



Alcohol Consumption in Russia (1998-2016)

Exploratory Data Analysis

Vinicius Mattoso

Context

This is Alcohol Consumption in Russia (1998-2016) Dataset. It contains values of consumption for wine, beer, vodka, brandy and champagne.

Content

This Dataset has 1615 rows and 7 columns. Keys for columns:

- "year" - year (1998-2016);
- "region" - name of a federal subject of Russia. It could be oblast, republic, krai, autonomous okrug, federal city and a single autonomous oblast;
- "wine" - sale of wine in litres by year per capita;
- "beer" - sale of beer in litres by year per capita;
- "vodka" - sale of vodka in litres by year per capita;
- "champagne" - sale of champagne in litres by year per capita;
- "brandy" - sale of brandy in litres by year per capita.

<https://www.kaggle.com/dwdkills/alcohol-consumption-in-russia>

Agenda

01

Data Loading

How do we read a file in csv format?

02

Plotly Library

Here we will briefly explain this library that allows interactive graphics.

03

Graphic Analysis

We will present some questions and we will answer by graphs.

04

Conclusion

Conclusions and insights generated from the analysis.

01

Data Loading



Loading the database and displaying some rows.

	year	region	wine	beer	vodka	champagne	brandy
694	2006	Jewish Autonomous Oblast	3.7	56.3	7.5	0.4	0.10
340	2002	Republic of Adygea	4.4	30.4	7.9	0.7	0.30
641	2005	Omsk Oblast	4.0	28.2	9.2	0.7	0.30
1212	2012	Kaluga Oblast	9.3	64.0	11.9	2.0	0.90
110	1999	Republic of Karelia	2.4	21.0	15.0	0.3	0.10
448	2003	Kamchatka Krai	7.4	39.1	25.1	1.8	1.20
1373	2014	Republic of Dagestan	0.4	2.2	2.9	0.2	0.30
109	1999	Karachay-Cherkess Republic	2.1	6.9	10.9	0.1	0.04
435	2003	Volgograd Oblast	2.7	46.7	13.2	1.8	0.50
460	2003	Lipetsk Oblast	2.7	54.9	10.1	0.7	0.20

```
#Loading pandas
```

```
import pandas as pd
```

```
#loading the .csv file
```

```
sales=pd.read_csv('russia_alcohol.csv')
```

```
#looking 10 samples
```

```
sales.sample(10)
```

Evaluating the data types in each column and checking for null values.

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 1615 entries, 0 to 1614  
Data columns (total 7 columns):  
 #   Column      Non-Null Count  Dtype    
---  -  
 0   year        1615 non-null   int64    
 1   region      1615 non-null   object   
 2   wine        1552 non-null   float64  
 3   beer        1557 non-null   float64  
 4   vodka       1554 non-null   float64  
 5   champagne   1552 non-null   float64  
 6   brandy      1549 non-null   float64  
dtypes: float64(5), int64(1), object(1)  
memory usage: 88.4+ KB
```

```
#Evaluating the data types in each column  
and checking for null values
```

```
sales.info()
```



02

Plotly Library

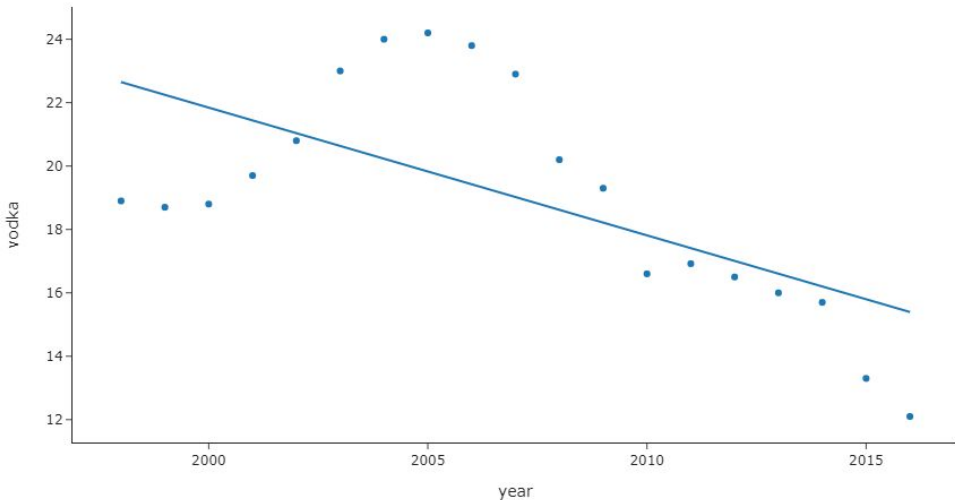
<https://plotly.com/graphing-libraries/>

03

Graphic Analysis



Assessing the trend over time for a product from a given region



```
# Assessing the trend over time  
drink="vodka"  
region='Moscow'  
df_aval_2=sales.copy()  
df_graph =  
df_aval_2[df_aval_2['region']==region]  
fig = px.scatter(df_graph, x="year",  
y=drink,trendline="ols",  
template="simple_white")  
fig.show()
```

Assessing the trend over time

```
fig = go.Figure()

fig.add_trace(go.Scatter(x=df_aval_1['year'], y=df_aval_1['wine'],
                        mode='lines+markers', name='Wine'))

fig.add_trace(go.Scatter(x=df_aval_1['year'], y=df_aval_1['beer'],
                        mode='lines+markers', name='Beer'))

fig.add_trace(go.Scatter(x=df_aval_1['year'], y=df_aval_1['vodka'],
                        mode='lines+markers', name='Vodka'))

fig.add_trace(go.Scatter(x=df_aval_1['year'], y=df_aval_1['champagne'],
                        mode='lines+markers', name='Champagne'))

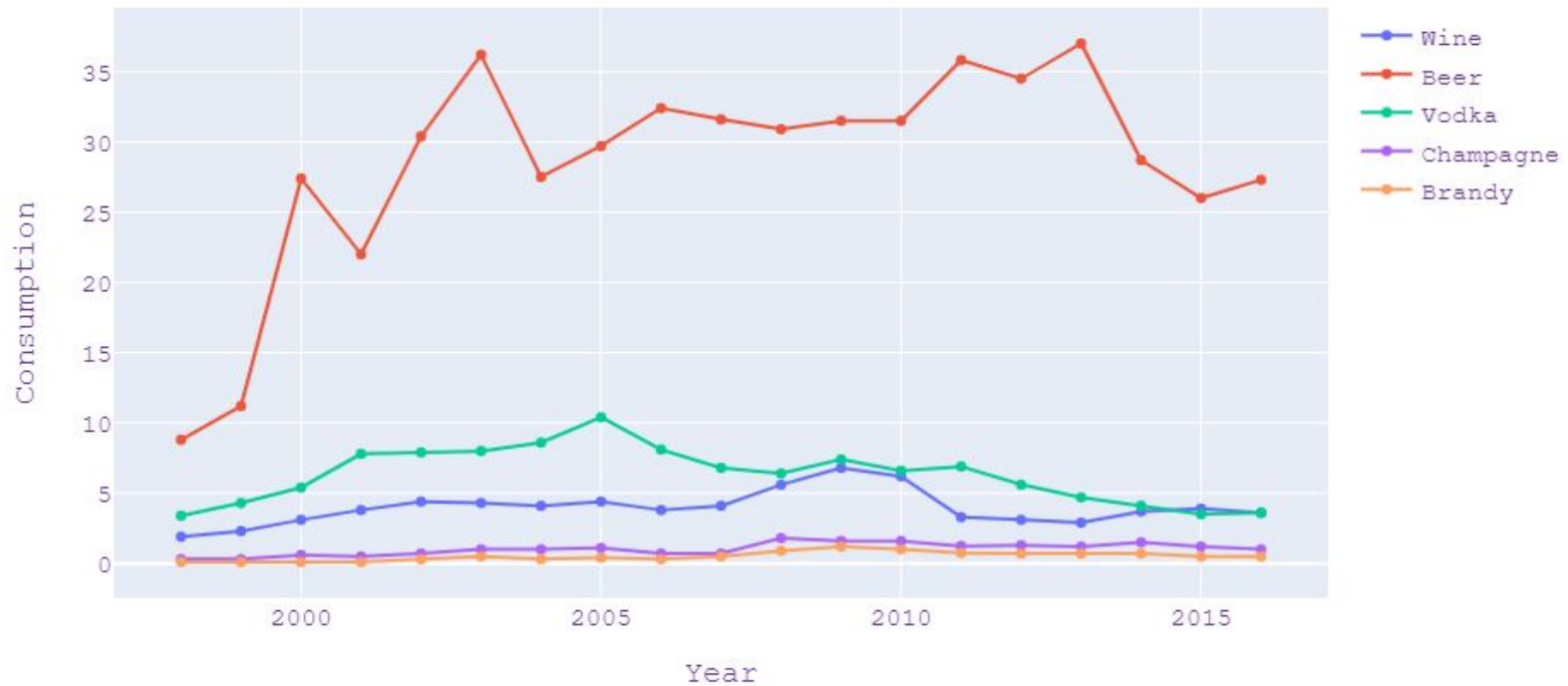
fig.add_trace(go.Scatter(x=df_aval_1['year'], y=df_aval_1['brandy'],
                        mode='lines+markers', name='Brandy'))

fig.update_layout(
    title="Consumption throughout the year",
    xaxis_title="Year",
    yaxis_title="Consumption",
    font=dict(
        family="Courier New, monospace",
        size=15,
        color="RebeccaPurple" ))

fig.show()
```

Assessing the trend over time

Consumption throughout the year



CHEERS!



