

# Alcohol Consumption

Coursera Capstone

# Datasets

- The Datasets were taken from the GHO website (the Global Health Observatory)
- These datasets were:
  - Alcohol Abstainers over the past 12 months : <https://apps.who.int/gho/data/node.main.A1042?lang=en>
  - Heavy episodic drinking over the past 30 days <https://apps.who.int/gho/data/node.main.A1047?lang=en>
  - 15 to 19 year olds current drinkers <https://apps.who.int/gho/data/node.main.A1214?lang=en>
  - Alcohol related road traffic crash deaths <https://apps.who.int/gho/data/node.main.A1095?lang=en>
  - Alcohol attributed fractions, liver cirrhosis deaths <https://apps.who.int/gho/data/node.main.A1094?lang=en>

# Initial Datasets

Out[ 8 ]:

Unnamed: 0		Alcohol-attributable fractions (15+), liver cirrhosis deaths (%)	Alcohol-attributable fractions (15+), liver cirrhosis deaths (%).1
0	NaN	Male	Female
1	Country	2016	2016
2	Afghanistan	5.5	1.1
3	Albania	70.3	45.6
4	Algeria	22.7	7

Out[ 3 ]:

Unnamed: 0		Unnamed: 1	15-19 years old, current drinkers (%)	15-19 years old, current drinkers (%).1	15-19 years old, current drinkers (%).2
0	Country	Year	Both sexes	Male	Female
1	Afghanistan	2016	0.3 [0.3-0.4]	0.5 [0.4-0.7]	0.1 [0.1-0.2]
2	Albania	2016	37.7 [34.8-40.5]	50.5 [46-54.9]	24.2 [21.1-27.6]
3	Algeria	2016	1.5 [1.3-1.8]	2.3 [1.9-2.8]	0.7 [0.6-0.9]
4	Andorra	2016	64.2 [-]	76.6 [-]	50.7 [-]

Out[ 9 ]:

Unnamed: 0		Unnamed: 1	Alcohol, heavy episodic drinking (15+), drinkers only, past 30 days (%)	Alcohol, heavy episodic drinking (15+), drinkers only, past 30 days (%).1	Alcohol, heavy episodic drinking (15+), drinkers only, past 30 days (%).2
0	Country	Year	Both sexes	Male	Female
1	Afghanistan	2016	2	2.3	0.5
2	Albania	2016	44.7	57.4	23.7
3	Algeria	2016	19.3	23.5	6.2
4	Andorra	2016	40.5	55	21.7

Out[ 6 ]:

Unnamed: 0		Unnamed: 1	Alcohol, abstainers past 12 months (%)	Alcohol, abstainers past 12 months (%).1	Alcohol, abstainers past 12 months (%).2
0	Country	Year	Both sexes	Male	Female
1	Afghanistan	2016	99.3	99.0	99.7
2	Albania	2016	49.0	35.0	62.4
3	Algeria	2016	96.8	95.2	98.5
4	Andorra	2016	25.4	14.8	35.7

Out[ 4 ]:

Unnamed: 0		Unnamed: 1	Age limits off-premise sales	Age limits off-premise sales.1	Age limits off-premise sales.2	Age limits on-premise service	Age limits on-premise service.1	Age limits on-premise service.2
0	Country	Year	Beer	Wine	Spirits	Beer	Wine	Spirits
1	Afghanistan	2016	Total ban	Total ban	Total ban	Total ban	Total ban	Total ban
2	Albania	2016	18	18	18	18	18	18
3	Algeria	2016	18	18	18	18	18	18
4	Andorra	2016	18	18	18	18	18	18

