Alex Litchfield

Data Analytics Term Project: Covid-19 Deaths vs Mental Health

AGENDA

01 Problem Area

02 Data Source/Application

03 The Analysis

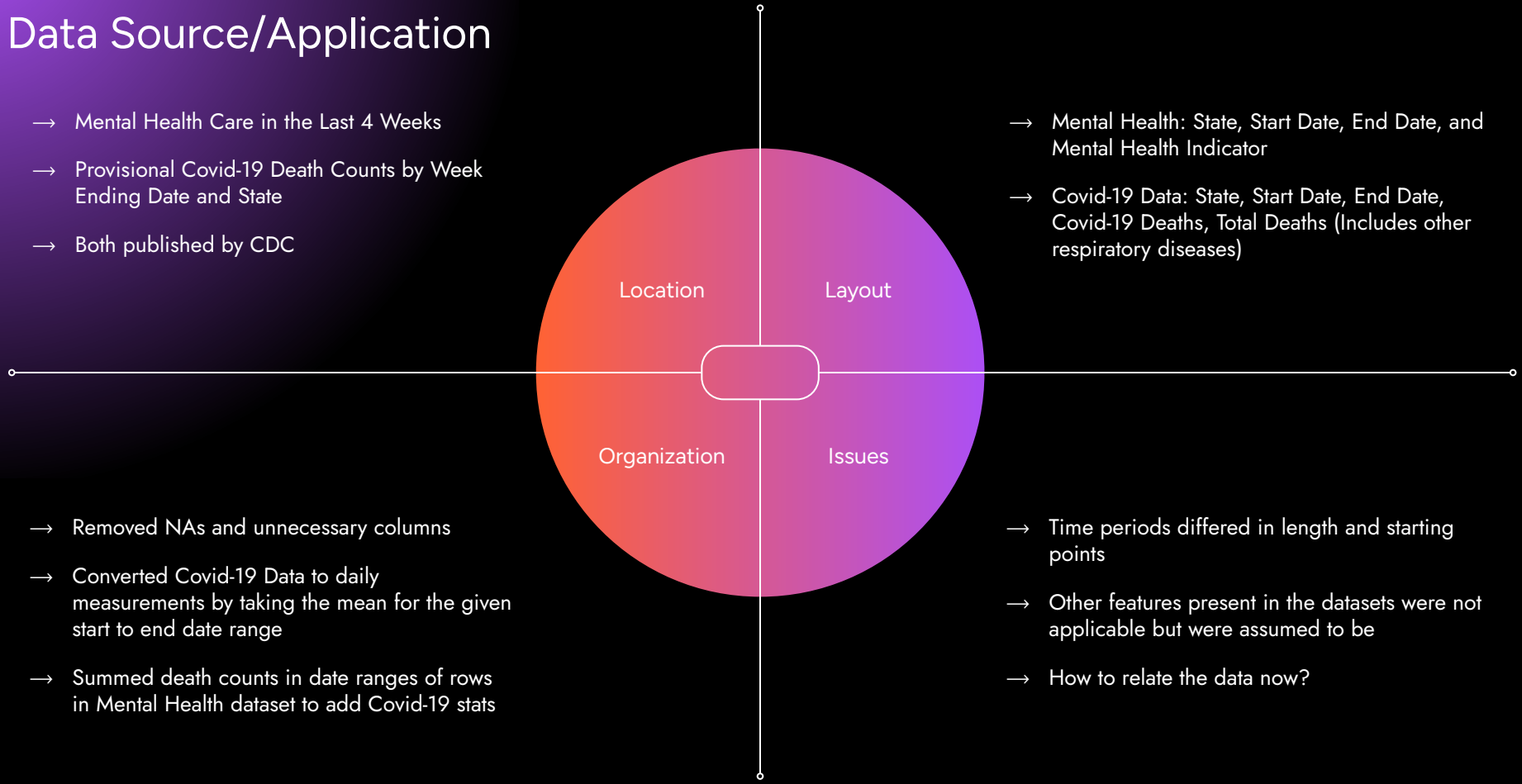
04 Outcome

Problem Area

U.S. states with higher Covid-19 mortality rates in a given week saw an increase mental health care.

