

Universiteti i Prishtinës “Hasan Prishtina”

Fakulteti i Inxhinierisë Elektrike dhe Kompjuterike



Dokumentim teknik i projektit

Lënda: Big Data

Titulli i projektit: Manipulim i të dhënave XML dhe vizualizim BI

Emri profesorit/Asistentit

Prof. Dr. Vigan Raça
MSc. Rafet Duriqi

Emri & mbiemri studentëve / email adresa

1. Fjolla Kadriu	fjolla.kadriu@student.uni-pr.edu
2. Leutrim Morina	leutrim.morina13@student.uni-pr.edu
3. Linda Hasanaj	linda.hasanaj@student.uni-pr.edu
4. Lum Hoxha	lum.hoxha@student.uni-pr.edu

Prishtinë, 2025

Përmbajtja

Abstrakti	3
1. Hyrje	4
2. Qëllimi i punimit	5
3. Pjesa kryesore	6
3.1. XML data manipulation	6
3.1.1 Instalimi dhe konfigurimi i BaseX	6
3.1.2 Importimi i databazës Mondial	6
3.1.3 Validimi i dokumentit XML duke përdorur xmllint	6
3.1.4 XQueries	7
3.1.5 XPath	8
3.2 BI tool visualization	10
3.2.1 Përzgjedhja e tool-eve	10
3.2.2 Lidhja e datasources	10
3.2.3 Krijimi i vizualizimeve	10
Vizualizimi për temperaturat vjetore në qytetet kryesore - Power BI	10
Vizualizimi i numrit të liqeneve për shtet - Power BI	11
Lumenjtë sipas shteteve (të filtruara nga një view) - Power BI	12
Përmbledhje globale e COVID-19 - Rastet e konfirmuara, të shëruarat dhe vdekjet sipas shteteve - Looker Studio	13
Analiza vizuale e faktorëve që ndikojnë në jetën studentore - Looker Studio	14
4. Sfidat dhe zgjidhjet	16
5. Konkluzione	17
Referencat	18

Abstrakti

Ky projekt fokusohet në manipulimin dhe analizën e të dhënave XML si dhe në krijimin e vizualizimeve kuptimplota përmes mjeteve të Business Intelligence, duke përdorur bazën e të dhënave Mondial dhe dataset-e të mëdha të hapura si ato për temperaturat ditore dhe rastet e COVID-19. Tema është interesante pasi përfshin si pjesën teknike të trajtimit të të dhënave XML (XPath, XQuery, validim me XSD), ashtu edhe analizën vizuale përmes Power BI dhe Looker Studio, duke ofruar një pamje të plotë mbi menaxhimin dhe interpretimin e të dhënave heterogjene.

Zgjidhja jonë kombinon përdorimin e xmllint për validim të saktë dhe efikas, BaseX për ekzekutim të pyetjeve XQuery/XPath, si dhe Power BI/Looker Studio për vizualizim të të dhënave në kohë reale. Kjo metodologji është më efektive pasi është lehtë e përsëritshme, funksionale me burime të ndryshme të dhënash, dhe e përshtatshme si për analizë teknike, ashtu edhe për prezantim të të dhënave për përdorues jo-teknikë.

1. Hyrje

Në epokën e të dhënave të mëdha, mënyra se si të dhënat strukturohen, analizohen dhe vizualizohen ka marrë një rëndësi kyçe në fusha të ndryshme, nga analiza ekonomike deri te shëndetësia dhe gjeopolitika. Të dhënat në formatin XML përfaqësojnë një nga strukturat më të përhapura për shkëmbim të informatave ndër-platformore, ndërsa mjetet e Business Intelligence (BI) si Power BI dhe Looker Studio e mundësojnë analizën vizuale në kohë reale dhe për përdorues jo-teknikë.

Ky projekt është motivuar nga nevoja për të ndërlidhur njohuritë teorike të fituara në lëndën e Big Data me aplikime praktike që shtrihen nga manipulimi i dokumenteve XML, validimi i tyre, ekzekutimi i pyetjeve të avancuara në XQuery/XPath, deri te ndërtimi i dashboard-eve interaktive për analiza kuptimplota. Si dataset kryesor është përdorur Mondial, një bazë të dhënash që përmban të dhëna gjeopolitike, dy dataset-e të mëdha për temperaturat globale dhe COVID-19, si dhe një dataset me të dhëna për studentët.

2. Qëllimi i punimit

Qëllimi i këtij punimi është të demonstrojë mënyrën se si të dhënat XML nga databaza Mondial mund të manipulohen, analizohen dhe vizualizohen përmes një kombinimi të mjeteve efikase si BaseX, xmllint, Power BI dhe Looker Studio. Përzgjedhja e këtyre veglave është bërë në bazë të funksionalitetit të avancuar që ofrojnë secila në faza të ndryshme të trajtimit të të dhënave.

BaseX është përdorur për ruajtjen dhe manipulimin e dokumentit XML, duke ofruar mbështetje të fuqishme për gjuhët XQuery dhe XPath, të cilat janë përdorur për të nxjerrë dhe filtruar të dhëna të rëndësishme nga struktura hierarkike e dokumentit. Përdorimi i xmllint për validim ka garantuar që struktura e dokumentit XML të jetë në përputhje me skemën XSD, duke siguruar integritetin dhe cilësinë e të dhënave.

