will also find in this paper an analysis of *Agora Market*.

This paper presents a research carried out on one of the largest marketplace (specially for drugs) on Internet, the ***AlphaBay Dark Web Market***. This web Market has caugth the attention of governemental agencies since two teenagers aged of 13 and 18 died after overdosing on a powerful synthetic opioid. Therefore, during our research, AlphaBay and Hansa, another dark web market, have been shut down on July, 2017 as a part of a law enforcement operation by the Federal Bureau of Investigation, the Drugs Enforcement Administration and European law enforcement agencies acting through Europol.[2] [3]

According to US Attorney General Jeff Sessions the aim of this action was to caution criminals from thinking that they could evade prosecution by using the dark web. However, it is widely believed that other web markets will take the place of *AlphaBay*. By the way, the popularity of AlphaBay can be explained by the shut down of *Silk Road 2.0* on 2013 since it has been launched on september 2014.

The understanding of such illegal market is crucial to fight it. Information gathered in those websites, allow to identify which are the most wanted ads for the consumer and where they come from. Therefore it might be possible to detect the footprint of each seller and, thus, help governmental agencies to identify recurrent sellers with various hidden identities.

Thus AlphaBay market will be analysed. Its nature, its different countries of origin, its main sellers, its predominance of items and so on will be investigated. During a first phase “Basic Statistics” will be carried out on the *Database*, in order to discover the marketplace and to point out its trends. Then, experimental results of *data mining techniques* will be discussed.

Code

Code

2 - Technical Implementation

2.1 - Code Repository

All the code is publicly available in the Github project *“Data Mining - Dark Web Market”*. The repository is accessible from the following link: https://github.com/SimonDele/Data-Ming-Dark-Web-Market (https://github.com/SimonDele/Data-Ming-Dark-Web-Market)

2.2 - Chosen Technologies

There is a number of technologies and programming languages that can be used for *Data Analysis*. Here are the ones that have been used for this project.

2.2.1 - R / RStudio

The 2 main progamming languages for this kind of research are *Python and R*. Python is the most used. But since R has obtained more attention for the last few years, that is the one that will be used here. Moreover R, which is a language dedicated for statistical and graphical presentation, will suit our needs.

Furthermore, R has a powerful dedicated environment (IDE) which is *RStudio* and a lot of *open-source libraries*. Here is a list of the main libraries used for this project.

stringr : String manipulation library.

units : Unit library including solution for convertion.

rpart : Package that contains a wide library for decision tree method.

arules : Used for association rules.

e1071 : Bayesian Naive implementation library.

bnlearn : library including solution for bayesian network creation and visualisation.

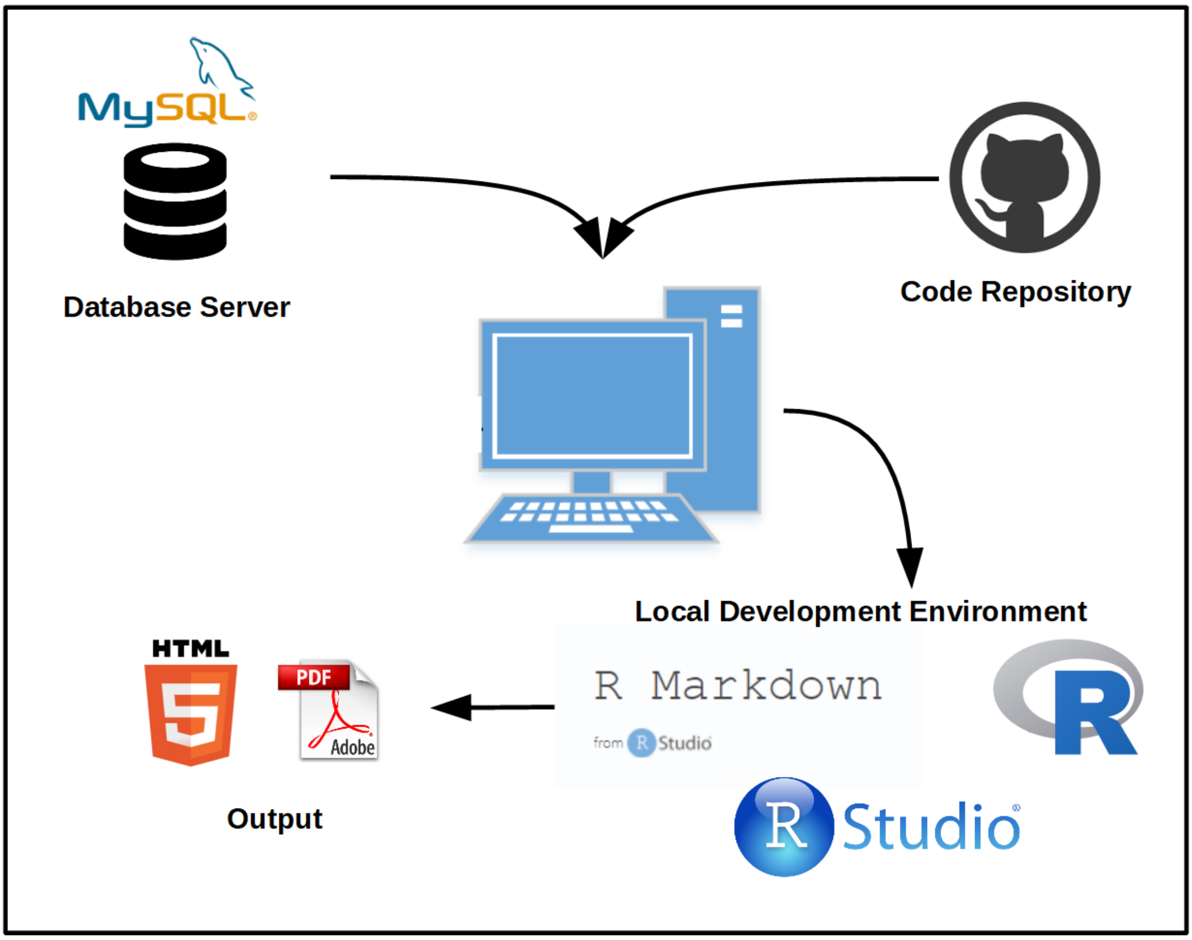
You can find the whole list of packages used in the *GitHub* repository.

2.2.2 - R Markdown Notebooks

This document has been produced by *R Notebooks*. R Notebooks is an *R Markdown* document with chunks that can be executed independently and interactively. It enables to produce the document in various formats such as *HTML, PDF and Word*. In addition, it can be used directly into the RStudio IDE.

2.3 - Detailed Diagram

This is the representation of the technical implementation taking place during this project:




```
Figure 1 - Technical implementation
```

3 - Data Retrieval

Thanks to Sin Wee Lee and Andres Baravalle data have been collected on the *AlphaBay Dark Web Market*. Each row of the data represents an ad and all the significant information can be find in the different columns : title, description, price (in USD), url link, seller, payment, origin, destination, category, timestamp that is to say the date when the ad was collected, creation date of the ad, number of product sold since this date and a link to the image.

The Data represents approximately *1/10* of the Web Market, but gives a pretty good representation since the uploaded ads were fairly distributed.

Thus, the first step was to clean the data (remove special characters, switch in lowercase …) and makes it readable in a computer way. That it's to say, to find in the title or description of the ads the amount (number and mass) of the product they are selling. Indeed, at the beginning these information were not given distinctly. Therefore an important work has been done on it in order to make analyses easier. With these information, the price of one unit of one dose (1 gram) of the product has been calculated and added.

Here is what the *Database* looks like :

Code

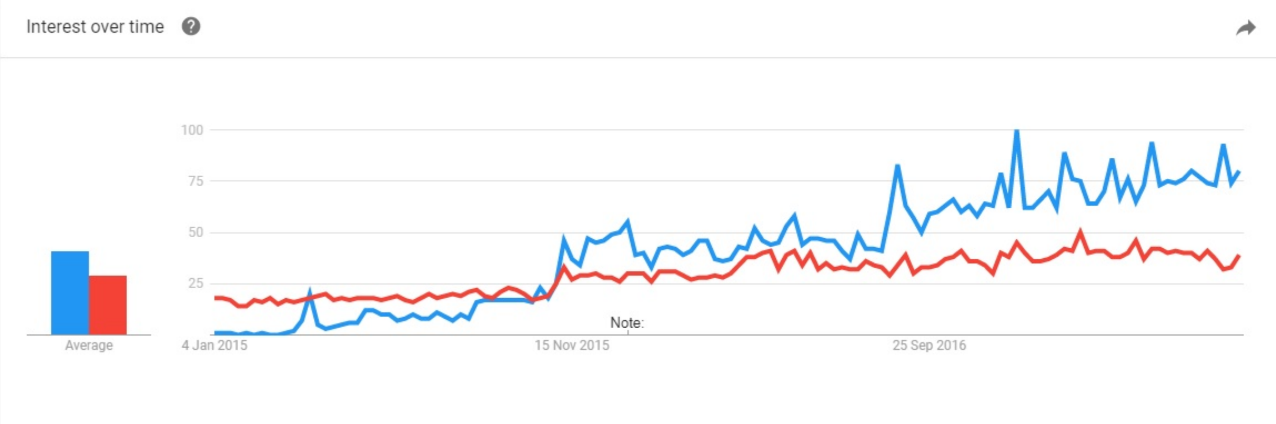
	id <int>	title <chr>
19409	226445	ritalin 10 mg 90 tablets brand name from novartis
19410	226448	tramadol 100mg 200 tablets
19411	226449	panam::fe 1g of amsterdam's ketamine **shipped from usa**

3 rows | 1-3 of 20 columns

```
Table 1 - Database Sample
```

4 - AlphaBay Market

As it has been said in the Introduction, *AlphaBay*, due to its popularity, drew the governmental attentions. As a matter of fact its reputation can be reflected by looking at *Google statistics* [4].

