



By Sebastian Zapata

Final Project for
ISBA_4715

PUTTING THE PIECES OF (NON)PROFIT

Github Repository Link:
https://github.com/sebz04/SQL_Final_Project

SEBASTIAN ZAPATA



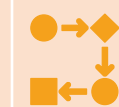
EDUCATION: SENIOR AT
LMU – ISBA MAJOR,
JOURNALISM MINOR



EXPERIENCE:
DATATHON
PARTICIPANT, SCHOOL
NEWSPAPER WRITER, SQL
+ AWS + DBT PROJECT
BUILDER



VALUES: DATA FOR
IMPACT — BUILDING
ANALYTICS SKILLS TO
TELL STORIES THAT
INFORM AND EMPOWER
THE PUBLIC.










SEGUE: ALL PATHS
CONVERGE IN MY FINAL
PROJECT — ANALYZING
NONPROFIT
ACCOUNTABILITY IN LA

WHAT IS A NONPROFIT?

an entity that operates
to benefit the public, not
to generate profits for
owners or shareholders

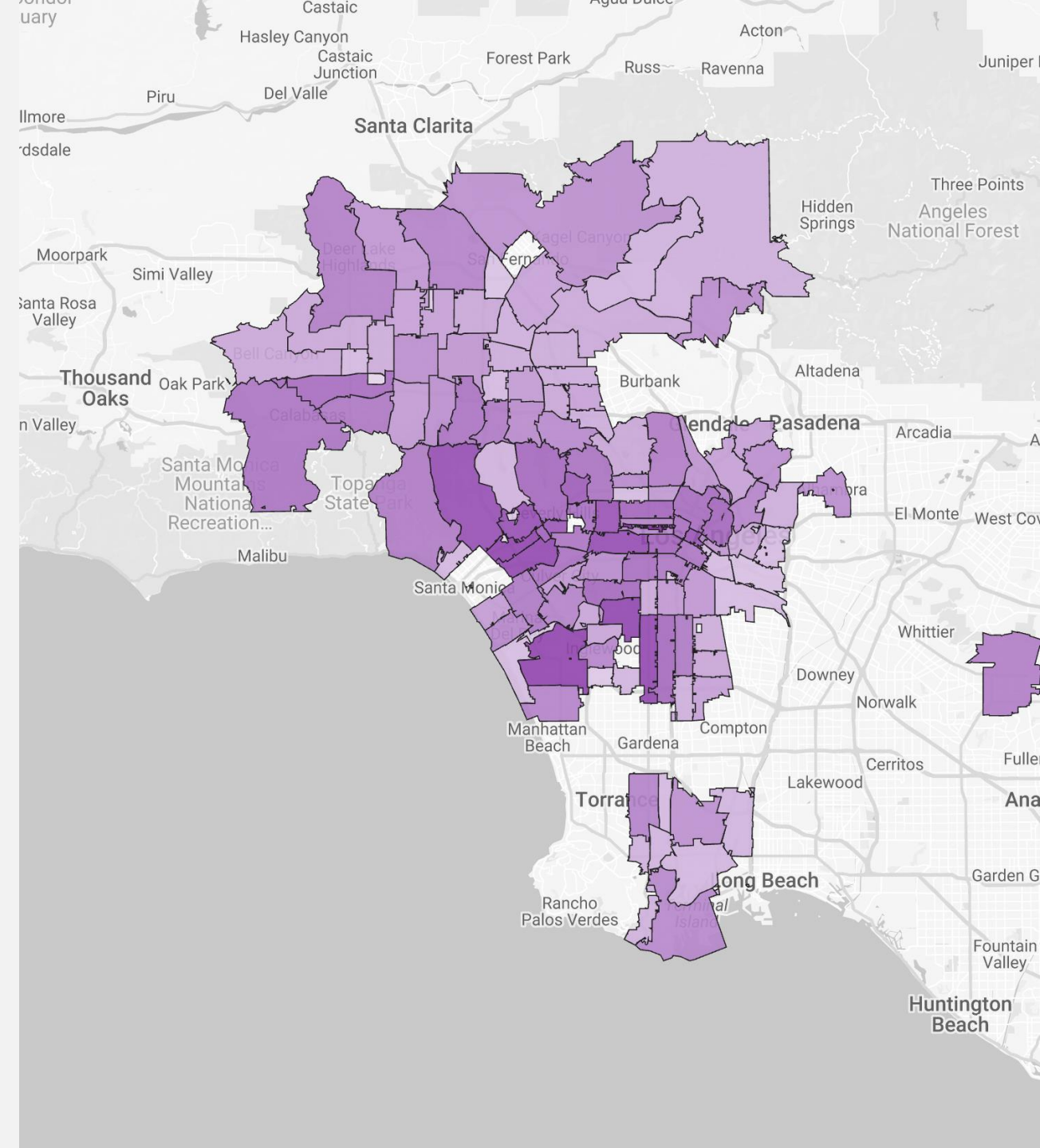
CALIFORNIA'S PREVALENCE IN NONPROFITS

State	# of Nonprofits	Highest Salary	Total Revenue
 California	209,051	\$22.7M	 \$478.4B
 New York	129,529	\$16.3M	 \$375.3B
 Pennsylvania	84,049	\$17.9M	 \$212.4B
 Texas	153,394	\$17.1M	 \$182.1B
 Massachusetts	46,058	\$20.1M	 \$181.9B
 Ohio	79,014	\$8.8M	 \$170.4B
 Illinois	78,062	\$17.1M	 \$153.4B
 Florida	116,100	\$23.8M	 \$148.3B
 Minnesota	40,789	\$8.6M	 \$110B
 Missouri	43,070	\$20.6M	 \$103.6B

- Highest Revenue among nonprofits nationwide
- Although valuable information, what does this mean for Los Angeles?
- Highlight L.A.'s role in revenue

LOS ANGELES NONPROFIT SECTOR

- Based on each organizations' financials:
- Revenue = \$53,144,146,331
- Expenses = \$50,691,586,717
- But these are only scratching the surface



PROBLEMS / STAKEHOLDERS

Who am I helping?

- *Watchdogs, journalists, donors, equity advocates*

What's the problem?

- Although financials are public, difficult to understand for non-technical journalists
- ProPublica search is not centralized to LA neighborhoods / zip codes

How am I solving it?

- *Through API and Web scraped data extraction,*
- *Cleaning and analyzing data*
