# Public Notice Analysis- Final Report

## I. INTRODUCTION

Public Notices are advertisements that are required to be published in the local Newspapers by law and are of importance to the social community. Public Notices are good information source for concerned citizens of a community and can help to build a public consensus on the important matters of a community. Recently, a company named CVS Caremark was running fraudulent activities against the state of Iowa and Newspaper notices helped a citizen to identify the activities of the company which ultimately led to the savings of tens of thousands of state dollars [1].

In this project, we attempt to analyze the public notices from the different states and try to find insights. To the best of our knowledge, there has not been any study on the public notice trends yet which leaves a huge opportunity to analyze this big chunks of data about society.

Our Contribution are listed below.

- *Category specific entity extractor*: We have identified different structures related to each public notice category and built feature extractor accordingly. We present our feature extractor in Section II.
- *Category classification*: We present our classification results in Section IV. Our deep learning model have an accuracy around 99% in classifying each public notice category.
- *Clustering*: Clustering results presented in Section V and VI were helpful in finding interesting insights from public notices.
- *Social Trends*: Finally, in Section VII, we present how public notices are related to different aspects of society.

## II. PRE-PROCESSING

We have used two data-set for this project. Our first data-set contains 27k legal notice with 26 tagged category. And our another data-set contains 9GB text data with 640k legal notice but no category. But for the both date-set we have state, county and date information. We have used our first data-set to train a model for the category classification task in the second data-set. Next, from the 640k data-set, we choose 10k legal notices randomly form 7 years, total 70k notices and predict the category using the trained model form the first data-set. Once we have the category, we append these 70k notices data-set with 27k notices data-set and finally we have a 97k notices data-set. We use these 97k notices data-set with 26 categories for our analysis 1 . We can't classify all the 640k notices due to resource scarcity. We do not have enough computing power to run the model on the 9GB text.

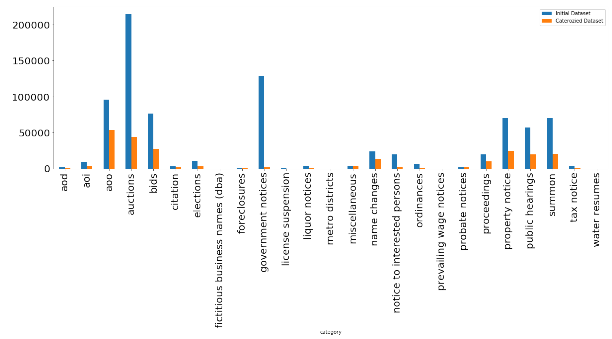



Fig. 1: Number of notices per categories, in both dataset

The outcomes for this project are described below.

## III. Entity Detection

The objective in this task is to find different entities mentioned in the notice. There are different types of notices, and every notices has different entities. For example, in a 'name change' notice, there are two entities, the old name and the new name, the 'auction' notice has the items name, date, time, address of the auction etc. So, we build several extractor for the entity detection form different kind of notices. First, we go through the notices to observe the pattern and then use some regular expression and string search technique to find the entities. Although, there are total 26 categories, we do not observe the pattern for every category. We are able to detect the entities for the 6 categories of notices described below.

*1) Article of Organization:* Article of organization(AOO) is used to outline the initial statements required to form a limited liability company. We read more than 100 notices in this categories, and find out that they have 3 to 10 sections in the notice, starting with a number 1 (Figure 2a). To extract all those entities we used regular expression to extract text in between two consecutive numbers. In many cases, instead of numerical digits, roman numerals are used. We also consider this case and out of 95754 AOO notices, we were able to extract information from 70941 notices. Figure 2b shows the extracted information form the sample AOO notice of Figure 2a.

*2) Name Changes:* We again go through more then 100 notices to find a pattern, but there is none. Next, we generate most common sub-string of length 3, 4 and 5 to check whether there are any words or phrases that are used frequently. We figured out that, some specific sub strings are occurred in-some specific states. For example, these sub strings *change petitioners name*, *change his her name*,

NOTICE OF FILING ARTICLES OF ORGANIZATION
OF THE CREATIVE
DUO, LLC
An Arizona Limited Liability Company
Pursuant to A.R.S. ï¿½29-635(C), notice is hereby given that the Articles of Organization of THE CREATIVE DUO, LLC, An Arizona Limited Liability Company, were filed in the Office of the Arizona Corporation Commission for:
1. Name: THE CREATIVE DUO, LLC
2. Registered Office Address: 6046 E. Hannibal, Mesa, AZ 85205
3. Statutory Agent: SHERI BAUMGART, 6046 E. Hannibal, Mesa, AZ 85205.
4. Management: Management of the limited liability company is at the time of the formation of the limited liability company and reserved for the member(s) whose name(s) and address(es) are as follows:
Member(s):
SHERI BAUMGART
6046 E. Hannibal
Mesa, AZ 85205
TEELI JONES
2133 E. Caspian
Queen Creek, AZ 85140
2/17, 2/22, 2/24/17
RR-2977006

(a) Sample 'AOO' notice

```
name: the creative duo  llc

registered office address: 6046 e. hannibal  mesa  az 85205

statutory agent: sheri baumgart  6046 e. hannibal  mesa  az 85205.

management: management of the limited liability company is at the time
of the formation of the limited liability company and reserved for the
member s  whose name s  and address es  are as follows:  member s :
sheri baumgart  6046 e. hannibal  mesa  az 85205  teeli jones  2133 e.
caspian  queen creek  az 85140  2 17  2 22  2 24 17  rr-29770
```

