Understanding Health and Wellness of Young Adults

Executive Summary

We, the members of “Data Science Corp” in this report analyzed food habits, lifestyle choices and health preferences of individuals to provide actionable insights for a food and beverages client (F&B) of SRG corporation. Instead of diving into the data analysis directly our approach was to understand the data and see if there is a problem that needs to be solved.

As most Graduate students are working professionals juggling between classes and work, we thought it would be interesting to study the lifestyle of young working adults and what impact it has on their food choices. We all have experienced at some point that stress and unbalanced workloads can lead to eating disorders. But we wanted to see this for ourselves, hypothesize it and then prove it through statistical analysis. And based on our analysis we recommend to our client F&B how they can produce healthier food and snack options for their young clients.

Our analysis suggested that stress is indeed high among young adults and there is a correlation between stress and their eating habits, such as eating out often at fast food restaurants and snacking more than usual. Our goal was to provide these young adults better options than the regular fast food or frozen options available in the market. To come up with a product we took a customer centric approach rather than a product centric approach. Instead of looking for people who would buy a product, we looked into what people might need or buy.

Based on our analysis of intent we came up with suggestions for F&B. We grouped people's preferences into two categories/intents (can be extended to multiple), and then we model people’s intentions based on income, age group ethnicity etc. So to summarize, we found a problem among a certain group, analyzed what they might need and then recommended solutions to our client F&B, which not only helped them open up a new market, but our solutions also helped young people with better food options. The rest of our report gives a detailed analysis of our approach and solutions.

We first classified the survey questions based on how they relate to the respondents to help us analyse the following categories:

1. **Health**
2. **Food/ Previous Meal Related**
3. **Personality/Goals/Motivations**
4. **Cooking Habits**
5. **Food Shopping Habits**
6. **Lifestyle**
7. **Personal/ Demographic**

Questions from the Lifestyle category helped us understand habits and formulate the problem. Questions from the Food Shopping Habits category helped hypothesize the intent of the customers shopping habits. Finally, we used demographic questions to find marketing solutions.

What drew our attention was how many people answered the below questions:
I often wish I had more energy - Q1rp
I am so busy, I often can't finish everything I need to in a day - Q1rq

My Contribution

We were given a survey data of 1000 individuals on their lifestyle, personality, and eating habits. The reason this project was different from any other was that this was not the tabular structured data we are used to working with, and along with the unstructured nature of the data, there was no specific problem to solve. We were expected to study the survey Q&A and come up with a case study useful for a consulting client.

Because of the open-ended nature of the project, I suggested we study and classify the survey questions to understand the behavior of each individual and then see if there is a problem to solve. This was well received by my fellow groupmates. Once we analyzed the data, we realized there seem to be a correlation between stress and a person exhibiting an eating disorder. But we were struggling to provide a meaningful direction to this analysis. After a round of discussion, We decided to come up with a product (a dinner box) that can be targeted to the demographics of young and stressed individuals who do not have enough time to cook and hence end up eating out often. Now that we have a product to offer, I took the initiative of designing a set of hypotheses to first validate our assumptions and define a target market. Few hypotheses, for example, look like this:

- People in younger age groups are more stressed than others.
- Stressed people have different eating habits than others.
- Price sensitivity varies among people of different educational backgrounds.

As I was familiar with each member's competencies and background in the group, I took the initiative of designing the hypothesis and assigning respective tasks to each group member.

Now that we had figured out a problem and listed down our approach to solving it, we were ready to dive into the statistical analysis part. However, we had one last bit of ambiguity to resolve. As the project itself was open-ended, we had no restriction on the tool, platform, or programming language that we want to use. And our team consisted of people from backgrounds ranging from finance, hospitality to music, so did their analytical skillsets (SAS, Python Excel, etc.). But our project required us to present all the analysis in one platform. I realized it would be very difficult for everyone to learn one tool and start working and suggested that we use the tool that is comfortable for each individual but we keep a shared dataset, so any modifications would be accessible to everyone. As I had strong proficiency in python, I volunteered to compile all analysis to one platform in the end. This ensured our project is easily readable.

