Group H: Atishé Chordia, Stafford Brunk, Tony Poor, Henry Wu, SeungJin Park

Final Project Plan

A. The problem (described in **B**) is interesting due to its relevance to a large number of consumers on CMU's campus. Moreover there are a tremendous number of pizza options in Pittsburgh and it is interesting to study why certain pizza places do better than the others and why certain pizza choices are popular among students. College students in particular would find the results intriguing for two reasons: first, pizza is a traditionally heavy part of their cuisine, and second, many of them are non-local and thus unfamiliar with the wide selection of local pizza restaurants. As an additional side effect, the survey could benefit small, high-quality pizza places that don't have the name brand to compete with recognized national chains. Our research question could potentially enable us to find some interesting outcomes that we can integrate with CMU's dining plan by suggesting an on campus dining location or allowing students to use dineX or Plaidcash to make it more convenient. This proposal is also doable as we have access to all students through various modes (described later) on campus. We expect to see a higher response rate since this is a topic very relevant to CMU students.

B. Some of the questions we propose to study are:

a. Determine how often people eat pizza and order for delivery, and where their favorite pizza place is
b. Compare the quality of pizza in chains to that of smaller, local places and what gives rise to this difference in quality
c. Examine each aspect of a pizza shop (pizza quality, non-pizza products, cost, business hours, delivery) and the influence of each on where one chooses to order from
d. Compare the importance of cost, delivery service, hours, etc. in the decision making of students vs. non-students?

These questions all tie up in to an overarching question: What characteristics of pizzas and pizza shops appeal to the undergraduate students enrolled in the Pittsburgh campus of Carnegie Mellon University?

C. The following are the list of a few sources that we have been researching. Some of the sources below present current literature and research on the topic and after looking at it thoroughly we have provided ways in which the results can be used in our project.

1. The Zagat Survey is naturally very relevant to our plans for a survey ranking pizza restaurants. We checked out how the survey works; it displays a comprehensive list of restaurants and allows the respondent to vote on any number of them, giving any restaurant a 4-point 'score' for food/decor/service. It allows respondents to enter optional information about price and a free-form "comments" field, while offering a large reward for "witty" and quote-worthy comments. These optional fields are likely used as additional information when Zagat releases their retail restaurant guides, rather than in the comparative ranking of restaurants. The criteria that Zagat uses to rank restaurants are interesting and we should consider it for our own survey, but perhaps decor is not as important in our case, considering that pizza is often delivered.

2. http://www.rateitall.com/t-10945-pittsburgh-pizza.aspx The following article presents some reviews of the pizza places in Pittsburgh, while the survey result helps us in understanding what some of the better questions are and the relationship between the survey responses and how the questions are phrased.

3. http://www.yelp.com/search?find_desc=Pizza&ns=1&find_loc=Pittsburgh%2C+PA This data will be interesting for a variety of reasons. There are many pizza places in Pittsburgh that specialize in different types of pizza. It would be interesting to find out what types of pizza most like, whether they are very loyal to one pizza parlor, what size pizza they usually order, etc. The survey's results could potentially show us the most popular pizza type in Pittsburgh. This website is more of a collection of community reviews than a full fledged survey or study. However, it is a collection of random people rating different pizza restaurants in the Pittsburgh area and should provide us with a good baseline for our survey.

4. http://community.post-gazette.com/blogs/citywalkabout/archive/2009/02/25/this-just-in-east-end-pizza-survey-s-top-vote-getters-revealed.aspx This is an interesting link to the Pittsburgh post gazette that has a few theories on how surveys are done, who the prime voters are, and what the rankings are. This survey also has some reviews on what determines good pizzas or good brands.

5. **Title: Program of mini surveys: pizza survey.** This journal was written by the food standard agency in UK and presents its results. We can learn dietary facts about various popular pizza locations and how it would have an impact on survey results.

D. *Sampling Frame*

The population that we will sample is clearly the ***sampling frame***. The sampling frame is different from part **E**. because, ideally we would like to sample our entire target population but due to several reasons (lack of access, etc) we may not be able to survey our entire anticipated target population. The sampling frame includes the chunk of the target population that we have access to and are in the scope of the people we would like to sample. Thus our sampling frame includes those Carnegie Mellon University undergraduate students from the Pittsburgh campus who are listed in the online CMU student directory.

E. *Target Population*

We want to make inferences about our ***target population***. The target population includes *all undergraduate students who are currently freshman, sophomores, juniors and seniors studying in Carnegie Mellon University's main campus in Pittsburgh.* Freshmen, who are on the dining plan, will also be surveyed to understand what proportion of them actually order pizzas from off campus. It will be interesting to see these results, since by understanding what pizza locations and pizza choices are preferred- we can then incorporate it in to the dining plan and make it useful for future freshmen students as well. Freshmen are not a central part of our study, but on collecting a data on them we can stratify based on year and draw further conclusions on freshmen. We are also keen to make

inferences about several sub groups of our target population. For instance: how does the most popular pizza choice or the reasons for dissatisfaction vary with race or major? We feel that this might reveal some interesting results. Section **K** on sampling scheme goes in to further detail about the sampling and non sampling errors and how they can be controlled for in our studies.

