**Introduction**

I have always been interested in firearms, even though my family has limited background in the hobby. I grew up knowing that my father and grandfather had firearms, but I never actually saw them use them. Through the Boy Scouts I came to develop a love for target shooting, and I began going with my father to a local state park to occasionally practice. As I came of age, I gained the right to keep and bear arms of my own, and I quickly developed quite a collection. I also developed a knack for repairing and building firearms, as the mechanics and engineering are very interesting to me. It has come to replace computers as my hobby, as computers are now what I do for work.

When I first learned of this project, I immediately thought of a data set that was gun related. I thought about gun crime statistics, but I was worried it would be hard to find data that would be relevant for analysis and would not be politically biased. I found an excellent candidate in National Firearms Act transfer times. The Act, described in detail in the next section, involves lengthy registration paperwork that can take months to get back. Statistics on this paperwork are self-reported by the firearms community to the website NFATracker.com, giving prospective buyers some idea of how long it will take before they can possess their item. I obtained a .CSV file from the website containing transfer statistics back to 2005. The following is an overview of the relevant terms and legislation, and an application of some topics from our class to analyze that data.

**Topic 0: An Overview of The National Firearms Act of 1934**

The National Firearms Act, or NFA, is a piece of Federal legislation regulating the sale, possession, and manufacture of various firearms and firearms components. Passed in1934 during the height of gangland violence, it was designed to prevent criminals from easily accessing concealable weapons. In practice it did not achieve this goal. It imposed a $200 tax and registration requirement for machine guns, short-barreled rifles and shotguns, suppressors, destructive devices (explosives and artillery), and “any other weapons”.

Machine guns are defined as: “weapons that shoot, are designed to shoot, or can be readily restored to shoot, automatically more than one shot without manual reloading by a single function of the trigger” (United States, p.9). In other words, one trigger pull, many shots. Since 1986 machine guns have been illegal for civilians to register, so all existing legal machine guns have been created before that date. The limited supply and extreme demand have led to legal machine gun prices skyrocketing, with items auctioning routinely in the tens of thousands of dollars.

Short barreled rifles and shotguns are defined as a: “weapon that has an overall length of less than 26 inches or a barrel or barrels of less than 16 inches in length” (United States, p.7). Short barreled rifles and shotguns must have at one time existed as regular firearms not applicable to the NFA, so pistols would not fall under this jurisdiction. Interesting to note, one cannot simply attach a shoulder stock or a vertical forward grip to a pistol, doing so would be illegally creating a short-barreled rifle or shotgun. However, if one was to purchase a pistol and attach a “stabilizing brace” to use as a shoulder stock, it is still considered a pistol. A “stabilizing brace” is a device designed to help shooters hold up a pistol with one hand, and so was not designed to be used as a stock (Figure 1). This apparently makes it kosher with the ATF. It is a loophole popularly exploited by firearms enthusiasts, as it requires no tax or paperwork.

Figure 1: The item on the left is a "stock", the item on the right is a "brace".

Suppressors, commonly referred to as silencers, are defined as “any device for silencing, muffling, or diminishing the report of a portable firearm” (United States, p.15). If a device decreases the report of gun by a fraction of a decibel, it is considered a silencer. Silencers are by far the most popular NFA item, as it is extremely useful for hunting and prolonged repetitive shooting. Contrary to their depiction in video games and movies, silencers do not completely mask the report of a firearm. It is still necessary to wear hearing protection when firing most calibers suppressed. The device may suppress the sound somewhat, but the crack of a bullet breaking the sound barrier can still damage your hearing.

“Any other weapons (AOWs)” are essentially guns designed to not look like guns such as pen guns, zip guns, knife guns, cane guns, and briefcase guns. AOWs are not very popular outside of hardcore collector circles. Interesting to note, the firearm in question has to not resemble a firearm only when in a fireable condition. So, pistols exist that look like pens or cell phones when holstered but transform into the shape of a handgun when ready to fire.

Despite being close to 85 years old, the NFA has remained mostly unchanged since it’s passing. The NFA is and has been considered to be outdated, ineffective, and burdensome by the American firearms enthusiast community. The initial tax of $200 was purposefully set to make it impossible for anyone but the extremely wealthy to possess regulated items. $200 in 1934 is the equivalent to $3,658.51 in today’s money, so buying an NFA item was like buying a car. It has only recently become affordable to own an NFA item due to inflation. The registration or transfer of an NFA item requires a grueling and exhaustive paperwork process involving fingerprints, passport photos, and an FBI background check. The chief law enforcement officer of the jurisdiction where the relevant entity or individual lives or is located must be notified when an NFA item is created or transferred. NFA items cannot be transported across state lines without first informing the Attorney General. Several states such as New York and New Jersey, have banned private possession of certain NFA items.

