ANES Data

Nathan Butler, Renee Waite, and Emmet Young

Dataset Information

- American National Election Studies (ANES) 2022
- Provides data about voting and the public opinion
- Found on ANES website
- We chose to focus on how demographics in the dataset relate to each abortion opinion
- We used unsupervised learning techniques
    - Apriori
    - Neural Nets
    - Clustering (Main priority)
    - Some bar charts to show distribution

Data Cleaning

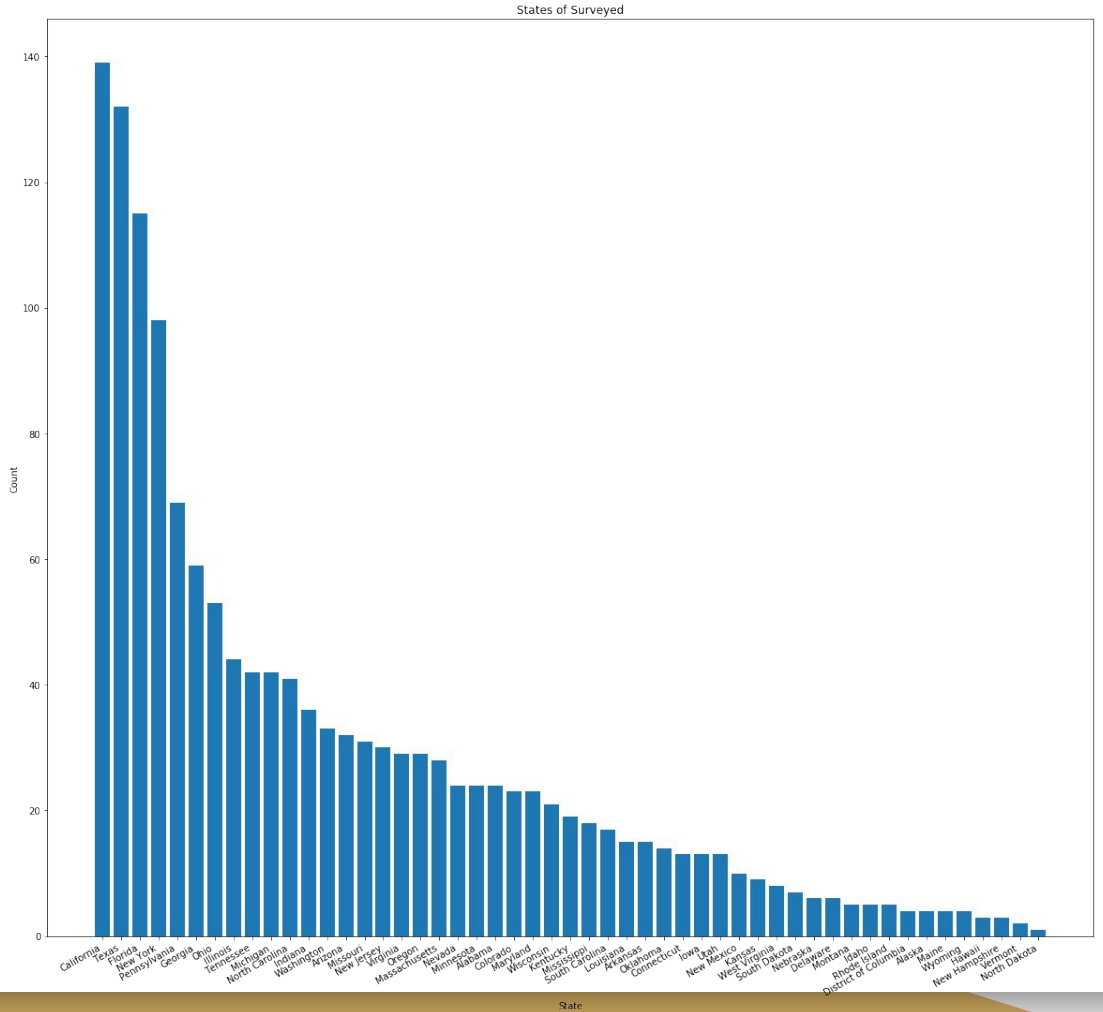
- Originally 1585 records and 577 columns
- Removed columns with more than 50% inapplicable
- Deleted the time data
  - Length of interview, questions, etc.
- Removed NAN columns
- Ended with 1585 records and 406 columns

Demographics Dataframe

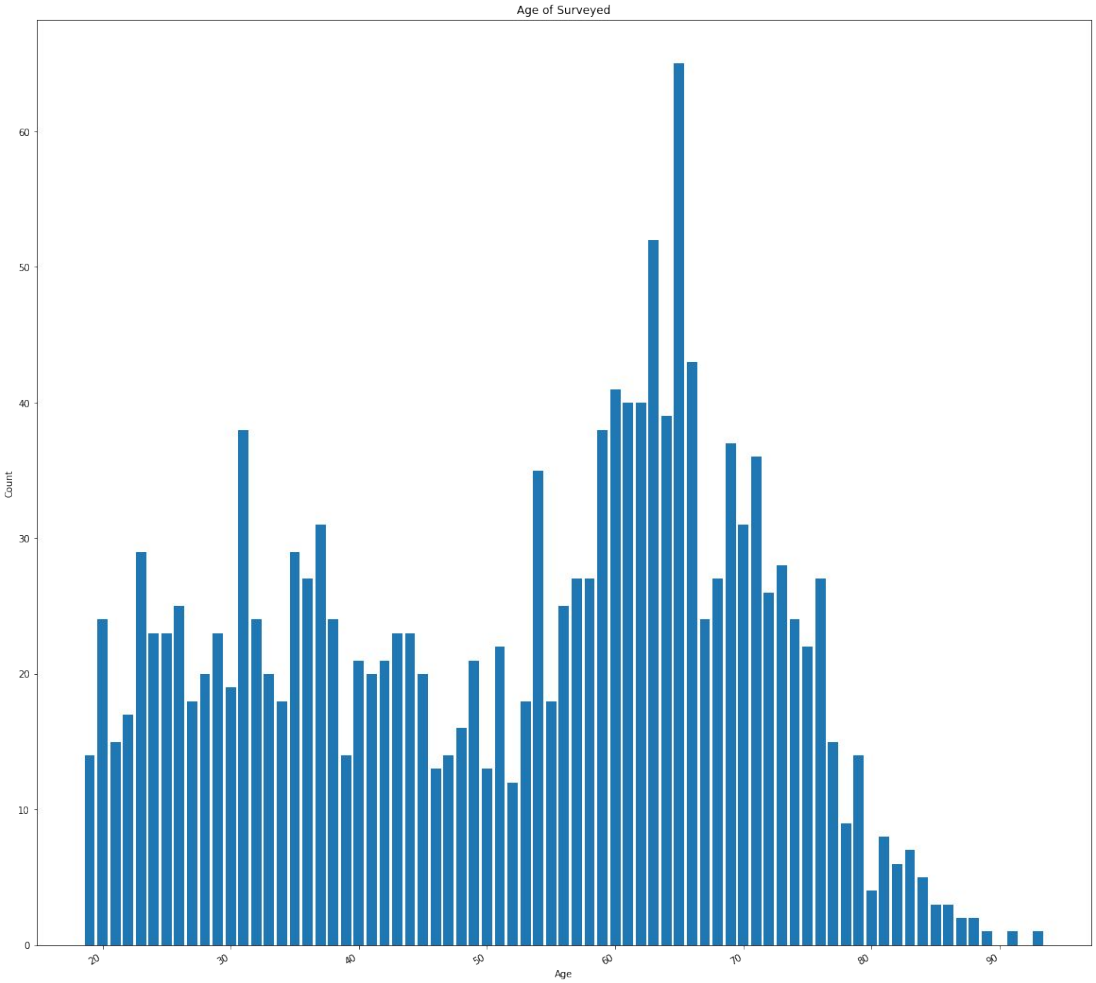
- Created a dataframe of some of the important demographics
  - Gender, age, state, party, past votes, religion, etc
  - 20 total columns
- Wanted to form clusters out of these demographics
- Data was in numeric coding
  - Uncoded each column using the ANES Userguide that has the questions and response options.
  - -1 were inapplicable answers and were named inap
- Focused on demographics with abortion opinions

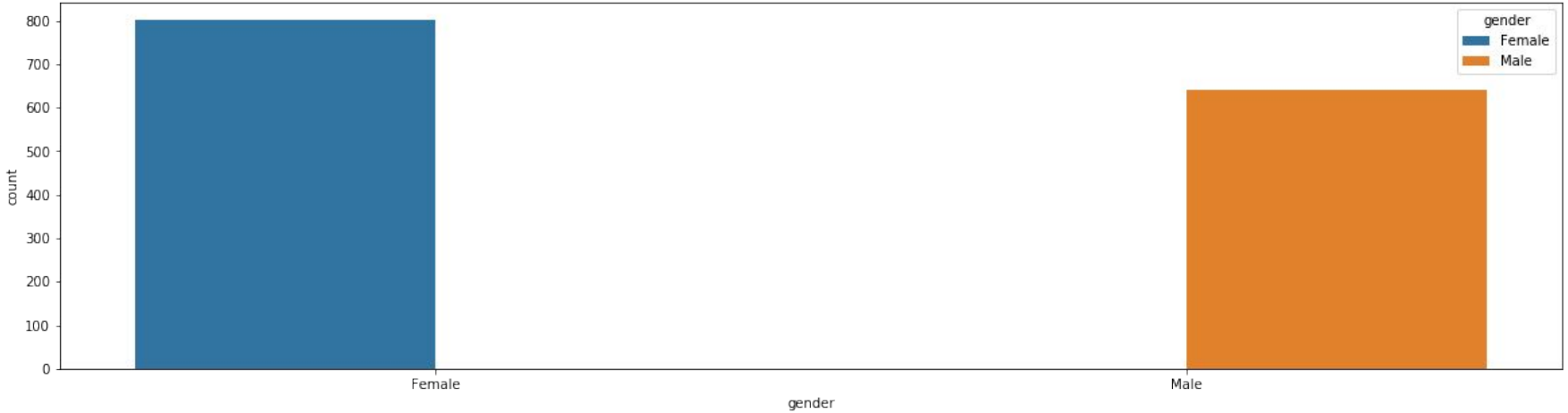
Barcharts (Distribution of Demographics)

