Business risk of cannabis

Ed Sarausad (@edsar)

# Introduction

Financial services companies (banks, credit unions, community banks, private lenders) assess the risk and reward of doing business with clients during their KYC (Know Your Customer) process. At Alacer, we have been working with many top-tier banks in Canada and Asia and mid-tier banks in the US helping them improve their Anti-Money Laundering practices to comply with the Bank Secrecy Act (often abbreviated as AML/BSA compliance). We are always looking for ways to help our clients be more proactive about assessing operational risk in order to put in place more effective and efficient processes and systems in place.

Washington recently legalized recreational use of marijuana, following the state of Colorado. As a previously illicit industry, connections to organized crime and illegally-operating businesses still exist and present significant risk to banking and financial services companies, subject to AML/BSA compliance regulations.

There are at least 22 other states with plans to legalize marijuana in very short order. In the face of this growing industry in the United States, the interesting business questions emerge:

* To what extent is a bank currently exposed to risk from customers involved with marijuana-related businesses? This question addresses current operational risk.
* Should a bank offer services to new or existing customers who are involved with marijuana-related businesses transactions? This question addresses balancing the risk and opportunity of banking a specific customer segment.
* To inform a de-marketing decision quantifiably, is there a way to predict the likelihood of business being a risk-based liability? This question addresses current operational risk through a predictive risk model.

We will help solve the problem of assessing the risk and reward of doing business with their new target customer segment of marijuana businesses. Financial services companies, traditionally, are very conservative, yet they constantly trade-off the risk and reward of any decision. The Banking client will decide whether to do business with or NOT do business with (called de-market) specific marijuana business owners.

# Deeper data dive

We will use public data acquired from the Washington Cannabis Board and Secretary of State:

* 1. Washington Secretary of State - Business registration data including trade name, address, owners, etc. This data when graphed in a network can help to understand the relationships between different businesses.
  2. Washington state Cannabis board - Companies who have applied for licenses including the state of the approval (pending, approved, etc) and longitudinal sales and violations (for those approved businesses).

In addition to this public data, we have also started to explore the use of proprietary data sources (notably from White Pages and IBM/Alchemy) to ask the following:

* Using directory services what do we know about specific business owners and phone numbers and locations associated with them?
* Using NLP on news articles, can we learn something new about   
  “busted” companies and related entities and the sentiment of articles, keywords and related entities?[[1]](#footnote-1)

We will focus on the analysis of the public data about which we can be more transparent with data and analysis-sharing. The applicant, sales, and violations data contain the following important fields and information:

|  |  |  |  |
| --- | --- | --- | --- |
| Data Source\* | # Records/ #Variables | Fields | Significance to analysis |
| [Sales](https://www.dropbox.com/s/yxgij2mtsou0lf0/mj-sales-transformed.csv?dl=1) | 6979 / 7 | Reporting period, total sales, excise tax due | Can analyze longitudinally sales and tax due by license |
| [Violations](https://www.dropbox.com/s/euwbe6mciu9oaw9/mj-violations-transformed.csv?dl=1) | 542 / 9 | Visit date, penalty type, case number, WAC code, violation type, | Can analyze longitudinally penalties (by type) and violations (by type) by license |
| [Applicants](https://www.dropbox.com/s/6guyvoo6a28d6vx/MarijuanaApplicants11032015.xls?dl=1) | 5903 / 29 | Full address, date created, email, day phone, night phone, owner name, privilege status, reason action, status date, trade name, UBI, type of business | Can analyze relationship of business type (producer, processer, retailer, medical), status of application, reason over time |

\*Since License number is provided in all data sources, it makes the the license a clean unit of analysis.

* + What are its limitations i.e. what are some questions that you cannot answer with this data set?

Some limitations include: the inability to address seasonality with only 1-2 years of data, the inability to generalize across states/countries with one Washington state data, and inability to correlate to demographic factors (general population, income, education levels) without US Census data.

* + What kind of cleaning and wrangling did you need to do?

We have kept all data in tact and compressed applicant data into a single CSV file denoting the business type. To facilitate specific date and time analysis, we added columns for booking sales at the end of the reporting period and a timestamp column required for certain visualization.

* + Are there other datasets you can find, use and combine with, to answer the questions that matter?

Other datasets from Colorado and Oregon could be interesting to combine with the Washington data to tease out differences between states. Having multiple states in the dataset would make the conclusions more generalizable to other states. Furthermore, data from other countries (Amsterdam, Holland) would also be interesting to analyze as there may be more years of data and less of a US-centric skew to the data.

# Preliminary exploration



Figure Overall View of Applicants By county



Figure Spokane county: Home of the Issued Licenses



Figure Tall, Green Unicorns: High-Grossing Businesses with few Violations

Based on the original exploration, we found an interesting concentration of licensed businesses in Spokane county which was interesting, especially since we have community bank clients from Spokane who are concerned about the problem we are investigating.

Furthermore, when applying this more broadly to a generalized population (eventually outside of a single county or state or country), we would like to test the following additional hypotheses. Additional US census data on population, economic output, or educational levels may be to add to our analysis, time permitting.

* Hypothesis 1: Businesses with severe violations (license suspension or cancellation) are significantly different from businesses without violations.
* Hypothesis 2: Business with healthy sales are in population centers or close to borders where excess demand from other states exist.
* Hypothesis 3: Business exhibit different risks (for retailers: sales to minors, for producers: theft and fraud) which can be measured.
* Hypothesis 4: Business health (measured by sales, as excise tax is a flat 25%\*) will differ by type of business (these are driven by factors out of scope such as acumen of business owners) [[2]](#footnote-2)

# Approach

**Regression/Classification** approaches will be used to tease out effects of business type, location, length of business, sales, previous violations on the labeled outcome of severe violation (suspension or cancellation of license).

**Supervised learning** methods including decision trees, Naïve Bayes, SVM

For unlabeled data requiring **unsupervised learning** methods, we’ll cluster (k-means) and look for other methods of finding patterns of risky behavior.

**Forecasting** market size by business type (producer, processor, retailer, medical) is interesting for banks and policy-makers alike. Banks may formulate specific business rules based on the type of business (producers may be less risky than retailers to violations such as sales to minors.

We will conduct numerous iterations on **feature selection** for the above and **evaluation** (using precision, recall and balanced F1 score).

# Resources

<https://github.com/edsarausad/sliderule/tree/master/capstone>

1. <http://kimatv.com/news/local/local-pot-stores-pass-compliance-checks-dont-sell-to-minors> lists 19 companies recently busted [↑](#footnote-ref-1)
2. Beginning July 1, 2015, the excise tax structure for recreational marijuana is changing.

   Under the new law ([2E2SHB 2136](http://app.leg.wa.gov/billinfo/summary.aspx?bill=2136&year=2015)), the marijuana excise tax is:

   No longer levied on the producers, processors and retailers. Instead, the tax is now imposed on retail customers.

   Going from 25 percent (at each level) to 37 percent on the retail customer.

   No longer part of the selling price subject to retail sales or B&O taxes. ( Source: <http://dor.wa.gov/Content/FindTaxesAndRates/marijuana/Default.aspx> ) [↑](#footnote-ref-2)