

Portland Gourmet



Group 5
Ying Chen
Zhao Qingying
Li Hui
Bai Yunqing
Yu Zhengtao



Table of Contents

- Introduction
- Dataset
- About the city
- Methodology
 - Data analysis
 - Web visualization
- Results
- Webpage

Project Introduction

- **Restaurants' atmospheres** could not only be regarded as a competitive advantage of the restaurants as well as the determinants of customer satisfaction (Agoes and Pasaribu, 2015). Different atmospheres will **attract** different consumer groups.
- **In this project**, we tried to explore the restaurant atmospheres in **Portland, US**. Through NLP methods and hierarchical clustering analysis, we identified several **specific themes** and also found significant **spatial differences** between different themes, which will help to learn the **distribution of local consumer groups**.
- With the analysis results, it might help the **restaurant** to better set its atmosphere and improve its service quality. And from the perspective of **cities**, it could impove city planning and management, such as the commercial facilities plan and safety supervision management.
- In addition, we also made an **interactive web page** to visualize our analysis results.



Table of Contents

- Introduction
- **Dataset**
- About the city
- Methodology
 - Data analysis
 - Web visualization
- Results
- Webpage

Dataset

- In recent years, people's **choice of restaurants** has become more and more dependent on the comments of other guests (Lim and van der Heide, 2015). Reviews can be regarded as customers' **impression of the restaurant**. Therefore, it could help to understand the atmospheres of the restaurants by analyzing these comments.
- Yelp is a **popular online review community**, which has been widely proved to have a great influence on consumers' decision-making (Ariyasriwatana and Quiroga, 2016). The Yelp dataset is a subset of businesses, reviews, and user data provided by Yelp website (Yelp, n.d.).



8,635,403 reviews



160,585 businesses



200,000 pictures



8 metropolitan areas

1,162,119 tips by 2,189,457 users

Over 1.2 million business attributes like hours, parking, availability, and ambience

Aggregated check-ins over time for each of the 138,876 businesses

Dataset

- After exploring Portland's restaurant data, there are **5,644 restaurants** in Portland with at least one yelp review, and **750,395 reviews** for those restaurants in total. However, due to computational limitations and meaningless reviews for too long ago, we selected only the **recent 4 year's reviews**, that leaves us **297,993 reviews** in Portland since 2017/1/1.

Dataset	Source	Format	Field	Mainly Useful Information
Business	Yelp website	json	12	Restaurant id, name, location, stars, review count for each restaurant
Review	Yelp website	json	9	Text content, stars, date, restaurant id for each review
Portland Neighborhood Boundary	Open Data	shapelfile	11	Name, location and size of each community for each neighborhood



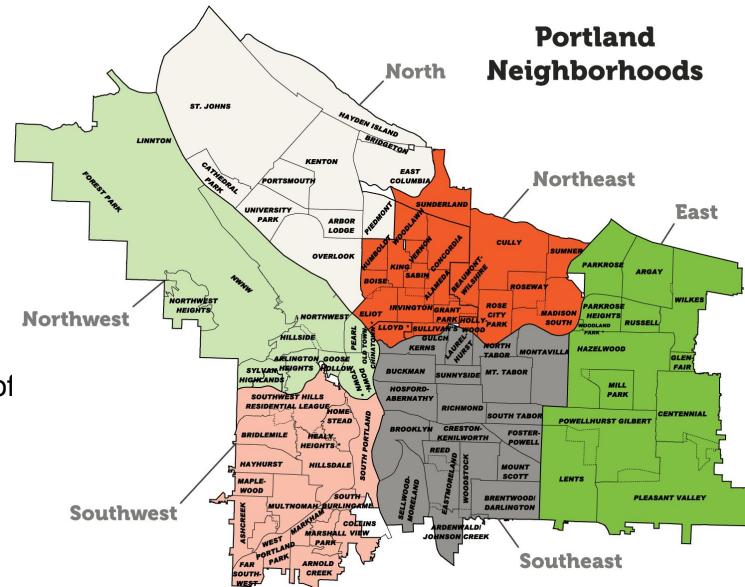
Table of Contents

- Introduction
- Dataset
- **About the city**
- Methodology
 - Data analysis
 - Web visualization
- Results
- Webpage

About the City

Portland, US

- Portland is the largest and most populous city in Oregon, 47% of Oregon's population lives in the Portland metropolitan area.
- Containing the highest number of parks per capita; the fourth largest food city in the United States (Washington Post, 2015).
- The Portland neighborhoods can be divided into:
 - North:** Major **industrial** area (heavy industry, manufacturing, logistics).
 - Northwest:** one of Portland's **most affluent neighborhoods**, featuring upscale art galleries, high-end businesses, restaurants and commercial retail.
 - Southwest:** primarily a single residential area for families. A small part of the southwest part of the **Downtown** area.
 - Northeast:** a mix of arts, commercial and residential areas. "Youth and cultural center".
 - East and Southeast:** mixed residential areas with a high concentration of stores and apartments.
- Portland's restaurant industry exhibits a very clear tendency to be "**centralized**". The largest food service is concentrated in the core of the city, including the **downtown, northwest and east areas**, and in corridors outside of the downtown business district (Schrock et al. 2015)



