**Universiteti i Prishtinës**

**Fakulteti Inxhinierisë Elektrike dhe Kompjuterike**



**Dokumentim teknik i projektit**

**Lënda: Big Data**

**Homework 3: Data Warehouse Design & BI Tool Visualization**

**Emri profesorit Emri & mbiemri studentëve / email adresa**

| Prof. Vigan Raca | 1. Gyltene Sfishta | gyltene.sfishta[@student.uni-pr.edu](mailto:filan.fisteku@student.uni-pr.edu) |
| --- | --- | --- |
| 2. Klajdi Gashi | klajdi.gashi[@student.uni-pr.edu](mailto:filan.fisteku@student.uni-pr.edu) |
| 3. Kleda Gashi | kleda.gashi[@student.uni-pr.edu](mailto:filan.fisteku@student.uni-pr.edu) |
| 4. Myhedin Vuciterna | myhedin.vuciterna[@student.uni-pr.edu](mailto:filan.fisteku@student.uni-pr.edu) |
|  | 5. Rinesa Hoxha | rinesa.hoxha1[@student.uni-pr.edu](mailto:filan.fisteku@student.uni-pr.edu) |

Prishtinë, 2024

Përmbajtja

[**Abstrakti 3**](#_30j0zll)

[**I. Hyrje 4**](#_1fob9te)

[**II. Qëllimi i punimit 5**](#_3znysh7)

[**III. Pjesa kryesore 5**](#_2et92p0)

[**IV. Konkluzione 9**](#_tyjcwt)

[**Referencat 9**](#_3dy6vkm)

# **Abstrakti**

Në këtë dokumentim do të përshkruhet mënyra e projektimit dhe implementimit të një ***Data Warehouse (DWH)*** që përmban tabela faktike dhe tabela dimensionale, duke përdorur indekse në tabelat dimensionale dhe automatizimin e procesit të dataflow nga databaza burimore në DWH.

Ky dokumentim do të trajtojë krijimin e skemës relacionale dhe zgjedhjen e tipit të skemës, krijimin e databazës target (***DWH***), dhe automatizimin e procesit të mbushjes së të dhënave.

Gjithashtu, do të përfshihet përdorimi i një mjeti të ***Business Intelligence (BI)*** për të krijuar vizualizime të avancuara të të dhënave. Ky proces përfshin lidhjen e disa burimeve të të dhënave, krijimin e grafeve dhe dashboard-eve të ndryshme.

Dokumentimi synon të ofrojë një përmbledhje të plotë të teknologjive të përdorura, procesit të zhvillimit dhe rezultateve të arritura gjatë kësaj eksperience.

# **Hyrje**

Një ***Data Warehouse (DWH)*** është një sistem që grumbullon, ruan dhe menaxhon të dhënat nga burime të ndryshme për t’i përdorur në analizat e biznesit dhe raportim. Ndryshe nga databazat operative që janë të optimizuara për përpunimin e transaksioneve të përditshme, një ***DWH*** është i optimizuar për leximet e mëdha dhe analizat komplekse.

Si karakteristika kryesore dallojmë:

* **Integrimi i të dhënave**: kombinon të dhëna nga burime të ndryshme në një strukturë uniforme.
* **Historiku i të dhënave:** ruajnë të dhëna historike për të mundësuar analizat afatgjata.
* **Konsistenca:** Përmirëson konsistencen e të dhënave duke përdorur një skemë unike për të gjitha burimet.
* **Performanca:** optimizohet për lexime të shpejta dhe analiza komplekse/
* **Analiza dhe raportimi:** mbështet analiza dhe raportime të avancuara për marrjen e vendimeve.

Vizualizimi i të dhënave me mjete të ***Business Intelligence (BI)*** përfshin përdorimin e softuerëve për të transformuar të dhënat e papërpunuara në grafikë, tabela. Këto mjete i’u ndihmojnë përdoruesëve të kuptojnë dhe analizojnë të dhënat në mënyrë të lehtë dhe intuitive.

Si karakteristika kryesore dallojmë:

* **Vizualizimi i të dhënave:** transformon të dhënat në grafika të ndryshme.
* **Analiza e të dhënave:** lejon përdoruesit të bëjnë analiza komplekse dhe të identifikojnë trende, modele dhe anomalitë.
* **Interaktiviteti:** përdoruesit mund të ndërveprojnë me vizualizimet për të eksploruar të dhënat në mënyrë më të detajuar.
* **Integrimi i burimeve të ndryshme të të dhënave:** lidh të dhënat nga burime të ndryshme.
* **Raportimi:** gjeneron raporte që mund të ndahen dhe publikohen për vendimmarrje të bazuar në të dhëna.