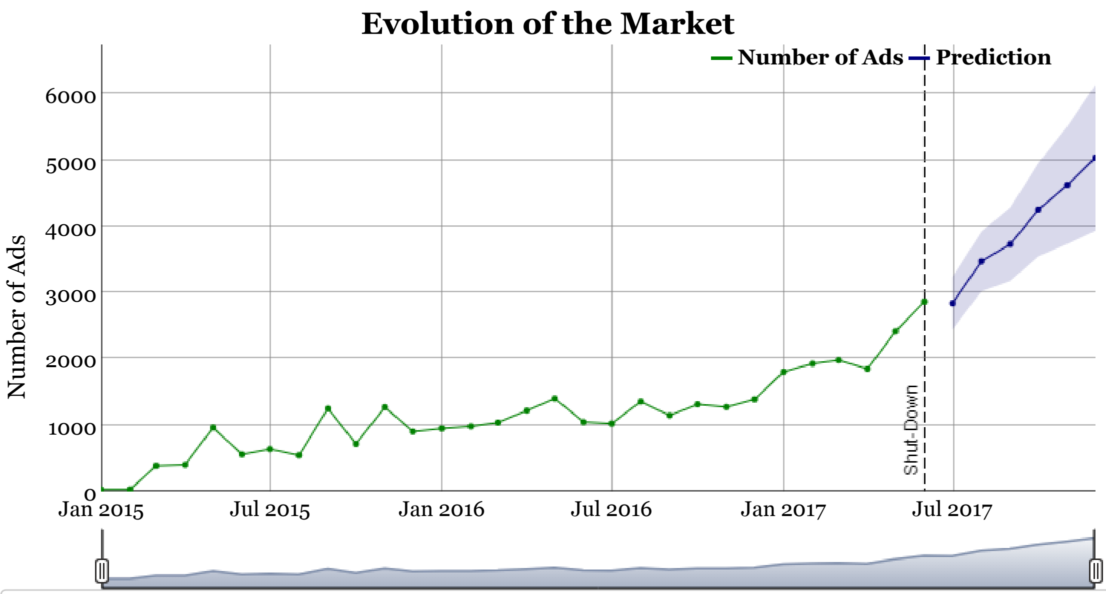



```
Figure 2 - Evolution of AlphaBay and Dream Market Google researches
```

On this graph, *AlphaBay* is in blue and *Dream Market* is in red, which is an other Dark Web Market still operating. This is showing the evolution of *Google* researches that are related to these two Web Markets from 2015 until June 2017. It should be noted that *AlphaBay* has become more and more popular for the last tree years, and that just before being shut down by the authority in july 2017, it was one of the most popular marketplaces.

Let's now try to look at the evolution of the market with the collected data on *AlphaBay*. Here you can see the number of ads posted per month from *January 2015* until *June 2017*.

Code




```
Figure 3 - Evolution of AlphaBay Web Market
```

The overall appearance and the growing popularity can be again pointed out with this graph. Between 2015 and 2016, there was a significant jump, the amount of ads rose from 7712 up to 14161. Nevertheless the most surprising thing is that the number of ads that have been posted during the six first months of 2017 (before the closing) is 12878 which is almost the same that in the whole 2016.

In order to see how the market would have looked like in the end of 2017 a prediction also has been added on this graph. Therefore, according to prognoses, the amount of ads would have reached a pick of 5,000 ads by the end of year.

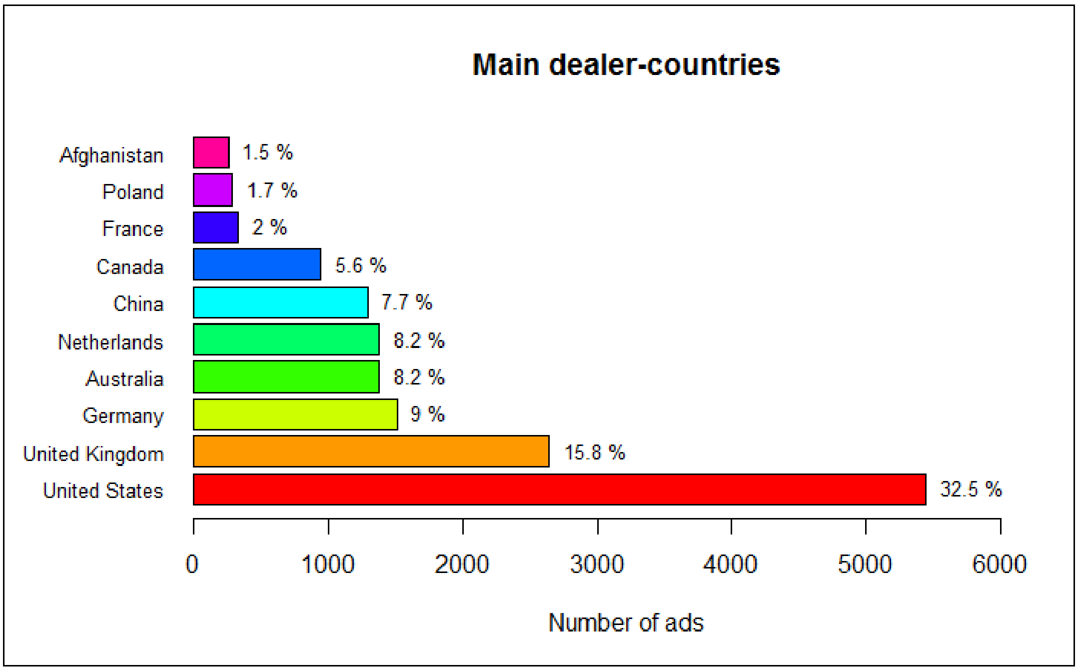
5 - Basic Statistics

5.1 - General distribution

As it has been said before, basic statistics have been first realized. Let's see the general distribution and trend of the market.

1. Global view of ads distribution

Code

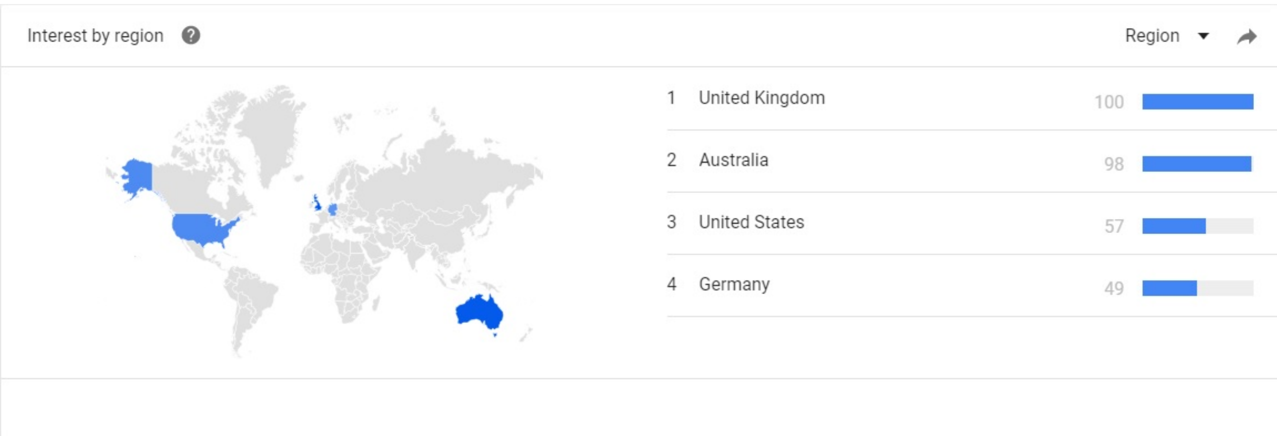



```
Figure 4 - Main Dealer Countries
```

This bar-chart represents the 10 main countries in the world regarding the number of ads. As we can see, *United States* are the biggest dealer far ahead of the rest. Their number of ads is more than twice as the number of the second one, which is *United Kingdom*.

Moreover, it is noticeable that most of these countries are economically powerful. For instance on these ten main countries, five belong to the Group of Seven (G7), only *Japan* and *Italy* are not present. And other ones are also located in powerful areas where a lot of trade are made with other countries.

Furthermore an interesting thing to point out is that the first four countries are exactly the one where the word *"AlphaBay"* is the most researched on *Google* [5].

