

Quantifying the Impact of the COVID-19 Pandemic on Mental Health Searches Using Euclidean Distance, Granger's Causality, and Regression

By: Anusha Dudella
McNeil High School

Mentors: Sara Del Valle and Selina Hulley
Los Alamos National Laboratory

Introduction

- **The COVID-19 pandemic increased mental health prevalence around the globe by around 25% (WHO, 2022)**
- **Many people do not seek help for mental health due to stigma**
- **Assessed the role of internet data in mental health prevalence in the US**

Hypotheses

- **COVID-19 increased anorexia, anxiety, and depression search interest in the US**

Hypotheses

- **COVID-19 increased anorexia, anxiety, and depression search interest in the US**
- **COVID-19 increased anorexia, anxiety, and depression search interest in Oregon more than in Ohio**

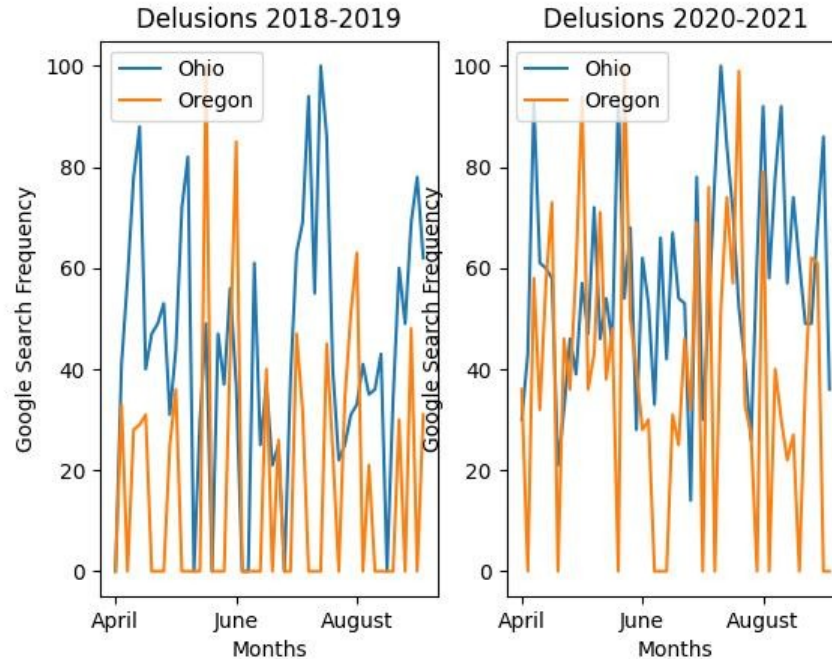
Hypotheses

- **COVID-19 increased anorexia, anxiety, and depression search interest in the US**
- **COVID-19 increased anorexia, anxiety, and depression search interest in Oregon more than in Ohio**
- **COVID-19 trends are predictive of ADHD search interests in the US**

Datasets

- **Used Google Trends for getting the frequencies of mental health searches**
- **PyTrends is the underlying API**
- **New York Times database for COVID-19 case counts**

