



Lënda: Big Data

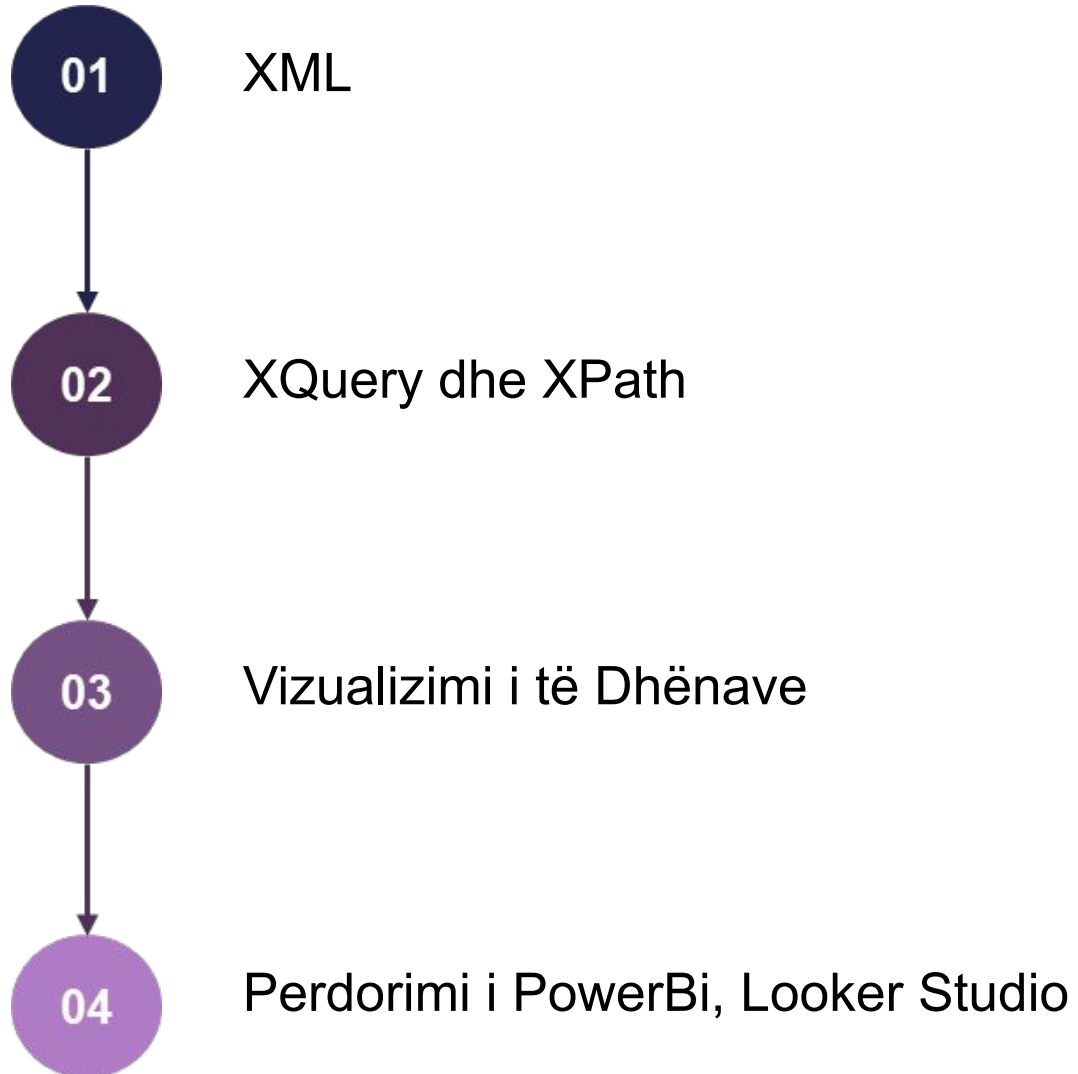
# **XML & Vizualizimi i të dhënave**

Prof. Dr. Vigan Raça  
Msc. Ass. Rafet Duriqi

Punuan:  
Endrit Kastrati  
Erza Gashi  
Jon Kuçi  
Kaltrina Krasniqi  
Mirgeta Gashi

# Përmbajtja

---



# XML

XML - përshkruan se si të ruhen të dhënat, në mënyrë strukturuar.

DTD - Per validim

```
<Lendet>
  <Lenda tipi="zgjedhore">
    <Emri>Big Data</Emri>
    <Profesor>Vigan Raca</Profesor>
    <Assistent>Rafet Duriqi</Assistent>
  </Lenda>
</Lendet>
```

## Validimi i një XML file

sh. Mondial.xml dhe Mondial.dtd

- Me anë të python..
- [Github Code](#)

```
3 class XMLValidator:
17     def validate_dtd(xmlPath, dtdPath=None):
45         print(" Unexpected error:", e)
46         return False
47
48 XMLValidator.validate_dtd("Resources/XML/mondial.xml", "Resources/XML/mondial.dtd")
49 XMLValidator.syntax_well_formed("Resources/XML/vaccinations.xml")
```

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

PORTS

Filter

```
[Running] python -u "c:\Users\jonku\Desktop\BigData_Project2\Util\XMLValidator.py"
XML is well-formed.
XML is valid against the DTD.
XML is well-formed.
```

# XML - XQuery dhe XPath

---

XQuery - filtron, rreshton, kombinon apo transformon rreshtat e XML.

- mbeshtet variable, kushtezime
- transformime dhe mundesi pafund per XML
- Ekzekutohet me ngadale

```
for $b in /books/book  
let $t := $b/title  
where $b/year > 1995  
return <bookTitle>{ $t/text() }</bookTitle>
```

XPath - filtron, navigon rreshtat e XML

- E thjeshte dhe e shpejte
- E shume perhapur ne bibloteka

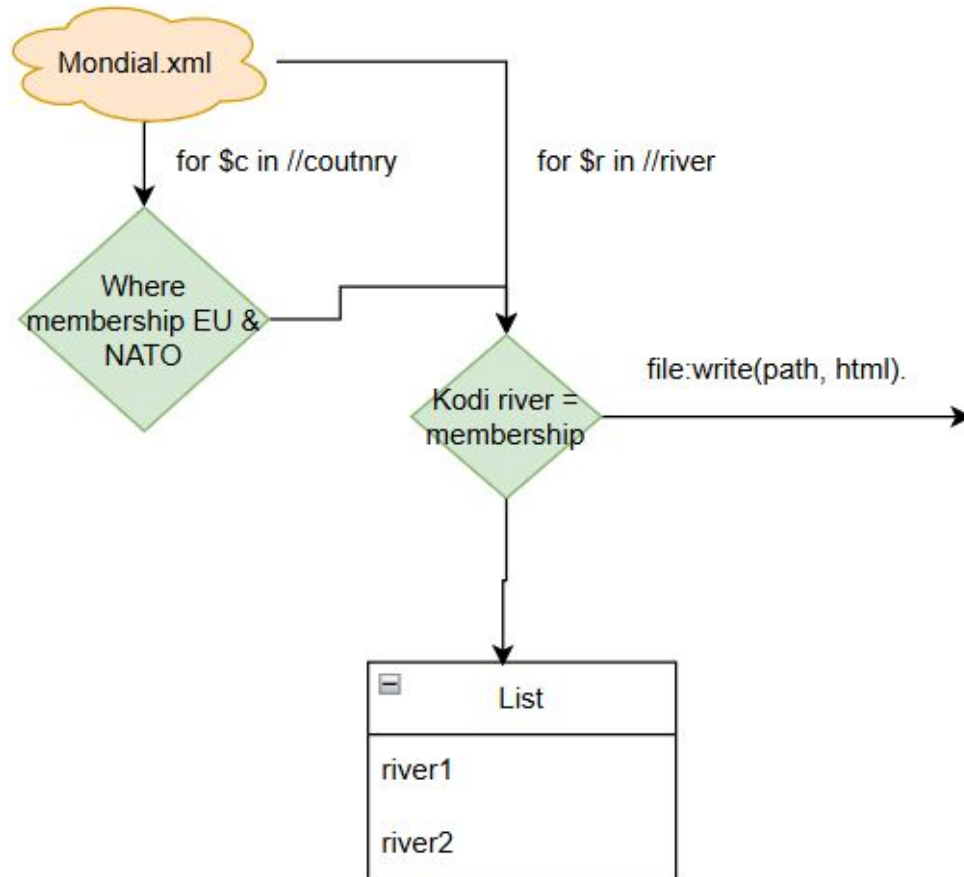
```
/library/book[year > 2000]/title
```