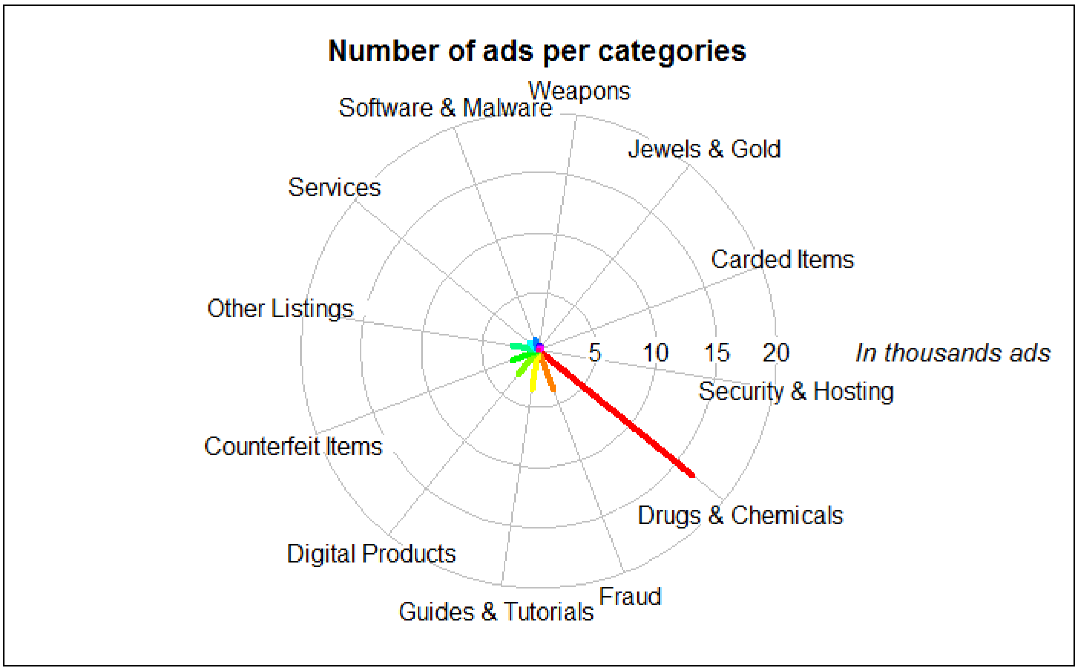



```
Figure 5 - AlphaBay world Google researches
```

2. Now let's have a look at the distribution of ads per category

Code

Code




```
Figure 6 - Distribution of the Market
```

Code

To begin with, it is observable that there are 12 main categories in this web marketplace. It appears that *"Drugs and Chemicals"* group is the largest one. By the way, it represents 45.64 % of the global market.

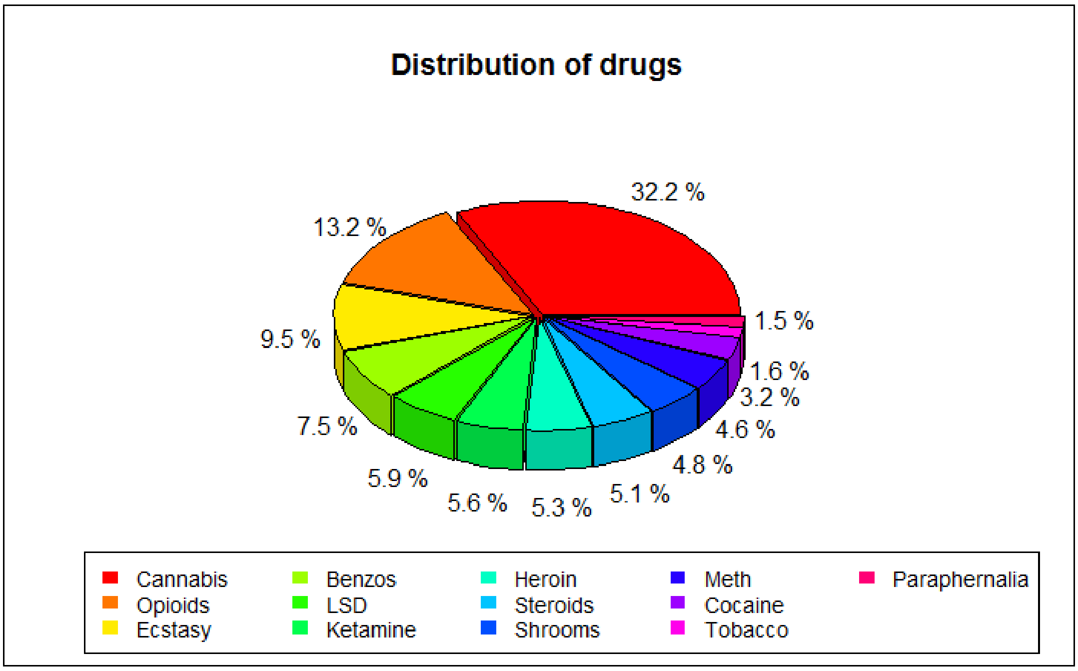
It is also worth noting that the second most popular category is *"Fraud"*, that is to say all the ads regarding impersonation, deception papers and accounts. It represents 13.5 % of the market. Eventually, all other items (digital product, weapons, jewelry …) represent a small rate of the marketplace.

5.2 - Drug Market

AlphaBay Web Market is a well known place for dealing drugs, this last chart has proved that. Thus, let's focus on the drug market.

1. Distribution of drugs

Code

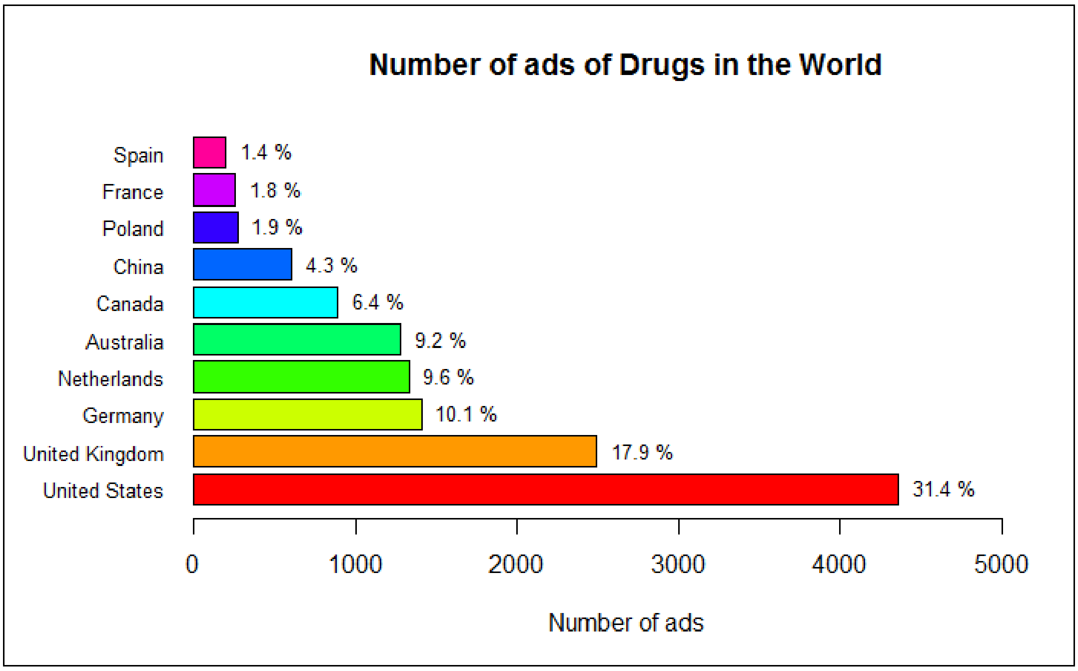



```
Figure 7 - Drug Distribution
```

A large range of drugs categories can be discerned. Nevertheless, *Cannabis*, *Opioids* and *Ecstasy* cover *more than 50 %* of the market. This is not surprising since Cannabis and Opioids can be *easily found* in some countries. Thus, it is easy for them to sell these drugs for the remaining countries where they are very requested. As a matter of fact, Cannabis is *legalised* in some countries and some Opioids can be obtained thanks to *prescription* by a doctor. The massive presence of this kind of drugs on the market raises the issue of how some of them are prescripted. The *remaining* of the market is splited by all other drugs.

2. World distribution of drugs

Code




```
Figure 8 - Main drug dealer countries
```

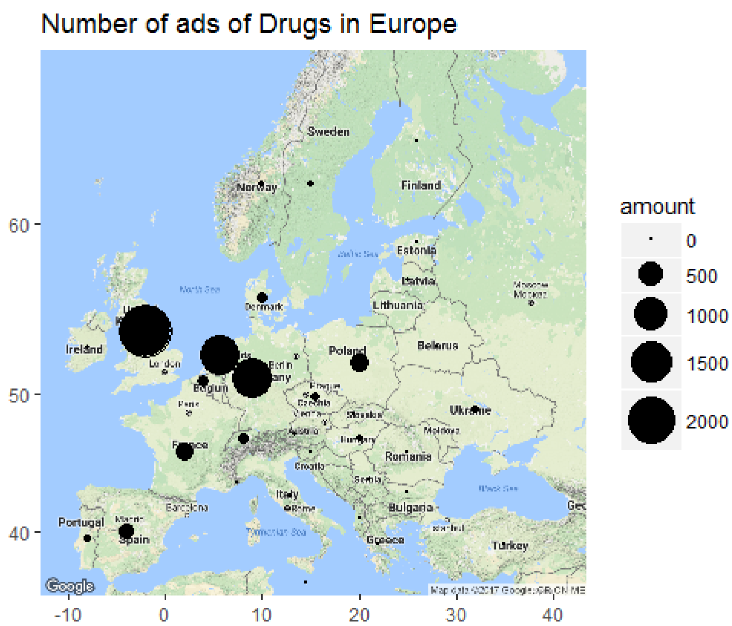
Once more, the 10 main dealer-countries in the world are plotted but this time regarding the number of drugs ads. At first sight, the chart looks like the first one. This is coherent, indeed, by comparing the ratio between drugs ads and the total number of ads, it is intelligible that they are mainly dealing drugs.

