

Project discussion : Dataset on religion survey

Charly Bion
Gözde Uğur
Moiroud Marine
Snell Pierre

June 9, 2018

1 Introduction

- Dataset
- Description

2 Data Cleaning

3 Exploratory analysis

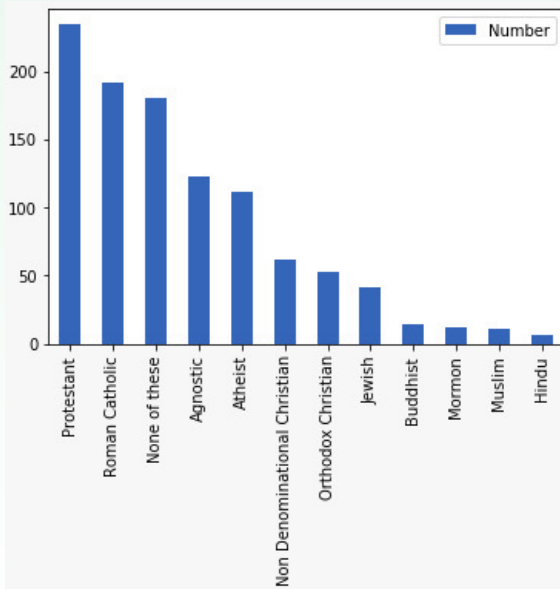
- Basic Analysis
- Standard deviation
- Demographic Analysis

4 Unsupervised learning

5 Supervised Learning

6 Conclusions

Dataset



- Dimensions : 1040 Rows, 48 Columns
- Social and governemental subject
- Data analysis friendly

Dataset on USA population, not representative of world wide population.

18. How comfortable do you feel when you: Ask or offer to pray with someone

- ☐ Extremely comfortable
- ☐ Very comfortable
- ☐ Somewhat comfortably
- ☐ Not so comfortable
- ☐ Not at all comfortable
- ☐ I don't do this

Type of answer:

- Multiple choices
- Yes or No
- Range based

- 1 Introduction
 - Dataset
 - Description
- 2 Data Cleaning
- 3 Exploratory analysis
 - Basic Analysis
 - Standard deviation
 - Demographic Analysis
- 4 Unsupervised learning
- 5 Supervised Learning
- 6 Conclusions

What is your present religion, if any? Unnamed: 1 Do you consider yourself to be an evangelical? Do you attend religious services

0	Response	None of these	Response	Response
1	None of these	Wesleyan Christian	No	Weekly or more
2	Atheist	NaN	NaN	NaN
3	Protestant	NaN	No	Weekly or more

We deleted one row and one columns.

What is your present religion, if any?

Do you consider yourself to be an evangelical?

Do you attend religious services

1	Non Denominational Christian	0	2
2	Atheist	-1	-1
3	Protestant	0	2
4	Muslim	0	2
5	Non Denominational Christian	0	2

Modifying the type of answers

We converted each ranged based answer with a number

- Non applicable → -1
- Not at all comfortable → 0
- Not so comfortable → 1
- Somewhat comfortable → 2
- Very comfortable → 3
- Extremely comfortable → 4

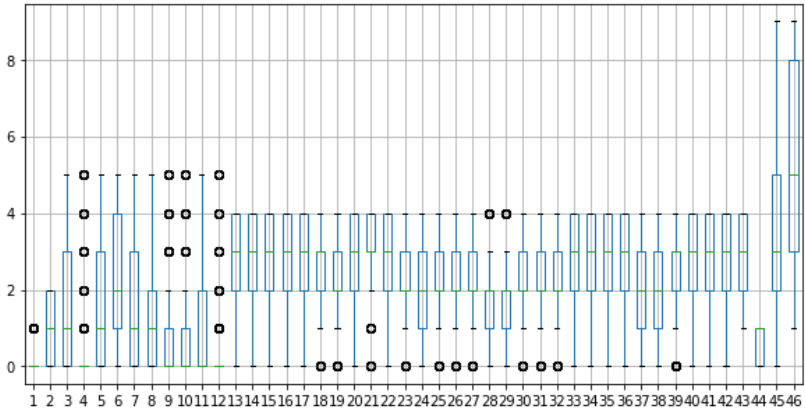
Summary

- 1 Introduction
 - Dataset
 - Description
- 2 Data Cleaning
- 3 Exploratory analysis
 - Basic Analysis
 - Standard deviation
 - Demographic Analysis
- 4 Unsupervised learning
- 5 Supervised Learning
- 6 Conclusions

Boxplot

A well spread dataset

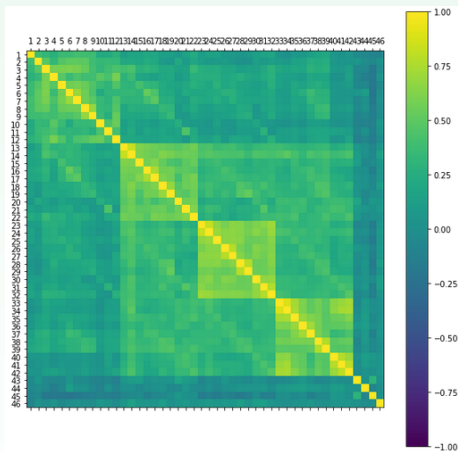
- All options are used
- Well distributed

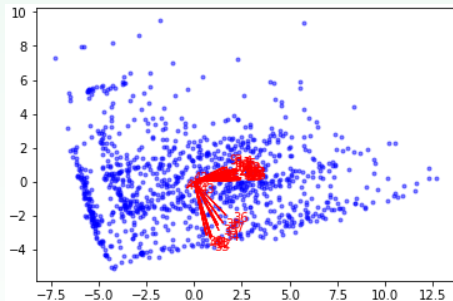


Correlation matrix

Outcomes

- 4 parts
- Age, gender, income and location are not correlated with the rest.