Due to the Bureau of Alcohol Tobacco and Firearms (ATF), the government agency assigned to handle NFA paperwork, being incompetent and inefficient, it can take anywhere from two months to two years for a person to have NFA paperwork approved. Many who wish to own rifles with shortened barrels simply use the aforementioned loophole of buying a pistol and adding a “brace”. Those who wish to own any other type of NFA item have no such recourse. Despite many half-hearted efforts to streamline the paperwork process, the dramatic uptick in NFA item popularity and affordability has left and ever-increasing backlog of registration requests.

The requests themselves are divided into several different types of form, two of which are relevant to an analysis. ATF Form 1 is an application to manufacture an NFA item, and ATF Form 4 is an application to transfer an NFA item. These forms can all be filed by individuals, corporations and LLCs, and trusts. For example, an LLC could file a Form 1 to manufacturer their own silencer, and an individual could become owner of it using a Form 4. Each time any ATF form is filed, a $200 tax must be paid. When paperwork comes back, a “tax stamp” is given to prove that the form was paid for and approved.

As you can see, the NFA is an extremely complicated and convoluted piece of legislation. It has many more intricacies and exception that are too numerous to be mentioned here. It is no wonder that it is recommended to have a firearms lawyer assist you when attempting to purchase or manufacture and NFA item. An expert on hand also helps applicant from making mistakes on paperwork that could result in forms being sent back unapproved, potentially adding months on to an already excruciating wait time.

United States, The Bureau of Alcohol, Tobacco, Firearms and Explosives, Office of Enforcement

Programs and Services. “The National Firearms Act Handbook.” The National Firearms

Act Handbook. www.atf.gov/firearms/national-firearms-act-handbook.

**Topic 1: Grooming Out Anomalous Entries**

Given the self-reported nature of the dataset I am using, I was expecting for most of the data points to be noisy. Indeed, most entries were not complete. These entries were either date of forms being sent, date of tax check being cashed, date of form approval, and date of form return. Out of the total 38704 entries, 23103 had to be purged due to incompleteness. That is a loss of close to 60%. Although much data was lost, it was a necessity as to not contaminate the results of my analysis.

**Topic 2: Irrelevant Features**

Each data entry has features that are statistically irrelevant. Users who wished to could add a description of their entry on NFATracker.com, but a majority of the entries simply read “Other”. The description of an entry has no impact on how fast paperwork is processed, so it was ignored in my analysis. Like the description, the name of the paperwork processor and the state in which the applicant resides were also ignored.

**Topic 3: Removing Noisy Data**

In order to ensure the integrity of an analysis, noisy data had to be removed from my entries. Also known as outliers or fliers, these points fall well outside the typical range for the rest of the data entries. My data set was small enough that noisy entries could be removed visually by looking at a graph.

**Topic 4: Data Preparation**

I removed any data entries with time elapsed between form sent and stamp received over 500 days. Although these entries may indeed be valid, I had to remove them to preserve the integrity of an analysis. They were greatly different from the rest of the data, and they were too few to justify keeping in the set. You would have to do something similar with data on TSA screening times. In a few cases, for whatever reason, it might take hours for a person to get through security. For the vast majority of people, the screening process takes at most 20 minutes. You would remove entries based on if their time exceeded a clearly excessive point.

**Topic 5: Entropy**

Calculating the entropy of a dataset is essential in order to determine its purity. It is always a good idea to run entropy on a set, as it can verify the validity of the findings of an analysis. The final, purged and culled data set has an entropy of 0.000894804, indicating an extremely pure set. This ensures that any of my results would be relevant and statistically accurate.

**Topic 6: Linear Regression**

In order to draw conclusions from my data, I had to use linear regression. Linear regression is an algorithm that determines the likelihood of correlation between two data features. For example, if I had data on health and fast food eating habits of the American public, I could use linear regression to determine if there was any correlation between BMI and total dollars spent on fast food. It is by far the most powerful tool I used for this project. Before I could begin the regression analysis, I had to add several extra properties to my data entries. Linear regressions can only interpret numeric data, so steps had to be taken to convert existing properties into numeric heuristics. For example, I had to calculate the total days between a form being sent and received for each of my entries. The following are the results of all regression analysis performed, as well as my thoughts on each.