This also matches with the second chart that shows that drugs are the main item in the market. There are few exceptions such as *Afghanistan* which has been replaced by *Spain* and *Canada* has been reversed with *China*.

But so far it is possible to conclude that the market of drugs is gathered in Europe and the north of America.

3. Take a global view of the drugs market in Europe with the following map

Code



```
Figure 9 - European Drug Dealers
```

Circles show the amount of ads concerning drugs. The map confirms previous assumptions that there are a lot of Drug dealers in Europe with, as major dealer countries, *United Kingdom*, *Netherlands* and *Germany*.

It appears that the principal dealer-countries are located on the Atlantic Coast and own huge harbours where there is important merchant shipping. Whereas on the East part there are not a lot of activities. This is probably due to the fact that dealers are using *international commercial maritime traffics* in order to dispatch their drugs all around the world. Maritime transport is an option increasingly used since it allows them to carry large quantities at one time. Drugs can be transported in small and fast boats (Go-Fast-Boat between countries border) or in containers on commercial vessels. Thus, significant seaports in Europe such as Rotterdam in Netherlands or Antwerp in Belgium are key points for this type of trafficking. In 2014 “Dutch police estimated that 25-50 % of the cocaine reaching Europe now enters via the port, which handles around 11 million containers a year.” [6]

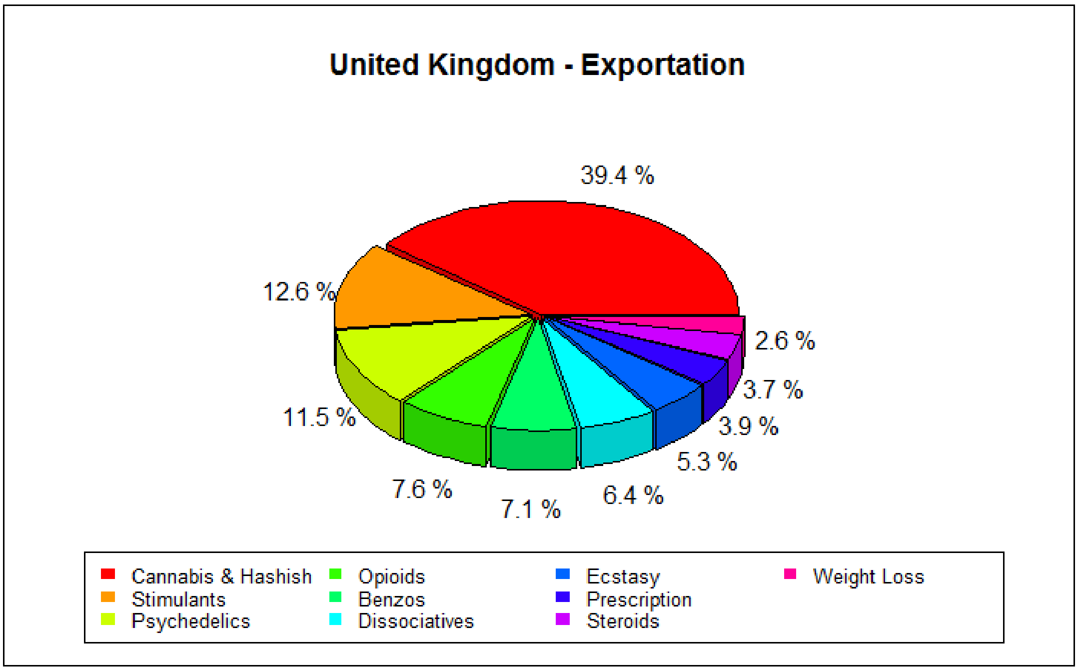
5.3 - Product Flow By Country

Let’s now focus more specifically on different countries and study their trend. To do so, export and import flows of the country have been investigated.

1. United Kingdom exportation

The pie below represents the repartition of each category that United Kingdom exports. Only the second subcategory has been kept because it appears to be the most relevant since the first one only gives information on the nature of the ad (for instance *“Drugs & Chemicals”*).

Code



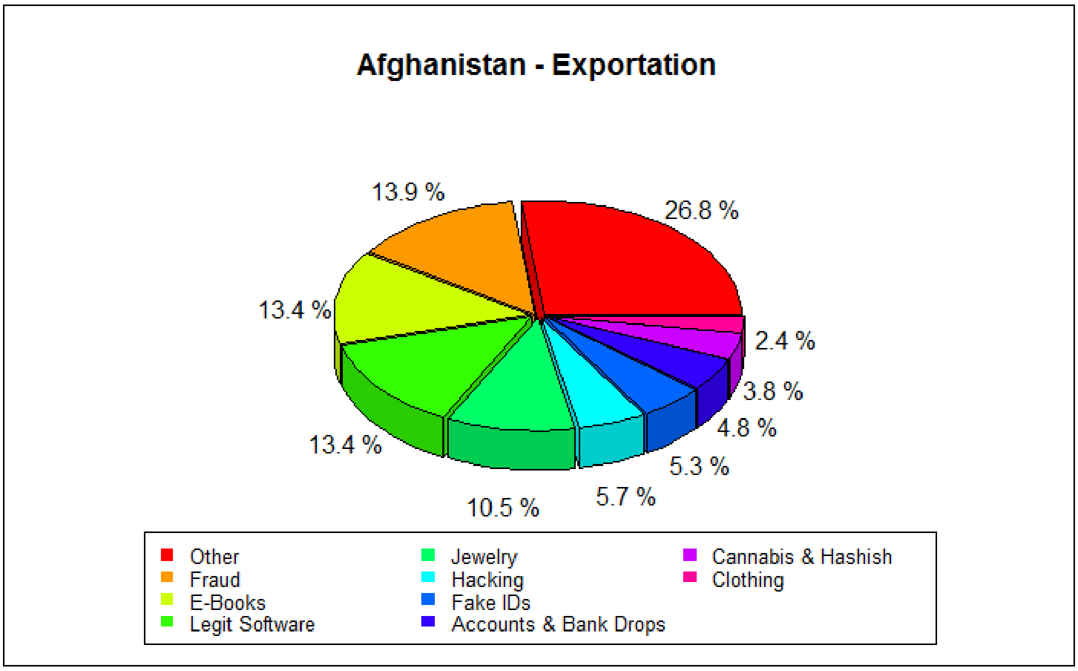

```
Figure 10 - United Kingdom exportation
```

Given that most of exported items are drugs, that is not surprising that they are the most sold product, as it has been seen before. Once again this pie chart shows the market diversity. Although a huge part concerns *“Cannabis & Hashish”* category, *“Stimulants”* and other highly dangerous drugs are significantly present as we.

Most of European countries follows the same rules as *United Kingdom* and this confirms previous assumptions.

2. Let’s have a look at the exportations of Afghanistan which seem different to United Kingdom

Code

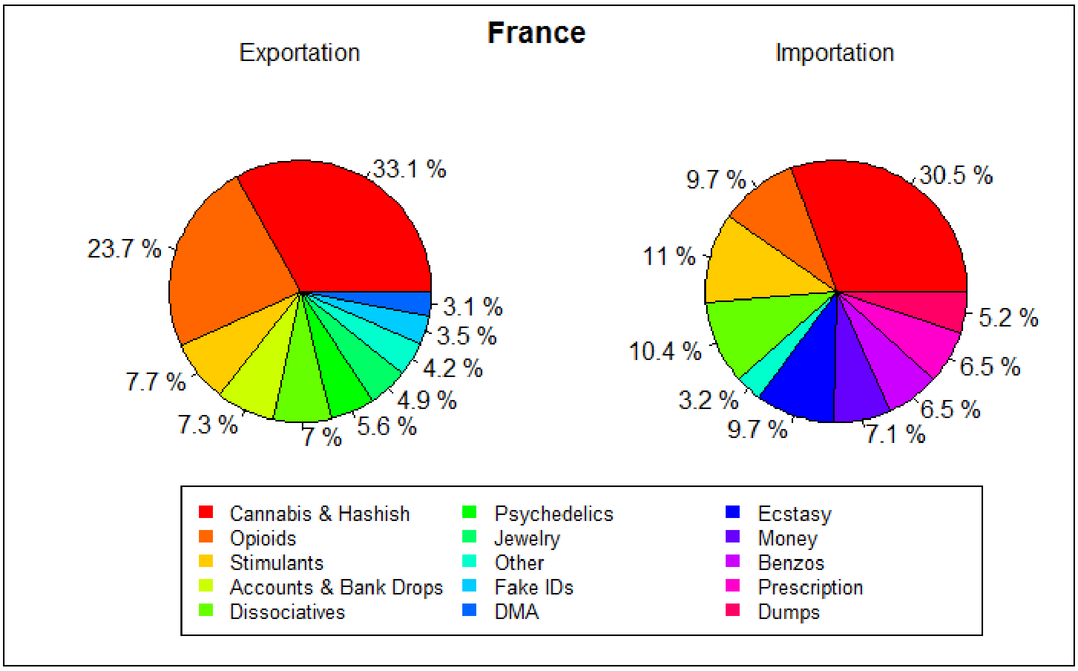



```
Figure 11 - Afghanistan exportation
```

What is suprising is that, unlike most of countries, *Afghanistan* doesn't really retail drugs on *AlphaBay Market*. Actually, a vast majority of exported products are false identity, deception account… *Afghanistan* is also dealing electronic devices or softwares.