(b) Extracted entity from the notice

Fig. 2: The extracted entity form the AOO notice

*change the name of the child*, *changing the petitioners legal name*, *present legal name of petitioner*, *change petitioners name* occurred frequently in the state of Nevada. For New Mexico, we found some other sub-strings. The old name and the changed name appear before and after these sub-string with the keyword from and to. Figure 3a shows an example. The text in the first red box is our key-phrase to localize the name change. Next, we detect the keywords from and to to help us detect the old name and new name. In some notices, the name of companies also change, and that followed the same format as we have observed in the AOO, AOI, AOD notices. Figure 3b shows both name change notices for a person and a company.

```
notice in the second judicial district court of the state of nevada in and
for the county of washoe case no. cv-1601731 dept. no. 8 in the matter of
the petition of vincent trieu for a change of name. on the 24th day of
august  2016  a petition was filed in the above-entitled court by vincent
trieu  requesting the court to legally change his her name from vincent
trieu to vincent chew. any opposition to this petition should be filed
with the above-entitled court within ten  10  days of the final
publication of this notice. dated this 24th day of august  2016  s
jacqueline bryant clerk of the court   illegible  deputy clerk no1535186
aug 26  sept 2  9  2016
```

(a) Sample 'Name Change' Notice

```
('mary anna', 'rkelson   mary ann  rkelson')
('leslie feth', 'leslie kalyn')
('tengonne vongpanya', 'teng vongpanya dante')
```

Change name of a person

```
entity name: galu energy  llc
a.c.c. file number: 1-1631246-5
entity name change: a group studio  llc  dated august 12  2016.  s  allison b.
felicioli  member.  jewish news 9 2  9 9  9

entity name: the mccoy team pllc
a.c.c. file number: p-20576664
entity name change: alexandra & keith mccoy pllc 7. professional services: to
render professional real estate services.  s  alexandra mccoy 3tc pub jan. 16
17  18  20

entity name: the mccoy team pllc
a.c.c. file number: p-20576664
entity name change: alexandra & keith mccoy pllc 7. professional services: to
render professional real estate services.  s  alexandra mccoy 3tc pub jan. 16
17  18  20
```

Change name of the company

(b) Extracted entities from the notice

Fig. 3: The extracted entity form the Name Change notice

NOTICE TO CREDITORS
Barbara L. Ayers, 1002 S. Duquesne Dr., Tucson, AZ 85710, was appointed as personal representative for the estate of Jack Eugene Ayers, Pima County Case No. PB20170120. Any claim not presented against the Estate within 4 months after the date that this notice is first published shall be forever waived.
PUBLISH: The Daily Territorial
Apr. 6, 13, 20, 2017

(a) Sample Probate Notice

```
legal no 5108  notice to creditors: darin leverl aulston  #9 cross road
granville  ma 01034  has been appointed as the personal representative of the
estate of jonathan gayle aulston

creditor notice james c ball  aka jim c ball or james craig ball passed away 10
20 16; ross vandersommen was appointed 11 2 16 as personal representative of
the estate  yavapai county superior court #p1300pb201600288
```

(b) Extracted entities from the notice

Fig. 4: The extracted entities form the Probate notice

*3) Probate Notice:* To detect the entities from probate notice, again, we find some frequent sub-strings, *appointed personal representative of*, *as personal representative*, *appointed as the personal representative* etc. These sub-strings indicate the person who is responsible for any claims against another person who is deceased (Figure 4a). We extracted the key part of the notice using those sub-string as there are lot of variations of the appearance of the deceased person name and the representative name (Figure 4b).

*4) Elections:* Our manual approach again fail to find any pattern. Again we generate the sub-strings of the length 3, 4 and 5 and we found that phrases like *will be held*, *to be held* occurred most of the time. These two phrases are also indicator of declaration of some kind of elections that will be held soon. Out of 1299 notices, we found out 898 notices has these two sub-strings. we need to investigate further to localize the types of elections and the date and place of elections.

*5) Liquor Notice:* In this category, the most frequent pattern is *applying to the* and there are also key-words such as *address* or *location* which indicate who is applying for liquor license.

*6) Other Categories:* We also tried to extract entities form the other categories, but there are no pattern to design any information extractor. The Auction and Bids are two highly frequent notice categories, but there are no specific pattern. However, we observed that in the Auctions most frequent words are *vin*, *vehicle identification number* and *property*.

