

Tax-exempt Organizations in the Mid-Atlantic & Great Lakes region

Group 4

Peipei Gao, Yike Liu, Jennifer Nguyen, Yue Zheng

This presentation will tell you...



The most and least popular tax-exempt activities in the region.



States that have the most profitable organizations operating in the top 2 popular activities



Most profitable not-for-profits in 2018 and their activities



Clues about interests and priorities of people in this region



When a non-profit should submit its tax-exempt application to Internal Revenue Service (IRS)

Why this Dataset?



Better understand the overall landscape of tax-exempt organizations in this region (activities, locations, financials, etc.)



Assess whether US tax-exempt orgs are doing well financially => clues about the quality of management and the preferences of American donors.



Prefer considerably tidy data with clear data description and reasonable dimension



Suitable for dimensional modeling and data analysis

Some Concerns & Expected Modifications



Large dataset => slow
processing time



Filter out attributes that are
not relevant to our questions



Many categorical attributes to
describe types of activities in
slightly different ways =>
only keep the most relevant
ones

Data Wrangling

- Filter necessary attributes based on our business questions and drop those unrelated columns (csvkit)
- Wrangling in Trifacta

- 1 Extract ending at '-' from ZIP
- 2 Rename ZIP1 to 'Five-digits ZIP'
- 3 Delete rows where SUBSECTION is one of SUBSECTION = 91 && SUBSECTION = 82 && SUBSECTION = 0
- 4 Replace matches of '00' from CLASSIFICATION with ''
- 5 Replace matches of '0' from CLASSIFICATION with ''
- 6 Split RULING at 1 positions
- 7 Rename RULING1 to 'RUNLING_YEAR'
- 8 Rename RULING2 to 'RULING_MONTH'
- 9 Replace mismatched values in RUNLING_YEAR with NULL
- 10 Replace matches of '00' from RULING_MONTH with ''
- 11 Replace missing values in RULING_MONTH with NULL
- 12 Replace cells which equal '0' in DEDUCTIBILITY with ''
- 13 Replace missing values in DEDUCTIBILITY with NULL
- 14 Split ACTIVITY 2 times starting after position 3 every 3 characters
- 15 Replace missing values in TAX_PERIOD with NULL
- 16 Replace missing values in INCOME_AMT with NULL
- 17 Replace missing values in REVENUE_AMT with NULL

Dimensional Modeling

- From this:

Exempt_Organization
EIN
NAME
STREET
CITY
STATE
ZIP
FIVE_DIGIT_ZIP
SUBSECTION
CLASSIFICATION
RULING_YEAR
RULING_MONTH
DEDUCTIBILITY
ACTIVITY1
ACTIVITY2
ACTIVITY3
STATUS
TAX_PERIOD
INCOME_CD
INCOME_AMT
REVENUE_AMT

Dimensional Modeling

- To this:
- Upload and examine the final dataset by Unix command-line tools
- Final dataset: 644,942 records and 20 attributes

Basic_Information
<u>Key</u>
EIN
NAME
SUBSECTION
CLASSIFICATION
DEDUCTIBILITY
ACTIVITY1
ACTIVITY2
ACTIVITY3
STATUS

Address
<u>Key</u>
STREET
CITY
STATE
ZIP
FIVE_DIGIT_ZIP

Exempt_Organization
basic_information_key
time_key
address_key
income_cd_key
INCOME_AMT
REVENUE_AMT

Time
<u>Key</u>
RULING_YEAR
RULING_MONTH
TAX_PERIOD

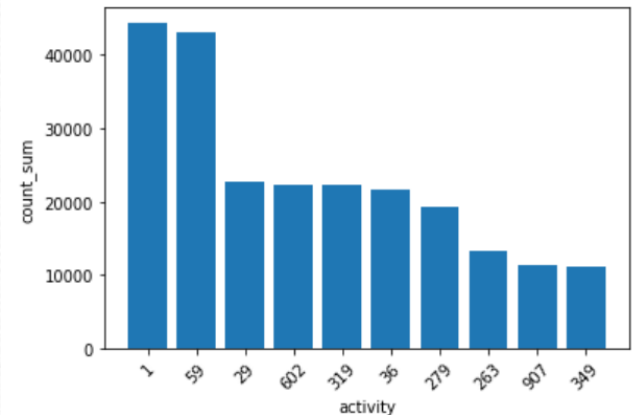
Income_classification
<u>Key</u>
INCOME_CD

Business Question #1:

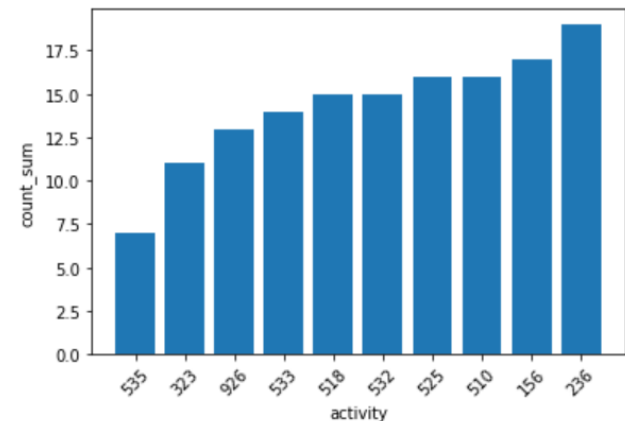
Which activities have the highest and lowest number of organizations? Among the most popular activities, what is the average income of the exempt organizations in each state?

- The five most popular activities
 - Activity 001 (Church, synagogue, etc., belong to religion activity) with the count of 44298
 - Activity 059 (Other school related activities, belong to education activity) with the count of 43079
 - Activity 029 (Other religious activities, belong to religion activity)
 - Activity 602 (Gifts, grants, or loans to other originations)
 - Activity 319 (Sports, Athletic Recreational and Social Activities)
- The two most unpopular activities
 - Activity 513 (Weapons systems, belong to Advocacy activity) with the count of 5
 - Activity 535 (Racial integration, belong to Advocacy activity) with the count of 7

The most popular activities



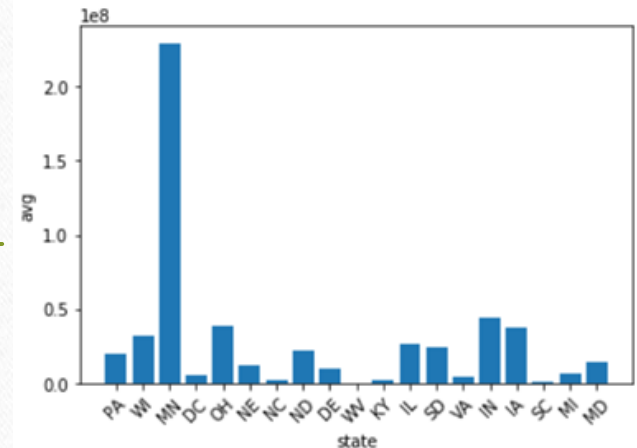
The least popular activities



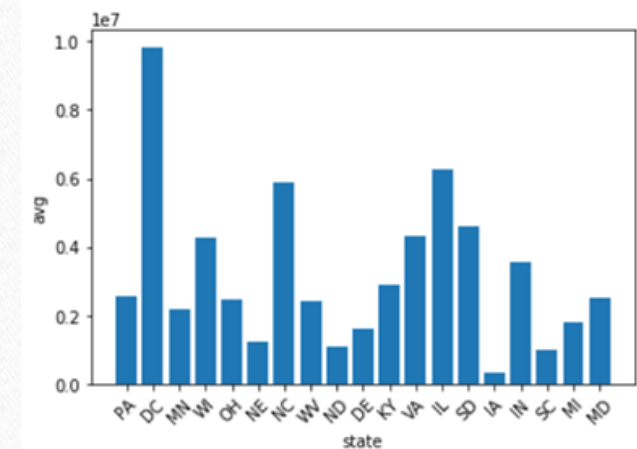
Among the two most popular activities, what is the average income of the exempt organizations in each state?

- Activity 001 (Church, synagogue, etc.): the variance of average net profits for these states is large
- Activity 059 (Other school related activities): DC has the highest average net profits in the Mid-Atlantic and Great Lakes Area

Average income amount for Activity 001 (Church, synagogue, etc.)