Në fazën e vizualizimit, Power BI është përdorur për të krijuar raporte interaktive dhe të pasura me grafikë bazuar në query-t e nxjerra nga dokumenti XML, si dhe për të kombinuar këto të dhëna me dataset-e të tjera si ai i temperaturave dhe COVID-19. Looker Studio është përdorur si mjet i dytë vizualizimi për të demonstruar fleksibilitetin e përdorimit të mjeteve të ndryshme për analitikë në web, veçanërisht për të dhëna me burime të bazuara në Google. Në Looker Studio është importuar dataseti me të dhëna për studentët, i cili është përdorur për të krijuar vizualizime të ndryshme që pasqyrojnë faktorët kryesorë që ndikojnë në performancën akademike dhe mirëqenien mendore të tyre.

Për të mbledhur informatat e nevojshme janë përdorur metoda praktike eksperimentale: janë shkarkuar dhe analizuar dokumente XML, janë zhvilluar query-t të personalizuar për raste të ndryshme, dhe janë testuar rezultate të ndryshme përmes vizualizimeve grafike. Të dhënat janë mbledhur si nga burime të hapura (p.sh. Mondial, Kaggle), ashtu edhe nga vetë rezultatet e query-ve të zbatuara në BaseX.

Kjo qasje e integruar synon të ilustrojë se si teknologjitë për trajtimin e të dhënave mund të kombinohen për të krijuar një zgjidhje të plotë që mbulon të gjithë ciklin: nga struktura XML deri tek raportimi vizual i insight-eve të rëndësishme.

3. Pjesa kryesore

3.1. XML data manipulation

3.1.1 Instalimi dhe konfigurimi i BaseX

Për të instaluar BaseX janë ndjekur këta hapa:

1. Hapëm faqen zyrtare të BaseX: <https://basex.org/download>.
2. Nga pjesa e shkarkimeve u zgjodh versioni për sistemin operativ përkatës, në rastin tonë Windows.
3. Shkarkuam BaseXZip Kit dhe e ekstraktuam në një path lokal.
4. Për të hapur aplikacionin, vetëm ekzekutohet file BaseX.exe.

Verifikimi i instalimit u bë pas hapjes së aplikacionit, u verifikua me sukses funksionimi i GUI-të. Në aplikacion u hap një dokument testues XML për të konfirmuar që aplikacioni funksionon siç duhet.

3.1.2 Importimi i databazës Mondial

Të dhënat XML i kemi shkarkuar nga faqja zyrtare e databazës Mondial: <https://www.dbis.informatik.uni-goettingen.de/Mondial/> . Naviguam në seksionin Mondial XML Data, dhe shkarkuam dokumentin mondial.xml si një dokument të tillë. Pas shkarkimit, për të importuar databazën ndjekëm këta hapa:

1. Hapëm BaseX.
2. Nga menu Database - New zgjodhëm opsionin për të krijuar një databazë të re dhe e emërtuam Mondial.
3. Zgjodhëm file mondial.xml si input path dhe dokumenti i importua me sukses.

3.1.3 Validimi i dokumentit XML duke përdorur xmllint

Resurset që u nevojitën për të kryer validimin e dokumentit xml janë: xmllint nga paketa libxml2, mondial.xml file XML me të dhëna dhe mondial.xsd, skema XSD për validim. Fajllat mondial.xml dhe mondial.xsd i vendosëm në një folder të përbashkët. Pasi verifikuam instalimin me sukses të xmllint, kaluam në direktorinë ku i kemi ata dy fajlla dhe verifikuam se kjo databazë është validuar me sukses, në figurën e mëposhtme është paraqitur kjo:

```

Microsoft Windows [Version 10.0.26100.3915]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Pc>xmllint --version
xmllint: using libxml version 20708
        compiled with: Threads Tree Output Push Reader Patterns Writer SAXv1 FTP HTTP DTDV
        alid HTML Legacy C14N Catalog XPath XPointer XInclude Iconv Unicode Regexprs Automata
        Expr Schemas Schematron Modules Debug Zlib

C:\Users\Pc>cd "C:\Users\Pc\OneDrive\Desktop\mondial"

C:\Users\Pc\OneDrive\Desktop\mondial>xmllint --noout --schema mondial.xsd mondial.xml
mondial.xml validates

C:\Users\Pc\OneDrive\Desktop\mondial>|

```

Figura 1: Validimi i databazës Mondial në formatin xml.

Arsyeja pse validimin e realizuam me xmllint është se ky është një mjeft i fuqishëm dhe i besueshëm për analizë dhe validim të fajllave në mënyrë automatike, mundëson një validim të plotë strukturor dhe sintaksor, detajim të saktë të gabimeve që lehtëson identifikimin e problemeve.

3.1.4 XQueries

Query 1: Listo të gjitha kryeqytetet në të cilat nuk kalon asnjë lum, mirëpo janë anëtare të paktën të një organizate.

Këtë query të implementuar në XQuery e gjeni në Aneksin A, kurse teknikisht çfarë bën kjo XQuery është:

- Përmes atributit @memberships ne gjejmë shtetet të cilat janë anëtare të paktën të një organizate,
- Marrim id e kryeqytetit dhe gjejmë qytetin me atë id dhe kontrollohet nëse brenda elementit located_at nëse atributi watertype është river.

Query 2: Listo lumenjtë të cilët rrjedhin nëpër shtete të cilat nuk janë anëtarë të NATO-s, kanë dalje në dete si dhe numri i lumenjëve në këto shtete është më i madh se 10.

Këtë query të implementuar në XQuery e gjeni në Aneksin A, kurse teknikisht se çfarë bën kjo XQuery është:

- Së pari tek elementi <sea> kemi gjetur të gjitha shtetet të cilat kanë dalje në det,
- Më pas marrim countryCode nga këto shtete të cilat kanë dalje në det dhe kontrollojmë cilat prej tyre nuk janë anëtarë të NATO-s,
- Më pas në elementin <river> kemi kontrolluar atë countryCode nëse në të kalojnë më shumë se 10 lumenjë.