## IV. Classification

We have a data-set containing 27k notices with categories. We use this tagged notices as our training data and predict the categories for the other 70k notices. This will enrich our categorized data-set for the analysis. We use separate classifier for the different category. It is a binary classification problem where one class is our target category and we randomly choose equal number of notices form all other categories. We use tf-idf [2] representation of the notices and the XGBoost [3], an optimized distributed gradient boosting library for the classification. When dividing the training and testing data-set, we make equal number of instances from the both classes in the test data, and use 5 fold k-validation. We achieve precision and recall around 0.99 (Table I) for all the major categories. The reason for these high accuracy is, the

| Categories | Number of test data | Precision | Recall | Accuracy |
|---|---|---|---|---|
| AOO | 53510 | 0.999113 | 0.999113 | 0.999112 |
| Auctions | 43661 | 0.995694 | 0.995683 | 0.995683 |
| Bids | 27460 | 0.99816 | 0.998161 | 0.998161 |
| Name Changes | 13768 | 0.998257 | 0.99826 | 0.998256 |
| Probate Notices | 2021 | 0.998558 | 0.998477 | 0.998514 |
| Property Notice | 24487 | 0.999448 | 0.999449 | 0.999449 |
| Proceedings | 10019 | 0.992912 | 0.992868 | 0.992863 |
| Summon | 20591 | 0.999951 | 0.999952 | 0.999951 |

TABLE I: Single class classification accuracy using deep learning model

notices for a particular category has a boilerplate pattern. So it is easy for the classifier to learn the pattern and make the accurate prediction. We also tried with Logistic Regression and Random forest Classifier model, but got the best results with XGBoost.

## V. Clustering

In this section, we present our clustering results on the public notices. We used TF-IDF score as our feature. Mini-batchKmenas is used as a more scalable version of K-means. We estimated the number of clusters among the notices by finding the mean silhouette scores for different k. k = 4 gave the best result. Figure 5 shows the distribution of public notices among the four clusters. It has to be mentioned here that, the clustering was done on fifty thousands representative notices. Also, some classes have very small in-cluster proportion so they are omitted from the plot for brevity. From the Figure 5 we can see that, Cluster 1 and 2 are comprised of AOO notices whereas Cluster 0 is almost fully comprised AOI notices. And Finally, Cluster 3 is mainly comprised of Auction and Proceeding notices. This clustering result further bolsters our intuition that different notice types have different structure among them.

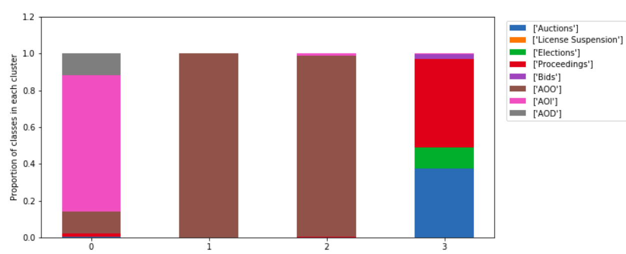
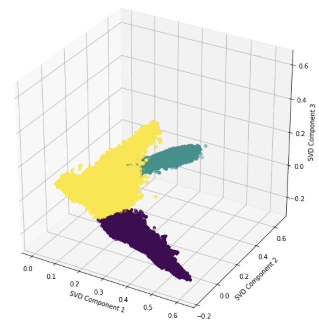


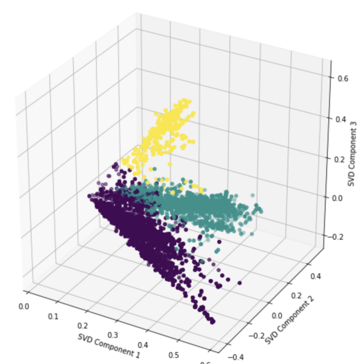

Fig. 5: Clustering Result on 50k Representative Notices

## VI. EDA on Different Notice Categories

In this section, we have tried to find interesting insights from the dataset. For this, we ran $k$-means clustering on each of the notice category. The intuition was that, there may be sub-categories among each type of notices. Although, most of the categories seem to be fairly pure and there is no apparent sub-categories among them, we found distinct clusters in AOI and Auction Notices. As it turns out, most of the AOI notices are of declaring a profitable or non-profitable organizations and significant amount of Auction notices are about auctioning vehicles.




(a) Auction clusters




(b) AOI Clusters

Fig. 6: 3D plot of $k$-means clustering on Auction and AOI notices. Each color represents a cluster

The clustering result on this two categories are presented in Figure 6.

*1) Auctions:* As mentioned above, one of the clusters in Figure 6b corresponds to auctions specific to vehicles. We found out that, each of these notices have the keyword *vin* in it, which means Vehicle Identification Number (VIN). A *vin* consists of 17 digits, where each digit has some information associated with it. Figure 7 shows the meaning of each digit in a *vin* number. We extracted all the *vin* numbers from these notices.

We used the first digit to find the place where the vehicle was built, second and third digits to find the manufacturer and tenth digit to find the year, when it was built. Next we map these extracted information with *vin* code [4] to find the human understandable form. We found out that the vehicles are built in all over the world including Africa,

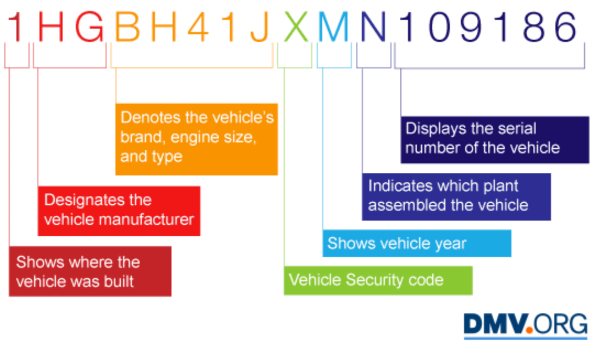



Fig. 7: Decoding a VIN

Australia, Europe, South America, Asia continent, although most of them are from North America. There are wide variety of manufacturers, but in the auctions list most of them are from Ford (Figure 8). We also observe that , most of the vehicles are manufactured in the year around 2000 (Figure 9).