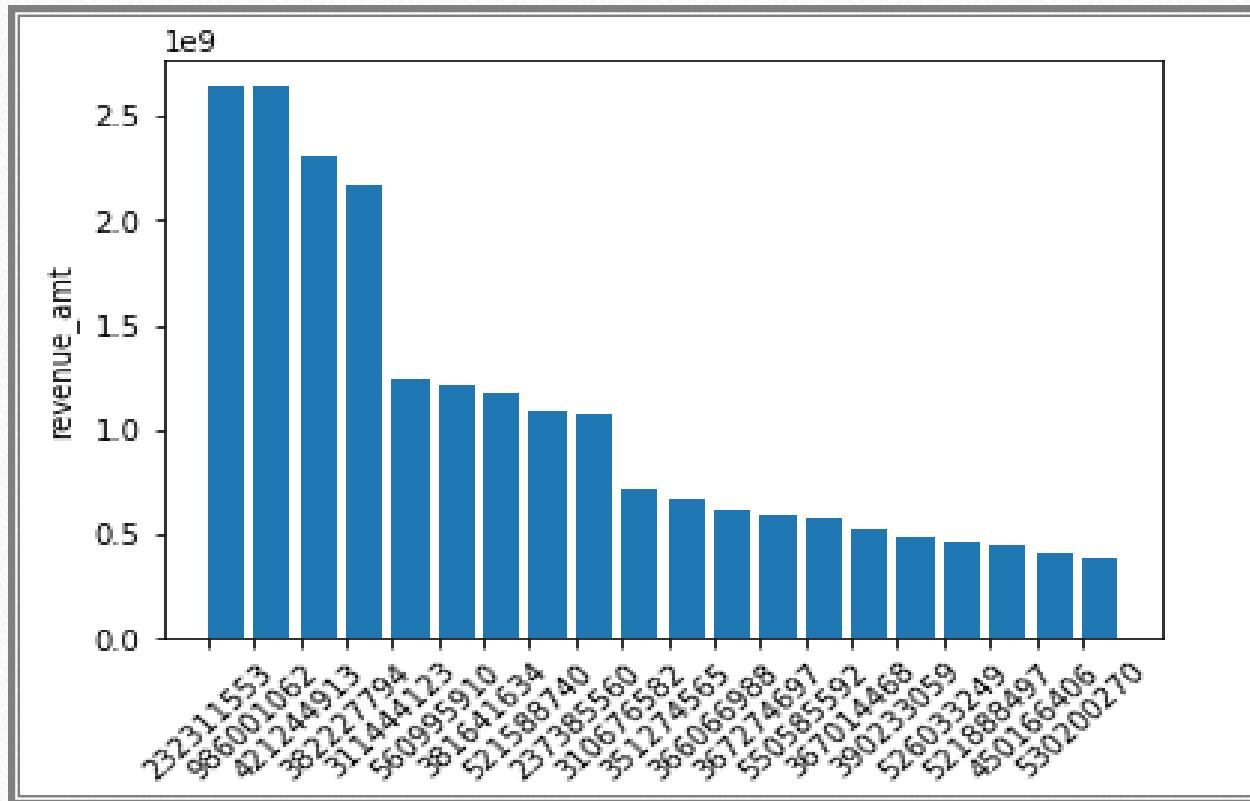


Average income amount for Activity 059 (Other school related activities)



Business question #2:

Which tax-exempt organizations to which contributions are not deductible have the highest revenue in the 2018 tax period?



Top 20 NFPOs with the highest revenue in 2018

The top 4 NFPOs

GEISINGER HEALTH
PLAN

CHANCELLOR
MASTERS & SCHOLARS
OF THE UNIVERSITY
OF OXFORD

PRAIRIE
MEADOWS RACE
TRACK & CASINO INC

MICHIGAN
CATASTROPHIC
CLAIMS ASSOCIATION
U S:



Healthcare in the top priority. Horse racing and gambling are popular.



Good business strategies must play a role in Geisinger's #1 ranking.



There is a high demand to study at Oxford. Oxford's alumni are generous.



The automotive industry in Michigan has vast resources.