# XML - XQuery 1 mondial

**Qëllimi:** Gjetja e lumenjeve që kalojnë në shtetet të cilat janë anëtare të NATO dhe EU.

Hyrja: Mondial.xml

Rezultati:



## Rivers in EU and NATO Countries

- Adda
  - Italy
- Aliakmonas
  - Greece
- Aller
  - Germany
- Alz
  - Germany
- Alzette
  - Luxembourg
- Ammer
  - Germany
- Aras
  - Turkey
- Argeş
  - Romania
- Arno
  - Italy
- Arrone
  - Italy
- Bann
  - United Kingdom

# XML - XQuery 2 mondial

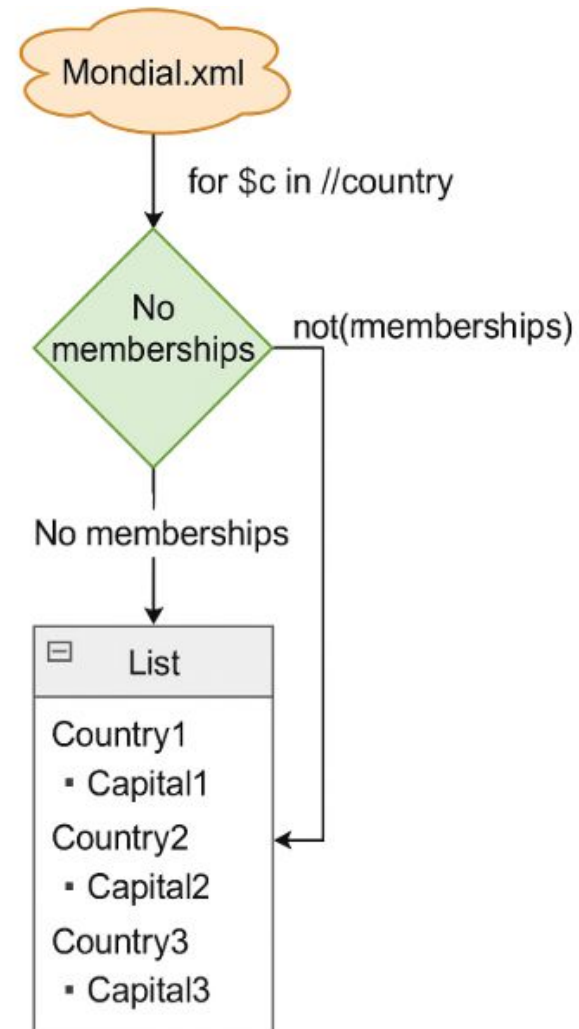
**Qellimi:** Gjetja e kryeqyteteve te shteteve qe nuk jane antare te asnje organizate boterore.

Hyrja: Mondial.xml

Rezultati:

## Countries Not in Any Organization and Their Capitals

- Bonaire
  - Capital: Kralendijk
- Ceuta
  - Capital: Ceuta
- Christmas Island
  - Capital: Flying Fish Cove
- Cocos Islands
  - Capital: West Island
- Gaza Strip
  - Capital: Unknown
- Melilla
  - Capital: Melilla
- Svalbard
  - Capital: Longyearbyen
- West Bank
  - Capital: Ramallah



# XML - XQuery - Shtesë

**Qëllimi:** Gjetja e e top 20 shteteve me te vakcinuara nga COVID 19. Popullata dhe shtete merren nga mondial, kurse numri i vakcinimeve eshte marre nga vaccinations.xml.

Dataset i Vaccinations u gjet si Json. Eshtë konvertuar në XML me python.

[Python Converter Github](#)

Hyrja: Mondial.xml dhe Vaccinations.xml

Rezultati:

**Top 20 Countries by People**

Rank	Country	Population
1	Tokelau	1499
2	Gibraltar	32577
3	Qatar	2404776
4	Brunei	393372

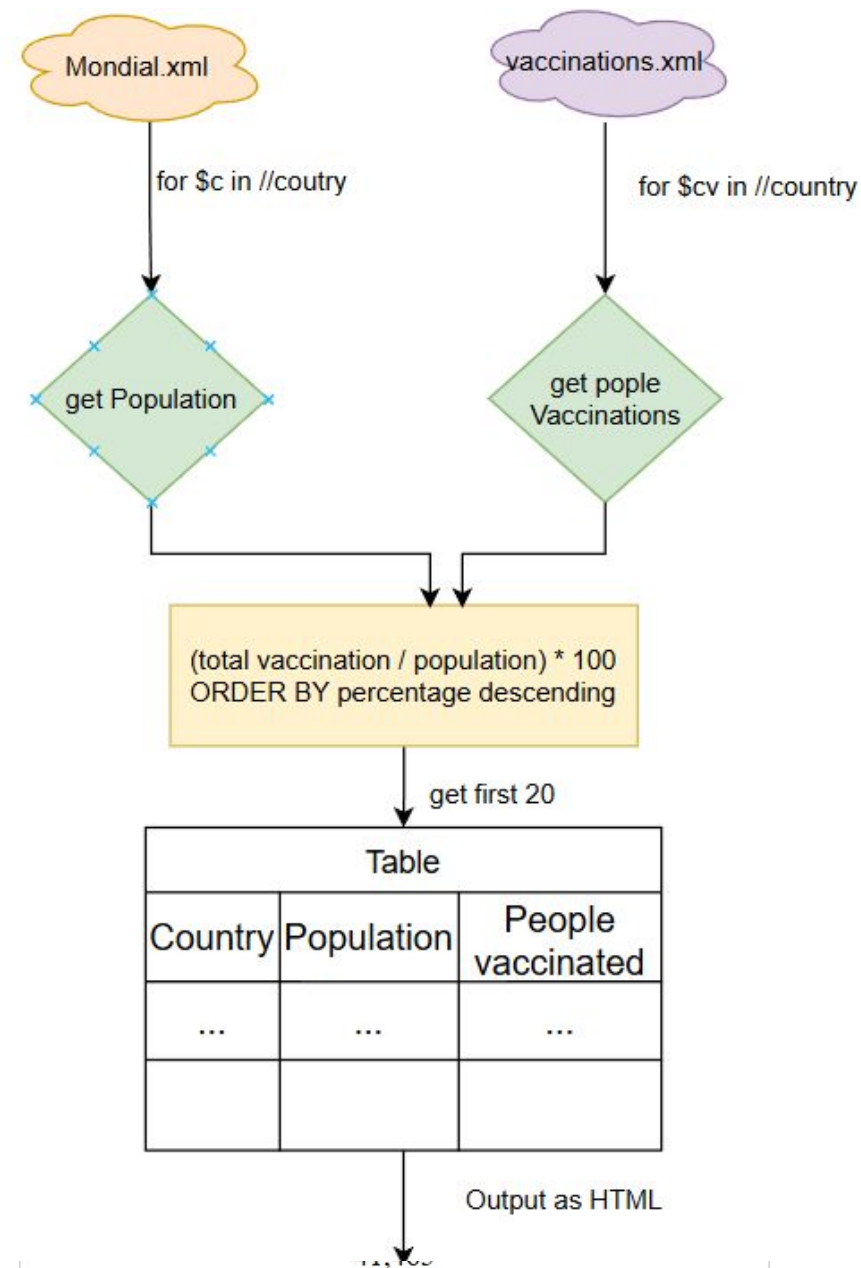


Table		
Country	Population	People vaccinated
...	...	...