Query 3: Të listohen të gjitha detet në të cilat nuk ka asnjë ishull mirëpo shtetet të cilat kufizohet ai det janë anëtare në NATO-së ose BE-së.

Këtë query të implementuar në XQuery e gjeni në Aneksin A, kurse teknikisht se çfarë bën kjo XQuery është:

- Zgjedhen të gjithë elementët <sea> që nuk kanë asnjë ishull të lidhur me to,

- Për secilin det, merren kodet e shteteve që kufizohen me të, të ndara me hapësirë,
- Nga atributi @country i detit, ndan kodet e shteteve me tokenize, sepse mund të jenë të ndara me hapësirë (p.sh. "AL GR IT" → ["AL", "GR", "IT"]),
- Kontrollon për secilin kod shteti nëse ai shtet është anëtar i NATO-s (org-NATO) ose anëtar i BE-së (org-EU) në atributin @memberships,
- Nëse numri i shteteve të anëtarësuar është i njëjtë me numrin e të gjitha shteteve që kufizohen me detin, atëherë krijon një element <result> që përmban emrin e detit.

Query 4: Të listohen të gjitha shtetet anëtare të NATO-s të cilat kanë dalje në të njëjtin det dhe liqen.

Këtë query të implementuar në XQuery e gjeni në Aneksin A, kurse teknikisht se çfarë bën kjo XQuery është:

- Së pari e kemi një funksion të personalizuar local:ordered-pair i cili merr dy kode vendesh dhe i kthen të bashkuara, si dhe përdoret për të shmangur duplikatet,
- Filtrojmë vetëm vendet të cilat janë anëtare të NATO-s,
- Gjejmë çiftet anëtare të NATO-s që kanë dalje në të njëjtin det, dhe më pas jinden çiftet anëtare të NATO-s që ndajnë të njëjtin liqen,
- Gjen vetëm ato çifte që janë njëkoësisht në të dy listat e lartëpërmendura.

Query 5: Të listohen 5 malet më të larta të cilat shtrihen në vendet e Ballkanit duke përjashtuar Serbinë.

Këtë query të implementuar në XQuery e gjeni në aneksin A, kurse teknikisht se çfarë bën kjo XQuery është:

- Së pari përcaktojmë vetë, shtetet e Ballkanit duke përjashtuar Serbinë,
- I kemi selektuar të gjitha elementet <mountain> dhe kontrollojmë nëse njëri nga kodet e gjendura lartë është në ndonjë mal,
- Më pas renditim malet nga më i larti tek më i ulëti.

3.1.5 XPath

Query 1: Listo emrat e shteteve të cilat janë anëtare të NATO-s dhe kanë popullsi më pak se 10000000 banorë të regjistruar në vitin 2011.

Këtë query e kemi implementuar në XPath dhe e gjeni në Aneksin A, kurse teknikisht se çfarë bën ky XPath është:

- Në çdo element <country> në dokumentin XML kontrollojam nëse atributi @memberships në të e përmban NATO-në,
- Gjetëm elementin <population> me atributin @year = 2011 dhe kontrollojam nëse plotësohet kushti për më shumë se 10000000 banorë,
- Dhe pasi që një vend i përmbush këto kushte në rezultat kthejmë elementin <name>.

Query 2: Listo emrat e kryeqyteteve të shteteve të cilat janë anëtare të EU-së dhe nëpër këto kryeqytete nuk kalon asnjë lum.

Këtë query e kemi implementuar në XPath dhe e gjeni në Aneksin A, kurse teknikisht se çfarë bën ky XPath është:

- Kontrollojmë nëse atributi @id i një qyteti përputhet me atributin @capital të një shteti, i cili është anëtar i EU-së,
- Më pas vendosim kushtin që nëpër atë qytet të mos kalojë asnjë lum përmes atributit @watertype,
- Në rezultat shfaqim vetëm emrin e këtij kryeqyteti, shteti i të cilit i plotëson të gjitha kushtet.

Query 3: Listo të gjitha shkretëtirat të cilat gjenden në Etiopi dhe janë më të mëdha se 150000m².

Këtë query e kemi implementuar në XPath dhe e gjeni në Aneksin A, kurse teknikisht se çfarë bën ky XPath është:

- Së pari selektohen të gjitha elementet <desert> kudo në dokumentin XML,
- Më pas kontrollojmë nëse atributi @country e përmabn codeCountry ETH dhe kontrolluam nëse vlera e elementit child <area> është më e madhe se 150000,
- Nga shkretëtirat të cilat plotësojnë kushtet e mësipërme kthehet vetëm emri i shkretëtirave përkatëse.

Query 4: Listo të gjitha aeroportet në perëndim të Shteteteve të Bashkuara dhe kanë lartësi mbidetare më të vogël se 50m.

Këtë query e kemi implementuar në XPath dhe e gjeni në Aneksin A, kurse teknikisht se çfarë bën ky XPath është:

- Kontrollojmë tërë dokumentin XML për elementin <airport> dhe më pas filtrojmë vetëm aeroportet të cilat e kanë atributin @country = 'USA',
- Vendosim kushtin që vlera e elementit child <longitude> të jetë më e vogël se -120 (koordinata gjeografike në hemisferën perëndimore),
- Vendosim kushtin që lartësia e aeroportit të jetë më pak se 50m mbi nivelin e detit tek child elementi <elevation> dhe në rezultat më pas kthejmë vetëm emrin e aeroportit.

Query 5: Listo emrat e vendeve në Evropë që kanë pasur një popullësi më të madhe se 10M banorë në vitin 2011.

Këtë query e kemi implementuar në XPath dhe e gjeni në Aneksin A, kurse teknikisht se çfarë bën ky XPath është:

- Zgjedhen të gjitha elementet <country> që ndodhen në kontinentin Evropë (kontrolluar përmes elementit encompassed),
- Pastaj filtrohen vetëm ato vende që kanë popullsi më të madhe se 10 milionë banorë për vitin 2011,
- Përfundimisht, kthehet emri i secilit vend që i plotëson këto kushte.

3.2 BI tool visualization

3.2.1 Përzgjedhja e tool-eve

Në përzgjedhjen e punojmë me tool-in Power BI dhe Looker Studio, meqenëse këto dy tools i kemi demonstruar edhe në ushtrime si dhe filtrimet dhe slicers janë më se mjaftueshëm të avancuara për të kryer kërkesat e projektit, integrimi i të dhënave mund të bëhet nga shumë burime të ndryshme në të parën, e në të dytën mjafton të kemi burime google based kurse publikimi mund të bëhet në format të PDF ose si link i thjeshtë në web.

3.2.2 Lidhja e datasources

Në Power BI i kemi importuar gjithsej 4 burime të ndryshme, databazën Mondial, datasetin e importuar nga detyra e parë që përmban 2 milionë rreshta me të dhëna në një tabelë të vetme dhe përmban të dhëna rreth temperatures mesatare ditore nëpër qytete të mëdha për çdo ditë nga viti 1995 deri në vitin 2020. Këtu kemi importuar edhe një view e cila korrespondon me një nga queryt e përcaktuara për grupin tonë. Kurse dataseti tjetër ka gjithsej 4 tabela dhe përmban të dhëna për COVID-19 rastet pozitive dhe raste të vdekjes nëpër shtete.

Kurse në Looker Studio kemi importuar vetëm datasetin e përdorur nga detyra e parë dhe datasetin tjetër që përmban të dhëna rreth COVID-19, si dhe një dataset me të dhëna për studentët, i cili përfshin informacione si nota, orë studimi, gjinia, gjendja mendore, gjumi, pjesëmarrja dhe të dhëna të tjera që ndihmojnë në analizimin e suksesit dhe mirëqenies së tyre.

3.2.3 Krijimi i vizualizimeve

Vizualizimi për temperaturat vjetore në qytetet kryesore - Power BI

Ky vizualizim paraqet ndryshimet e temperaturave mesatare vjetore në qytete të ndryshme nga viti 1995 deri në vitin 2020. Qëllimi kryesor ishte të vëzhgohet trendi i temperaturës për çdo qytet dhe të identifikohen rritje apo ulje të mundshme gjatë viteve.

Elementet e përdorura për vizualizim:

1. Column chart (vertikal) - paraqet temperaturën mesatare të qytetit për secilin vit nga 1995 deri më 2020, lejon krahasim të drejtpërdrejtë ndër vite për një qytet të vetëm

2. Line chart - ilustron trendin e temperaturës mesatare në mënyrë të vazhdueshme, për të vëzhguar më qartë luhatjet dhe rritjet graduale apo të papritura, ky chart ka një titull dinamik që ndryshon në varësi të qytetit të përzgjedhur në slicer.
3. Slicer - i lejon përdoruesit të zgjedh një qytet nga lista për të pasur një analizë më të përqendruar, përdorimi i slicer-it e bën vizualizimin më interaktiv dhe më të qartë.

Kurse measure e përdorur për titullin dinamik në line chart duke e bërë këtë chart më dinamik rreth qytetit të përzgjedhur është një DAX me përmbajtjen e paraqitur në Aneksin A, ku kjo measure e shfaq automatikisht emrin e qytetit që është përzgjedhur në slicer.

Në figurën e mëposhtme paraqitet ky vizualizim, kurse në zip folder e gjeni me emrin PowerBIVisualization.pbix.

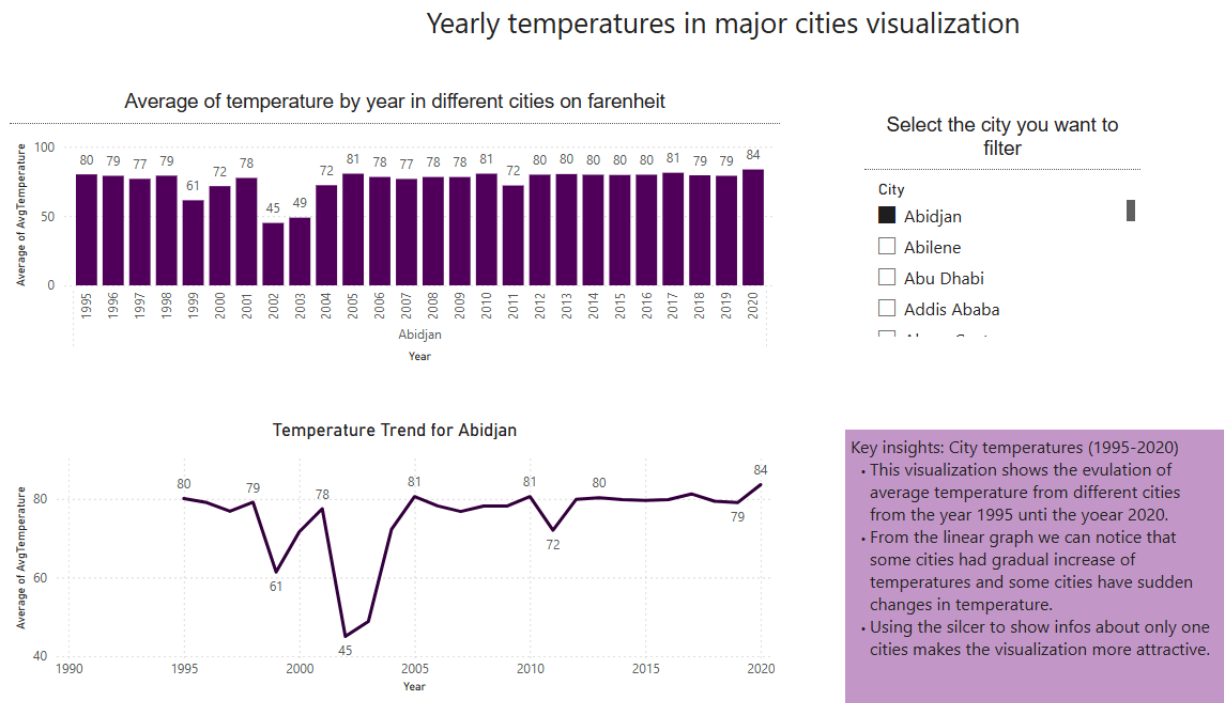


Figura 2: Vizualizimi rreth temperaturave mesatare vjetore.

Vizualizimi i numrit të liqeneve për shtet - Power BI

Ky vizualizim paraqet shpërndarjen e liqeneve sipas shteteve, duke u bazuar në të dhëna të importuara nga databaza Mondial. Qëllimi i vizualizimit është të identifikojë shtetet me më shumë apo më pak liqene dhe të analizojë përqindjen që ato zënë në krahasim me shtet tjera.

Elementet e përdorura për vizualizim:

1. Column chart (horizontal) - paraqet numrin e liqeneve për secilin shtet, lehtëson krahasimin vizual midis vendeve me më shumë ose më pak liqene.
2. Pie chart - tregon përqindjen e liqenve që ka secili shtet në raport me totalin, jep një pamje të përgjithshme të shpërndarjes së burimeve ujore të liqeneve.

3. Cards (KPI) - kemi krijuar 4 karta të cilat shfaqin shtetin e më së shumti liqene, numrin përkatës të liqeneve për këtë shtet, shtetin me më së paku liqene, numrin përkatës të liqeneve për këtë shtet. Për krijimin e këtyre cards kemi krijuar 4 measures të cilat i kemi paraqitur në Aneksin A.

Në figurën e mëposhtme paraqitet ky vizualizim, kurse në zip folder e gjeni me emrin PowerBIVisualization.pbix.

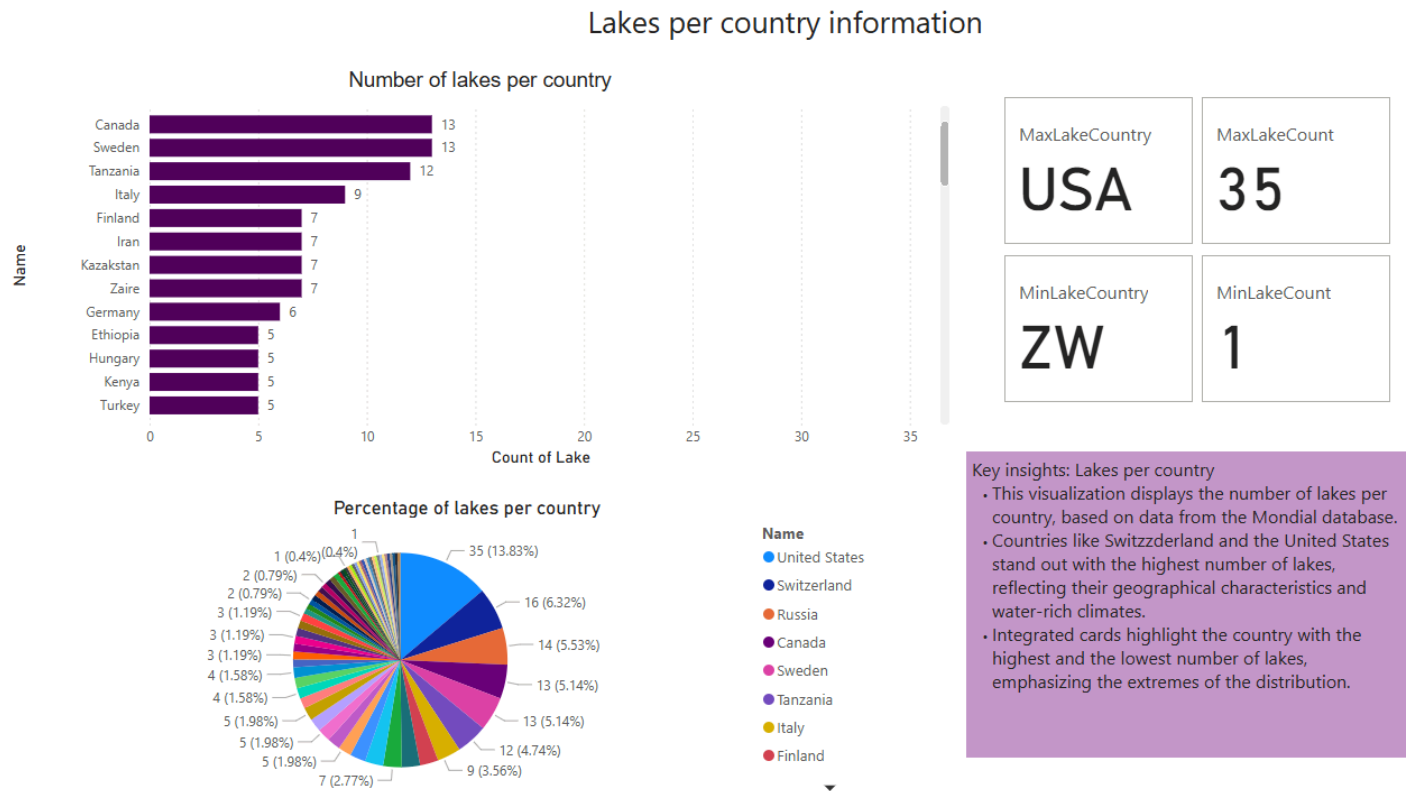


Figura 3: Vizualizimi rreth numri të liqeneve për shtete.

Lumenjtë sipas shteteve (të filtruara nga një view) - Power BI

Për këtë vizualizim kemi përdorur një view paraqet shpërndarjen e lumenjëve sipas vendeve që plotësojnë tri kritere të specifikuar: nuk janë anëtare të NATO-s, kanë dalje në det dhe ka më shumë se 10 lumenjë që kalojnë në këto shtete. Kjo view është paraqitur në Aneksin A.

Elementet e përdorura për vizualizim:

1. Donut chart - paraqet përqindjen e lumenjëve që i përkasin shtetit të filtruar, në këtë rast kemi vetëm dy shtete që përmbushin kriteret duke përfaqësuar 100% të të dhënave të filtruara.
2. Table - shfaqen emrat e lumenjve dhe shtetin përkatës se ku kalojnë, përdoret për të ofruar informacion të detajuar mbi të dhënat e përfshira në grafik.

3. Slicer - lejon përdoruesin të filtrojë të dhënat për vetëm një shtet dhe fokusohet në lumenjtë e tij, ky slicer ndikon edhe në tabelë edhe në grafik.

Në figurën e mëposhtme paraqitet ky vizualizim, kurse në zip folder e gjeni me emrin PowerBIVisualization.pbix.

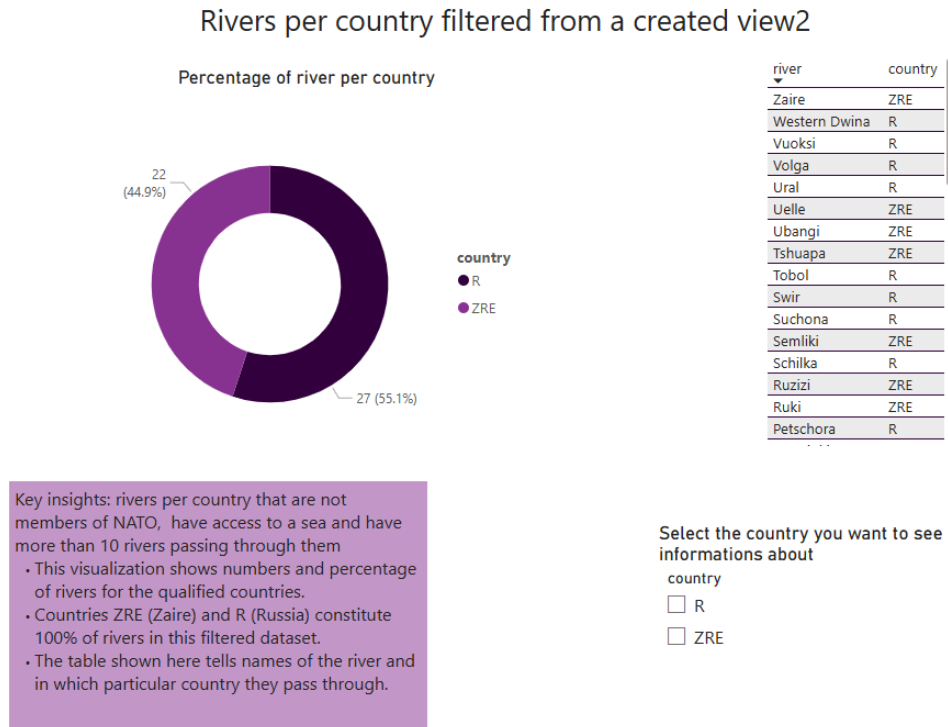


Figura 4: Vizualizimi për lumenjtë që kalojnë nëpër shtetet të cilat i plotësojnë kushtet e lartpërmendura.

Përmbledhje globale e COVID-19 - Rastet e konfirmuara, të shëruarat dhe vdekjet sipas shteteve - Looker Studio

Ky vizualizim paraqet një analizë të situatës globale të COVID-19 duke përdorur të dhëna të detajizuara për numrin e rasteve të konfirmuara, të shëruara, aktive dhe numrin e vdekjeve sipas shteteve.

Elementet e përdorura për vizualizim:

1. Table - rendit shtetet dhe informatat me kolona të rasteve të konfirmuara, të vdekjeve, të shëruara dhe aktive.
2. Column chart - tregon 10 shtetet me më së shumti raste të konfirmuara.
3. Pie chart - tregon përqindjen e vdekjeve të konfirmuara nga 10 shtetet më të prekura.
4. Cards - tregon numrin total të rasteve të shëruara në botë.

Në figurën e mëposhtme është paraqitur ky vizualizim, kurse në zip folder e gjeni me emrin LookerStudioVisualization.pdf.

Pandemic Summary by Country – Confirmed, Recovered, Active & Deaths

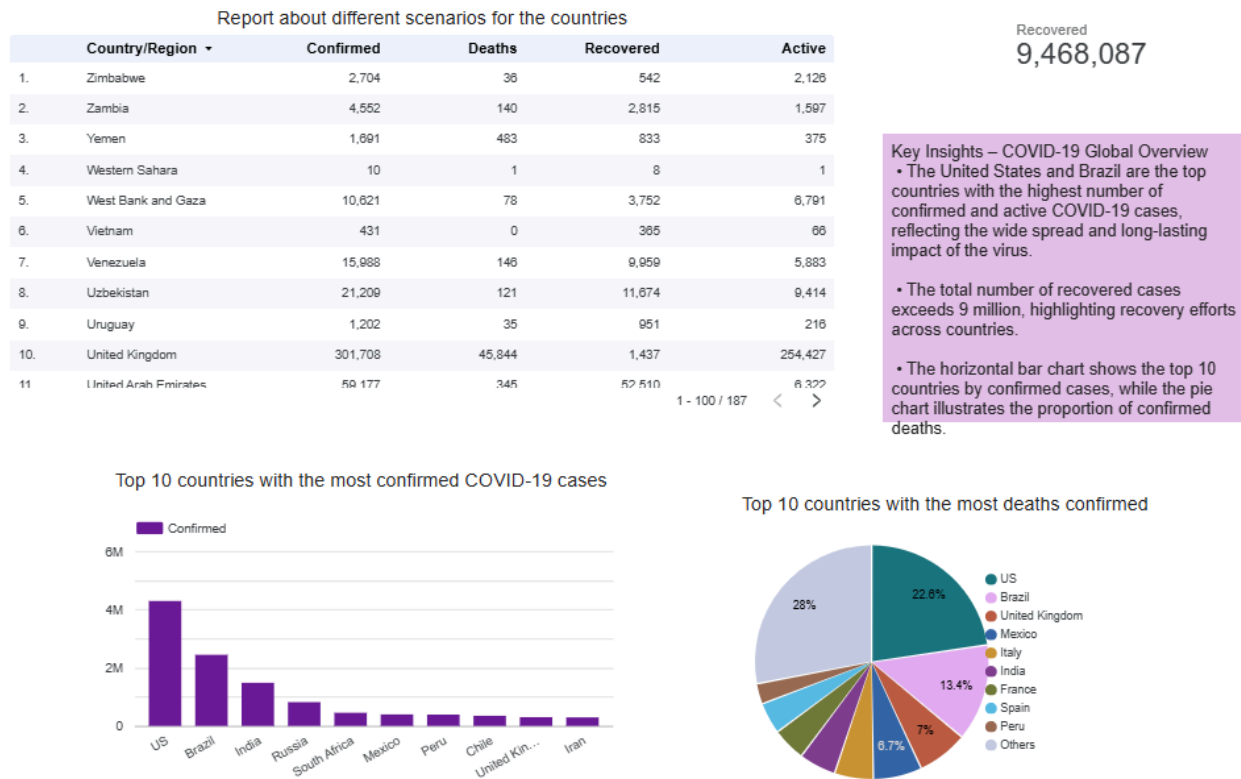


Figura 5: Vizualizimi në lidhje me informata të shteteve dhe COVID-19.

Analiza vizuale e faktorëve që ndikojnë në jetën studentore - Looker Studio

Ky vizualizim paraqet një analizë gjithëpërfshirëse të faktorëve që ndikojnë në suksesin akademik dhe mirëqenien e studentëve, duke përfshirë të dhëna për performancën në provime, stilin e jetesës dhe shprehitë e studimit.

Elementet e përdorura për vizualizim:

1. Table – rendit studentët më të suksesshëm bazuar në notat, orët e studimit, pjesëmarrjen dhe përqindjen e pjesëmarrjes në mësim.
2. Bar chart – analizon ndikimin e cilësisë së ushqimit në shëndetin mendor dhe gjumin e studentëve.
3. Pie chart – paraqet përqindjen e studentëve sipas gjinisë (Meshkuj, Femra, Tjetër).
4. Bar chart – tregon ndikimin e nivelit të arsimit të prindërve në suksesin e studentëve.

Në figurën e mëposhtme është paraqitur ky vizualizim, ndërsa dokumenti përkatës ndodhet në folderin ZIP me emrin LookerStudioVisualization1.pdf.

