

TaxData 2



Maleehah Zafar

CS 4624 Multimedia, Hypertext and Information Access

Dr. Fox

Virginia Tech, Blacksburg VA 24061

12/13/2022

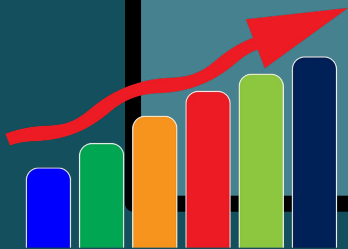


Table of Contents

01 Project Overview

02 Important Terms

03 Deliverables

04 Timeline

05 Tools

06 Work Completed

07 Future Work

08 Challenges

09 Solutions

10 Lessons Learned

11 Acknowledgements

12 References

Project Overview

Converting 20 years of tax data in flat file and XML file format into a readable format and run some preliminary analysis

Form 990-EZ	Short Form		OMB No. 1545-0047
	Return of Organization Exempt From Income Tax		2019
Under section 501(c), 527, or 4947(a)(1) of the Internal Revenue Code (except private foundations)		Open to Public Inspection	
<p>▶ Do not enter social security numbers on this form, as it may be made public.</p> <p>▶ Go to www.irs.gov/Form990EZ for instructions and the latest information.</p>			
A For the 2019 calendar year, or tax year beginning , 2019, and ending , 20			
B Check if applicable: <input type="checkbox"/> Address change <input type="checkbox"/> Name change <input type="checkbox"/> Initial return <input type="checkbox"/> Final return/terminated <input type="checkbox"/> Amended return <input type="checkbox"/> Application pending	C Name of organization Number and street (or P.O. box if mail is not delivered to street address) Room/suite City or town, state or province, country, and ZIP or foreign postal code		D Employer identification number E Telephone number F Group Exemption Number ▶
	G Accounting Method: <input type="checkbox"/> Cash <input type="checkbox"/> Accrual Other (specify) ▶		H Check <input type="checkbox"/> if the organization is not required to attach Schedule B (Form 990, 990-EZ, or 990-PF).
	I Website: ▶		
	J Tax-exempt status (check only one) – <input type="checkbox"/> 501(c)(3) <input type="checkbox"/> 501(c) () ◀ (insert no.) <input type="checkbox"/> 4947(a)(1) or <input type="checkbox"/> 527		
K Form of organization: <input type="checkbox"/> Corporation <input type="checkbox"/> Trust <input type="checkbox"/> Association <input type="checkbox"/> Other			
L Add lines 5b, 6c, and 7b to line 9 to determine gross receipts. If gross receipts are \$200,000 or more, or if total assets (Part II, column (B)) are \$500,000 or more, file Form 990 instead of Form 990-EZ. ▶ \$			
Part I Revenue, Expenses, and Changes in Net Assets or Fund Balances (see the instructions for Part I) Check if the organization used Schedule O to respond to any question in this Part I <input type="checkbox"/>			

Form 990 (Rev. January 2020)	Return of Organization Exempt From Income Tax		OMB No. 1545-0047
	Under section 501(c), 527, or 4947(a)(1) of the Internal Revenue Code (except private foundations)		2019
Department of the Treasury Internal Revenue Service		Open to Public Inspection	
<p>▶ Do not enter social security numbers on this form as it may be made public.</p> <p>▶ Go to www.irs.gov/Form990 for instructions and the latest information.</p>			
A For the 2019 calendar year, or tax year beginning , 2019, and ending , 20			
B Check if applicable: <input type="checkbox"/> Address change <input type="checkbox"/> Name change <input type="checkbox"/> Initial return <input type="checkbox"/> Final return/terminated <input type="checkbox"/> Amended return <input type="checkbox"/> Application pending	C Name of organization Doing business as Number and street (or P.O. box if mail is not delivered to street address) Room/suite City or town, state or province, country, and ZIP or foreign postal code		D Employer identification number E Telephone number G Gross receipts \$
	F Name and address of principal officer:		H(a) Is this a group return for subsidiaries? <input type="checkbox"/> Yes <input type="checkbox"/> No H(b) Are all subsidiaries included? <input type="checkbox"/> Yes <input type="checkbox"/> No If "No," attach a list. (see instructions)
	J Tax-exempt status: <input type="checkbox"/> 501(c)(3) <input type="checkbox"/> 501(c) () ◀ (insert no.) <input type="checkbox"/> 4947(a)(1) or <input type="checkbox"/> 527		H(c) Group exemption number ▶
	I Website: ▶		
K Form of organization: <input type="checkbox"/> Corporation <input type="checkbox"/> Trust <input type="checkbox"/> Association <input type="checkbox"/> Other ▶		L Year of formation:	M State of legal domicile:
Part I Summary			

