

So you want to be a Street Performer?



Have License?



Objective

- Understand Historical **Trend** for different Business Activity
- Geographic coverage for various business
- **Competitor** location
- Identify ideal location and business area for aspiring **new** business owner

Methodology

- Data Source: City of Chicago
- Socrata Open Data API, Web query with Json format
- 1 Million+ License, 2003-- Present
- Storage :  MongoDB.
-  Streamlit app.

Issues

- Non Local Business, like DALLAS
- Recording keeping start since 2012 for some feature
- Missing vital info, like GPS location / Business Area
- Typo “CHICAGP” and non-sensical data
- Back Dated License

Data Pipeline

- Clean to usable standard
- showcase industrial **trend** for many category
- Higher performance for **indexed** searches
- Business **leader** identified for each category
- Functions to generate data from source in python shell, only load **prepared** data for Streamlit app.

Conclusion

- Choose a license code, up to 10,000 business
- Or business area, 30 business in particular category
- Bar chart provide top level info summary
- Plot number of business over the decade
- individual business leader detailed info.
- Map to locate competitors

Future Work

- Study the license renewal pattern via ML
- Save certain query results to speed up App response time.

Demo

- License Code cover too large an area, not enough detail
- Business activity are missing a lot data, especially for historical data.