Tax Data Project

Nicky Huynh, Kaleb Barker, Long Nguyen, Jack Lacey

CS 4624 – Multimedia, Hypertext, and Information Access Capstone
Dr. Edward Fox, Virginia Tech, Blacksburg, VA 24061
Tuesday, May 3, 2022
Project Recap

- **Motivation:** Form 990 data is not easy to access or analyze
- **Goal:** Create a database of 990 tax filings
- **Phases:** Retrieval, processing, and analysis
Objectives & Deliverables

Analysis & Documentation
- Analyzing trends
- Example Queries
- User Guide

990 Database
- Tables:
  - Standard 990
  - 990-EZ
  - Officers
  - Business Master Files

Data Conversion
- CSV library of tourism office filings
Data Processing

- Over one million XML files
- Single threaded extraction of data tag while preserving path tags
- Processed using ARC job submission with Python scripts
- Output to CSV for database ingestion
990 Database – Design

- **eo_combined**
  - EIN
  - NAME
  - STREET
  - CITY
  - ZIP
  - ACTIVITY

- **tax_990**
  - Filer-EIN
  - ReturnType
  - CYTotalRevenueAmt
  - PYTotalRevenueAmt
  - CYInvestmentIncome
  - PYInvestmentIncome
    - Year

- **tax_990ez**
  - Filer-EIN
  - ReturnType
  - TotalRevenueAmt
  - TotalExpenseAmt
  - InvestmentIncomeAmt
    - Year

- **eo_states**
  - EIN
  - NAME
  - STREET
  - CITY
  - ZIP
  - ACTIVITY

- **eo_regions**
  - EIN
  - NAME
  - STREET
  - CITY
  - ZIP
  - ACTIVITY

- **officer_990**
  - PersonNm
  - TitleTxt
  - AverageHrsPerWk...
  - CompensationAmt
  - HighestCompensatedInd

- **officer_990ez**
  - PersonNm
  - TitleTxt
  - AverageHrsPerWk...
  - CompensationAmt
  - EmployeeBenefitProgramAmt
    - EIN
    - TaxYr
990 Database – Queries

```python
queryTop_1_2020 = """"SELECT DISTINCT `Filer-EIN`, `TaxYr`, `Filer-BusinessName-BusinessNameLine1`, `CYTotalRevenueAmt` FROM tax_990 WHERE `Filer-EIN` = 840382392 ORDER BY `TaxYr` Asc"
DenverMetroCYTotalRevenueAmt = pd.read_sql(queryTop_1_2020, conn)
```

<table>
<thead>
<tr>
<th>Filer-EIN</th>
<th>TaxYr</th>
<th>Filer-BusinessName-BusinessNameLine1</th>
<th>CYTotalRevenueAmt</th>
</tr>
</thead>
<tbody>
<tr>
<td>840382392.0</td>
<td>2014.0</td>
<td>Denver Metro Convention &amp; Visitors Bureau</td>
<td>23171751.0</td>
</tr>
<tr>
<td>840382392.0</td>
<td>2015.0</td>
<td>Denver Metro Convention &amp; Visitors</td>
<td>24373195.0</td>
</tr>
<tr>
<td>840382392.0</td>
<td>2016.0</td>
<td>Denver Metro Convention &amp; Visitors</td>
<td>26060785.0</td>
</tr>
<tr>
<td>840382392.0</td>
<td>2017.0</td>
<td>Denver Metro Convention &amp; Visitors</td>
<td>32213522.0</td>
</tr>
<tr>
<td>840382392.0</td>
<td>2018.0</td>
<td>Denver Metro Convention &amp; Visitors</td>
<td>39623731.0</td>
</tr>
<tr>
<td>840382392.0</td>
<td>2019.0</td>
<td>Denver Metro Convention &amp; Visitors</td>
<td>40992367.0</td>
</tr>
<tr>
<td>840382392.0</td>
<td>2020.0</td>
<td>Denver Metro Convention &amp; Visitors</td>
<td>12927884.0</td>
</tr>
</tbody>
</table>
### Preliminary Database Analysis

**Top 6 Tourism Offices in 2019**

- Mostly large cities in large population states
- Top 30 is relatively constant over from 2013-2019