# Cleaned Dataset

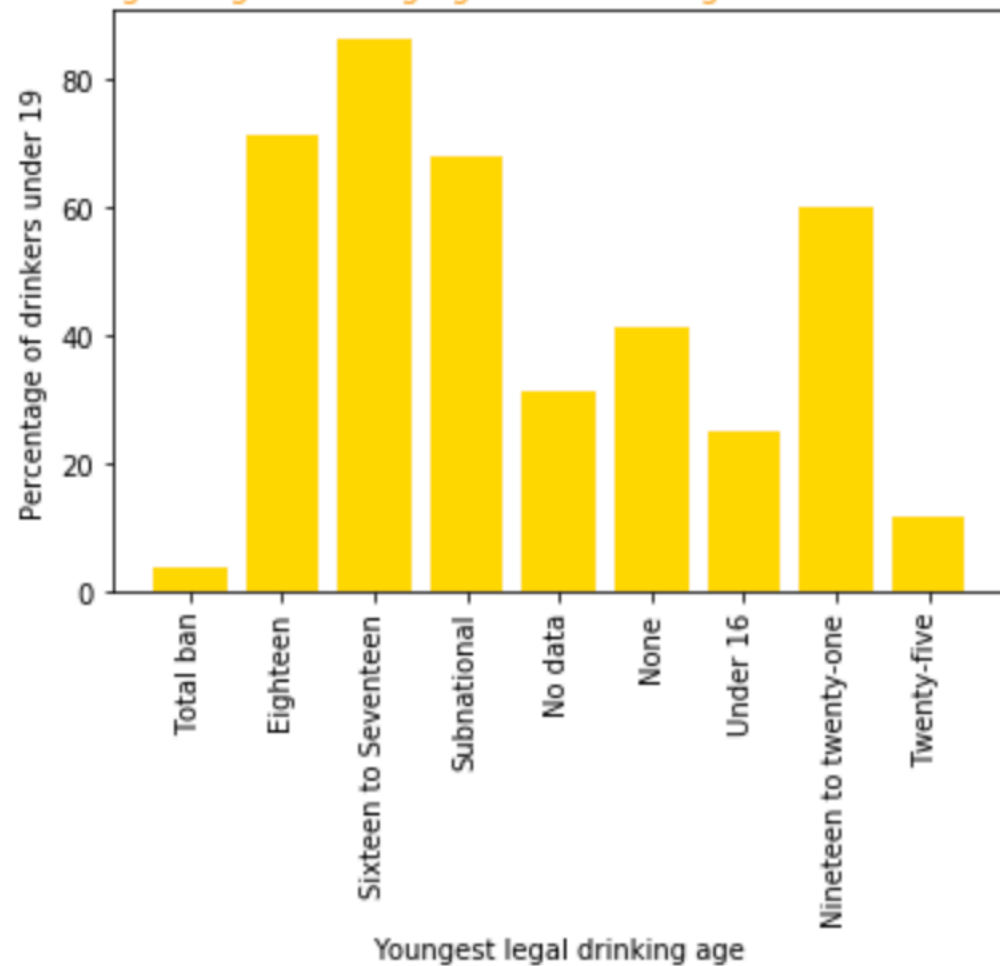
07]:

	Year	Total Heavy Episodic Drinking	Male Heavy Episodic Drinking	Female Heavy Episodic Drinking	Male liver cirrhosis deaths	Female liver cirrhosis deaths	Total Abstainers	Male Abstainers	Female Abstainers	Male road traffic deaths	Female road traffic deaths	Total drinkers under 19	Male drinkers under 19	Female drinkers under 19	Youngest legal drinking age
Unnamed: 0															
<b>Afghanistan</b>	2016	2.0	2.3	0.5	5.5	1.1	99.3	99.0	99.7	0.7	0.5	0.3	0.5	0.1	Total ban
<b>Albania</b>	2016	44.7	57.4	23.7	70.3	45.6	49.0	35.0	62.4	37.9	28.0	37.7	50.5	24.2	Eighteen
<b>Algeria</b>	2016	19.3	23.5	6.2	22.7	7.0	96.8	95.2	98.5	3.1	2.0	1.5	2.3	0.7	Eighteen
<b>Andorra</b>	2016	40.5	55.0	21.7	50.0	50.0	25.4	14.8	35.7	50.0	50.0	64.2	76.6	50.7	Eighteen
<b>Angola</b>	2016	81.6	90.1	66.8	60.9	40.0	52.3	38.4	65.7	48.7	42.0	33.5	46.0	21.1	Eighteen