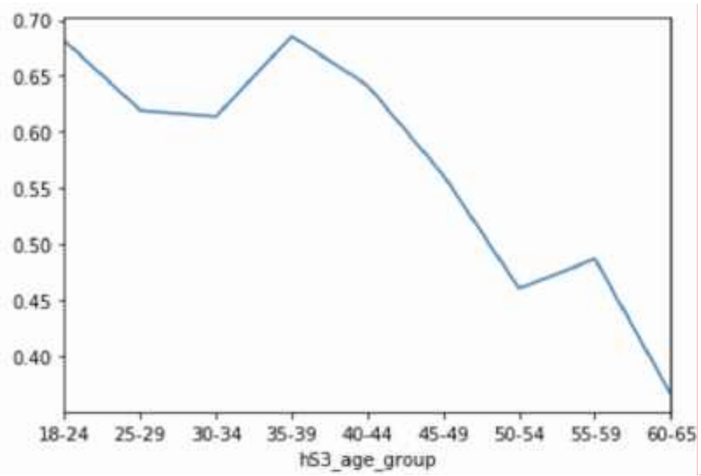
I believe my analytical skills to break down a business problem and people skills to succeed in a collaborative environment make me a great fit for the job.

Analysis

Stress keeps me from being the type of person I really want to be - Q1ra

We took these as signs of stress and wanted to see if there is correlation between age and stress.

Hypothesis 1: People in younger age groups are more stressed than others.



Commented [1]: data.groupby(['hS3_age_group'])['Q1ra_is_stressed'].mean().plot()

Table of hS3_age_group by Q1ra_is_stressed			
hS3_age_group	Q1ra_is_stressed		
Frequency Percent Row Pct Col Pct	0	1	Total
18-24	36 3.59 31.86 8.29	77 7.68 68.14 13.53	113 11.27
25-29	43 4.29 38.05 9.91	70 6.98 61.95 12.30	113 11.27
30-34	44 4.39 38.60 10.14	70 6.98 61.40 12.30	114 11.37
35-39	33 3.29 31.43 7.60	72 7.18 68.57 12.65	105 10.47

Statistic	DF	Value	Prob
Chi-Square	8	43.4008	<.0001
Likelihood Ratio Chi-Square	8	43.5937	<.0001
Mantel-Haenszel Chi-Square	1	33.1581	<.0001
Phi Coefficient		0.2080	
Contingency Coefficient		0.2037	
Cramer's V		0.2080	

Commented [2]:
```
ods rtf file='Age_group_Stress_ChiSquared';
proc freq data=srg_data;
tables hS3_age_group*Q1ra_is_stressed/ChiSq;
run;
ods rtf close;
```

As suggested by our descriptive analysis and Chi-Squared test above we see there is strong correlation between age-group and stress. As we can see, age-groups 18-24 and 35-39 have the highest percentages of stressed people in them (68.1% and 68.6% respectively). Now let's see if stressed people eat differently than others.

Hypothesis 2: Stressed people have different eating habits than others.

Table of Q1ra_is_stressed by type_of_restaurant					
Q1ra_is_stressed	type_of_restaurant				
Frequency Percent Row Pct Col Pct	At-home?	Carry_Out_or_Delivery_Pizz	Casual Dining	Fast_Casual	Fast_Food_Restaurant
0	288 28.71 66.36 44.44	20 1.99 4.61 36.36	20 1.99 4.61 52.63	10 1.00 2.30 28.57	56 5.58 12.90 39.16
1	360 35.89 63.27 55.56	35 3.49 6.15 63.64	18 1.79 3.16 47.37	25 2.49 4.39 71.43	87 8.67 15.29 60.84
Total	648 64.61	55 5.48	38 3.79	35 3.49	143 14.26

Commented [3]:
```
ods rtf file='Stressed_peole_eating_out_ChiSquared';
proc freq data=srg_data;
tables Q1ra_is_stressed*type_of_restaurant/ChiSq;
run;
ods rtf close;
```

Statistic	DF	Value	Prob
Chi-Square	8	12.6265	0.1254
Likelihood Ratio Chi-Square	8	12.8058	0.1187
Mantel-Haenszel Chi-Square	1	0.0660	0.7973
Phi Coefficient		0.1122	
Contingency Coefficient		0.1115	
Cramer's V		0.1122	

Though statistically not very significant, the crosstab analysis shows that people who are stressed usually eat out more than others, also they prefer carryout pizza and fast food more than others.

Apart from eating out, another way to measure eating disorder is by knowing how often people snack.

Hypothesis 3:Do people who feel stressed-out also snack more?