- *Visualizing findings*



Built architecture & SQL models



Scraped + cleaned with Python



Created Queries in Postgres



Automated with GitHub Actions



Designed charts in Looker



Deployed data to AWS RDS

ROLES AND RESPONSIBILITIES

As a data reporter, you will:

- Plan and execute quantitative analyses — and interpret results — in support of investigative stories and visualizations
- Make phone calls, conduct interviews and do on-the-ground reporting
- Assist fellow reporters with data brainstorming, acquisition, cleaning, integrity-checking and analysis
- Bulletproof and spot-check data work done by other journalists on the team and in the newsroom
- Wrangle data, including FOIA'ing for it, scraping it, cleaning it, analyzing it and preparing it for publication
- Teach and encourage best practices and newsroom data policies

The ideal candidate will have:

- Experience working on data projects in a newsroom *or* experience doing statistical or data-scientific analysis, either in academia or in another research setting
- Excellent interpersonal skills, solid editorial judgment and the ability to write quickly and cleanly
- A solid understanding of the tools of data analysis, such as database management systems (SQL Server or PostgreSQL), statistics software (R), Microsoft Excel, etc.

DATA REPORTER -- PROPUBLICA

DATA PIPELINE



PROPUBLICA NONPROFIT EXPLORER API

Used Requests to call API in
Python Script

No API key needed – Free API

Plethora of Data on nonprofits
across the Country

Shows foundational
understanding of Database
Management Systems

TABLES FROM API CALL

Table: raw_nonprofits → staging_nonprofits

EIN	Name	Address	City	State	Zipcode	Subsecdd	Subsecdd Category	NTEE Code	NTEE Code Major	NTEE category
-----	------	---------	------	-------	---------	----------	----------------------	--------------	-----------------------	------------------