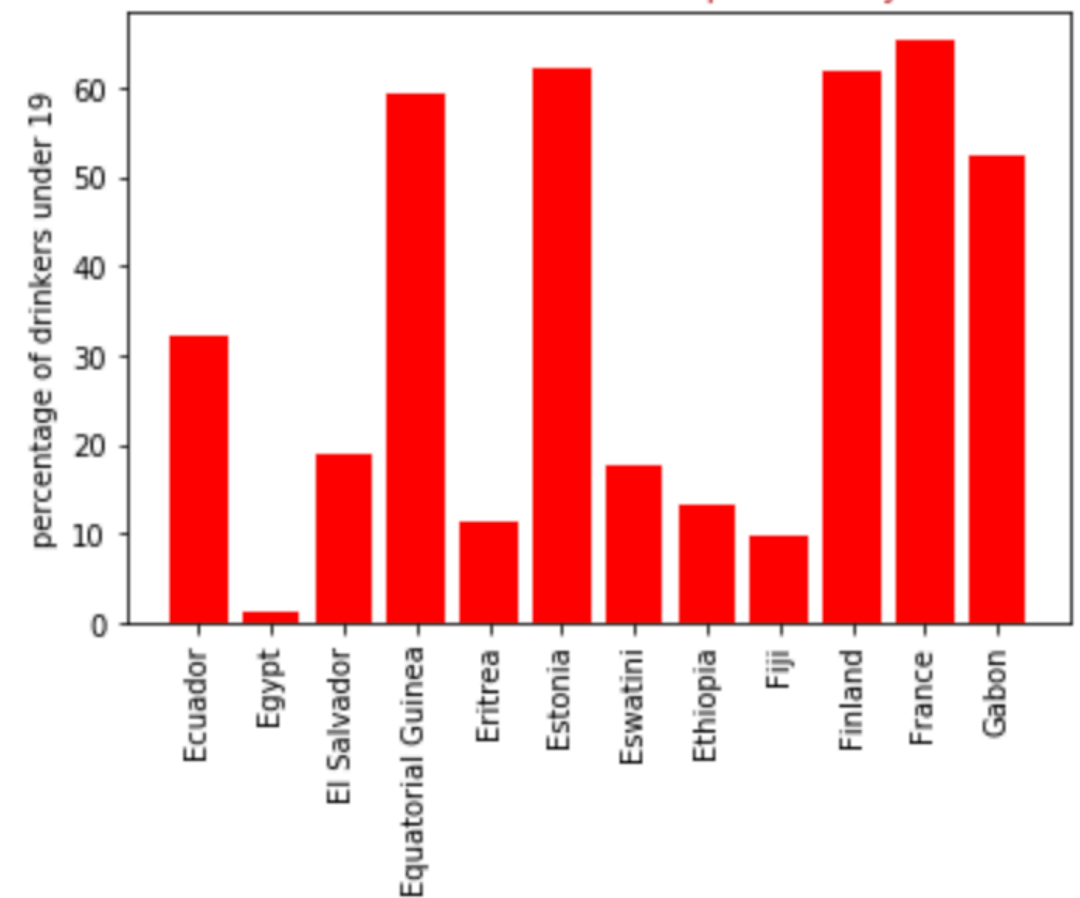
- The data has been merged into one DataFrame.
- The labels have been changed to make them simpler.
- The columns with different spirits, wines and off licence beer have been deleted and one column with the youngest legal drinking age has replaced them.
- The youngest legal drinking age has been changed into a number of categories.
- The columns with the drinkers under 19 have been edited to show numbers only.
- The countries have become the index.

