**Michael Grigoriev**

**Problem Set 5**

**Monday:**

*1. Starbucks wants to evaluate whether their mobile pay solutions are having a positive impact on customer service. Outline how they might collect data to answer this question using:*

A. **An Observational Study** – Starbucks can do an observational study where they watch customers use the mobile pay solutions and note how they react to using the technology. This method would differentiate itself from other methods like focus groups because they are watching customers in natural settings. This could serve as both a quantitative and qualitative study.

B. **Focus Groups-** An example of a focus group study that Starbucks could do is by gathering a group of people and asking them to use the mobile pay solutions. They may ask the focus groups to use a couple different types of technologies to make sure the participants don’t favor one or another based on biases. Using this research method would allow Starbucks to measure how quickly, accurately, and efficiently people use the mobile pay solutions. This would most likely serve as a qualitative study.

C. **An Online Survey**

Starbucks could create an online survey and ask their customers to take it for something like a gift card to encourage participation. The questionnaire would ask questions in regard to their new and past mobile pay solutions. The questionnaire would serve as a quantitative study.

Be sure to identify what the procedure would look like, what data you would obtain, and what questions you could answer using this data.

**Wednesday:**

*2. In 2014, Facebook conducted their infamous Emotional Contagion study where they manipulated users' newsfeeds to contain differing amounts of positive and negative content. Name the (a) research question, (b) independent variables, and (c) dependent variables that Facebook used in this study.*

*A.* The Research Question- Facebook conducted the infamous Emotional Contagion Study to experiment whether emotional contagion occurs between users on Facebook. Emotional contagion is when one person’s emotions and behaviors directly trigger similar emotions in other users.

*B.* The independent variable in this experiment is the presence of emotion on people’s News Feed.

*C.* The dependent variable was people’s emotional expression responses.

3. **What kind of data collection strategies would you use to:**

A. *Determine user perceptions of a social media campaign-* I think a good data collection strategy to gauge user perceptions of a social media campaign would be a quantitative study using surveys and questionnaires. This data collection strategy would be most effective because it’s important to see how many users would be opposed to a social media campaign and how many would support it.

B. *Assess the effectiveness of a web redesign-* I think to conduct a web redesign you would need to put together focus groups to asses the effectiveness of the change. If there is an overwhelming negative response to the situation, then I think a web redesign could be effective. However, a web redesign would have to be done in increments because other wise it may deter a lot of users if it creates a drastic change.

C. *Decide whether the next iPhone will live up to its hype (and turn a profit)-* I think that to asses whether the next iPhone will live up to its hype it will need multiple different data collection methods. I think a quantitative study should be done to see if people would pay for a new IPhone based on new features. This could be done by a questionnaire that asks if they would pay for the new iPhone for a certain price. Then they could conduct a focus group to test a prototype of the iPhone they will be selling.

Be certain to provide a sufficient justification for why you feel this method is correct.

**4. For the three scenarios above, describe how you would conduct each investigation (i.e., not only the type of method, but what question would you answer, what procedure would you use, what kind of data would you collect, and how would you analyze that data to answer your question).**

**A**. I would conduct a survey and or questionnaire. I would ask businesses questions like: do you currently advertise on our social media platform, has it improved sales for your business, and other questions for businesses. Then I would ask users whether they buy things or are affected by businesses that advertise on the platform, do they dislike social media platforms that allow businesses to advertise on their site, and some others. This would allow me to collect data and analyze whether social media campaigns could be successful. If users say that they are influenced by businesses that have social media campaigns then I would suggest businesses conduct these efforts.

**B**. I would asses the effectiveness of a web redesign by putting together focus group. I would start by asking the participants how they feel about the current website. Then I would show them the new design and see which design they prefer. I would put together another group of people who don’t currently use the site and see whether they like the new design or old more without telling them which one is which. This could eliminate some possible biases. Based on the replies from the participants I would redesign certain features of the website.

**C.** I would conduct a questionnaire that would ask questions like, if we were to come out with an iPhone with these certain features, how much you be willing to pay? Then I would put together a focus group and have them use a prototype of the iPhone that they will be coming out with and see how they like it. If their response were overwhelmingly positive, I would increase the price of the phone. I would also ask the focus group participants how much they would be willing to pay for a phone with the new features. All of this data would allow me to determine the price of the phone I will be creating.

**5. Describe how you might use data fusion to determine how the location people are in influences the ways they communicate with others. Assume you have full access to data from all technologies that a person may be carrying with them or that may be embedded in the environment.**

Data fusion would allow me to collect data through various sources. This would allow me to determine whether the location people are in, influences how they communicate with others. For example people on a subway in inner city New York may be communicating with others much more than in a school perhaps. I would collect this data through users’ phone gps data. I would then continue to collect people’s gps data and track their communication on social media, text, or phone calls. I could create a geographical map with plots of where there is communication via smartphones. I could then infer that certain locations have more phone activity based on these plots of communication. The graph I would use would be similar to thermal map. Having this data could be very powerful for businesses to advertise, as one example of how this would be useful. For example in a place where there is a high tendency of phone use, businesses can advertise when people’s gps marks them in that area.

**6. Format Parsing: Use the skeleton code in Problem\_Set\_5.zip to construct parsers to work with Unstructured, List, and Hierarchical data formats to analyze a text dataset. Directions are within the file.**

import string

def countWordsUnstructured(filename):

wordCounts = {}

datafile = open(filename).read()

data = datafile.split()

for word in data:

for mark in string.punctuation:

word = word.replace(mark,"")

if word in wordCounts:

wordCounts[word] = wordCounts[word] + 1

else:

wordCounts[word] = 1

return wordCounts

# This function should count the words in an unstructured text document

# Inputs: A file name (string)

# Outputs: A dictionary with the counts for each word

# +1 bonus point for removing punctuation from the wordcounts

bush1989 = countWordsUnstructured("./Bush\_1989.txt")

print (bush1989)