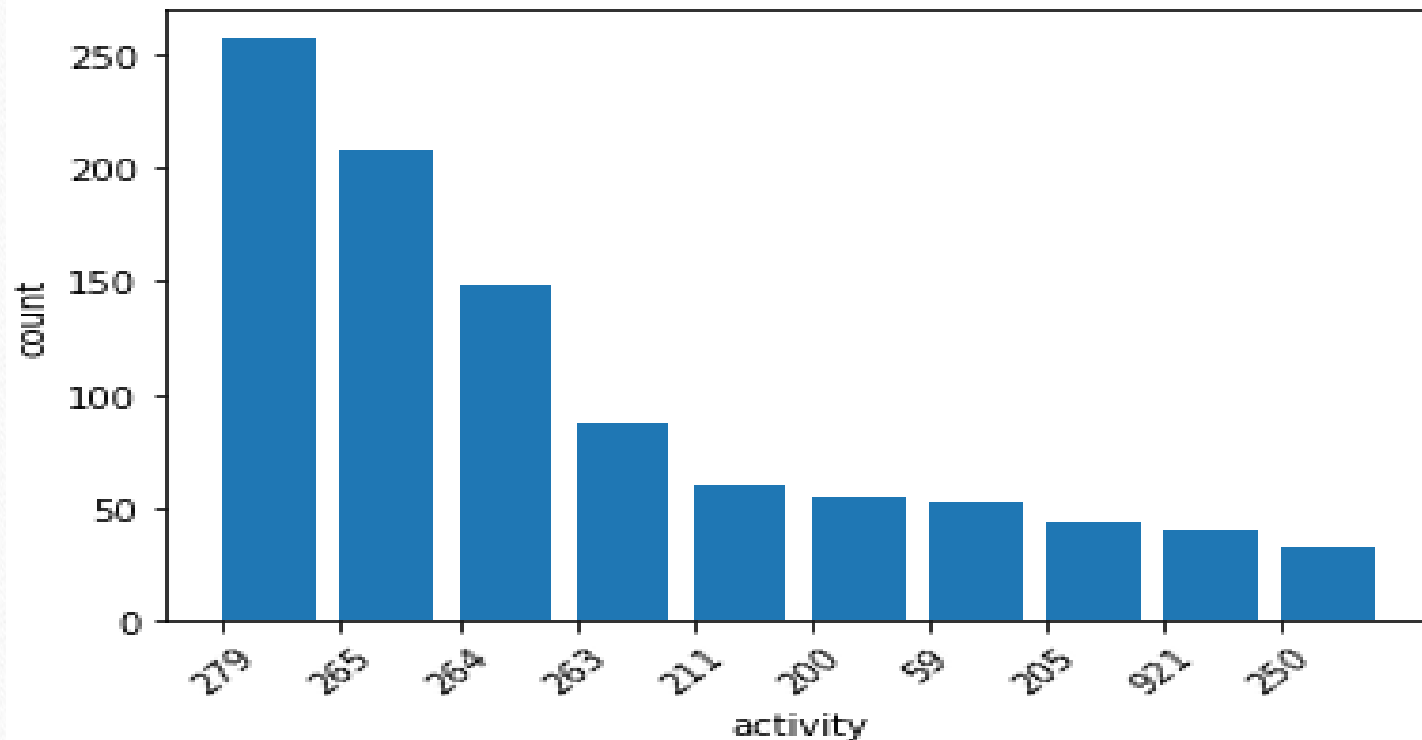
Most popular activities for top 1000 NFPOs:

#1 ~ #4: under Employee of Membership Benefit Organizations

#5 : under Business and Professional Organizations

#6: chamber of commerce, business league, etc.

#7: Schools, Colleges and Related Activities

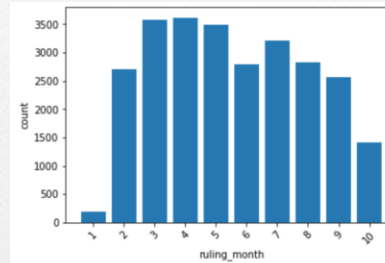


Business Question #3:

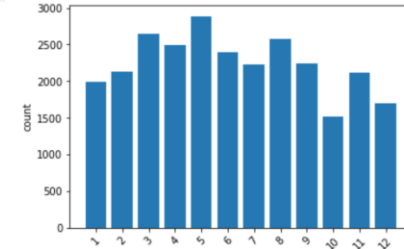
Which months have the highest number of organizations that first received the exempt status?

If there is a trend of high possibility to be exempted on certain months?

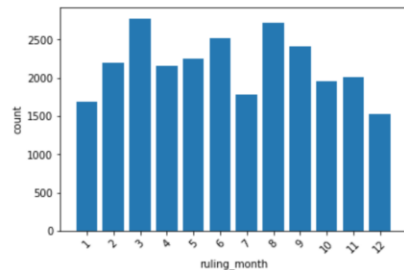
- Many orgs tend to be exempt on Feb, Mar, Apr, Aug, and Sep
- Least possibility of receiving exempt status: Jan, Nov, Dec
- **Suggestion:** apply for tax-exempt from from late Sep to Jan



2019

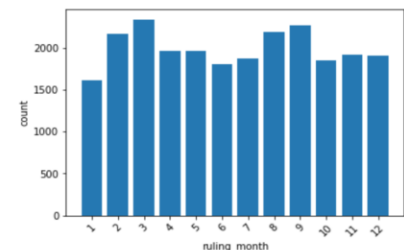
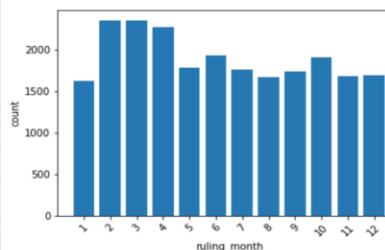


2018



2017

2015



2016

QUESTION TIME!



[Image source](#)