



# Project 1: Let's Eat!

---

TEAM MEMBERS:

KASIM AHMAD, ALEX BLAKEMORE, ANN NGUYEN,  
LISA (ISLAS) TAYLOR, & EDWIN VARGAS



MERCYHURST  
UNIVERSITY

## **Background:**

Using survey data provided by college students of Mercyhurst University, we want to explore how food preferences are related to a student's background. It would be interesting to explore if there's a correlation between this sample size's exercise habits, GPAs, and other factors in relation to what they prefer to eat. We plan to organize this survey data, explore it further with our questions, and perform calculations/visualizations to present this to our class.

# **Hypothesis:**

**How does a person's external environment and influences affect people's eating habits and perception of food?**

# Questions????

1. How often do they cook/exercise?
  - a. Any correlation between gender and exercise regimen?
  - b. Does playing sports affect how often you exercise?
  - c. Does living situation affect how often someone cooks?
2. How do parents education influence their eating habits?
  - a. Any correlations between parent's education and student's eating habits?
  - b. Any correlations between parent's education and cooking at home?
  - c. Compare student cooking, parent cooking, or student's eating out?





# How did we get here????



1. Found data sets recommendations from Git Lab.
2. Looked at Kaggle & found lots of different data sets.
3. As a team we chose food, because everyone loves food.
  1. When we came across this data set, we felt there was enough information to start answering questions we had for this data.
  2. The information would help us approve or disprove our hypothesis

# How it started, how it ended.....

---



	GPA	Gender	breakfast	calories_chicken	calories_day	calories_scone	coffee	comfort_food	comfort_food_reasons	comfort_food_reasons_coded	...	s
0	2.4	2	1	430	NaN	315.0	1	none	we dont have comfort	9.0	...	
1	3.654	1	1	610	3.0	420.0	2	chocolate, chips, ice cream	Stress, bored, anger	1.0	...	
2	3.3	1	1	720	4.0	420.0	2	frozen yogurt, pizza, fast food	stress, sadness	1.0	...	
3	3.2	1	1	430	3.0	420.0	2	Pizza, Mac and cheese, ice cream	Boredom	2.0	...	
4	3.5	1	1	720	2.0	420.0	2	Ice cream, chocolate, chips	Stress, boredom, cravings	1.0	...	
...	...	...	...	...	...	...	...	...	...	...	...	
120	3.5	1	1	610	4.0	420.0	2	wine. mac and cheese, pizza, ice cream	boredom and sadness	NaN	...	
121	3	1	1	265	2.0	315.0	2	Pizza / Wings / Cheesecake	Loneliness / Homesick / Sadness	NaN	...	
122	3.882	1	1	720	NaN	420.0	1	rice, potato, seaweed soup	sadness	NaN	...	
123	3	2	1	720	4.0	420.0	1	Mac n Cheese, Lasagna, Pizza	happiness, they are some of my favorite foods	NaN	...	
124	3.9	1	1	430	NaN	315.0	2	Chocolates, pizza, and Ritz.	hormones, Premenstrual syndrome.	NaN	...	

125 rows × 61 columns

RAW DATA

```
# Remove unnecessary columns
food_df = food_df.drop(["comfort_food", "comfort_food_reasons", "diet_current", "eating_changes", "father_profession", "fav_cuisine"])
food_df
```

```
# Replace values in the first column and last column
food_df["GPA"] = food_df["GPA"].replace(["Personal", "3.79 bitch", "Unknown"], ["3.6", "3.79", "3.7"])
food_df["weight"] = food_df["weight"].replace(["Not sure, 240", "144 lbs"], ["240", "144"])
food_df
```

```
# Remove the third row and fill the NaN cells with the previous value from the next row
food_df = food_df.drop([2])
food_df = food_df.reset_index(drop=True)
food_df = food_df.ffill(axis=0)
```

```
# Fill the NaN cells in the first row with the values chose to replace it
values = {"calories_day": 3.0, "cuisine": 1}
food_df = food_df.fillna(value=values, limit=1)
food_df
```

