horizontal line

**Group 5**

Breakout Room 5

Northwestern Bootcamp, IL 60001

(123) 456-7890

Boot Camp Project 1

**5th October 2021**

# GROUP 5 MEMBERS

Charles Kung  
Hai Pham  
Matt Keeley  
Shay Masood

# TOPIC

We are looking at proving or disproving the following hypothesis:

**Diabetes prevalence is higher in counties that have more fast food restaurants per capita.**

# PROJECT OUTLINE

We will be pulling Diabetes Prevalence data from the following sources as a dataset:

1. Collect dataset for Fast Food Restaurants in the US today (2018, 2019, 2021): <https://www.kaggle.com/khushishahh/fast-food-restaurants-across-us>
2. Collect dataset for Diabetes Prevalence in Illinois by County (2018, 2019 and 2021) <https://www.countyhealthrankings.org/app/illinois/2021/downloads>
3. County / Zip Code conversion: We added FIPS and county name with zip code so we can merge them. We would like to do the report at county level.
4. Use Pandas to clean up the datasets
5. Use Pandas to group and slice and dice
6. Use Matplotlib to create some interesting charts
7. Use a hypothesis test to prove or disprove the hypothesis
8. If time allows, apply scipy.stats linear regression to predict 2020 values.
9. Create summary
10. Present Jupyter Notebooks and discuss results

# LIMITATIONS

* Time delay for cause and effect of food and diet and diabetes prevalence
* There may be other factors besides fast food restaurants that may impact the diabetes prevalence

# RESEARCH QUESTIONS

## Does the prevalence of fast food restaurants in a county increase the percentage of adults with diabetes?

## We are missing data for 2020. We can predict data for 2020 through a linear regression. Can the pandemic explain a discrepancy?