Output as HTML

...

2,852,178

446,714

Vaccination %
146.96%
127.28%
118.6%
113.56%

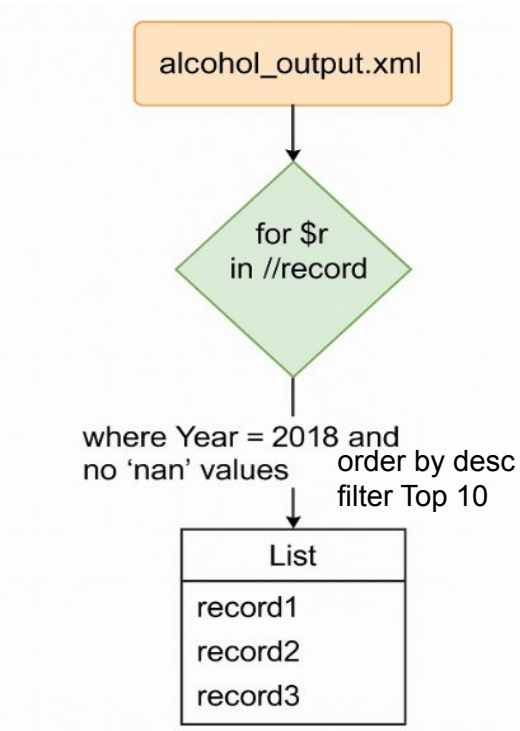


# XML - XQuery - Shtesë

**Qëllimi:** Gjetja e top 10 shteteve me konsumin më të lartë të alkoolit përgjatë vitit 2018. Output-i krijohet si një dokument HTML me tabelë të stiluar.

Dataset-in e Alcohol Consumption e gjetëm si CSV file. Është konvertuar në XML me një skriptë në python.

[Convert CSVtoXML](#)



Top 10 Countries by Alcohol Consumption in 2018

Rezultati:

Country	Continent	Alcohol Consumption (liters)	GDP per Capita	Population
Seychelles	Unknown	20.5	27,342.16	97,094
Uganda	Unknown	15.09	2,121.97	42,729,032
Czechia	Unknown	14.45	39,932.99	10,665,679
Lithuania	Unknown	13.22	35,427.46	2,801,270
Luxembourg	Unknown	12.94	113,589.67	604,244
Germany	Unknown	12.91	53,463.40	83,124,413
Ireland	Unknown	12.88	84,303.34	4,818,694
Latvia	Unknown	12.77	30,036.09	1,928,461
Spain	Unknown	12.72	40,312.55	46,692,863
Bulgaria	Unknown	12.65	22,208.45	7,051,610



# XML - XQuery - Shtesë

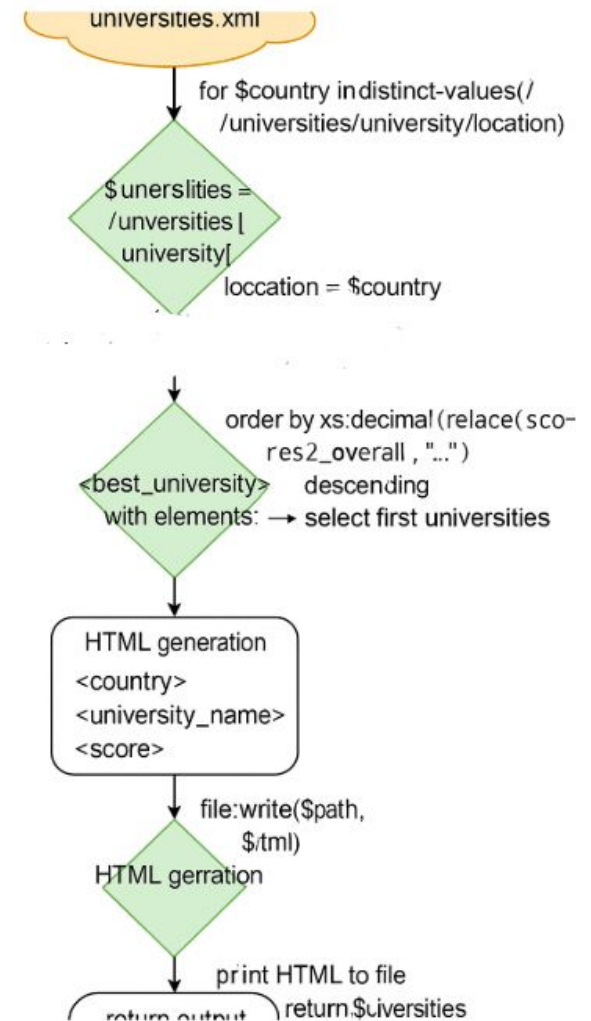
**Qëllimi:** Identifikimin dhe paraqitjen e universitetit më të mirë për secilin vend në formë të tabelës  
Output-i krijohet si një dokument HTML me tabelë të stiluar, duke paraqitur vendin , emrin e universitetit dhe total score në bazë të cilës janë krahasuar

Dataset-in e University Ranking e gjetëm si CSV file. Eshtë konvertuar në XML me një skriptë në python.

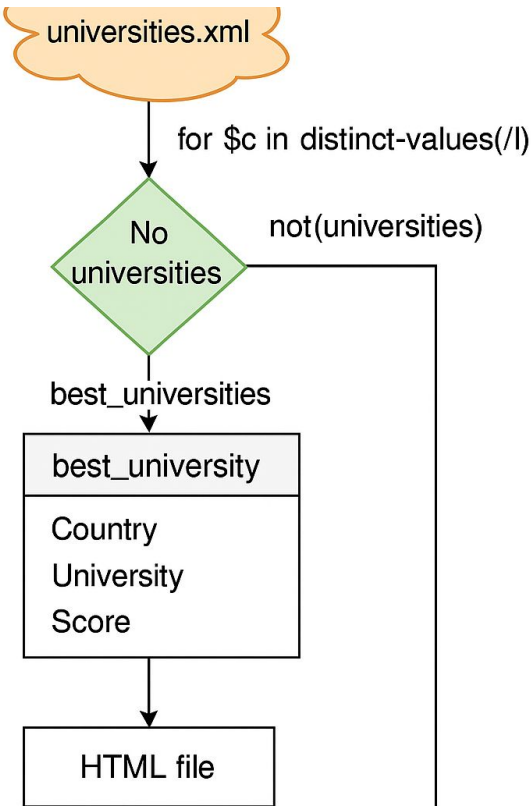
[Convert CSVtoXML](#)

Rezultati:

Rank	Country	University Name	Score
1	United States of America	Harvard University	96.1
2	United Kingdom	University of Oxford	94.2
3	Switzerland	ETH Zurich – Swiss Federal Institute of Technology Zurich	88.3
4	Canada	University of Toronto	83.9
5	Hong Kong	University of Hong Kong	79.2
6	Japan	University of Tokyo	78.3
7	South Korea	Pohang University of Science and Technology	75.1
8	Singapore	National University of Singapore	79.2
9	Australia	University of Melbourne	77.9
10	China	Peking University	72.0



# XML - XQuery - Shtesë



**Qëllimi:** Identifikimin dhe paraqitjen e universitetit më të mirë për secilin vend ne vitin 2024 me qëllim që të krahasonhen me ato të vitit 2011-2016

Output-i krijohet si një dokument HTML me tabelë të stiluar.