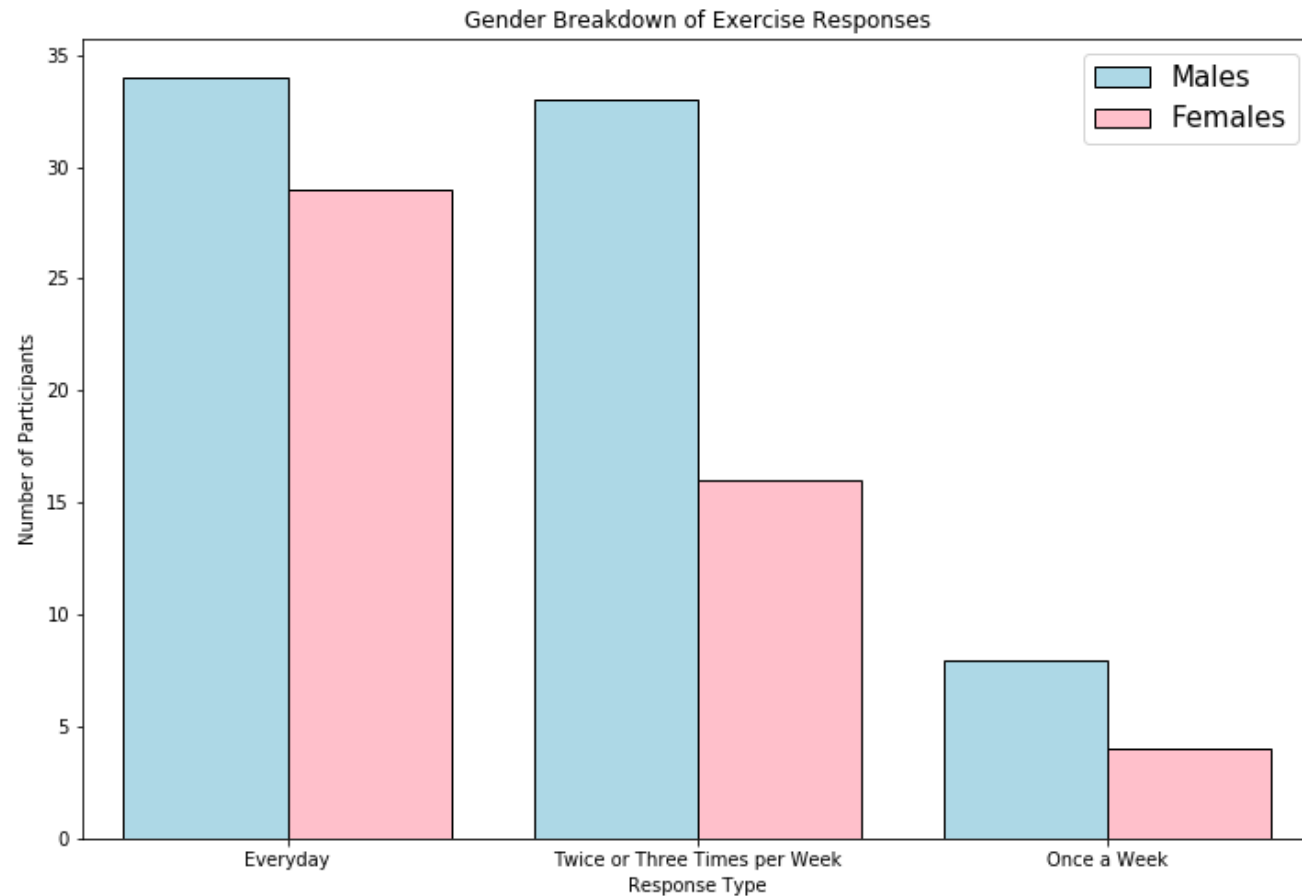
```
# Count the values
food_df.count()
```