F. We will take a stratified random sampling approach (based on year-freshmen, sophomores, juniors, seniors and departments-CIT, MCS, SCS, Tepper, HSS, CFA) in order to obtain a representative sample, where we randomly select x_1 SCS freshmen, x_2 SCS sophomores, etc. We will define the sample size of each strata based on the proportion to the student body, and use the email address listed on the CMU student directory to contact our randomly selected students for an **email-and-web survey.**

Using the CMU student directory instead of the C-Book offers us the advantage of being able to easily create a program that randomly selects students based on the strata that we defined. However, it does come with a few caveats: for example, the student directory includes ineligible units like Qatar and Silicon Valley students. Thus, our program will select students from these campuses as well, even if they aren't included in our target population. This is the coverage error that can be fixed by using a survey question that asks for their home campus and filters out non-Pittsburgh students.

While this methodology has a large non-response bias, we will be able to offer some kind of incentive (gift card drawing or something) for filling out the survey. We will also be able to spread the availability of the survey to many different areas where many eyes will see it (Facebook, misc.market, fliers on campus, etc). These will form a medium of advertisement and may encourage people to fill out the web surveys. These strategies should enable us to overcome the non-response bias and get a good sample.

Two comments:
1. If you do a raffle, you need to allow people to opt out of the raffle to protect anonymity
2. Please keep *SEPARATE* responses from the strat rand sample (with followup emails!) in F above, vs responses from general adverts in Facebook, misc.market, fliers, etc. Combining leads to uknown biases. Separate analyses lets you look at what the biases actually are.

G.

Variables descriptive of student's behavior encompasses how CMU undergrads eat pizza: Are they loyal to one pizzeria? Do they vary their purchases based on day of the week, occasion, price, etc? What types of pizzas are they ordering? How many pizzas are they typically ordering? How often? What is their favorite pizzeria? All of these variables are various attributes of pizza eating and are thus the targets of our survey. These are only variables that can be derived from student behavior characteristics. In addition to this, we propose to measure variables such as whether the pizza arrives hot, how well it has been cooked (over cooked or under baked), freshness of ingredients, overall presentation, and generosity with toppings. Combining both sets of variables helps us gain a better understanding of the research question.

H. Not required.

I. [Submitted separately]

J. [Submitted separately]

K. *Sampling Scheme*

We will take a stratified random sampling approach (based on year-freshmen, sophomores, juniors, seniors and departments-CIT, MCS, SCS, Tepper, HSS, CFA) in order to obtain a representative sample, where we randomly select x_1 SCS freshmen, x_2 SCS sophomores, etc.

After creating our survey and pretesting it out (field tests, experts in the topic) we will start the survey process by sending out an email to the selected sample and asking them to complete an email-and-web survey. After getting our preliminary results we can identify the nonresponders of the initial survey. We will start by sending them a friendly reminder e-mail regarding the survey and reiterate our need of a quick response from them, as the first wave of nonresponders may have intended to complete the survey, but forgot about it as time passed. This will be sent out 3 days after the first survey and it will highlight that the survey is designed to take very little time to complete. We will send more proactive follow up e-mails (every three days for the next 2 weeks or 4-5 emails) with repeat nonresponders to try to gauge their leverage and determine the effort and incentives we need to give to persuade them to complete the survey. We will try to increase their interest in the topic and reassure them of the legitimacy of our survey and the possibility of incorporating the survey results to allow for better pizza quality and service.

After respondents provide data to ensure that the quality of our estimates are high we may need to make post survey adjustments, which include weighting underrepresented populations or imputing missing values in the surveys.

If the stratified sample is not representative of the target population, we will use our program to randomly select an additional sample from underrepresented strata that will produce enough expected respondents to make our overall sample reflective of the target population. We will then send out surveys and follow up on the new sample using the same methods we applied to the original sample. If there is inadequate time to complete an additional sample, then we will use a simple weighted approach for the existing answers, putting more weight on the underrepresented strata.

L. Questionnaire attached.

M. *Sample Size estimation*

We are planning on using a sample size of about 211.

We also did a quick estimation of the stratified **margin of error**, where we assumed that each strata had an equivalent proportion (in other words, $W_i = W_j$, etc.). Since our total population size is about 5280, we found that a sample size of about 4% (or about 211) would allow us to achieve a margin of error of about 6.8%.