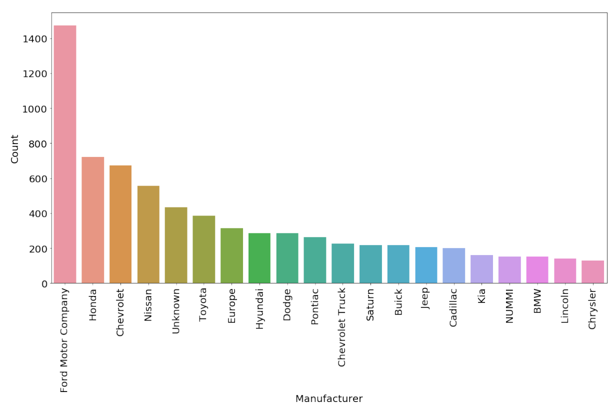



Fig. 8: Manufacturer Top 20

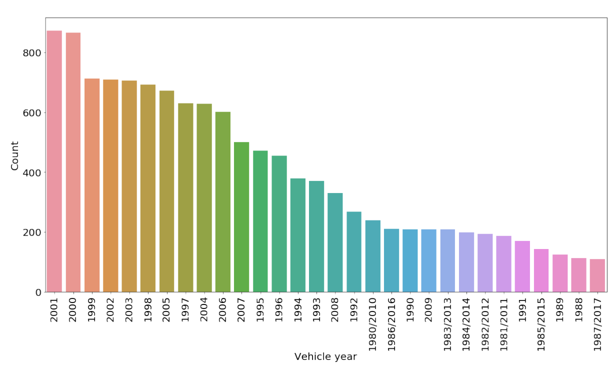



Fig. 9: Vehicle manufacturing year

*2) AOO:* The articles of organization are a document that outlines the initial statements required to form a limited liability company in many U.S. states. This is one of the most frequent types of notices as every new business has to file a AOO to open their business. We tried to find what kind of business are frequent in Arizona as most of the AOO notices are from Arizona. We extracted all the company name from the AOO notices using simple string matching operation and formed a word cloud from them. Figure 10 shows some of the popular business categories in AOO notices.


Fig. 10: Word cloud of Business types in AOO notices

Most popular business categories are *Construction, Consulting, Investment organization, Management, Services, Law firms, Real estate,* and *Property*.

*3) License Suspension:* Next, we investigate the License Suspension notices. Even though there were only 288 notices of this particular category, the notices were almost evenly distributed among all the 8 states, meaning this type of notices indeed are rare. Rare data often contain interesting patters and as we read through the notices, we found out that, there were four categories of license suspension notices: Nursing/Medical License Suspension, Driving License Suspension, Law Practice License Suspension and Business/Industry License Suspension notices. Figure 11 shows the distribution of License Suspension notices among the four categories.

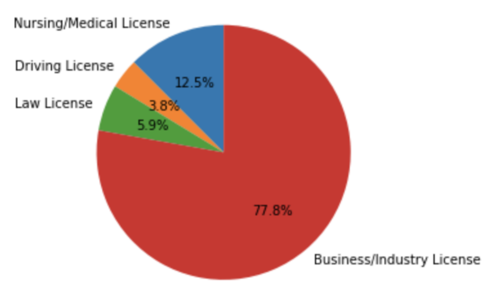



Fig. 11: License Suspension Distribution

In Figure 12, we show the state wise distribution of the license suspension notices.

As evident from Figure 12, a lot of industries are loosing their license in New Mexico. We investigated those notices and found out that, they are loosing their license because of violating air/water/dust pollution limits. Also, 16 nurse/practitioner's lost their license in Arizona.

*4) AOI:* As we have said earlier, Article of Organization (AOI) notices can be broadly divided into two categories - declaration of profitable and non-profitable organization.

| | state | Nursing/Medical | Driving | Law | Business/Industry |
|---|---|---|---|---|---|
| 0 | Arizona | 16 | 0 | 0 | 26 |
| 1 | Arkansas | 1 | 0 | 0 | 2 |
| 2 | Louisiana | 2 | 2 | 6 | 15 |
| 3 | Mississippi | 2 | 0 | 0 | 0 |
| 4 | Nevada | 1 | 0 | 0 | 1 |
| 5 | New Jersey | 7 | 1 | 11 | 30 |
| 6 | New Mexico | 2 | 3 | 0 | 133 |
| 7 | Tennessee | 5 | 0 | 0 | 4 |
| 8 | Wyoming | 0 | 5 | 0 | 13 |

Fig. 12: State wise License Suspension Distribution

Almost all of the AOI notices are from Arizona state in our dataset. Figure 13 shows the count of AOI notices in different counties of Arizona. Maricopa, which is the fourth largest county in the USA, bigger than 23 states in population, has the highest number of AOI notices in all categories.