Conclusions

- We can project on two axes
- Relevant with the groups of the correlation matrix
- Gives a base of "answer space"

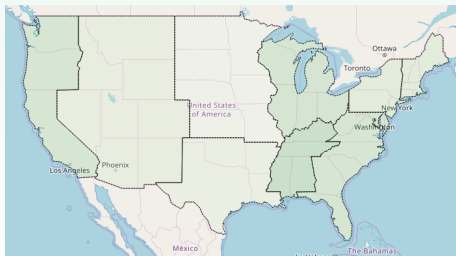
Standard deviation

Question with the highest standard deviation

How often do you pray in public with visible motions (sign of the cross, bowing, prostration, shokeling, etc)?

Result

- Exclude Atheists
- Average: a few times a year
- Muslims: many times per week
- Hindu, Mormon, Orthodox Christian and Roman Catholic people: a few times per month
- Buddhist, Agnostic and Jewish people : less than once a year

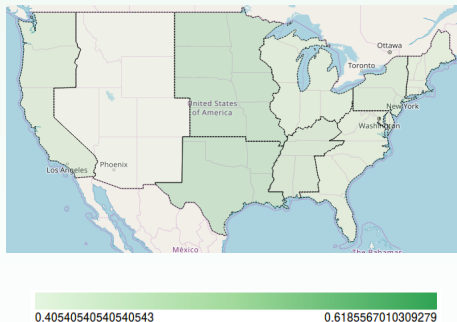


2.356164383561644

2.759259259259259

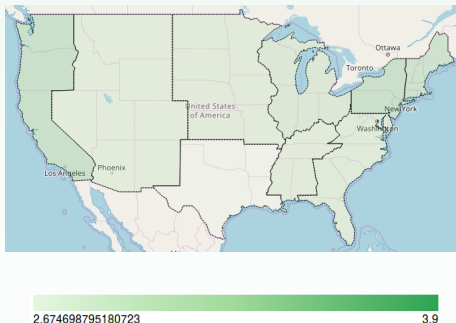
Results

- Old people are discreet
- The older you are, the easier you accept food
- Oldest & youngest are less disturbed when they are declined than the middle aged
- Young people don't like to pray/be prayed for/by others



Results

- Men display public religious behaviour more than women
- Women pray for others more
- Women are more comfortable to witness others' religious behaviours



Results

Less income people:

- Are more evangelical
- Attend religious services more than people with greatest income
- Pray with visible motions/objects than others
- Are more comfortable when they talk about religions, are witnessed, are offered to pray for

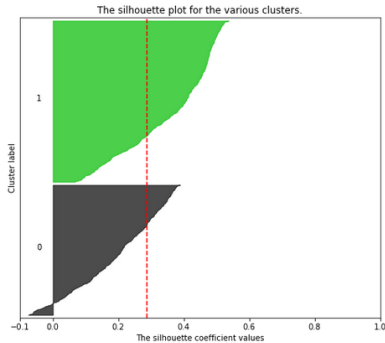
Summary

- 1 Introduction
 - Dataset
 - Description
- 2 Data Cleaning
- 3 Exploratory analysis
 - Basic Analysis
 - Standard deviation
 - Demographic Analysis
- 4 Unsupervised learning
- 5 Supervised Learning
- 6 Conclusions

Unsupervised learning

K-Means & Silhouette plot

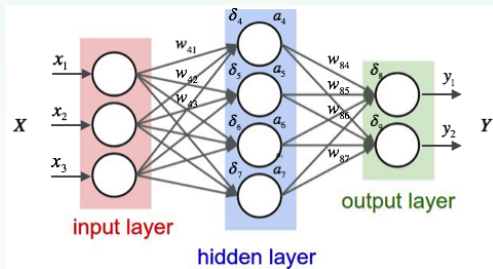
Silhouette analysis for KMeans clustering on sample data with $n_clusters = 2$



Summary

- 1 Introduction
 - Dataset
 - Description
- 2 Data Cleaning
- 3 Exploratory analysis
 - Basic Analysis
 - Standard deviation
 - Demographic Analysis
- 4 Unsupervised learning
- 5 Supervised Learning
- 6 Conclusions

Supervised Learning



Results of training :

- Trainable to guess a religion (+90% Accuracy)
- Trainable for all other columns
- Few data for some religions

Summary

- 1 Introduction
 - Dataset
 - Description
- 2 Data Cleaning
- 3 Exploratory analysis
 - Basic Analysis
 - Standard deviation
 - Demographic Analysis
- 4 Unsupervised learning
- 5 Supervised Learning
- 6 Conclusions

Conclusions

- Need a global and diverse data set
- Other possible uses of our program
- Other possible correlations