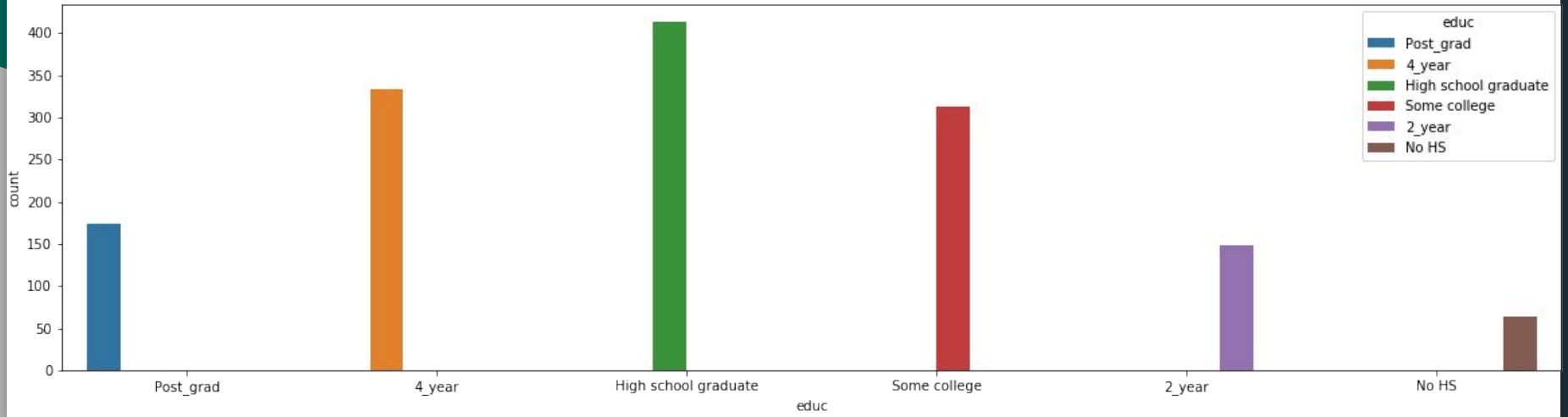
State

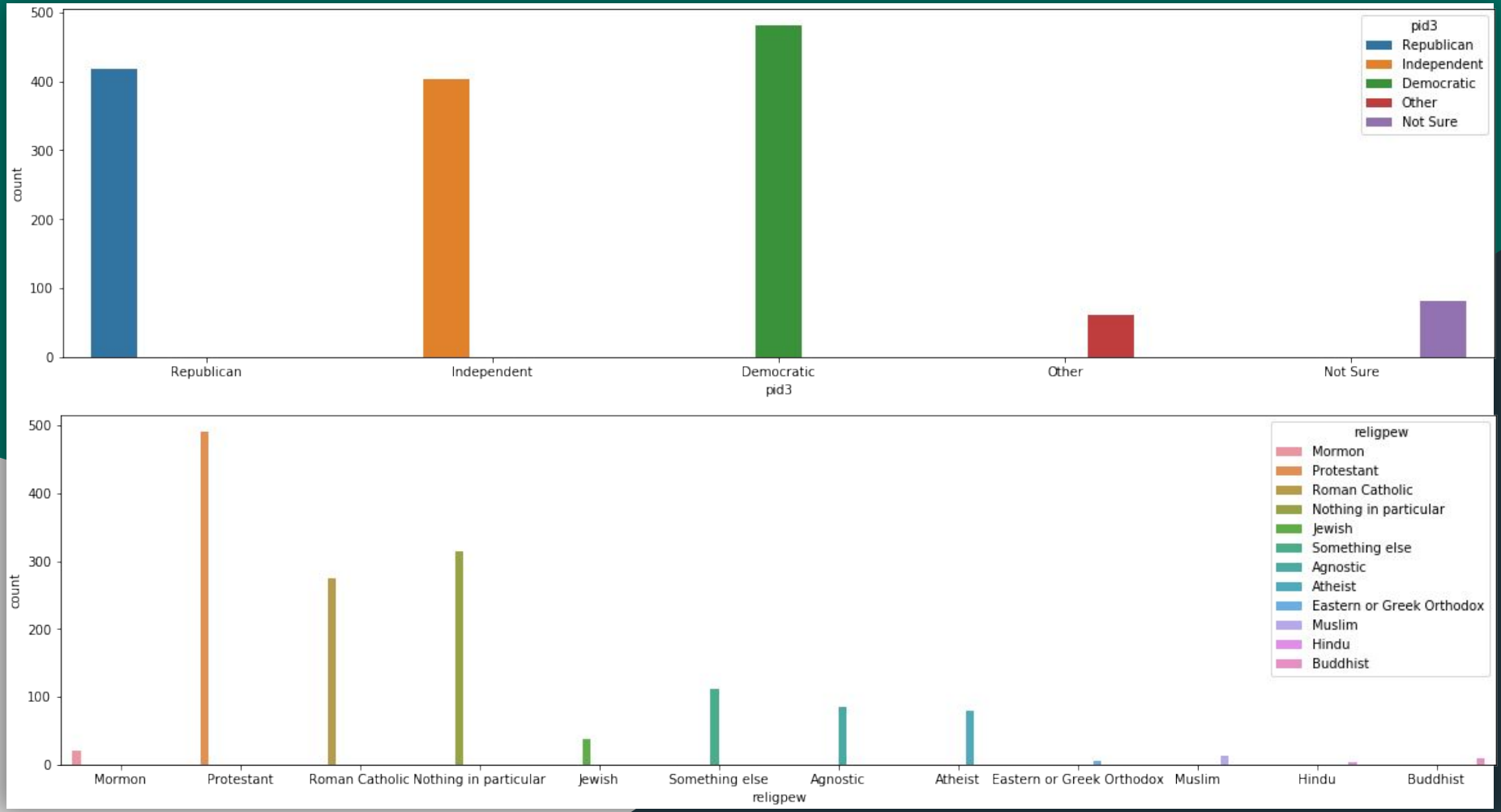

States of Surveyed
Count
State
0
20
40
60
80
100
120
140
California
Texas
Florida
New York
Pennsylvania
Georgia
Ohio
Illinois
Tennessee
Michigan
North Carolina
Indiana
Washington
Arizona
Missouri
New Jersey
Virginia
Oregon
Massachusetts
Nevada
Minnesota
Alabama
Colorado
Maryland
Wisconsin
Kentucky
Mississippi
South Carolina
Louisiana
Arkansas
Oklahoma
Connecticut
Iowa
Utah
New Mexico
Kansas
West Virginia
South Dakota
Nebraska
Delaware
Montana
Idaho
Rhode Island
District of Columbia
Alaska
Maine
Wyoming
Hawaii
New Hampshire
Vermont
North Dakota

Age


Age of Surveyed
Count
Age
0
10
20
30
40
50
60
20
30
40
50
60
70
80
90


gender
Female
Male
count
800
700
600
500
400
300
200
100
0
Female
Male
gender

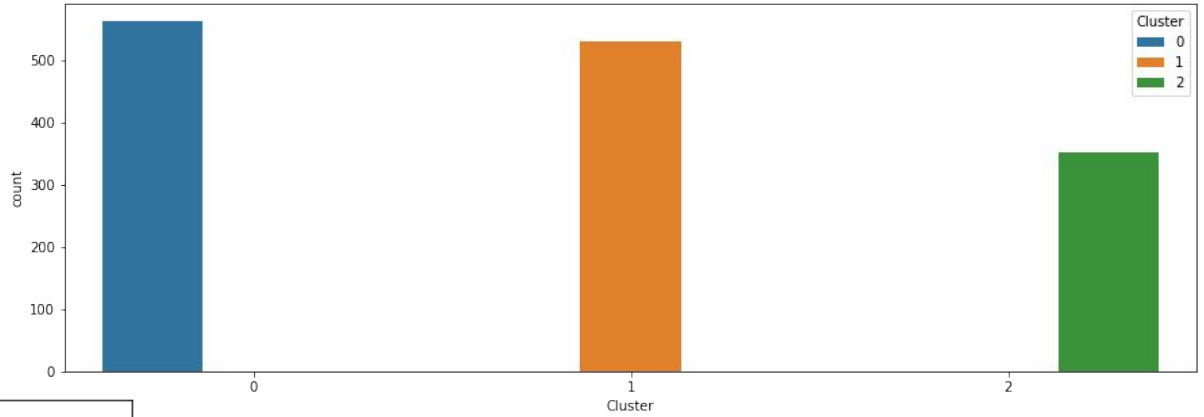

educ
Post_grad
4_year
High school graduate
Some college
2_year
No HS
count
400
350
300
250
200
150
100
50
0
Post_grad
4_year
High school graduate
Some college
2_year
No HS
educ

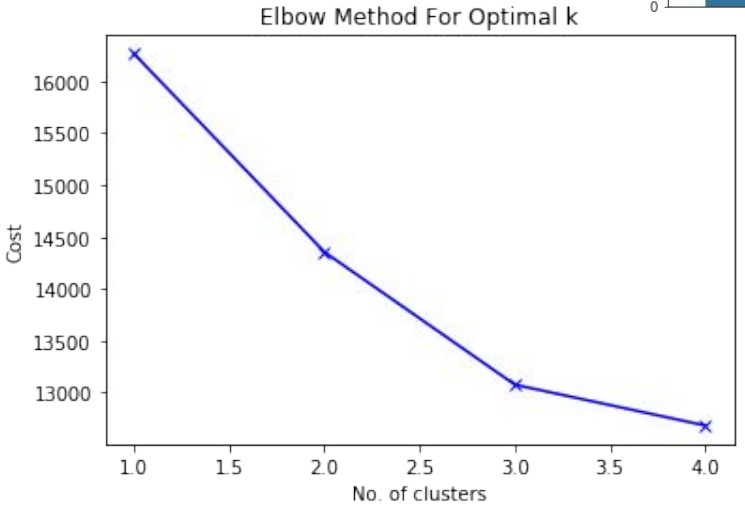

pid3
Republican
Independent
Democratic
Other
Not Sure
count
pid3
religpew
Mormon
Protestant
Roman Catholic
Nothing in particular
Jewish
Something else
Agnostic
Atheist
Eastern or Greek Orthodox
Muslim
Hindu
Buddhist
religpew

Demographic Clustering

Demographic Clusters

- Kmodes clustering
  - Categorical data

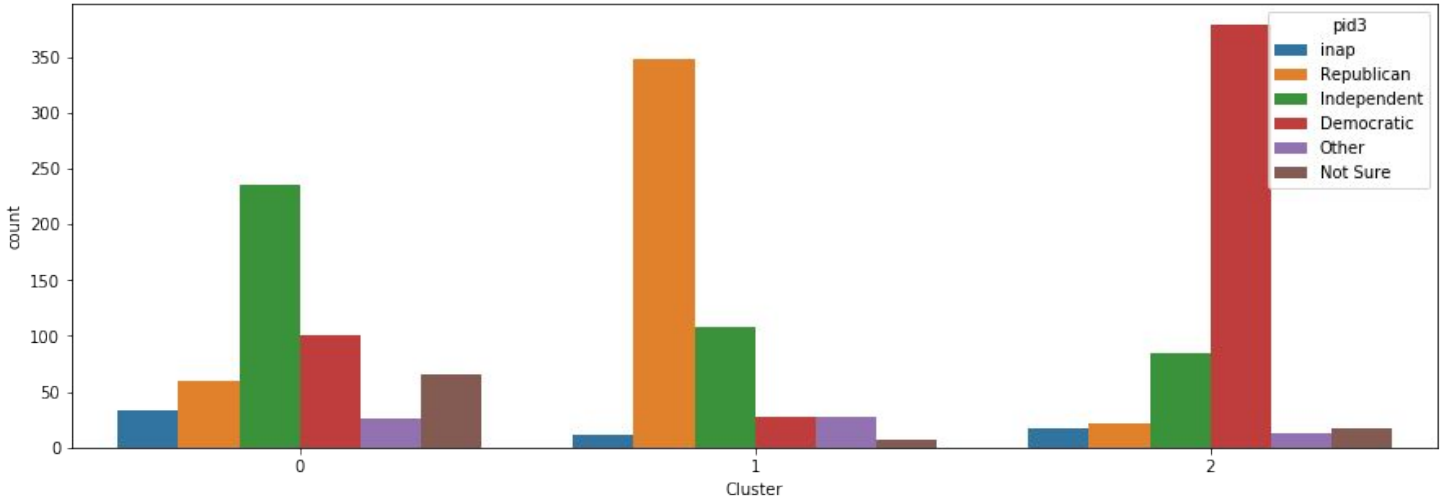




- Most optimal k is k = 3
  - Attempted to use more and delete columns

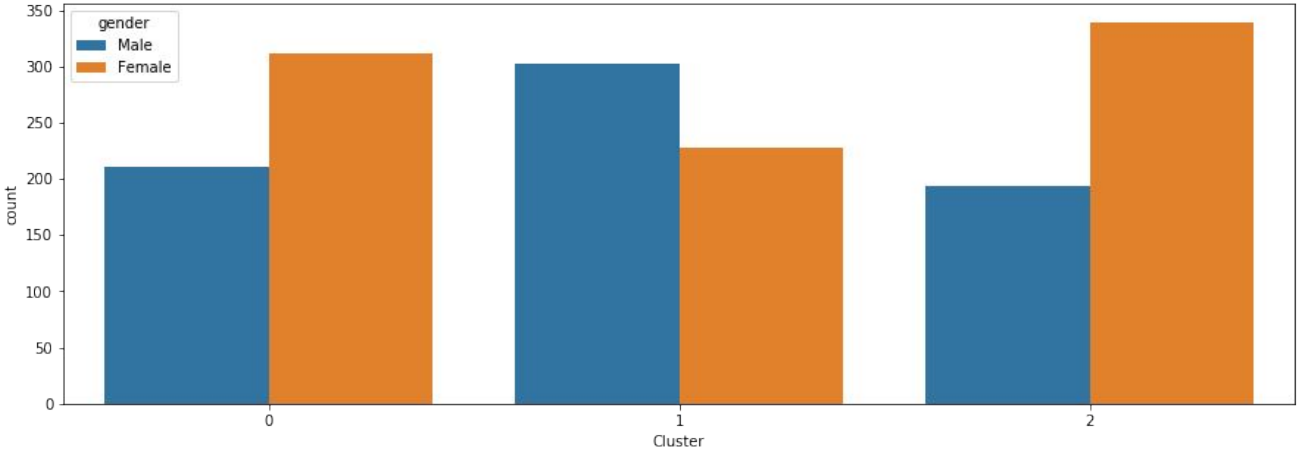
Party Affiliation

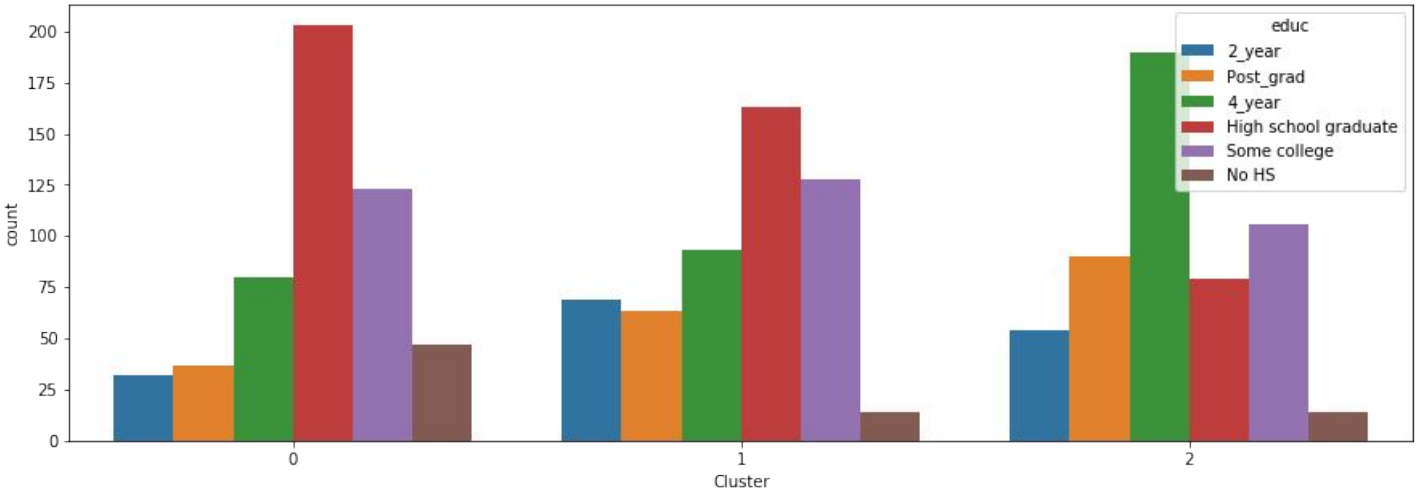
- Cluster 0 leans independent
- Cluster 1 is heavily republican
- Cluster 2 is majority Democratic



Gender

- Cluster 1 is the only male dominant cluster.

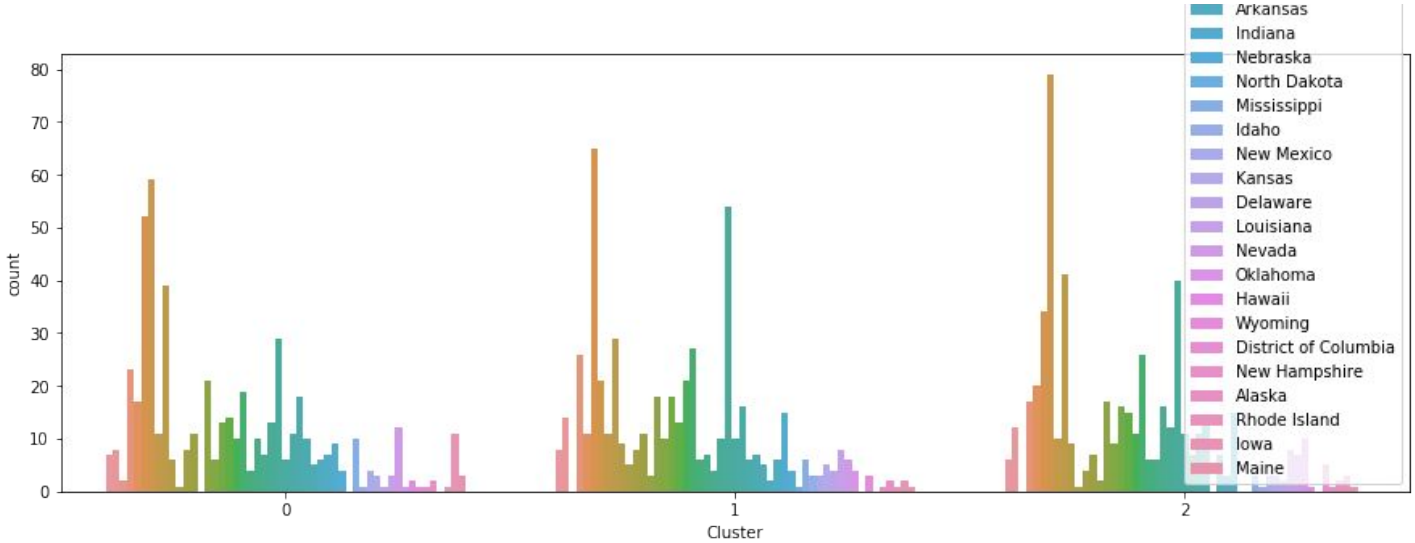




Education

- Cluster 0 seems less likely to attend a 4 year or post grad school.
- Cluster 1 seems the most balanced
- Cluster 2 has highest level of 4 year and post grad

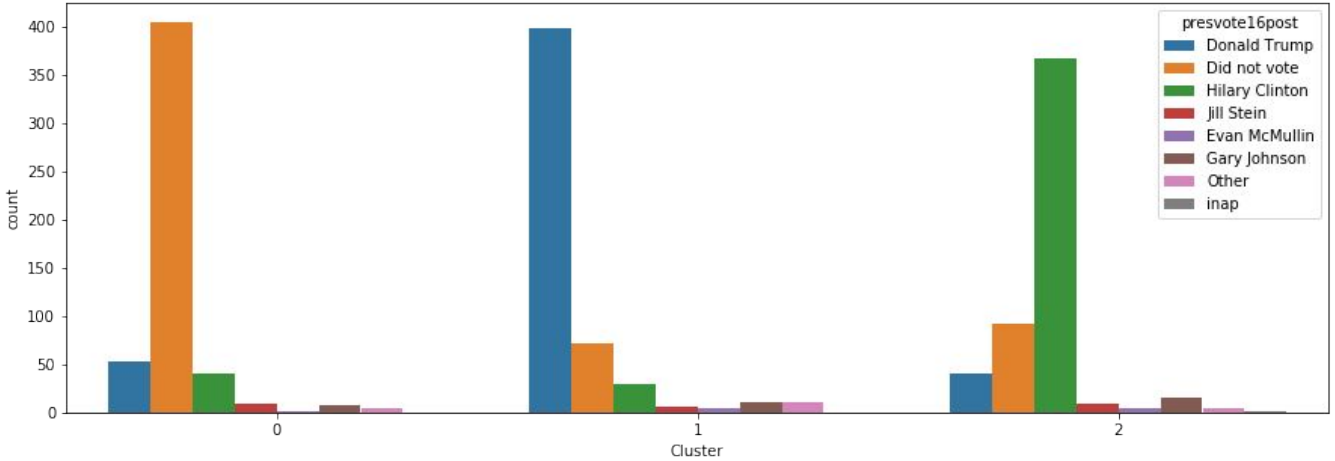
State Cluster

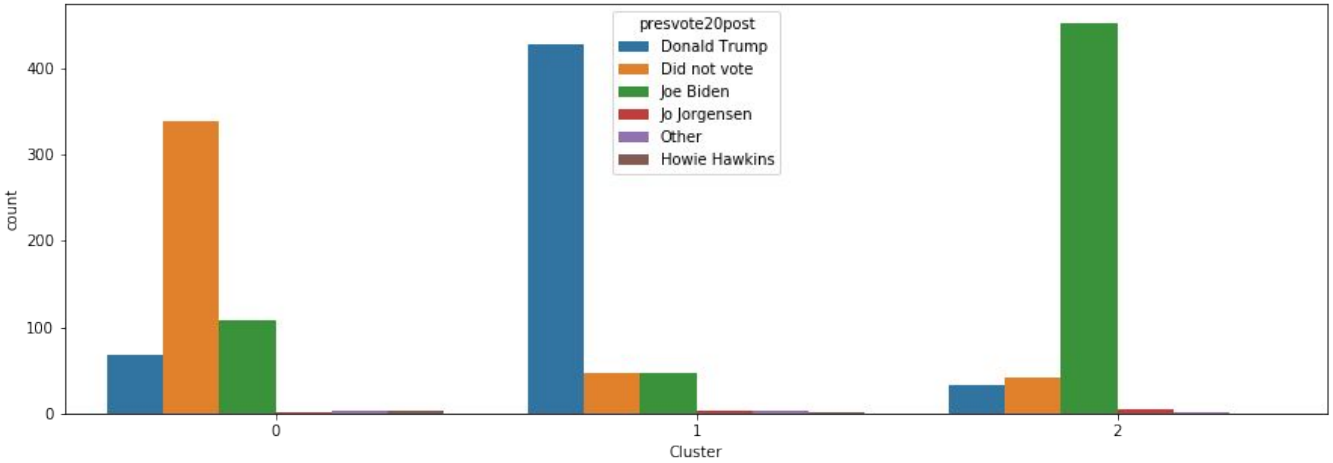


- Top 3 from 0
  California-41
  Texas-31
  New York - 24
- Top 3 from 1
  Florida-68
  Texas-52
  New York - 30
- Top 3 from 0
  California-79
  Texas-49
  New York - 44

2016 Election

- Cluster 0 did not vote
- Cluster 2 voted Donald Trump
- Cluster 3 Voted Clinton
- Lines up with party affiliation



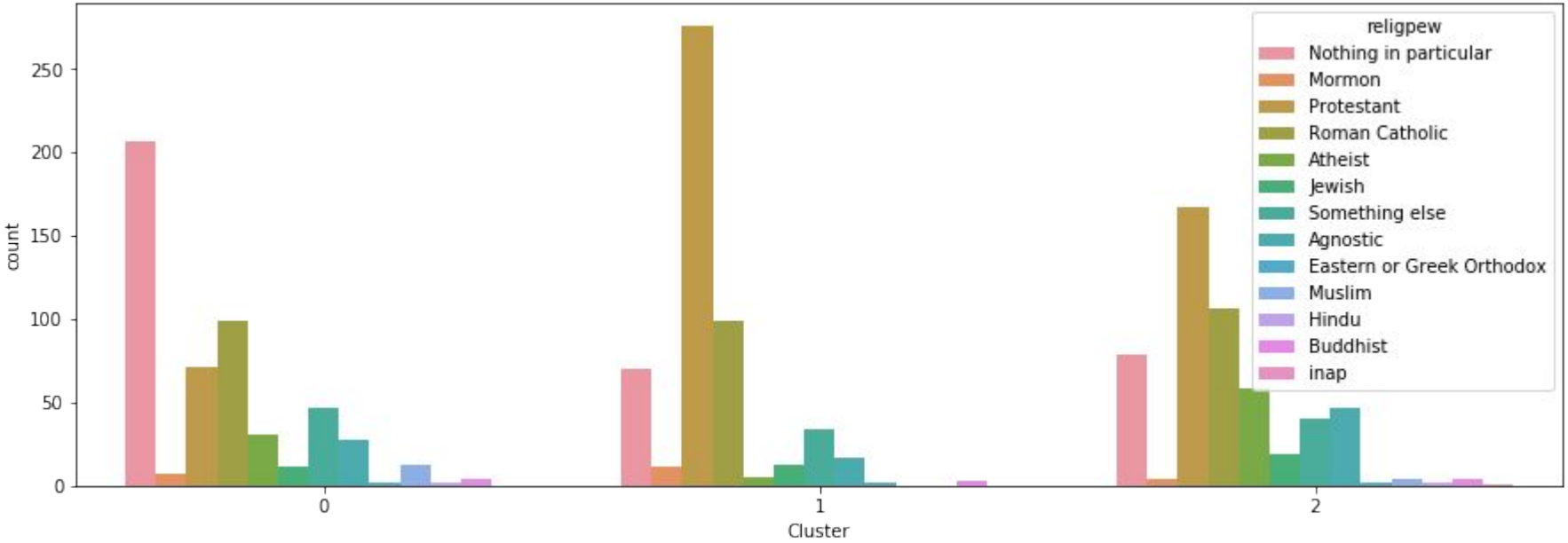


2020 Election

- Cluster 0 did not vote
- Cluster 2 voted Donald Trump
- Cluster 3 Voted Joe Biden
- Lines up with party affiliation
- Voters tend to vote same party.