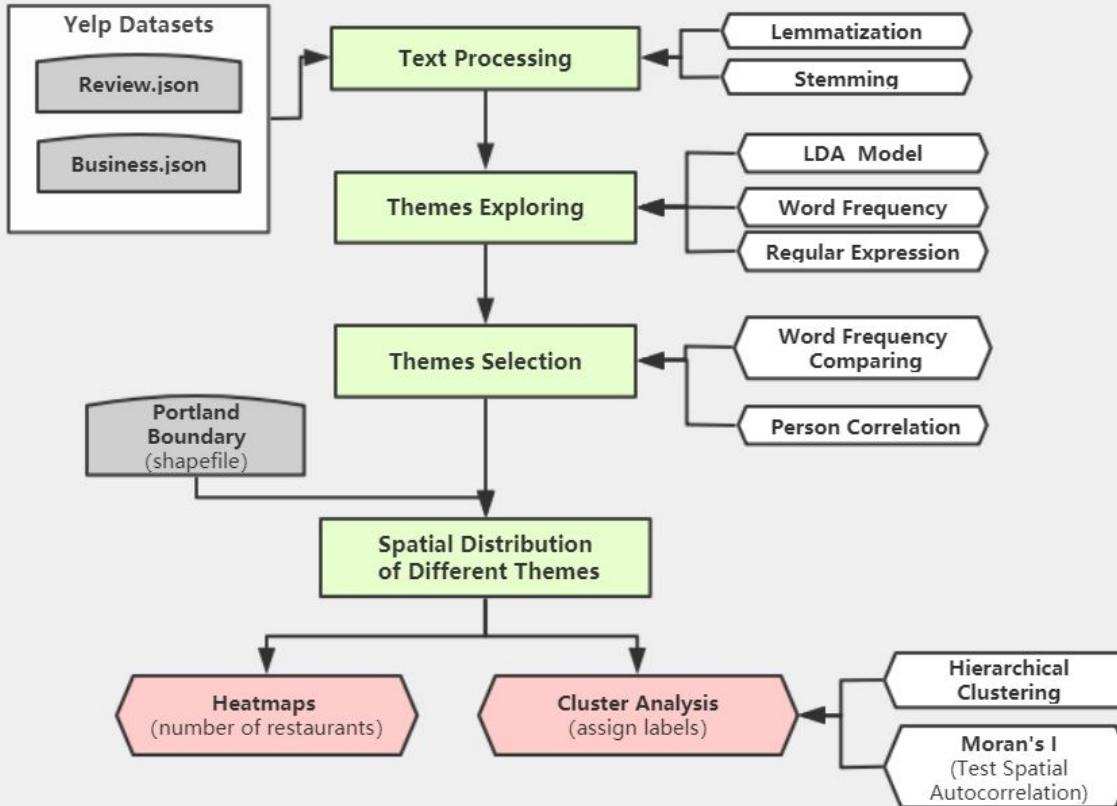
Source: Technoccult (2014)
<http://www.technoccult.net/2014/11/16/mutation-vectors-east-bound-but-not-yet-down-edition/>



Table of Contents

- Introduction
- Dataset
- About the city
- **Methodology**
 - Data analysis
 - Web visualization
- Results
- Webpage