Table 2 of Is Stressed by Gender			
Controlling for Works to much=Yes			
Is Stressed			
Frequency Percent Row Pct Col Pct	Female	Male	Total
No	65 14.25 48.15 31.10	70 15.35 51.85 28.34	135 29.61

Commented [4]: ODS RTF File = 'Stressed.rtf';
PROC FREQ DATA=srg_data NLEVELS;
TABLE Q1rb_Work_too_much*Q1ra_is_stressed *S2;
RUN;
ODS RTF CLOSE;

Yes	144 31.58 44.86 68.90	177 38.82 55.14 71.66	321 70.39
Total	209 45.83	247 54.17	456 100.00

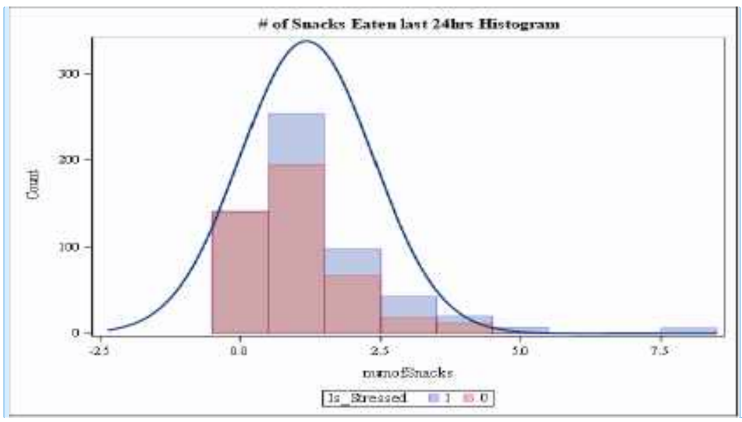
It shows that 68.9% of females and 71.7% of males who said to have worked too much also felt stressed.

Table 4 of Snacked in the last 24hrs by Gender			
Controlling for Works to much=Yes Is Stressed=Yes			
Frequency Percent Row Pct Col Pct	**Female**	**Male**	**Total**
No	31 9.66 40.79 21.53	45 14.02 59.21 25.42	76 23.68

Commented [5]: ODS RTF File = 'Work to much.rtf';

```
PROC FREQ DATA=srg_data NLEVELS;
   TABLE
Q1rb_Work_too_much*Q1ra_is_stressed*Q25 *S2;
RUN;
ODS RTF CLOSE;
```

Yes	113 35.20 46.12 78.47	132 41.12 53.88 74.58	245 76.32
Total	144 44.86	177 55.14	321 100.00



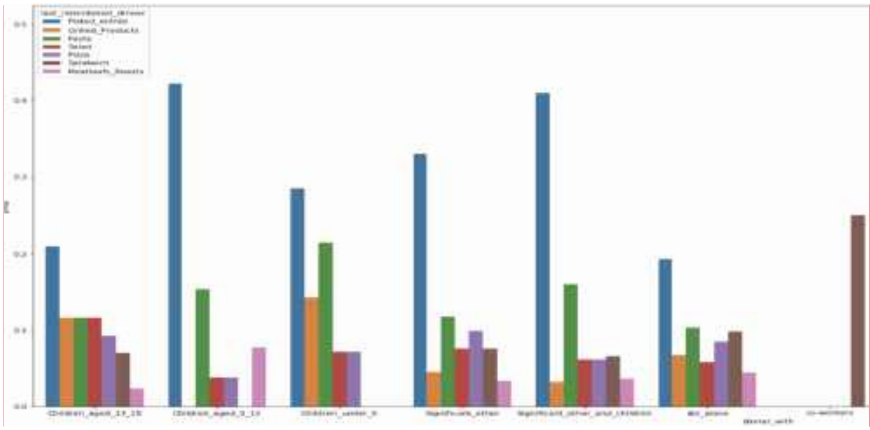
Commented [6]: ODS RTF FILE = 'Snack Histogram.rtf';
PROC sgplot data=srg_data;
histogram numofSnacks / scale = count binstart = 0 binwidth = 1 transparency=0.5 group =Q1ra_is_stressed;
density numofSnacks / type = normal;
title '# of Snacks Eaten last 24hrs Histogram';
run;
ODS RTF CLOSE;

This graph and table above suggest that people who feel stressed also snack more than others.

Now that we have shown people who feel they are stressed also have some sort of eating disorder- either eating out too much or snacking more than regular- rather than advising them to stop snacking or eating out, as it is very difficult to induce a habit, we suggest healthier snack and dinner options.

But before we suggest coming up with a new product, we want to see what people prefer right now. Of course, our analysis is based on the assumption that what people had last night was a representation of their regular behavior.

How and what people eat?