# **Qëllimi i punimit**

Projekti ka për qëllim krijimin e një ***Data Warehouse*** ***(DWH)*** që përmban tabela faktike dhe tabela dimensionale, me indeksim të përshtatshëm për dimensionet. Procesi i dataflow do të automatizohet për të mbushur të dhënat nga databaza burimore në DWH. Po ashtu ka për qëllim vizualizimin e të dhënave duke përdorur mjete ***BI (Business Intelligence)*** për të krijuar grafika dhe dashboard-e informative.

# **Pjesa kryesore**

Hapat e ndjekur për implementimin e ***DWH***:

1. **Krijimi i Source DB**

* Krijimi i një databaze e cila përmban më shumë se 20 tabela/objekte.
* Databaza e zgjedhur është e emërtuar: **mondial**
* Sigurimi se databaza ka të dhëna të mjaftueshme dhe sktruktura të ndryshme.

1. **Nxjerrja e Skemës Relacionale**

* Identifikimi i tabelave që do të përdoren për ndërtimin e Fact Tables dhe Dimension Tables.
* Zgjedhja e tipit të skemës

1. **Krijimi i Target DB (DWH)**

* Krijimi i një databaze të re e cila ka emrin me prapashtesën **DWH**
* Databaza e krijuar është e emërtuard: **dwh\_mondial**
* Krijimi i tabelave faktike dhe dimensionale
* Emërtimi i tabelave në mënyrë që të tregojnë entitetin e tabelës:
* Query për DWH: query.txt e ruajtur në zip file

1. **Automatizimi i DataFlow**

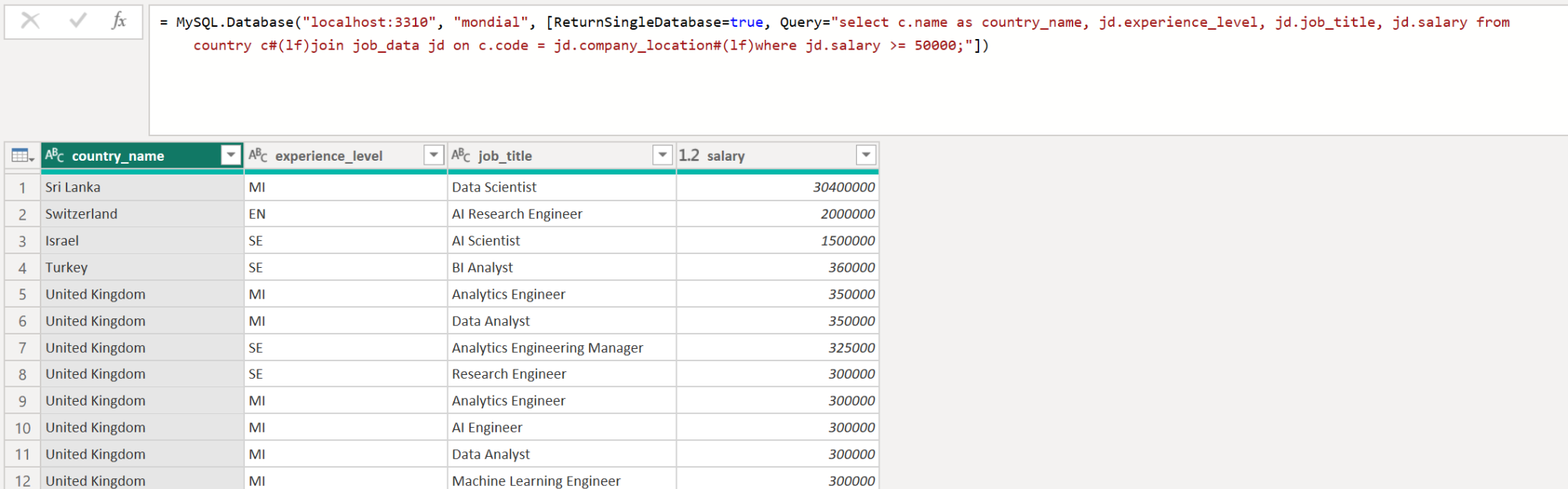
* Implementimi i procesit të dataflow për të mbushur të dhënat nga Source DB në DWH.
* Konfigurimi i dataflow në mënyrë që të ekzekutohet për qdo 5 ose 10 minuta.
* Përdorimi i **ETL** **(Extract, Transform, Load)** mjete për automatizimin e procesit.

| import mysql.connector import time  # Database connection details source\_db\_config = {  'user': 'root',  'password': '1234',  'host': 'localhost',  'database': 'mondial',  'port': 3310 }  dwh\_db\_config = {  'user': 'root',  'password': '1234',  'host': 'localhost',  'database': 'dwh\_mondial',  'port': 3310 }  def transfer\_data():  # Connect to source and DWH databases  source\_conn = mysql.connector.connect(\*\*source\_db\_config)  dwh\_conn = mysql.connector.connect(\*\*dwh\_db\_config)    source\_cursor = source\_conn.cursor(dictionary=True)  dwh\_cursor = dwh\_conn.cursor()   source\_cursor.execute("SELECT \* FROM country")  for row in source\_cursor:  dwh\_cursor.execute("""  INSERT INTO DimensionCountry (CountryCode, CountryName,Capital, Area, Population)  VALUES (%s, %s, %s, %s, %s)  ON DUPLICATE KEY UPDATE CountryName=%s, Capital=%s, Area=%s, Population=%s  """, (row['Code'], row['Name'], row['Capital'], row['Area'], row['Population'],  row['Name'], row['Capital'], row['Area'], row['Population']))    source\_cursor.execute("SELECT \* FROM province")  for row in source\_cursor:  dwh\_cursor.execute("""  INSERT INTO DimensionProvince (ProvinceName, CountryCode, Population, Area, Capital)  VALUES (%s, %s, %s, %s, %s)  ON DUPLICATE KEY UPDATE CountryCode=%s, Population=%s, Area=%s, Capital=%s  """, (row['Name'], row['Country'], row['Population'], row['Area'], row['Capital'],  row['Country'], row['Population'], row['Area'], row['Capital']))   source\_cursor.execute("SELECT \* FROM city")  for row in source\_cursor:  dwh\_cursor.execute("""  INSERT INTO DimensionCity (CityName, CountryCode, ProvinceName, Population, Longitude, Latitude)  VALUES (%s, %s, %s, %s, %s, %s)  ON DUPLICATE KEY UPDATE CountryCode=%s, ProvinceName=%s, Population=%s, Longitude=%s, Latitude=%s  """, (row['Name'], row['Country'], row['Province'], row['Population'], row['Longitude'], row['Latitude'],  row['Country'], row['Province'], row['Population'], row['Longitude'], row['Latitude']))   source\_cursor.execute("""  SELECT c.Code AS CountryCode, ct.Name AS CityName, p.Name AS ProvinceName, ct.Population, c.Area  FROM city ct  JOIN country c ON ct.Country = c.Code  JOIN province p ON ct.Province = p.Name  """)  for row in source\_cursor:  dwh\_cursor.execute("""  INSERT INTO FactPopulation (CountryCode, CityName, ProvinceName, Population, Area)  VALUES (%s, %s, %s, %s, %s)  ON DUPLICATE KEY UPDATE Population=%s, Area=%s  """, (row['CountryCode'], row['CityName'], row['ProvinceName'], row['Population'], row['Area'],  row['Population'], row['Area']))   dwh\_conn.commit()  source\_cursor.close()  dwh\_cursor.close()  source\_conn.close()  dwh\_conn.close()  print("Data transfer completed.")  while True:  transfer\_data()  time.sleep(86400 / 144) # 10 minutes |
| --- |

*Skripta e përdorur për procesin e automatizimit*

1. **Lidhja e Datasources në BI Tool**

* Lidhja e datasource:
  + Mondial DB
  + Imported DB
  + Një datasource tjetër me të paktën 10,000 rekorde.
  + Query