Data Source/Application

- → Mental Health Care in the Last 4 Weeks
- → Provisional Covid-19 Death Counts by Week Ending Date and State
- → Both published by CDC



- → Mental Health: State, Start Date, End Date, and Mental Health Indicator
- → Covid-19 Data: State, Start Date, End Date, Covid-19 Deaths, Total Deaths (Includes other respiratory diseases)

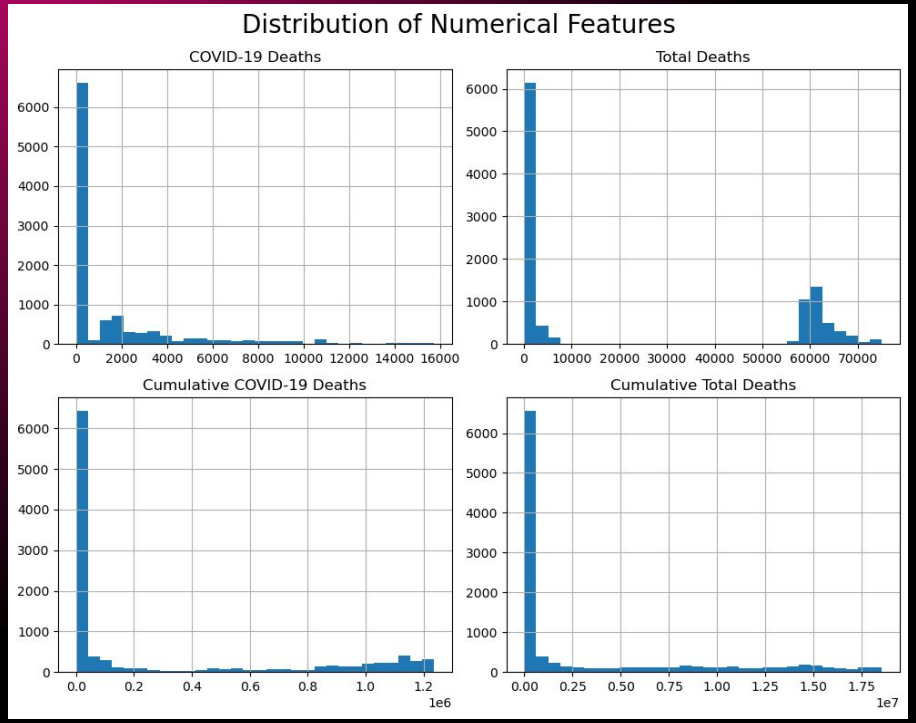
- → Removed NAs and unnecessary columns
- → Converted Covid-19 Data to daily measurements by taking the mean for the given start to end date range
- → Summed death counts in date ranges of rows in Mental Health dataset to add Covid-19 stats

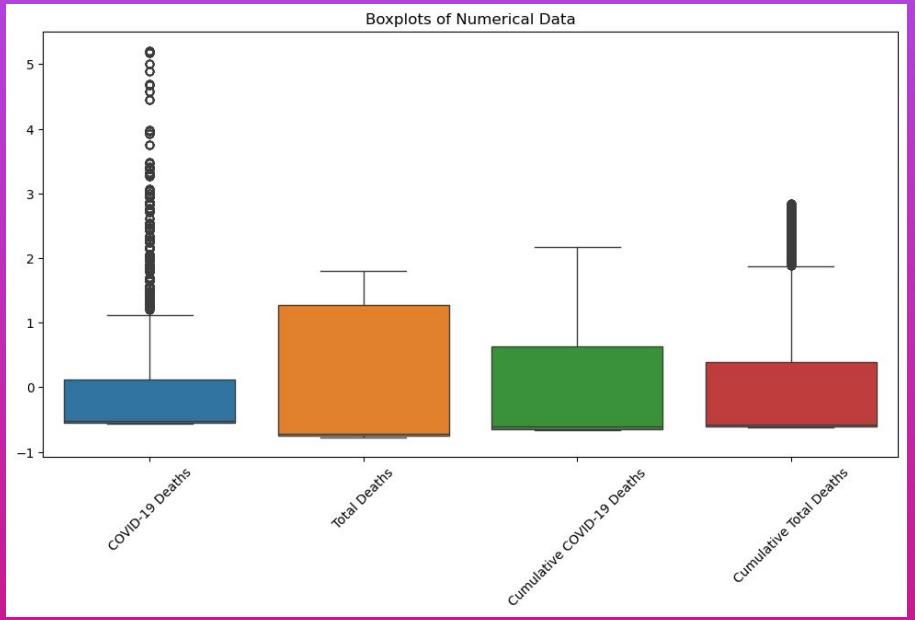
- → Time periods differed in length and starting points
- → Other features present in the datasets were not applicable but were assumed to be
- → How to relate the data now?

