World alcohol consumption analysis (2010-2019)

Master in BI & Big Data Analytics (2022/2023)
Data Ingestion, ETL & Data Quality, Data Visualization

Fabio Jr Lorenzini
Milano, August 2023

Agenda

1. Facts
2. Needs & Objectives
3. BI Architecture
4. Data Source
5. Data Profiling & ETL
6. Data Visualization
7. Achieved Objectives
8. Limits Encountered and Next Steps
9. Conclusions

1. Facts

*“The **global alcoholic drinks market size** was valued at **1448,2 billion $** in 2021 and is expected to expand at a compound annual **growth rate** of **10,3%** from 2022 to 2028.”* (Grand Review Research, 2021)

“*Worldwide, **3 million deaths every year** result from harmful use of alcohol. This represents 5,3% of all deaths.*” (World Health Organization, 2022)

“***Alcohol: alarm in Europe***”. European citizens drinks around 15 litres of alcohol per capita each year. Beer is the preferred drink (44%) in central Europe, followed by wine (34%) in southern Europe and spirits (22%). (Istituto Superiore di Sanità, 2020)

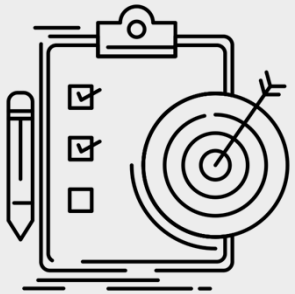
“*In **Italy**, the 49,4% of male and 44,4% of female aged between 11 and 25 years old consumed at least one alcoholic drink over the year. […]*
***Binge drinkers** -who drinks at least 5 alcohol beverages in few hours- **between 11 and 25** are almost **1 million**.*” (ISTAT, 2022)

2. Needs & Objectives

Alcohol abuse is cause of several physical issues (e.g. liver-gastrointestinal, cardiovascular and neurological problems, cancer), mental and behavioral disorders. Alcoholism can also lead to commit crimes, suicides and to increase road incidents risk.

This work is intended as a brief **research** for **health workforce** in general (e.g. doctors, psychotherapist, social workers, etc.), but also to **inform** all **adult and young people** about alcohol consumption trends.

General purpose of the investigation is to **analyze the worldwide alcohol consumption**, moving along a period **from 2010 to 2019** to obtain a global framework.



Main goals:

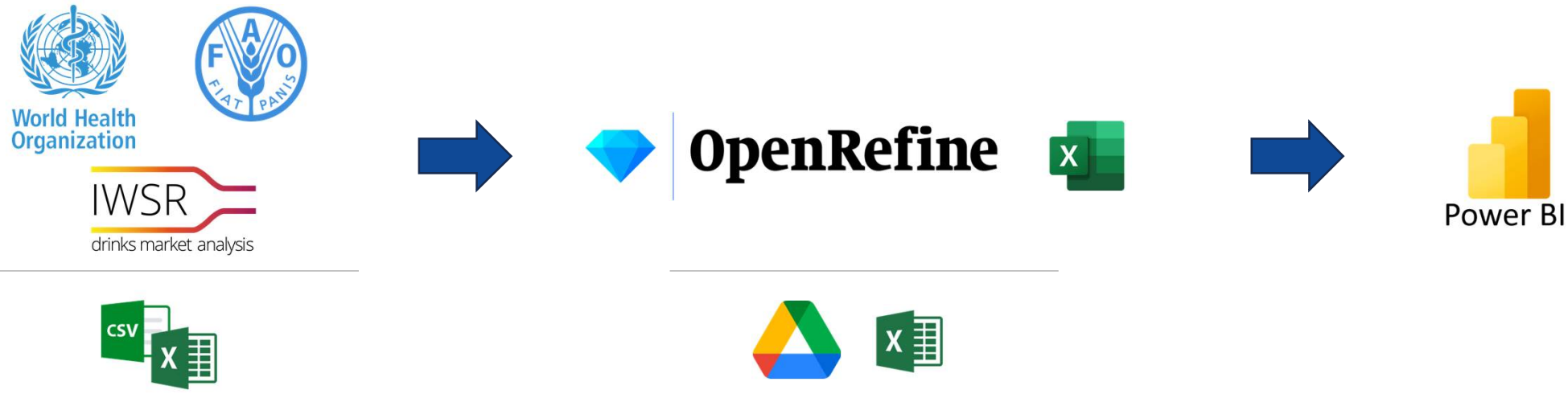
1. **Processing** raw data as to **normalize** the **dataset** to be used.
2. **Analyzing and visualizing** data to gain useful insights. Examples:
   - **Alcohol consumption per capita by country** (20 countries with highest/lowest consumption).
   - **Focus on Italy** alcohol consumption (by category and year).
3. **Validation** of the main **alcohol consumption trend** and information reported by world leading health organizations.

3. BI Architecture

Data Source

Data Profiling & ETL

Data Visualization



4. Data Source

Raw dataset has been downloaded from World Health Organization (WHO) official website.

Data gathered by WHO, Food and Agriculture Organization (FAO) in collaboration with IWSR.
(IWSR is the most trusted company/source in terms of beverage alcohol trends in the world.)

Raw dataset (xlsx, csv files) comprehends **180 countries** with data collected according to **type of alcohol beverage** for each year (from 2010 to 2019).

The measurement unit used in dataset is ***alcohol per capita consumption (APC)* – in litres of pure alcohol –** considering people above 15 years old, over a calendar year in a country.

			Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol)									
			2019	2018	2017	2016	2015	2014	2013	2012	2011	2010
Afghanistan	(FAO) 1961-2015; (FAO, IWSR) 2016-2018	All types	0.01	0.01	0.01	0.01	0	0.01	0.01	0.01	0.01	0.02
		Beer	0	0	0	0	0	0	0	0	0	0.01
		Wine	0	0	0	0	0	0	0	0	0	0
		Spirits	0.01	0.01	0.01	0.01	0	0.01	0	0	0	0
		Other alcoholic beverages	0	0	0	0	0	0	0.01	0.01	0.01	0.01
Albania	(FAO) 1962-1999; (Global Data, IWSR, OIV) 2000-2018	All types	4.4	4.44	4.39	4.38	4.33	4.1	4.28	4.43	5.03	4.88
		Beer	1.75	1.75	1.64	1.52	1.48	1.29	1.34	1.34	1.88	1.71
		Wine	1.15	1.12	1.11	1.17	1.1	0.99	1.07	1.12	1.08	1.08
		Spirits	1.43	1.5	1.56	1.61	1.68	1.74	1.79	1.88	1.99	1.97
		Other alcoholic beverages	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.09	0.09	0.11

Illustrative raw dataset extraction – source: WHO

Source (and details regarding the method of measurement)
WHO: https://apps.who.int/gho/data/node.main.A1039?lang=en
IWSR: https://www.theiwsr.com/

5. Data Profiling & ETL 1/3

Raw dataset needed to be processed as to be consumed for data analysis and visualization purposes.

1. Raw data did not present all categories (e.g. alcohol beverages, country) organized by columns.
   - A transpose operation was necessary.
   - *Row_ID* column has been added.
   - Final dataset included 1880 records.

			Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol)									
			2019	2018	2017	2016	2015	2014	2013	2012	2011	2010
Afghanistan	(FAO) 1961-2015; (FAO, IWSR) 2016-2018	All types	0.01	0.01	0.01	0.01	0	0.01	0.01	0.01	0.01	0.02
		Beer	0	0	0	0	0	0	0	0	0	0.01
		Wine	0	0	0	0	0	0	0	0	0	0
		Spirits	0.01	0.01	0.01	0.01	0	0.01	0	0	0	0
		Other alcoholic beverages	0	0	0	0	0	0	0.01	0.01	0.01	0.01
Albania	(FAO) 1962-1999; (Global Data, IWSR, OIV) 2000-2018	All types	4.4	4.44	4.39	4.38	4.33	4.1	4.28	4.43	5.03	4.88
		Beer	1.75	1.75	1.64	1.52	1.48	1.29	1.34	1.34	1.88	1.71
		Wine	1.15	1.12	1.11	1.17	1.1	0.99	1.07	1.12	1.08	1.08
		Spirits	1.43	1.5	1.56	1.61	1.68	1.74	1.79	1.88	1.99	1.97
		Other alcoholic beverages	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.09	0.09	0.11

Illustrative raw dataset extraction – source: WHO

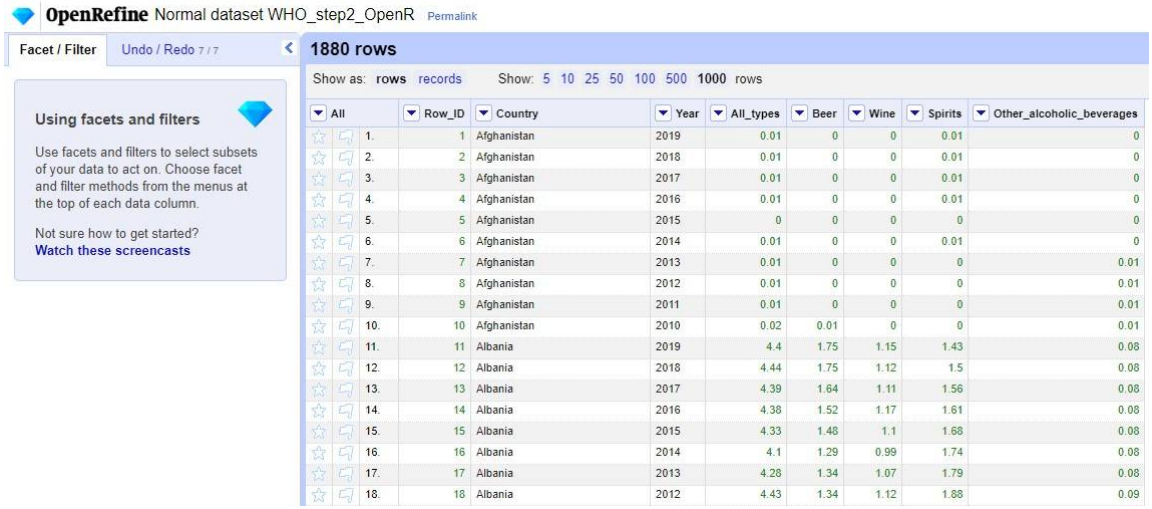
Row_ID	Country	Year	All_types	Beer	Wine	Spirits	Other_alcoholic_beverages
1	Afghanistan	2019	0.01	0	0	0.01	0
2	Afghanistan	2018	0.01	0	0	0.01	0
3	Afghanistan	2017	0.01	0	0	0.01	0
4	Afghanistan	2016	0.01	0	0	0.01	0
5	Afghanistan	2015	0	0	0	0	0
6	Afghanistan	2014	0.01	0	0	0.01	0
7	Afghanistan	2013	0.01	0	0	0	0.01
8	Afghanistan	2012	0.01	0	0	0	0.01
9	Afghanistan	2011	0.01	0	0	0	0.01
10	Afghanistan	2010	0.02	0.01	0	0	0.01
11	Albania	2019	4.4	1.75	1.15	1.43	0.08
12	Albania	2018	4.44	1.75	1.12	1.5	0.08

Illustrative normalized dataset extraction

All_types = Beer + Wine + Spirits + Other_alcoholic_beverages

5. Data Profiling & ETL 2/3

2. Normalized dataset was analyzed and adjusted by using **OpenRefine** tool:
   - Converted from text to number columns *All_types*, *Beer*, *Wine*, *Spirits*, *Other_alcoholic_beverages.*
   - Converted from number to text column *Year*.
   - Trimmed out leading and trailing whitespace.
   - Collapsed any consecutive whitespace.
   - Checked for any null value.



All	Row_ID	Country	Year	All_types	Beer	Wine	Spirits	Other_alcoholic_beverages
1.	1	Afghanistan	2019	0.01	0	0	0.01	0
2.	2	Afghanistan	2018	0.01	0	0	0.01	0
3.	3	Afghanistan	2017	0.01	0	0	0.01	0
4.	4	Afghanistan	2016	0.01	0	0	0.01	0
5.	5	Afghanistan	2015	0	0	0	0	0
6.	6	Afghanistan	2014	0.01	0	0	0.01	0
7.	7	Afghanistan	2013	0.01	0	0	0	0.01
8.	8	Afghanistan	2012	0.01	0	0	0	0.01
9.	9	Afghanistan	2011	0.01	0	0	0	0.01
10.	10	Afghanistan	2010	0.02	0.01	0	0	0.01
11.	11	Albania	2019	4.4	1.75	1.15	1.43	0.08
12.	12	Albania	2018	4.44	1.75	1.12	1.5	0.08
13.	13	Albania	2017	4.39	1.64	1.11	1.56	0.08
14.	14	Albania	2016	4.38	1.52	1.17	1.61	0.08
15.	15	Albania	2015	4.33	1.48	1.1	1.68	0.08
16.	16	Albania	2014	4.1	1.29	0.99	1.74	0.08
17.	17	Albania	2013	4.28	1.34	1.07	1.79	0.08
18.	18	Albania	2012	4.43	1.34	1.12	1.88	0.09

Dataset view by using OpenRefine

Row_ID	Country	Year	All_types	Beer	Wine	Spirits	Other_alcoholic_beverages
1	Afghanistan	2019	0,01	0	0	0,01	0
2	Afghanistan	2018	0,01	0	0	0,01	0
3	Afghanistan	2017	0,01	0	0	0,01	0
4	Afghanistan	2016	0,01	0	0	0,01	0
5	Afghanistan	2015	0	0	0	0	0
6	Afghanistan	2014	0,01	0	0	0,01	0
7	Afghanistan	2013	0,01	0	0	0	0,01
8	Afghanistan	2012	0,01	0	0	0	0,01
9	Afghanistan	2011	0,01	0	0	0	0,01
10	Afghanistan	2010	0,02	0,01	0	0	0,01
11	Albania	2019	4,4	1,75	1,15	1,43	0,08
12	Albania	2018	4,44	1,75	1,12	1,5	0,08

Final Normalized dataset extraction, after OpenRefine adjustments

5. Data Profiling & ETL 3/3

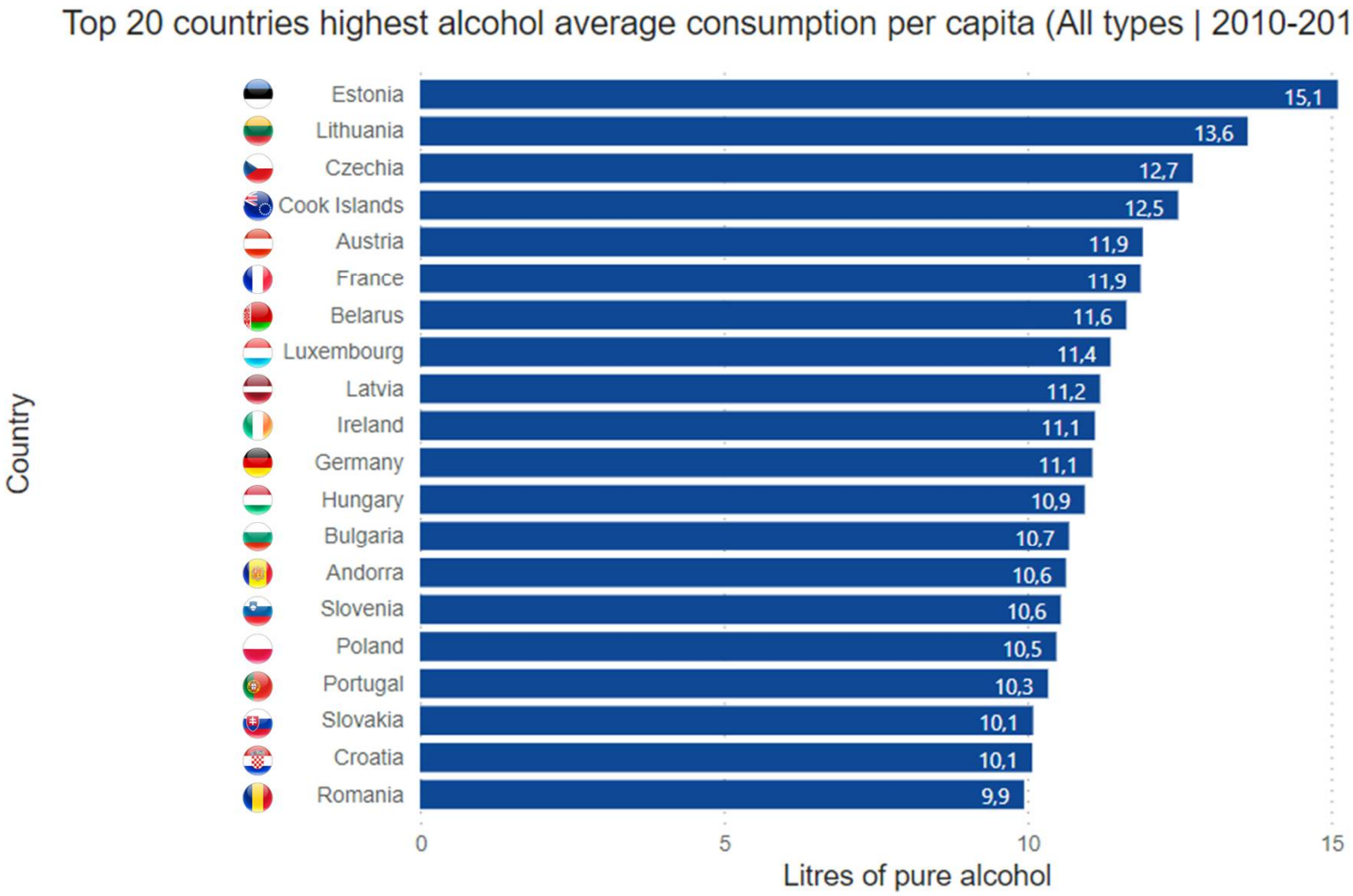
In terms of data quality, since information comes from leading authorities like WHO, FAO, IWSR, the accuracy can be assumed 100%. Also completeness and consistency in dataset result reliable.

Volatility, currency and timeliness cannot be calculated or considered, since the dataset represents a snapshot collecting information from 2010 to 2019.