# Percentage Of drinkers under 19

Youngest legal drinking age vs. Percentage of drinkers under 19

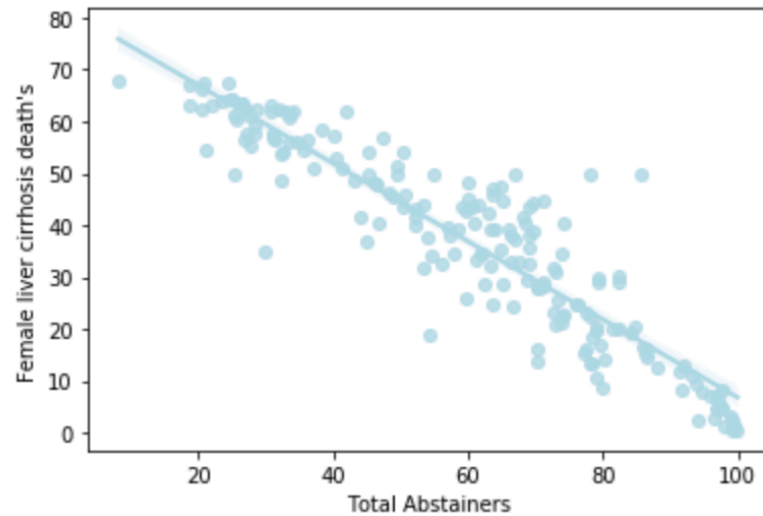


Total drinkers under 19 per country

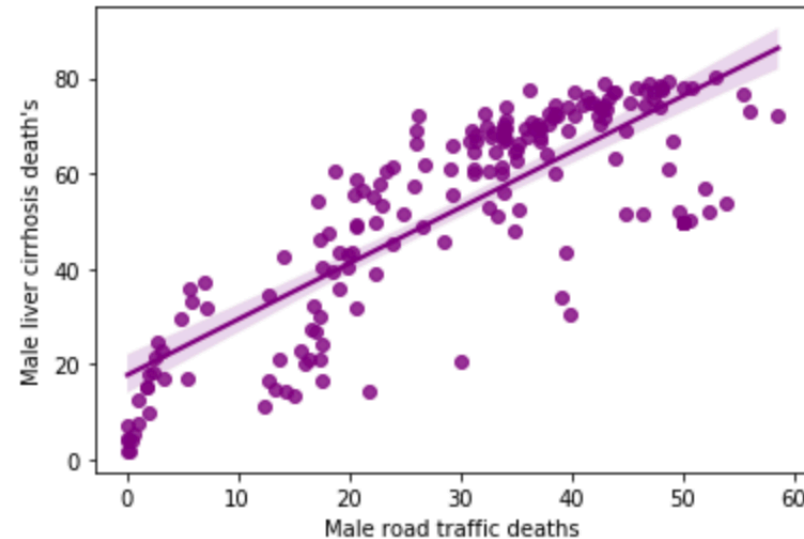


# *(Features which combine to do the same thing???)*

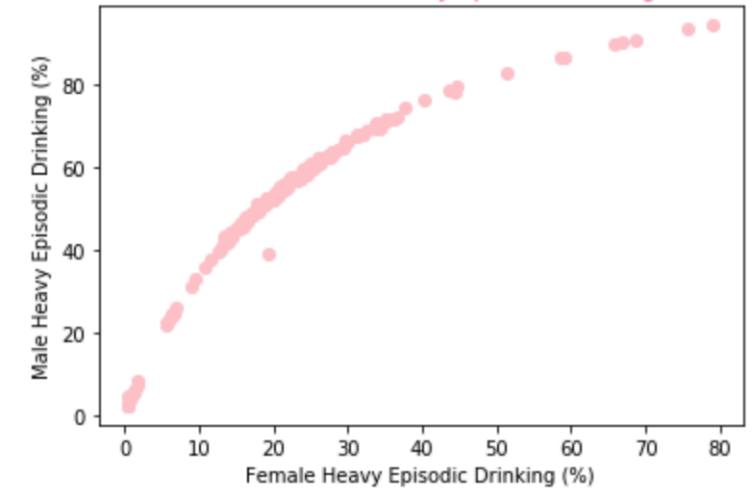
Total abstainers vs. female liver cirrhosis death's



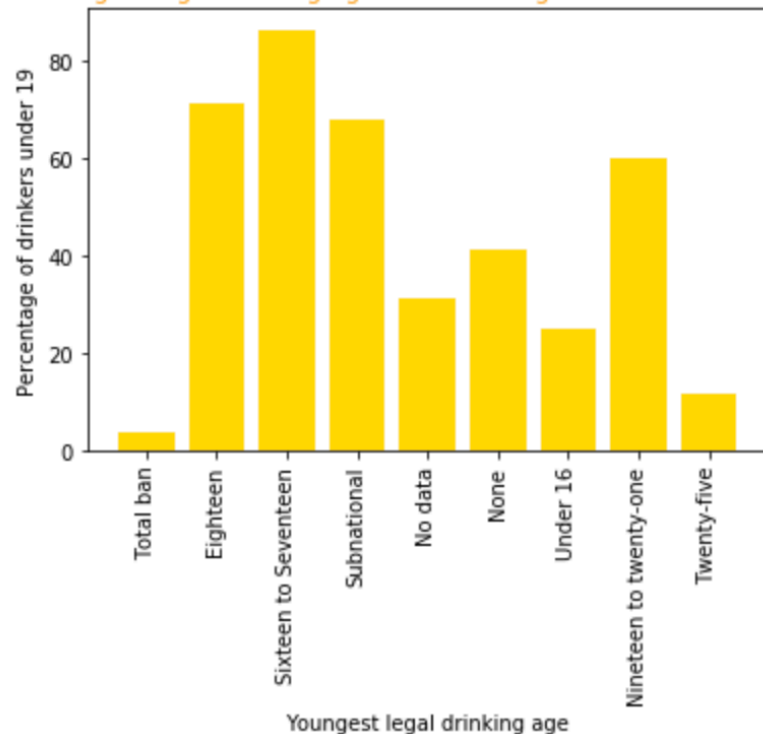
Male road traffic deaths vs. Male liver cirrhosis death's