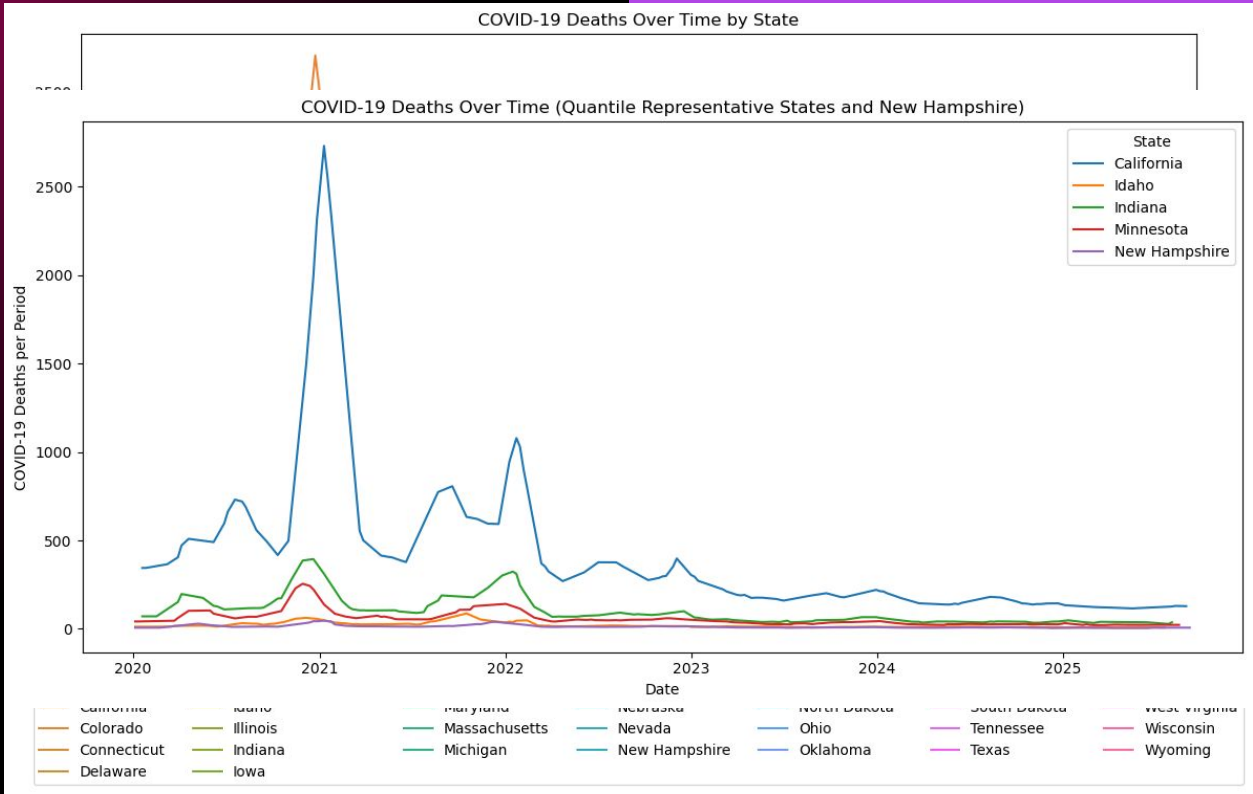
Analysis

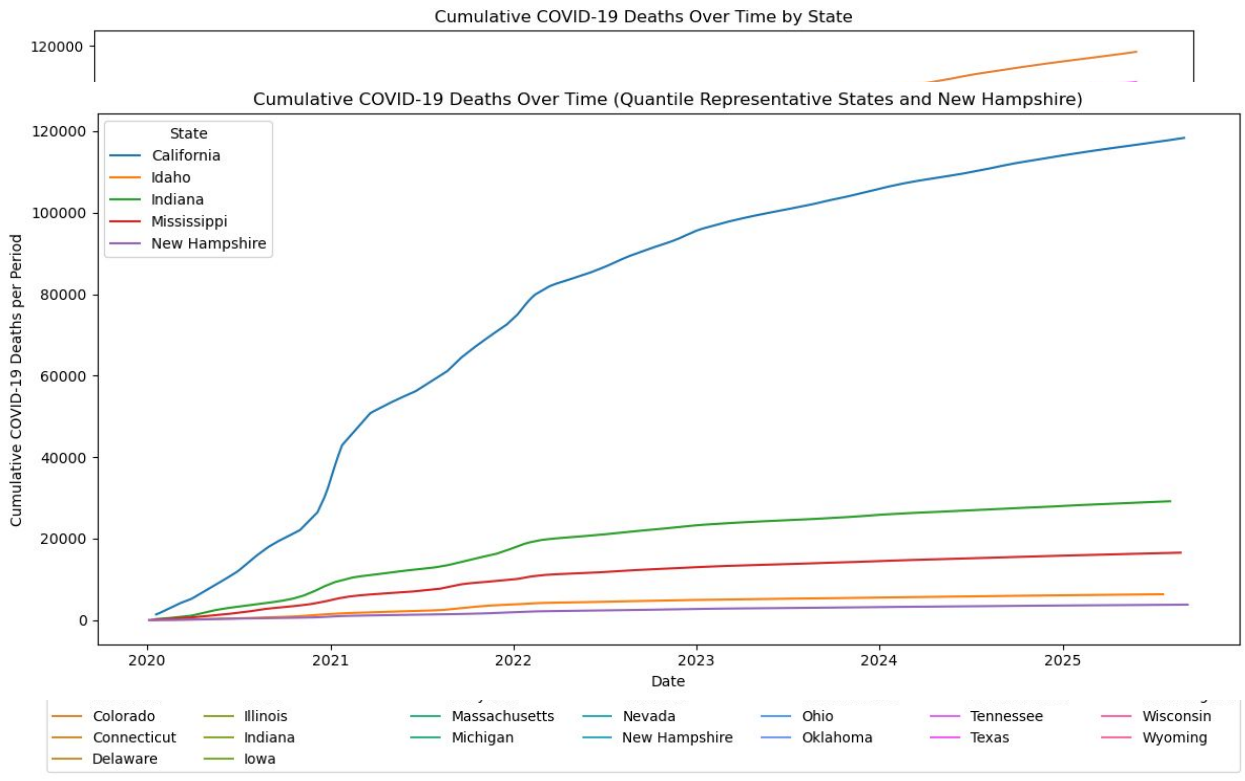
How many times does each state occur in the dataset? Why do we care?

State	
United States	3672
Alabama	132
Alaska	132
Arizona	132
Arkansas	132
California	132
Colorado	132
Connecticut	132
Delaware	132
District of Columbia	132
Florida	132
Georgia	132
Hawaii	132
Idaho	132
Illinois	132
Indiana	132
Iowa	132
Kansas	132
Kentucky	132
Louisiana	132
Maine	132
Maryland	132
Massachusetts	132
Michigan	132
Minnesota	132
Mississippi	132
Missouri	132
Montana	132
Nebraska	132
Nevada	132
New Hampshire	132
New Jersey	132
New Mexico	132
New York	132
North Carolina	132
North Dakota	132
Ohio	132
Oklahoma	132
Oregon	132
Pennsylvania	132
Rhode Island	132
South Carolina	132
South Dakota	132
Tennessee	132
Texas	132
Utah	132
Vermont	132
Virginia	132
Washington	132
West Virginia	132
Wisconsin	132
Wyoming	132