Nëse i krahasojmë me rezultatet të viteve 2011-2016 kemi:

- Harvard ka qenë në krye ndër vite, por në 2024 është zëvendësuar nga Stanford.
- Oxford ka kaluar nga vendi i 2-të në vendin e 1-rë me një rritje në pikë.
- ETH Zurich mbetet në të njëjtin pozicion (#3) por ka përmirësuar rezultatin.
- National University of Singapore ka pasur rritje në renditje dhe pikë

## Rezultati:

Rank	Country	University Name	Score
1	United Kingdom	University of Oxford	98.5
2	United States	Stanford University	98
3	Switzerland	ETH Zurich	93.1
4	China	Tsinghua University	92.4
5	Singapore	National University of Singapore	90
6	Canada	University of Toronto	88.6
7	Japan	The University of Tokyo	83.1
8	Germany	Technical University of Munich	82.5
9	Hong Kong	University of Hong Kong	80.3
10	Australia	University of Melbourne	79.2

# XML - XQuery - Shtesë

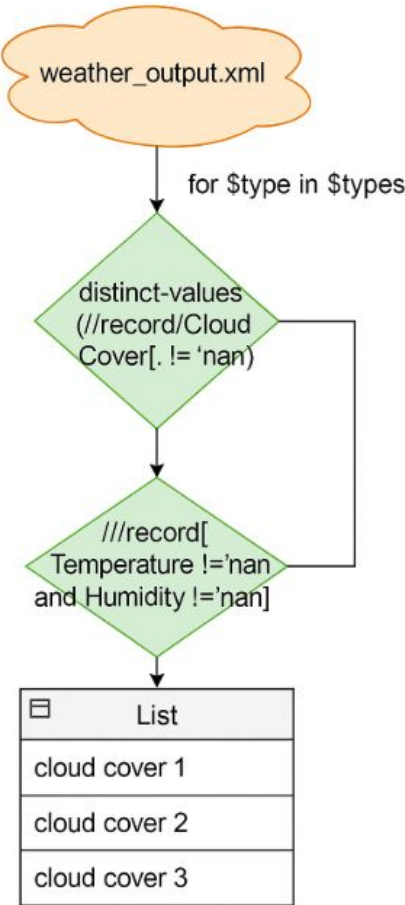
**Qëllimi:** Analiza e Mesatares së Temperaturës dhe Lagështisë sipas Llojit të Vranësirës.

Ky XQuery ka për qëllim të analizojë ndikimin e llojit të vranësirës në dy indikatorë meteorologjikë shumë të rëndësishëm:

- Temperatura mesatare (°C)
- Lagështia mesatare (%)

Dataset-i është gjetur si CSV file. Eshtë konvertuar në XML me një skriptë në python.

[Convert CSVtoXML](#)



Average Temperature and Humidity per Cloud Cover Type

Cloud Cover	Avg. Temperature (°C)	Avg. Humidity (%)
clear	34.3	52.0
partly cloudy	21.4	67.7
cloudy	13.0	49.2
overcast	12.5	76.6

Rezultati:

# XML - XPath 1 dhe 2

**Qellimi:** Gjendja e shteteve te pavarshara qe kane qene me te prekura nga sundimi i Britanise se madhe "Great Britain" duke filtruar ne gjuhen folese 'English' mbi 50%.

```
1 //country/city[
2   located_at[@watertype='sea'] and
3   number(elevation) < 10 and
4   population[last()][number(.) > 500000]
5 ]/
6 concat(..name[1]/text(), ' -> ', name[1]/text())
7
```

OK

14 Results, 307 b

Latvia -> Rīga  
Qatar -> Doha  
Yemen -> Aden  
Sri Lanka -> Colombo  
Taiwan -> Kaohsiung  
South Korea -> Changwon  
Thailand -> Bangkok  
Egypt -> Bur Said  
20 Egypt -> Al Iskandariyah  
Egypt -> As Suwayh

```
4 //country[
5   indep_date[@from='GB'] and
6   language[. = 'English' and number(@percentage) > 50]]/name/text()
7
```

OK

6 Results, 63 b

Ireland  
Australia  
New Zealand  
Canada  
United States  
Bahamas

**Qellimi:** Gjenja e qyteteve me te rrezikuara nga permbytjet duke filtruar sipas qyteteve qe kane dalse ne det, kane lartesi mbidetare me te ulet se 10 metra dhe kane popullsi mbi 500 mije banore. Kthehet rezultati sipas pattern "qyteti -> shteti".

# Vizualizimi i të Dhënave

---

**Vizualizimi i të dhënave** ndihmon në kuptimin e shpejtë dhe intuitiv të trendeve, lidhjeve dhe anomalive në dataset.

## Power BI

- Jo falas, por ofron analiza të avancuara, filtra dinamikë dhe vizualizime të detajuara.
- Ka një learning curve më të lartë, por është shumë i fuqishëm për analiza profesionale.

## Looker Studio

- Falas dhe shumë i lehtë për t'u përdorur.
- Ideale për raporte të thjeshta, shpërndarje online, dhe vizualizime të shpejta.
- E përshtatshme për përdorues pa shumë eksperiencë teknike.