Commented [7]: import matplotlib.pyplot as plt
import seaborn as sns
from matplotlib.ticker import PercentFormatter

dinner_list = ['Plated_entrée','Pasta','Pizza','Salad','Sandwich','Grilled_Products','Meatloafs_Roasts']

x, y, hue = "dinner_with","prop","last_remmbered_dinner"
from matplotlib.pyplot import show

f, axes = plt.subplots(ncols=1, nrows=2,sharey=True,figsize=(30,30))

prop_df = (data[hue]
    .groupby(data[x])
    .value_counts(normalize=True)
    .rename(y)
    .reset_index())

sns.barplot(x=x, y=y, hue=hue, data=prop_df[prop_df['last_remmbered_dinner'].isin(dinner_list)], ax=axes[1])

- People in general prefer a Plated entree, which is usually a balanced meal of protein and carbohydrates (Rice with chicken, Chicken with vegetables/potatoes etc.). While other options such as Pizza, Pasta, Sandwiches and Grilled products are among other popular choices.
- We can see that Plated entree(balanced meal) consists of up to 42% of meals in family dinners (people who dined with children aged 5-12), while it's only 19% in case of people who ate alone. On the other hand, people who eat alone, their dinner choices include more variety than regular people.
- There are a few reasons why "**people who ate alone**" are of particular interest to us. Usually, people who eat alone tend to be more depressed rather than those who eat with family, friends or a colleague. And while people who ate with others probably did not have complete freedom to make their dinner choices (or influenced by others), the people who ate alone are the people who made their own dinner choices, meaning they have more freedom to choose their dinner.
- This is beneficial for us due to two reasons. First, we will get a very accurate estimate of the person's preference. A person buying pizza for himself as opposed to a spouse choosing a restaurant to eat.
- Secondly, because this group has more freedom to buy for themselves, they make a good target group for marketing experiments.

Commented [8]: data[data['dinner_with'] == 'Children_aged_5_12']['last_remmbered_dinner'].value_counts(normalize=True)

Commented [9]: data[data['dinner_with'] == 'ate_alone']['last_remmbered_dinner'].value_counts(normalize=True)

Dinner choice of people who eat alone

Plated_entrée	0.192825
Pasta	0.103139
Sandwich	0.098655
Pizza	0.085202
Grilled_Products	0.067265
Salad	0.058296
Meatloafs_Roasts	0.044843
Mexican_food	0.044843
Soups_stews_chili_gumbo	0.035874
Fast_Food_Restaurant	0.035874

Dinner choice of Overall Population

Plated_entrée	0.310451
Pasta	0.134221
Pizza	0.075820
Sandwich	0.072746
Salad	0.069672
Grilled_Products	0.051230
Meatloafs_Roasts	0.036885
Fast_Food_Restaurant	0.028689
Mexican_food	0.028689
Soups_stews_chili_gumbo	0.021516

Commented [10]: data[(data['dinner_with'] == 'ate_alone')]['last_remmbered_dinner'].value_counts(normalize=True)

data[(data['hS3_age_group'].isin(['18-24','35-39'])) & (data['dinner_with'] == 'ate_alone')]['last_remmbered_dinner'].value_counts(normalize=True)

Commented [11]: data['last_remmbered_dinner'].value_counts(normalize=True)

Do people who eat alone and are depressed eat differently than others?

- We noticed that, among the people who ate alone a lower percentage prefered a plated entree (19.2%, even lower among stressed adults who eat alone - 16%) compared to the overall group (31%).

- Top dinner choices overall and among people who feel stressed and eat alone are the following:

  Plated_entree
  Pasta
  Sandwich
  Pizza
  Grilled products

- Top snack choices among people who feel stressed are the following:

Commented [12]: data[(data['hS3_age_group'].isin(['18-24','35-39'])) & (data['dinner_with'] == 'ate_alone')]['last_remmbered_dinner'].value_counts(normalize=True)

Snacks chosen by those who feel stress keeps them from being the type of person they want to be		
Snack Type	Count	%
Salty snacks	25	16%
Sweet snacks	21	14%
Fruit and Vegetables	15	10%
Dairy	13	8%
Dips	12	8%
Nuts, Grains and Seeds	11	7%
Meat snack	11	7%
Breakfast snack	11	7%
Beverage Snack	10	6%
Heat and eat snacks (Frozen)	8	5%
Bars	6	4%
Some other snack	6	4%
Sandwich Type	5	3%
Total	154	100%

Based on these top dinner and snack choices of our target group we came up with the following products to test the intent of buying.

Dinner Product 1: Archer Farms (Premium Brand) Multigrain Sandwich 200 calories $4.00 (Healthy Sandwich)
Dinner Product 2: Great Value (Walmart Brand) Totino's triple meat Pizza 250 calories $2.50 (Tasty Pizza)

Snack Product 1: Welch's Fat & Gluten free Mixed Fruit Snacks 80 calories/serving $3.99
Snack Product 2: Tostitos Scoops! Tortilla Chips, Party Size 140 calories/serving $2.98

The reason behind choosing pizza and sandwich is that they represent the eating out option (as opposed to Plated entree) . Also, people who eat with their family and spend a long time cooking can be targeted using healthy quick meals as average cooking time for someone who dined with his/her family and children is about 9 minutes higher than usual

Commented [13]: print('Avg Cooking time overall :' ,data['total_time_spent_cooking_int'].mean())
print('Avg Cooking time for people who had dine with family with children:',data[data['dinner_with'] == 'Significant_other_and_children']['total_time_spent_cooking_int'].mean())
print('Avg Cooking time for people who had dine with family without children:',data[data['dinner_with'] == 'Significant_other']['total_time_spent_cooking_int'].mean())

Now, we test our hypothesis that these products will actually be bought by our target group. For this we modeled the intent of the user based on the survey questions they have answered.

