

Universiteti i Prishtinës “Hasan Prishtina”

Fakulteti i Inxhinierisë Elektrike dhe Kompjuterike



Dokumentim teknik i projektit

Lënda: Big Data

Titulli i projektit: Manipulimi i të Dhënave XML dhe Vizualizimi

Emri profesorit/Asistentit	Emri & mbiemri studentëve / email adresa	
Prof. Dr. Vigan RACA MSc. Rafet DURIQI	1. Endrit Kastrati	endrit.kastrati3@student.uni-pr.edu
	2. Erza Gashi	erza.gashi11@student.uni-pr.edu
	3. Jon Kuçi	jon.kuci@student.uni-pr.edu
	4. Kaltrina Krasniqi	kaltrina.krasniqi31@student.uni-pr.edu
	5. Mirgeta Gashi	mirgeta.gashi@student.uni-pr.edu

Prishtinë, 2024

Përmbajtja

Abstrakti	3
I. Hyrje	4
II. Qëllimi i punimit	5
III. Pjesa kryesore	6
3.1. Manipulimi i të dhënave XML	6
3.1.1. Instalimi, konfigurimi dhe importimi	6
3.1.2. Validimi i dokumenteve XML dhe trajtimi i gabimeve	7
3.1.3. Përdorimi i XQuery	7
3.1.4. Përdorimi i XPath	7
3.2. Vizualizimi me mjetet BI	9
3.2.1. Zgjedhja e Mjeteve	9
3.2.2. Krijimi i Vizualizimeve me Power BI	9
3.2.2.1. Vizualizimi i Mondial DB	9
3.2.2.2. Vizualizimi i CSV Dataset-it nga faza I	9
3.2.2.3. Vizualizimi i Dataset-it me 10,000 rekorde	9
3.2.2.4. Vizualizimi i Dataset-it ...ERZA	9
3.2.3. Krijimi i Vizualizimeve me Looker Studio	10
3.2.3.1. Vizualizimi i Dataset-it	10
3.2.3.2. Vizualizimi i Dataset-it	10
IV. Sfidat dhe Zgjidhjet	11
V. Konkluzione	12
Referencat	13

Abstrakti

Ky projekt integron dy komponentë të rëndësishëm të teknologjisë së të dhënave: manipulimin e të dhënave XML dhe ndërtimin e vizualizimeve përmes mjeteve të Business Intelligence (BI). Në pjesën e parë është realizuar analiza dhe përpunimi i dokumenteve XML duke përdorur gjuhët XQuery dhe XPath. Janë zhvilluar query të personalizuar mbi skemën e Mondial DB, duke u fokusuar në nxjerrjen e të dhënave relevante si vendet, qytetet, popullsia, dhe marrëdhëniet mes tyre. Për validimin e skemave XML janë përdorur mjete si xmllint dhe BaseX, duke siguruar integritetin dhe strukturën korrekte të dokumenteve.

Në pjesën e dytë të projektit, janë përdorur mjete BI si Power BI dhe Looker Studio për të ndërtuar dashboards interaktive mbi disa burime të të dhënave, përfshirë dataset-in Mondial, një CSV të importuar, një dataset me mijëra rekorde dhe query e view të nxjerra nga to.

Janë krijuar vizualizime të ndryshme garfike duke theksuar interaktivitetin përmes komponentëve dhe parametrave tjerë. Projekti ka ofruar një përvojë të vlefshme praktike në përpunimin e të dhënave dhe vizualizimin analitik, duke thelluar njohuritë mbi transformimin e të dhënave të strukturuar në informacion të përdorshëm për vendimmarrje.

I. Hyrje

Në epokën moderne të teknologjisë së informacionit, përpunimi dhe vizualizimi i të dhënave janë kthyer në komponentë kyç në marrjen e vendimeve të mençura dhe logjike.

Pavarësisht zhvillimeve të fundit drejt formateve më të lehta si JSON, XML (eXtensible Markup Language) vazhdon të mbetet një standard i rëndësishëm për shkëmbimin dhe strukturimin e të dhënave në shumë sisteme ekzistuese dhe institucione ndërkombëtare. Struktura e tij e rreptë dhe e mirëorganizuar e bën të përshtatshëm për ruajtjen e të dhënave komplekse hierarkike, sidomos në skenarë që kërkojnë validim formal.

Në të njëjtën kohë, Business Intelligence (BI) është bërë thelbësore për shndërrimin e të dhënave të papërpunuara në informacione të dobishme, që ndihmojnë organizatat të identifikojnë trende, zbulojnë anomali dhe marrin vendime të bazuara në analiza të sakta. Mjetet si Power BI dhe Looker Studio ofrojnë mundësi të avancuara për vizualizimin interaktiv të të dhënave nga burime të ndryshme.

II. Qëllimi i punimit

Qëllimi i këtij projekti është të integrojë përpunimin e të dhënave XML dhe vizualizimin e tyre përmes mjeteve të vizualizimit si PowerBI për të ndërtuar një zgjidhje të plotë që shndërron të dhënat e strukturuar në informacione të dobishme për vendimmarrje.

Në pjesën e parë, projekti synon të zhvillojë aftësi për manipulimin e dokumenteve XML duke përdorur XQuery dhe XPath për të nxjerrë të dhëna relevante nga baza e të dhënave Mondial, si dhe të sigurojë validimin e saktë të këtyre dokumenteve përmes mjeteve si BaseX.

Në fazën e dytë, projekti fokusohet në krijimin e vizualizimeve interaktive dhe analitike duke përdorur Power BI dhe Looker Studio, të cilat mundësojnë paraqitjen e të dhënave nga burime të ndryshme (Mondial DB, CSV, dhe dataset-e me mijëra rekorde) në formë dashboard-esh tërheqës dhe funksionalë. Përmes kësaj qasjeje, projekti synon të demonstrojë aftësitë praktike në përpunimin dhe analizën e të dhënave, duke kontribuar në zhvillimin e njohurive për teknologjitë moderne të informacionit dhe vendimmarrjen e bazuar në të dhëna.

III. Pjesa kryesore

3.1. Manipulimi i të dhënave XML

3.1.1. Instalimi, konfigurimi dhe importimi

3.1.1.1. Instalimi dhe konfigurimi i tool-it të përzgjedhur (BaseX)

Për përpunimin e të dhënave XML, është përdorur BaseX, një platformë e fuqishme dhe e lehtë për menaxhimin dhe analizën e dokumenteve XML. Procesi i instalimit dhe konfigurimit të BaseX është kryer si vijon:

Shkarkimi i BaseX:

- Nga faqja zyrtare e BaseX (<http://basex.org>).
- Zgjedhim versioni më të fundit të BaseX (p.sh., BaseX 10.7) për sistemin operativ të përdorur (Windows/Linux/Mac).

Konfigurimi:

- Pas instalimit, hapim UI e BaseX
- Është verifikuar konfigurimi bazë, duke siguruar që Java Runtime Environment (JRE) të jetë i instaluar, pasi BaseX kërkon Java për të funksionuar.
- Janë vendosur parametrat fillestarë për menaxhimin e bazave të të dhënave XML, si madhësia e memorjes dhe formatet e input/output.

BaseX u konfigurua për të mbështetur query-t XQuery dhe XPath, duke mundësuar importimin dhe analizën e skemës Mondial XML në fazat e mëvonshme të projektit.

3.1.1.2. Importimi i skemës Mondial XML

Skema e Mondial XML është shkarkuar nga burimi zyrtar ([Mondial XML](#)).

Në BaseX, është importuar përmes opsionit "Create Database", pastaj duke ngarkuar fajllin XML (mondial.xml). Struktura e të dhënave është verifikuar për integritet, duke mundësuar përpunimin e mëtejshëm me XQuery dhe XPath.

3.1.2. Validimi i dokumenteve XML dhe trajtimi i gabimeve

3.1.2.1. Validimi përmes python

Për të garantuar saktësinë dhe përputhshmërinë strukturore të dokumenteve XML të përdorura në projekt, është zhvilluar një skriptë e personalizuar në gjuhën Python. [Kodi në Github](#)

Skripta është ndërtuar duke përdorur bibliotekën “lxml.tree” dhe përfshin dy funksione kryesore:

- `syntax_well_formed(xmlPath)`: Kontrollon nëse dokumenti është *well-formed* (i formuar në mënyrë të saktë sipas rregullave të sintaksës XML).
- `validate_dtd(xmlPath, dtdPath)`: Validon përmbajtjen dhe strukturën e XML duke e krahasuar me një fajll DTD të caktuar ose me DTD-në inline në vetë dokumentin XML.

Të gjitha XML file të cilat janë përdorur përgjatë projektit janë validuar me në sintaksë dhe kanë rezultuar të jenë pa gabime.

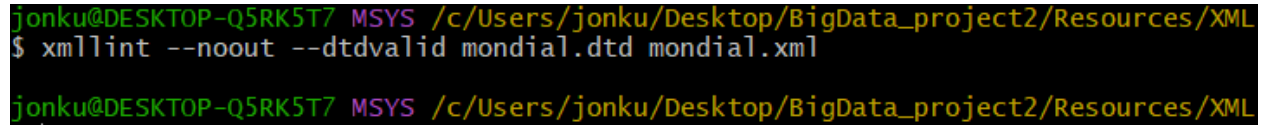
Për “mondial.dtd” është shkarkuar Document Type Definition (DTD) nga linku: [Modial.dtd](#)

XML file i mondial është krahasuar se a i përmbahet strukturës së definuar (dtd) dhe validimi është bërë me sukses duke mos pasur asnjë gabim.

3.1.2.2. Validimi përmes xmlini

Projekti gjithashtu është validuar ndërmjet xmlini, një mjet komandash i cili vjen me bibliotekën libxml2. Megjithëse xmllint nuk është i disponueshëm natyrshëm në Windows, përmes instalimit të ambientit MSYS2 u mundësua përdorimi i tij në një mjedis të ngjashëm me Linux. Pas instalimit të MSYS2 dhe ekzekutimit të komandës “pacman -S libxml2”, komanda xmllint u bë e përdorshme nga terminali.

Komanda e përdorur për të kryer validimin ishte:



```
jonku@DESKTOP-Q5RK5T7 MSYS /c/Users/jonku/Desktop/BigData_project2/Resources/XML
$ xmllint --noout --dtdvalid mondial.dtd mondial.xml
jonku@DESKTOP-Q5RK5T7 MSYS /c/Users/jonku/Desktop/BigData_project2/Resources/XML
```

Figura 1. Ekzekutimi i komandës për validim XML dhe rezultati bosh që është sukses

Kjo komandë kontrollon dokumentin XML duke e krahasuar me fajllin DTD (mondial.dtd). Nëse dokumenti nuk përmban gabime, xmllint nuk shfaq asnjë mesazh, duke nënkuptuar që është valid. Në rast gabimi, vegla ofron mesazhe të detajuara të cilat ndihmojnë në lokalizimin dhe korrigjimin e problemeve strukturore.

Kjo qasje siguroi që dokumenti të ishte sintaksisht i formuar mirë dhe strukturisht valid, përpara se të importohet në BaseX për analizë të mëtejshme me XQuery dhe XPath.

3.1.3. Përdorimi i XQuery

XQuery (XML Query Language) është një gjuhë standarde e krijuar për të kërkuar, manipuluar dhe transformuar të dhëna në formatin XML. Është e ngjashme në qëllim me SQL (që përdoret për baza të dhënash relacionale), por e përshtatur për strukturën hierarkike të XML. Ajo i mundëson përdoruesit të eksplorojë dokumentet XML në mënyrë fleksibile dhe të gjenerojë rezultatet në formate të ndryshme (XML, HTML, tekst).

XQuery mbështet operatione si filtrimi, renditja, ndërtimi i elementeve të rinj, si dhe bashkimi i të dhënave nga burime të ndryshme XML. Skriptat XQuery zakonisht ruhen në fajlle me prapashtesën “.xquery”.

3.1.3.1. XQuery 1: Kryeqytetet e shteteve jo-anëtarë të organizatave botërore nga Mondial

Qëllimi dhe struktura: Kjo query është ndërtuar me synimin për të identifikuar shtetet që nuk janë anëtare të asnjë organizate ndërkombëtare, pra që nuk kanë të definuar atributin @memberships. Për secilin nga këto shtete, nxirret edhe kryeqyteti përkatës përmes lidhjes me elementin capital. Rezultatet renditen alfabetikisht sipas emrit të shtetit për të mundësuar një paraqitje të strukturuar dhe të lehtë për analizim.

Hyerje: Dokumenti XML: mondial.xml, i ngarkuar në ambientin BaseX.

Dalje:

- Një strukturë XML e personalizuar me root <non_member_countries>, e cila përmban emrat e shteteve pa anëtarësi dhe kryeqytetet e tyre.
- Gjenerimi i një file HTML (xquery2_result.html) që paraqet në mënyrë vizuale listën e shteteve të filtruar bashkë me kryeqytetet përkatëse.

Shembull rezultati

```
<non_member_countries>
  <country>
    <name>Bonaire</name>
    <capital>Kralendijk</capital>
  </country>
  ....
</non_member_countries>
```

Lokacioni i kodit: GitHub: [XQuery 1](#)

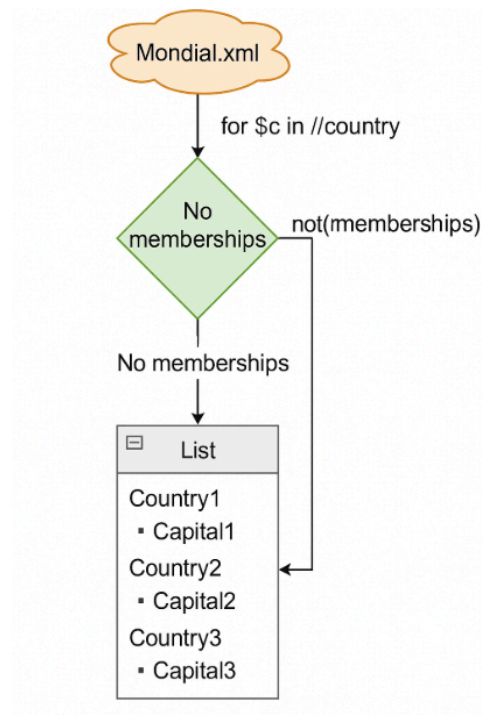


Figura 2. Diagrami logjik i ekzekutimit të XQuery1

3.1.3.2. XQuery 2: Lumenjtë në shtetet anëtare të NATO-s dhe BE-së (pa Suedinë/Francën) Mondial

Qëllimi dhe struktura: Kjo query është ndërtuar me qëllim që të identifikojë lumenjtë që kalojnë nëpër shtete të cilat janë anëtare të organizatave ndërkombëtare NATO dhe BE. Pjesë e logjikës së kërkesës është përjashtimi i dy shteteve specifike nga analiza: Suedia (code='S') dhe Franca (code='F'). Query-ja filtron shtetet me @memberships që përfshijnë NATO ose BE, dhe më pas kontrollon se cilët lumenj kalojnë nëpër ato shtete. Lumenjtë dhe shtetet përkatëse renditen në mënyrë alfabetike për të ofruar një strukturë të lexueshme dhe të organizuar.

Hyerje: Dokumenti XML: mondial.xml, i ngarkuar në ambientin BaseX.

Dalje:

Një strukturë XML e personalizuar me root <rivers_in_EU-Nato_countries>, e cila përmban emrat e lumenjve dhe shteteve ku ata kalojnë.

Gjenerimi i një file HTML (xquery1_result.html) që paraqet në mënyrë të organizuar listën e lumenjve dhe shteteve përkatëse.

Shembull rezultati

```
<rivers_in_EU-Nato_countries>
  <river>
    <name>Guadiana</name>
    <country>Portugal</country>
    <country>Spain</country>
  </river>
  ...
</rivers_in_EU-Nato_countries>
```

Lokacioni i kodit: GitHub: [XQuery2](#)

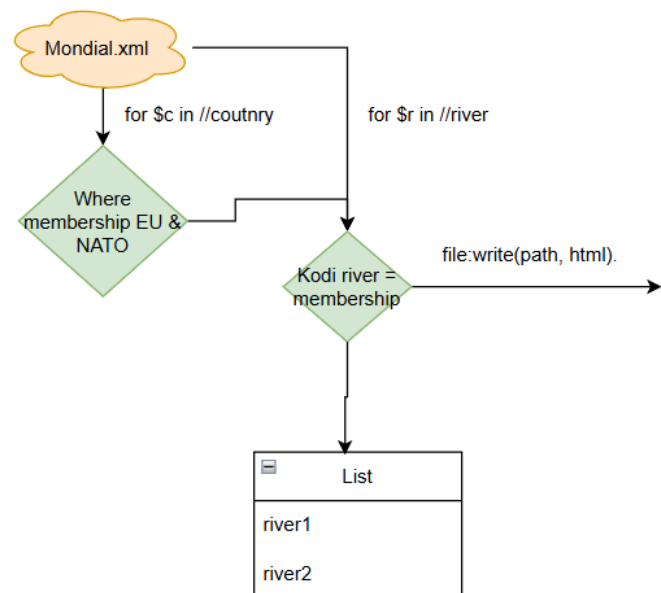


Figura 3. Diagrami logjik i ekzekutimit të XQuery2

3.1.3.3. XQuery 3: Top 20 shtetet më të vaksinuara kundër COVID-19 (Shtesë)

Qëllimi dhe struktura: Kjo query ka për qëllim të identifikojë 20 shtetet kryesore me përqindjen më të lartë të popullsisë plotësisht të vaksinuar kundër COVID-19. Të dhënat për popullsinë dhe emrat e shteteve merren nga dokumenti mondial.xml, ndërsa numri i vaksinimeve merret nga një dataset i konvertuar në XML nga formati JSON përmes një skripti Python (linku në GitHub). Për çdo shtet, llogaritet maksimumi i njerëzve të vaksinuar plotësisht gjatë ditëve të raportuara dhe pastaj përqindja e tyre në raport me popullsinë. Rezultatet renditen në mënyrë zbritëse sipas përqindjes së vaksinimit dhe shfaqen si një tabelë HTML.

Hyerje:

- Dokumenti mondial.xml me të dhënat për shtetet dhe popullsinë
- Dokumenti vaccinations.xml (i konvertuar nga JSON në XML përmes një tool-i Python – shiko GitHub: [vaccination_converter.py](#))
- Të dy dokumentet të ngarkuar në ambientin BaseX

Dalje:

- Një strukturë XML me root <Top20>, e cila përmban informacionin për shtetet me përqindjen më të lartë të vaksinimit.
- Gjenerimi i një file HTML (vaccination_report.html) që paraqet një tabelë me të dhënat e mëposhtme për 20 shtetet e para: emri i shtetit, popullsia, numri i personave të vaksinuar plotësisht dhe përqindja e vaksinuar.

Shembull rezultati

```
<Top20>
<result>
  <country>Gibraltar</country>
  <population>33,000</population>
  <people_fully_vaccinated>35,000</people_fully_vaccinated>
  <vaccination_percentage>106%</vaccination_percentage>
</result>
...
</Top20>
```

Lokacioni i kodit GitHub: [XQuery3](#)

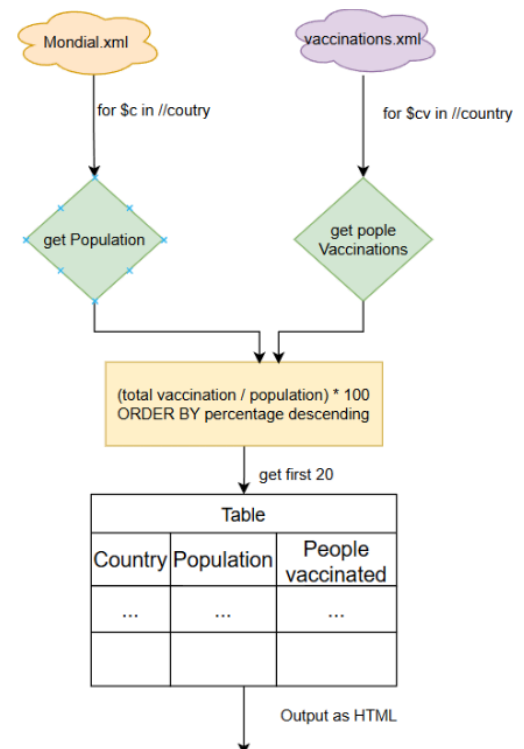


Figura 4. Diagrami logjik i ekzekutimit të XQuery shtesë

3.1.4. Përdorimi i XPath

XPath (XML Path Language) është një gjuhë për navigimin nëpër dokumente XML. Ajo përdor shprehje rrugësh për të selektuar nyje (nodes) brenda një dokumenti dhe është shumë efektive për kërkime të thjeshta dhe filtrime pa nevojën për të ndërtuar struktura të reja XML. XPath mbështetet nga shumë mjete XML, përfshirë BaseX, dhe është përdorur në këtë projekt për të bërë pyetje direkte mbi të dhënat e skemës Mondial, duke ofruar rezultate të menjëhershme dhe të lexueshme.

3.1.4.1. XPath 1: Shtetet e pavarura nga Britania e Madhe me përqindje të lartë të anglishtes

Qëllimi dhe struktura: Kjo query identifikon shtetet që janë pavarësuar nga Britania e Madhe (me atributin @from='GB') dhe në të cilat më shumë se 50% e popullsisë flet anglisht. Filtrimi bëhet duke përdorur funksionet krahasuese brenda XPath dhe kthen vetëm emrat e shteteve që plotësojnë të dy kushtet.

Hyerja: Fajlli i mondial.xml, i ngarkuar në BaseX.

Dalja:

Lista e elementeve <name> që përfaqësojnë emrat e shteteve të filtruar.

Shembull rezultati

```
<name>Ireland</name>
<name>Australia</name>
<name>New Zealand</name>
<name>Canada</name>
<name>United States</name>
<name>Bahamas</name>
```

...

Lokacioni i kodit GitHub: [XPath 1](#)

3.1.4.2. XPath 2: Qytetet bregdetare nën 10 metra dhe me mbi 500,000 banorë

Qëllimi dhe struktura: Kjo query identifikon qytetet që ndodhen pranë detit (me atributin @watertype='sea'), kanë lartësi nën 10 metra (elevation < 10) dhe kanë popullsi mbi 500,000 banorë. Shprehja XPath përdor disa kushte të kombinuara për të filtruar të dhënat dhe përdor funksionin concat() për të formuar rezultatin në formatin "Shteti → Qyteti".

Hyerja: Fajlli i mondial.xml, i ngarkuar në BaseX.

Dalja:

Strukturë tekstuale me formatin "Shteti → Qyteti", e nxjerrë direkt nga query XPath.

Shembull rezultati

Latvia → Riga
Qatar → Doha
Yemen → Aden
Sri Lanka → Colombo
Taiwan → Kaohsiung
South Korea → Changwon
Thailand → Bangkok

Lokacioni i kodit GitHub: [XPath2](#)

3.1.5. Kur të përdorim XQuery e kur XPath

XPath përdoret në situata kur:

- Kërkohen të dhëna direkte nga një dokument XML pa nevojë për transformime apo manipulime të avancuara, qasje e lehtë dhe e shpejtë.
- Duhet të filtrohen ose të selektohen elementë bazuar në attribute ose vlera specifike (p.sh., të gjitha qytetet me më shumë se 1 milion banorë).

XQuery përdoret kur:

- Kërkohet ndërtimi i strukturave të reja XML apo HTML në bazë të të dhënave të marra.
- Duhet të kombinohen të dhëna nga disa dokumente XML ose të llogariten vlera të përpunuara (si mesatare, përqindje, renditje).
- Output-i duhet të formatohet në mënyrë të personalizuar, si tabela HTML, raporte të grupuara apo struktura të reja me emërtime të përcaktuara nga përdoruesi.

