Mining Beer DataSet
Project Description

Abstract

To explore and mine hidden patterns within beer datasets (brewery, beer, beer review) and answer questions which come into mind when one thinks about a variety of beers.

Introduction

Covid-19 The first dataset which comes into one's mind in today's time, is very important but we are already looking at its analysis on the news like 10 times a day. Which made me think in the opposite direction and seek datasets which can be challenging as well as fun at the same time to display my skills gained in R.

It reminded me of last year's spring break, me and my friends took a road trip South. Towards Virginia, North Carolina to enjoy the best nature had to offer i.e. "The Smoky Mountains". And taste different varieties of beer we could come across. During this trip we would stop to seek out famous local breweries and try different varieties of beer they had to offer. The process would always be to seek out breweries, in the area we were in, via the internet or through recommendation by locals; and learn about their types of beers, ratings and so on before visiting the place out. Now thinking back to those times I wanted to take a data Scientist approach to the same questions and seek answers using the below data sources.

Data Sources

- https://www.kaggle.com/nickhould/craft-cans#beers.csv (beers, breweries)
- https://data.world/socialmediadata/beeradvocate (beer_ratings)

Questions

Some of the questions which I wish to answer using the data set:

- Which are the highest Rated breweries and where are they located? City wise, State wise? Plot this on a US map.
- Which are the highest Rated beers and the location they are produced? Plot this on a US map.
- What kind of beer would you find most in a particular region?
- Which type of beer has a higher rating?
- Which breweries are producing the most rated beers?
- Which state has the most number of breweries based on different pallets, etc.