**Public Health-Related Preliminary Project Proposal**

1. Prediction oftheoverall utilization rate of mental health services at the state level for the upcoming year from Google Trends.

***Basics***

* Who is Team Lead this week? **Collins Njagi**
* Who is Recorder? **Jessica Reidy**
* Who is Spokesperson? **Rospide Eddy**
* **A preliminary project title**: Forecasting Mental Health Service Utilization: A Google Trends-Based Approach for State-Level Prediction

***Background & Question***

* A defined research question that serves a need or fills a niche
  + **What is the question?**

Can Google Trends data be used to accurately Predict the overall utilization rate of mental health services at the state level for the upcoming year?

* + **What need does it fill?**

1. Enhanced public health planning: Accurate predictions can inform policy decisions, program development, and intervention strategies.
2. Improved mental health resource allocation by the state: By accurately predicting mental health service utilization rates, a state can allocate resources more effectively.
   * **Why is it worth your time/effort to explore this question?**
3. Google Trends offers a creative approach: Google Trends provides real-time data, thus timely and informed insights into mental health concerns.
4. Current methods limitations: traditional methods for predicting mental health services utilization, mostly rely on historical data.
5. Mental health is a growing concern in modern times; therefore, an accurate real-time prediction can help address this situation.  
     
     
     
   * **Is your question novel / original? Yes**

* Unique application of Google Trends: Google Trends application in predicting mental health at the state level is very new even though Google Trends has been used in various public health studies.
* Potential for improving prediction accuracy: By incorporating Google Trends, this study may identify new patterns that can enhance prediction accuracy, thus, resulting in a new approach to forecasting mental health service utilization.
* An identified **stakeholder**

1. State health departments: Accurate predictions of health mental service utilization will inform resource allocation and planning.
2. State policymakers and legislators: an accurate mental health service utilization forecast will inform policy and program developments.
3. Mental health service providers: Hospitals and private practices will be in a position to predict future demand for their services.

* A **hypothesis:** There is a significant positive correlation between Google Trends data related to mental health concerns and utilization rate of mental health services, this is because increased online searches of mental health information reflect growing public awareness and concern about mental health issues which in turn drives demand for mental health services.
* A **prediction:** Using Google Trend data, we predict that states with higher search volumes for mental health-related terms (e.g., depression, therapy) will experience a higher utilization rate of mental health services in the upcoming year as compared to the previous year.

***Data & Analysis***

* **What data set(s)** have you found that you think are a good match for your question? Why?

[Mental Health Service Utilization Google Trends - Explore](https://trends.google.com/trends/explore?date=all&geo=US-CT&q=mental%20health&hl=en)

[Mental Health Treatment Services Dataset](https://www.samhsa.gov/data/data-we-collect/mh-cld-mental-health-client-level-data/datafiles) This is a Public-use Files (PUF) released by the Substance Abuse and Mental Health Service Administration (SAMHSA)

* **What response / outcome variable will you use?**

Mental Health Service Utilization Rate: I intend to measure this as the percentage of individuals receiving mental health treatment among those who need it at the state level for the upcoming year.

* **What predictor variable(s) will you use**?

1. Demographic variables: Age, sex, ethnicity, and race will provide insights into how different demographic groups utilize mental health services.
2. Mental health diagnoses: Anxiety disorder reported, Depressive disorder reported, and bipolar disorder reported.
3. Substance use-related variables: Substance use problem, Alcohol, and substance use diagnosis might reveal insights into the impacts of substance use on mental health utilization.
4. Service utilization variables: Residential treatment centers, state psychiatric hospital services, and psychiatric inpatient will reveal the type of services utilized.

* **What is your tentative analysis plan**?

**Step 1: Data preparation:**

* Merge Google Trend data set and mental health treatment service
* Transform variables

**Step 2**: **Exploratory data analysis:**

* Summary statistics for each variable
* Correlation analysis
* Visualizations to show relationships between variables

**Step 3**: **Feature engineering:**

* Extract relevant features from Google Trend

**Step 4**: **Model development:**

* Linear regression
* Machine learning algorithms

**Step 5**: **Model evaluation:**

* Train and test split data
* Evaluate model performance
* Compare models

**Step 6**: **Interpretation and visualization:**

* Correlation Heatmaps & Time-Series Plots
* **Are there any pitfalls you can see with this plan? Anything that could topple your idea?**

1. Data quality issues especially from Google Trends, data might be biased based on terminologies (e.g., “depressed”, “mental health clinic near me” used.
2. Feature engineering: identifying relevant features from Google Trends data and creating effective interaction terms will require some level of expertise.

* **How will you know if your question is answered**?
* High predictive accuracy of my model in forecasting mental health service utilization rate using Google Trends data.
* Consistent patterns in relationships between Google Trends data and mental health service utilization rates across different states within the same period.
* **How will you know if your hypothesis is supported?**
  + Statistically significant coefficient (p-value > 0.05) for Google Trends data in regression analysis.
  + High correlation between Google Trends data and mental health service utilization (r) > 0.7.

***Technical Details***

* **What language do you plan to code in?**

R-programming

* **Are there any other resources you will need?**

Google Trends Data

* **What is the link to your GitHub repo?**

<https://github.com/JessRiedy/DSE6311OM_SP2025R2_Data-Science-Capstone>

**References**

Substance Abuse and Mental Health Services Administration. (2024). Mental Health Client-Level Data (MH-CLD) 2022: Public Use File (PUF) Codebook. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services