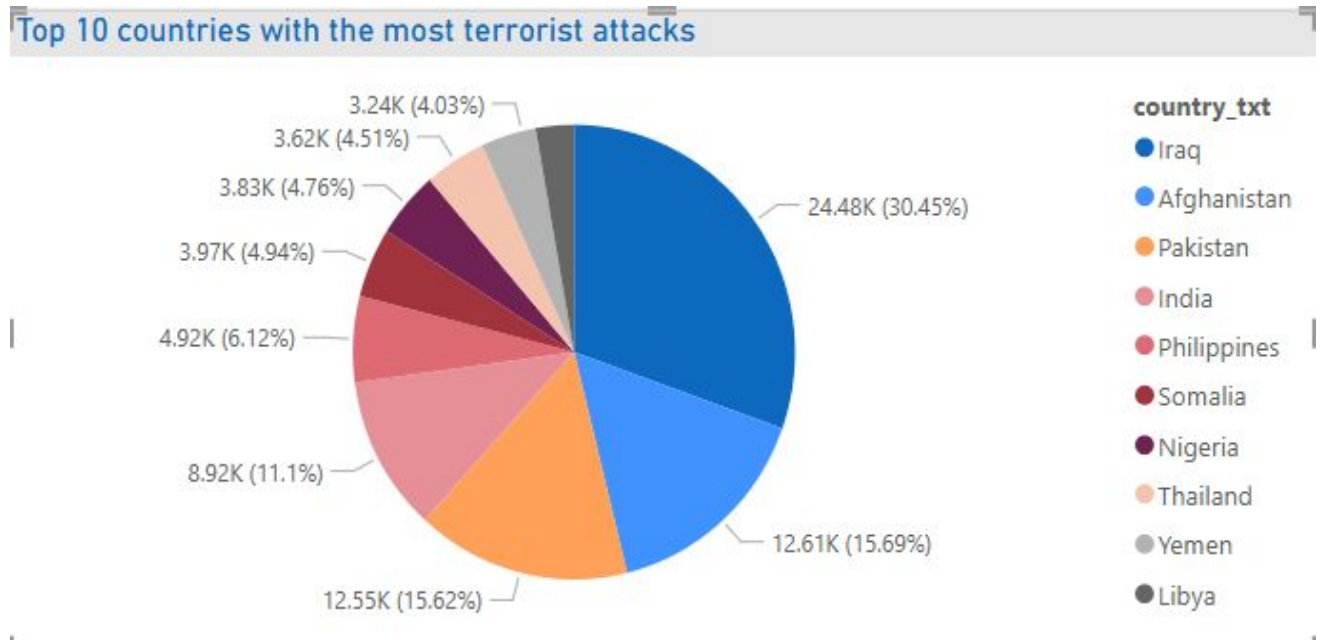
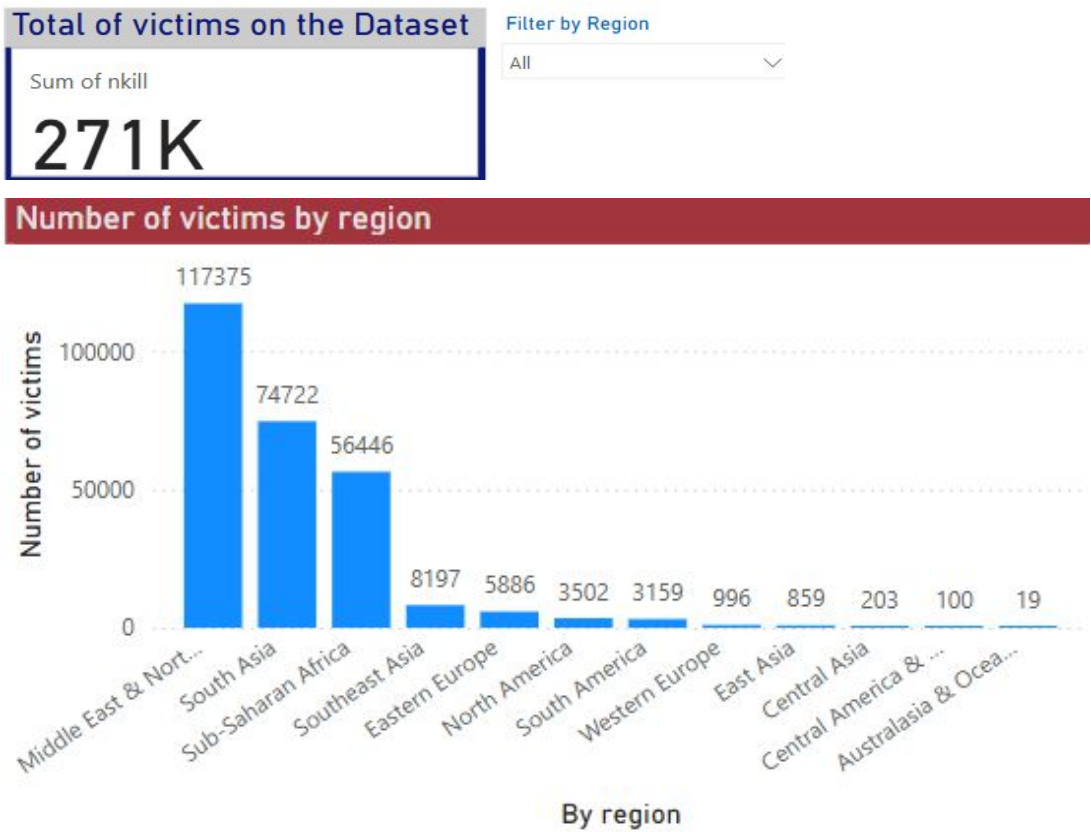


# Vizualizimi - Dataset i Terrorizmit Global [Power Bi]

**Përshkrimi:** Ky dashboard vizualizon të dhëna nga Global Terrorism Dataset për periudhën pas vitit 2000, duke analizuar ndikimin e sulmeve terroriste në nivel global.

## Key insights:

Iraku, Afganistani dhe Pakistani përbëjnë mbi 50% të sulmeve të regjistruara në top 10-she. Rajoni i Lindjes së Mesme dhe Azisë Jugore ka numrin më të lartë të viktimave.



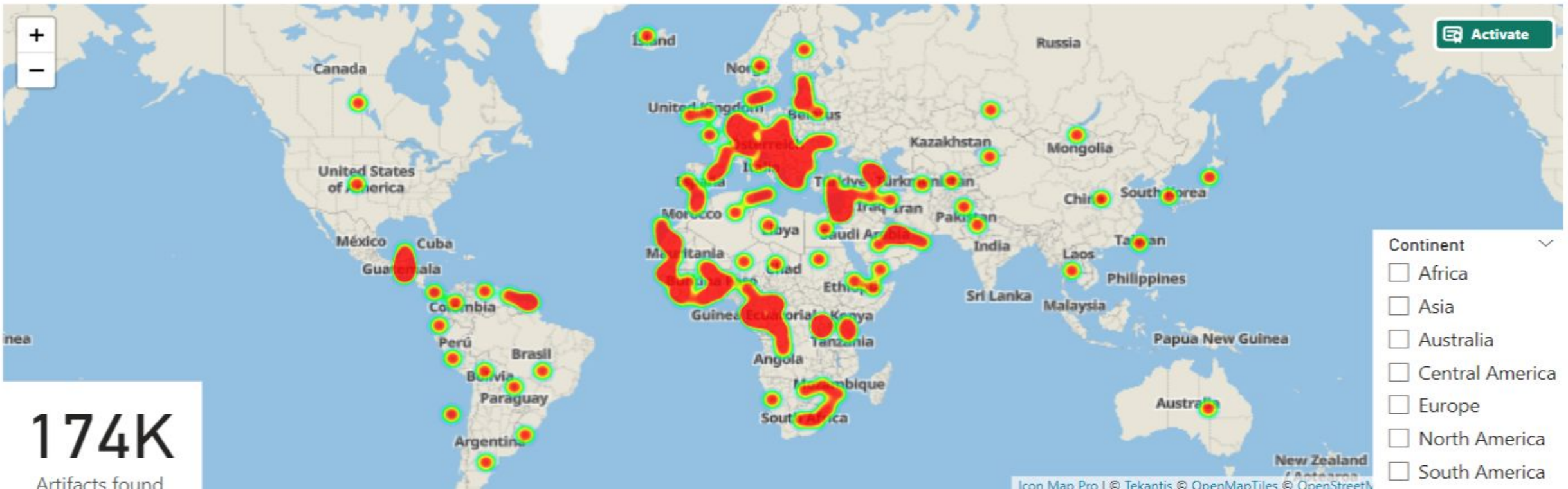
# Vizualizimi - Dataset i Gjendjeve Arkeologjike [Power Bi]

**Përshkrimi:** Pikat e nxehura në hartë përfaqësojnë vendndodhjet e gjetjeve arkeologjike të hershme të njerëzimit, të datuara me metoda radiometrike.

## Key Insights:

- Civilizimet e hershme duken në ekuador shkaki klimës.
- Shumica e gjetjeve arkeologjike janë të përqendruara në **Afrikë**, duke mbështetur teorinë e origjinës së njerëzimit në këtë kontinent

Heat Map of location of findings



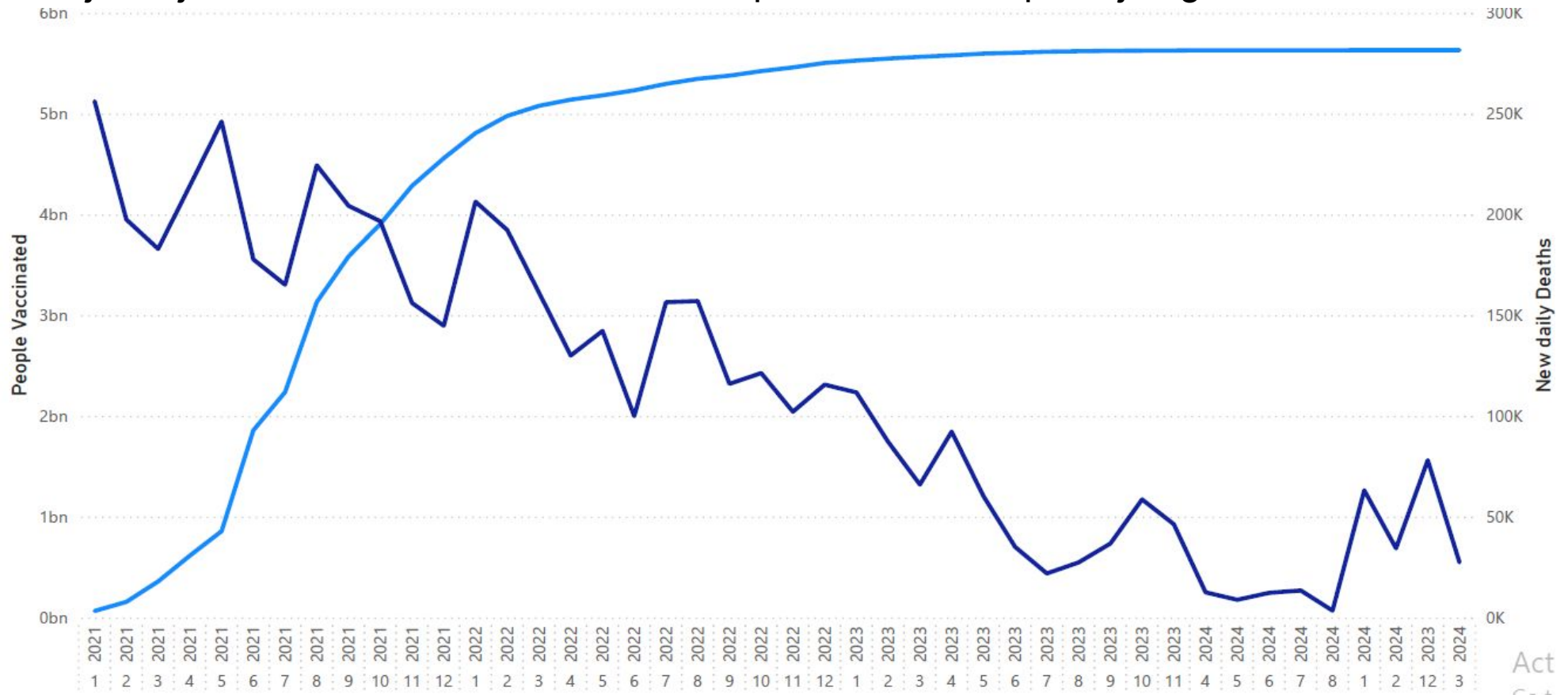


# Vizualizimi - Dataset i Covid 19 Vaksinimeve dhe Vdekjeve [Power Bi]

Nga dy databaza të ndryshme: COVID Vaccination dhe COVID Deaths.  
Krijuar view me keto kolona: month, year, new deaths, people vaccinated

**Përshkrimi:** Gjendja e marrëdhënies dhe vlerësimi i efektit të vaksinave në vdekjet mujore.

**Key insights:** Me rritjen e numrit të personave të vaksinuar, është vërejtur një ulje e numrit të vdekjeve mujore. Kjo sugjeron se arritja e një mbulimi të lartë me vaksina ka qenë thelbësore për uljen globale të vdekshmërisë nga COVID-19

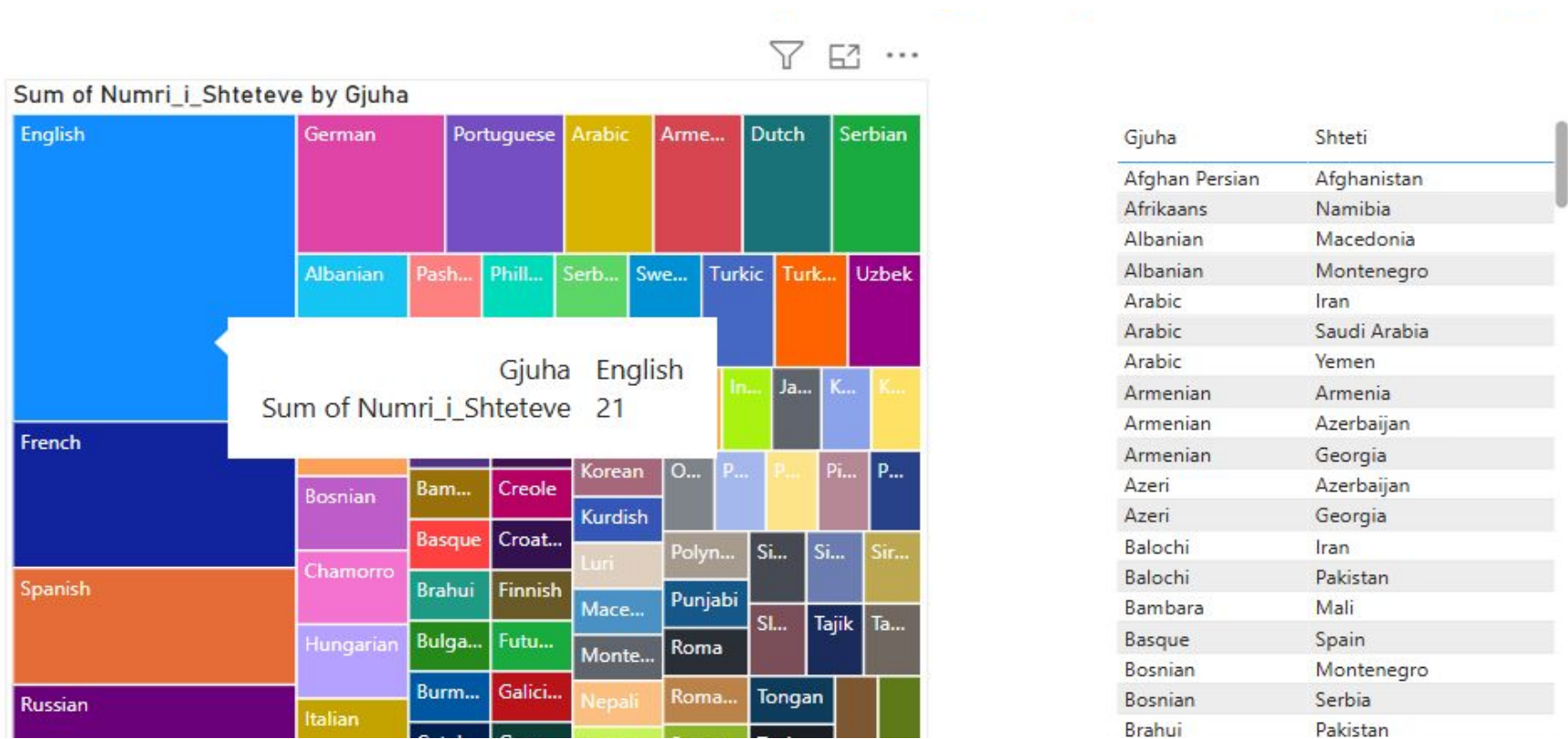


# Vizualizimi - Dataset i Gjuhesise se Shteteve [Power Bi]

Nga Mondial Databaza.

**Përshkrimi:** Vizualizimi i numrit total të shteteve që flasin gjuhën përkatëse dhe tabela Gjuhë-Shtetë.

**Key insights:** Shihet qe dukshem gjuha Angleze,Spanjolle dhe Frengje dominojn ne kete fushe si shkak i kolonizimit ne te kaluaren.



# Vizualizimi - Dataset Mondial [Power Bi]

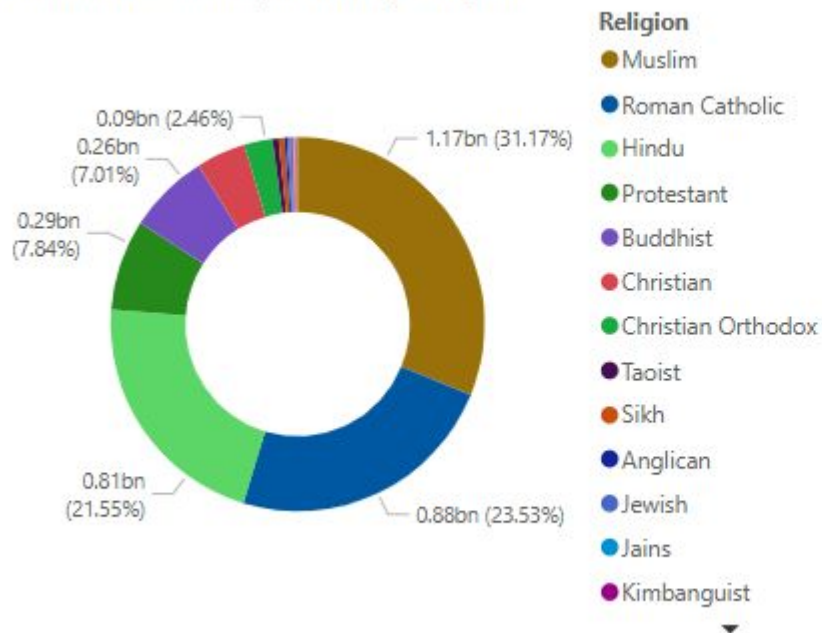
**Përshkrimi:** Në këtë vizualizim është përdorur dataset-i Mondial, e cila përmban informacione të detajuara mbi vendet e botës, përfshirë popullsinë, ekonominë, religjionet, qytetet dhe gjeografinë fizike.

## Key Insights:

- Feja me më shumë ndjekës në nivel global është Islami, e ndjekur nga Katolicizmi Romak dhe Hinduizmi. Katolicizmi dhe Islami janë të pranishëm në mbi 100 shtete secili, duke dëshmuar një dominim të gjerë global.

37  
Total Religions

Estimated Global Population by Religion



# Vizualizimi - Dataset i World University Rankings [Power Bi]

## Përshkrim:

Grafiku tregon cilësinë akademike të universiteteve në botë nga 2011 deri në 2016, me ngjyrë më të errët për vendet me cilësi më të lartë.

## Key insights:

- Universitetet në Amerikën e Veriut dhe Evropë kanë nivelin më të lartë akademik gjatë periudhës 2011–2016, me ngjyrë më të errët që tregon këtë trend.
- Rajonet si Azia dhe Australia po përmirësojnë gradualisht cilësinë e tyre akademike, ndërsa Afrika dhe pjesa më e madhe e Amerikës së Jugut kanë nivele më të ulëta akademike.



# Vizualizimi - Dataset i World University Rankings [Power Bi]

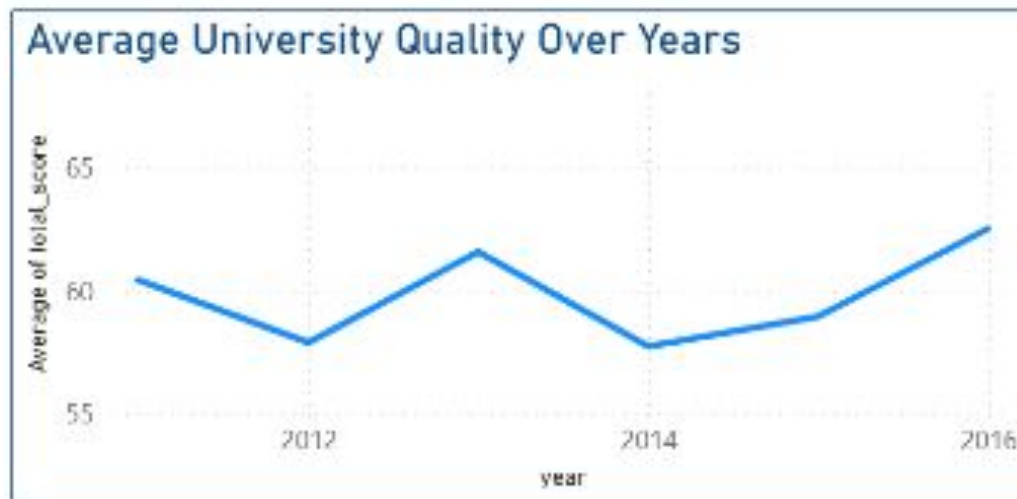
## Përshkrim:

Grafiku i parë paraqet shtetet kryesore me cilësinë më të lartë të universiteteve

Grafiku i dytë tregon evolucionin e mesatares së cilësisë universitare globale gjatë viteve 2011 deri 2016,

## Key Insights:

- Top vendet në cilësinë e universiteteve si Singapori, SHBA dhe Kina përfaqësojnë rajonet me zhvillim të fuqishëm akademik dhe teknologjik. Vlerat janë të ngushta rreth mesatares 60, që tregon konkurrencë të fortë në cilësinë universitare.
- Cilësia ka luhatje vjetore, por rritet nga rreth 58 në vitin 2011 në mbi 62 në vitin 2016. Kjo tregon përmirësim të vazhdueshëm të standardeve akademike globalisht.





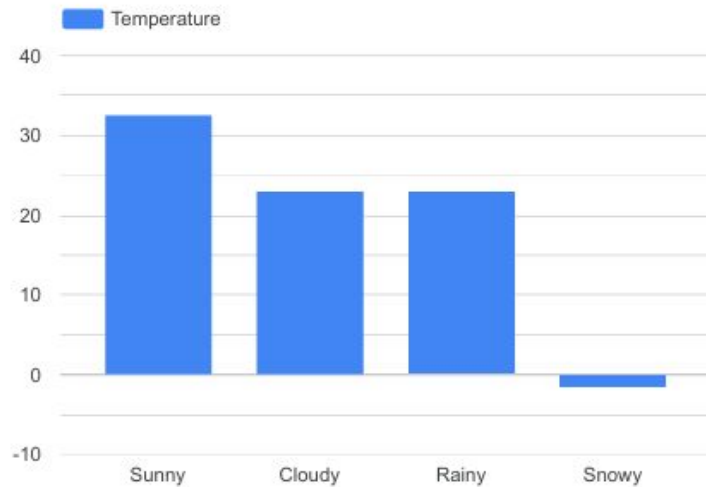
# Vizualizimi - Dataset për klasifikimin e motit [Looker Studio]

**Përshkrimi:** Në këtë vizualizim është përdorur një dataset i cili trajton klasifikimin e motit sipas tipareve klimatike si temperatura, lagështia dhe mbulimi me re.

