

STAT 579 Final Project

Chelsey Legacy, Lindong Zhou, Evan Pete Walsh*

November 21, 2014

Abstract

Analysis of the Yelp Academic Dataset.

*Statistics graduate students at Iowa State University of Science and Technology

Contents

1	Introduction	1
2	Data	1
3	Restaurants	1
3.1	Breakfast	1
3.2	Lunch	1
3.3	Dinner	1
4	Bars	1
5	Hotels	2
6	Fitness	2
7	Conclusion	2

1 Introduction

Lindong

2 Data

The Yelp Academic Dataset¹ provides data enthusiasts with the exciting opportunity to explore an incredible collection of information regarding the characteristics and quality of hundreds of businesses across the United States. Specifically, the data includes details and reviews on 250 of the closest business to 30 large universities, including Columbia, MIT, Purdue, University of Michigan - Ann Arbor, and the University of Pennsylvania, to name a few. The raw data is in json format and contains three different types of json objects: 'business,' 'review,' and 'user.'

Describe each object....

For the scope of this study, we will only be using the 'business' dataset. Our first concern was manipulating the data into a more usable format.

Go through method of cleaing....

3 Restaurants

3.1 Breakfast

3.2 Lunch

3.3 Dinner

4 Bars

Chelsey

¹https://www.yelp.com/academic_dataset

5 Hotels

Lindong

6 Fitness

Lindong

7 Conclusion

Chelsey