Important Terms

Form 990 and Form 990-EZ

Tax-exempt organizations, nonexempt charitable trusts, and section 527 political organizations file this form. Depending on the amount the organization earned, they will fill out either Form 990 or Form 990-EZ.

Schedule A-O, R

If an organization fills out Form 990, then they may be required to fill out one of the many Schedules which gives more information about the organization.

- Schedule J gives information about compensation of officers at organizations

Deliverables



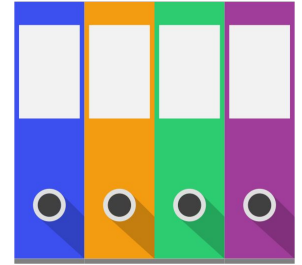
Program a universal parser for all types of forms



Create a new database with at least IRS Form 990 and Schedule J



Visualize the data in a meaningful way



Add comments and instructions in the repository

Timeline

Initial contact with
Dr. Zach
Make
improvements to
previous work

October
2022

Test parser on
other tax forms
Data Analysis

December
2022

September
2022

Parse 1 Schedule
J file
Parse a directory
of Schedule J files

November
2022

Add
documentation

Tools

Programming Language

Python

Python Libraries

xml, lxml, sqlite3, pandas

Storage System

Advanced Research Computing (ARC)



Work Completed

Universal Parser

- Created a parser that can parse IRS 990, IRS 990-EZ, IRS 990-PF, and Schedules A-O/R
- Tested the parser on a subset of files in order to ensure correct functionality

New Database

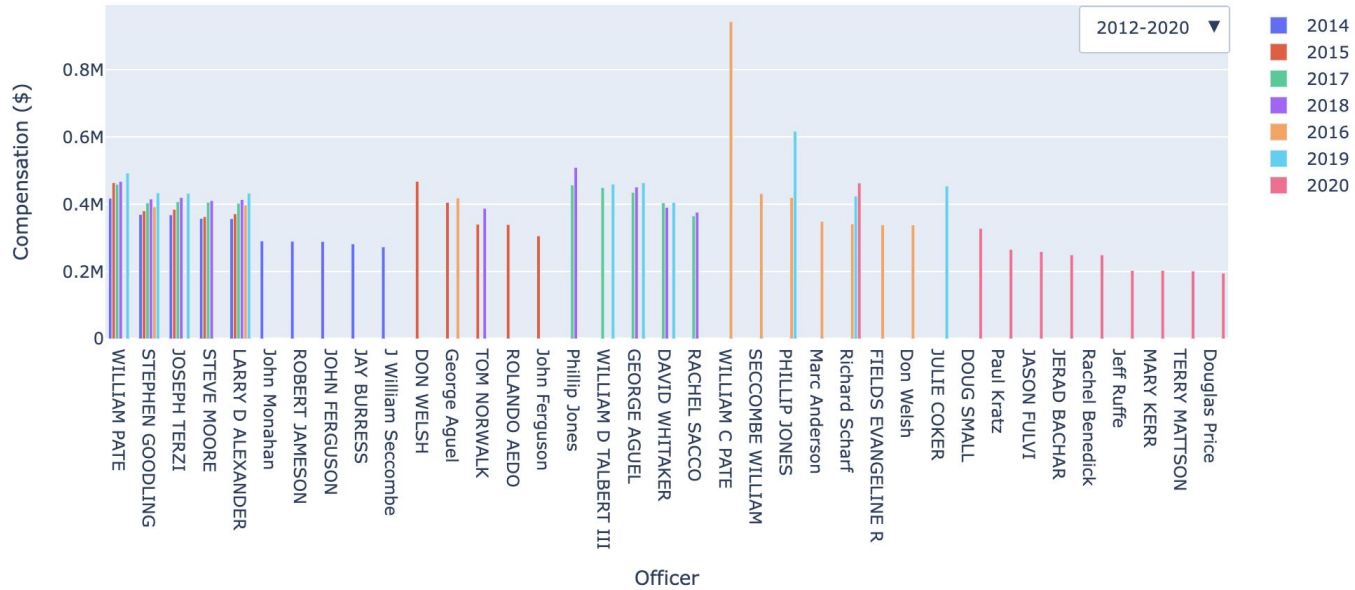
- Created a new database called NPTR (nonprofit tax records) in ARC
- Uploaded data from Form 990, Form 990-EZ, and Schedule J