Distribution of Numerical Features
COVID-19 Deaths
Total Deaths
Cumulative COVID-19 Deaths
Cumulative Total Deaths
1e6
1e7


Boxplots of Numerical Data
COVID-19 Deaths
Total Deaths
Cumulative COVID-19 Deaths
Cumulative Total Deaths


COVID-19 Deaths Over Time by State
COVID-19 Deaths Over Time (Quantile Representative States and New Hampshire)
State
California
Idaho
Indiana
Minnesota
New Hampshire
COVID-19 Deaths per Period
2500
2000
1500
1000
500
0
2020
2021
2022
2023
2024
2025
Date
Colorado
Connecticut
Delaware
Illinois
Indiana
Iowa
Massachusetts
Michigan
Nevada
New Hampshire
Ohio
Oklahoma
Tennessee
Texas
Wisconsin
Wyoming


Cumulative COVID-19 Deaths Over Time by State
120000
Cumulative COVID-19 Deaths Over Time (Quantile Representative States and New Hampshire)
State
California
Idaho
Indiana
Mississippi
New Hampshire
120000
100000
80000
60000
40000
20000
0
Cumulative COVID-19 Deaths per Period
2020
2021
2022
2023
2024
2025
Date
Colorado
Connecticut
Delaware
Illinois
Indiana
Iowa
Massachusetts
Michigan
Nevada
New Hampshire
Ohio
Oklahoma
Tennessee
Texas
Wisconsin
Wyoming

Classifier Model Results (Including “U.S” as a State)

	Accuracy	Recall	Precision	F1 Score
Logistic Regression	28.2%	28.2%	28.2%	28.1%
Random Forest	43.7%	43.7%	43.7%	43.7%
KNN	43.7%	43.9%	43.7%	43.6%
SVM (Poly)	31.0%	38.3%	31.0%	28.7%

Classifier Model Results (Removing “U.S” as a State)

	Accuracy	Recall	Precision	F1 Score	
Logistic Regression	39.8%	39.7%	39.8%	39.4%	
Random Forest	43.2%	43.2%	43.2%	43.2%	
KNN	46.6%	46.8%	46.6%	46.4%	
SVM (Poly)	33.6%	36.9%	33.6%	31.9%	

-- Logistic Regression

	0	1	2	3
0	[illegible]	83	138	143
1	126	[illegible]	131	123
2	137	111	[illegible]	116
3	140	142	105	[illegible]

-- Logistic Regression (without U.S. as a State)

	0	1	2	3
0	[illegible]	42	65	57
1	52	[illegible]	105	61
2	90	67	[illegible]	37
3	104	89	42	[illegible]

-- Random Forest Classifier

	0	1	2	3
0	[illegible]	67	118	106
1	73	[illegible]	90	132
2	94	131	[illegible]	64
3	127	100	69	[illegible]

-- Random Forest Classifier (without U.S. as a State)

	0	1	2	3
0	[illegible]	30	78	82
1	41	[illegible]	63	75
2	80	81	[illegible]	46
3	68	68	53	[illegible]

-- KNN Classifier

	0	1	2	3
0	[illegible]	81	108	77
1	92	[illegible]	83	107
2	115	151	[illegible]	55
3	129	113	61	[illegible]

-- KNN Classifier (without U.S. as a State)

	0	1	2	3
0	[illegible]	40	70	53
1	53	[illegible]	43	57
2	72	106	[illegible]	28
3	82	67	48	[illegible]