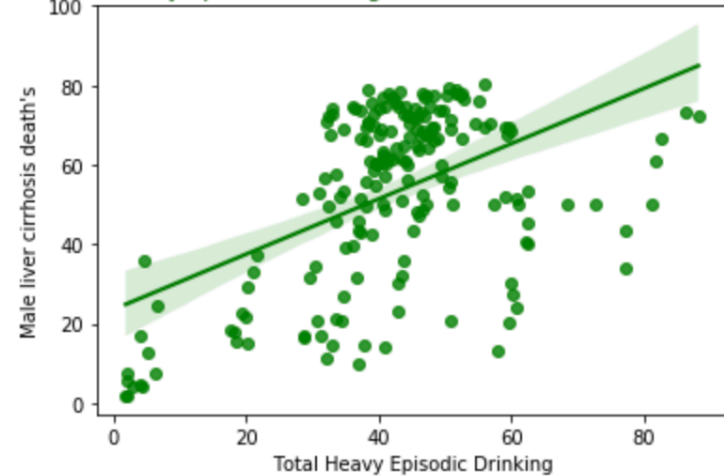
Male vs. Female heavy episodic drinking



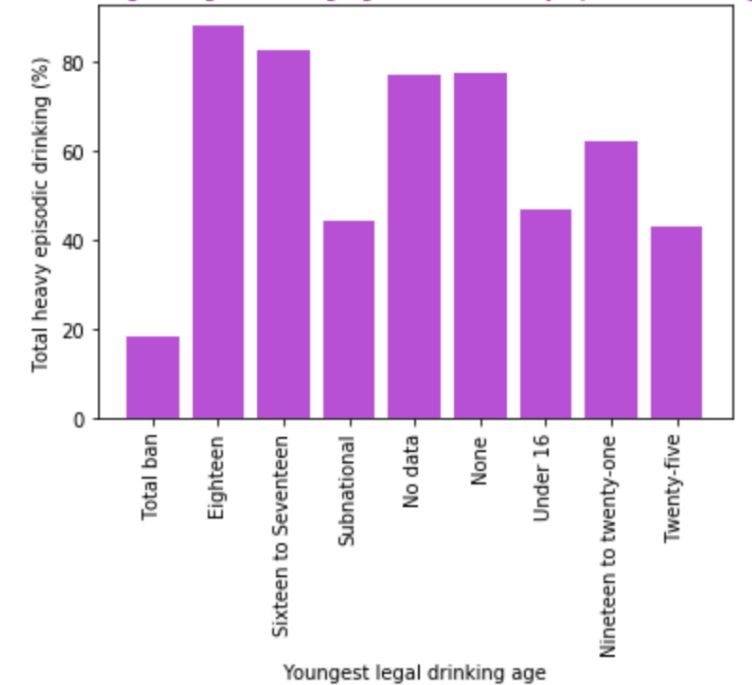
Youngest legal drinking age vs. Percentage of drinkers under 19