Hypothesis 4: What factors influence intention of buying Product 1 and Product 2 among stressed people.

Model Intention of customer based on the following factors:

Price,Store Brand, Taste, Quality, Natural Product, Calorie

1. When it comes to food, I'm primarily a price shopper(**Price**) **Q30rbb_food_price_shopper**
2. I prefer to buy store brands(**store brand**) **Q30ram**
3. I eat for taste enjoyment more than for health purposes (**Preference to taste**) **Q30rbm**
4. I buy based on quality, not price(**Quality**) **Q30rbc**
5. I go out of my way to buy products that are all natural(**Natural Product**) **Q30rae**
6. I don't allow junk food in my home(**Calorie**) **Q30rao**

If the user answered Questions 1,2,3 as agreed/strongly agreed , then he has shown intention to buy product 1 (choice =1)

If the user has answered Questions 4,5,6 as agreed/strongly agreed , then he has shown intention to buy product 2 (choice =2)

If the user answered similarly for both question sets, then Unsure (choice = 3) .

Commented [14]:
```
def create_y_label(x):
    profile_1 = x['Q30rbb_food_price_shopper'] + x['Q30ram_prefer_store_brand'] + x['Q30rbm_prefer_taste_to_health']
    profile_2 = x['Q30rbc_prefer_quality']+x['Q30rae_prefer_natural_product']+x['Q30rao_calorie_conscious']

    if profile_1 > profile_2:
        return "Tasty_Pizza"
    elif profile_1 < profile_2:
        return "Healthy_sandwich"
    else:
        return "Unsure"

data['choice'] = data.apply(lambda x: create_y_label(x),axis=1)
```

We ran a generalized Logistic model with Three choices(Tasty_Pizza,Healthy_Sandwich,Unsure) with demographic features such as Gender,income group,age etc.

Model: Choice = Gender + income_group + age + is_stressed + is_stressed*age

Odds Ratio Estimates				
Effect	choice	Point Estimate	95% Wald Confidence Limits	
S2 Male vs Female	Healthy_sandwich	0.466	0.170	1.280
S2 Male vs Female	Tasty_Pizza	0.612	0.272	1.377
D6 $100,000 but less than $200,000 vs Less than $15,000	Healthy_sandwich	1.520	0.255	9.040
D6 $100,000 but less than $200,000 vs Less than $15,000	Tasty_Pizza	0.893	0.196	4.064
D6 $15,000 but less than $25,000 vs Less than $15,000	Healthy_sandwich	1.583	0.358	6.999
D6 $15,000 but less than $25,000 vs Less than $15,000	Tasty_Pizza	1.802	0.538	6.032
D6 $200,000 but less than $300,000 vs Less than $15,000	Healthy_sandwich	<0.001	<0.001	>999.999
D6 $200,000 but less than $300,000 vs Less than $15,000	Tasty_Pizza	0.164	0.009	3.082
D6 $25,000 but less than $35,000 vs Less than $15,000	Healthy_sandwich	0.273	0.025	3.003
D6 $25,000 but less than $35,000 vs Less than $15,000	Tasty_Pizza	1.575	0.420	5.900
D6 $300,000 but less than $500,000 vs Less than $15,000	Healthy_sandwich	>999.999	<0.001	>999.999
D6 $300,000 but less than $500,000 vs Less than $15,000	Tasty_Pizza	>999.999	<0.001	>999.999
D6 $35,000 but less than $50,000 vs Less than $15,000	Healthy_sandwich	3.294	0.508	21.354
D6 $35,000 but less than $50,000 vs Less than $15,000	Tasty_Pizza	2.946	0.568	15.293
D6 $50,000 but less than $75,000 vs Less than $15,000	Healthy_sandwich	0.349	0.056	2.186
D6 $50,000 but less than $75,000 vs Less than $15,000	Tasty_Pizza	1.053	0.343	3.229
D6 $75,000 but less than $100,000 vs Less than $15,000	Healthy_sandwich	4.518	0.676	30.182
D6 $75,000 but less than $100,000 vs Less than $15,000	Tasty_Pizza	2.552	0.468	13.898
D6 Decline to answer vs Less than $15,000	Healthy_sandwich	1.391	0.150	12.897
D6 Decline to answer vs Less than $15,000	Tasty_Pizza	1.547	0.262	9.120