Religion

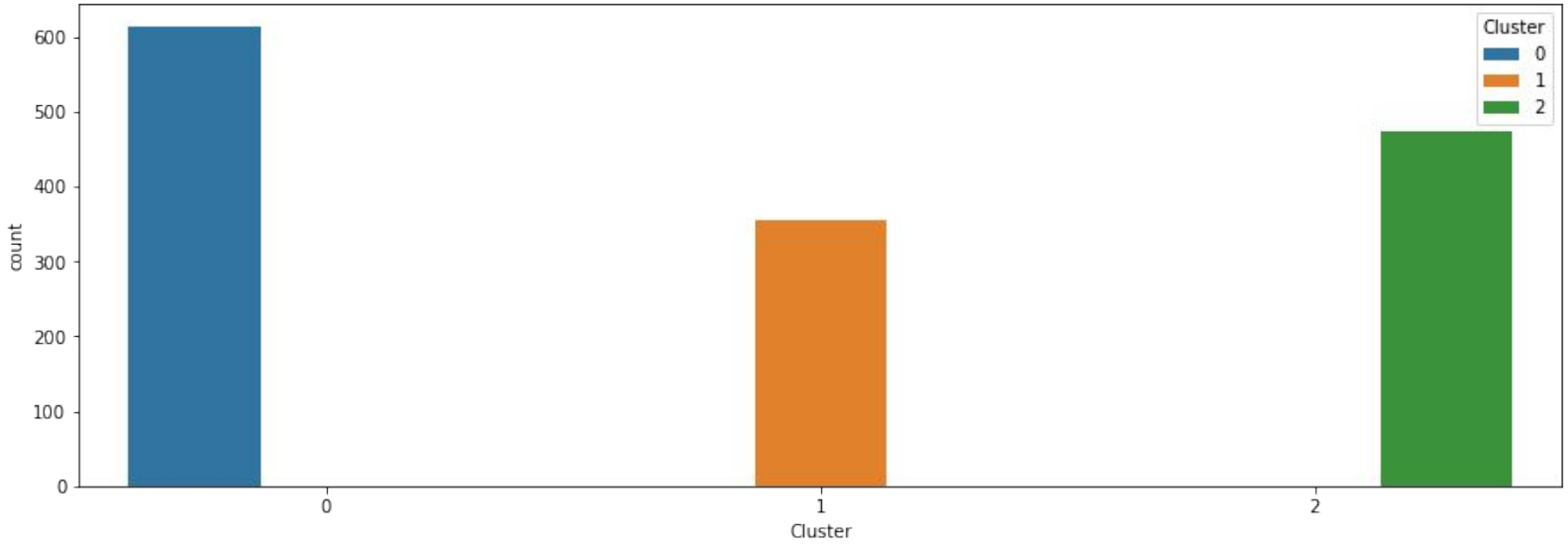
- Cluster 0 was mainly non-religious
- Cluster 1 was mainly Protestant
- Cluster 2 was most balanced
- Nearly 50% of the US is protestant



Demographics with Abortion Data

- We merged our demographic data with a four opinionated abortion questions.
  - Which party deals with abortion best?
  - When is abortion okay when concieved by rape, unwanted, or the women's life is at risk?
- Used kmodes clustering and bar charts to show data.
  - Using elbow method k = 3