Total heavy episodic drinking vs. Male liver cirrhosis death's.pdf

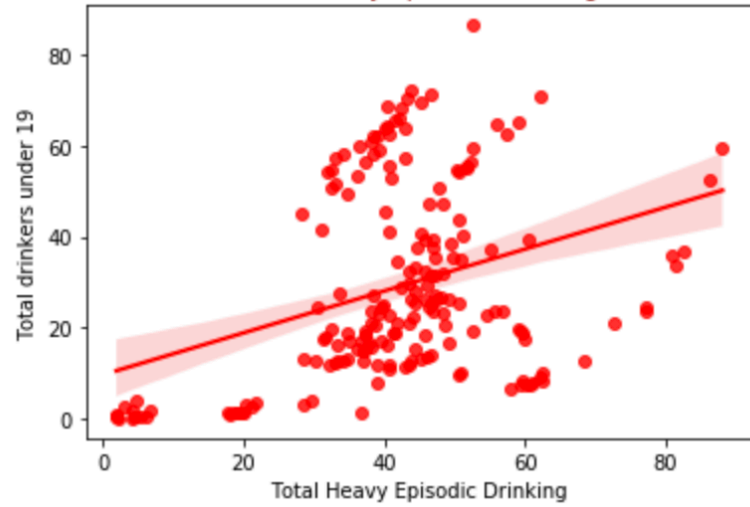


Youngest legal drinking age vs total heavy episodic drinking

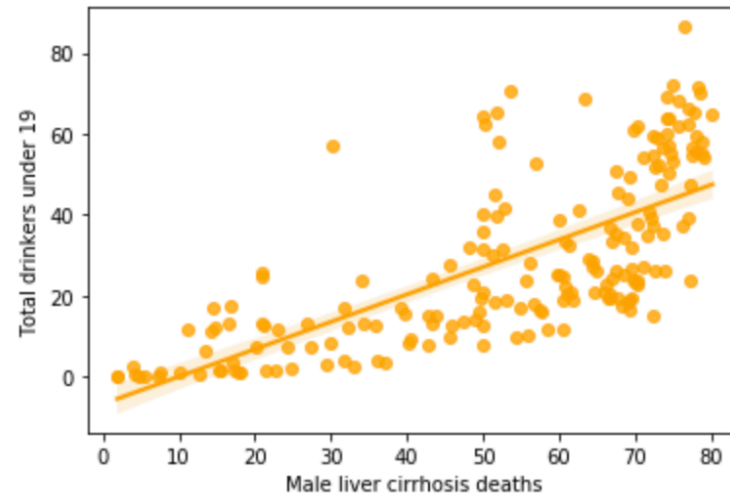


# Correlations

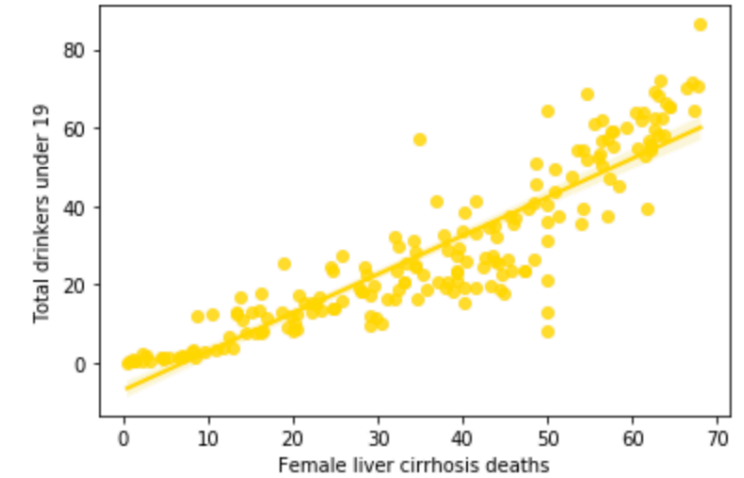
Total Heavy Episodic Drinking



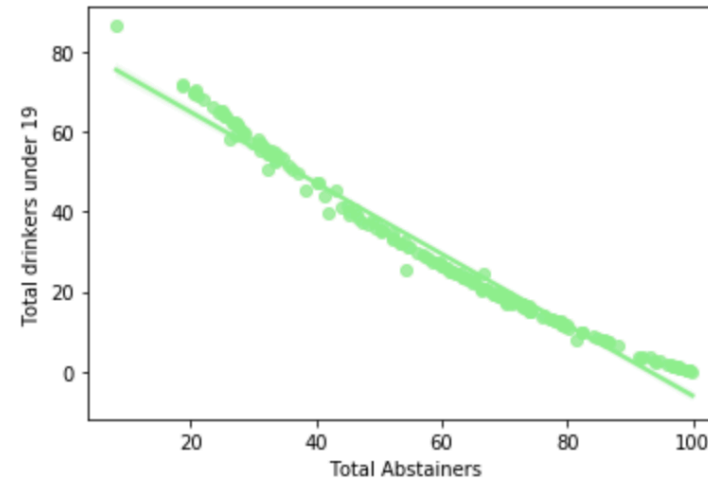
Male liver cirrhosis deaths



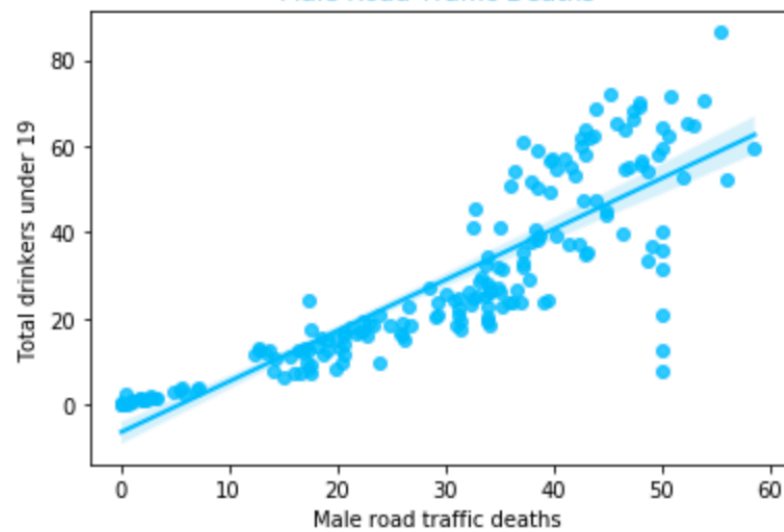
Female liver cirrhosis deaths



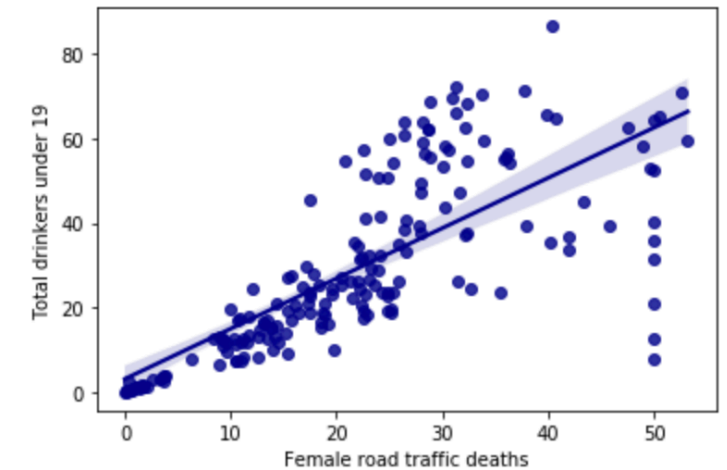
Total Abstainers



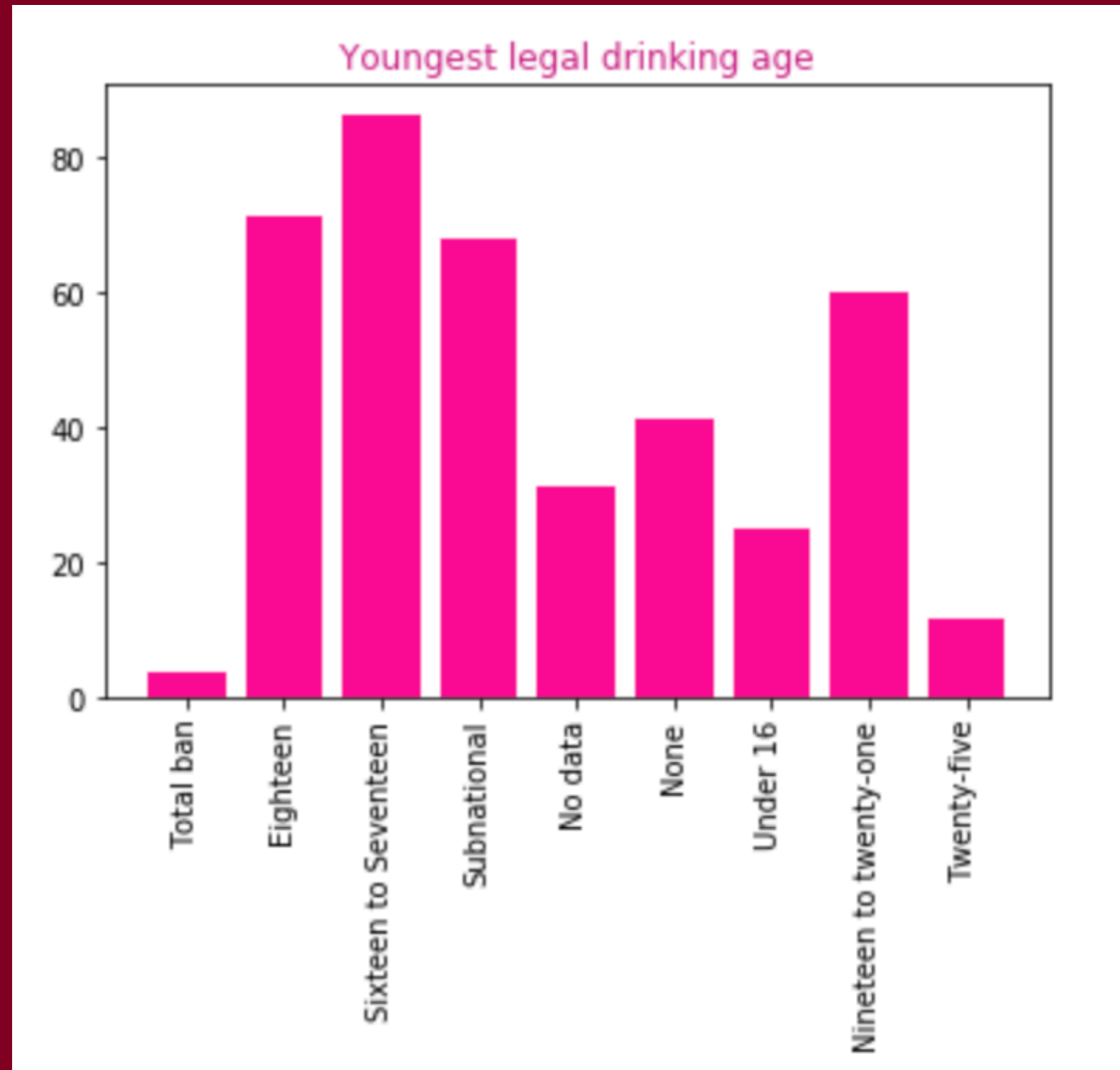