Në praktikë, XPath është i përshtatshëm për query të shpejta dhe të fokusuara, ndërsa XQuery ofron më shumë fleksibilitet dhe kontroll për detyra të avancuara.

3.2. Vizualizimi me mjete të BI

3.2.1. Zgjedhja e Mjeteve

Për ndërtimin e vizualizimeve janë përzgjedhur dy nga mjetet më të njohura dhe të fuqishme të Business Intelligence: Power BI dhe Looker Studio. Zgjedhja është bërë duke marrë parasysh faktorë si lehtësia në përdorim, mbështetja për burime të ndryshme të të dhënave, mundësitë e personalizimit, si dhe aksesueshmëria përmes platformave të ndryshme.

Power BI është një platformë BI e zhvilluar nga Microsoft, e cila ofron aftësi të avancuara për analizë të të dhënave, lidhje me një gamë të gjerë burimesh (SQL, Excel, CSV, Web, API), dhe ndërtim të dashboards shumë të personalizueshëm me KPI, slicers, filtra dinamikë dhe grafikë interaktivë. Për më tepër, Power BI lejon përpunimin paraprak të të dhënave përmes Power Query dhe mbështet modelimin e avancuar me DAX.

3.2.1.1. Instalimi dhe konfigurimi i Power BI

Për ndërtimin e dashboards në këtë projekt, është përdorur Power BI Desktop, versioni falas i ofruar nga Microsoft për përdoruesit e sistemit operativ Windows. Instalimi dhe konfigurimi janë kryer duke ndjekur hapat e mëposhtëm:

Shkarkimi i aplikacionit:

- Vizitimi i faqes zyrtare.
- Zgjedhja e opsionit “Download free” për Power BI Desktop.
- Alternativisht, mund të shkarkohet përmes Microsoft Store në Windows.

Instalimi:

- Ekzekutimi i skedarit .exe të shkarkuar.
- Ndjekja e udhëzimeve të instaluesit (Next > Accept > Install).
- Pas përfundimit, aplikacioni mund të hapet direkt nga desktop-i.

3.2.1.2. Konfigurimi i Looker Studio

Looker Studio, nga ana tjetër, është një mjet BI falas nga Google, i cili lehtëson krijimin e raporteve vizuale përmes një ndërfaqeje të thjeshtë dhe të aksesueshme nga shfletuesi. Ai është shumë efikas për lidhje të shpejtë me Google Sheets, BigQuery, CSV, dhe burime të tjera të mbështetura online. Për shkak të integritetit natyral me Google Cloud Platform dhe mjete të tjera Google, është ideal për shpërndarje të shpejtë të raporteve interaktive online.

Qasja bëhet përmes një llogarie Google nga çdo shfletues modern, duke vizituar adresën zyrtare. Të gjitha raportet ruhen online dhe mund të shpërndahen përmes lidhjeve interaktive në kohë reale.

3.2.2. Krijimi i Vizualizimeve me Power BI

3.2.2.1. Vizualizimi i Mondial DB

Për këtë vizualizim është përdorur baza e të dhënave Mondial, e cila përmban informacione të detajuara mbi vendet e botës, përfshirë popullsinë, ekonominë, religjionet, qytetet dhe gjeografinë fizike.

Për të vizualizuar të dhënat e bazës Mondial DB në Power BI, është përdorur SQL Server-i si ndërmjetës për ruajtjen dhe menaxhimin e të dhënave. Në Power BI, është zgjedhur opsioni “Get Data > SQL Server” dhe janë zgjedhur tabelat dhe views pastaj është kontrolluar Data Model për të siguruar lidhjet ndërmjet tabelave.

Nga kjo bazë të dhënash janë nxjerrë të dhëna që lidhen me shpërndarjen e religjioneve globale dhe numrin e ndjekësve.

Qëllimi kryesor i këtij dashboard-i është të paraqesë në mënyrë vizuale:

- Popullsinë e përllogaritur globale sipas religjioneve.
- Numrin total të religjioneve të regjistruara në të dhënat e Mondial DB.
- Shpërndarjen kontinentale të religjioneve (përmes slicer-it për kontinentin).

Grafikët dhe mjetet e përdorura për vizualizim janë:

1. **Column Chart vertikal** - tregon numrin e vendeve ku çdo religjion është i përhapur.
2. **Pie Chart** - vizualizon ndarjen e popullsisë globale sipas religjioneve.
3. **KPI Card** - është përdorur për paraqitjen e numrit total të religjioneve të regjistruara.
4. **Slicer** - lejon përdoruesin të filtrojë të dhënat sipas kontinenteve për të analizuar më mirë shpërndarjen gjeografike të religjioneve.

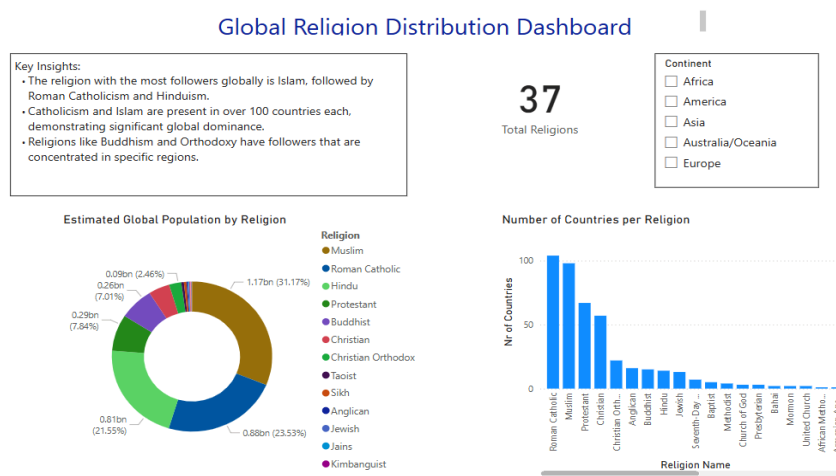


Figura 5. Dashboard i Mondial Dataset-it për shpërndarjen e religjioneve

KPI Card është përdorur për paraqitjen e numrit total të religjioneve të regjistruara ndërsa Slicer lejon përdoruesin të filtrojë të dhënat sipas kontinenteve për të analizuar më mirë shpërndarjen gjeografike të religjioneve.

Gjetjet Kryesore:

- 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.
- Fe si Budizmi dhe Ortodoksia kanë ndjekës që janë të përqendruar kryesisht në rajone të caktuara.

Fajlli i dashboardit të dataset-it Mondial: [Power BI Vizualizimi](#)

3.2.2.2. Vizualizimi i CSV Dataset-it të importuar

Për këtë vizualizim është përdorur një dataset i importuar në format CSV, i cili përmban informacione mbi lokacionet e gjetjeve arkeologjike të hershme të njerëzimit. Dataset-i është konvertuar nga formati JSON në XML përmes një skripti të ndërtuar në gjuhën Python dhe më pas është ruajtur si CSV për qëllime të analizës. Për transformimin dhe pastrimin e të dhënave është përdorur Power Query brenda Power BI, duke mundësuar ndërtimin e një modeli të qartë për vizualizim. Të dhënat janë marrur nga: [Github - Covid Vaccinations](#).

Qëllimi kryesor i këtij dashboard-i është të paraqesë në mënyrë vizuale:

- Numrin total të artefakteve të zbuluara sipas kontinenteve.
- Lokacionet më të rëndësishme të gjetjeve arkeologjike në hartë.
- Përqendrimin gjeografik të civilizimeve të hershme sipas të dhënave të datuara.

Grafikët dhe mjetet e përdorura për vizualizim janë:

1. **Pie Chart** – tregon shpërndarjen e artefakteve në çdo kontinent.
2. **Heat Map** – paraqet lokacionet gjeografike të gjetjeve me intensitet vizual sipas përqendrimit.
3. **Card** – pasqyron numrin total të artefakteve të identifikuara për kontinent.
4. **Slicer** – lejon filtrimin sipas kontinenteve për eksplorim më të detajuar.