Data Analysis



Data Visualization

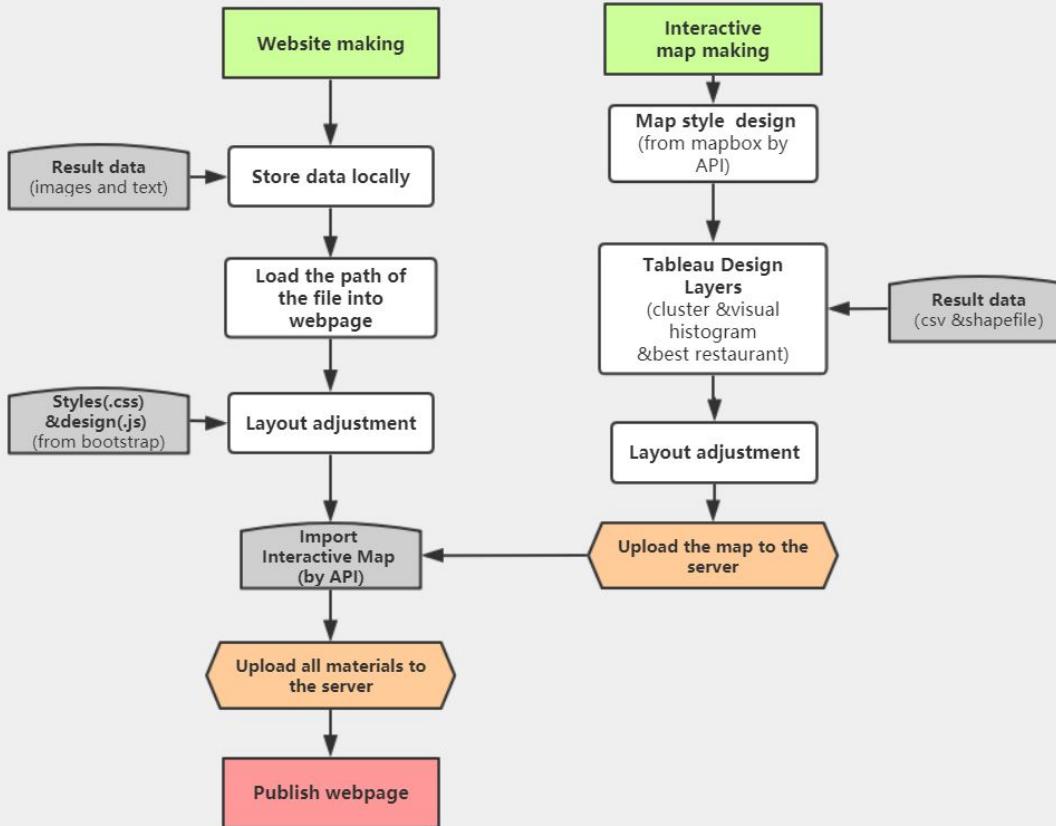




Table of Contents

- Introduction
- Dataset
- About the city
- Methodology
 - Data analysis
 - Web visualization
- **Results**
- Webpage

WordCloud

Text processing results using lemmatization:



Theme Exploring

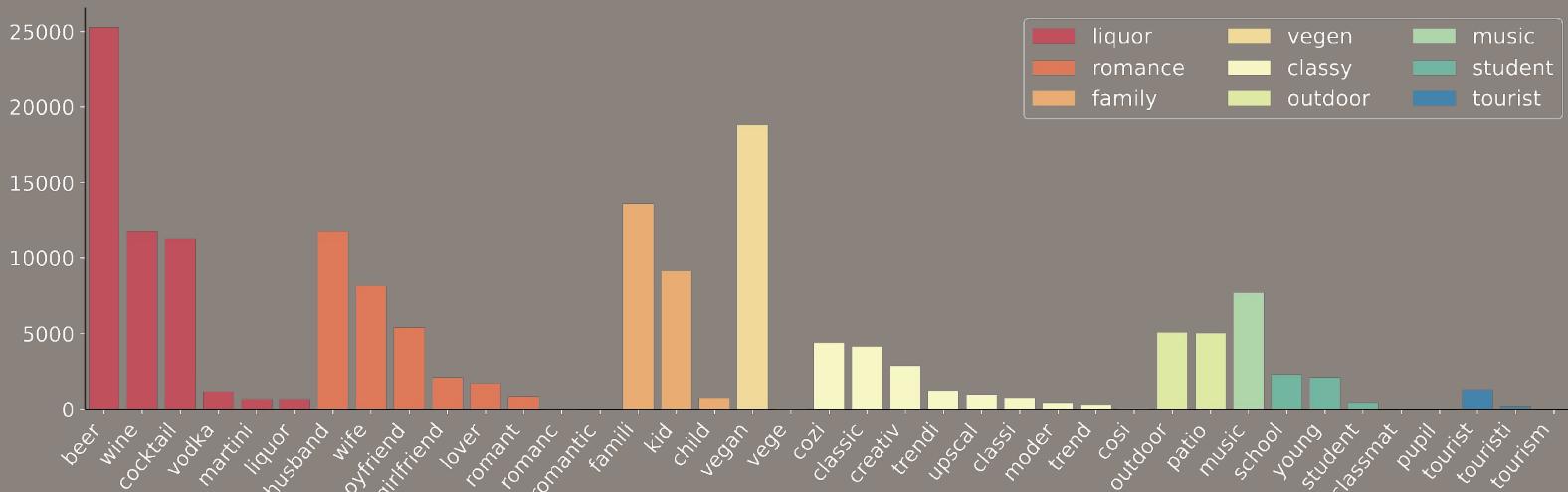
LDA model results for 5 topic clusters:

Table 1. Terms with the highest beta values for 5-star restaurants									
Topic cluster 1		Topic cluster 2		Topic cluster 3		Topic cluster 4		Topic cluster 5	
Term	Beta value	Term	Beta value	Term	Beta value	Term	Beta value	Term	Beta value
vegan	0.035	service	0.048	breakfast	0.02	fresh	0.022	chicken	0.057
option	0.015	experience	0.021	brunch	0.017	café	0.016	pizza	0.048
sandwich	0.013	staff	0.021	taco	0.017	coffe	0.015	cheese	0.023
vegetarian	0.011	friendly	0.016	egg	0.016	bowl	0.015	pie	0.013

Table 2. Terms with the highest beta values for 1-star restaurants									
Topic cluster 1		Topic cluster 2		Topic cluster 3		Topic cluster 4		Topic cluster 5	
Term	Beta value	Term	Beta value	Term	Beta value	Term	Beta value	Term	Beta value
service	0.037	customer	0.023	location	0.037	pizza	0.118	chicken	0.031
terrible	0.012	business	0.022	service	0.03	cold	0.011	sauce	0.012
horrible	0.012	owner	0.015	customer	0.03	delivery	0.009	flavor	0.01