Male Road Traffic Deaths



Female Road Traffic Deaths



# Correlation between youngest legal drinking age





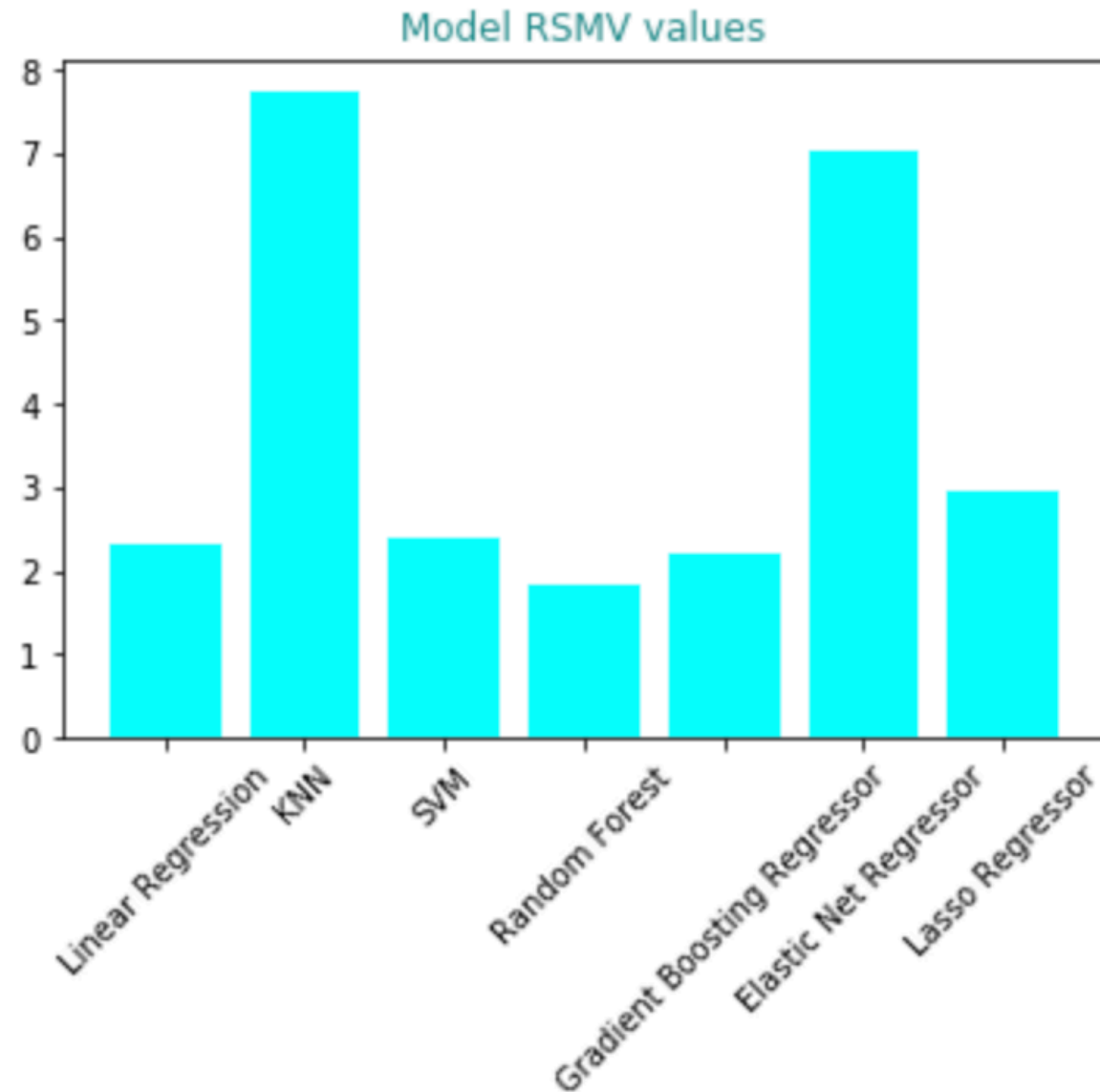
# Correlation between features

<u>Feature</u>	<u>Pearson R Correlation score</u>
Total Heavy Episodic Drinking	0.34796
Male liver cirrhosis deaths	0.70956
Female liver cirrhosis deaths	0.8975
Total Abstainers	-0.98884
Male road traffic deaths	0.85391
Female road traffic deaths	0.75527

# Regression Models

- We made regression models because the numbers being investigated were the percentage of drinkers under 19.
- The models I decided to make were:
  - Support Vector Regression
  - K-Nearest Neighbour Regressor
  - Linear Regression
  - Gradient Boosting Regression
  - Random Forest Regression
  - Elastic Net Regression
  - Lasso Regression

# Model Evaluation



4 ] :

	Models	RMSE
0	Linear Regression	2.326989
1	KNN	7.745028
2	SVM	2.395010
3	Random Forest	1.850992
4	Gradient Boosting Regressor	2.200069
5	Elastic Net Regressor	7.018678
6	Lasso Regressor	2.954112

- The Random Forest model performed the best and the KNN model performed the worst.