| | county | profit | non-profit | others |
|---|---|---|---|---|
| 0 | Apache | 7 | 15 | 7 |
| 1 | Cochise | 32 | 41 | 12 |
| 2 | Coconino | 99 | 60 | 30 |
| 3 | Gila | 33 | 31 | 12 |
| 4 | Graham | 1 | 3 | 13 |
| 5 | Greenlee | 0 | 1 | 3 |
| 6 | La Paz | 5 | 12 | 6 |
| 7 | Maricopa | 3539 | 1508 | 1894 |
| 8 | Mohave | 182 | 58 | 66 |
| 9 | Navajo | 21 | 24 | 15 |
| 10 | Pima | 309 | 186 | 142 |
| 11 | Pinal | 85 | 140 | 31 |
| 12 | Santa Cruz | 21 | 28 | 16 |
| 13 | Yavapai | 162 | 148 | 69 |
| 14 | Yuma | 121 | 24 | 38 |

Fig. 13: AOI count for different counties

*5) Property Notices:* The most common phrase in this category is shown in the fig 14. It appears that 'claim against' phrase is in the all the notices. This type of notices are used to pay the debt of an decedent person. Although there are many other types of property notices, but in our data-set all of them are of this particular type.

## VII. Time Series Analysis and Some Social Phenomenons

### *A. Time series*

We have performed some time series analysis on the labelled dataset to find trends and seasonality (differential patterns) on the frequency distribution of the notice categories over time. Specifically, how the trends and seasonal patterns are changing over time for each notice category, and how they compare to each other, aggregated on both the

| | phrase | count |
|---|---|---|
| 0 | claims matured or unmatured against the estate | 17460 |
| 1 | claims against the estate are required to | 15666 |
| 2 | claims against said estate to present the | 3659 |
| 3 | claims against this estate are required to | 2708 |
| 4 | claims matured or unmatured against his or | 2286 |
| 5 | claims against the estate must exhibit them | 2084 |
| 6 | claims against said estate are required to | 1481 |
| 7 | claims against the estate must be presented | 1459 |
| 8 | claims matured or unmatured against said estate | 1342 |
| 9 | claims matured or unmatured against this estate | 1077 |

Fig. 14: The count of different phrases in the Property notices

states-level and the full dataset. Such analyses reveal which states had frequent notices of which categories on respective periods, how their trends are moving over time, and lays the foundational framework for future design and analysis of time series forecasting models for predicting counts of notice categories. Aggregating socioeconomic data for the corresponding time-period of the time series, we would be able to link and/or correlate the categorical trends to the social phenomenons happened over the period.

The time series for the frequency distribution of the notice categories is provided at fig. 15, and a clearer plotting of the top-5 frequent categories are present at fig. 16. We immediately notice that there is seasonality present for the categories: each month there are either four or five jumps present. Also, there seems to be trends present for most of the categories, which might not be obvious right away. In other words, it looks like there are trends and seasonal components to these time series, which we investigate further.

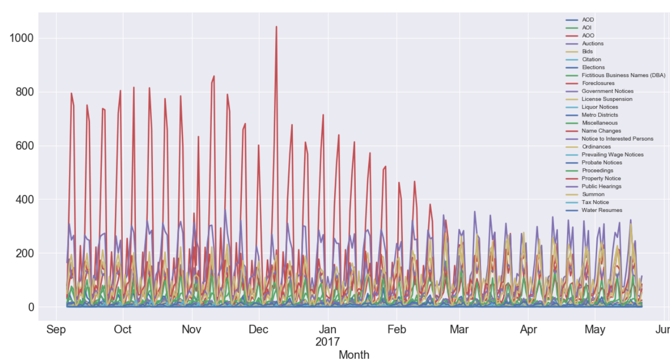



Fig. 15: Frequency distribution of the notice categories over time.

The time series for the most frequent category at the date-uniform dataset, AOO, is plotted at fig. 17 as a clear visual representative of the trend and seasonality. We find that the trend is on the decline, whereas the seasonality remains similar.

### *B. Trends in the notice categories*

To identify the trends of the categories, we take the rolling averages of their frequencies, i.e. at each time point, we compute the average of the points on either side of it where

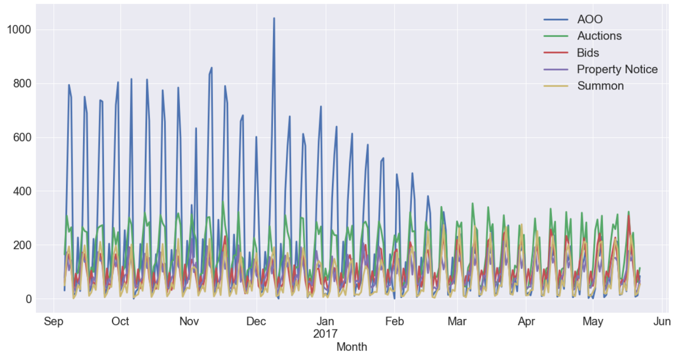



Fig. 16: Frequency distribution of the top-5 notice categories over time.