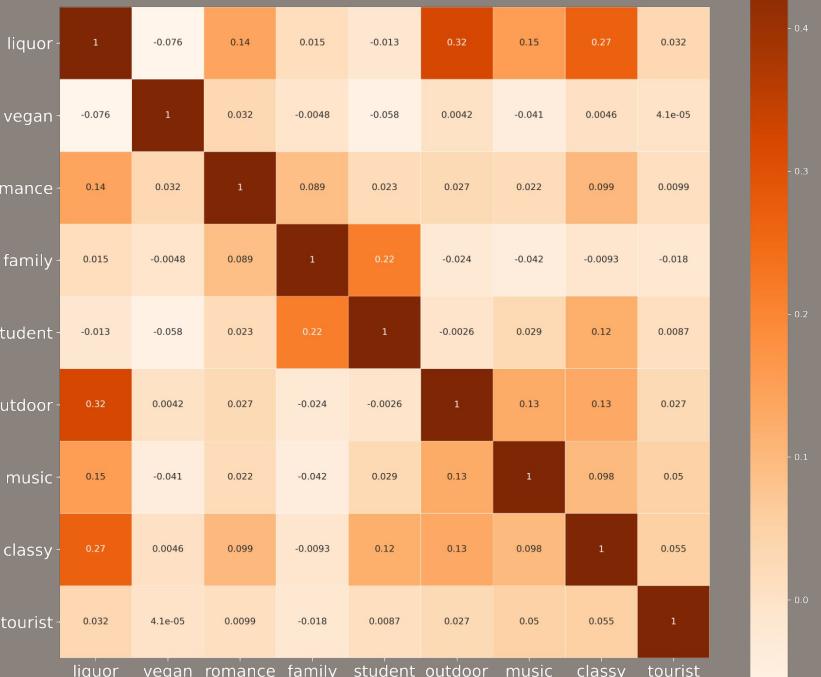
Theme Exploring

Word frequency chart:



Theme Selection

Pearson Correlation Matrix:



Theme Intensity:
selected reviews for a theme /
total reviews for a restaurant

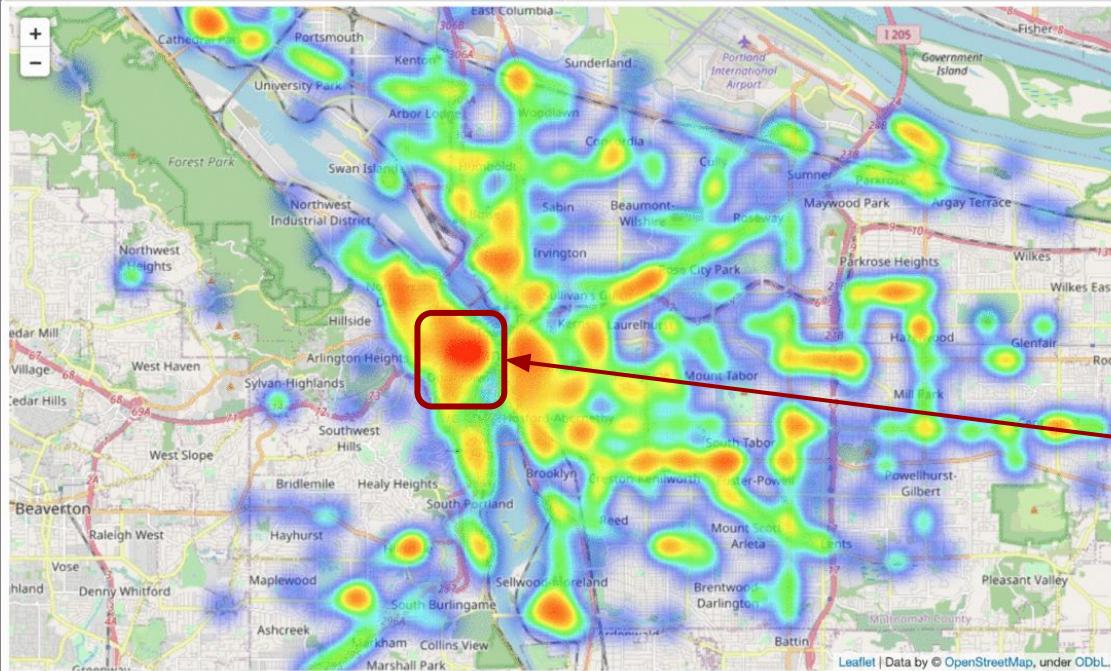
business_id	restaurant_name	address	stars	boroughs	vegan_rate	liquor_rate	family_rate	student_rate	romance_rate	outdoor_rate
M1cf085pyT-UYfMr9i4J6g	Mars Wine & Kitchen	2393 NE Fremont St	4.0	ALAMEDA	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
I1xHKh_MkYzDDaq3XOm8Wg	Lucca	3449 NE 24th Ave	4.0	ALAMEDA	0.0430	0.2688	0.1398	0.0108	0.1505	0.0108
Rn1_0J6qcl5_kF_y7pVzUg	Gulder	2393 NE Fremont St, Ste B	4.5	ALAMEDA	0.0952	0.0833	0.0357	0.0000	0.0595	0.0357

Positively related themes:

- Liquor & Outdoor
- Student & Family

Heat Map

Ranking Stars



Final Themes:

- Classy
- Family
- Liquor
- Romance
- Vegan

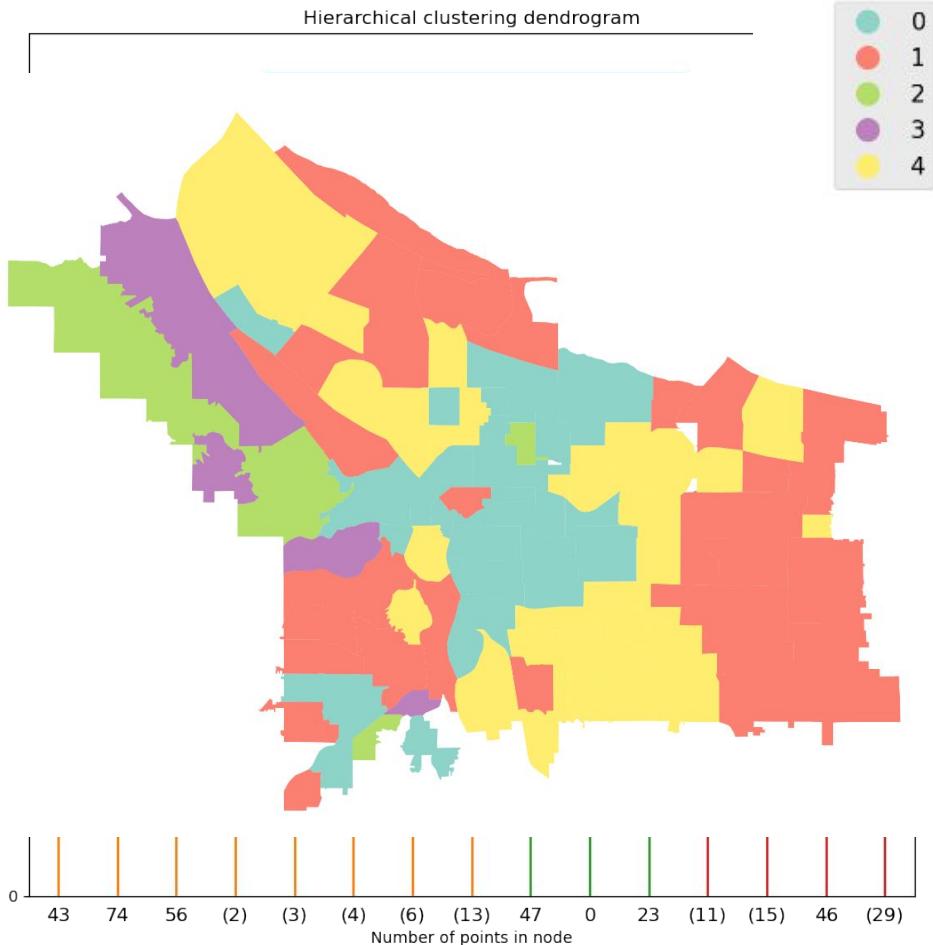
Common hotspot at
Downtown Portland.

Different sub-hotspot
across the city.

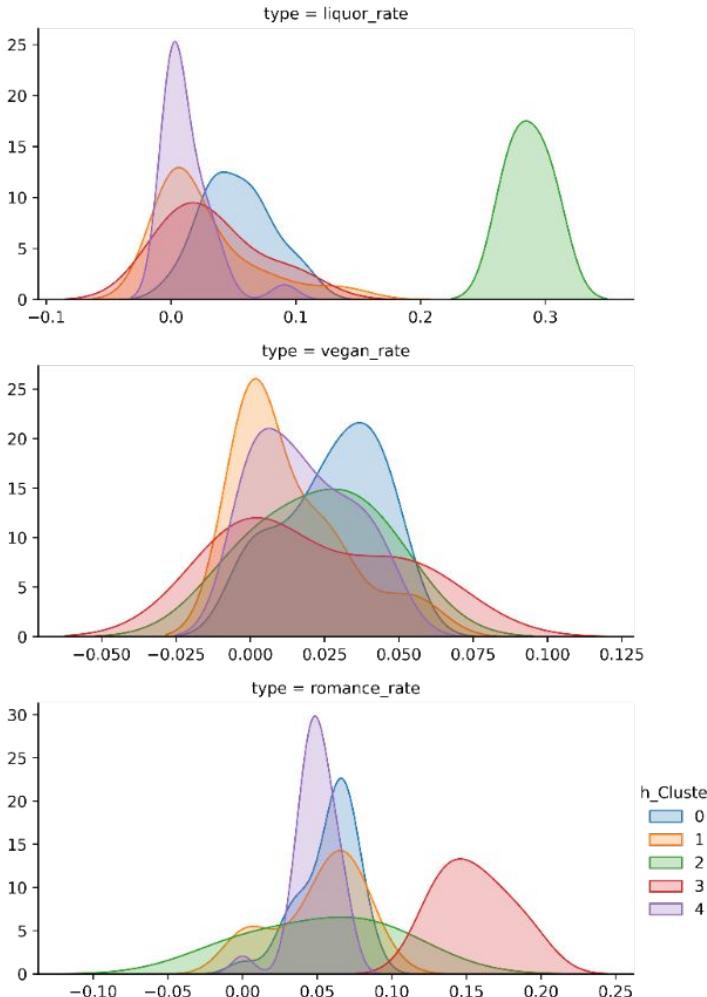
Results: Hierarchical Clustering

The obvious spatial differences and clusters among Portland by the restaurant ambience themes can be identified:

- Cluster 0 is dominant in the centre of Portland, with small parts distributed in outer Portland.
- Cluster 1 tends to group together in the outer edges of Portland.
- Cluster 2 and Cluster 3 majorly located in the western part of the city.
- Cluster 4 randomly distributed across the city.