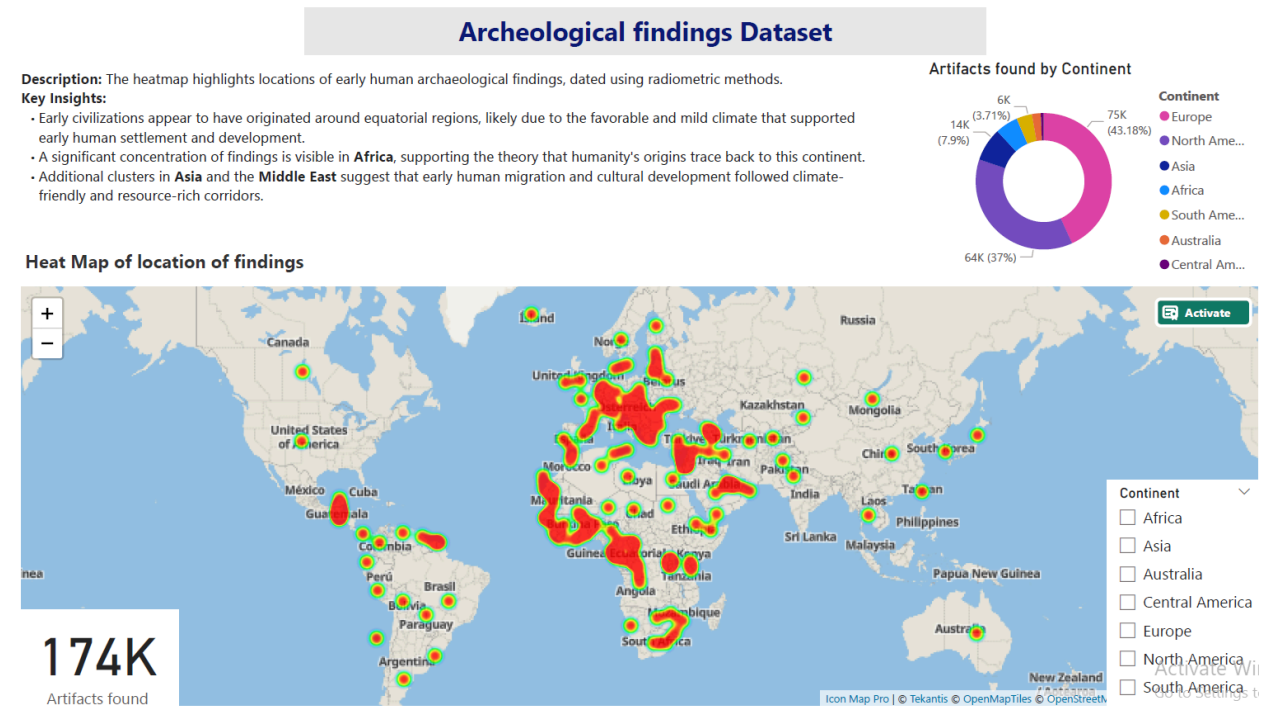


Figura 6. Dashboard i Dataset-it Arkeologjik për shpërndarjen e artefakteve

Card është përdorur për paraqitjen e numrit total të gjetjeve, ndërsa Heat Map dhe Slicer mundësojnë analizë ndërvepruese të shpërndarjes së tyre sipas rajonit.

Gjetjet Kryesore:

- Rajonet me dendësinë më të lartë të artefakteve janë Afrika dhe Azia, duke përfshirë teorinë e origjinës njerëzore në këto kontinente.
- Zbulimet arkeologjike janë më të përqendruara në brezin ekuatorial, për shkak të kushteve klimaterike të favorshme për zhvillimin e hershëm të jetës. Vendet me më pak gjetje si Australia dhe Amerika Qendrore tregojnë mungesë të kërkimeve intensive ose dokumentimit historik në këto zona.

Fajlli i dashboard-it të dataset-it CSV: [Power BI Vizualizimi](#)

3.2.2.3. Visualizimi i Query/View të nxjerrur nga databazat ekzistuese

Për këtë vizualizim janë përdorur dy dataset-e të përditshme që përfshijnë numrin e njerëzve të vaksinuar dhe numrin e vdekjeve ditore nga COVID-19 në të gjithë botën. Të dhënat janë kombinuar dhe përpunuar në MySQL dhe janë vendosur në Power BI përmes “Import View”, duke u lidhur përmes muajit dhe vitit të raportimit për të analizuar marrëdhënien mes vaksinimeve dhe numrit të vdekjeve. Të dhënat janë marrur nga

Për të vizualizuar ndikimin e vaksinimit në uljen e fataliteteve gjatë pandemisë, është ndërtuar një dashboard dinamik që paraqet në mënyrë të integruar numrin total të vaksinimeve dhe vdekjet e reja mujore.

Qëllimi kryesor i këtij dashboard-i është të paraqesë në mënyrë vizuale:

- Numrin total të njerëzve të vaksinuar për çdo muaj që nga fillimi i vaksinimit.
- Trendet e vdekjeve ditore në të njëjtën periudhë kohore.
- Ndikimin e mbulimit me vaksina në uljen e fataliteteve globale.

Grafikët dhe mjetet e përdorura për vizualizim janë:

1. **Line Chart** – paraqet në një grafik të vetëm trendin e vaksinimeve dhe vdekjeve mujore në periudhë shumëvjeçare.
2. **Card** – tregon totalin e vaksinimeve të regjistruara dhe numrin total të vdekjeve të reja gjatë periudhës së analizuar.
3. **Slicer** – lejon përzgjedhjen e periudhës sipas vitit dhe muajit për analizë më të saktë.

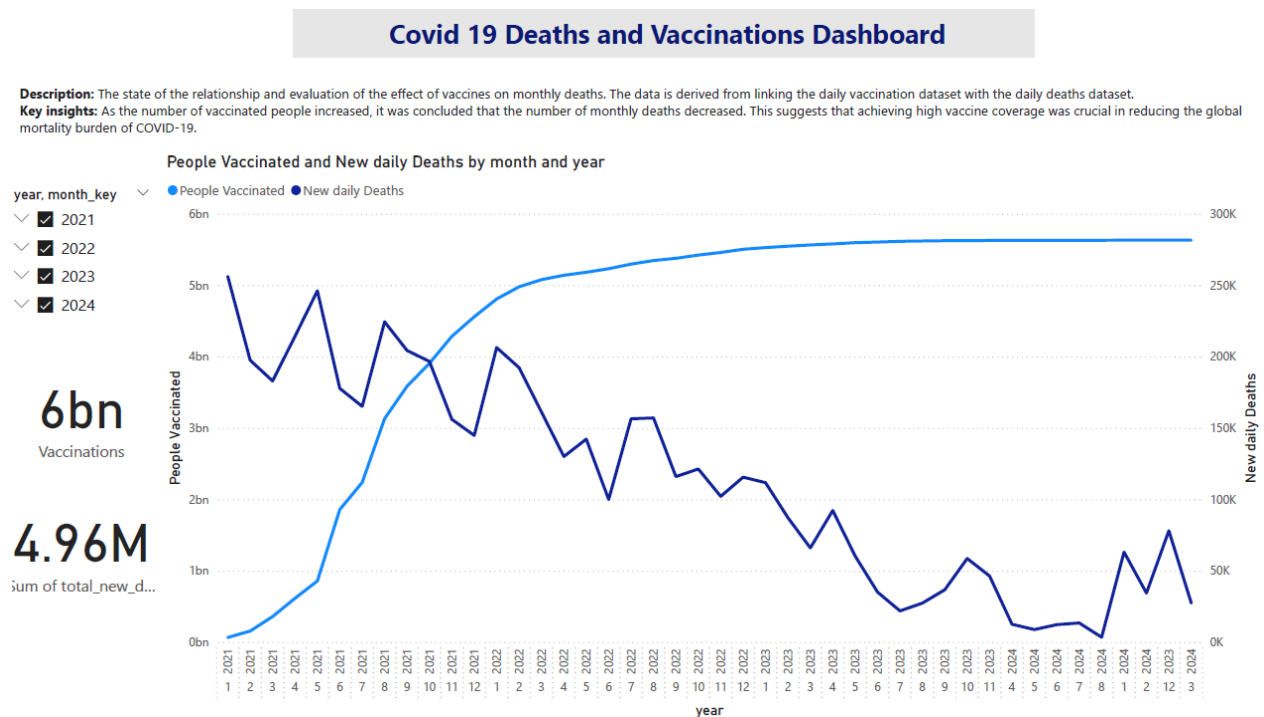


Figura 7. Dashboard i COVID-19 për vaksinat dhe vdekjet globale

Grafiku i linjës paraqet qartë një rënie të vazhdueshme të vdekjeve mujore si rezultat i rritjes së numrit të vaksinimeve, ndërsa KPI Card ofron një pasqyrë të përmbledhur të ndikimit total të pandemisë.