3. Let's compare France export & import flows and see if there is a difference between them

Code




```
Figure 12 - France exportation and importation
```

It is noticeable that both pies are different. The percentages of each category are not equal and some of them don't appear systematicaly in the other pie chart. However it is obvious that significant exported drugs are also imported. Morevover, *France* is importing some drugs that are not local.

Nevertheless these conclusions should be moderated since targeting one particular country reduces significantly the number of information used for statistics.

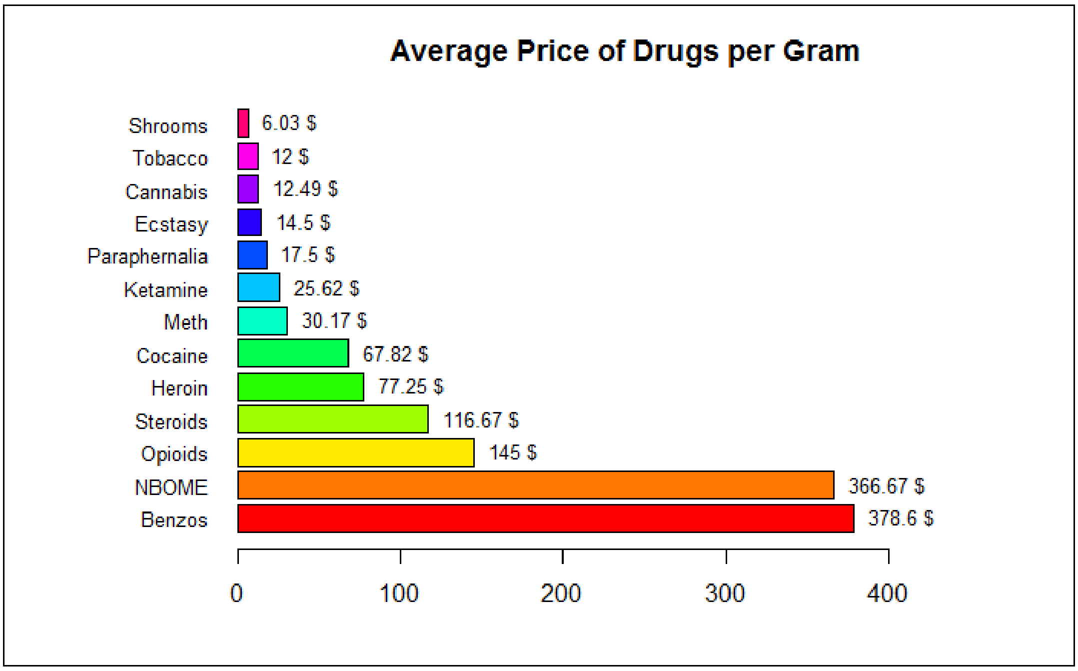
5.4 - Market Prices

After analysing general trend and flows, one interesting topic to analyse is market prices. One may ask if sold products in *AlphaBay* are cheaper than in the streets.

1. Average prices on the *AlphaBay Web Market*

Firstly, the average price of one gram of the most common drugs has been calculated and results below has been obtained.

Code

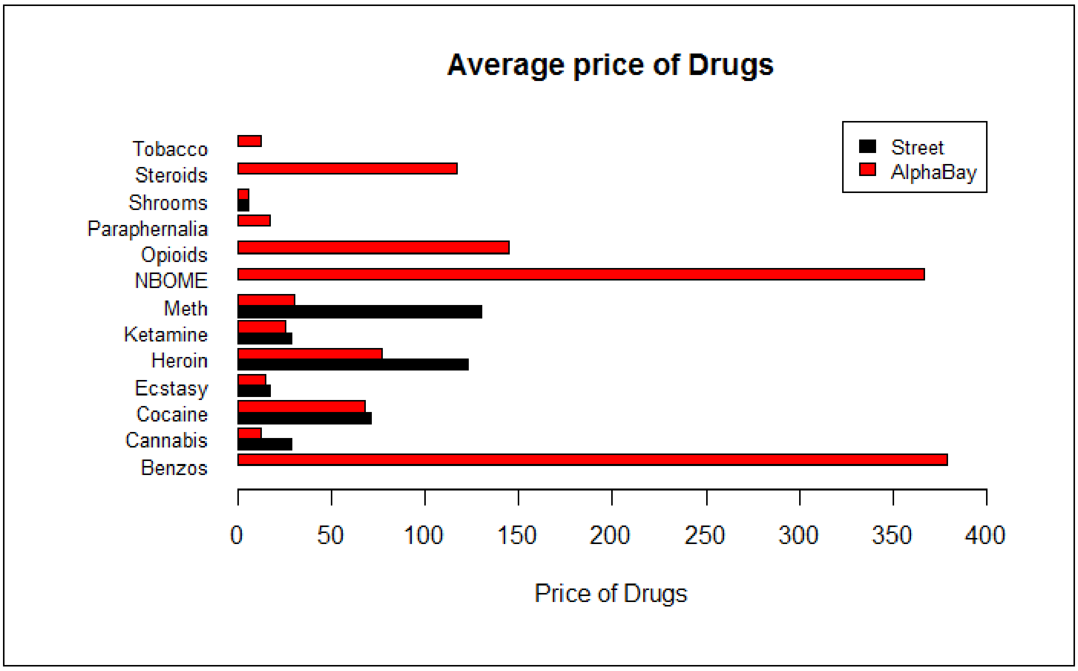



```
Figure 13 - AlphaBay Drug Prices
```

2. Comparison with the "street"

Finally, some information about prices of street sellers have been collected, for same drugs as above. Thus, prices in articles and websites have been gathered. Unfortunately, values on some drugs are missing. Below are results of these investigations.

Code




```
Figure 14 - Comparison of Prices between AlphaBay and Street
```

Globally, it appears that prices of *street sellers* are often largely higher than *AlphaBay* ads. In few cases both prices tend to be similar.(Please find in the last section all references used : [7], [8], [9], [10], [11])

6 - Data mining

Secondly, data mining techniques have been performed in order to discover *hidden rules* and *correlations* in the database. Another goal is to predict the value of one variable given other values.
Please note that for all this part the analysis is only on *Drug Market*.

6.1 - Sellers Predictions

The first thing to wonder is how to guess the seller of an ad. To answer this question, different data mining methods have been used, especially *Decision Tree* and *Bayesian classification*.

Algorithm has been run on a subset of the database with by rows ads and by columns the origin, category, seller and price. The aim is to predict who is selling each ads. By training the algorithm on one half of the data, predictions could be made on other half. Given that most of sellers own just few ads (occasional advertisements) only the main ones were selected, which represent at best the market. Otherwise, data mining techniques will fail in finding rules for them.

To check efficiency of the algorithm a measure of accuracy must be calculated. It is obtain by comparing the prediction of decision tree method with the true value.

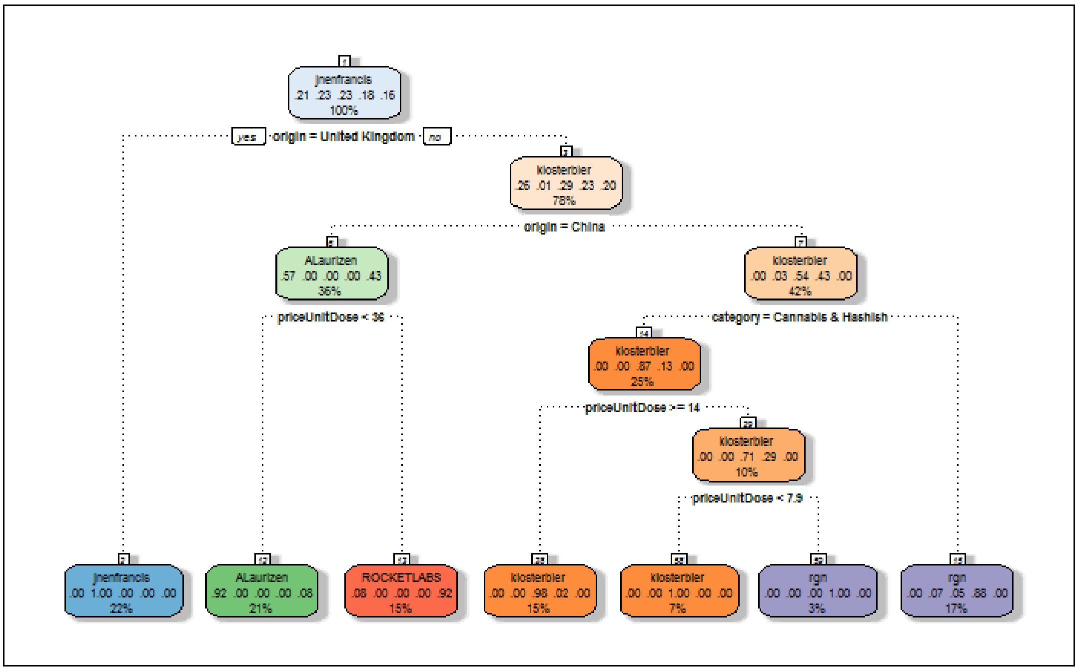
1. Decision tree

Using *rpart package*, which is based on the *CART Alorithm*, a decision tree has been created. Thanks to it, predictions of the seller could be made. Prognoses on the five most significant sellers and the related tree can be found below.