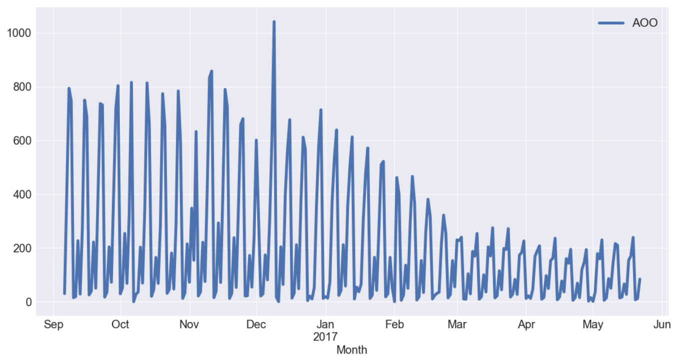



Fig. 17: Frequency distribution of the AOO (Article of organization) category over time.

the number of points to be averaged is specified by a window size. We set the window size to a bi-monthly period (60 days). For the top-5 frequent categories (AOO, Auctions, Bids, Property Notice, and Summon), the trend-plot on rolling averages is found at fig. 18. Averaging the frequencies tends to smooth out most of the noise and seasonality, and clearly interpretable trends are found out.

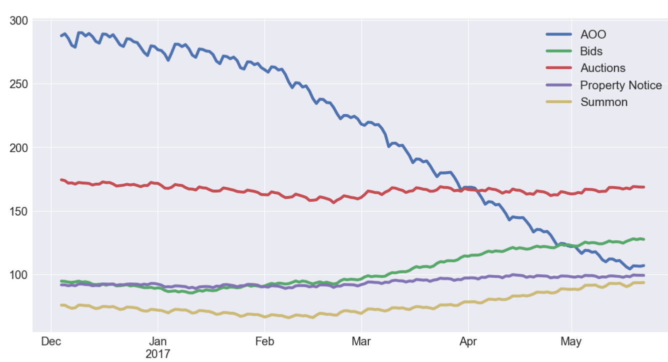



Fig. 18: Trends of the top-5 categories over time.

From fig. 18, we observe that although AOO (Article of Organization) starts out as the most frequent notice type at the turn of the business year, it diminishes towards the end of the business year. The other categories tend to remain fairly consistent in terms of occurrences, with an increasing trend in the bids category.

### C. Seasonality in the notice categories

For the full dataset, we then analyze the nature of seasonality for the notice categories, which is the nature of repetition for the frequency time series. In order to separate the seasonality, we may either remove the trend from the original signal, or perform differencing, where the differences are computed between successive data points as representatives of the nature of repetition. The seasonality plots of the top-5 most frequent categories is found at fig. 19. Now that we have removed much of the trend, we can observe constant numbers (4 or 5) of peaks in each month for each category. This corresponds to the fact that there is a very strong weekly-pattern present in each of the categories, i.e. each category peaks each week exactly once, and that too are usually on some fixed weekday, as we have found out from autocorrelation analysis, following shortly.

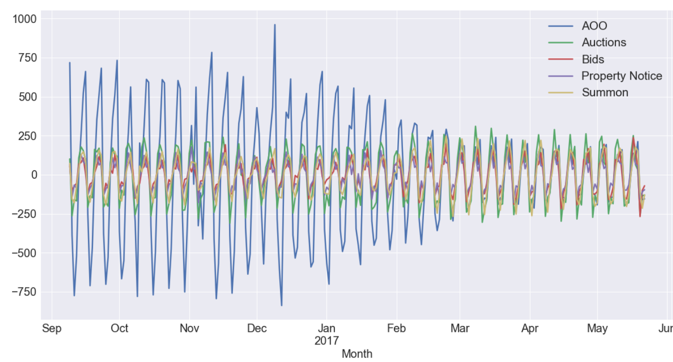



Fig. 19: Seasonality of the top-5 categories over time.

### D. Autocorrelations in frequency distribution of the notice categories

For all the notice categories, we have found an almost identical autocorrelation plot. Hence, we show just one autocorrelation analysis at fig. 20 and fig. 21 for the auctions category, as representative of all the categories. For a time series, an autocorrelation plot provides how much it is correlated with itself at various lags. Thus, if a time series approximately repeats itself every $n$-days, there would be peaks present in the autocorrelation function at every $n$ day. Observing the plot at fig. 21 with a constrained lag limit of 30 days, we find huge spikes at the autocorrelation function at 7 days: the frequency distribution of the auctions category is very strongly correlated with itself lagged/shifted 7 days. We have more peaks at 14, 21, etc., i.e. at other multiples of 7, with progressively lesser magnitudes. The maximum spike at lag 0 corresponds to the correlation of 1 for the time series with itself.

Note that, we have observed identical phenomenon for all the considerable notice categories.

We have also performed similar time-series analysis on the notice categories grouping the notices by their corresponding states; then performing state-level analysis. The plottings are not provided here due to volume.

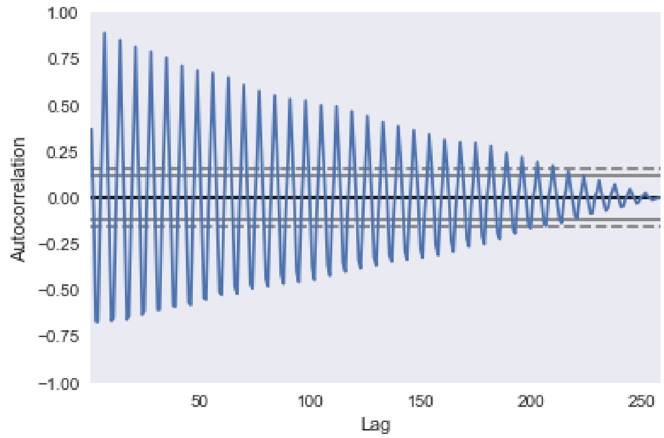



Fig. 20: Autocorrelation of the time series for the auctions category.