Results: Clusters' Performances

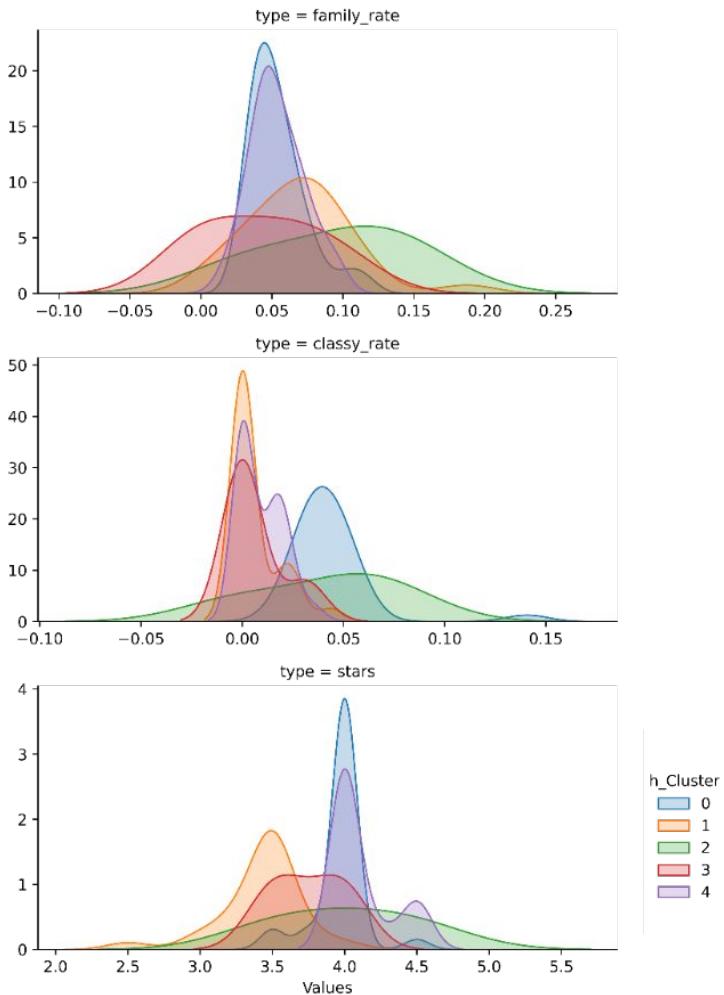


Distributions of Each Vibe of 5 Clusters

Based on the result of hierarchical clustering we assign each cluster with labels below:

- **Cluster 0** - Featured vegetarian food, high classy vibe with high stars
- **Cluster 1** - family gatherings, medium-low stars
- **Cluster 2** - prominent liquor
- **Cluster 3** - Highlighted romantic atmosphere
- **Cluster 4** - high stars

Results: Clusters' Performances



Distributions of Each vibe of 5 Clusters

Compared to the other clusters, Cluster 0 has a relatively high level of classy ambience with a high mean star rate 4.0, meanwhile, this group owns many restaurants providing vegan food. The average star rate in Cluster 1 is slightly lower than others, but the vibe is suitable for family gatherings. The Cluster 2 owns widest review rates in general except for the beverages, which has most outstanding performance in the 5 clusters. The Cluster 3 owns the best romantic vibes and Cluster 4 has the highest average restaurant star rate.

Discussions

Certified
B
Corporation



Figure 2. MCANDSCHMICKS, Portland



Figure 3. PONZI Vineyards, Portland

- Our results are helpful for understanding the diverse consumer groups in neighbourhoods of Portland and might potentially attract different client groups (Rahimi, Andris and Liu, 2017).
- The clustering results also indicate which places have nice bars for nightlife, and which place has good infrastructures but lacking in additional attractions for visitors.
- For instance, despite that the Forest Park neighbourhood is well equipped with amenities such as markets and shops, along with unique visions of the urban park, the number of restaurants is less than neighbourhoods of centre city and they are less attractive for family groups.

All figures from: *Travel Portland Visitors Guide, 2020*
https://issuu.com/travelportland/docs/travel_portland_visitors_guide_2019-20

Further Suggestions

Suggestions for Restaurants

It would be conducive for restaurants to **redefine the atmosphere** more suitable for the restaurant and improve the service quality of the restaurant. For example, for the northeast area of the downtown, it has good commercial infrastructure and public facilities. And this area labelled with high family and low stars. For restaurants in this area, more consideration can be given to setting the atmosphere which is suitable for families to eat together. At the same time, they could focus more on the quality of dishes and service, which may be different from other restaurants in the area and obtain more customers.

Suggestions for City Government

The analysis results of this project will be helpful for the government and company to **plan commercial facilities**. For example, it could first consider the neighborhood with high romance to built an romance theme park. Because the proportions of couples in the consumer groups are relatively higher in these neighborhoods.

Also, it can provide some suggestions for **city management**. For example, the three neighborhoods labelled with high liquor might have more safety and noise issues during the night because of the danger of alcohol. Thus, it would be better to increase police patrols at night in these three areas.