Commented [15]: /*
S2 - Gender
D6 - income group
S3 - age
*/
ods rtf file='Choice Intention Generalized Logit regression';
proc logistic data=srg_filtered_data;
class S2(ref='Female') D6(ref='Less than $15,000') Q1ra_is_stressed /param = ref;
model choice(event='Tasty_Pizza') = S2 D6 S3 Q1ra_is_stressed S3*Q1ra_is_stressed / link=glogit;
run;
ods rtf close;

Maximum Likelihood Estimates of Age

S3		Healthy_sandwich	1	0.00816	0.0261	0.0979	0.7543
S3		Tasty_Pizza	1	0.0211	0.0210	1.0107	0.3147

- People who earn between 50K-75K are more likely to buy Tasty Pizza (Odds ratio = 1.053) compared to People who earn less than 15K, while they are less likely (Odds Ratio = 0.34) to buy a healthy sandwich compared to a lower income group, despite having higher income.

- On the other hand, people who earn between 75K and 100K are more likely to buy a healthy sandwich (Odds Ratio of 1.96 = 4.518 - 2.552) compared to the tasty Pizza.

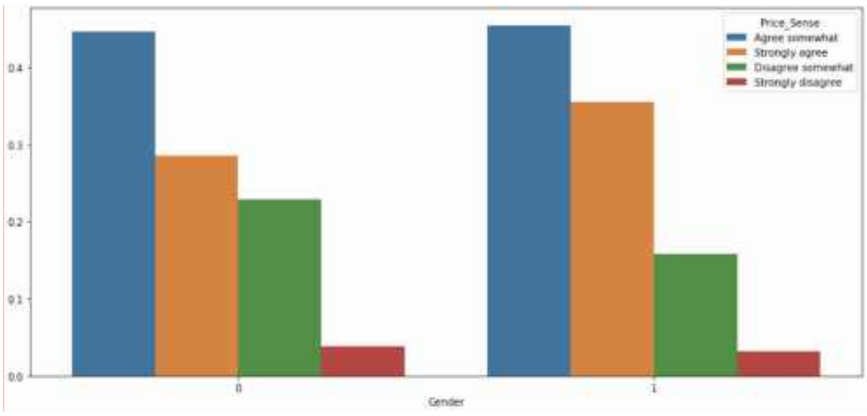
- Also Men are less likely to buy either a Healthy sandwich or tasty Pizza compared to Women and with increase in age the log odds ratio of Tasty Pizza is higher than Healthy Sandwich.

- Among all the people who eat alone and are stressed, the log odds ratio of intention to buy a “Tasty Pizza” will be higher among people who earn between 75K-100K and lower among people who earn between 200K-300K.

- So our target group is young females who earn between 75K-100K, this is the group that has shown more intention to buy a healthy sandwich (Quality, less preference to price) and therefore prime target customers for F&B. For other customers who are still keen on having tasty pizza, F&B should introduce a healthier but tastier pizza option that is also cheap.

Now we can market our product to these population groups based on their price sensitivity.

Does price sensitivity vary between genders?

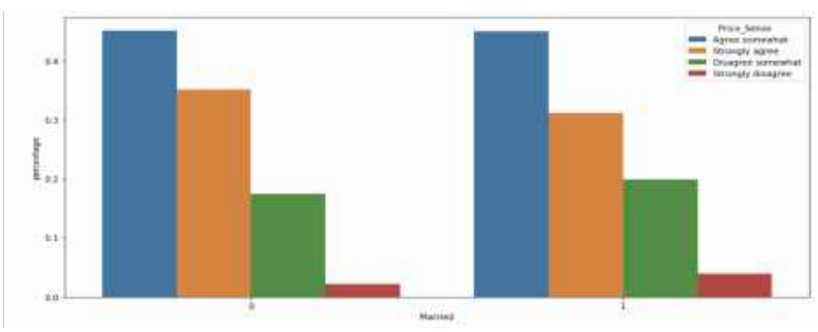


Commented [16]: x, y, hue = "Gender","percentage","Price_Sense"

```
from matplotlib.pyplot import show

f, axes = plt.subplots(ncols=1, nrows=2,sharey=True,figsize=(15,15))

prop_df = (df5[hue]
          .groupby(df5[x])
          .value_counts(normalize=True)
          .rename(y)
          .reset_index())

sns.barplot(x=x, y=y, hue=hue, data=prop_df, ax=axes[1])
```

More females (1) believe they agree to shopping based on price (Orange bar), while males (0) there are more who disagree with buying on price. So, women in general are more price sensitive.