Code

```
            pred
             ALaurizen jnenfrancis klosterbier rgn ROCKETLABS
  ALaurizen         53           0           0   0          4
  jnenfrancis        0          74           0   2          0
  klosterbier        0           0          69   4          0
  rgn                0           0           0  73          0
  ROCKETLABS         8           0           0   0         62

       Table 2 - Sellers Prediction / Decision Tree method
```




```
Figure 15 - Seller prediction / Decision Tree
```

```
[1] "The accuracy is : 94.84 %"
```

It is striking to realise that predictions are very reliable since the accuracy is very high.

In the *Table 2* you can find by columns our Prognoses and by rows the real sellers. In other words the diagonal shows the number of correct predictions and all other values are mistakes made by the algortihm.

The tree enables to have a good visual aspect on it and gives a lot of essential information on these five main sellers.

Besides, with quick basic statistics, it is possible to validate this tree and its branchs, that is to say how it is splited. Indeed let's have a look at these five dealers. This array contains the main country, main category and median price of their ads.

Code

seller <fctr>	origin <chr>	category <chr>	price <dbl>
jnenfrancis	United Kingdom	Cannabis & Hashish	8.68
klosterbier	Worldwide	Cannabis & Hashish	25.54
rgn	Worldwide	Stimulants	17.75
ALaurizen	China	Stimulants	6.96
ROCKETLABS	China	Stimulants	247.50

5 rows

```
Table 3 - Seller Analyses
```

Obviously, these information are exactly the ones that are in the tree. *jnenfrancis* is guessed thanks to origin *United Kingdom*. For the sellers dealing mainly in *China*, that is to say *ALaurizen* and *ROCKETLABS*, the spliting is made regarding the price (*ALaurizen* is obviously cheaper). For other sellers the algorithm use the category *Cannabis & Hashish*.

Predictions have been done here on only five sellers in order to have a readable tree (otherwise the size of the tree is too big for being plotted). However, prognoses with more sellers can be made and with still a good accuracy. For instance, with 10 sellers, the results below are obtained.

Code

Sellers <fctr>	Prediction <fctr>	Freq <int>
ALaurizen	ALaurizen	57
empireteam	ALaurizen	0
Fapppylicious	ALaurizen	0
FelixUK	ALaurizen	0
GreenLeafLabs	ALaurizen	0
jnenfrancis	ALaurizen	0
klosterbier	ALaurizen	0
optiman	ALaurizen	0
rgn	ALaurizen	0
ROCKETLABS	ALaurizen	7

1-10 of 100 rows Previous 1 2 3 4 5 6 ... 10 Next

```
Table 4 - Sellers Prediction / Decision Tree method
```

```
[1] "The accuracy is : 84.72 %"
```

Once more, this array presents the number of mistakes and correct predictions.

2. Bayesian classification - Naive alogrithm

Bayesian Classification has been used in the same objective as decision tree, making predictions. Here are results when running *Bayesian Naive Algorithm* with the same data.

Code

Sellers <fctr>	Prediction <fctr>	Freq <int>
jnenfrancis	jnenfrancis	59
klosterbier	jnenfrancis	0
rgn	jnenfrancis	0
ALaurizen	jnenfrancis	0
ROCKETLABS	jnenfrancis	0
empireteam	jnenfrancis	0
GreenLeafLabs	jnenfrancis	0
optiman	jnenfrancis	0
Fapppylicious	jnenfrancis	0
FelixUK	jnenfrancis	2

1-10 of 100 rows Previous **1** 2 3 4 5 6 ... 10 Next

```
Table 5 - Sellers Prediction / Naive Bayesian Classification
```

```
[1] "The accuracy is : 55.01 %"
```

The accuracy is 55.01 % which is not very good comparing to decision tree. However one way to improve the prognosis is to add new variables to the data which could be relevant like the number of ads already sold and the creation date of the ad.

Code

Sellers <fctr>	Prediction <fctr>	Freq <int>
jnenfrancis	jnenfrancis	67
klosterbier	jnenfrancis	0
rgn	jnenfrancis	0
ALaurizen	jnenfrancis	0
ROCKETLABS	jnenfrancis	0
empireteam	jnenfrancis	0
GreenLeafLabs	jnenfrancis	0
optiman	jnenfrancis	0
Fapppylicious	jnenfrancis	0
FelixUK	jnenfrancis	4

1-10 of 100 rows Previous **1** 2 3 4 5 6 ... 10 Next

```
Table 6 - Sellers Prediction / Naive Bayesian Classification
```

```
[1] "The accuracy is : 84.21 %"
```

Results show that the algorithm succeeds in predicting most of the sellers. Therefore, the accuracy is 84.21 %. Which is still a little bit less than with decision tree. However with more sellers (i.e more than 40 sellers for instance), this algorithm tends to be more accurate than the one based on decision tree method.

Later in the section *Prediction of profitability* , it will be discussed how to exploit at best this two values : the *creation date of ads* and the *number of sold products*.

3. Text Mining

One last method, but not least : Text mining. This is a fashion method since with the power of our computer, we are now able to perform wide analyses on *text data*. It can be used for text classification (determinating the topic, the tone, etc…), but here it will be used for predicting who has written the text (i.e sellers).

The main interest comparing to previous methods is that it is only based on the words used by the sellers. Thus, it is possible to identify the seller according to his *writting style* in other type websites, such as Social Network. What previous methods can not do because they are trained with variables specific to AlphaBay. Furthermore, there is no need to explain the interest of identifing in Social Network illegal sellers …

The method consists in training the algortihm with words from the ad description of each seller. Then, it is tested with a part of the data which has not been used for training. The algorithm choosen for this method is *Support Vector Machine* but other alogrithms have been tested such as decision tree and random forest and results are similar. The result below has been obtained with a data containing the ads description from the 50 main sellers.

Code

```
[1] "The accuracy is : 95.53 %"
```

The accuracy is 95.53 % which is very high. For more sellers, (which own enough ads, in order to have some of them in the training dataset) the accuracy decreased but it is still very high. For instance for 150 main sellers it is around 90%.

However, this accuracy has to be taken with precaution. Indeed, most of the ads descriptions from one seller are very similar. The reason is that they create various ads for the same product with different quantities. Thus, ads description are almost the same. That is why, it is not surprising that the accuracy is so high.

Anyway, that show us that the algorithm is working properly and it can be used in order to find hidden identities of one seller in AlphaBay or in other Website and eventually discover his real identity.

6.2 - Clustering of drugs

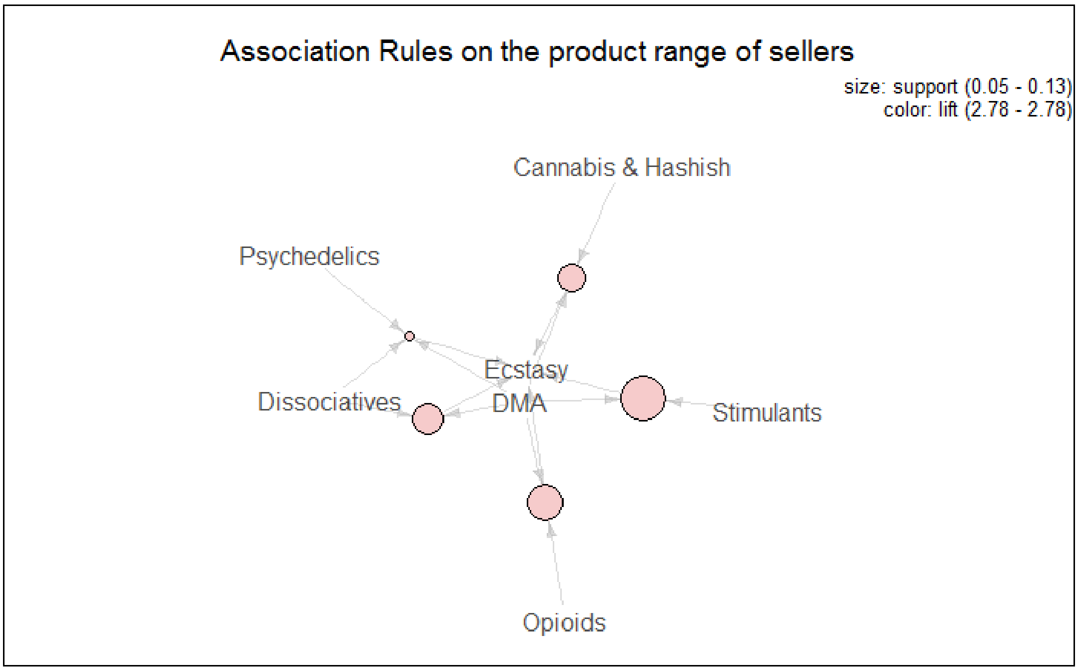
Secondly, one can wonder if there were links between some drugs. That is to say, if this is possible to cluster some drugs.