Sparse Data



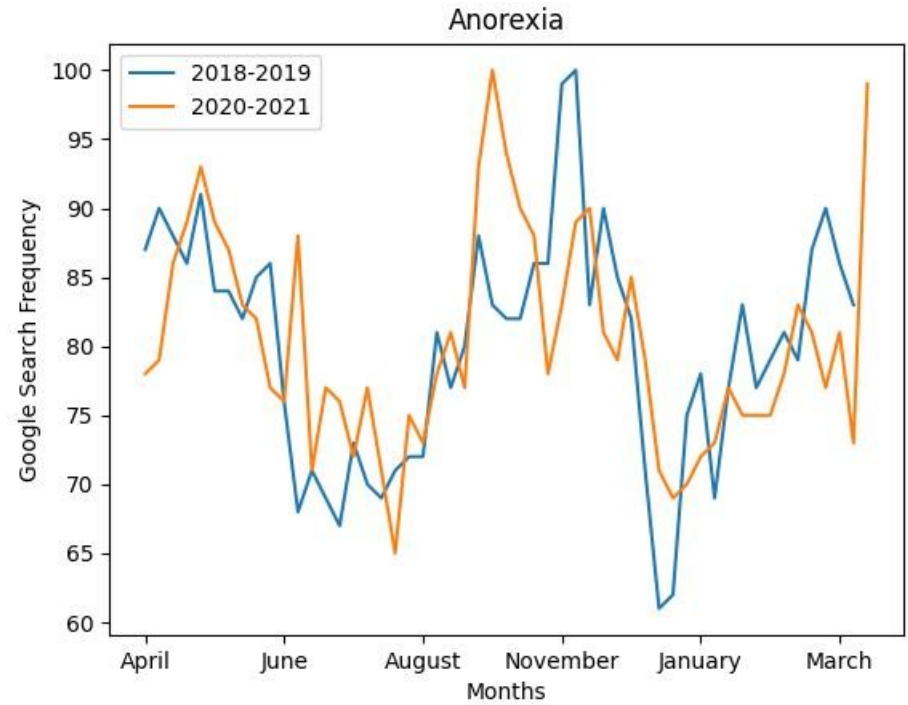
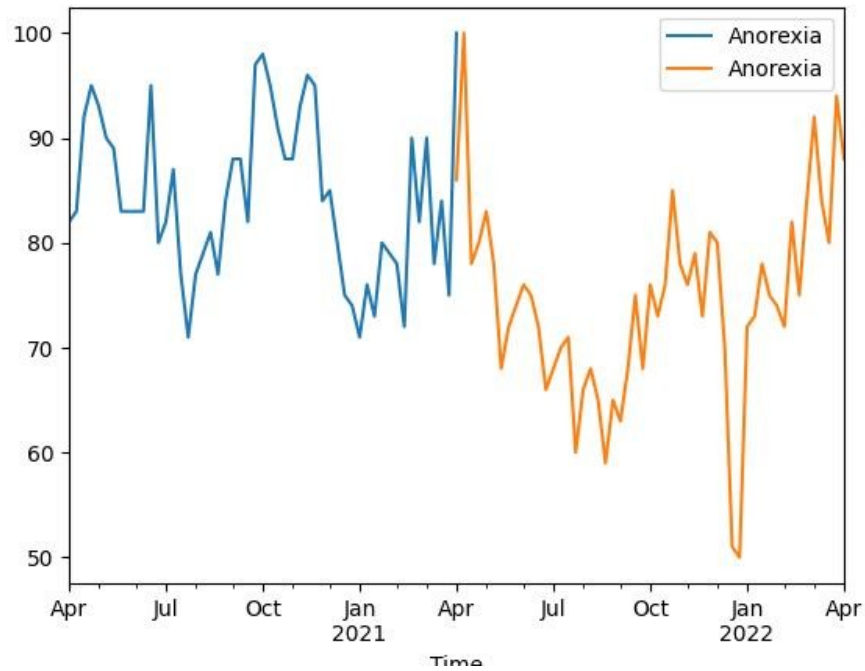
Clean data is pivotal!

Data Alignment and Normalization

- **Converted daily rolling averages for NY COVID case data to weekly rolling averages**
- **Used min-max scaling in order to normalize NY COVID case data to Google Trends frequencies**