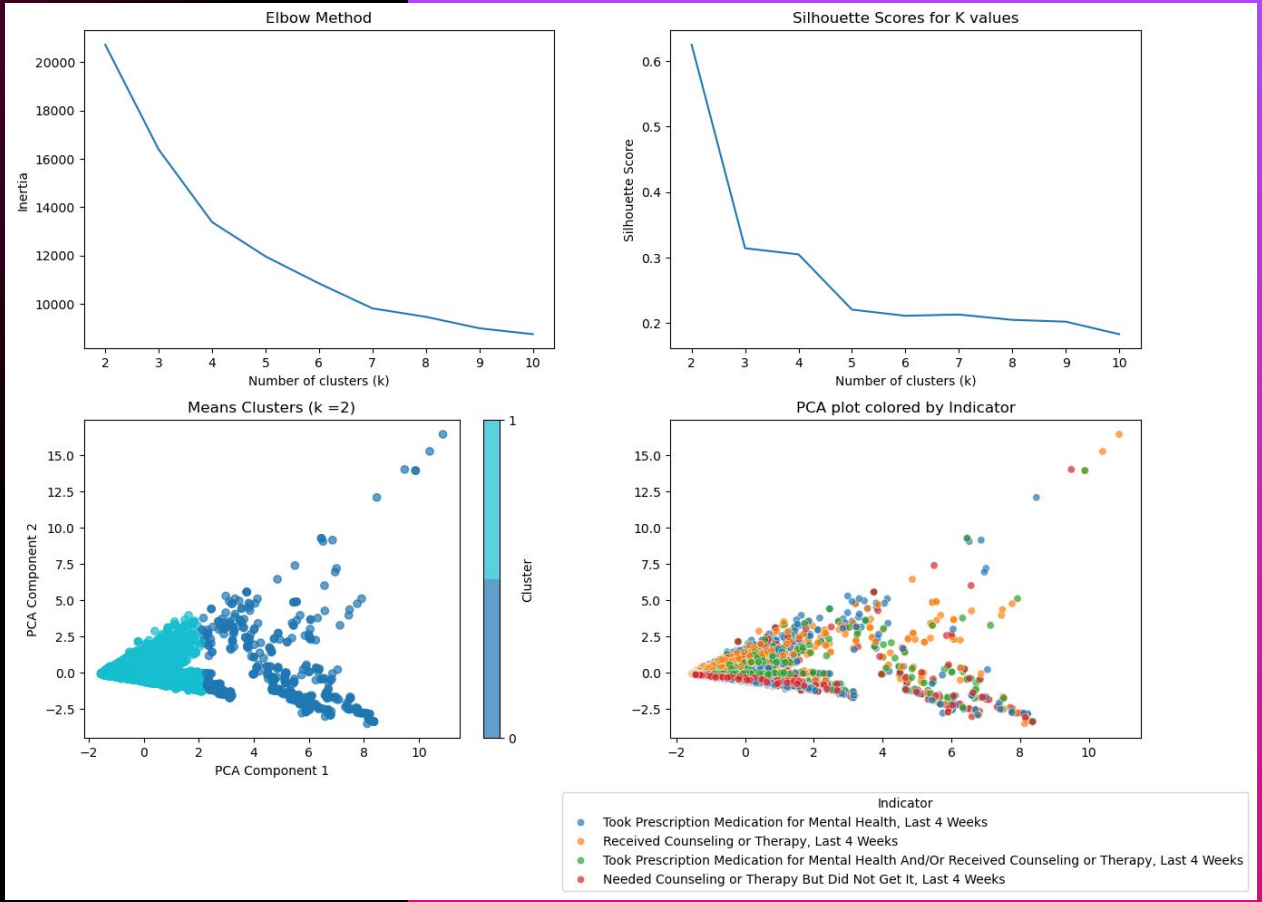
-- SVM

	0	1	2	3
0	[illegible]	315	29	67
1	45	[illegible]	18	109
2	36	330	[illegible]	72
3	87	312	15	[illegible]

-- SVM (without U.S. as a State)

	0	1	2	3
0	[illegible]	50	185	56
1	11	[illegible]	153	73
2	18	66	[illegible]	64
3	24	80	115	[illegible]

K-Means Clustering + PCA Visual Results



CONCLUSION

1. Data didn’t have enough features to train on

2. Replace datasets or add additional datasets in the future with relevant features

3. Hypothesis was proven wrong due to poor model performance and lack of useable features

THANK YOU

Any questions?