<table>
<thead>
<tr>
<th>Filer-BusinessName-BusinessNameLine1</th>
<th>Filer-USAddress-CityNm</th>
<th>Filer-USAddress-StateAbbr</th>
<th>CYTotalRevenueAmt</th>
</tr>
</thead>
<tbody>
<tr>
<td>OrlandoOrange County Convention &amp; Visitors Bureau Inc</td>
<td>Orlando</td>
<td>FL</td>
<td>80,519,755</td>
</tr>
<tr>
<td>FLORIDA TOURISM INDUSTRY MARKETING CORP INC</td>
<td>TALLAHASSEE</td>
<td>FL</td>
<td>48,182,560</td>
</tr>
<tr>
<td>SAN DIEGO CONVENTION AND TOURIST BUREAU</td>
<td>SAN DIEGO</td>
<td>CA</td>
<td>41,202,845</td>
</tr>
<tr>
<td>Denver Metro Convention &amp; Visitors Bureau</td>
<td>Denver</td>
<td>CO</td>
<td>40,992,367</td>
</tr>
<tr>
<td>ATLANTA CONVENTION AND VISITORS BUREAU INC</td>
<td>ATLANTA</td>
<td>GA</td>
<td>35,729,622</td>
</tr>
<tr>
<td>HAWAII VISITORS &amp; CONVENTION BUREAU</td>
<td>HONOLULU</td>
<td>HI</td>
<td>34,731,961</td>
</tr>
</tbody>
</table>
Preliminary Database Analysis

Top 6 Tourism Offices in 2020

- Smaller population states
- No Florida tourism offices are in top 30
- Tallahassee and Orlando offices did not submit tax form for 2020

<table>
<thead>
<tr>
<th>Filer-BusinessName-BusinessNameLine1</th>
<th>Filer-USAddress-CityNm</th>
<th>Filer-USAddress-StateAbbreviationCd</th>
<th>CYTotalRevenueAmt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denver Metro Convention &amp; Visitors Bureau</td>
<td>Denver</td>
<td>CO</td>
<td>12,927,884</td>
</tr>
<tr>
<td>CONVENTION &amp; VISITORS BUREAU OF GREATER KANSAS CITY</td>
<td>KANSAS CITY</td>
<td>MO</td>
<td>9,870,365</td>
</tr>
<tr>
<td>SAINT PAUL RIVERCENTRE CONVENTION AND VISITORS AUTHORITY</td>
<td>ST PAUL</td>
<td>MN</td>
<td>9,717,353</td>
</tr>
<tr>
<td>BUFFALO NIAGARA CONVENTION &amp; VISITORS BUREAU INC</td>
<td>BUFFALO</td>
<td>NY</td>
<td>6,896,089</td>
</tr>
<tr>
<td>GRAND RAPIDS KENT COUNTY CONVENTION &amp; VISITORS BUREAU DBA EXPERIENCE GRAND RAP</td>
<td>GRAND RAPIDS</td>
<td>MI</td>
<td>6,496,669</td>
</tr>
<tr>
<td>VISITGREENVILLESC</td>
<td>GREENVILLE</td>
<td>SC</td>
<td>6,156,731</td>
</tr>
</tbody>
</table>
Sample Visualizations

Top 3 Revenues in 2020

- Denver Metro Convention & Visitors Bureau
- Visitors Bureau of Greater Kansas
- Saint Paul Rivercentre Convention

Year


Total Revenue Amount (in Million)

10 15 20 25 30 35 40 1e7
Sample Visualizations

Schools, Colleges, and Related Activities

Number of Organizations

- Fraternity and sorority
- College, trade school, etc.
- College athletic association
- Scholarships

Activity
Challenges

- Creating a unified schema for hundreds of tags across several years
- Processing Array Data: Organization Officers & Schedule Forms

Lessons Learned

- Consistent communication between team members and client
- Improve organization & planning

<table>
<thead>
<tr>
<th>PersonNm</th>
<th>TitleTxt</th>
<th>AverageHoursPerWeekRt</th>
<th>IndividualTrusteeOrDirectorInd</th>
<th>OfficerInd</th>
</tr>
</thead>
<tbody>
<tr>
<td>FINNY A VARGHESE</td>
<td>PASTOR</td>
<td>40</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>TOM KURIAN</td>
<td>SECRETARY</td>
<td>20</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>THOMAS GEORGE</td>
<td>TREASURER</td>
<td>20</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Future Work

- Appending Tax Year 2021 data
- Processing Forms 990-PF & 990-N
- Processing other Form 990 Schedules
- Creation of Web Dashboard
Acknowledgements

- Client: Professor Florian Zach, Ph.D., Howard Feiertag
  Department of Hospitality and Tourism Management,
  Pamplin College of Business
- Mentor: Dr. Edward A. Fox
- VT Advanced Research Computing (ARC)
References

AWS S3 IRS 990 Bucket: https://registry.opendata.aws/irs990/

IRS Website Form 990 Data Download:

SQLite Documentation: https://www.sqlite.org/index.html

Business Master File Information: