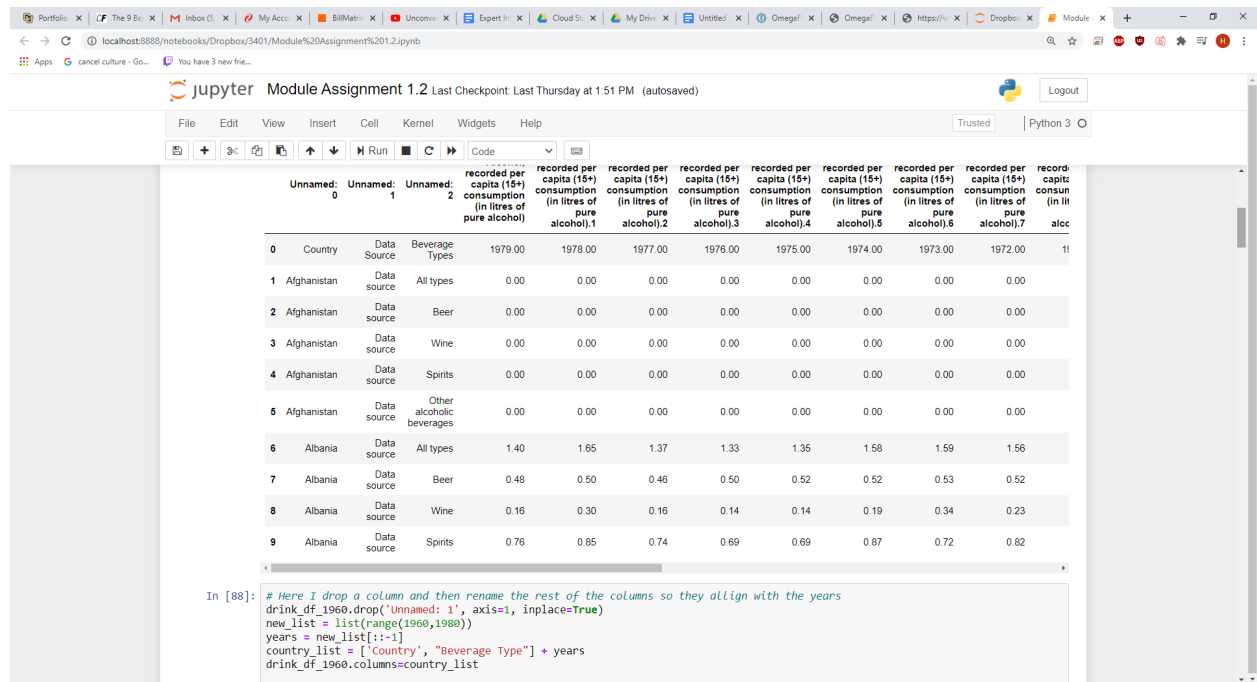


Data Tidying and Visualizing

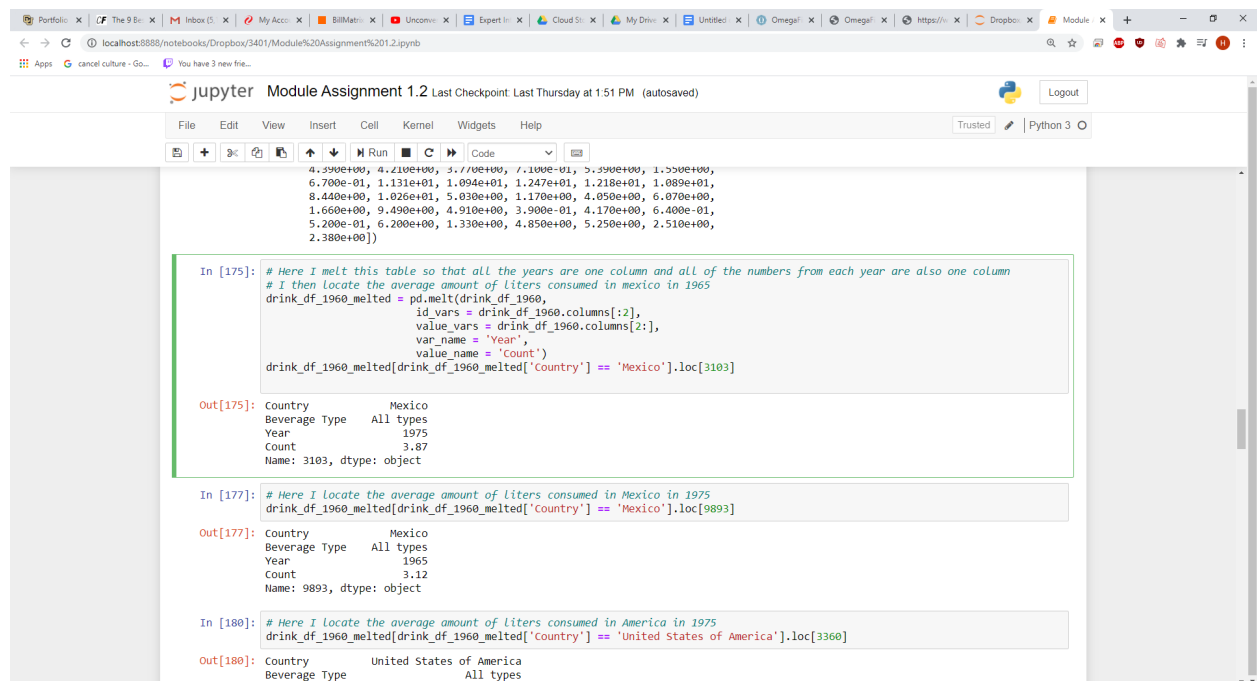


Unnamed: 0 Unnamed: 1 Unnamed: 2 recorded per capita (15+) consumption (in litres of pure alcohol).1 recorded per capita (15+) consumption (in litres of pure alcohol).2 recorded per capita (15+) consumption (in litres of pure alcohol).3 recorded per capita (15+) consumption (in litres of pure alcohol).4 recorded per capita (15+) consumption (in litres of pure alcohol).5 recorded per capita (15+) consumption (in litres of pure alcohol).6 recorded per capita (15+) consumption (in litres of pure alcohol).7 recorded per capita (15+) consumption (in litres of pure alcohol).8

	Country	Data source	Beverage Types	1979.00	1978.00	1977.00	1976.00	1975.00	1974.00	1973.00	1972.00	1971.00
0												
1	Afghanistan	Data source	All types	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	Afghanistan	Data source	Beer	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	Afghanistan	Data source	Wine	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	Afghanistan	Data source	Spirits	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	Afghanistan	Data source	Other alcoholic beverages	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	Albania	Data source	All types	1.40	1.65	1.37	1.33	1.35	1.58	1.59	1.56	1.56
7	Albania	Data source	Beer	0.48	0.50	0.46	0.50	0.52	0.52	0.53	0.52	0.52
8	Albania	Data source	Wine	0.16	0.30	0.16	0.14	0.14	0.19	0.34	0.23	0.23
9	Albania	Data source	Spirits	0.76	0.85	0.74	0.69	0.69	0.87	0.72	0.82	0.82

```
In [18]: # Here I drop a column and then rename the rest of the columns so they align with the years
drink_df_1960.drop('Unnamed: 1', axis=1, inplace=True)
new_list = list(range(1960,1980))
years = new_list[::-1]
country_list = ['Country', "Beverage Type"] + years
drink_df_1960.columns=country_list
```

For the first step of this project I use the `.drop` function to get rid of useless columns. I then concatenate the two string columns with the integer columns for the years 1960 - 1979. This tidy's up the data and makes the table much easier to read.



```
In [175]: # Here I melt this table so that all the years are one column and all of the numbers from each year are also one column
# I then locate the average amount of liters consumed in Mexico in 1965
drink_df_1960_melted = pd.melt(drink_df_1960,
                               id_vars = drink_df_1960.columns[:2],
                               value_vars = drink_df_1960.columns[2:],
                               var_name = 'Year',
                               value_name = 'Count')
drink_df_1960_melted[drink_df_1960_melted['Country'] == 'Mexico'].loc[3103]
```

```
Out[175]: Country      Mexico
Beverage Type  All types
Year          1975
Count         3.87
Name: 3103, dtype: object
```

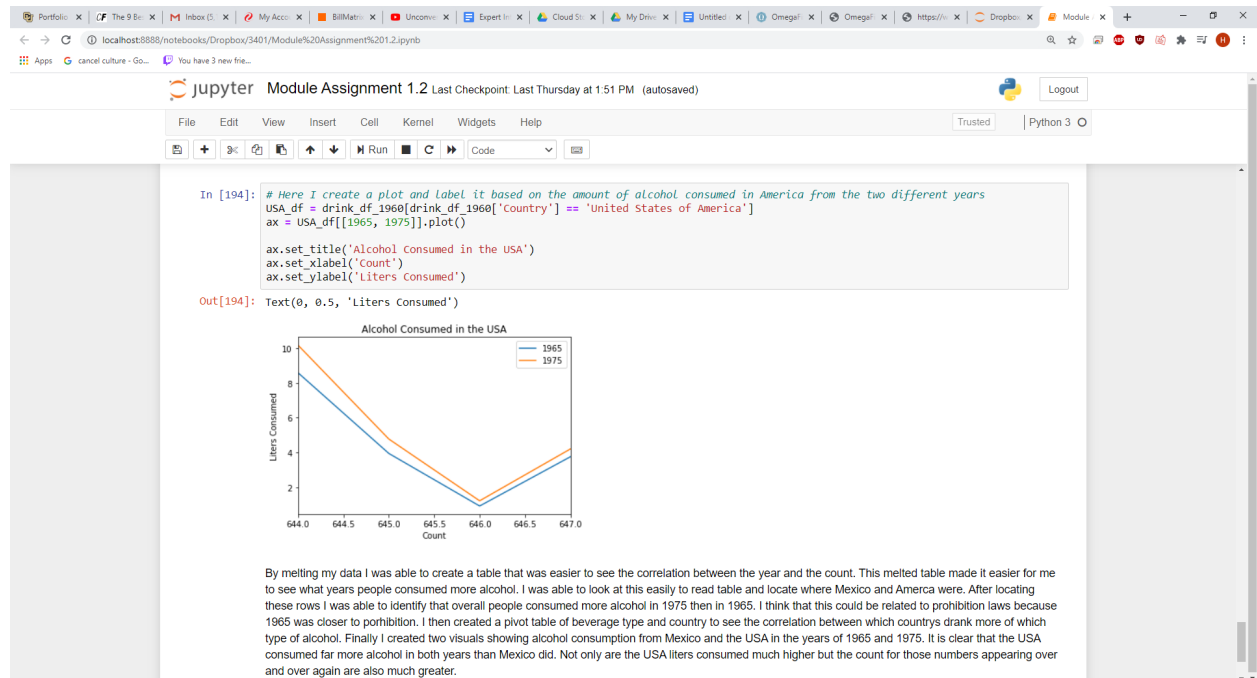
```
In [177]: # Here I locate the average amount of liters consumed in Mexico in 1975
drink_df_1960_melted[drink_df_1960_melted['Country'] == 'Mexico'].loc[9893]
```

```
Out[177]: Country      Mexico
Beverage Type  All types
Year          1965
Count         3.12
Name: 9893, dtype: object
```

```
In [180]: # Here I locate the average amount of liters consumed in America in 1975
drink_df_1960_melted[drink_df_1960_melted['Country'] == 'United States of America'].loc[3360]
```

```
Out[180]: Country      United States of America
Beverage Type  All types
Year          1975
Count         3.87
Name: 3360, dtype: object
```

Above I am melting the data so that I can add up all of the years into one count column.



Here I visualize my findings by comparing the alcohol consumption between the USA and Mexico for the years 1965-1975