	GPA	Gender	breakfast	calories_chicken	calories_day	calories_scone	coffee	comfort_food_reasons_coded	cook	comfort_food_reasons_coded.1	...
0	2.4	2	1	430	3.0	315.0	1	9.0	2.0	9	...
1	3.654	1	1	610	3.0	420.0	2	1.0	3.0	1	...
2	3.2	1	1	430	3.0	420.0	2	2.0	2.0	2	...
3	3.5	1	1	720	2.0	420.0	2	1.0	1.0	1	...
4	2.25	1	1	610	3.0	980.0	2	4.0	3.0	4	...
...	...	...	...	...	...	...	...	...	...	...	...
119	3.5	1	1	610	4.0	420.0	2	2.0	3.0	2	...
120	3	1	1	265	2.0	315.0	2	2.0	3.0	3	...
121	3.882	1	1	720	2.0	420.0	1	2.0	3.0	3	...
122	3	2	1	720	4.0	420.0	1	2.0	3.0	7	...
123	3.9	1	1	430	4.0	315.0	2	2.0	3.0	5	...

124 rows × 49 columns

# Any correlation between gender and exercise regimen?



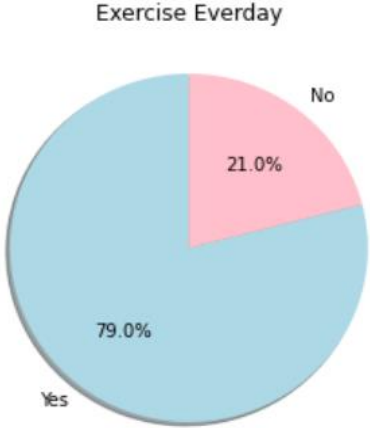
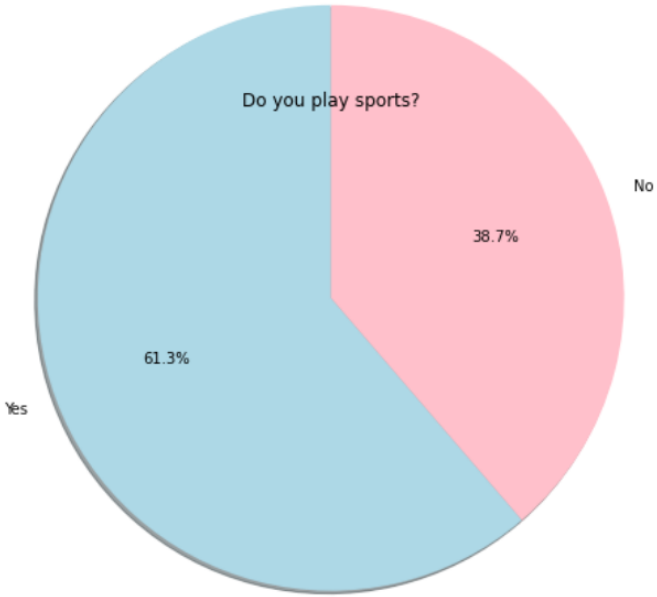
	Gender	on_off_campus	sports	exercise	cook
0	2	1	1	1	2
1	1	1	1	1	3
2	1	1	2	3	2
3	1	1	1	1	1
4	1	1	2	2	3
...	...	...	...	...	...
119	1	3	1	2	3
120	1	1	1	2	3
121	1	1	2	2	3
122	2	1	2	1	3
123	1	1	2	2	3

124 rows × 5 columns

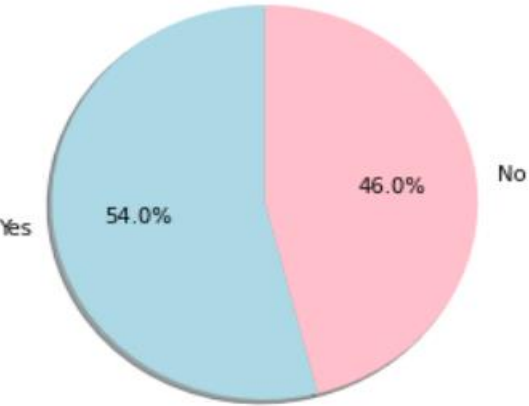
**Findings:** Males: 34, 33, 8. Females: 29, 16, 4

Less female participates than males. Responses show that they're even. There was no significant correlation to gender at the rate they exercise.