Table: raw_financial_history → staging_financial_history

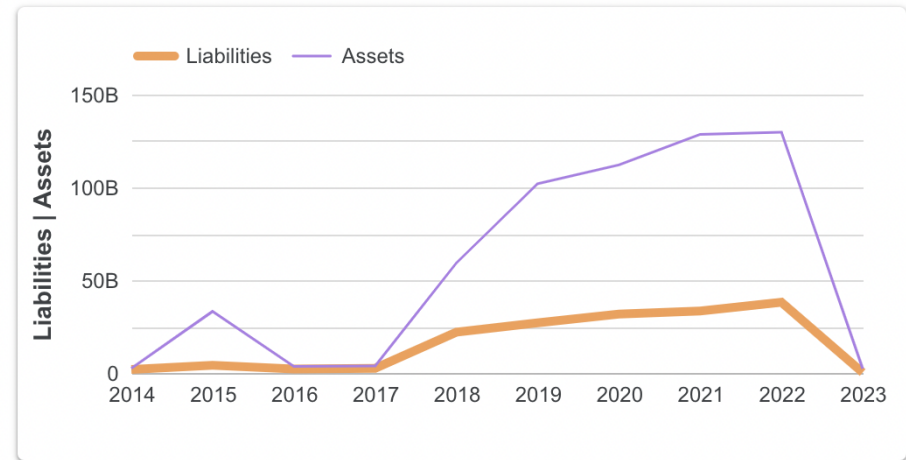
EIN	Year	Total Revenue	Total Expense	Total Assets	Total Liabilities
-----	------	------------------	------------------	-----------------	----------------------

API DESCRIPTIVE ANALYTICS QUERY

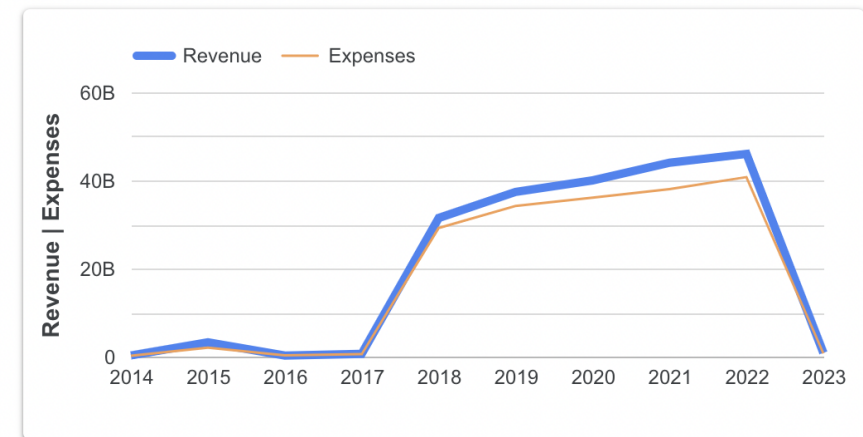
What does the nonprofit sector look
like in Los Angeles?

How financial stable are these
nonprofits?

Nonprofit Sector Financials



Nonprofit Sector Financials



DESCRIPTIVE QUERY TAKEAWAYS

• **Insight:**

- Revenue grew: \$13M → \$83M
- Loss years: 2010, 2016
- High debt ratios: 2013, 2017

Recommendation:

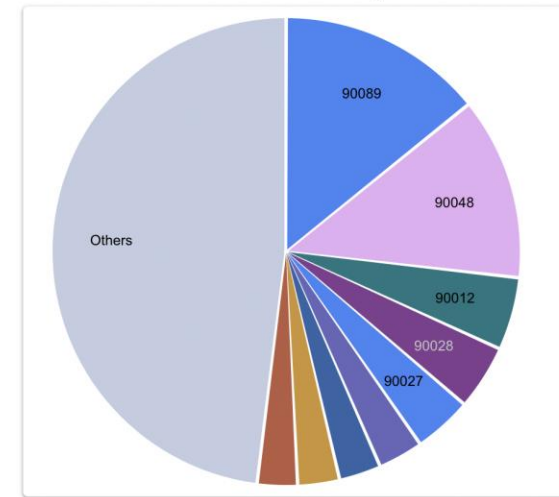
- Investigate liability spikes
- Track debt vs. program growth
- Look at nonprofit financial health throughout the years to find pattern

Prediction:

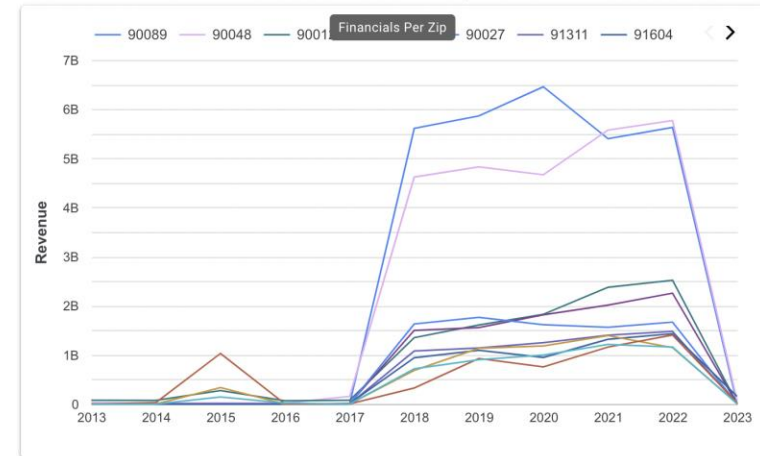
- Small nonprofits may merge
- Efficiency metrics prioritized
- Exec pay vs. impact questioned

API DIAGNOSTIC ANALYTICS QUERY

Financials Per Zip



Financials Per Zip



DIAGNOSTICS QUERY TAKEAWAYS

• **Insights:**

- Outliers: USC (90089), Cedars (90048)
- Skewed revenue, asset growth
- Growth seen across ZIP codes
- Assets rise more than revenue

Recommendation:

- Examine asset vs. revenue growth
- Identify sources: property, donors
- Study use of assets as leverage
- Link assets to program spending

Predictions:

- Refined zipcode analysis needed
- Equity-based funding gains support
- Zip-level insights will expand
- Funding tied to local needs

WEB SCRAPED DATA SOURCE

- Used Requests to get HTML of each nonprofit's webpage
- Used BeautifulSoup to parse through text
- Shows I can scrape information, clean it, analyze, and prepare the data for publication

Nonprofit Explorer

By [Andrea Suozzo](#), [Alec Glassford](#), [Ash Ngu](#) and [Brandon Roberts](#). Updated April 23, 2025

Browse millions of annual returns filed by tax-exempt organizations with ProPublica's Nonprofit Explorer. See details like executive compensation, revenue, expenses and more. Search for an organization or a person, or search the full text of filings.

Nonprofits ▼

Search for a nonprofit, keyword, city or EIN

Q

Examples: [ProPublica](#), [university](#), [Minneapolis](#) or [14-2007220](#)