|  |  |  |
| --- | --- | --- |
| Properties Compared | Multiple R | R Square |
| Whether or not an entry was electronically filed VS days between form was sent and stamp received. | 0.419784577023598 | 0.176219091106881 |
| Form type VS days between form was sent and stamp received. | 0.18188865006572 | 0.0330834810227301 |
| Examiner VS days between form was sent and stamp received. | 0.10534033001475 | 0.0110965851276165 |

I performed the first test (Whether or not an entry was electronically filed VS days between form was sent and received.), because I wanted to know if a form is electronically filed actually has any impact on how fast the paperwork is processed. Based on the results of the analysis, it appears that e-filing does not have a consistent impact on how long processing takes. I am not surprised by these results, as it is typical for a government agency to be slow and inefficient, regardless of what technology is involved. E-filing was supposed to be available to the general public for all types of forms years ago, but as of writing it is only available to dealers and certain types of forms. ATF could have had a paper-free filing system with wait times in the days but making things easy for gun enthusiasts is on the bottom of their priority list, if it is even on there at all.

The second analysis performed (Form type VS days between form was sent and received.) was conducted in order to prove whether processing time had anything to do with the type of form submitted. Based on the extremely low multiple R and R square, I would say there is no correlation between the two at all. This essentially confirms my suspicion, that there is no organization going on at the AFT NFA division. I would not be surprised if mailed forms were thrown in a box for examiners to pick from. Whatever gets done is what gets done. After all, it’s not their money that gets wasted.

My third and final regression analysis (Examiner VS days between form was sent and stamp received.) was performed in order to determine if certain paperwork examiners performed better worse others. Based on the results, I would say that all of the examiners are equally inefficient. To be honest, I am not very surprised at these results. Perhaps they all have a pact together in order to preserve their jobs? If some people were more efficient than others, then the less efficient ones could get cut. Perhaps there is an NFA paperwork examiner union?

**Topic 7: Sampling**

A sampling of 1000 entries was selected randomly from my set to perform a sampled analysis. Sampling can be useful to determine if phenomenon observed in the whole data set is observed in part of the data set. My first analysis on the sample (Form type VS days between form was sent and stamp received.) was actually worse than the aggregate analysis (0.0287728255456999 multiple R). This indicated to me that it would not be worth it to analyze any other features in the sample for correlation.

**Topic 8: Binning**

One of the things that became readily apparent when viewing my data on a graph is that there are massive spikes in processing time around the years 2011-2013, and 2016. These were caused by a deluge of new forms being sent in because of the Sandy Hook shooting and the reelection of Obama, and an ATF rule change requiring more intensive paperwork and fingerprints. In the case of the first events, there was fear of a possible ban on NFA items which did not materialize. The 2016 rule change caused many to put in paperwork all at once in order to save hassle later.

|  |  |  |  |
| --- | --- | --- | --- |
| Date Range | Properties Compared | Multiple R | R Square |
| Bin 1: 12/14/12 to 7/13/16 | Whether or not an entry was electronically filed VS days between form was sent and stamp received. | 0.467191704979214 | 0.218268089201385 |
|  | Form type VS days between form was sent and stamp received. | 0.181735300363353 | 0.033027719398158 |
|  | Examiner VS days between form was sent and stamp received. | 0.0680611486299598 | 0.00463231995282948 |
| Bin 2: 1/1/16 to 10/10/18 | Whether or not an entry was electronically filed VS days between form was sent and stamp received. | 0.371437255911829 | 0.13796563507931 |
|  | Form type VS days between form was sent and stamp received. | 0.145380500646273 | 0.0211354899681609 |
|  | Examiner VS days between form was sent and stamp received. | 0.0779743881585343 | 0.00608000520869777 |

Based on the results from the tests, it would appear that the results for many of the bins are as bad, or worse than the results for the whole. The main exception would be the electronic filing analysis, with some marginal evidence of correlation. I am surprised to see that the multiple R went down over time on that test, as supposedly the electronic filing system received many updates and improvements in the last few years. But remember, this was from the same people who brought you Healthcare.gov, and we saw how well that one worked.

**Topic 9: Simple Practical Implementation**

Given that the class is called machine learning, I figured it would be interesting to attempt a simple program to determine how long NFA paperwork should take based on date of submission. It is capable of predicting wait time based on various properties of a submission. It even can perform a multilinear regression to predict an entire hypothetical submission. It can even predict how long an entry should take (in the context of the total average) given a particular date of the check being cashed. I am proud of myself for the features I was able to add in such a short period of time, and I am surprised it worked as nicely as it did. The source code for the entire project, as well as copies of my final data sets can be found on my GitHub (https://github.com/msimanski/NFApredictor).