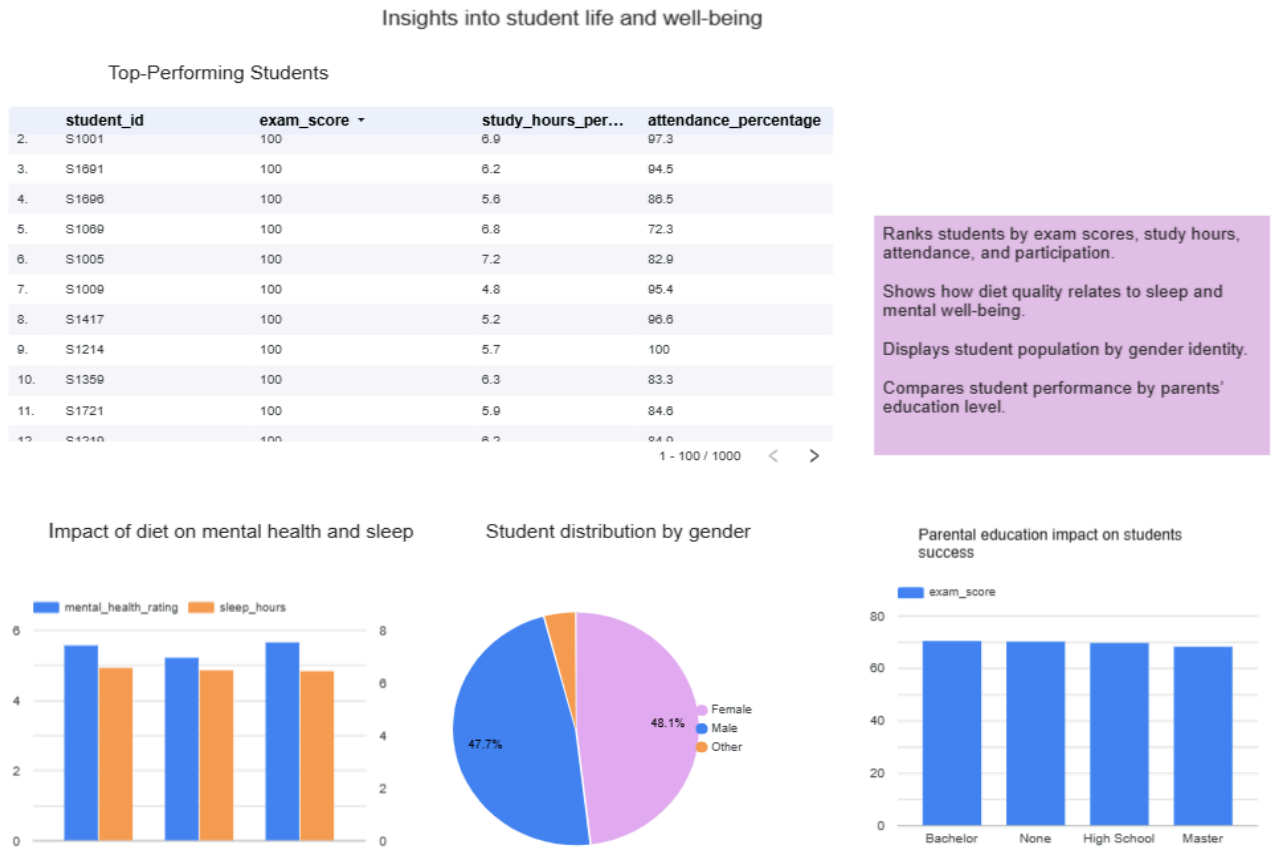


Figura 6: Vizualizimi në lidhje me studentët dhe performancën e tyre.

4. Sfidat dhe zgjidhjet

Gjatë realizimit të këtij projekti, ekipi është përballur me disa sfida teknike dhe praktike, të cilat janë kapërcyer me analiza të kujdesshme dhe zbatim të zgjidhjeve efektive. Më poshtë janë përshkruar sfidat kryesore dhe mënyrat e adresimit të tyre:

1. Validimi i dokumentit XML me xmllint në Windows.

Sfidë:

Instalimi i xmllint në sistemin Windows kërkonte manualisht tri paketa të ndara dhe konfigurim korrekt të path-it.

Zgjidhje:

U shkarkuan dhe vendosën në një folder të përbashkët tri komponentët: libxml2, zlib, dhe iconv. U testua suksesshëm validimi duke përdorur komandën `xmllint --noout --schema mondial.xsd mondial.xml`.

2. Zhvillimi i query-ve komplekse me XQuery/XPath.

Sfidë:

Formulimi i query-ve që kombinojnë shumë kushte (si shtetet që kanë dalje në det, nuk janë në NATO dhe kanë më shumë se 10 lumenj) kërkonte menaxhim të kujdesshëm të kushteve.

Zgjidhje:

Query-t u ndanë në pjesë logjike (modularizim), dhe u testuan pjesë-pjesë në BaseX GUI, duke përmirësuar gabimet sintaksore ose logjike përpara bashkimit të plotë të tyre.

5. Konkluzione

Në këtë punim u trajtua cikli i plotë i manipulimit dhe vizualizimit të të dhënave XML duke përdorur mjete të ndryshme si xmlint për validim, BaseX për manipulim të të dhënave me XQuery/XPath, si dhe Power BI dhe Looker Studio për krijimin e vizualizimeve të avancuara. Gjatë punës u importua me sukses databaza Mondial, u zhvilluan pyetje të personalizuar mbi temat si liqenet, lumenjtë, malet dhe shtetet anëtare të NATO-s, dhe më pas u vizualizuan të dhënat përmes grafikëve të ndryshëm (pie, column, line chart etj.).

U arrit qëllimi i punimit duke kombinuar teknika të validimit, analizës dhe prezantimit të të dhënave në mënyrë të qartë dhe interaktive. U ndërtuan vizualizime të dobishme për analiza klimatike, gjeografike dhe statistikore mbi rastet COVID-19, me qëllim që të tregohet fuqia e këtyre teknologjive në projekte reale të të dhënave.

Në të ardhmen mund të zgjerohet analiza me më shumë skema XSD dhe validime më të thelluara semantike. Po ashtu, mund të integrohen më shumë burime të dhënash heterogjene, dhe të ndërtohen dashboards dinamike me interaktivitet më të madh duke përdorur DAX më të avancuar dhe parametrizime në Looker Studio.

Referencat

- <https://basex.org/download/>
- <https://www.dbis.informatik.uni-goettingen.de/Mondial/>
- <https://www.zlatkovic.com/pub/libxml/iconv-1.9.2.win32.zip>
- <https://www.zlatkovic.com/pub/libxml/zlib-1.2.5.win32.zip>
- <https://www.zlatkovic.com/pub/libxml/libxml2-2.7.8.win32.zip>
- <https://www.dbis.informatik.uni-goettingen.de/Mondial/mondial.xsd>
- <https://www.kaggle.com/datasets/imdevskp/corona-virus-report>
- <https://www.kaggle.com/datasets/sudalairajkumar/daily-temperature-of-major-cities>
- <https://www.kaggle.com/datasets/jayaantanaath/student-habits-vs-academic-performance>