Using $f_h = .04$, $W_h = 1/24$, and $n_h = (5280/24)f_h = 8.8$:

$$\begin{aligned}\text{MOE} &= 1.96 \times \sqrt{\sum W_h^2(1-f_h)\tfrac{.25}{n_h-1}} \\ &= 1.96 \times \sqrt{24(W_h^2(1-f_h)\tfrac{.25}{n_h-1})} \\ &= 1.96 \times \sqrt{24((1/24)^2(1-.05)\tfrac{.25}{7.8})} \\ &= .068\end{aligned}$$

Of course, this is certainly only an estimation. The proportions are not set according to the actual class/major distributions at CMU (and naturally, for example, we couldn't have $n_h = 8{:}8$), but it remains a reasonable approximation. Meanwhile, the sample size is a good number, simply because it allows us to break down and poll from each of our strata.

Since we are using email with a strong incentive program, we are hoping for a 30% response rate. This would imply that we would need to reach out to about **650** students.

Questionnaire

1. What is your age? _____
2. What is your gender?
   a. Male
   b. Female
3. What is your year in school?
   a. Freshman
   b. Sophomore
   c. Junior
   d. Senior
4. What is your primary school?
   a. CIT
   b. CFA
   c. HSS
   d. TSB
   e. MCS
   f. SCS
   g. Heinz
   h. Other: _____
5. What is your home state or country? _____
6. How many times per week do you eat food from a campus dining restaurant (Entropy, trucks, hot dog stand, all other dining locations included)?
   a. 0
   b. 1-2
   c. 3-5
   d. 6-10
   e. 10-15
   f. 15+

7. How often do you eat pizza?
    a. Daily
    b. Weekly
    c. Monthly
    d. Annually
    e. Never
8. What kind of pizza do you like the most?
    a. Thin crust
    b. Thick crust
    c. Deep dish
    d. Sicilian
    e. Other: _____
9. What methods do you use to order pizza?
    a. Internet
    b. Telephone
    c. Text
    d. Walk-In Order
    e. Other: _____
10. When you order pizza, what factor influences your order the most?
    a. Price
    b. Quantity
    c. Quality
    d. Other: _____
11. When you order pizza, how often do you order from the same pizzeria?
    a. Always
    b. Often
    c. Sometimes
    d. Occasionally
    e. Never
12. How much does the proximity of a pizzeria affect your decision to order from it?
    a. Completely
    b. Somewhat
    c. Neutral
    d. Doesn't really
    e. None
13. What delivery time is acceptable to you?
    a. Less than 15 minutes
    b. 15-30 minutes
    c. 30-45 minutes
    d. 45-60 minutes
    e. More than 1 hour (don' really care as long as it comes)
14. How much does the price of the pizzeria affect your decision to order from it?
    a. Completely
    b. Somewhat

c. Neutral
d. Doesn't really
e. None

15. How much does the occasion for eating the pizza (ie just for dinner, for a party, etc) affect your decision on what pizzeria to order from?
    a. Completely
    b. Somewhat
    c. Neutral
    d. Doesn't really
    e. None
16. What time of day do you typically order pizza? Mark all that apply.
    a. 4:00 a.m. – 10:00 a.m.
    b. 10:00 a.m. – 2:00 p.m.
    c. 2:00 p.m. – 6:00 p.m.
    d. 6:00 p.m. – 11:00 p.m.
    e. 11:00 p.m. – 4:00 a.m.
17. How much do business hours affect your decision to order from a pizzeria?
    a. Completely
    b. Somewhat
    c. Neutral
    d. Doesn't really
    e. None
18. On what day of the week do you typically order pizza? Mark all that apply.
    a. Monday
    b. Tuesday
    c. Wednesday
    d. Thursday
    e. Friday
    f. Saturday
    g. Sunday
19. How do you typically pay for pizza?
    a. Cash
    b. Check
    c. Credit Card
    d. Dinex
    e. PlaidCa$h
20. How many of each size pizza do you typically order in one order?
    a. Small: _____
    b. Medium: _____
    c. Large: _____
    d. Extra Large: _____
    e. Extra Extra Large: _____
21. Which pizzeria do you feel is the best in Pittsburgh (including on campus dining)?
    a. Vocelli's

b. Papa John's
c. Domino's
d. Pizza Palerma
e. Tartan's Pavillion
f. Antoons
g. Mineos
h. Lucci's
i. Other: _____

22. Do you care that your pizza comes hot when delivered?
    a. Yes
    b. No
23. If yes, on a scale from 1-10 rate how important it is to get hot pizza
    _____
24. From the following criteria: Freshness of ingredients, generosity with sauce/toppings, presentation, crust thickness, arrives hot, please list in ascending order in the order of importance.
    _____

Did survey get cut off here?