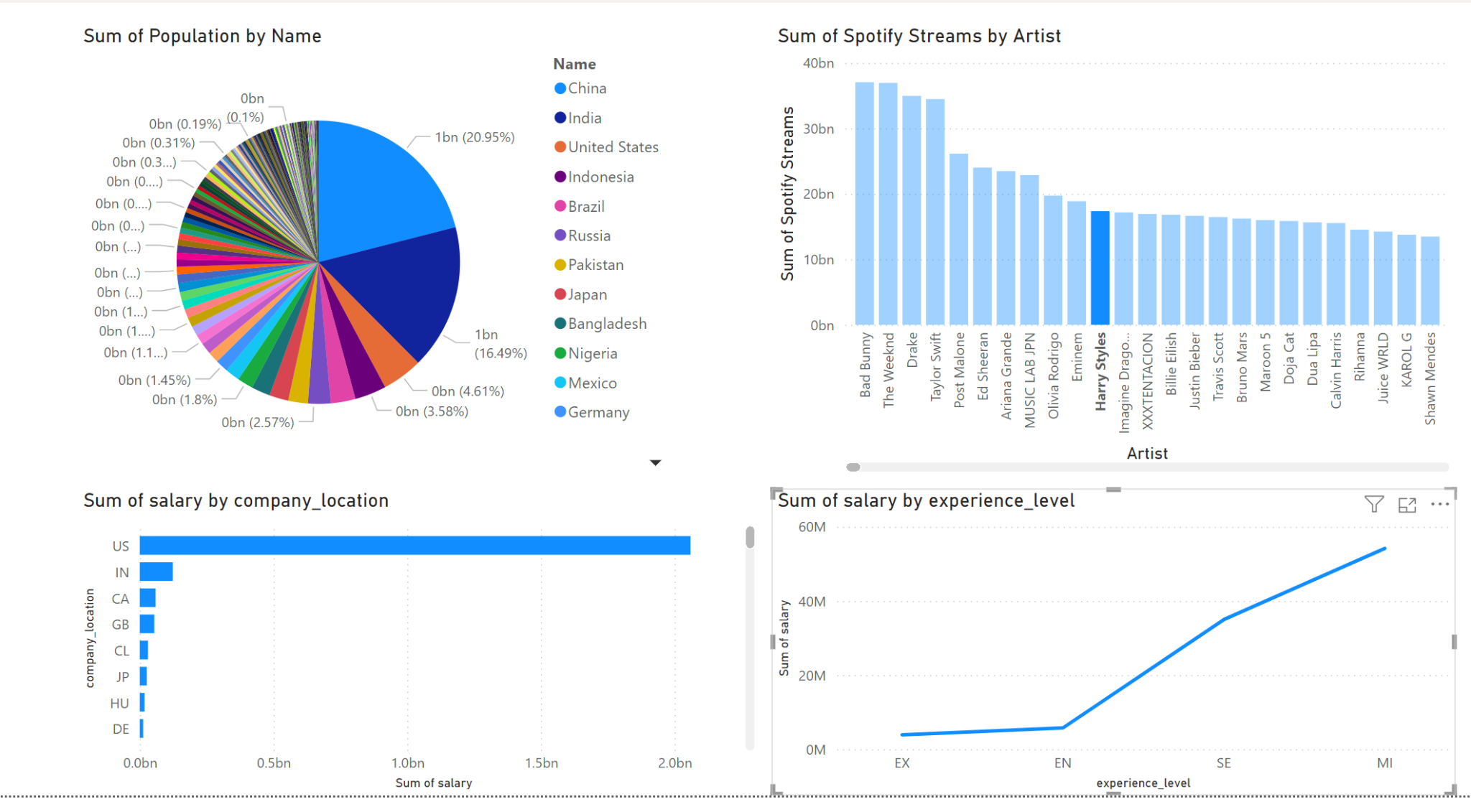
**select c.name as country\_name, jd.experience\_level, jd.job\_title, jd.salary from country c#(lf)join job\_data jd on c.code = jd.company\_location#(lf)where jd.salary >= 50000;**

1. **Krijimi i Grafeve**

* Krijimi i grafeve për secilin dataset.
* Grafet janë të llojeve të ndryshme si: **Chart Column Vertical, Chart Column Horizontal, Pie, Linear Graphs.**

1. **Dokumentimi dhe Ruajtja**

Ruajtja e grafit si .**pdf** ne zip file te derguar dhe figura me grafet:

****

*Figura [1]*  Paraqitja e grafit të punuar me **BI**.

# **Konkluzione**

Pas përfundimit të këtij projekti dhe pas përdorimit të një sërë metodash për zhvillimin e tij, janë arritur disa perfundime të rëndësishme. Përdorimi i një game të gjerë metodash ka kontribuar në pasurimin e përmbajtjes së projektit dhe në sigurimin e një analize të tij.

Përdorimi i metodave hulumtuese dhe shpjeguese ka lejuar ekipin të kuptojë thellësisht konceptet dhe teknologjitë që janë përdorur në projekt.

# **Referencat**

Të gjitha informatat e përdorura në këtë dokumentim janë autentike dhe bazuar në ligjërata dhe ushtrime.

* (*What Is a Data Warehouse?*, n.d.)
* (FRANKENFIEL, JAKE. 2024. *What Is Business Intelligence (BI) Types, Benefits, and Examples*.)