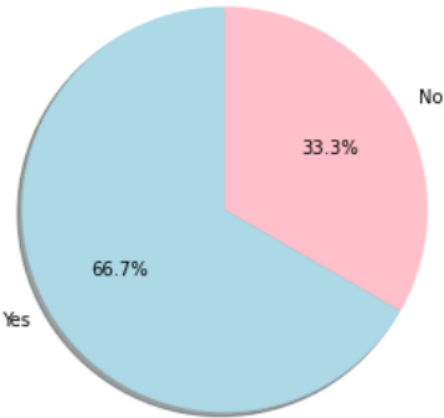
# Does playing sports affect how often you exercise?



Exercise Twice or Three Times per Week



Exercise Once a Week

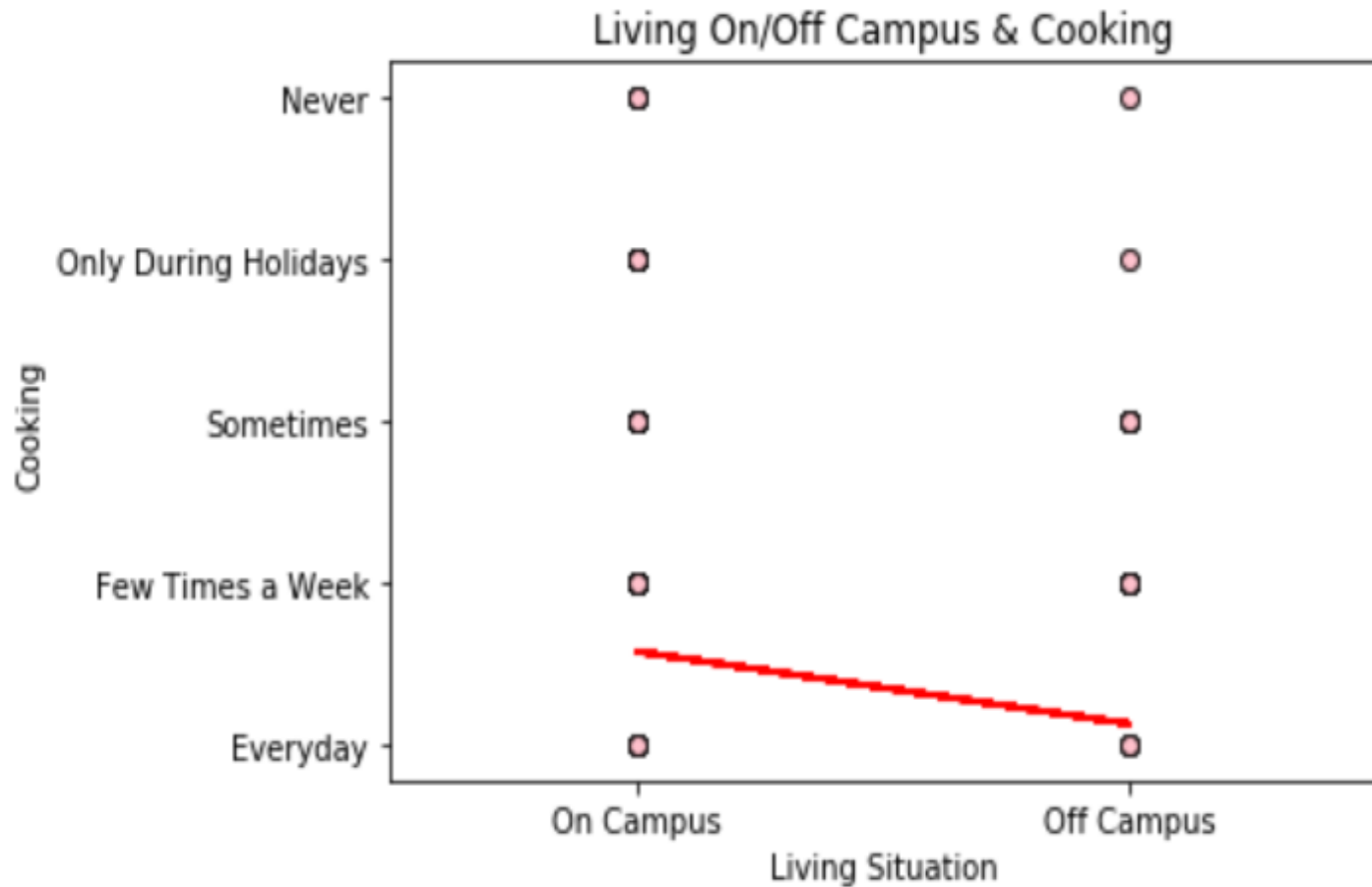


	Gender	on_off_campus	sports	exercise	cook
0	2	1	1	1	2
1	1	1	1	1	3
2	1	1	2	3	2
3	1	1	1	1	1
4	1	1	2	2	3
...	...	...	...	...	...
119	1	3	1	2	3
120	1	1	1	2	3
121	1	1	2	2	3
122	2	1	2	1	3
123	1	1	2	2	3

124 rows × 5 columns

**Findings:** Did the participants play sports? The sub categories included if they exercised daily? Or how often they exercised? 1st pie chart 62 participants, 2<sup>nd</sup> pie chart 50 participants, 3<sup>rd</sup> pie chart 12 participants.

# Does living situation affect how often someone cooks?



	Gender	on_off_campus	sports	exercise	cook
0	2	1	1	1	2
1	1	1	1	1	3
2	1	1	2	3	2
3	1	1	1	1	1
4	1	1	2	2	3
...	...	...	...	...	...
119	1	3	1	2	3
120	1	1	1	2	3
121	1	1	2	2	3
122	2	1	2	1	3
123	1	1	2	2	3

