horizontal line

## Ironhack

Whitney Alli

2023 Candidate for Data Analytics

Final Project Plan

Project Plan **U.S. Abortion Care**

Dates Updated\*\*

**BUSINESS OBJECTIVE Week 1**

Research April 15

Questions - 10 questions to shape data April 15

Create data table, and begin cleaning, create data map May 1

* Remove null values
* Review data types
* Drop columns
* Scale to numerical values
* Create Data Linkage

**ANALYSIS Week 1**

Create histogram and clusters - make sense of the visuals May 4

Create comparisons May 5

Create correlation table and heatmap May 6

**TRANSFORM DATA Week 2**

Standardization May 9

Normalizing May 11

Train model May12

Modeling - k cluster/ kmeans, analysis of each

**PRESENTATION Week 3**

Tableau - create visualizations May 15

Create Presentation in Rise360 May 16

***Presentation May 20***

# 

# 

# OVERALL ANALYSIS

## Question

## **Analyzing in a Post Roe v, Wade Era the Short and Long Term Implications of Abortion Bans in the United States**

On June 24, 2022, the Supreme Court issued a ruling in *Dobbs v. Jackson Women’s Health Organization* that overturned the constitutional right to abortion as well as the federal standards of abortion access, established by prior decisions in the cases *Roe v. Wade* and *Planned Parenthood v. Casey*. Prior to the *Dobbs* ruling, the federal standard was that abortions were permitted up to fetal viability. That federal standard has been eliminated, allowing states to set policies regarding the legality of abortions and establish limits. Access to and availability of abortions varies widely between states, with some states banning almost all abortions and some states protecting abortion access.

## Data Source

Sources were gathered to show how the opinion on abortion is trending within the United States since the landmark case of Roe V, Wade. Public perception being allowed to decide vital medical care. These decisions have immediate and long term implications both directly on women’s health, the economy, and overall GDP. The analytics is a meta analysis over decades worth of data on the sentiments of Americans

| **Data set** [**https://osf.io/td7mk**](https://osf.io/td7mk)  **Talking Points** [**https://www.kff.org/womens-health-policy/report/key-facts-on-abortion-in-the-united-states/**](https://www.kff.org/womens-health-policy/report/key-facts-on-abortion-in-the-united-states/) |
| --- |

# PROCESS & ANALYSIS

## Project Process

**Methodology: Analysis Unsupervised Learning**

1. Collecting Data - stated above
2. Preparing the Data: cleaning data in efforts to get numerical
3. Choosing a Model: Predictive and diagnostic modeling
4. Training the Model: train test
5. Evaluating the Model:
6. Parameter Tuning:
7. Making Predictions:

## Weekly Meetings

Check in Saturday mornings on weekly progress and then again on May 1st to determine the final project deadline. Tentatively May 19th.

# ONBOARDING TASKLIST

## Creation of Unsupervised Learning Model

* Machine Learning (4 types)
* GitHub/Make repository and pushes
* Collect Data : Meta analysis of Abortion since RoevWade made law till present
* Visualization of Data
* Training Model Selection
* Evaluating Model
* Predictions

# RESOURCES

## Data Analytical Process

* Collecting Data: Sourced from Gallup and Guttenbach. Data is a meta-analysis of abortion, pregnancies, and births from the induction of Roe v. Wade to the removal.
* Preparing the Data
* Choosing a Model: [https://www.analytics8.com/blog/what-are-the-four-types-of-analytics-and-how-do-you-use-them/#:~:text=Modern%20analytics%20send%20to%20 fall,diagnostic%2C%20 predictive%2C%20and%20 prescriptive](https://www.analytics8.com/blog/what-are-the-four-types-of-analytics-and-how-do-you-use-them/#:~:text=Modern%20analytics%20tend%20to%20fall,diagnostic%2C%20predictive%2C%20and%20prescriptive).
* Training the Model: Sklearn for a linear regression model that lends itself to predictive analysis
* Evaluating the Model
* Parameter Tuning:
* Making analysis and visualzation:

## CODE Restart

**%conda install -c conda-forge kneed**

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn.linear\_model import LinearRegression

from sklearn.preprocessing import StandardScaler

from sklearn.preprocessing import OneHotEncoder

from sklearn.metrics import mean\_absolute\_error, mean\_squared\_error, r2\_score

from kneed import DataGenerator, KneeLocator

from sklearn.datasets import make\_blobs

from sklearn.cluster import KMeans

from sklearn.metrics import silhouette\_score

from sklearn.preprocessing import StandardScaler

from collections import Counter

**data = pd.read\_csv('/Users/testuser/Downloads/NationalAndStatePregnancy\_PublicUse.csv')**

data.head()

**data.shape**

**#912 rows and 103 columns, initial set of data**

data.dtypes

**data = data.drop(['pregnancyratelt15', 'pregnancyrate1517' , 'pregnancyrate1819','pregnancyrate1519','pregnancyrate2024', 'pregnancyrate2529', 'pregnancyrate3034', 'pregnancyrate40plus'], axis=1)**

**#removal of pregnancyrate columns**

data.head()

**Data.shape**

data = data.drop(['pregnancyratetotal'], axis=1)

#removal of pregnancyrate columns, missed in initial removal

**data = data.drop(['abortionratelt15', 'abortionrate1517' , 'abortionrate1819','abortionrate1519','abortionrate2024', 'abortionrate2529', 'abortionrate3034', 'abortionrate40plus', 'abortionratetotal'], axis=1)**

**#removal of abortionrate columns**

data = data.drop(['birthratelt15', 'birthrate1517' , 'birthrate1819','birthrate1519','birthrate2024', 'birthrate2529', 'birthrate3034', 'birthrate3539', 'birthrate40plus', 'birthratetotal'], axis=1)

#removal of birthrate columns

**data = data.drop(['abortionratiolt15', 'abortionratio1517' , 'abortionratio1819','abortionratio1519','abortionratio2024', 'abortionratio2529', 'abortionratio3034', 'abortionratio3539', 'abortionratio40plus', 'abortionratiototal'], axis=1)**

**#removal of abortionratio columns**

data = data.drop(['pregnancyratelt20', 'pregnancyrate3539' , 'abortionratelt20','abortionrate3539','birthratelt20', 'abortionratelt20'], axis=1)

**data = data.drop(['abortionratiolt20'], axis=1)**

data = data.drop(['notes', 'versiondate'], axis=1)

#removal of informational columns <ALL COLUMNS NOT NEEDED HAVE BEEN DROPPED>

**print(data.describe())**

print(data.groupby('state').size())

**data.select\_dtypes([np.number]).hist(figsize=(100,100))**

**plt.show()**

numericals = data.select\_dtypes(np.number)

categoricals = data.select\_dtypes(object)

**x, y = DataGenerator.figure2()**

X\_lin = data.drop(['population1544'], axis =1)

y\_lin = data[['population1544']]

X\_lin\_train, X\_lin\_test, y\_lin\_train, y\_lin\_test, train\_test\_split (X\_lin, y\_lin, test\_size=0.3, random\_state=42 )

**print([round(i, 3) for i in x])**

**print([round(i, 3) for i in y])**

kneedle = KneeLocator(x, y, S=1.0, curve='concave', direction='increasing')

print(round(kneedle.knee, 3))

print(round(kneedle.elbow, 3))

**plt.style.use('ggplot')**

**kneedle.plot\_knee\_normalized()**