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Good Graphs



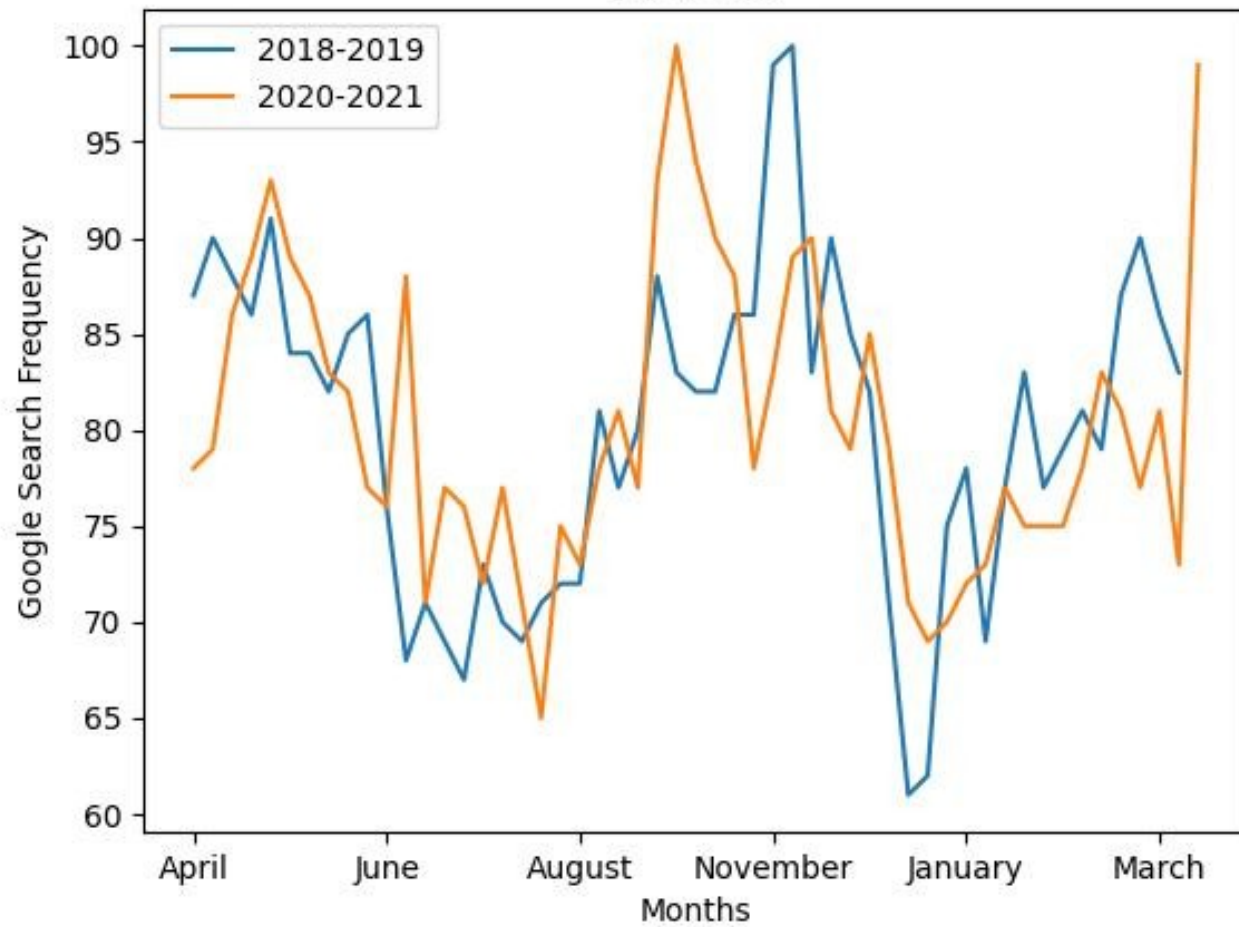
Hypothesis 1 Approach

- **COVID-19 increased anorexia, anxiety, and depression search interest in the US**
- **Data visualizing: compared frequencies of keywords for before (2018-2019) and during COVID (2020-2021)**