124 rows × 5 columns

The r-squared is: -0.1784371563011233

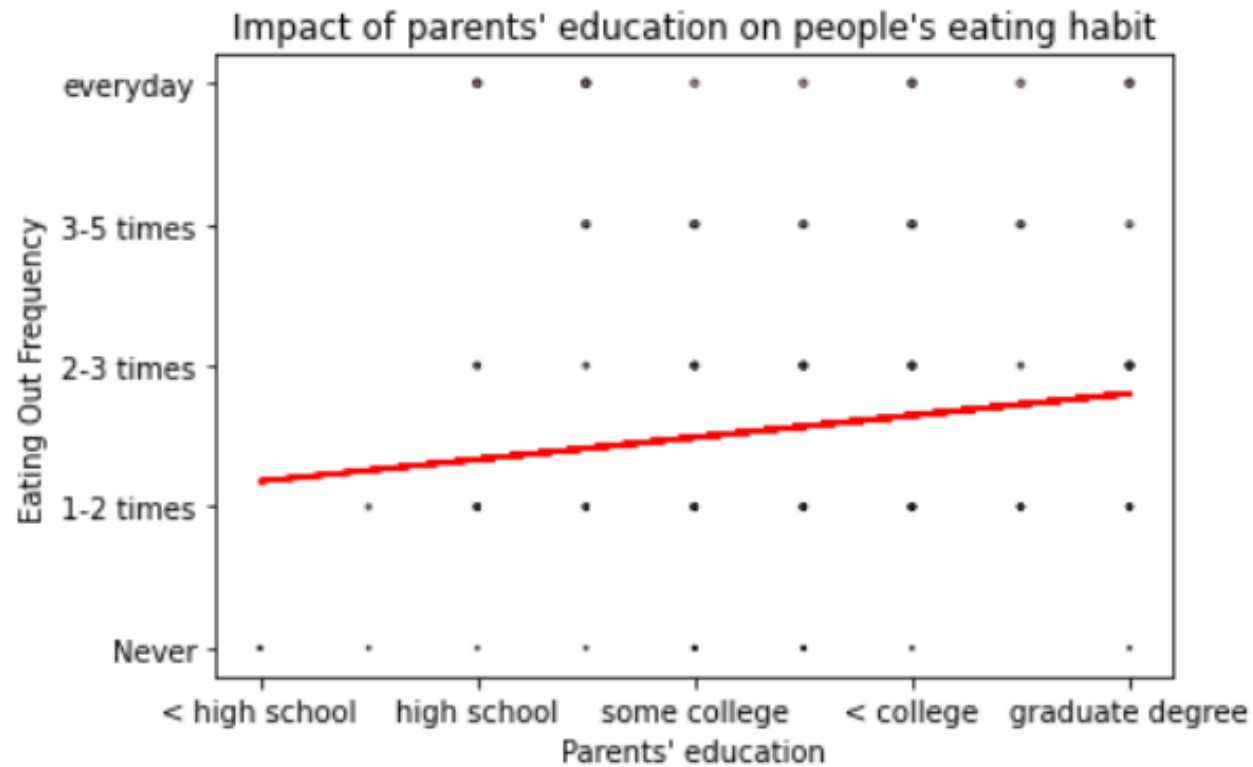
The equation of the regression line is:  $y = -0.45x + 3.35$

**Findings:** Linear regression at which rate students cooked. There is a small correlation, that if you lived off campus that you cooked more, vs living on campus

# Any correlations between parent's education and student's eating habits?

The r-squared is: 0.12938009729910624

The equation of the regression line is:  $y = 0.16x + 2.02$



	father_education	mother_education	parents_cook	eating_out	cook	weight
0	5	1	1	3	2	187
1	2	4	1	2	3	155
2	2	4	1	2	2	240
3	4	5	1	2	1	190
4	1	1	2	1	3	190
...	...	...	...	...	...	...
119	4	3	1	2	3	156
120	5	2	3	4	3	180
121	5	1	2	3	3	120
122	3	2	2	5	3	135
123	4	3	3	1	3	135