Table of Contents

- Introduction
- Dataset
- About the city
- Methodology
 - Data analysis
 - Web visualization
- Results
- **Webpage**

Web Structure

INTRODUCTION

Today's restaurants not only provide food to meet the needs of the taste buds, but also provide service and atmosphere to their customers to satisfy their inner needs. Restaurants' atmospheres could not only be regarded as a competitive advantage of the restaurants as well as the determinants of customer satisfaction (Agoes and Pasaribu, 2015). Different atmospheres will attract different consumer groups.

In this project, we tried to explore the restaurant atmospheres in Portland, US. Through NLP methods and hierarchical clustering analysis, we identified several specific themes and also found significant spatial differences between different themes, which will help to learn the distribution of local consumer groups. With the analysis results, it might help the restaurant to better locate its atmosphere and improve its service quality. And from the perspective of cities, it could help the government to better plan the commercial facilities and improve safety supervision.

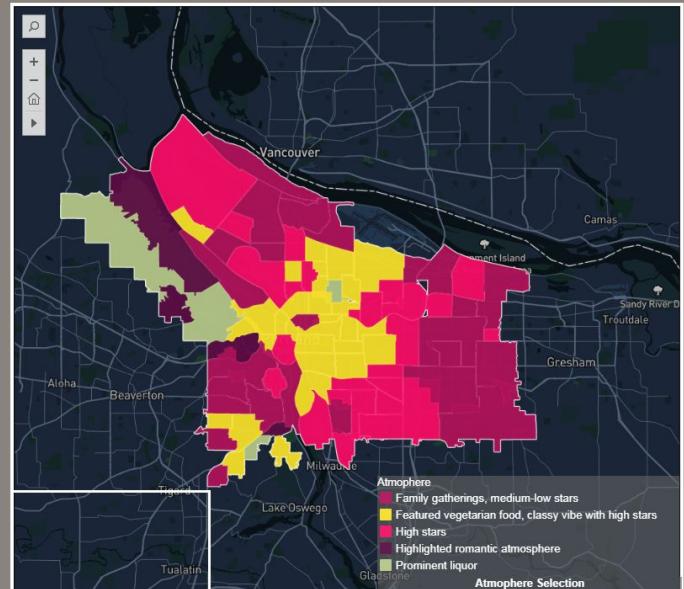


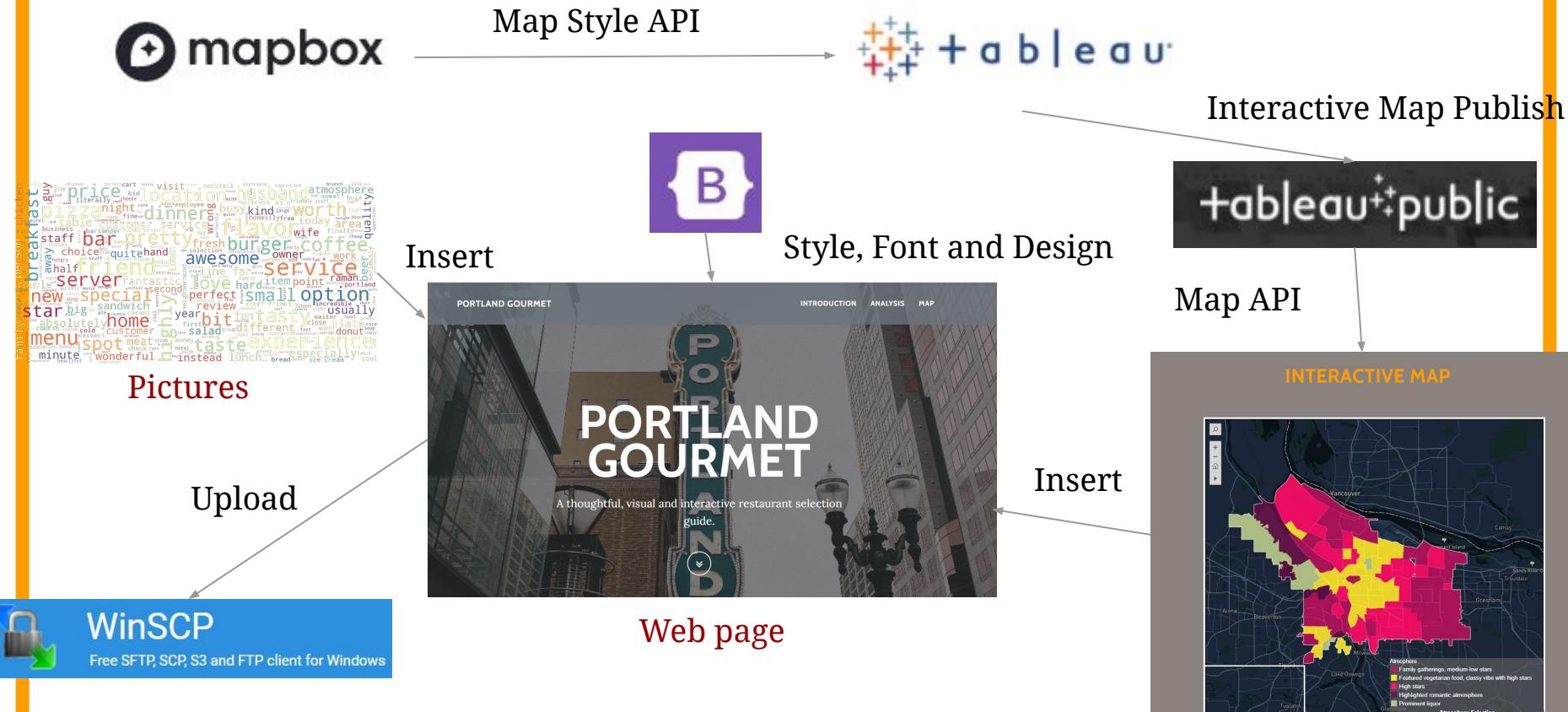
ANALYSIS

Data collection, cleaning and cluster analysis.



INTERACTIVE MAP





Limitation

Topic selection: Although the unsupervised LDA Algorithm is used, the effect of topic classification was not ideal. We mainly chose the topics that we think were more meaningful from the high-frequency words artificially. More classification methods, such as Supervised LDA, may be needed.

Cluster Analysis: In cluster analysis, hierarchical clustering was used without considering the influence of spatial format. By calculating the global Moran's I of each theme, there was significant spatial autocorrelation with themes vegan(0.166) and stars(0.152). In further study, it might need to impose spatial constraints on clusters.

Map visualization: The map views opened on different devices may not match. This is because the map format of the inserted page cannot exactly match the screens of different devices.

User feedback: This visualized result is only for the analysis of restaurant reviews, so it might be **not friendly to web viewers**. For example, the result of the final screening can only give the user one restaurant in the selected area, which will limit the user's choice. It still needs to collect more user feedback data to improve the interactivity of web pages and maps.

Reference

- Agoes, A. & Pasaribu, N. F., 2015. Kajian Pustaka Mengenai Restaurant Atmosphere. *Binus business review*, 6(1), pp.110–116.
- Ariyasriwatana, Weronuj & Quiroga, Luz Marina, 2016. A thousand ways to say 'Delicious!—Categorizing expressions of deliciousness from restaurant reviews on the social network site Yelp. *Appetite*, 104, pp.18–32.
