



Diversity Comments 项目说明文档

Thanks for watching this!

This Sway is the official document for the Project: Diversity Comments.

Please go ahead to check what we have done!

Group Members (listed in family name)

Peng, Ziwei from University of Electronic Science and Technology of China

Sheng, Kaiwen from Southeast University

Xu, Yi from Northeast Normal University

Yang, Linfeng from Southeast University

Zhao, Jiajing from University of Electronic Science and Technology of China

Goals & Applications

分析世界不同区域人群和美国各州人群对酒店住宿的评价，提取评价中的关键词来观察他们最关注的酒店组成部分。通过获取结果，分析不同区域人群对酒店评价的差异，以及他们最关注的三个关键词，利用地图的形式进行呈现，从而可以为酒店的服务提供参考和针对性。

通过对酒店的评价进行自然语言处理，得到全球 27 个国家以及全美 50 个州的用户的酒店体验偏好，了解到当不同地区的游客入住酒店时关注度排在前三位的指标，据此酒店可以针对用户偏好指定个性化的服务，给用户更好的入住体验，同时使得酒店的服务更优质，实现双赢。

Data Sources

515k-hotel-reviews-data.zip (from Kaggle.com)

<https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe>

USReviews.csv (from Kaggle.com)

<https://www.kaggle.com/datafiniti/hotel-reviews>

Work Division

Peng, Ziwei:

Idea design, Application of data visualization, data processing;

Sheng, Kaiwen:

Alogorithm Design and implement, data processing, data verification;

Xu, Yi:

Idea design, data collection, data processing, location verification;

Yang, Linfeng:

Application of data visualization, data processing, data verification;

Zhao, Jiajing:

Data collection, location mapping, make slides, try to spider information.

(listed in in no particular order)

Achievements #1: Analysis on scores and sentiments of reviews

Formula:

$$\text{ExpectedScore}(\text{Country}) = \frac{\sum \text{Score}}{n}$$

$$\text{ExpectedPolarity}(\text{Country}) = \frac{\sum \text{sentiment}(\text{comment}).\text{polarity}}{n}$$

$$\text{ExpectedSubjectivity}(\text{Country}) = \frac{\sum \text{sentiment}(\text{comment}).\text{subjectivity}}{n}$$