124 rows × 6 columns

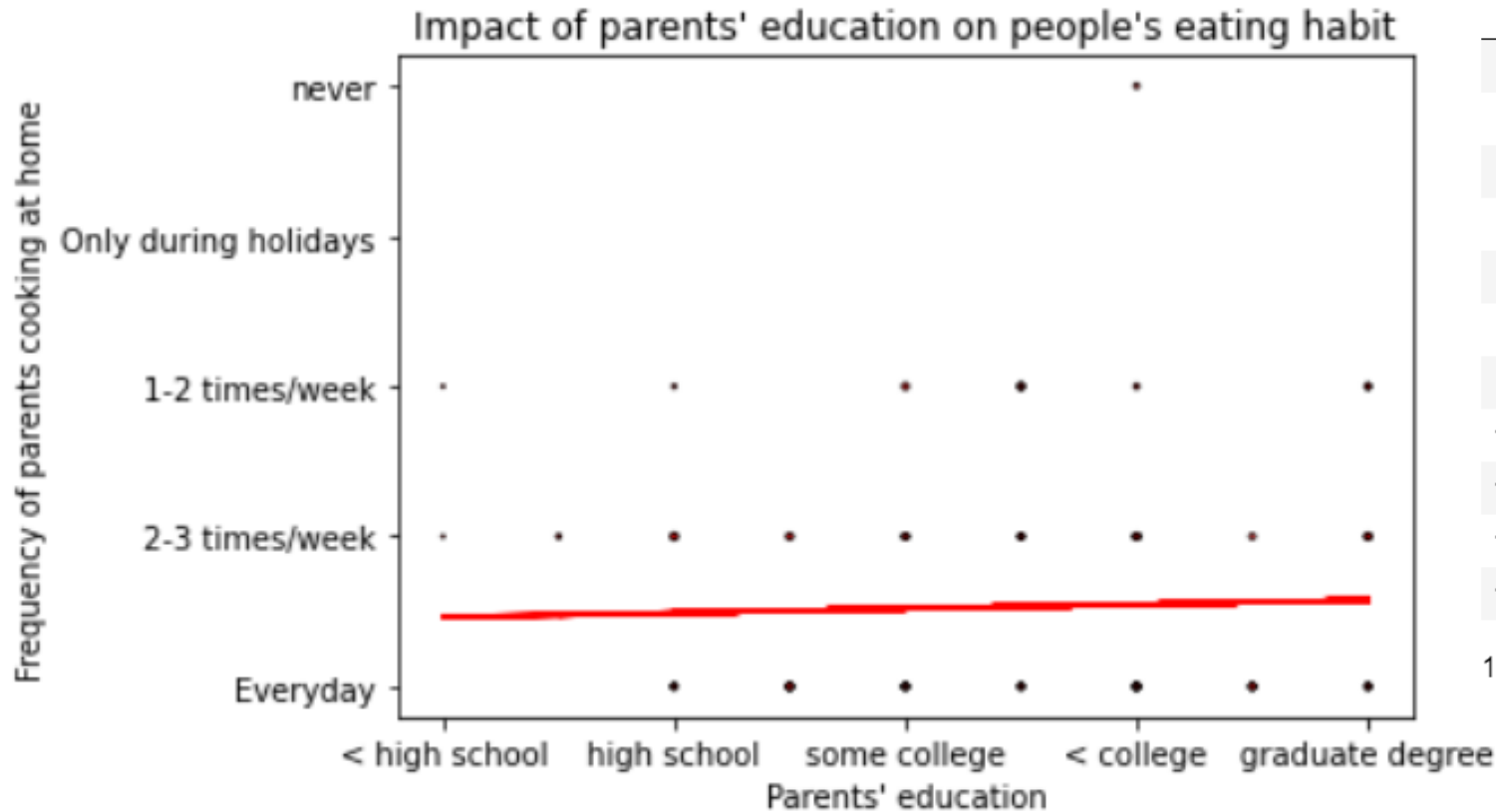
**Findings:** There is minimum relevance between a person's eating out habit and parents' education level. However, we can see that the higher level the parents' education is, the more often will people eat out



