# **Teacher Notes for Yelp Business Data Analysis** **– Online Learning**

## **Motivation and Essential Understandings**

Businesses and other organizations often collect data on their users through apps. This data might include information about the user, such as their age, as well as information about their interactions with businesses, such as commonly purchased items. The goal of this lesson is for students to gain exposure to the kinds of data commercial apps collect from users and to think critically about how the data can be used.

* How do businesses manipulate user data for their benefit?
* What are the ethical implications of collecting user data?

## **Context and Dataset**

The application *Yelp* allows users to leave ratings and comments about businesses they’ve visited. The company uses the data collected on its users and businesses for its own analysis and has published a subset of it for personal, educational, and academic purposes. You will be guiding your students to understand and evaluate possible uses of this dataset.

## **Learning Outcomes**

Students will be able to:

* Understand how data is organized and shared
* Think critically about how data can be used for marketing purposes
* Understand how visualizations can be used to interpret geographical data
* Think critically about users’ data in the context of ethics and privacy

## **Analytics Concepts and Skills**

1. Hypothesis building / “asking the right questions”
2. Developing insights from exploratory data analysis

## **Students**

This lesson is for undergraduate Students, aimed towards those interested in business and marketing. No prior knowledge is necessary for this lesson, however, some exposure to excel may be useful for completing the challenge problem.

## **Time to Teach this Lesson**

This lesson can be taught in a 1-hour online learning session with little or no TA assistance.

**Lesson Materials**

|  |  |  |
| --- | --- | --- |
| **Materials** | **File** | **Description** |
| Lecture 1 | Instructor proprietary | PPT Lecture TBD / as determined by professor |
| Original dataset | yelp\_businesses.json | Cleaned Yelp dataset |
| Challenge Problem data | NC\_yelp.xlsx | Subset of cleaned Yelp dataset containing only NC data, as excel file |
| Solution to challenge problem (walk through) | instructions\_challengeProb.docx | Step-by-step instructions with screen shots to obtain challenge problem solution |
| Introduction to dataset | Yelp\_Worksheet\_Intro.docx | Introduction to data set and data dictionary |
| Interactive Learning Guide | Yelp\_learning\_guide.docx | A guide for online interactive learning containing Q&A and discussion questions |
| Student Worksheet | Yelp\_Worksheet\_student.docx | A worksheet for students to complete as an in-class activity |
| Teacher Worksheet | Yelp\_Worksheet\_teacher.docx | Sample answers to the student worksheet |
| Teacher Notes | Yelp\_Lesson\_Teacher\_Notes.docx | Lesson planner with links to resources |
| Power Point Slides | Yelp\_presentation.pptx | Slides to more easily screen share learning guide concepts in a 1 hour online lesson |

## **Teaching Strategies**

## **Lesson Narrative**

|  |
| --- |
| **Module 1: Interactive Learning (Virtual Classroom)** |

Note that as you move through questions in the guide, the questions should be shared with the students in whatever medium you are teaching. For instance, you might display poll questions on PowerPoint slides while screen sharing and post them in a zoom chat as they come up.

**Introduction:**

Walk through the concept of *data dictionaries* with your students and explain the context of the Yelp data set. Explain that Yelp has been collecting data on its user and business base since its inception. Keep in mind that the data given is just a small portion of the data collected.

1. Look at the data in the sample dataset. What are the attributes (different categories of information) that Yelp collects?
2. Discuss the usefulness of consumer data in business / marketing. Consider: How does this relate to the other material in your class?

**Section 1: Poll**

Now that students have an understanding of the data, use a virtual poll to ask students question(s) in section 1 of the worksheet (**POLL 1**).

Note that questions can be used with any “yes/no” or polling functionality by asking “answer *yes* if you think A. is correct, or *no* if you think answer B. is correct.” When students have all answered, explain that answer A. is correct because units are commonly confusing in data sets, and that this is a great example of the importance of understanding your data.

**Section 2: Break out rooms**

Organize students into appropriately sized groups given your class size and place groups into virtual breakout rooms. Students should be able to access the worksheet and the worksheet introduction (“Yelp\_Worksheet\_Intro.docx”) and work through the problems in the Interactive Learning Guide together in class.

Begin by introducing section 2. You can do this by following the guide’s explanations, such as *One important way that businesses can take advantage of this data is by comparing themselves to similar businesses. This involves grouping businesses by particular categories. Now, we’ll consider how a business might utilize their customer data.*

Ask each group to discuss answers to at least 1 question in this section (**Questions 1,2,3**). After each section, ask the questions to the class and have each group explain the answer(s) they came up with. Discuss and answer questions according to the Interactive Learning Guide.

**Section 3: Poll, Break out rooms**

Introduce section 3 and screen-share the image of the bar chart. Poll students on the state that seems to have the most saturated market (**POLL 2**). Explain that the most saturated market in this set is AZ, as you can see in the graph it contains around 60,000 businesses. Remind the students that this graph is only a subset of Yelp data.

Next, screen share the heatmap of businesses by location and ask students which city seems to have the highest concentration of businesses (**POLL 3)**. Explain that Las Vegas, NV has the highest number of businesses by location. Discuss how this result differs from the answer to poll 2 – this is a good place to explain why it matters *how* you look at your data. Here, data is aggregated by city rather than state.

If there is time, move to breakout rooms and ask questions (**Questions 4,5**). If you are short on time, consider explaining these questions as an open discussion with the class.

**Section 4: Break out rooms**

Introduce section 4 and ask students to discuss data privacy in breakout groups. Ask each group to answer at least one question in this section (**Questions 6,7**). Discuss students’ answers.

**Wrap up**

When the lesson is complete, discuss the role of data science in business and marketing. Consider discussing the note at the end of the worksheet:

* Why might it be useful for students to hone their analysis skills?
  + If they understand their data and learn to use software meant for handling large data sets, students can answer many business and marketing questions with their own analyses and exploration.
* What tools might be necessary or useful to consider learning more about to handle large data sets like this one?
  + For example, tools such as R, Python, and SQL are useful for handling large datasets.
  + Encourage students to share knowledge / experiences if they have used these tools or seen them used in your domain.

|  |
| --- |
| **(Optional) Module 2: Challenge Problem.** |

Provide student groups with excel file of North Carolina Yelp data (NC\_yelp.xlsx). As a take-home activity, have them filter data according to instructions under “Challenge Problem” section in the worksheet. For added challenge, ask students to identify another category or city in NC and explain or create a visualization to communicate their findings.