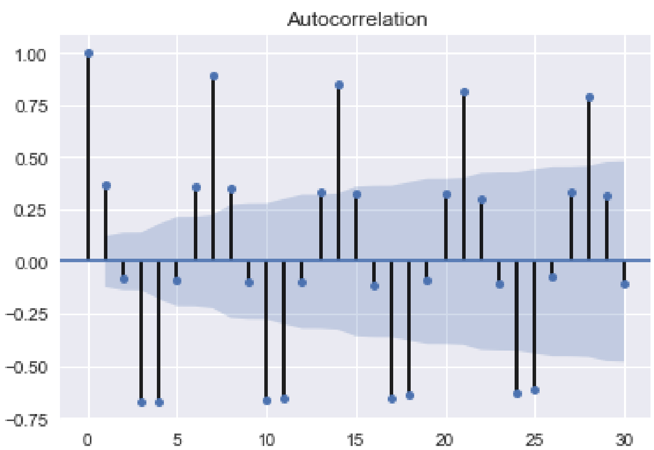



Fig. 21: Autocorrelation of the time series for the auctions category.

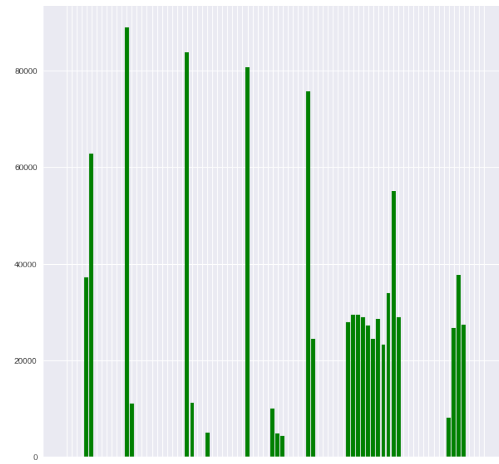



Fig. 22: Frequency distribution of notice dates for the full 05/2012 - 08/2018 range.

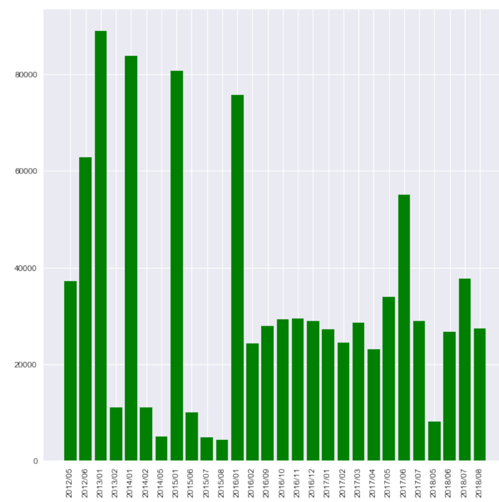



Fig. 23: Frequency distribution of the notice dates for just the months present at the dataset.

### E. *Distribution of dates over the larger dataset*

The larger public notice dataset provided has a very uneven distribution regarding dates. For a total of 7 years (2012 - 2018), the dataset contains notices spanning just 28 months. The frequency distribution of the months spanning over the total 84 months time-range, and spanning just over the 28 months actually present existent at the dataset are provided at the Figs. 22 and 23 respectively.

As evident from Fig. 22, the dates distribution is very much uneven for much of the full time-range, peaking at a few number of months, and totally absent at most others. Thus, correlating certain social phenomenons with this dataset would generate spurious results. Hence, we have focused on the more uniform portion of the data (2016/09 - 2017/05) to relate with socio-economic data.

### F. *Some social phenomenon*

For each county, the notice counts for various categories have been aggregated. So each county provides a vector of counts for the categories. We also have the socio-economic data formatted on the county-level. Thus, the frequency vectors of the categories per county is then matched with county-level socio-economic data to find interesting social phenomenon. A few of the interesting finds in county-level spatial correlations are noted below.

- Poverty has a strong correlation with bids count. The Spearman correlation coefficient is 0.64. The corresponding scatter plot showing the correlation along with the confidence intervals for bids vs total poverty on county-level is provided at Fig. 24.
- Unemployment is strongly correlated with auctions, with a Spearman correlation coefficient of 0.7. The corresponding scatter plot on county-level is provided at Fig. 25.
- Number of personnel with high school degrees has a strong correlation with the count of public hearings. The Spearman correlation coefficient is 0.64. The corresponding scatter plot on county-level is provided at

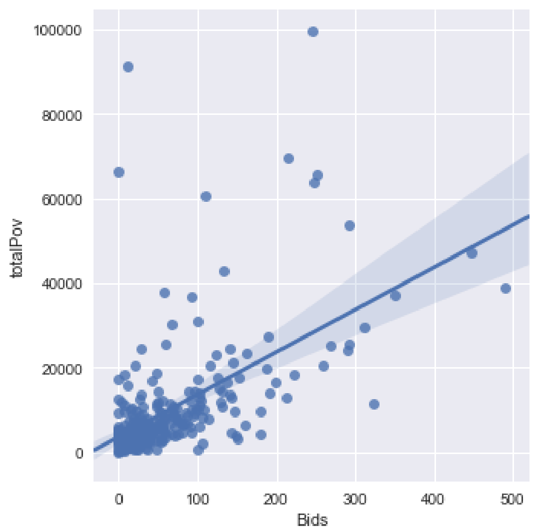



Fig. 24: Scatter plot showing bids vs. total poverty per county.