# Any correlations between parent's education and cooking at home?

The r-squared is: 0.03580824435819679

The equation of the regression line is:  $y = 0.03x + 1.43$

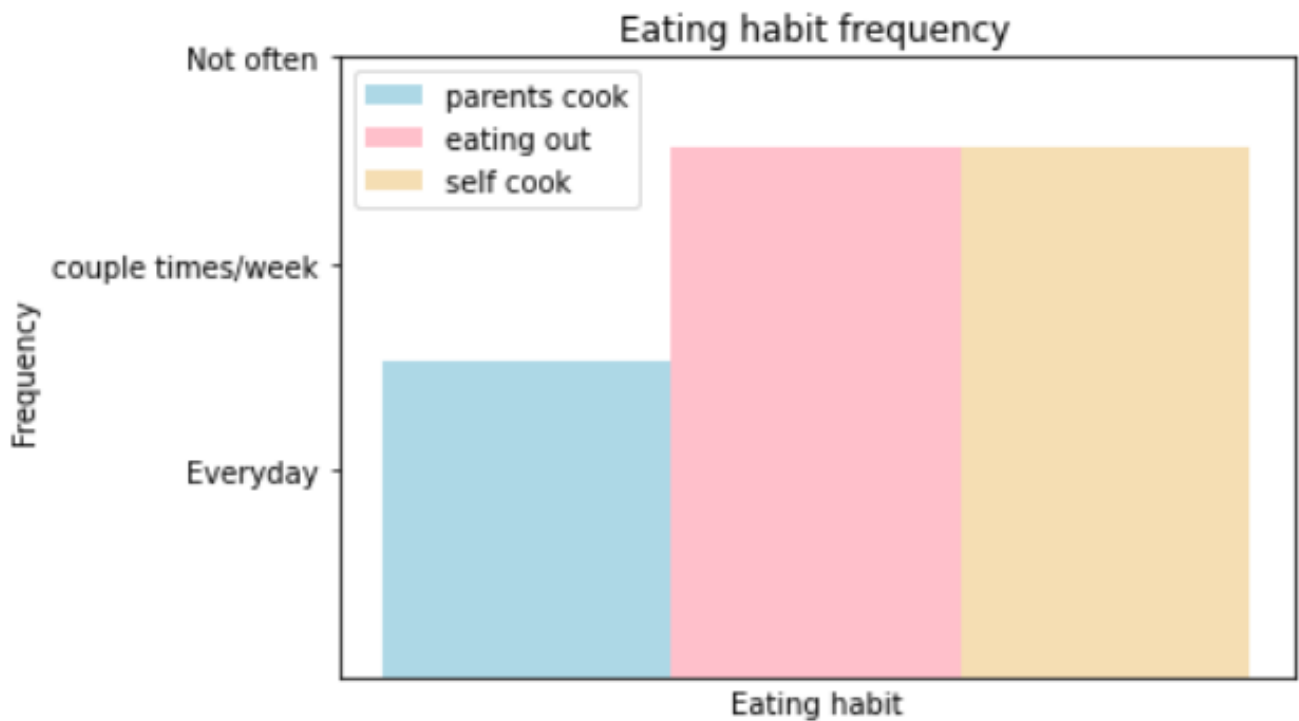


	father_education	mother_education	parents_cook	eating_out	cook	weight
0	5	1	1	3	2	187
1	2	4	1	2	3	155
2	2	4	1	2	2	240
3	4	5	1	2	1	190
4	1	1	2	1	3	190
...	...	...	...	...	...	...
119	4	3	1	2	3	156
120	5	2	3	4	3	180
121	5	1	2	3	3	120
122	3	2	2	5	3	135
123	4	3	3	1	3	135

124 rows × 6 columns

Findings: Most parents cook at home almost everyday or at least 2-3 times a week regardless of the education level

Compare student cooking, parent cooking, or student's eating out?



	father_education	mother_education	parents_cook	eating_out	cook	weight
0	5	1	1	3	2	187
1	2	4	1	2	3	155
2	2	4	1	2	2	240
3	4	5	1	2	1	190
4	1	1	2	1	3	190
...	...	...	...	...	...	...
119	4	3	1	2	3	156
120	5	2	3	4	3	180
121	5	1	2	3	3	120
122	3	2	2	5	3	135
123	4	3	3	1	3	135

124 rows × 6 columns

Findings: People have the tendency of eating out or having their parents cook instead of self-cooking

SO, WHAT IS THE DATA TELLING US?  
WELL, AS OF THIS POINT HERE, WE'RE  
NO LONGER ON SPEAKING TERMS.



SMAN

ModernAnalyst.com

## Data Analysis Conclusion:

Based on the conclusions found, our data suggests that there is no correlation between any of the relationships that we were examining.

- Any correlation found was insignificant ( $>0.01$ )

What we could do more time and resources:

- Find another data set that is related where we can merge the data for more further exploration and grows our sample population
- Find a different way to deal with the NaN values that would allow us to further explore the data