Are married and single people more price sensitive?

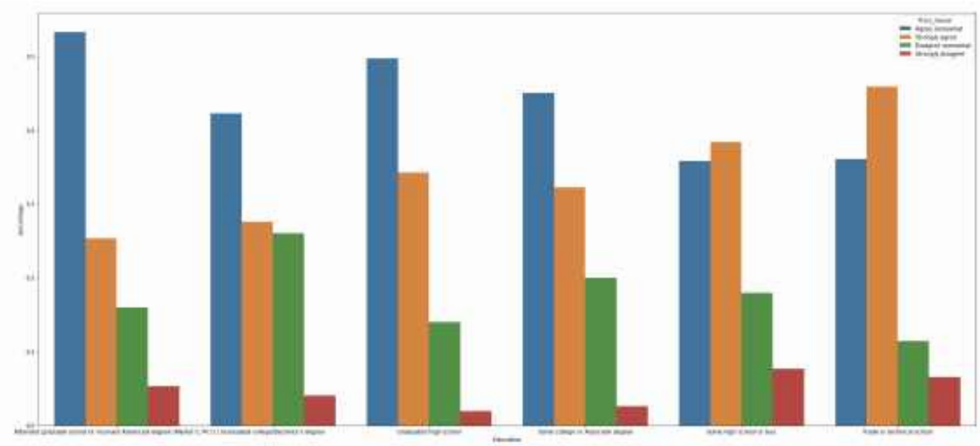


Commented [17]: x, y, hue = "Married","percentage","Price_Sense"

```
from matplotlib.pyplot import show

f, axes = plt.subplots(ncols=1, nrows=2,sharey=True,figsize=(15,15))

prop_df = (df5[hue]
          .groupby(df5[x])
          .value_counts(normalize=True)
          .rename(y)
          .reset_index())

sns.barplot(x=x, y=y, hue=hue, data=prop_df, ax=axes[1])
```

People who are married/living with their partner (1) seem to be less price sensitive compared to single people, though the difference does not seem to be significant.

Does price sensitivity vary among people of different educational backgrounds?



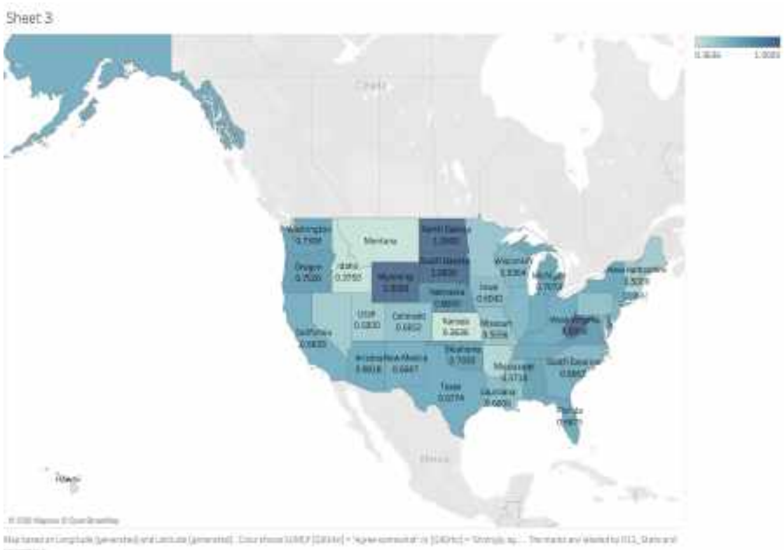
Commented [18]: x, y, hue = "Education","percentage","Price_Sense"
from matplotlib.pyplot import show

```
f, axes = plt.subplots(ncols=1, nrows=2,sharey=True,figsize=(30,30))

prop_df = (df5[hue]
        .groupby(df5[x])
        .value_counts(normalize=True)
        .rename(y)
        .reset_index())

sns.barplot(x=x, y=y, hue=hue, data=prop_df, ax=axes[1])
```

Does price sensitivity vary among states?