COVID-19 vs Anorexia

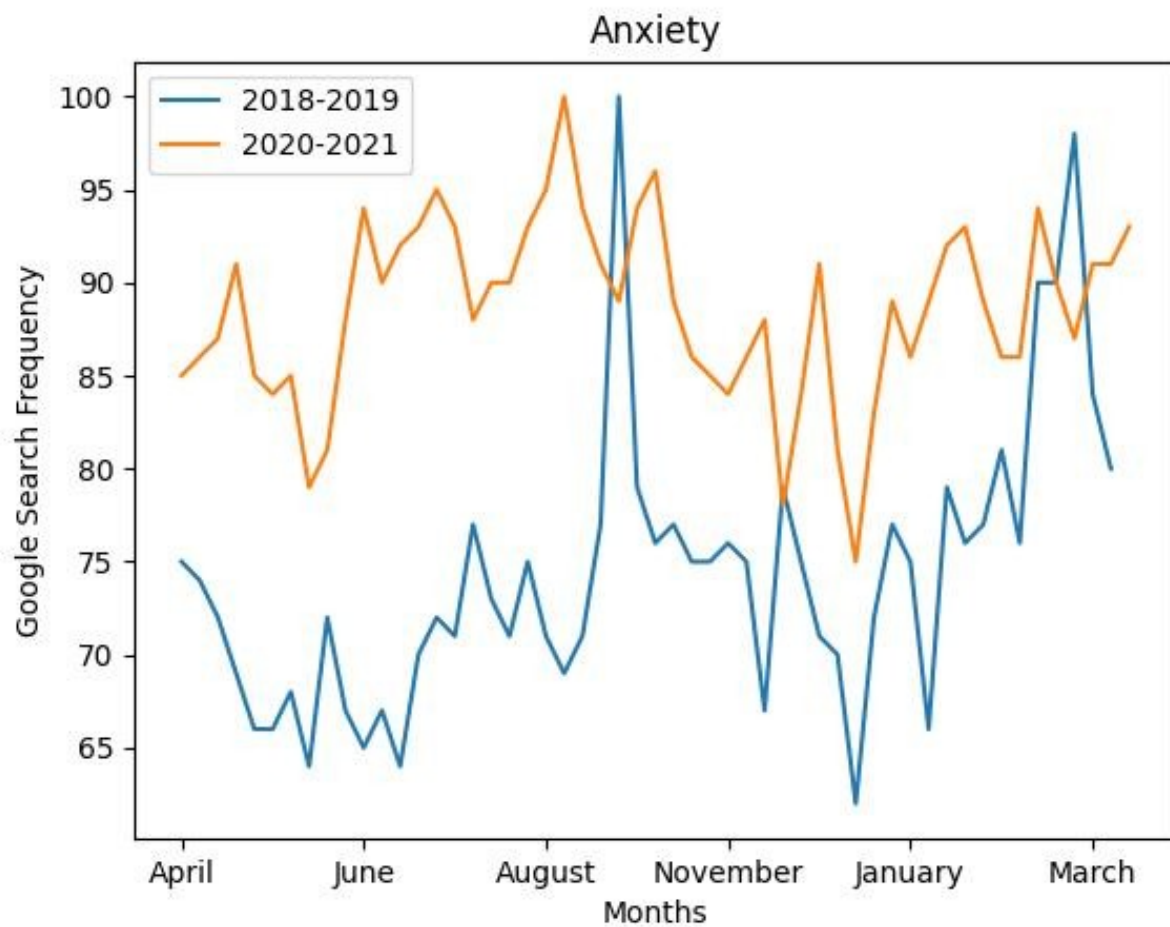
- **Cannot establish whether COVID-19 caused an increase in anorexia searches.**
- **Cyclic appearance in both pre-COVID and COVID years. Seasonal increase in the spring and fall.**