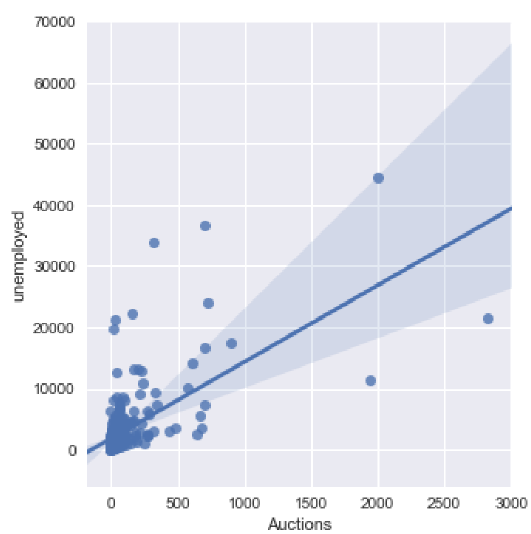



Fig. 25: Scatter plot showing auctions count vs. total unemployment per county.

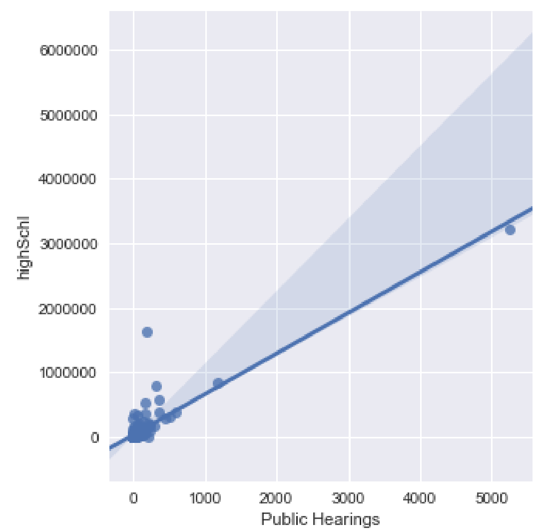



Fig. 26: Scatter plot showing public hearings vs. total high school degree per county.

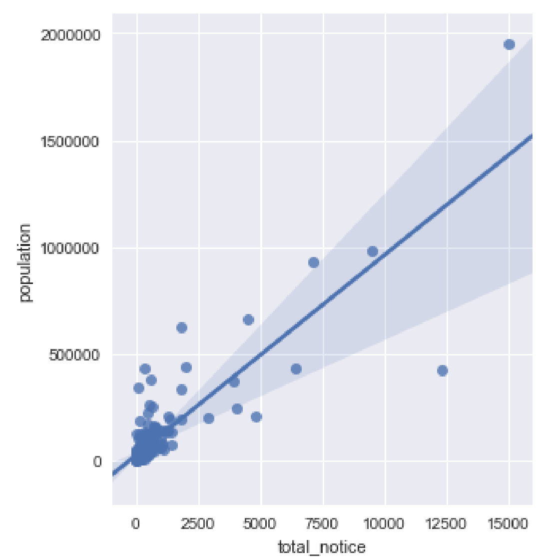



Fig. 27: Scatter plot showing total notice count vs. total population per county.

Fig. 26.

- To confirm the intuitive idea that the total notice count per county should be strongly correlated with the county-population, we tested the hypothesis and found it true, with a Spearman correlation coefficient of 0.73. The scatter plot is provided at Fig. 27.

## VIII. Conclusion

Public notices are rich source of data that can provide insights in various perspective. Classification of these notices is the first step towards finding any kind of insights from these notices. Empirical results on our classification model shows, it can correctly identify notice category with high accuracy, precision, and recall. One constraint we had that, we only had data of eight states with uneven sample from each year. This made it hard to infer strong social trends using these notices. More even data across all the fifty states would have helped us in finding more insightful patterns. One possible future work would be to incorporate a more comprehensible dataset at first for more insightful analysis. We have identified few possible hypothesis to test in future such as find out how the type of business declared in AOO notices relates to the geographical information of counties and states, what are the causes behind name changes notices, Why people lose their license in different fields and so on. Working on these hypothesis would be our next step.

## References

[1] Publicx Notice Resource Center, "Newspaper notice saves county tens of thousands of dollars,"

2018. [Online]. Available: https://www.pnrc.net/2018/10/02/newspaper-notice-saves-county-tens-of-thousands-of-dollars/
[2] "sklearn.feature_extraction.text.tfidfvectorizer scikit-learn 0.20.1 documentation," https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html, (Accessed on 12/06/2018).
[3] "Python api reference xgboost 0.81 documentation," https://xgboost.readthedocs.io/en/latest/python/python_api.html, (Accessed on 12/06/2018).
[4] "Vehicle identification numbers (vin codes)/world manufacturer identifier (wmi) - wikibooks, open books for an open world," https://en.wikibooks.org/wiki/Vehicle_Identification_Numbers_(VIN_codes)/World_Manufacturer_Identifier_(WMI), (Accessed on 12/06/2018).