Advanced

Looking for filings for political nonprofits? **[Try the new 527 Explorer.](#)**

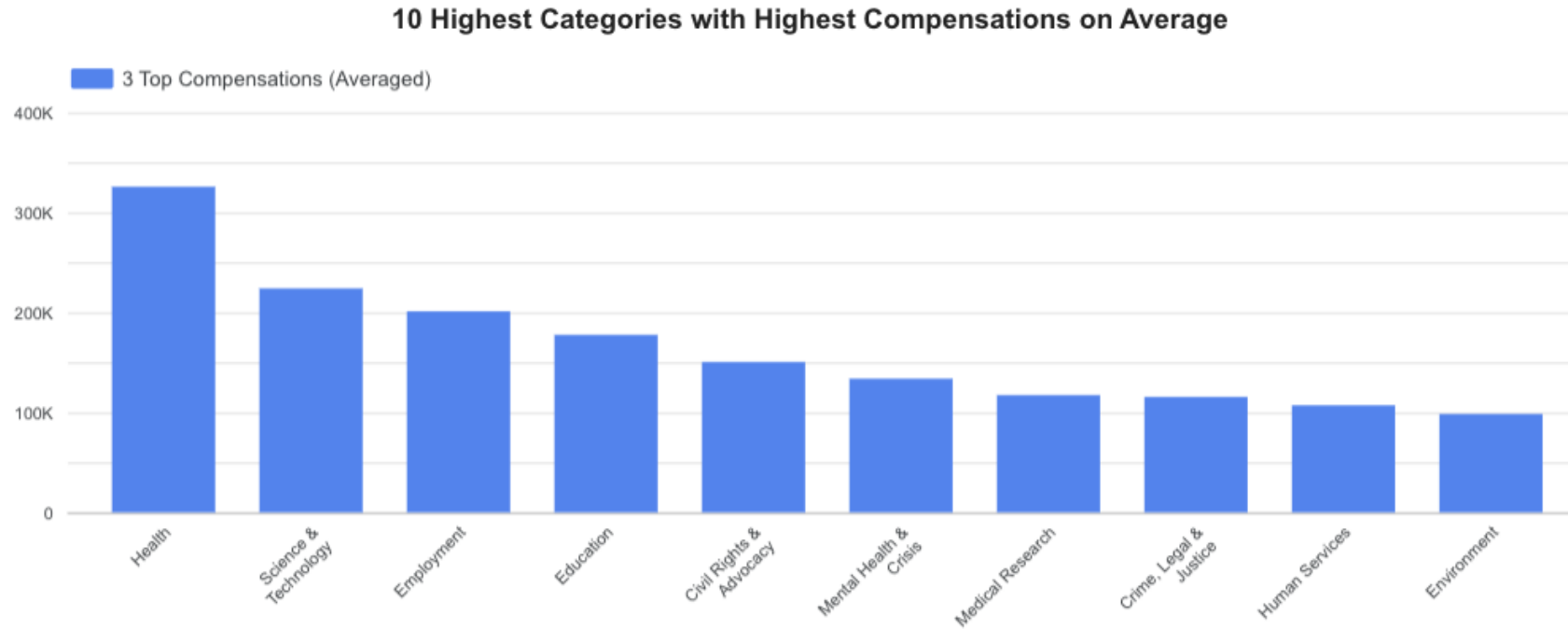
TABLES FROM WEB SCRAPING

**Table: raw_nonprofit_financials →
staging_nonprofit_financials**

EIN	Name	Category	Year	Revenue	Expenses	Net Income	Net Assets	Has Financials	Has Category	Year_int
-----	------	----------	------	---------	----------	------------	------------	----------------	--------------	----------

Table: raw_org_comp → staging_org_comp

EIN	First Emp Name	First Emp Title	First Emp Comp	Second Emp Name	Second Emp Title	Second Emp Comp	Third Emp Name	Third Emp Title	Third Emp Comp	Has Emps
-----	----------------	-----------------	----------------	-----------------	------------------	-----------------	----------------	-----------------	----------------	----------



WEB SCRAPE DESCRIPTIVE ANALYSIS

Who are the highest earners? How do these earners compare among categories?

DIAGNOSTICS QUERY TAKEAWAYS

• **Insights:**

- Health, Sci/Tech lead in pay
- Education: high revenue, low pay
- Civil Rights, Mental Health mid-tier
- Religion, Sports pay < \$25K
- Signals funding and mission bias

Recommendation:

- Benchmark by mission, not revenue
- Evaluate leadership vs. outcomes
- Question low-pay, high-impact sectors
- Flag high-pay, low-revenue outliers

Predictions:

- Pay tied to mission complexity
- More sector-specific benchmarks
- Funders add compensation conditions
- Low-pay sectors risk stagnation

High Compensation by Organization				
Name	NTEE Category	3 Highest Compensations	top_3_comp_p... ▾	
American Industrial Real Estate Association	Unknown	\$711,853	119.96	
American Federation Of State County & Municipal Emplo...	Unknown	\$682,448	108.61	
Good Samaritan Institute For Research And Education	Diseases & Disorders	\$327,527	101.96	
International Alliance Theatrical Stage Employees & Movl...	Unknown	\$3,182,286	101.82	
Southern Christian Leadership Conference Of Southern C...	Civil Rights & Advocacy	\$130,000	100.16	
Global Technopolitics Forum	International	\$19,850	96.67	
Cotsen Foundation For Academic Research	Arts, Culture & Humanities	\$103,000	95.68	
Unite A Nation Inc	Community Improvement	\$175,158	95	
California Institute For Yiddish Language And Culture	Arts, Culture & Humanities	\$38,400	92.38	

1 - 100 / 7301 < >

WEB SCRAPE DIAGNOSIS ANALYSIS

How is compensation spread among nonprofits? Who benefits the most? How does it compare to overall expenses?

DIAGNOSTICS QUERY TAKEAWAYS

• **Insights:**

- Some orgs pay >100% in comp
- Dozens spend 70–100% on execs
- Examples: AIREA (120%), GSI (102%)
- Raises red flags on nonprofit use

Recommendation:

- Check for deficits, delayed pay
- Compare comp vs. program output
- Spot patterns by sector, zip, size
- Flag recurring vs. one-time spikes
- Review if outside 501(c)(3) scope

Predictions:

- Demand for comp justification grows
- Pay tied to program impact

CONCLUSION – DATA WITH PURPOSE

Built full pipeline: API + web

Used SQL + Python for insight

Analyzed financial health +
equity

Tackled real nonprofit
challenges

Ready to drive data-driven
impact