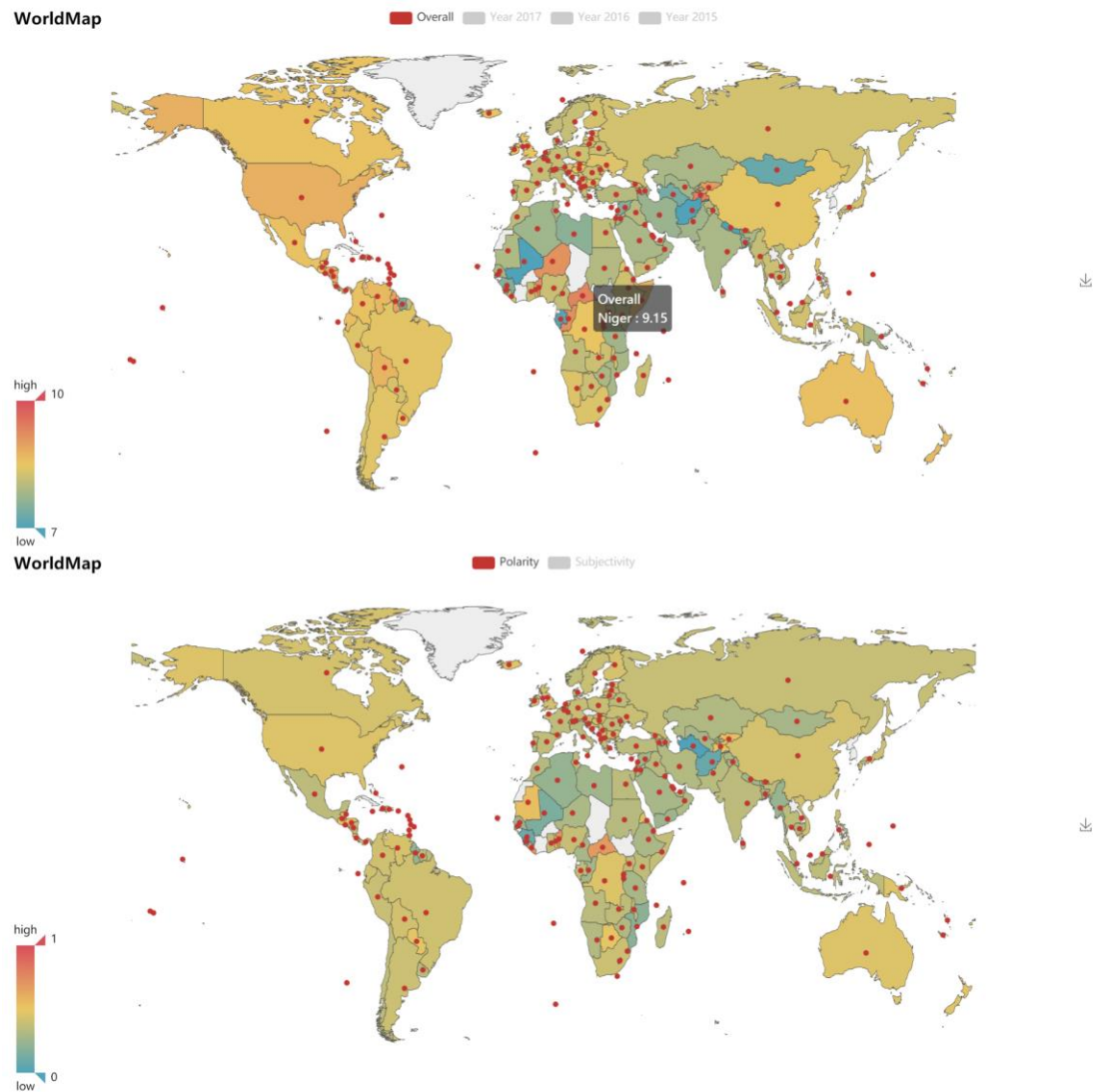


Figure 1 The average score and sentiments analysis of the worlds

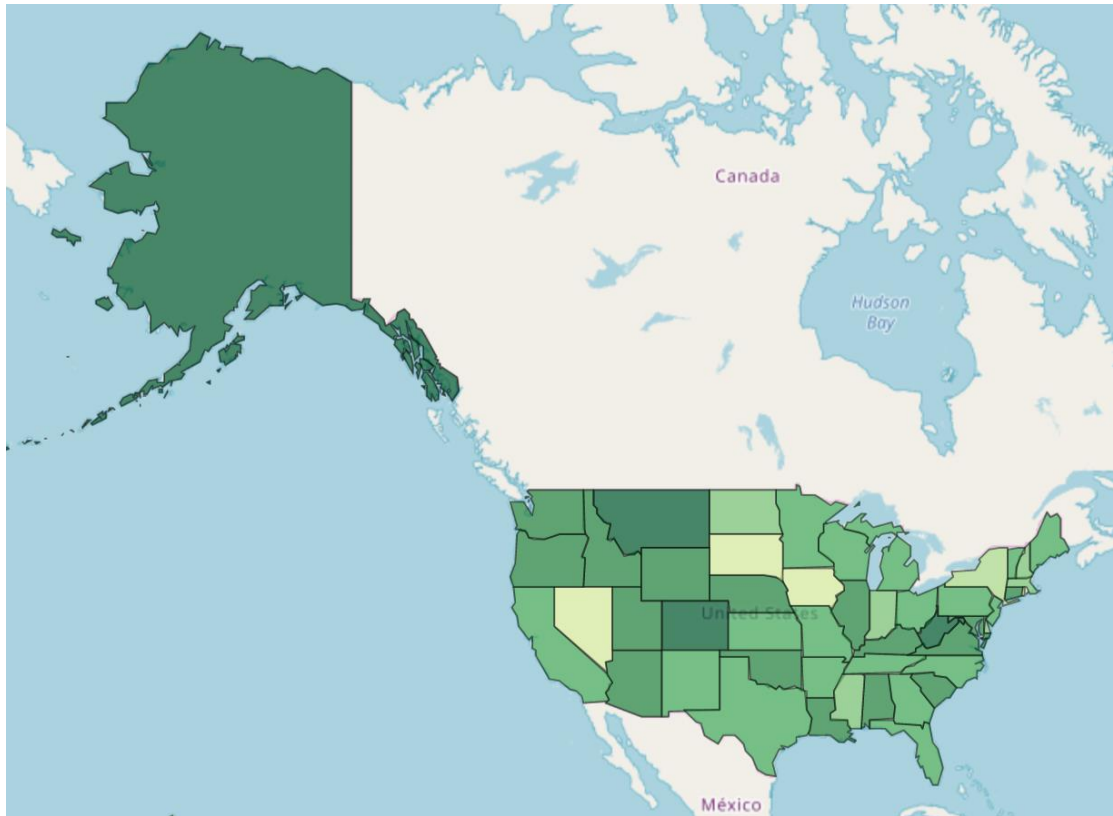


Figure 2 The average rating in US

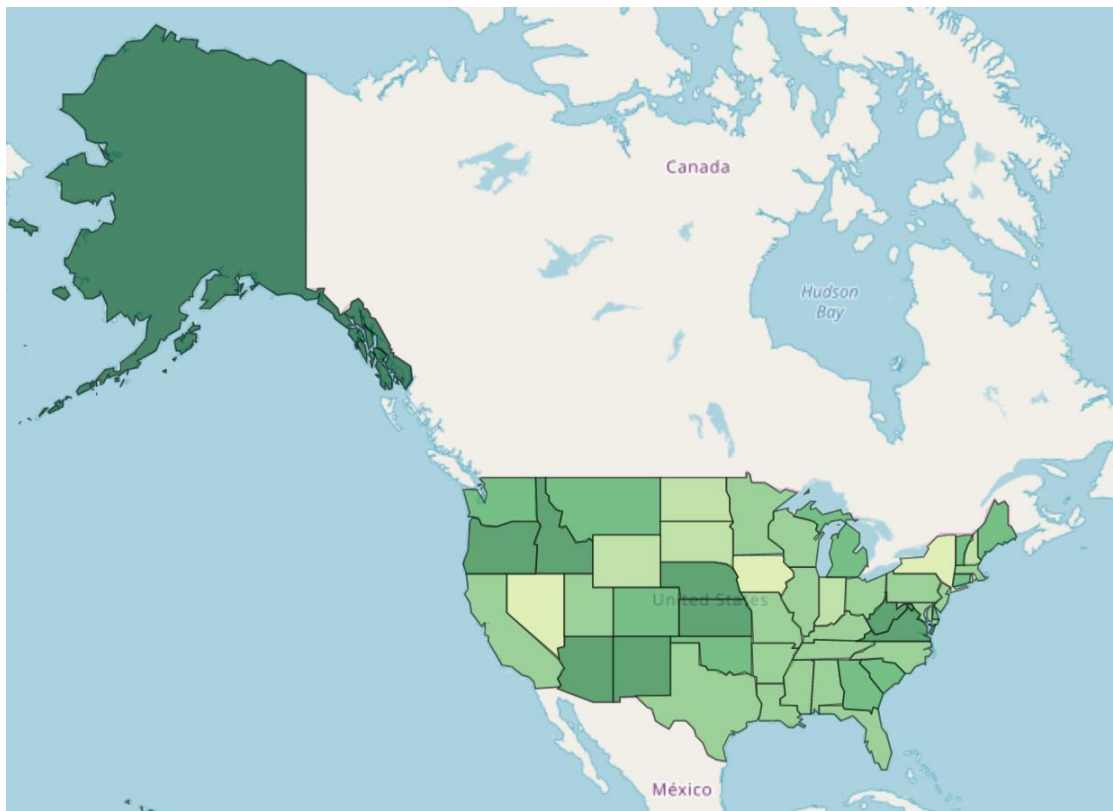


Figure 3 The polarity of comments in US

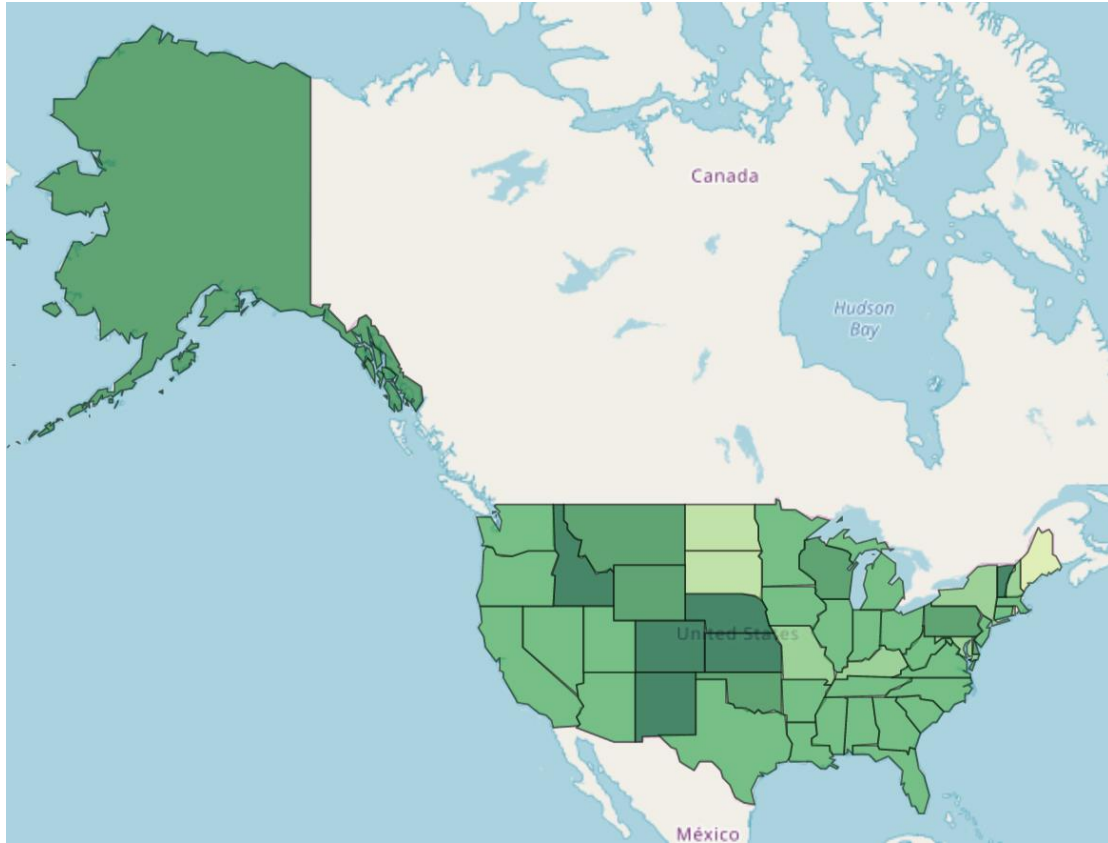


Figure 4 The subjectivity of comments in US

See more details in [SeminarA/Results/overall.html](#) and [SeminarA/Results/US_show*.html](#)

Findings:

1. The ratings are closely related to the polarity of comments.
2. Africa and the Middle East tend to give low rating than other areas around the world.
3. Some states tend to give low rating than other states in US.

Achievements #2: Key words extraction in reviews

RAKE Algorithm

Rapid Automatic Keyword Extraction algorithm; Stop words (and, the, of); analyzing the frequency of word appearance and its co-occurrences with other words in the text.

Formula:

$$\text{WordScore} = \text{WordDegree}(w) / \text{WordFrequency}(w)$$



Figure 5 Customers' preference of all over the world

See more details in [SeminarA/Results/marker_chart.html](#) and the txt files are generated keywords. We verified the data and removed the duplicated words.

Findings:

1. Of 27 countries with more than 3000 samples:
 - 26 consider 'location' as one of the three most key points in hotel;
 - 20 consider the room (comfortable, size, quiet) as one of the three most key points in hotel;
 - 16 consider 'staff services' as one of the three most key points in hotel;
 - 10 consider breakfast and other meals as one of the three most key points in hotel.
 2. Of 50 states in US:
 - 33 consider the room (comfortable, size, quiet) as one of the three most key points;
 - 29 consider breakfast and other meals as one of the three most key points in hotel;
 - 13 consider 'location' as one of the three most key points in hotel.
- also mentioned: furnishings, gardens, pool, elevator, (toilet?), wifi, pickup service.....