MACHINE LEARNING FUNDAMENTALS – CAPSTONE PROJECT

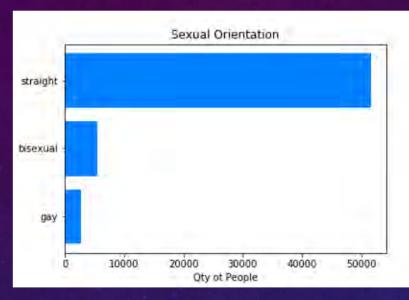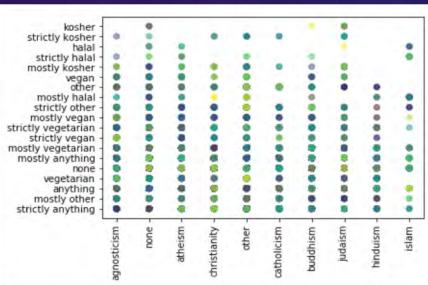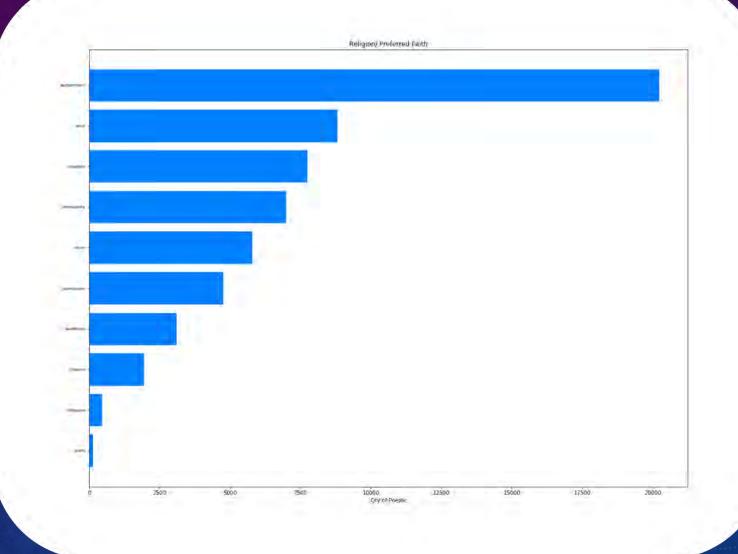CODECADEMY CAPSTONE PROJECT 2018-09 – 2018-11

GRAPHS EXPLORING THE DATASET RECEIVED FROM OKCUPID

Does religion and diet effect orientation or does orientation effect the Religion a person will choose?

QUESTIONS

Question 1

Can you predict the religion preference by orientation?

I expect that people's sexual orientation is one of the major deciding factors when deciding on religious path.

Question 2

Can you predict diet by the religion a person follows?

I expect that people's diet is decided by their religious path.

EXPLANATION OF NEW COLUMNS ADDED TO THE DATASET - RELIGION

Something to note:

In the following dataset all NaN values have been removed and replaced by NONE. No choice was made by the user on profile.

Also, religion data was condensed by only taking the main religion into account – example: agnosticism for all agnosticism but not too serious about it, etc. Done to prevent to many variability in the dataset.

Ordinal mapping for Religion is as follows:

- 0 → agnosticism
- 1 → atheism
- 2 → buddhism
- 3 → Catholicism
- 4 → Christianity
- 5 → hinduism

- 6 → islam
- 7 → judaism
- 8 → other
- 9 → none

This was mapped to column – "religion_code"

EXPLANATION OF NEW COLUMNS ADDED TO THE DATASET - DIET

Something to note:

In the following dataset all NaN values have been removed and replaced by NONE. No choice was made by the user on profile.

Ordinal mapping for diet is as follows:

- 0 → strictly anything
- 1 → mostly other
- 2 → anything
- 3 → vegetarian
- 4 → mostly anything
- 5 → mostly vegetarian

- 6 → strictly vegan
- 7 → strictly vegetarian
- 8 → mostly vegan
- 9 → strictly other
- 10 → mostly halal
- 11 → other

- 12 → vegan
- 13 → mostly kosher
- 14 → strictly halal
- 15 → halal
- 16 → strictly kosher
- 17 → kosher
- 18 → none

This was mapped to column – "diet_code"

EXPLANATION OF NEW COLUMNS ADDED TO THE DATASET - ORIENTATION

Something to note:

In the following dataset all NaN values have been removed and replaced by NONE. No choice was made by the user on profile.

Ordinal mapping for orientation is as follows:

➢ 0 → straight

➢ 1 → bisexual

➢ 2 → gay

This was mapped to column – "orientatiion_code"

CLASSIFICATION: K-NEAREST NEIGHBORS
RELIGION GIVEN ORIENTATION – QUESTION ONE

K-Nearest Neighbors

Fit execution time: 0.6192s

Accuracy: 0.8614

Precision: 0.11

Recall: 0.06

F1-Score: 0.06

K-NEIGHBORS REGRESSOR
RELIGION GIVEN ORIENTATION – QUESTION ONE

Support Vector Machines

Fit execution time: 23.401s

Accuracy: 0.8630

Precision: 0.74

Recall: 0.86

F1-Score: 0.80

REGRESSION: LINEAR REGRESSION
DIET GIVEN RELIGION – QUESTION TWO

Multiple Linear Regression

Fit execution time: 0.0018s

Accuracy: 0.0113

K-NEIGHBORS REGRESSOR
DIET GIVEN RELIGION – QUESTION TWO

K-Neighbors Regressor

Fit execution time: 0.704s

Accuracy: -0.192

CONCLUSION OF DATA

Question 1: Can you predict the religion preference by orientation?

With the condensed data done on religion, it seems as if it is possible to predict that a certain type of orientation chooses a certain type of religion.

Question 2: Can you predict diet by the religion a person follows?

According to the data, it look as if religion is not such a good driver for a preferred diet, but this could be due to the amount of option – I might get a better result if I clean the data, for example: stay with the main option – VEGAN, VEGETARIAN, KOSHER. Rather than the with words, mostly, strictly.

I would use the same type of function than that of religion.

The dataset is rich and many more options exist, you can add diet, orientation and religion and play with all three option to see what the outcome will be?

You can also tackle income and hob correlation as well as sex and orientation.