Gjetjet Kryesore:

- Me rritjen e mbulimit global me vaksina, numri i vdekjeve mujore ka pësuar një rënie të konsiderueshme.
- Viti 2021 përmban pikat më të larta të vaksinimit dhe gjithashtu një rënie progresive të fataliteteve pas çdo vale të re.
- Vaksinimi masiv ka rezultuar në një ulje të barrës së përgjithshme të vdekjeve në shkallë globale.

Fajlli i dashboard-it të dataset-it COVID-19: [Power BI Vizualizim](#)

3.2.2.3. Vizualizimi i Dataset-it me mbi 10,000 rekorde

Për këtë vizualizim është përdorur **Global Terrorism Dataset**, një dataset me mbi 180,000 rreshta, i cili përmban të dhëna të detajuara për sulmet terroriste në mbarë botën që nga vitet 1970 deri në vitin 2017 i marrë nga uebfaqja [Kaggle](#). Nga ky dataset u filtruan të dhënat për periudhën pas vitit 2000, për të analizuar ndikimin e sulmeve në kohët moderne.

Qëllimi kryesor ishte të analizohej:

- numri total i viktimave gjatë kësaj periudhe,
- trendi i sulmeve në kohë,
- shpërndarja e viktimave sipas rajoneve,
- dhe cilat janë 10 vendet më të prekura nga sulmet terroriste.

Grafikët e përdorur në dashboard:

1. **KPI Card** – Tregon numrin total të viktimave.
2. **Line Chart** – Tregon numrin e sulmeve terroriste nga viti në vit
3. **Bar Chart horizontal** – Paraqet numrin e viktimave sipas rajonit.
4. **Pie Chart** – Paraqet 10 shtetet me më shumë sulme terroriste.
5. **Slicer** – Lejon filtrimin e vizualizimeve sipas rajonit.

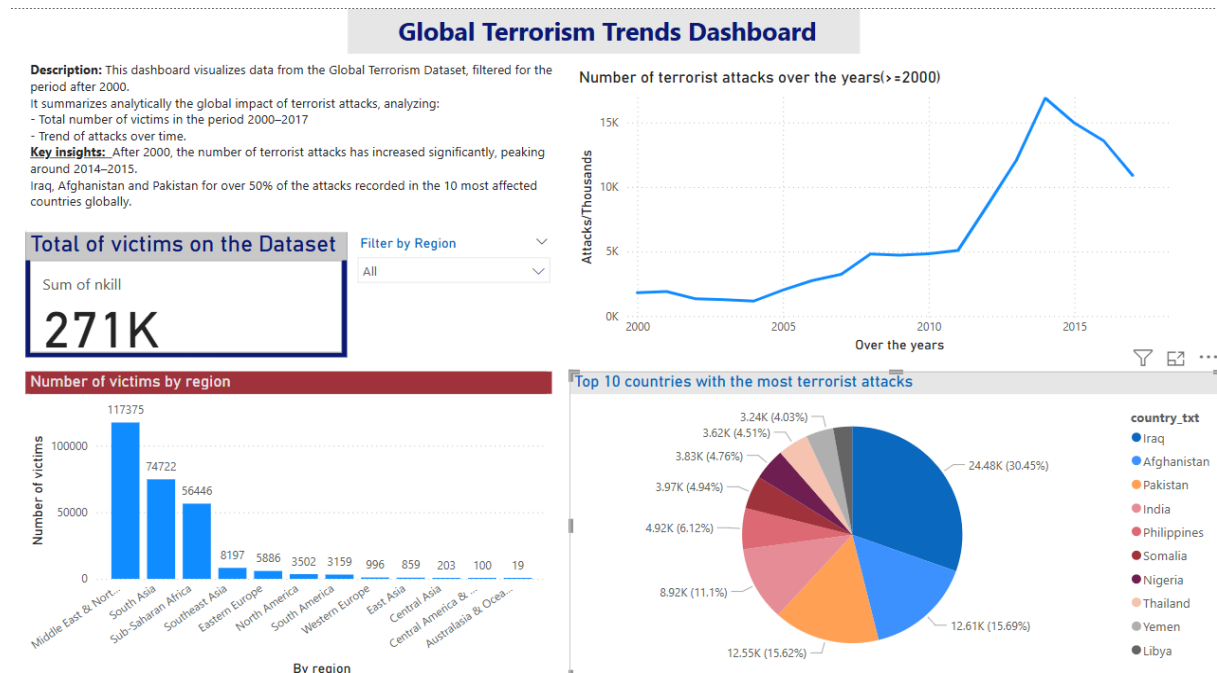


Figura 6. Dashboardi i Dataset-it për Terrorizmin Global

Gjetjet kryesore:

- Numri i sulmeve terroriste u rrit ndjeshëm pas vitit 2010, me kulm në vitet 2014–2015.
- Vendet më të prekura janë Iraku, Afganistani dhe Pakistani, që së bashku përbëjnë mbi 50% të sulmeve të regjistruara në top 10-she.
- Rajoni i Lindjes së Mesme dhe Azisë Jugore ka numrin më të madh të viktimave.

Kjo vizualizim demonstroi fuqinë e Power BI në analizimin e një dataset-i të madh dhe kompleks përmes ndërfaqes interaktive dhe vizuale.

Skedari i dashboardit të dataset-it me 10,000 rekorde: [Power BI Vizualizimi](#)

3.2.2.4. Vizualizimi i Dataset-it për performancën akademiket universitare

Në këtë vizualizim është përdorur një dataset që përmban rangimet globale të universiteteve për periudhën 2011-2016, me fokus të veçantë në cilësinë akademike dhe performancën shkencore i marr nga uebfaja [Kaggle](#). Qëllimi i këtij dashboard-i është të paraqesë në mënyrë të përmbledhur dhe vizuale:

- Mesataren globale të performancës universitare si indikator kryesor.
- Tendencat e cilësisë akademike gjatë viteve.

- Shtetet kryesore sipas cilësisë universitare në një diagramë krahasuese.
- Shpërndarjen gjeografike të universiteteve më të mira në një harta interaktive.

Grafikët dhe Mjetet e Përdorura:

- **KPI Card** – Paraqet mesataren globale të rangimit të universiteteve
- **Line Chart** – Ilustron ndryshimet e mesatares së cilësisë universitare gjatë viteve.
- **Bar Chart Horizontal** – Evidenton dhe krahason shtetet kryesore bazuar në cilësinë akademike të universiteteve të tyre.
- **Harta Gjeografike** – Vizualizon shpërndarjen rajonale dhe kontinentale të universiteteve më të renditura, duke nxjerrë në pah fokuset kryesore akademike në botë.

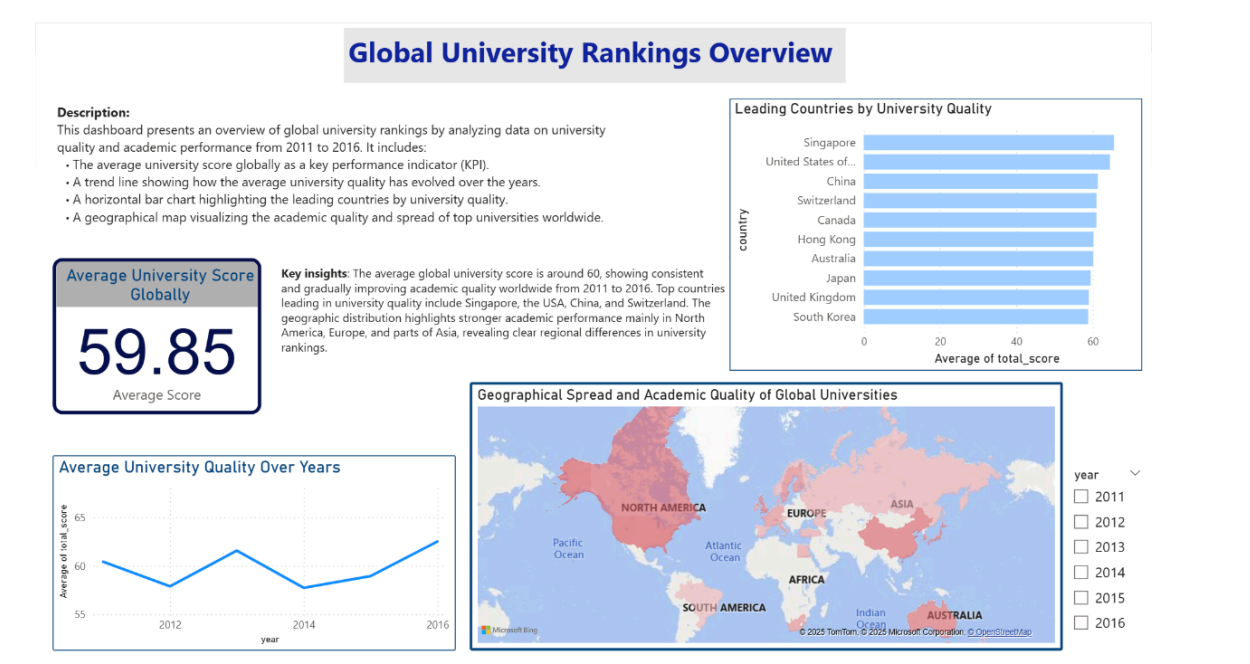


Figura x. Dashboardi i performancën akademike universitare (2011–2016)

Gjetjet Kryesore:

- Mesatarja globale e cilësisë universitare qëndron rreth 60, duke reflektuar një përmirësim të qëndrueshëm në cilësinë akademike në periudhën 2011–2016.
- Shtetet kryesore në cilësinë universitare përfshijnë Singaporin, Shtetet e Bashkuara të Amerikës, Kinën dhe Zvicrën.

- Shpërndarja gjeografike tregon performancë më të fortë akademike kryesisht në Amerikën e Veriut, Evropë dhe pjesë të Azisë.
- Rajonet kanë dallime të qarta në renditjen akademike, me një tendencë të përgjithshme rritëse të cilësisë gjatë viteve të analizës. Megjithatë, gjatë këtij intervali u vunë re disa luhajtje, përfshirë rënie të përkohshme në vitet 2012 dhe 2014, që sugjerojnë ndryshime apo sfida të përkohshme në performancën universitare në nivel global.

3.2.3. Krijimi i Vizualizimeve me Looker Studio

3.2.3.1. Vizualizimi i Dataset-it për llojin e motit

Për këtë vizualizim është përdorur një dataset i marrë nga Kaggle, i cili trajton klasifikimin e motit sipas tipareve klimatike si temperatura, lagështia dhe mbulimi me re.

Burimi i dataset-it është uebfaqja [Kaggle](#).

Dataset-i është shkarkuar në format CSV dhe më pas është importuar në Looker Studio përmes lidhjes me Google Sheets, ku paraprakisht janë pastruar dhe riformatuar disa kolona për përputhje me kërkesat e vizualizimit.

Qëllimi i këtij vizualizimi është të analizohet ndikimi i llojit të motit në indikatorë si temperatura mesatare, lagështia dhe shpërndarja e reve, për të krijuar një përmbledhje të tendencave më të rëndësishme klimatike që vërehen në të dhëna.

Grafikët dhe mjetet e përdorura për vizualizim janë:

1. **Column Chart vertikal** – tregon temperaturën mesatare sipas llojit të motit.
2. **Combo Chart** – paraqet temperaturën dhe lagështinë për secilin lloj moti në të njëjtin grafik.
3. **Pie Chart** – vizualizon përqindjen e shpërndarjes së llojit të qiellit (clear, cloudy, partly cloudy, overcast).

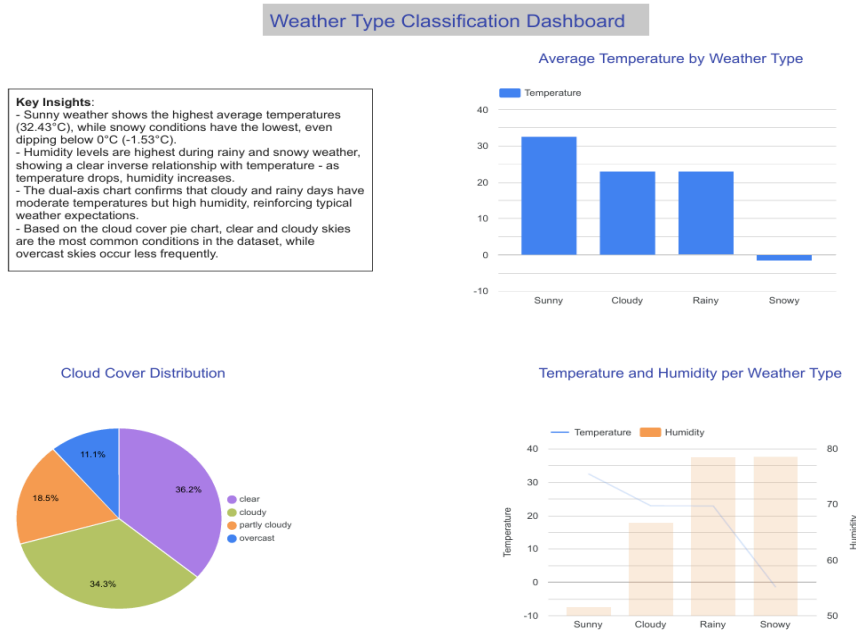


Figura x. Dashboard i Dataset-it për llojin e motit.

Gjetjet Kryesore:

- Motet me diell kanë temperaturat më të larta mesatare (32.43°C), ndërsa motet me borë kanë temperaturat më të ulëta, duke rënë edhe nën 0°C (-1.53°C).
- Lagështia është më e lartë gjatë motit me shi dhe borë, duke treguar një lidhje të kundërt me temperaturën.
- Moti me re dhe me shi ka temperatura mesatare, por lagështi të lartë.
- Grafiku me rreth (pie chart) për mbulimin me re tregon që moti i kthjellët dhe me re janë më të zakonshmit, ndërsa "overcast" (re të dendura) janë më të rrallë.

Skedari i dashboardit të dataset-it të motit: [Looker Studio Vizualizimi](#)

3.2.3.2. Vizualizimi i Dataset-it has për përdorimin e alkoolit në botë

Në këtë vizualizim është përdorur një dataset i pasur që kombinon informacion mbi konsumimin e alkoolit për banorë me tregues ekonomikë, si GDP për capita dhe popullsia, për vendet e ndryshme të botës nga viti 2000 deri në 2018. Të dhënat janë përpunuar fillimisht në SQL Server, ku është krijuar një view e quajtur [AlcoholGDP View](#), që përmban kolonat e nevojshme të dëshiruara dhe të filtruara për vizualizim.

Rezultati i View-së është ruajtur si skedar CSV dhe është importuar në Looker Studio, ku janë ndërtuar dashboards interaktiv për analizimin vizual të këtyre marrëdhënieve.

Qëllimi i vizualizimit ishte:

- analizimi i shpërndarjes globale të konsumimit të alkoolit,
- lidhja mes GDP për capita dhe konsumit të alkoolit,
- identifikimi i vendeve me konsumin më të lartë të raportuar,
- analiza e tendencave sipas viteve për vendet e ndryshme.

Grafikët e përdorur në dashboard:

1. **Scatter Plot** – Paraqet shpërndarjen e vendeve sipas GDP per capita dhe konsumit të alkoolit. Tregon mungesë korrelacioni të qartë.
2. **Line Chart** – Tregon konsumimin e alkoolit sipas vendeve në mënyrë të renditur (nga më i larti në më të ultin).
3. **Column Chart** – Paraqet 10 vendet me konsumimin më të lartë të alkoolit.
4. **Dropdown Filter** – Lejon filtrimin e të dhënave sipas vitit për analiza dinamike.

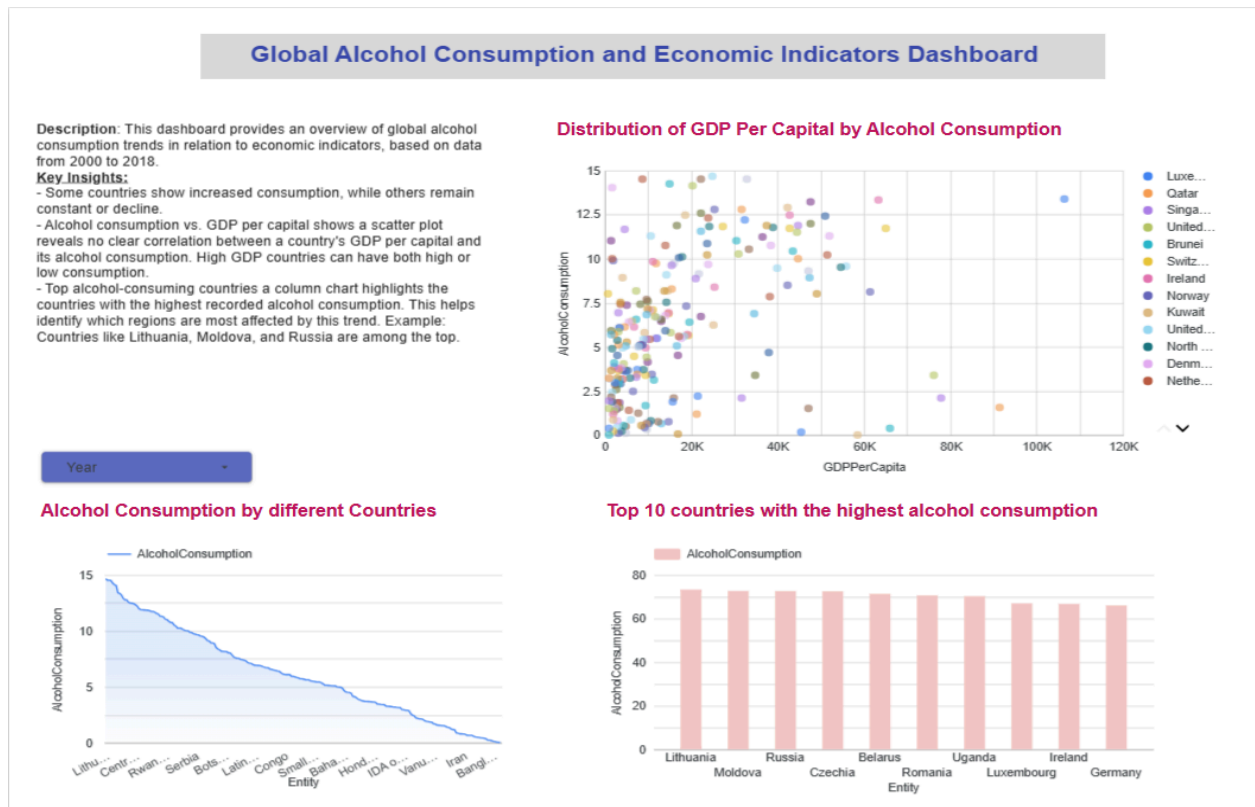


Figura x. Dashboard i Dataset-it për konsumimin e alkoolit në botë.

Gjetjet kryesore (Key Insights):

- Vendet me GDP të lartë nuk kanë domosdoshmërisht konsum më të lartë alkooli; disa kanë konsum të ulët.
- Vendet si Lituania, Moldavia dhe Rusia rezultojnë ndër më të prekura nga konsumimi i lartë.
- Shumë vende të Evropës Lindore dhe Qendrore janë ndër më të ekspozuarat ndaj këtij fenomeni.
- Në disa vende konsumi qëndron konstant në vite, ndërsa në të tjera ka ulje ose rritje të dukshme.

Skedari i dashboardit të dataset-it për përdorimin e alokoolit: [Looker Studio Vizualizimi](#)

IV. Sfidat dhe Zgjidhjet

Gjatë realizimit të projektit janë hasur disa sfida teknike, veçanërisht në fazën e konvertimit të formateve të ndryshme të dataset-eve, në mënyrë që ato të ishin të përshtatshme për pjesët teknike të projektit si validimi XML, manipulimi me XPath/XQuery dhe importimi në mjetet BI.

Për ta zgjidhur këtë sfidë, u zhvilluan skripta të dedikuara në Python të cilat automatizojnë konvertimin mes këtyre formateve përkatësisht:

1. Nga JSON në CSV [[Kodi në Github](#)]
2. Nga JSON në XML [[Kodi në Github](#)]
3. Nga CSV në XML [[Kodi në Github](#)]

Këto skripta lexojnë strukturën e të dhënave burimore (p.sh. objekte JSON me lista të brendshme), i transformojnë ato në forma të sheshta ose të strukturuara siç kërkohet, dhe i eksportojnë në formatin e synuar me etiketa dhe renditje të saktë.

Ky qasje:

- eliminoi nevojën për manipulim manual të të dhënave,
- garantoi konsistencë dhe vlefshmëri strukturore (veçanërisht për XML), dhe
- përshpejtoi përgatitjen e të dhënave për secilën fazë të projektit.

Përdorimi i këtyre skriptave automatizoi dhe standardizoi procesin e përgatitjes së të dhënave, duke garantuar integritetin, vlefshmërinë dhe kompatibilitetin me mjetet teknologjike të përdorura në projekt .

V. Konkluzione

Ky projekt demonstroi rëndësinë e integritit të përpunimit të të dhënave XML me mjetet moderne të vizualizimit të Business Intelligence si Power BI dhe Looker Studio. Përmes përdorimit të XQuery dhe XPath, është realizuar një analizë e thelluar mbi të dhënat komplekse të Mondial XML dhe disa dataset-ave të tjerë të marrë për analizë, duke siguruar validim të saktë dhe nxjerrje të qëllimshme të informacionit. Në fazën pasuese, këto të dhëna janë shndërruar në dashboards interaktive që ofrojnë njohuri të vlefshme për temat si demografia, pandemia, arsimi, moti dhe terrorizmi global. Projekti thekson se aftësia për të pastruar, transformuar dhe vizualizuar të dhëna është thelbësore për marrjen e vendimeve të informuara dhe logjike në epokën digjitale.

Referencat

1. Baza e të dhënave Mondial – Universiteti i Leipzig.
E disponueshme në: <https://www.dbis.informatik.uni-goettingen.de/Mondial/>
2. Baza e të dhënave “Terrorizmi Global”.
E disponueshme në: <https://www.kaggle.com/datasets/START-UMD/gtd>
3. Baza e të dhënave “Konsumimi i Alkoolit në botë”.
E disponueshme në: <https://www.kaggle.com/datasets/pralabhpoudel/alcohol>
4. Baza e të dhënave “Klasifikimi i motit”.
E disponueshme në: <https://www.kaggle.com/datasets/nikhil7280/weather-type>
5. Baza e të dhënave “Rankimi i Universiteteve në botë”.
E disponueshme në: <https://www.kaggle.com/datasets/mylesoneill/world-university>
6. Baza e të dhënave “Covid Vaccination”.
E disponueshme në:
<https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations>
7. Baza e të dhënave “Covid Deaths”.
E disponueshme në:
<https://www.kaggle.com/datasets/dhruvildave/covid19-deaths-dataset>