1. Overview

Team 5 was tasked with analyzing data provided by Right Price to help determine if the use of competitor prices would serve as a valid substitute for consumer data obtained from surveys. This would allow the company to increase efficiency by removing time-consuming and labor-intensive research methods, without sacrificing the quality of the price recommendations. To do so, we decided to first visualize the overall data as well as individual products, and then establish hypothesis tests to better understand what variables influence price according to each product. After that, we would present additional recommendations and build models that Right Price could potentially use to estimate future optimal prices.

Objective: Analyze what key variables influence price and determine if Right Price may substitute their survey-based research with research of the competitor prices.

2. Key Findings

Gender - Surface level exploration of the data set led us to believe that younger people were more lenient with their spending. We had some belief that as age increased, people's WTP decreased. To further explore this, I used GGPLOT to view the relationship between age and WTP across the entire data set, including all products. Although it wasn't the cleanest visualization, it supported our initial thoughts. I then took my product (Oreo Shake) to see if things aligned. Creating another graph, I saw the same negative correlation with age and WTP. I even decided to facet by race for further testing—my findings were the same as they'd been thus far.

Age - After looking at age, we decided to explore gender. Without having any prior assumptions about this variable, we went into our research with open minds—we wanted to find anything that seemed interesting. At first, we believed females had a higher WTP than males. Our visualizations confirmed that females had a higher *range* of WTP, but on average, the WTP of each gender were very similar per race. Even when I narrowed the visualizations to my product (Oreo Shake), it resulted in the same findings.

Ethnicity - Another variable we looked at was Ethnicity**.** From a first glance, based on the WTP data, all ethnicities price each product proportionally. which shows that the population has a standard on how they price each product which is that food items are priced higher than drinks. Then, looking at the combined products and how the different ethnicities price them, it can be seen that the Black ethnicity group has the lowest median in pricing the products. To further explore this, we graphed location with price and observed that Ladera Heights price products way lower than Fullerton. However, after adding the product type into the graph, it shows that only drinks were surveyed at Ladera Heights and only food at Fullerton. This explains why Fullerton's pricing was higher and can also explain why the Black ethnicity group had the lowest median in pricing. Since 71.51% of Ladera Heights population is Black/African American, it can skew how they price products as most of the Black ethnicity group surveyed could have been from Ladera Heights which only surveyed drinks-which is generally a lower priced item (for entire population) according to the data.

Based on the visualizations, we decided to establish 2 hypothesis tests. We compared WTP prices and competitor prices to identify differences and help us determine if Right Price can utilize the competitor prices data to replace survey-based research. Our H0 was "The prices according to WTP and Competitor's Price are the same " and our H1 was "The Competitor's Prices are higher than the WTP prices". We tested the hypothesis with the overall data and within each product. After obtaining the observed statistic, creating and visualizing the null distribution, and getting the p-value, we got to the conclusion that we had significant evidence to reject the null hypothesis, and to accept that "the competitor's prices are higher than the WTP prices".

The initial visualizations made showed that Female and Younger surveyors had higher WTP prices. To test if the variables Gender and Age actually influenced the prices, we established our second hypothesis test. The H0 was that "The WTP prices are the same for Female and Male" and our H1 was that "The WTP prices are higher for Female than for Male". We used the same methodology as with the first test and concluded that there isn't enough evidence to reject the null hypothesis. Finally, we filtered the data set into Young, Middle and Old, repeated the second hypothesis test, and found that Age was not a confounding factor.

3. Recommendations

To delve into how prices could be used to estimate demand we created 2 different linear regression models. The first model predicts demand based on the WTP prices. The linear model was significant with an adjusted R-squared value of 0.2911 and a negative relationship between the variables. Even though this was significant, we wondered if we could make the demand estimate better. So for our second model, we decided to use profit instead of just price to estimate demand. We calculated profit by subtracting the unit costs from the WTP prices in the demand dataset. This linear model was also significant with an adjusted R-squared value of 0.3321 (higher than our previous model). The equation for the model was (Demand = 95.559 + -8.569*Profit) indicating a negative relationship between these variables too. The hypothesis test on the slope indicated that profit does have an effect on demand with a p-value of 0. The confidence interval told us that we're 95% confident that the impact is going to be negative - as profit for the client goes up, the demand for the product will decrease. Therefore, we believe that if the company used competitor prices when calculating profit, they can use this model as an alternative to estimate the demand. Since we're adding unit costs of the products into the mix, it will also help reduce variability in the demand function which was one of the goals.

Next, to estimate the optimal price, we decided to create a profit regression line for the whole dataset. We created a linear function that predicts profit based on the WTP prices. We then plotted the profit line (positive sloping) and the demand line (negative sloping) with price as the independent variable for both and looked for the intersection to estimate the optimal price. From our visualization, the general price was between $6-$7. This is an alternate method we explored to find a price to set for these products. We suggest using these two lines to find the optimal price as it accounts for both the profit that the client could make and the demand for that product. For better results, we also suggest fine tuning these two lines to individual products instead of a generalized model.

Lastly, based on our first regression model, we believe future regression models could be created that predict WTP based on competitor prices. Right now, there is a different number of datapoints for WTP and competitor price which makes creating a linear regression model not feasible. However, if we could equate the number of datapoints between the two sources, the company could use competitor prices to estimate WTP without having to survey people. In addition, based on the data provided, we observed that the number of people and amount of competitors became more similar after $5.551 (mean WTP price). Because of that, we recommend to utilize competitor price to estimate WTP only with values higher than $5.551, which will yield a higher accuracy.

4. Further considerations

As an alternative to competitors' prices, we believe RightPrice can use data from food delivery apps to learn the prices of the competitors and how many sales each product has. If the cost of purchasing the data is something the company can afford, it would be a useful alternative for data collection.

In the WTP data, we found that the date and location data points did not offer much insight, as the amount of data collected is too limited. For instance, if surveys were collected year-round, then the collection date could affect seasonal pricing. Similarly, if a client was considering multiple locations, then the location data could ultimately impact where they will open a store. However, since this data is not present, it is difficult to draw any conclusions from this set. One place where we found that the location data could be helpful is when comparing similar products in the same area (e.g. chocolate shakes and oreo shakes in Ladera Heights).

We also determined that there are additional data points RightPrice should collect if they continue to use surveys. First, learning more information about the products, such as the size of the item (e.g. a small, medium, or large drink) or whether it is from a more upscale restaurant or a fast food chain is important to help determine if there is a certain size beverage that people are more attracted to and what the optimal price would be.

Additionally, if the client is an expanding company interested in opening a new location, an interesting data point would be whether potential customers have heard of the brand or purchased anything from the client before, which could influence pricing by capitalizing on name recognition or trends. Alternatively, if they are a start-up, it would be interesting to know if prospective customers have seen any advertisements.