- City of Portland, 2021, 'May 2021 – Bureau of Planning and Sustainability (BPS) Updates', Available at:
<https://www.portland.gov/bps/news/2021/5/4/may-2021-bureau-planning-and-sustainability-bps-updates> (Accessed: 15 May 2021).
- Lim, Young-shin & Van Der Heide, Brandon, 2015. Evaluating the Wisdom of Strangers: The Perceived Credibility of Online Consumer Reviews on Yelp. *Journal of computer-mediated communication*, 20(1), pp.67–82.
- Melina, M. (2015). "The search for America's best food cities: Portland, Ore". *The Washington Post*.
- Open data, n.d. 'Neighborhood Boundaries'.Available at:https://gis-pdx.opendata.arcgis.com/datasets/1ef75e34b8504ab9b14bef0c26cade2c_3/data?geometry=-122.965%2C45.478%2C-122.223%2C45.646 (Accessed: 24 April 2021).
- Pedregosa et al., 2011. Scikit-learn: Machine Learning in Python, *JMLR* 12, pp. 2825-2830.
- Schrock, G., Green, J., and Liu, J. 2015. 'Portland's Food Economy: Trends and Contributions', *Institute for Sustainable Solutions Publications and Presentations*, 88.
- Travel Portland. 2021. 'Explore diverse shopping, dining and cultural offerings in downtown Portland.', Available at:
<https://www.travelportland.com/neighborhoods/downtown/> (Accessed: 15 May 2021).
- Vallurupalli, Vamsi & Bose, Indranil, 2020. Exploring thematic composition of online reviews: A topic modeling approach. *Electronic markets*, 30(4), pp.791.
- Yelp, n.d. 'Yelp Open Dataset'.Available at:<https://www.yelp.com/dataset> (Accessed: 24 April 2021).
- USCB, 2019, Annual Estimates of the Resident Population for Incorporated Places of 50,000 or More, *Population Division*. Retrieved May 15, 2021

Teamwork Distribution

Data analysis part

- **Chen Ying:** data cleaning, text pre-processing, NLP and LDA map production
- **Li Hui:** focusing on clustering, exploring algorithms, results visualizations.
- **Zhao Qingying:** organizing structure/logics, result analysis, debug and organizing our code to GitHub.

Web creation part

- **Bai Yunqing:** Web frameworks, interaction and visual design and database links
- **Yu Zhengtao:** Web page production, map production, data visualization and interactive system design

Webpage display

The image shows the homepage of a website called "PORTLAND GOURMET". The background is a photograph of a city street with tall buildings, a bridge, and a street lamp. Overlaid on the center is a large, ornate sign that reads "PORTLAND GOURMET" in white letters. Above the main title, there is a smaller section with the text "A thoughtful, visual and interactive restaurant selection guide." At the bottom center of the page is a circular navigation button with a downward arrow.

POR
TLAND
GOURMET

A thoughtful, visual and interactive restaurant selection guide.

INTRODUCTION ANALYSIS MAP

Thank You !