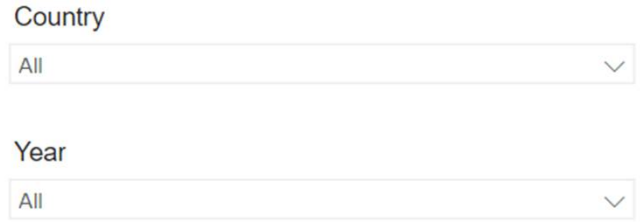
Accuracy	Completeness	Volatilty	Currency	Timeliness	Consistency

Column	Data Type	Accuracy	Completeness	Consistency
Country	String	100%	100%	100%
Year	String	100%	100%	100%
All_types	Number	100%	100%	100%
Beer	Number	100%	100%	100%
Wine	Number	100%	100%	100%
Spirits	Number	100%	100%	100%
Other_alcoholic	Number	100%	100%	100%

6. Data Visualization



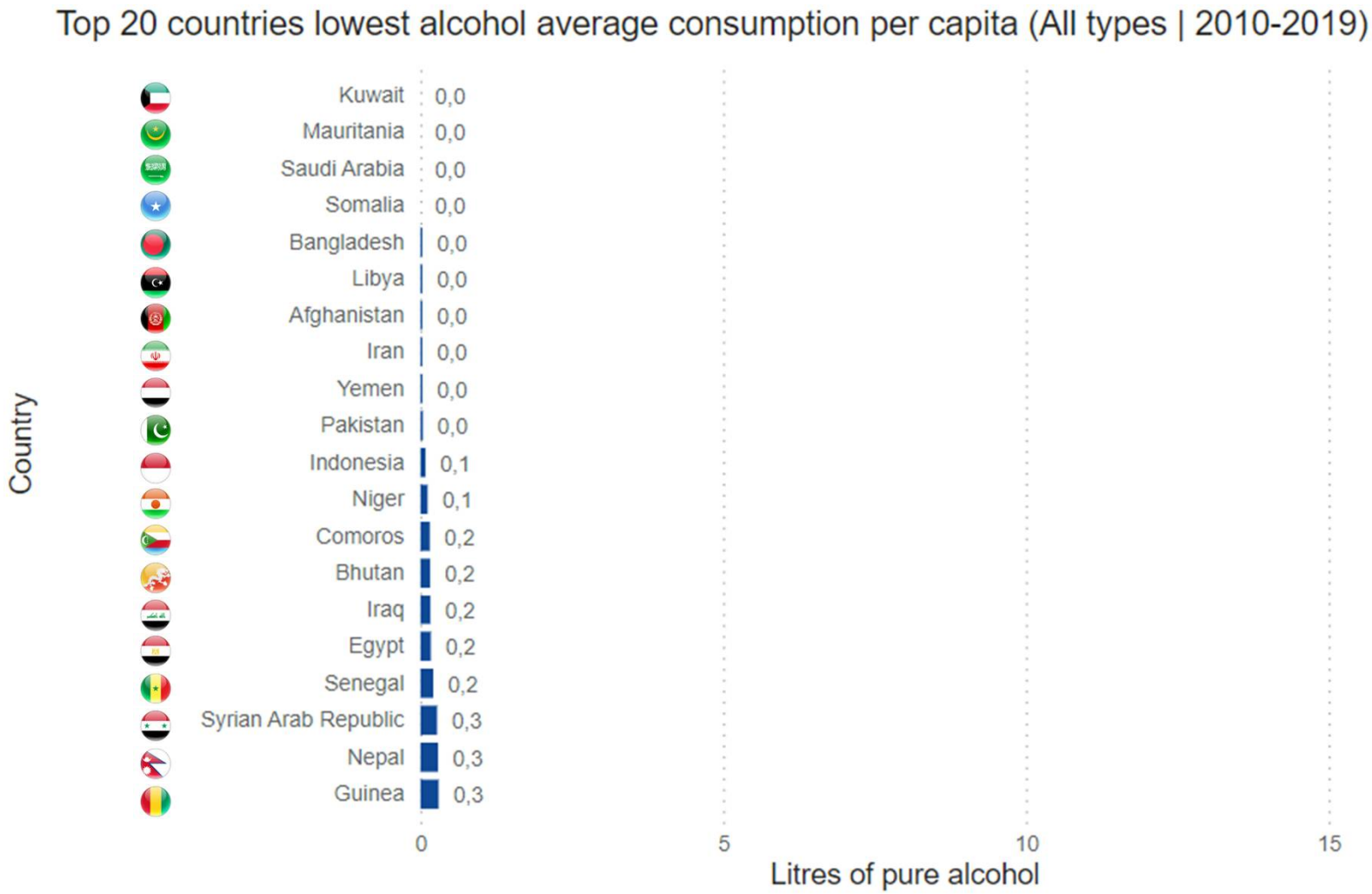
Screenshot from MS PowerBI

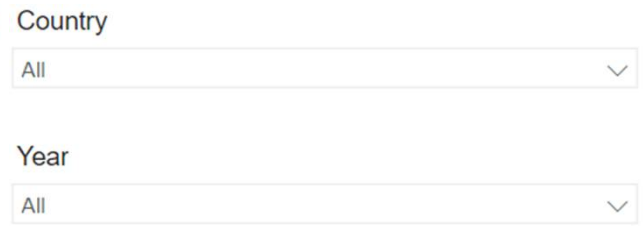


Looking at the bar chart immediately stands out how **European countries** have the **highest** levels of **alcohol consumption** per capita.