Anorexia



COVID-19 vs Anxiety

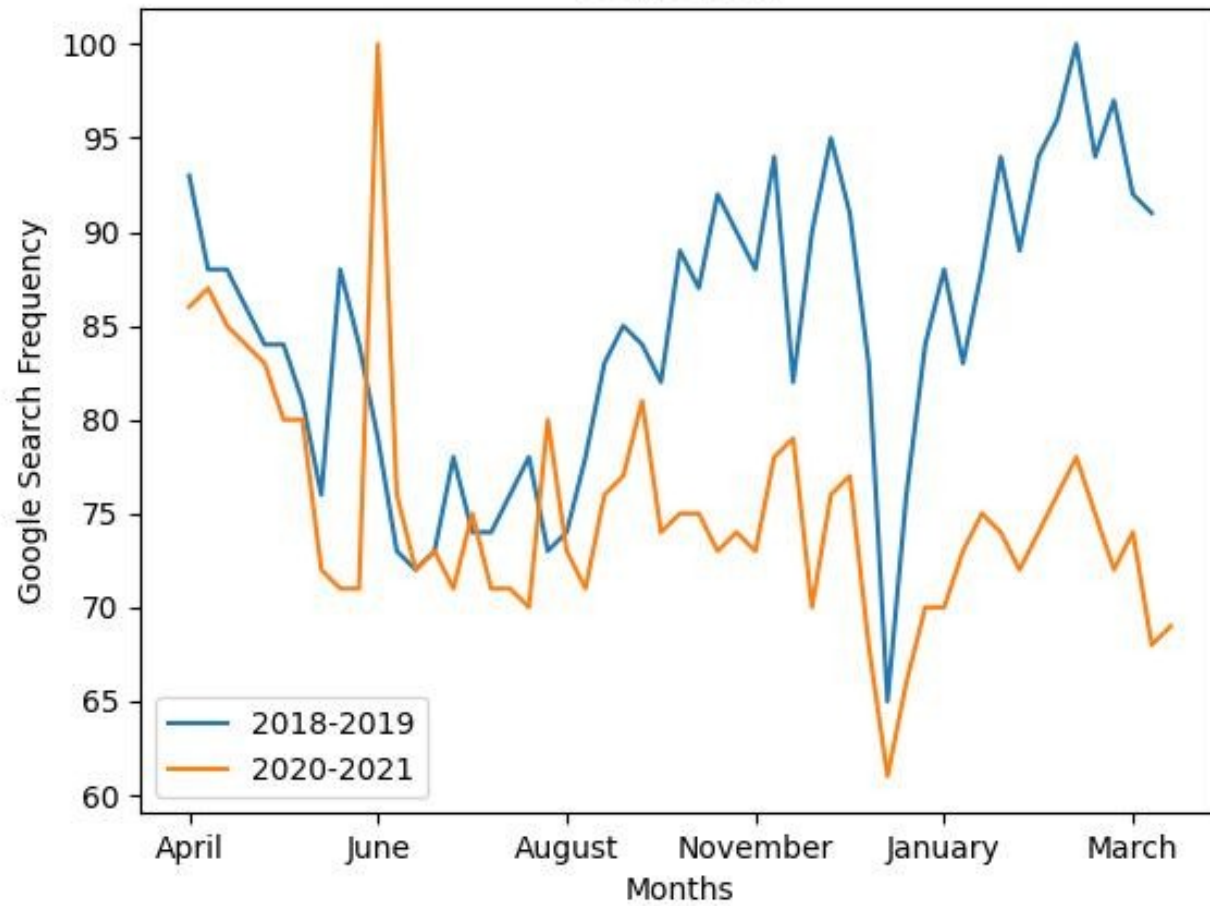
- **There was an increase in search frequencies for anxiety in 2020-2021 (COVID) than in 2018-2019.**



