



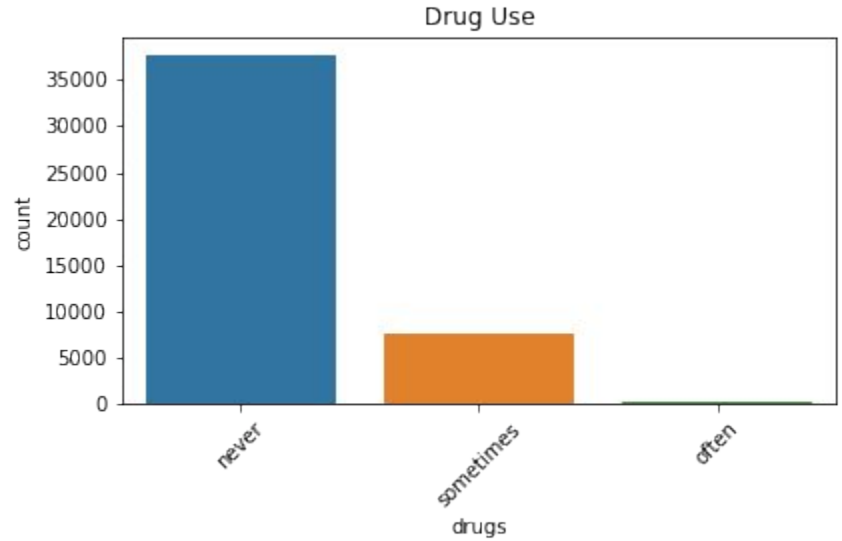
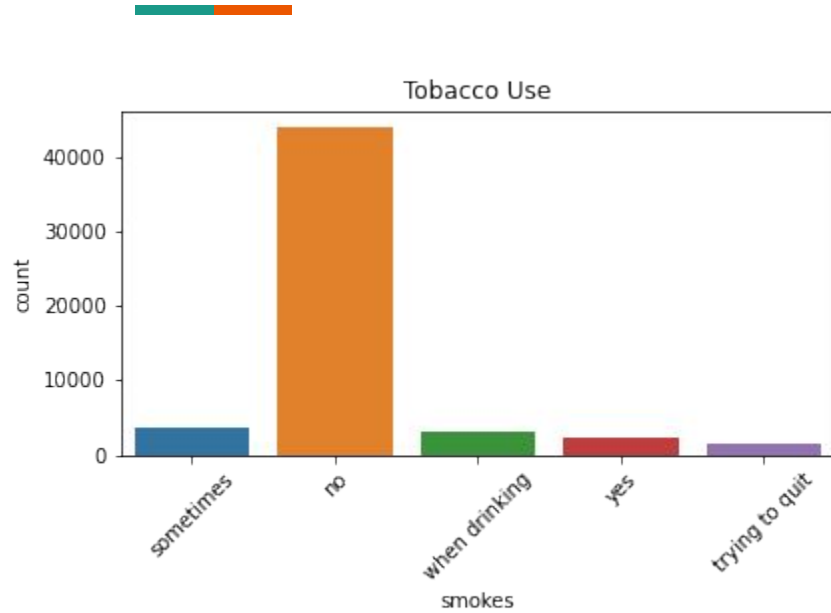
# **Codecademy OkCupid Data Scientist Project**

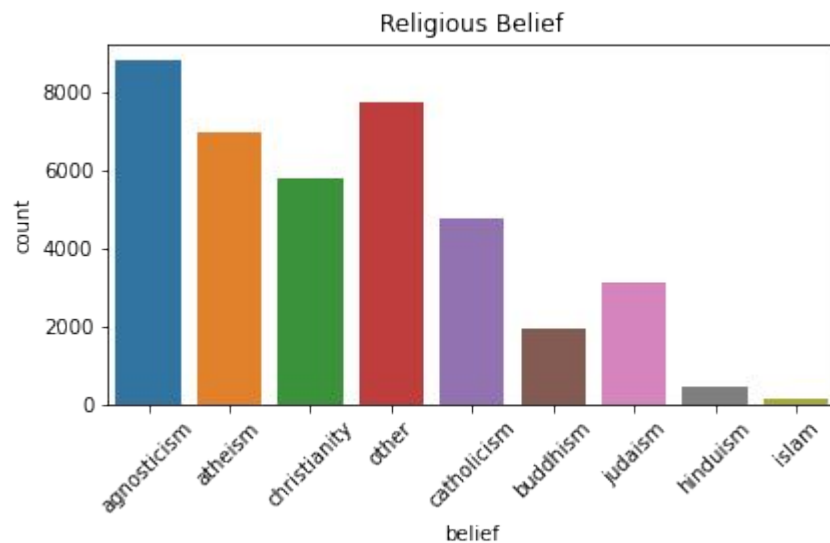
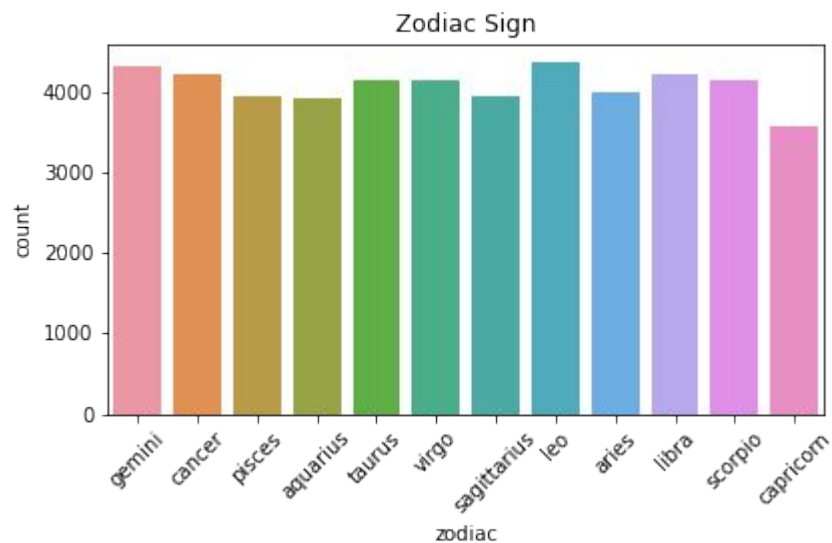


## Goals of the Project

Data from OkCupid was analyzed with the purpose of determining if a user's zodiac sign could be accurately predicted from other data said user had provided.

The data showed a wide breadth of responses in relation to drug use, smoking and drinking habits, religious belief, and zodiac sign (for those users who did provide the information).







Logistic regression, nearest neighbor algorithm (with 8 nearest neighbors considered), and random forest were the machine learning methods used to attempt to predict zodiac signs. Aspects considered from users included religious belief, drug use, smoking habits, drinking habits, and diet. Of the three machine learning algorithms, random forest was the most successful, with an accuracy of approximately 25%



	precision	recall	f1-score	support
aquarius	0.09	0.02	0.03	1082
aries	0.11	0.07	0.08	1158
cancer	0.11	0.19	0.14	1227
capricorn	0.11	0.10	0.10	1043
gemini	0.10	0.27	0.14	1256
leo	0.11	0.15	0.13	1225
libra	0.12	0.09	0.10	1152
pisces	0.09	0.02	0.03	1108
sagittarius	0.15	0.01	0.03	1091
scorpio	0.11	0.14	0.12	1163
taurus	0.09	0.05	0.06	1171
virgo	0.09	0.10	0.10	1190
accuracy			0.10	13866
macro avg	0.11	0.10	0.09	13866
weighted avg	0.11	0.10	0.09	13866

## Logistic Regression

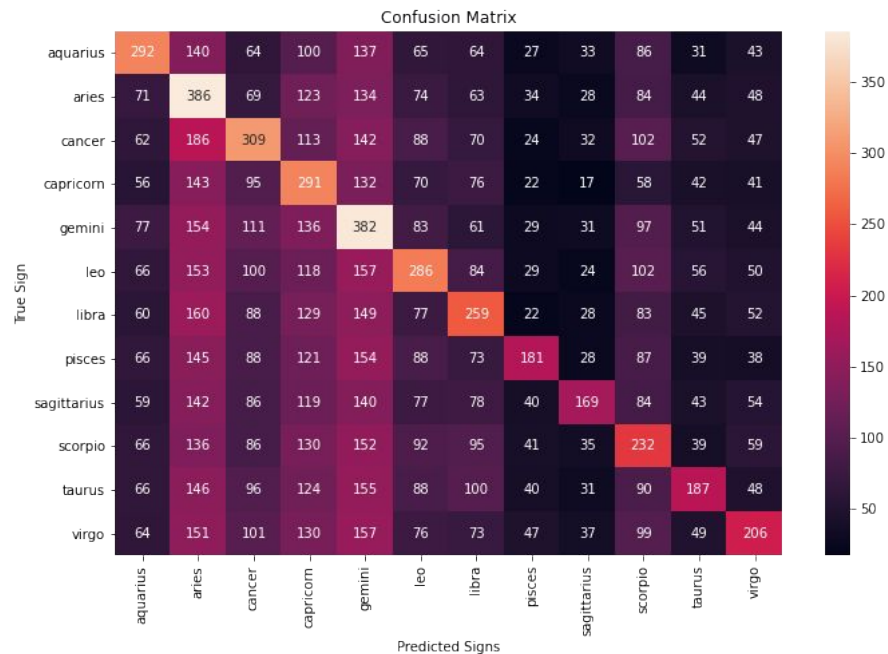
	precision	recall	f1-score	support
aquarius	0.16	0.19	0.17	1082
aries	0.12	0.28	0.17	1158
cancer	0.20	0.14	0.16	1227
capricorn	0.16	0.14	0.15	1043
gemini	0.12	0.32	0.17	1256
leo	0.18	0.16	0.17	1225
libra	0.20	0.11	0.14	1152
pisces	0.18	0.09	0.12	1108
sagittarius	0.23	0.10	0.14	1091
scorpio	0.14	0.12	0.13	1163
taurus	0.18	0.10	0.13	1171
virgo	0.24	0.08	0.12	1190
accuracy			0.15	13866
macro avg	0.17	0.15	0.15	13866
weighted avg	0.17	0.15	0.15	13866


## K Nearest Neighbor (k=8)



	precision	recall	f1-score	support
aquarius	0.29	0.27	0.28	1082
aries	0.19	0.33	0.24	1158
cancer	0.24	0.25	0.25	1227
capricorn	0.18	0.28	0.22	1043
gemini	0.19	0.30	0.24	1256
leo	0.25	0.23	0.24	1225
libra	0.24	0.22	0.23	1152
pisces	0.34	0.16	0.22	1108
sagittarius	0.34	0.15	0.21	1091
scorpio	0.19	0.20	0.20	1163
taurus	0.28	0.16	0.20	1171
virgo	0.28	0.17	0.21	1190
accuracy			0.23	13866
macro avg	0.25	0.23	0.23	13866
weighted avg	0.25	0.23	0.23	13866

Random Forest





Overall the predictions for zodiac signs were not accurate enough to draw any conclusions between the data and zodiac signs. To increase the accuracy, more elements of user data could be used, including body type, education level, pet ownership, sexual orientation, whether or not the user has children, or even languages spoken.