The only outlier is represented by Cook Islands.

6. Data Visualization



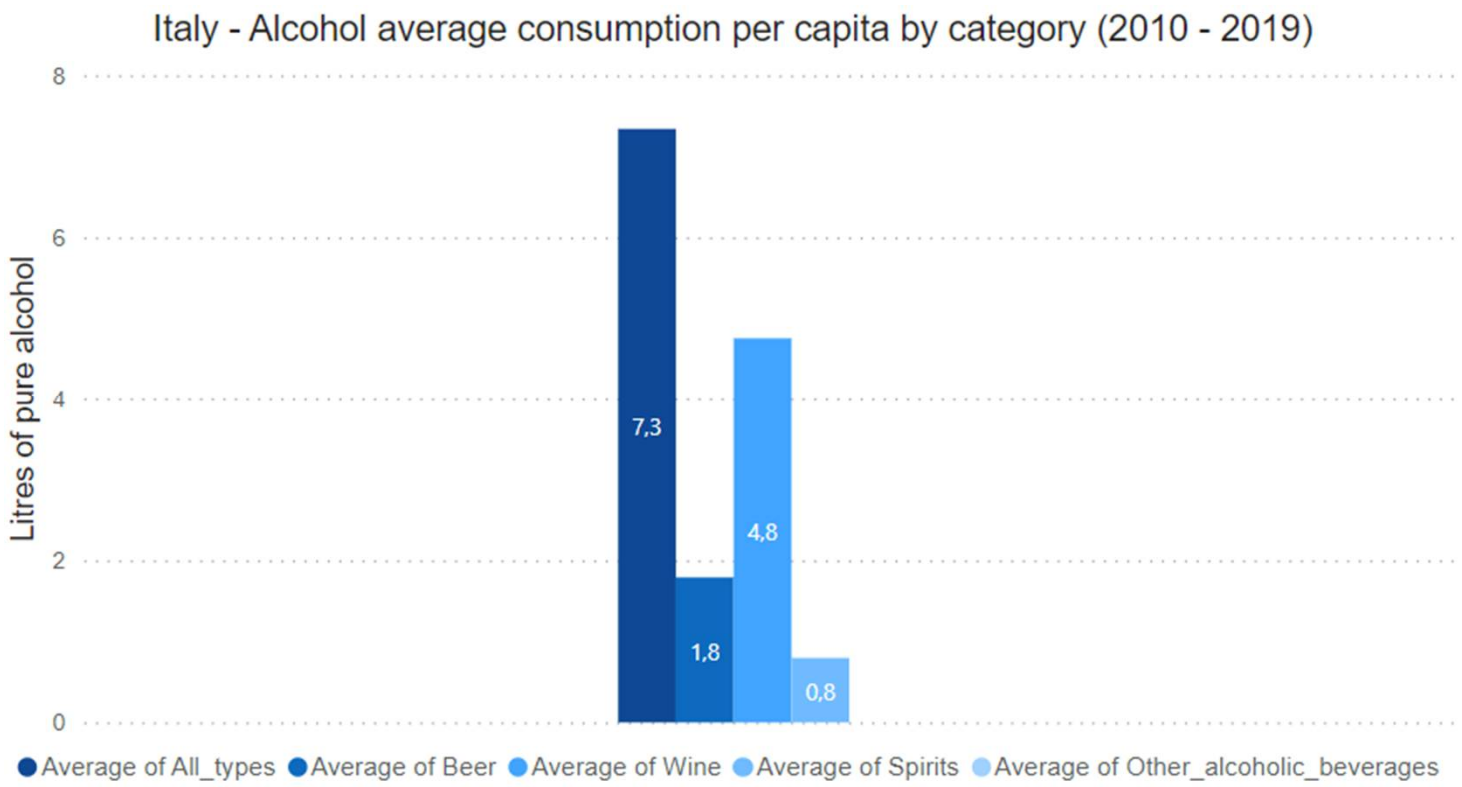


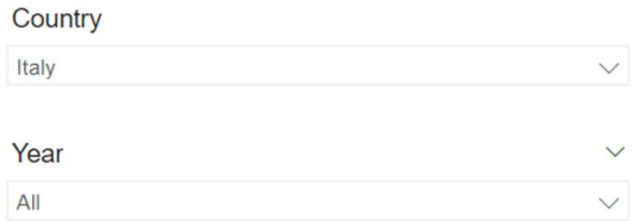
The bar chart shows that **countries** with **lowest alcohol consumption** seem to be those in which are in force hard **law** and/or **religious restrictions**.

E.g.: especially in Saudi Arabia, Kuwait, Iran, Libya, Bangladesh alcohol sales and consumption are illegal and banned. Also other Islamic countries have similar restrictions.