COVID-19 vs Depression

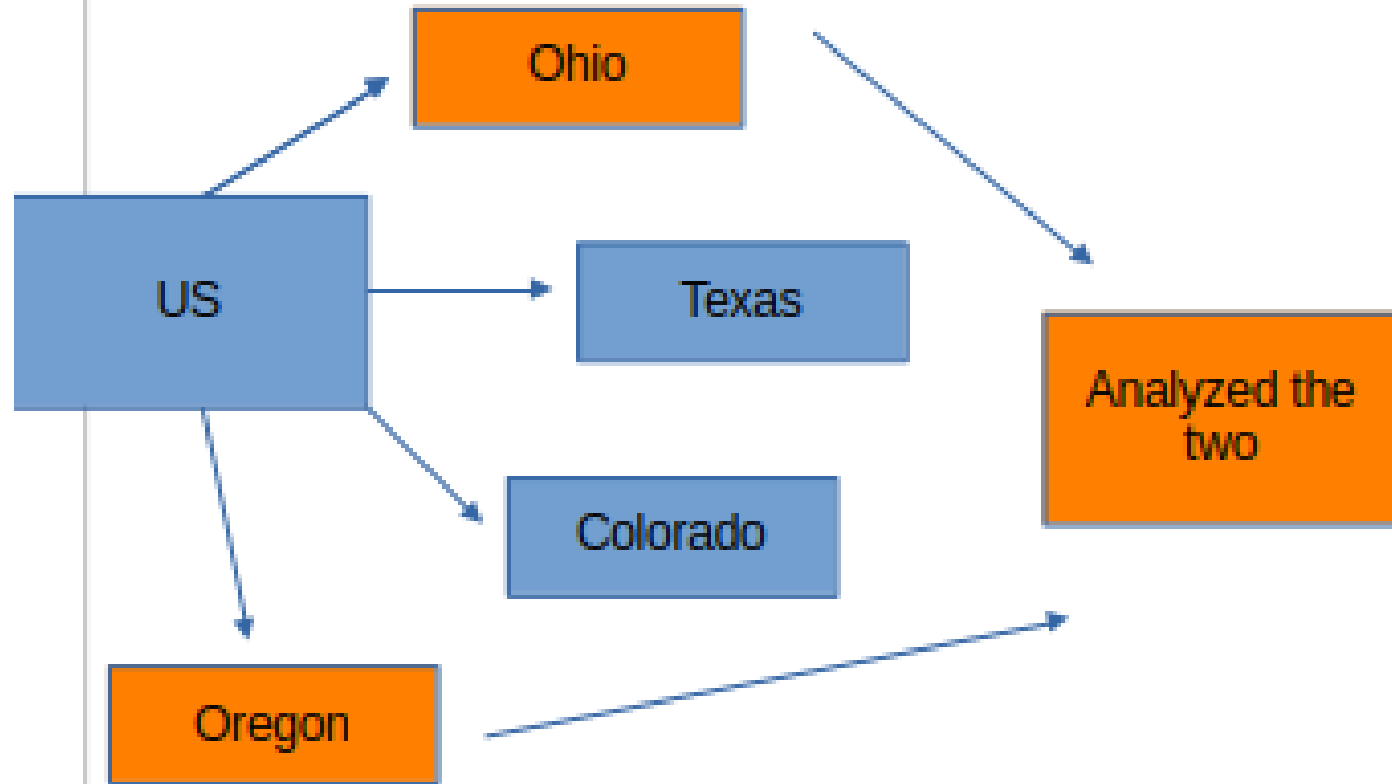
- **There was an decrease in search frequencies for depression in 2020-2021 (COVID) than in 2018-2019.**

Depression

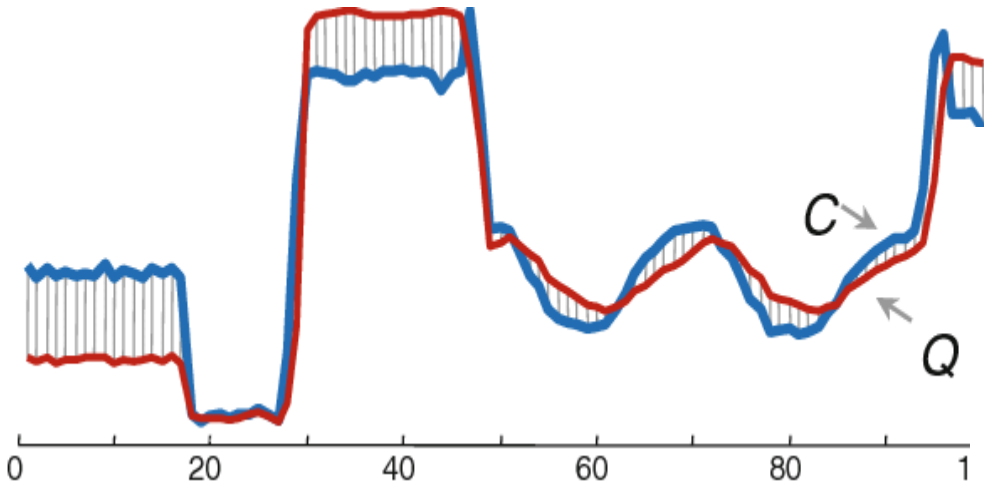


Hypothesis 2 Approach

- **COVID-19 increased anorexia, anxiety, and depression search interest in Oregon more than in Ohio**
- **Looked at regional data and then selected the two most differing regions using Euclidean distance**



Euclidean Distance



$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

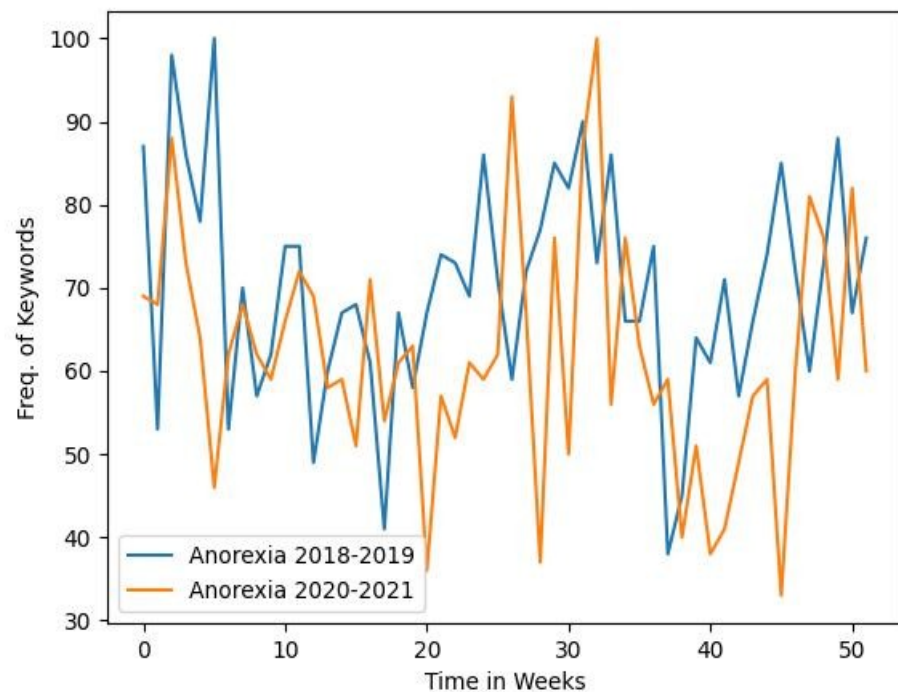
Euclidean Distance = square root of the sum of the vertical lines