Connecticut	0.7692	10
Oregon	0.7500	9
Virginia	0.7500	9
Alabama	0.7333	11
Oklahoma	0.7333	11
Washington	0.7308	19
Michigan	0.7073	29
New York	0.7015	47
Tennessee	0.6957	16
Florida	0.6875	44
Arizona	0.6818	15
Indiana	0.6818	15
Kentucky	0.6800	17

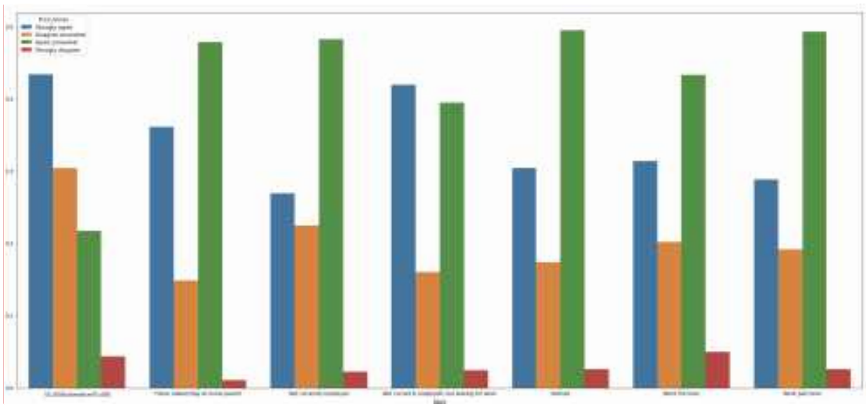
Illinois	0.6250	25
New Jersey	0.6250	15
North Carolina	0.6176	21
Iowa	0.6000	6
Louisiana	0.6000	6
Mississippi	0.5714	4
[illegible]	0.5714	32
[illegible]	0.5652	13
[illegible]	0.5556	15



We can see that upto 70-80% people buy based on quality not price in most states, the exceptions are Pennsylvania, Colorado, Missouri where only 55-60% people buy on quality. Which means our pricing of the product should be different in these states.

People with bachelor's or more advanced degrees seem to be less price sensitive compared to others, which is expected as they are more likely to have a higher income.

Does price sensitivity vary among people of different work status?



As expected unemployed people are more price sensitive than people who are currently employed.

Now, we want to test which of the above factors has the most impact on price sensitivity.

Hypothesis 5: Which demographic factors has the most influence on price sensitivity.

Price_sensitivity = Age + Gender + ethnicity + Marital_status + Educational_background + location

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Gender 1 vs 0	1.458	1.049	2.027
ethnicity African_American vs White	0.938	0.581	1.515
ethnicity American_Indian vs White	0.562	0.093	3.391
ethnicity Asian/Oriental vs White	0.554	0.278	1.104
ethnicity Decline_to_answe vs White	0.753	0.174	3.265
ethnicity Hispanic vs White	2.267	0.825	6.225
ethnicity Hispanic_White vs White	0.856	0.290	2.525
ethnicity other vs White	1.488	0.147	15.026
Married 1 vs 0	0.854	0.575	1.267
Education 1 vs 0	1.871	0.801	4.370
Education 2 vs 0	1.515	0.541	4.241
Education 3 vs 0	1.206	0.542	2.682
Education 4 vs 0	0.903	0.395	2.064
Education 5 vs 0	1.428	0.543	3.756
Work 1 vs 0	0.703	0.371	1.334
Work 2 vs 0	0.652	0.300	1.417
Work 3 vs 0	0.935	0.401	2.179
Work 4 vs 0	0.757	0.371	1.546
Work 5 vs 0	0.933	0.430	2.026

Commented [19]: x, y, hue = "Work","percentage","Price_Sense"
```
from matplotlib.pyplot import show

f, axes = plt.subplots(ncols=1, nrows=2,sharey=True,figsize=(30,30))

prop_df = (df5[hue]
        .groupby(df5[x])
        .value_counts(normalize=True)
        .rename(y)
        .reset_index())

sns.barplot(x=x, y=y, hue=hue, data=prop_df, ax=axes[1])
```

Education	1	1	0.3792	0.1978	3.6756	0.0552
Education	2	1	0.1684	0.3056	0.3037	0.5815
Education	3	1	-0.0598	0.1542	0.1502	0.6983
Education	4	1	-0.3495	0.1691	4.2739	0.0387

This reinforces a few of our assumptions. For example, females are more price sensitive than male. We also noticed a difference in price sensitivity among different ethnic groups; People with Hispanic ethnic background being more price sensitive compared to people with white background. And as expected people with higher educational background show lower price sensitivity.

Recommendations:

- To summarize, F&G should introduce healthy sandwiches as a premium product in their outlet or in premium supermarkets in a city (busy working and single people live there) , where the demographics has a median income of around 75-100K. Additionally, the pricing should be sensitive to the location the demographic is in. For example, in Philadelphia, PA the pricing should be lower compared to New York, NY though they are located nearby the price sensitivity is different between the two states. This group is the target customer for F&G as they have a higher intention and less price sensitivity compared to other groups.

- For income groups higher than 100K, F&G can afford to experiment with costlier products which are of even higher quality and have more personal branding.

- For low income groups which are even more sensitive to price F&G can introduce healthy snacks of varying price range with more weightage to taste and price, and less weightage to quality.