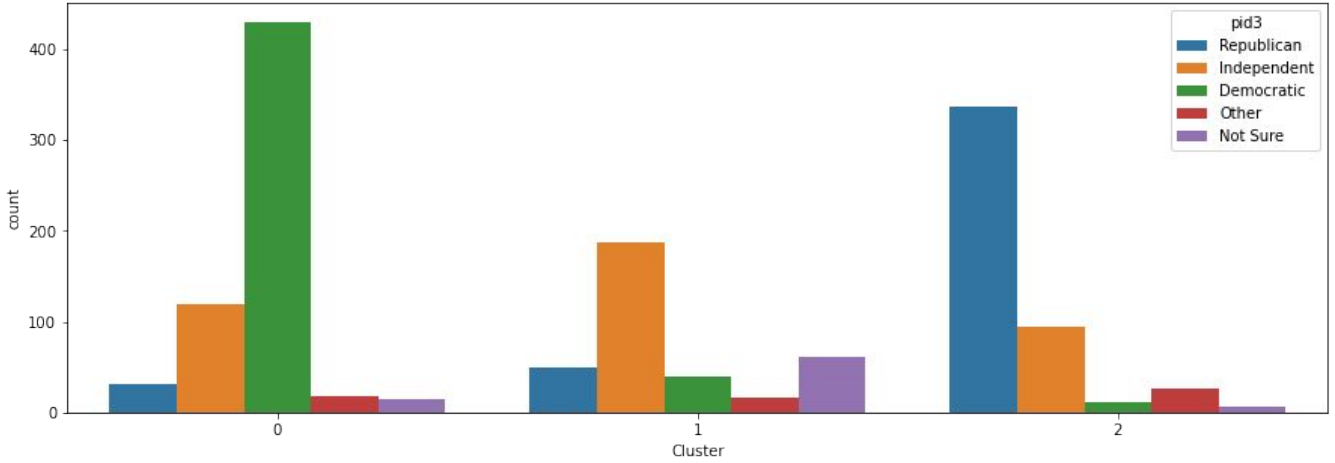
Clusters with Abortion Added

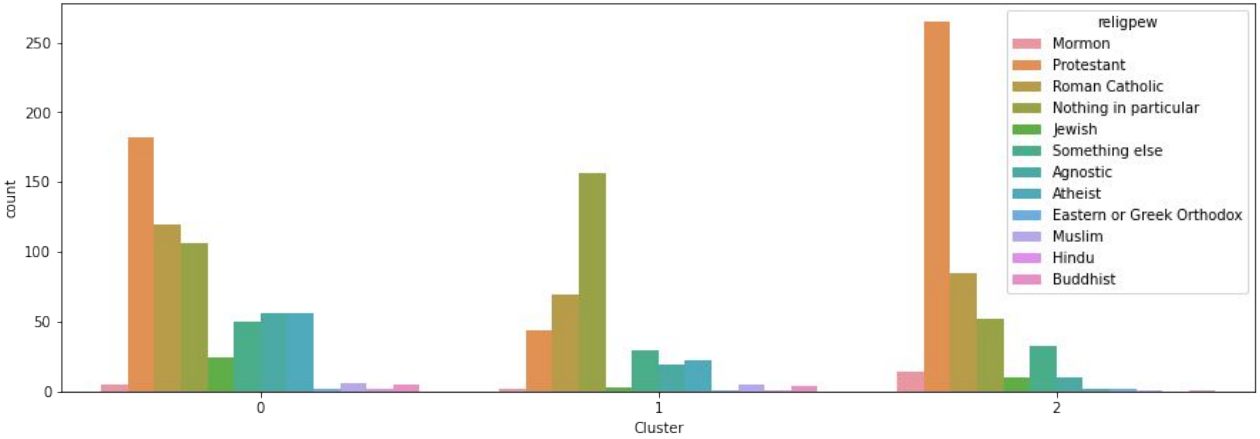


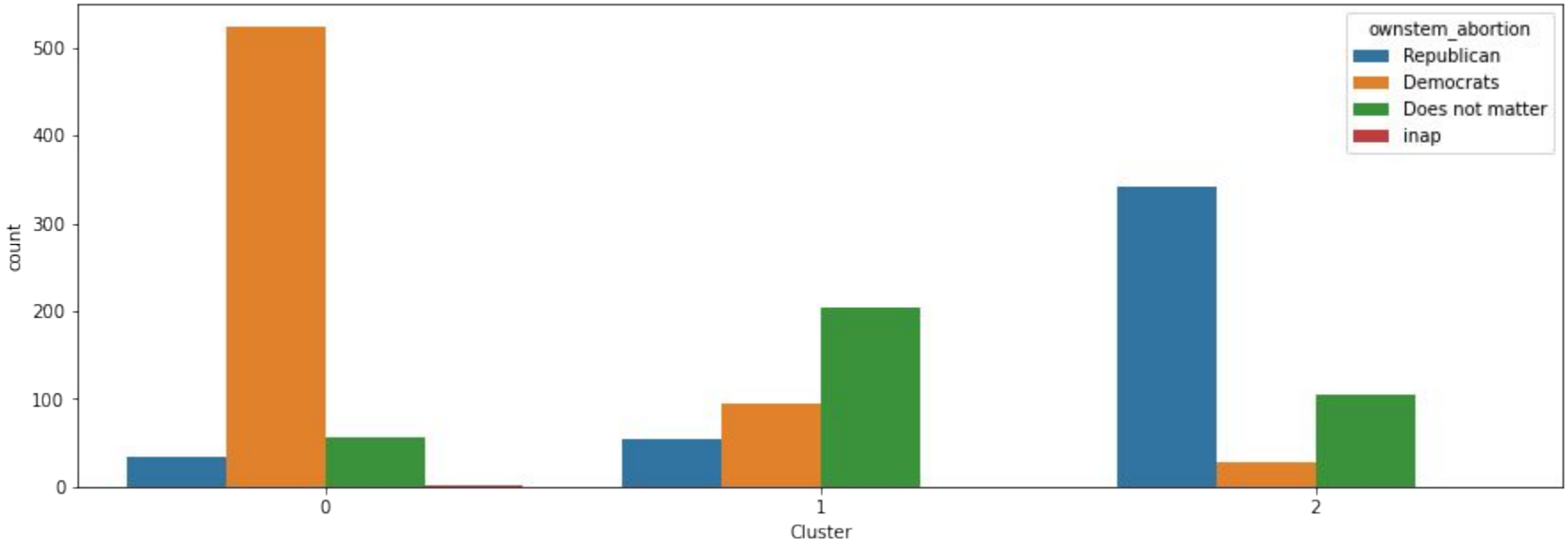
*Clusters are different from prior clusters

Party Affiliation

- Again clusters split by parties

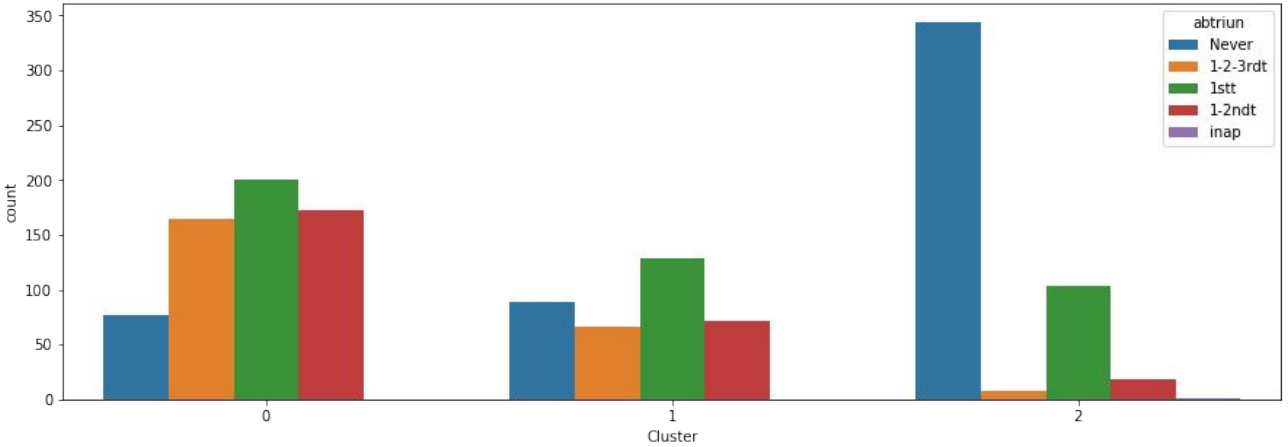


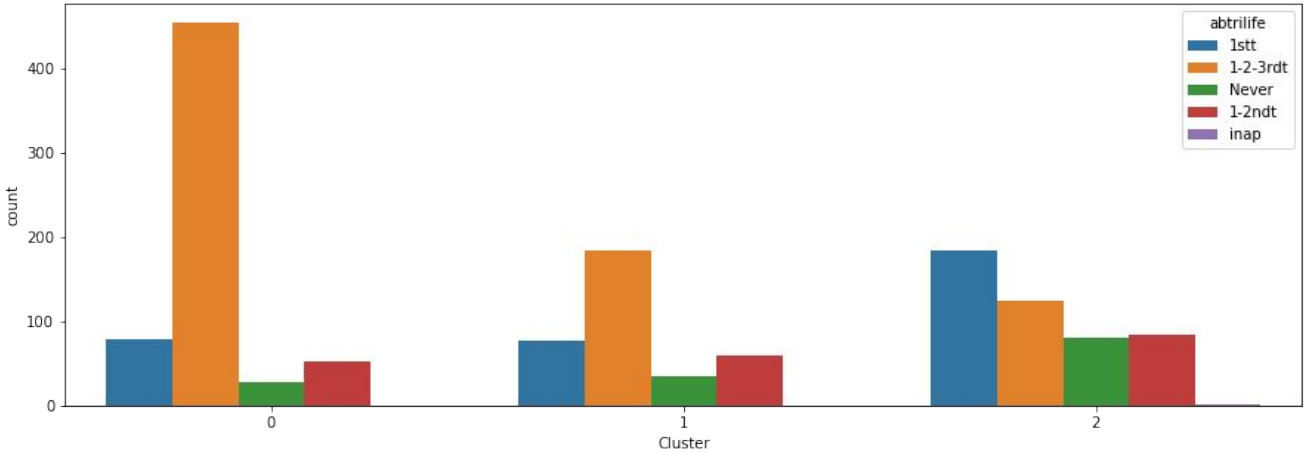




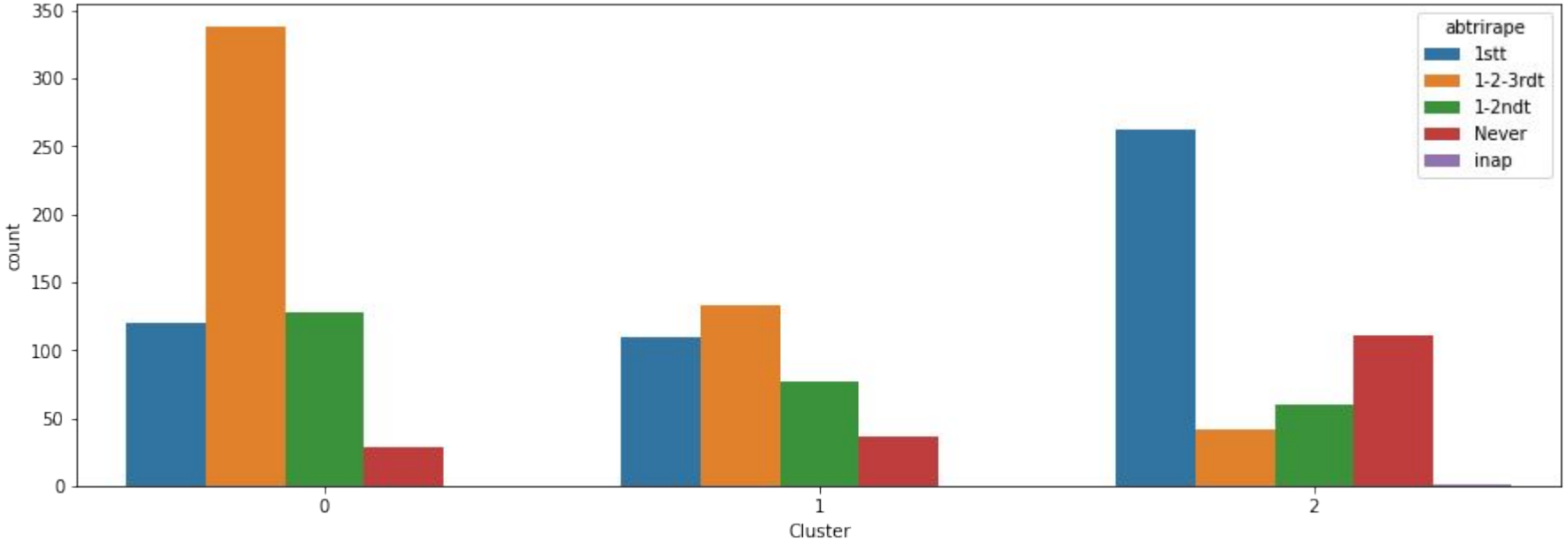
- Ownstem_abortion
  - Which party deals with abortion the best
  - Follows which party they belong to.