Screenshot from MS PowerBI

6. Data Visualization



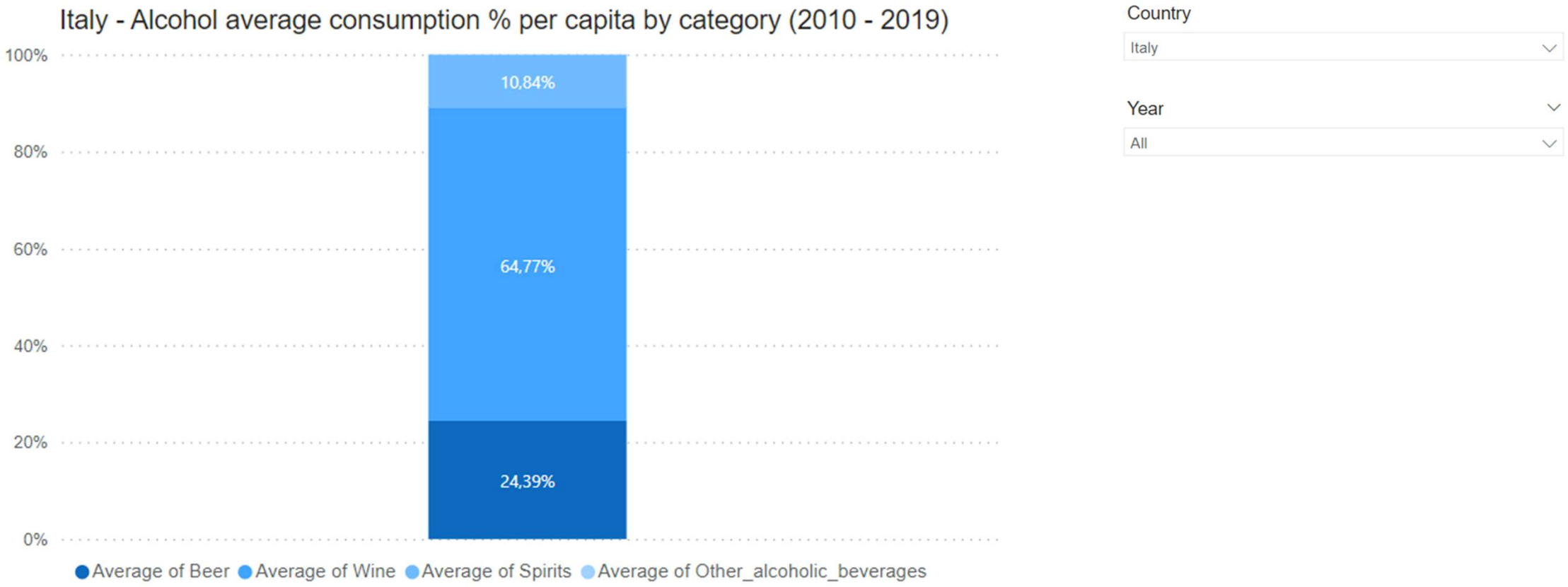


Screenshot from MS PowerBI

In **Italy** the average alcohol consumption per capita is about **7,3 litres**.

Most of the alcohol sales/use is represented by **wine** (~**4,8 litres**).