Documentation

- Added a detailed README in git repository

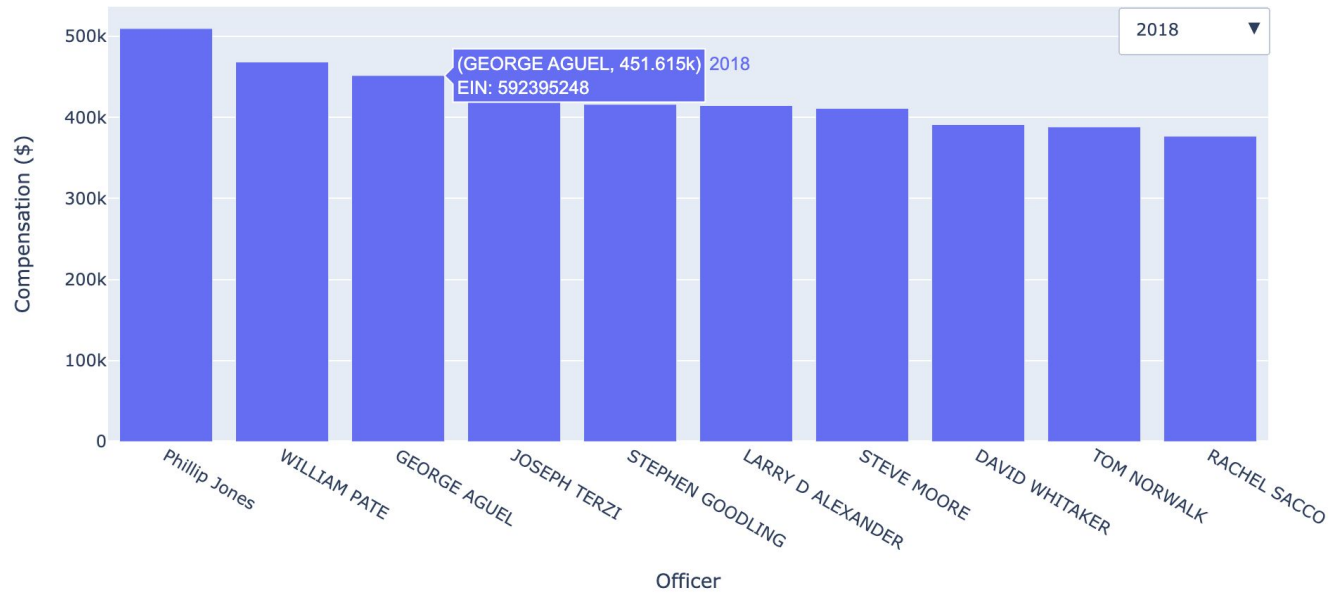
Work Completed

Visualization: Top 10 Highest Paid Officers in Tourism Offices



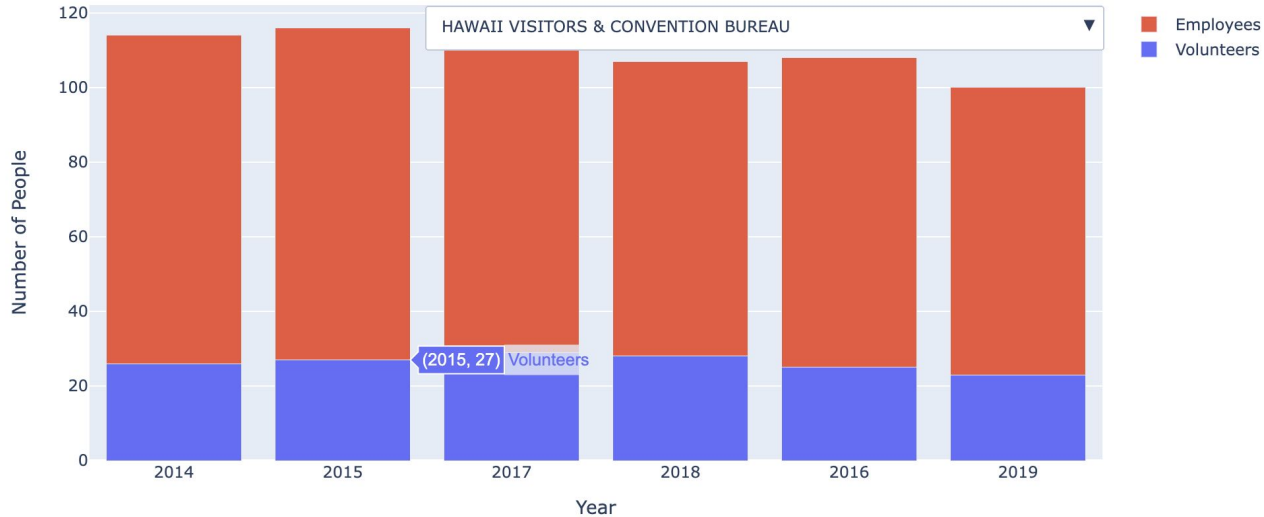
Work Completed

Visualization: Top 10 Highest Paid Officers in Tourism Offices



Work Completed

Visualization: Number of Employees and Volunteers in Tourism Offices



Future Work

Update Database

Parse the remaining files which are IRS 990-PF and the other Schedules

Nonprofit Tax Records Website

Design a website where a user can search for specific information about nonprofit organizations

- Searching for a nonprofit organizations Schedule J file from 2015

ML Algorithm

Create a machine learning algorithm that can predict which nonprofit organization is a tourism office

Challenges

Minimal Documentation

The team in Spring 2022 did not document their code well

Ineffective Database

The database created in Spring 2022 was not well structured

Duplicate Data

While working on the visualizations, it was discovered that there is duplicate data in the new database

Solutions

Minimal Documentation

Try to document what they have done and document my code extensively

Ineffective Database

Change deliverables of project to create a universal parser and create a new database

Duplicate Data

Create a python script that goes into the database and removes any duplicate data

Lessons Learned

Documentation

Having well-documented code is vital to any project because other people need to know what you did

Flexibility

Being able to pivot in a different direction if the original objectives of a project are unattainable

Acknowledgements

Dr. Florian Zach

an assistant professor in the Pamplin College of Business, specifically the Howard Feiertag Department of Hospitality and Tourism Management, at Virginia Tech

Huihui Zhang

a PhD student in the Pamplin College of Business

Dr. Edward A. Fox

References

<https://aws.amazon.com/public-datasets/irs-990/>

<https://www.irs.gov/charities-non-profits/form-990-series-downloads>

<https://images.google.com/>

<https://www.irs.gov/forms-pubs/about-form-990>

<https://arc.vt.edu/>