- Abrtriun
  - Unwanted pregnancy
  - Cluster 0: pro abortion
  - Cluster 1: mixed feelings
  - Cluster 2: very anti





- Abrtilife
  - Women's life is at risk
  - Cluster 0: Anytime abortion
  - Cluster 1: Anytime abortion
  - Cluster 2: first trimester, but very distributed

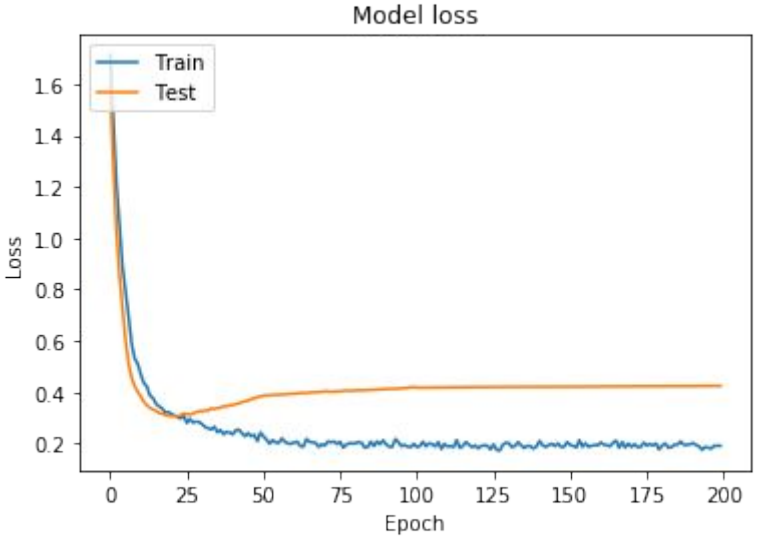


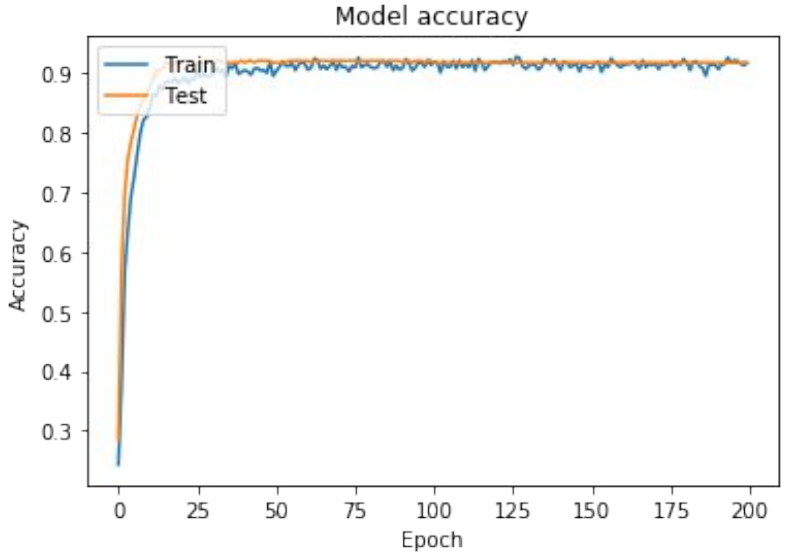
- Arbrtirape
  - Abortion if pregnancy is due to rape
  - Cluster 0: anytime abortion
  - Cluster 1: anytime abortion or first trimester
  - Cluster 2: first trimester abortion or never

Neural Nets

Neural Networks

- We created neural nets using “pid3”, political party affiliation as the end goal.
- Our accuracy for determining the political party was 91.70% +/- 0.27%



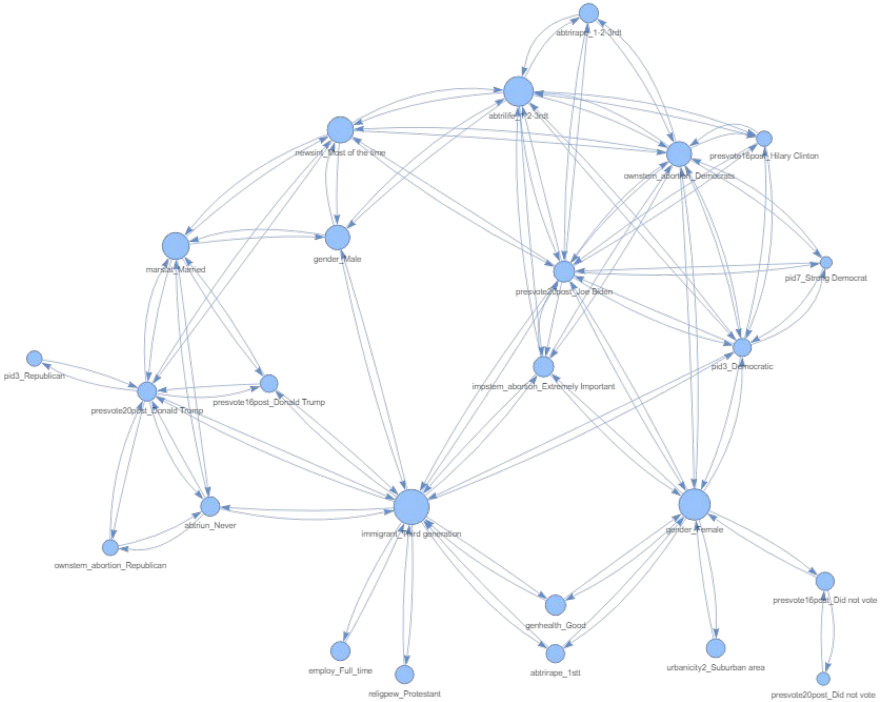


Apriori

Apriori

- Found support, lift, leverage, confidence of the demographics and used it to create a Apriori Graph of Support.
- Best support was usually the more popular yes/no answers
  - 0.700139 (child18_No, votereg_Yes)
  - 0.612188 (ownhome_Own, votereg_Yes)
  - 0.545014 (immigrant_Third generation, votereg_Yes)
  - 0.527008 (ownhome_Own, child18_No)
  - 0.495152 (immigrant_Third generation, child18_No)
  - 0.488227 (votereg_Yes, gender_Female)
  - 0.460526 (votereg_Yes, newsint_Most of the time)

Apriori Graph of Support-With Abortion



Conclusions

Data Conclusion

- The biggest determination of cluster is based on political beliefs and how someone voted.
- The clusters were basically separated into republicans, democrats, and non-voters.
  - Expected outcome
  - Republicans are more anti-abortion
  - Democrats are more pro-abortion
  - Non-voters seemed to be mostly independent.
- Without political parties the data is much harder to cluster.
  - Seems to be random

Project Issues

- Optimal cluster amount was three
  - This made it very politically based
    - Only true way we found there to be useful clusters
- Going from one shell to another
- Zeppelin pyvis
- Apriori visualizations

Thank You!

Questions?