To do so, firstly, a new data frame has been created with by rows sellers and by columns different sub-categories of *"Drugs & Chemicals"*. In each cell, value is True or False if the dealer has already sold something in this sub-category or not. Then, *Apriori algorithm* of *Association Rules* has been used. One drug has been selected that must be in the itemset. Here it is Ecstasy and results can be seen below.

Code

```
     lhs                     rhs       support confidence lift
[1]  {Cannabis & Hashish,
      DMA }               => {Ecstasy}    0.09       1.00 2.78
[2]  {Dissociatives,
      DMA }               => {Ecstasy}    0.10       1.00 2.78
[3]  {DMA ,
      Opioids}            => {Ecstasy}    0.11       1.00 2.78
[4]  {DMA ,
      Stimulants}         => {Ecstasy}    0.13       1.00 2.78
[5]  {Dissociatives,
      DMA ,
      Psychedelics}       => {Ecstasy}    0.05       1.00 2.78
[6]  {DMA ,
      Psychedelics,
      Stimulants}         => {Ecstasy}    0.06       1.00 2.78
[7]  {Cannabis & Hashish,
      Dissociatives,
      DMA }               => {Ecstasy}    0.05       1.00 2.78
[8]  {Cannabis & Hashish,
      DMA ,
      Opioids}            => {Ecstasy}    0.06       1.00 2.78
[9]  {Cannabis & Hashish,
      DMA ,
      Stimulants}         => {Ecstasy}    0.07       1.00 2.78
[10] {Dissociatives,
      DMA ,
      Opioids}            => {Ecstasy}    0.05       1.00 2.78
[11] {Dissociatives,
      DMA ,
      Stimulants}         => {Ecstasy}    0.09       1.00 2.78
[12] {DMA ,
      Opioids,
      Stimulants}         => {Ecstasy}    0.08       1.00 2.78
[13] {Cannabis & Hashish,
      Dissociatives,
      DMA ,
      Stimulants}         => {Ecstasy}    0.05       1.00 2.78
[14] {Cannabis & Hashish,
      DMA ,
      Opioids,
      Stimulants}         => {Ecstasy}    0.05       1.00 2.78
[15] {Dissociatives,
      DMA ,
      Opioids,
      Stimulants}         => {Ecstasy}    0.05       1.00 2.78
[16] {DMA }               => {Ecstasy}    0.19       0.95 2.64
[17] {DMA ,
      Psychedelics}       => {Ecstasy}    0.11       0.92 2.55
[18] {Dissociatives,
      Psychedelics}       => {Ecstasy}    0.09       0.82 2.27
[19] {Cannabis & Hashish,
      Opioids}            => {Ecstasy}    0.09       0.82 2.27
[20] {Psychedelics,
      Stimulants}         => {Ecstasy}    0.12       0.80 2.22
[21] {Cannabis & Hashish,
      Dissociatives}      => {Ecstasy}    0.08       0.80 2.22
[22] {Cannabis & Hashish,
      Opioids,
      Stimulants}         => {Ecstasy}    0.08       0.80 2.22

             Table 7 - Cluster of drugs / Association Rules
```




```
Figure 16 - Cluster of drugs / Association Rules
```

The algorithm succeeds in finding some rules in the data frame. That means that some drugs can effectively be clustered. The support is between 5% and 15% so it is frequent to have these itemsets. Moreover the confidence is more than 80%. In other words, if there is the itemset on the left we are most likely to have the drug on the right.

These rules can be interpreted as follows : sellers often deal more than one product. And these products can be clustered by type.

6.3 - Prediction of ads origin

After trying to make predictions on sellers (cf. Sellers Predictions), Prognoses on the origin of the ads have been made using Decision Tree method and Association Rules.

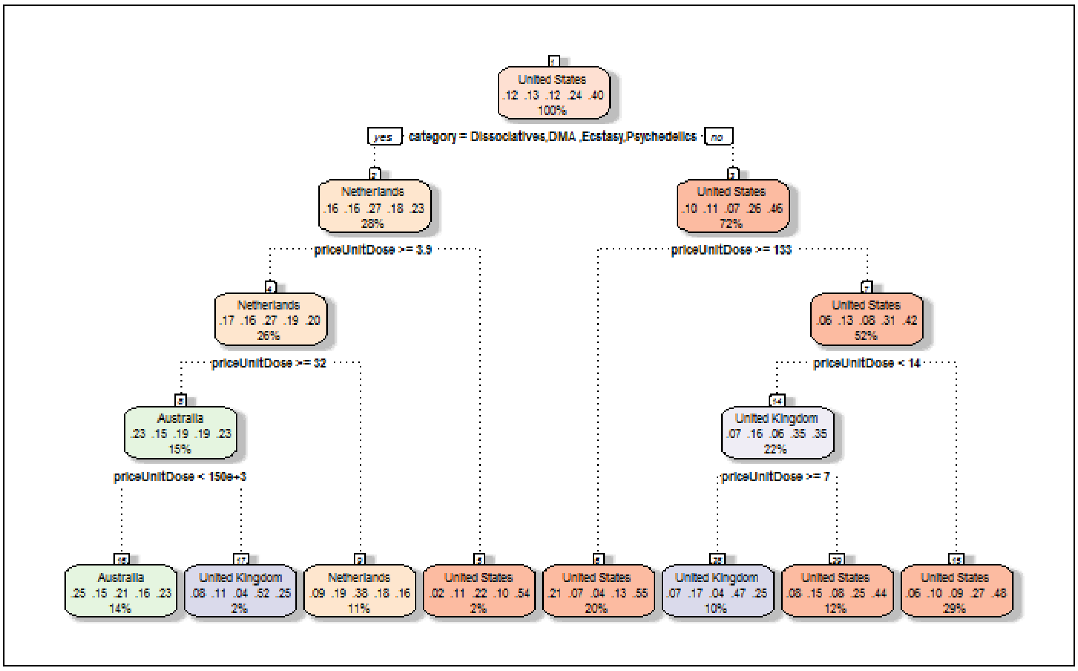
1. Decision tree

Looking at prices and categories, a *Decision Tree* has been created in order to predict the origin. In the same way the algorithm has been trained on one half of the data, and predictions made on the other half.

Code

```
                pred
                 Australia Germany Netherlands United Kingdom United States
  Australia            151       0          68             32           376
  Germany              122       0         131             97           369
  Netherlands          159       0         174             35           295
  United Kingdom       109       0         122            270           704
  United States        191       0         106            161          1733

         Table 8 - Origins Prediction / Decision Tree method
```




```
Figure 17 - Origins prediction / Decision Tree
```

Results don't seem to be very good, the accuracy is 43.07 % which is lower than previously. It turns out that without sellers, which give a lot of information on the origin, prognoses are not very reliable.

Now let's see if correlations can be found between categories and origins.

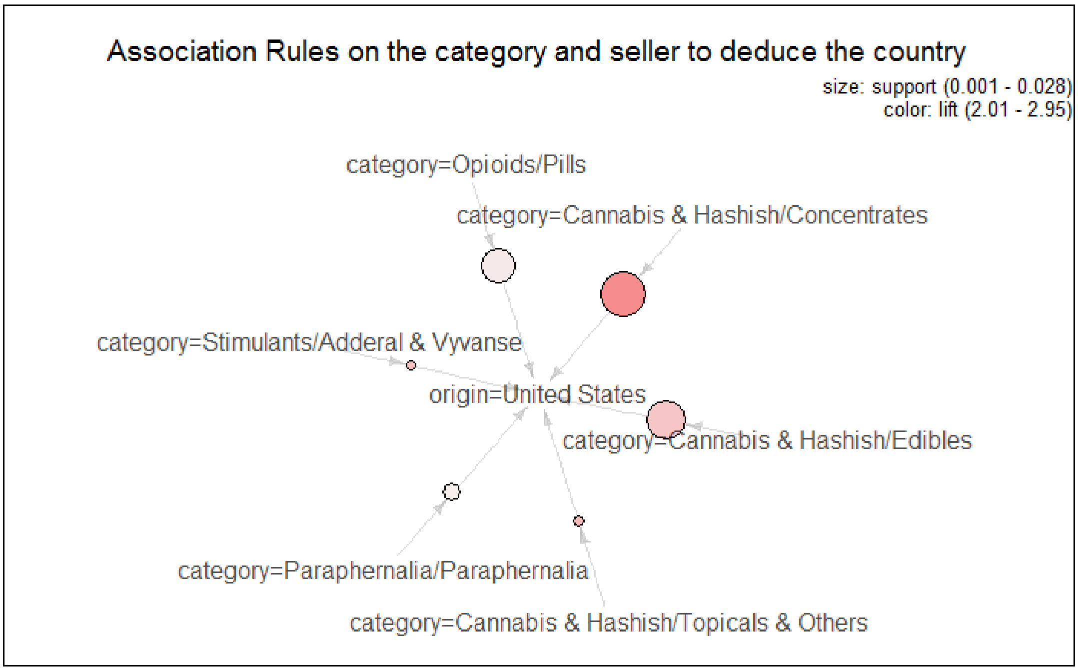
2. Rules to determine cateogries of ads coming from United States

To do so, *Association Rules Method* has been run with 2 variables : category and origin. Thus, one may be able to make a link with predictions of above decision tree. One country has been fixed, here United States.

