# DATASCI415 Final Project Proposal

1. **Team Members:**
- Jiahao Cheng(jiahaoch@umich.edu)
- Kevin Fang(kevfang@umich.edu)
- Vy Dang(ivydang@umich.edu)

2. **Project Overview:**
   The goal of this project is to predict and understand an individual's risk for depression based on data from the National Health and Nutrition Examination Survey (NHANES). The project contains predictive modeling and inference analysis.

3. **Key Questions:**
- **Prediction Question:** Can we predict whether an individual is depressed based on people's lifestyle? (Example: Diet, Smoke, Alcohol)
- **Inference Question:** How do lifestyle factors (Example: Diet, Smoke, Alcohol) affect the odds of developing depression, and which of these factors are statistically significant?

4. **Methods:**
- **For Predictive Modeling:** Random Forest, Gradient Boosting, AdaBoost
- **For Inference:** Logistic Regression with Odd Ratio
- **Reasons for Choosing the Above Methods:**
   Random Forest: Aggregates multiple trees, reduces risks of overfitting, and we can use feature importance to select important predictors, can model robustly the complex relationships between health, lifestyles, and demographic factors.
   Boosting methods: The strength of boosting is comparable to that of a Random Forest, and we'd need that strength for the complex data that we'd like to model. Note that a risk of overfitting is present, so maximum iterations must be limited in practice.
   Logistic Regression: This is a straightforward and very interpretable model, making it the ideal choice to conduct inference on.

5. **Timeline:**

| Week 1 (Mar 29 - Apr 4) | Finalize questions, methods, and responsibilities; submit proposal |
| --- | --- |
| Week 2(Apr 5 - Apr 11 ) | EDA, build predictive models, perform model tuning, draft final report |
| Week 3(Apr 12- Apr 18) | EDA, Inference analysis, draft final report |
| Week 4 (Apr 19 - 21) | Final polish, individual contribution summary, submit report and code |

6. **Responsibilities:**

| Work Section | Jiahao Cheng | Kevin Fang | Vy Dang |
| --- | --- | --- | --- |
| **Project Proposal** | Work together | | |
| **EDA** | Code and Visualize | Code and Visualize | Interpretation |
| **Predictive Question** | Build, evaluate, and tune the Gradient Boosting Model | Build, evaluate, and tune the AdaBoost Model | Build, evaluate, and tune the Random Forest Model |
| | Compare three models' performance | | |
| **Inference Analysis** | Code and Visualize | Code and Visualize | Interpretation |
| **Final Report** | Each person writes the section corresponding to the work they did | | |