Euclidean Distances

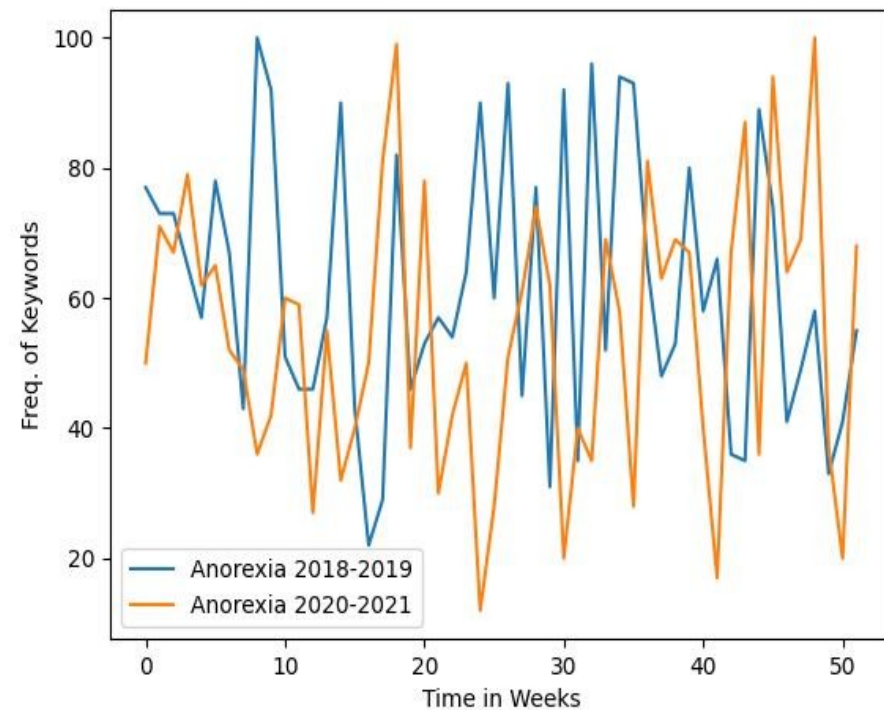
	Colorado	Texas	Oregon	Ohio
Anorexia	249.79	160.18	359.47	143.82
Depression	247.64	146.66	428.55	72.09
Anxiety	259.92	142.18	438.30	57.96

Distance of 438.30 indicates distance between 2018-2019 and 2020-2021 line plots.

Ohio



Oregon



Hypothesis 3 Approach

- **COVID-19 trends are predictive of ADHD search interests in the US**
- **Keywords, R squared, tested for causality, implemented regression model**
- **ADHD > 0.5**

Granger Causality

- **Test that proves whether one time series is useful in forecasting another**
- **Does not account for any confounding/external variables**

Granger Causality

- **H0: “Is not predictive” HA: “Is predictive”**
- **P-value: measures how confident we are about assumption given observed data**
- **P val \leq 0.05: reject H0**
- **P val $>$ 0.05: “fail to reject” H0**

Granger Causality Results

	ADHD_X	US_Cases_X
ADHD_Y	1.0	0.0
US_Cases_Y	0.0	1.0

Implementation of Linear Regression

- **Trained regression model on 70% of data**
- **Tested the model with 30% of data**

$$y = \beta_0 + \beta_1 X + \varepsilon$$

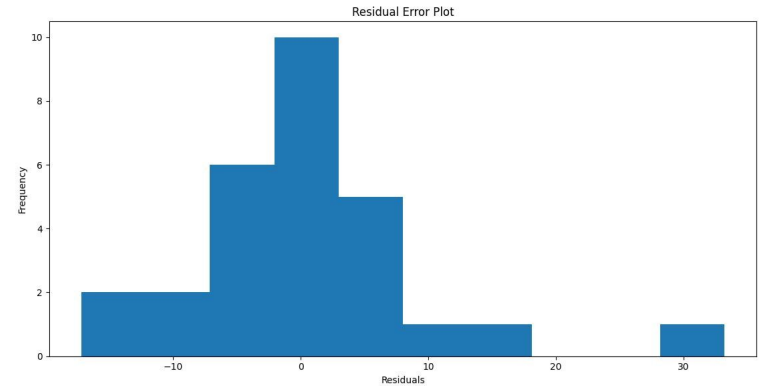
RMSE/Residual Error

RMSE (root mean squared)

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

Average error rate is around 4%

Residual Error Plot



$$\hat{e}_i = Y_i - \hat{Y}_i$$

To conclude..

- **Hypothesis 1: COVID-19 increased anorexia, anxiety, and depression search interest in the US**
- **Hypothesis 2: COVID-19 increased anorexia, anxiety, and depression search interest in Oregon more than in Ohio**
- **Hypothesis 3: COVID-19 trends are predictive of ADHD search interests in the US**

Key Takeaways

- **Clean data**
- **Using the right model for your dataset/hypothesis**
- **Making meaningful graphs**
- **Deleting useless files**



Questions?