6. Data Visualization

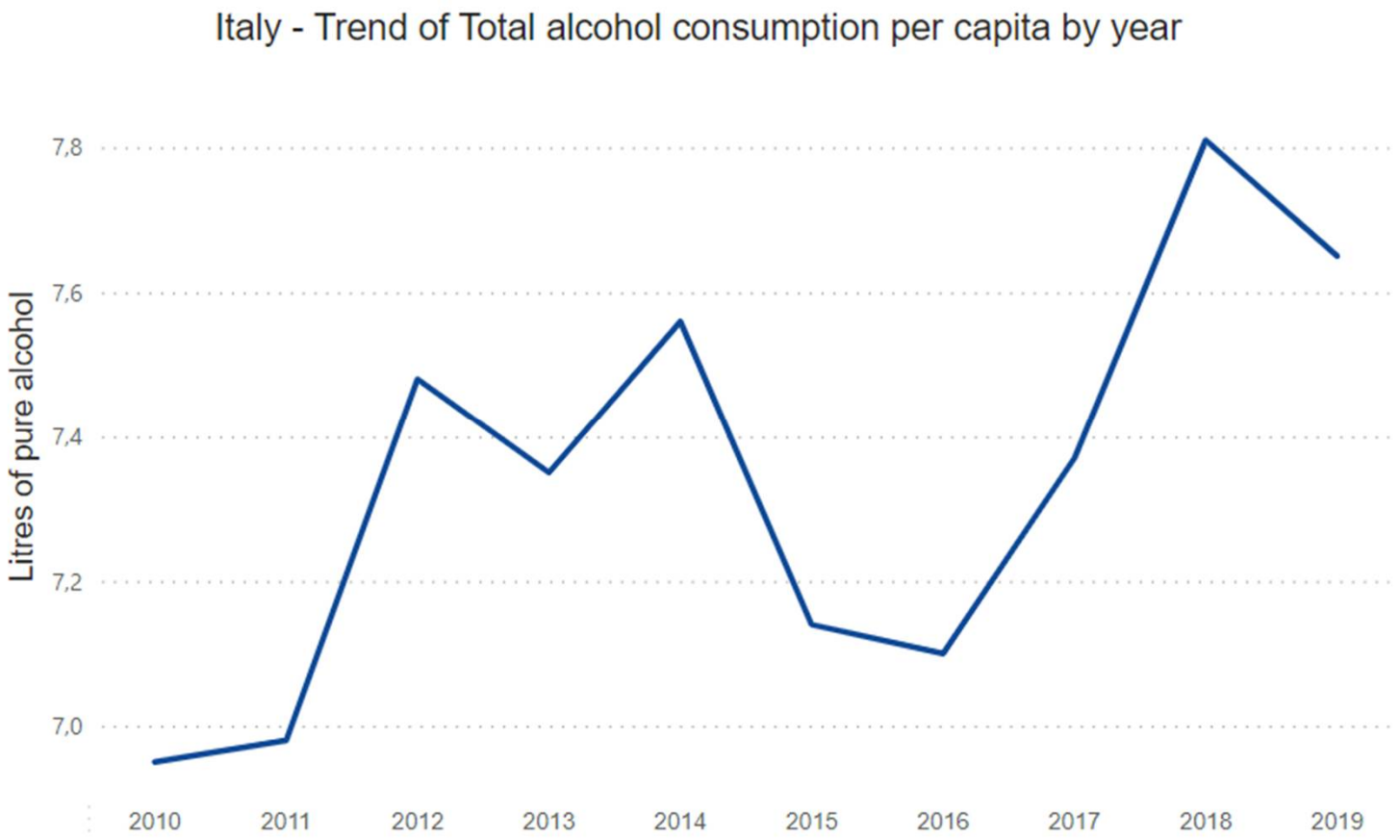


Screenshot from MS PowerBI

In Italy the alcohol beverage most consumed is definitely **wine** (**~64,8%**).

Almost 2 out of 3 people prefer wine, followed by **beer** (**~24,4%**) and **spirits** (**~10,8%**).

6. Data Visualization



Screenshot from MS PowerBI

Country

Italy

Year

All

Although in **Italy** the annual average **alcohol consumption** per capita is about 7,3 litres, the diagram shows a general **ascending trend in the last decade**.

Years 2015 and 2016 represent exceptions, but just after it is possible to notice a new alarming sharp rise for alcohol consumption.

7. Achieved Objectives

Here follow the main goals achieved in this research:

- ✓ Analyzed and visualized average alcohol consumption per capita by country (considering a time span of ten years). Found out the countries with highest/lowest alcohol consumption levels.
- ✓ Analyzed and visualized average alcohol consumption per capita in Italy. Found out Italians preference in terms of alcoholic drinks: wine (~ 2 out of 3 people).
- ✓ Verified the Italian alcohol consumption trend in the decade 2010-2019: this has generally increased over the years, especially from 2016.
- ✓ **Confirmed what reported by ISS organization: Europe presents the highest alcohol consumption per capita**. In the ranking of top 20 countries with highest alcohol consumption, 19 of them belong to European area.

8. Limits Encountered and Next Steps

- Dataset limited up to year 2019.
- Dataset limited to 180 countries.

Looking ahead:

- Extend the research from 2020 to nowadays.
- Understand how the Covid-19 pandemic (quarantine/isolation) has affected the global alcohol sales and consumption, also among different age range.
- Deepen the study to find correlation between alcohol consumption trend and physical/mental diseases, death (among adult and young people).

9. Conclusions

Working on this research the following considerations emerged:

- It is not so easy to find out good and/or enough data sources to rely on as to respond to business needs. Once found out data collected by leading authorities, this allowed to proceed with less uncertainty.

- Normalizing and manipulating a dataset was quite demanding and time consuming, but it was a necessary step in order to have set up material to work on.

- The use of a data visualization tool like MS PowerBI permitted to gain useful and interesting insights about the global alcohol consumption trends in the last decade. Also it was possible to confirm what major authorities reported: alcohol consumption and, hence, related physical/mental diseases are constantly increasing (especially in Europe).

Thanks.

Questions?

Feel free to reach out...

fabiojr.lorenzini@gmail.com

https://www.linkedin.com/in/fabiojrlorenzini/

https://github.com/FabioJrLorenzini