



Yelp Dataset Analysis

for Service-Rich Area Identification

Team #23: Michael Boykin, Zofia Jozefkowicz, Kelly Motley, Darryl Novotny

https://github.com/dcnconsult/yelp_db.git

Identify **service-rich areas** in the Tampa metropolitan area

- **Evaluate business density, ratings, and reviews** to pinpoint high-quality service clusters and map business concentrations
- **Provide actionable insights** for homebuyers, real estate agents, and urban planners to make data-driven location decisions.

A person wearing a white button-down shirt is pointing their right index finger towards the right side of the frame. The background is slightly blurred, showing more of the person's torso and arm.

Business Problem

Real estate agents, investors, and homebuyers struggle to find locations that meet their needs based on proximity to key services.



Exploratory Data Analysis

Exploratory **Data Analysis**

Analysis of the Yelp dataset revealed strong coverage in **Florida** ranked **2nd** among all states with **26,330** businesses



Tampa Bay metropolitan area shows robust business representation

- Tampa: 9,048 businesses
- Clearwater: 2,221 businesses
- Saint Petersburg: 1,663 businesses
- Brandon: 1,033 businesses
- Largo: 1,002 businesses
- Other significant areas include Palm Harbor, New Port Richey, Lutz, and Riverview



Justification for Tampa Bay Focus

- Large data sample with over 9,000 businesses in Tampa alone
- Diverse metropolitan area including urban, suburban, and coastal communities
- Strong representation across multiple cities providing comprehensive coverage
- Growing market with dynamic business landscape
- Mix of established and developing areas offering varied analysis opportunities

Exploratory Data Analysis

Data Suitability for Business Questions

Comprehensive Coverage

- **Large Scale:** 868,606 reviews covering 19,674 businesses provide a robust foundation for analysis
- **Geographic Representation:** Well-distributed across major Tampa Bay metropolitan areas including Tampa (9,048 businesses), Clearwater (2,221 businesses), and St. Petersburg
- **User Base:** 279,252 unique customers offering diverse perspectives and experiences

Data Richness

- **Review Depth:** Average review length of 49.9 words provides substantial content for analysis
- **Vocabulary Range:** 171,254 unique words in the full dataset enables detailed textual analysis
- **Rating Distribution:** Average rating of 3.75 stars with full spectrum coverage indicates balanced feedback
- Multiple years of data capture seasonal variations and long-term trends
- Allows for analysis of business development patterns and neighborhood evolution

NLP Application

Market Analysis for Real Estate Investors and
Service Clustering



Sentiment Analysis

Positive Reviews (4-5 stars)

- Characterized by words like "excellent," "amazing," "fantastic"
- Often shorter and more concise
- Frequent mentions of specific staff members
- Focus on exceeding expectations

Negative Reviews (1-2 stars)

- Longer, more detailed explanations
- Focus on specific incidents
- More frequent use of time-related terms
- Emphasis on customer service issues

Geographic Themes

Location-Based Patterns

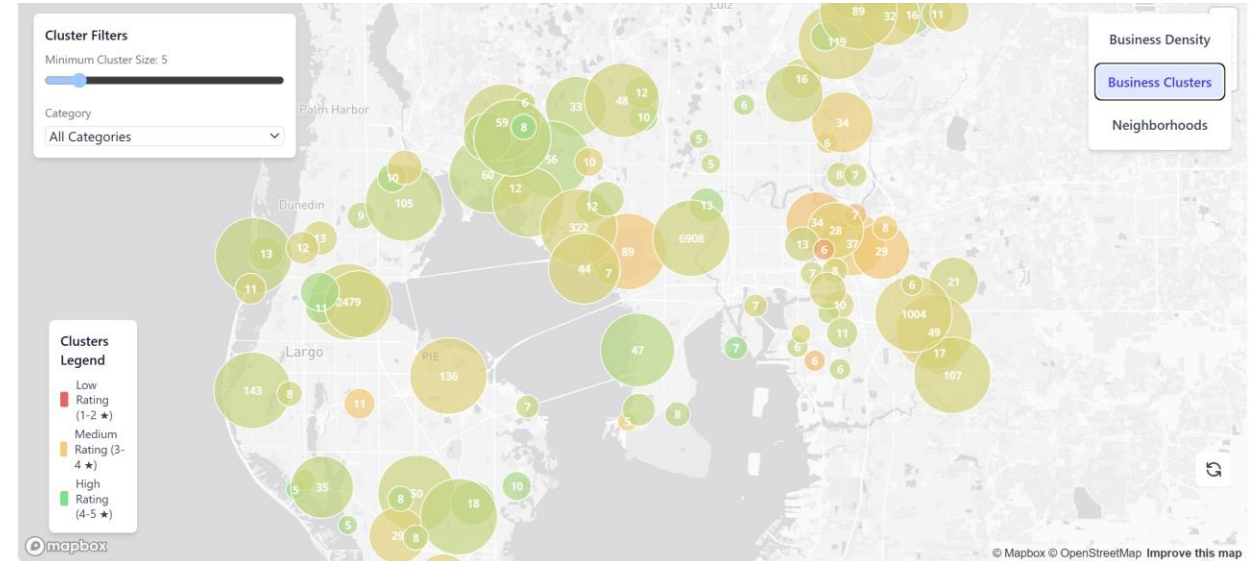
- Downtown Tampa reviews emphasize accessibility and parking
- Coastal area reviews focus more on atmosphere and views
- Suburban reviews show greater emphasis on service speed and convenience

Neighborhood-Specific Trends

- Different service expectations across neighborhoods
- Varying emphasis on price and value
- Distinct cultural preferences in different areas

Geographic Information System (GIS)

- Kernel Density Estimation for service concentration
- Spatial autocorrelation analysis
- Service accessibility mapping
- Business relationship network modeling
- Service cluster connectivity analysis
- Competition proximity analysis



https://github.com/dcnconsult/yelp_db.git

Primary Focus

1. Service-rich area identification
2. Business category distribution maps
3. Service quality predictions
4. Neighborhood development recommendations

Secondary Insights

1. Business success factors
2. Service gap identification
3. Competition analysis
4. Growth opportunity mapping



Results and **Business Insights**

Tampa, Clearwater, and Saint Petersburg
are among the **best areas for essential
services** in Florida

Results and Business Insights

Yelp Data Analysis



Tampa (9,048 businesses)

Clearwater (2,221 businesses)

Saint Petersburg (1,663 businesses)

- These areas have high business density and strong customer ratings across **restaurants, healthcare, and shopping**



Real estate agents and homebuyers

Can use these insights to find properties near top amenities, ensuring locations align with lifestyle needs.



Yelp Dataset Analysis

for Service-Rich Area Identification

Team #23: Michael Boykin, Zofia Jozefkowicz, Kelly Motley, Darryl Novotny

https://github.com/dcnconsult/yelp_db.git