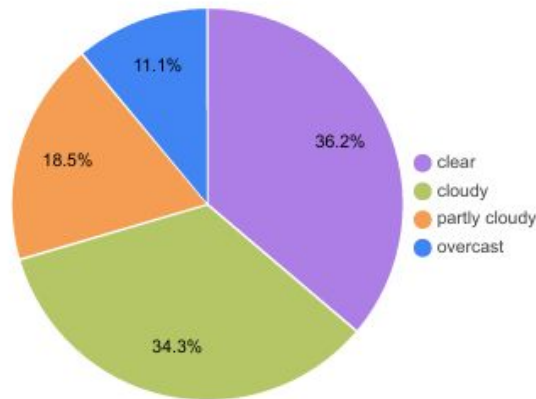
## Key Insights:

- Motet me diell kanë temperaturat më të larta mesatare ( $32.43^{\circ}\text{C}$ ), ndërsa motet me borë kanë temperaturat më të ulëta, duke rënë edhe nën  $0^{\circ}\text{C}$ . Lagështia është më e lartë gjatë motit me shi dhe borë, duke treguar një lidhje të kundërt me temperaturën.

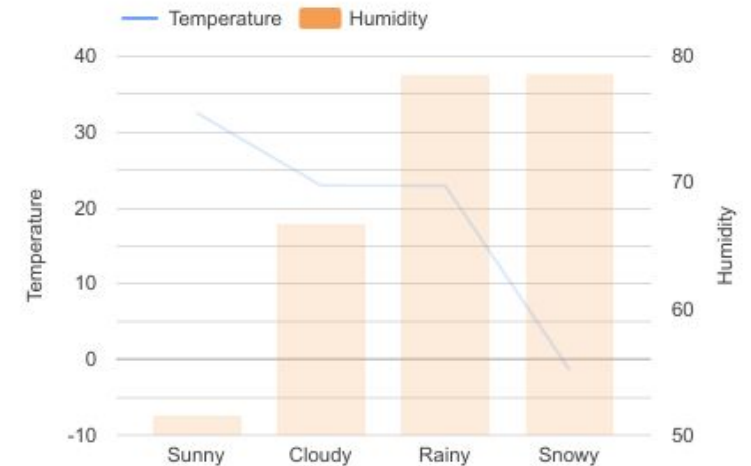
Average Temperature by Weather Type



Cloud Cover Distribution



Temperature and Humidity per Weather Type



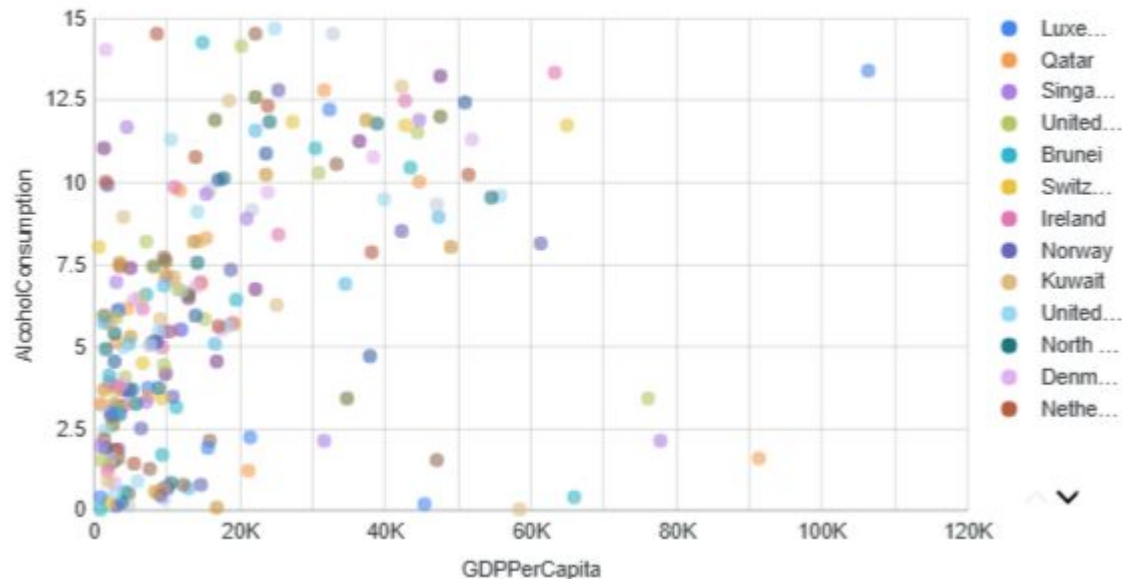
# Vizualizimi - Dataset i Konsumit të Alkoolit në nivel global [Looker Studio]

**Përshkrimi:** Në këtë vizualizim është përdorur një dataset që kombinon konsumin e alkoolit për banorë.

**Key insights:** Vendet me GDP të lartë nuk kanë domosdoshmërisht konsum më të lartë alkooli.

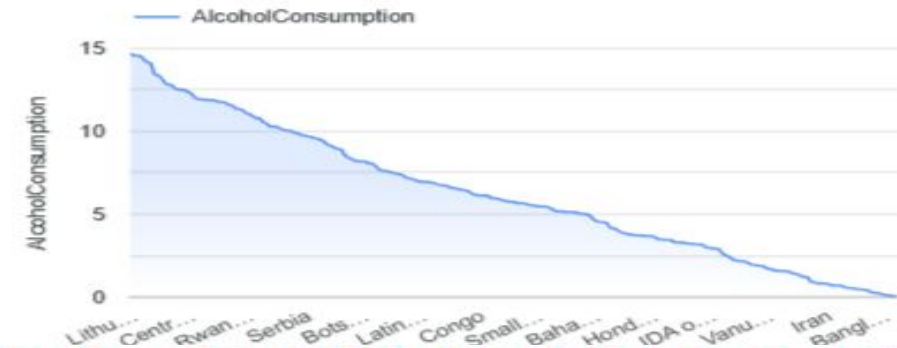
Lituania, Moldavia dhe Rusia kanë nivelin më të lartë të konsumit.

Distribution of GDP Per Capita by Alcohol Consumption

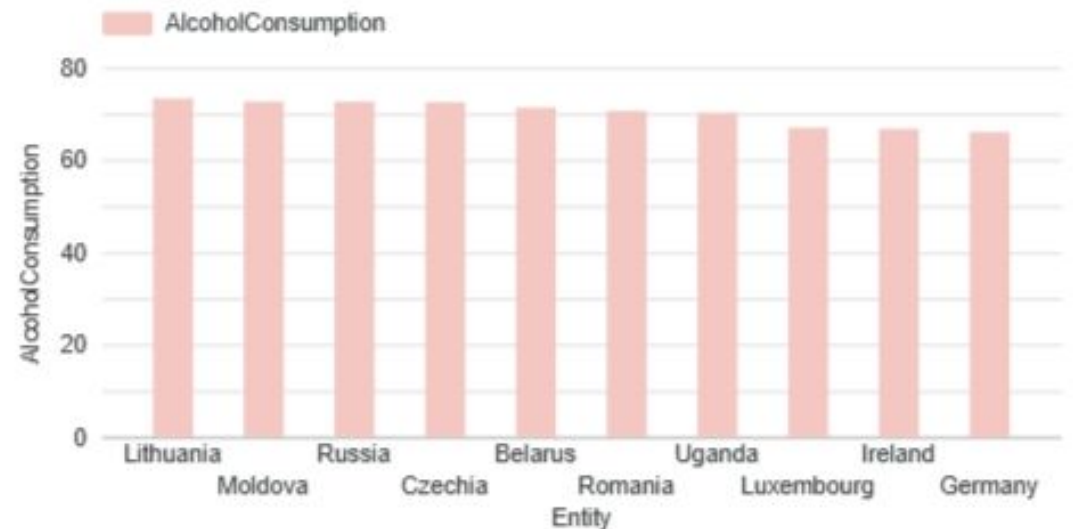


Year

Alcohol Consumption by different Countries



Top 10 countries with the highest alcohol consumption





Faleminderit për Vëmendjen!

---