Code

```
    lhs                                                   rhs                     support confidence l
ift
[1] {category=Cannabis & Hashish/Concentrates}       => {origin=United States}   0.028          0.76
 2.95
[2] {category=Cannabis & Hashish/Topicals & Others} => {origin=United States}   0.002          0.68
 2.63
[3] {category=Stimulants/Adderal & Vyvanse}          => {origin=United States}   0.001          0.67
 2.58
[4] {category=Cannabis & Hashish/Edibles}            => {origin=United States}   0.023          0.65
 2.51
[5] {category=Opioids/Pills}                          => {origin=United States}   0.020          0.53
 2.07
[6] {category=Paraphernalia/Paraphernalia}           => {origin=United States}   0.006          0.52
 2.01

               Table 9 - Origin Prediction / Association Rules
```




```
Figure 18 - Origin Prediction / Association Rules
```

The results show that when there is an ad of the category on the left, it is likely to come from United States with a confidence higher than 50%. Thus, ads from United States are often on Cannabis & Hashish, this can be easily confirmed by plotting a pie chart of United States exportations, as in the section before.

It is striking to see that Cannabis & Hashish seems to be the main rule. This can be explained by the legalisation of Cannibis in some states. Thus, the sales of these products is easy in United States and may interest people from other countries where they are not legalized.

6.4 - Prediction of ads profitability

Secondly, it worths to predict the profitability of an ad. That is to say, given an ad to predict if it will be sold a lot or not. Each ad have information on the category, origin, seller, price and a rate of profitability. This rate is caculated by dividing the number of product sold by the current lifetime of the ad and by times 30 to have a number of ads sold monthly. Bayesian Neural Network algorithm has beed accomplished on this new data and the results obtained are below.

Code

Profitability <fctr>	**Category** <fctr>	**Probability** <dbl>
[0.0000, 0.0896)	Benzos	0.466666667
[0.0896, 0.8795)	Benzos	0.296969697
[0.8795, 3.1915)	Benzos	0.042424242
[3.1915, 9.1691)	Benzos	0.060606061
[9.1691,480.0000]	Benzos	0.133333333

1-5 of 70 rows Previous **1** 2 3 4 5 6 ... 14 Next

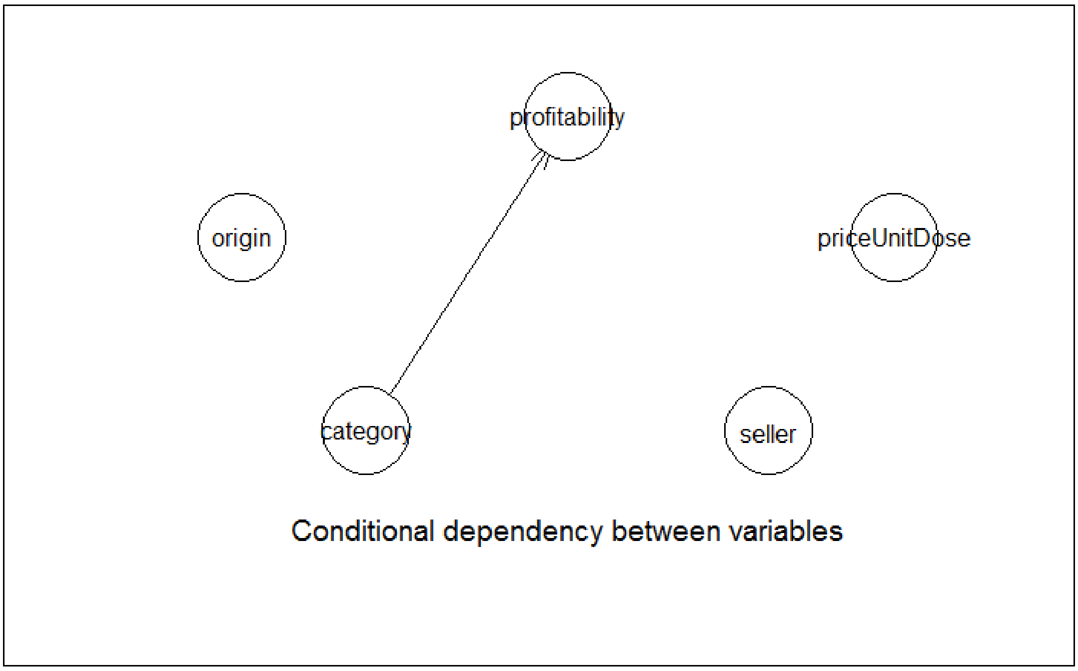
```
Table 10 - Categories Profitability
```

Category <fctr>	**Expectancy** <fctr>
Benzos	[1.48 , 30.75]
Cannabis & Hashish	[2.88 , 114.89]
Dissociatives	[1.77 , 74.75]
DMA	[1.56 , 84.02]

Category <fctr>	Expectancy <fctr>
Ecstasy	[3.24 , 103.68]

1-5 of 14 rows Previous 1 2 3 Next

```
Table 11 - Expectancies of category Profitability
```




```
Figure 19 - Variable Dependencies / Bayesian Neural Network
```

Neural Network shows that the profitability is conditionnaly dependant to category. That is to say category has a significant impact on profitability. Conditional probabilities are shown in the array. It is surprising that price and seller have no impacts on profitability.

Furthermore, expectancy of each event profitability X given category Y has been calculated. Thus, we can find the most profitable products to sold. Apparently it seems to be presciption, steroid and opioids. May be they are more profitable because they are not “common products” (and still wanted) contrary to Cannabis or Cocaine which can be found more easily on the street.

7 - References

[1] Baravalle A, Lopez MS, Lee SW. " Mining the dark web: Drugs and fake ids ".

[2] Baraniuk C. " BBC - alphabay and hansa dark web markets shut down " 20/07/2017. [Online]. [Accessed: 24/07/2017]. Available: http://www.bbc.co.uk/news/technology-40670010 (http://www.bbc.co.uk/news/technology-40670010).

[3] Kopan T. " CNN - doj announces takedown of dark web market alphabay " 20/07/2017. [Online]. [Accessed: 24/07/2017]. Available: http://edition.cnn.com/2017/07/20/politics/doj-takes-down-dark-web-marketplace-alphabay/index.html (http://edition.cnn.com/2017/07/20/politics/doj-takes-down-dark-web-marketplace-alphabay/index.html).

[4] " Google trends - explore: AlphaBay, dream market ". [Online] [Accessed: 22/07/2017]. Available: https://trends.google.co.uk/trends/explore?date=2015-01-01%202017-07-01&q=alphabay,Dream%20Market (https://trends.google.co.uk/trends/explore?date=2015-01-01%202017-07-01&q=alphabay,Dream%20Market).

[5] " Google trends - explore: AlphaBay ". [Online] [Accessed: 22/07/2017]. Available: https://trends.google.co.uk/trends/explore?date=2014-12-01%202017-07-01&q=alphabay (https://trends.google.co.uk/trends/explore?date=2014-12-01%202017-07-01&q=alphabay).

[6] " Cocaine - trafficking and supply (eu drug markets report) ". Paragraph: Modes of Transport [Online]. [Accessed: 24/07/2017]. Available: http://www.emcdda.europa.eu/publications/eu-drug-markets/2016/online/cocaine/trafficking-and-supply (http://www.emcdda.europa.eu/publications/eu-drug-markets/2016/online/cocaine/trafficking-and-supply).

[7] " DrugWise - how much do drugs cost? ". [Online] [Accessed: 12/07/2017]. Available: http://www.drugwise.org.uk/how-much-do-drugs-cost/ (http://www.drugwise.org.uk/how-much-do-drugs-cost/).

[8] " The telegraph - the cost of street drugs in britain. ". [Online] [Accessed: 12/07/2017]. Available: http://www.telegraph.co.uk/news/uknews/crime/11346133/The-cost-of-street-drugs-in-Britain.html (http://www.telegraph.co.uk/news/uknews/crime/11346133/The-cost-of-street-drugs-in-Britain.html).

[9] " RehabCenters.net - the average cost of illegal street drugs. ". [Online] [Accessed: 12/07/2017]. Available: http://www.rehabcenter.net/the-average-cost-of-illegal-drugs-on-the-street/ (http://www.rehabcenter.net/the-average-cost-of-illegal-drugs-on-the-street/).

[10] " Canada.com - what illegal drugs cost on the street around the world. ". [Online] [Accessed: 12/07/2017]. Available: http://o.canada.com/business/interactive-what-illegal-drugs-cost-on-the-street-around-the-world (http://o.canada.com/business/interactive-what-illegal-drugs-cost-on-the-street-around-the-world).

[11] " The student pocket guide - facts about drugs. ". [Online] [Accessed: 12/07/2017]. Available: http://www.thestudentpocketguide.com/2012/01/student-life/health-and-relationships/facts-about-drugs/ (http://www.thestudentpocketguide.com/2012/01/student-life/health-and-relationships/facts-about-drugs/).

[12] " R markdown from r studio - r markdown tutorial " . [Online] [Accessed: 29/06/2017]. Available: http://rmarkdown.rstudio.com/lesson-1.html (http://rmarkdown.rstudio.com/lesson-1.html).

[13] Tan P-N, Steinback M, Kumar V. " Introduction to data mining ". Addison Wesley; 2006.