



Mise en place d'un indicateur de santé globale " WHS INDICATOR "

Réalisé par : Mr Anas Mohammadi

Membres du jury: Mr. Kamal Omari
Mr . Mohamed Atibi

Soutenue le: 03/10/2022



Plan

METHODOLOGIE

ARCHITECTURE

ALIMENTATION/ETL

CONCLUSION ET PERSPECTIVES



PLANNING
PREVISIONNEL

MODELISATION

REPORTING

Problématique

Pour l'OMS, Les différents indicateurs de la santé mondiale sont divisés sur plusieurs catégories. Ainsi on se retrouve avec plusieurs sources de données réparties sur des tables mais certes ca reste incompréhensible pour la vaste majorité du public

Suite de la **Problématique**

1


L'Indicateur de santé

Comment pourrions nous mesurer l'état de la santé mondiale avec ces différents indicateurs ?

2

Source des informations

Si la source l'OMS(WHO) est infidèle, que serait considéré comme des sources fiables ?



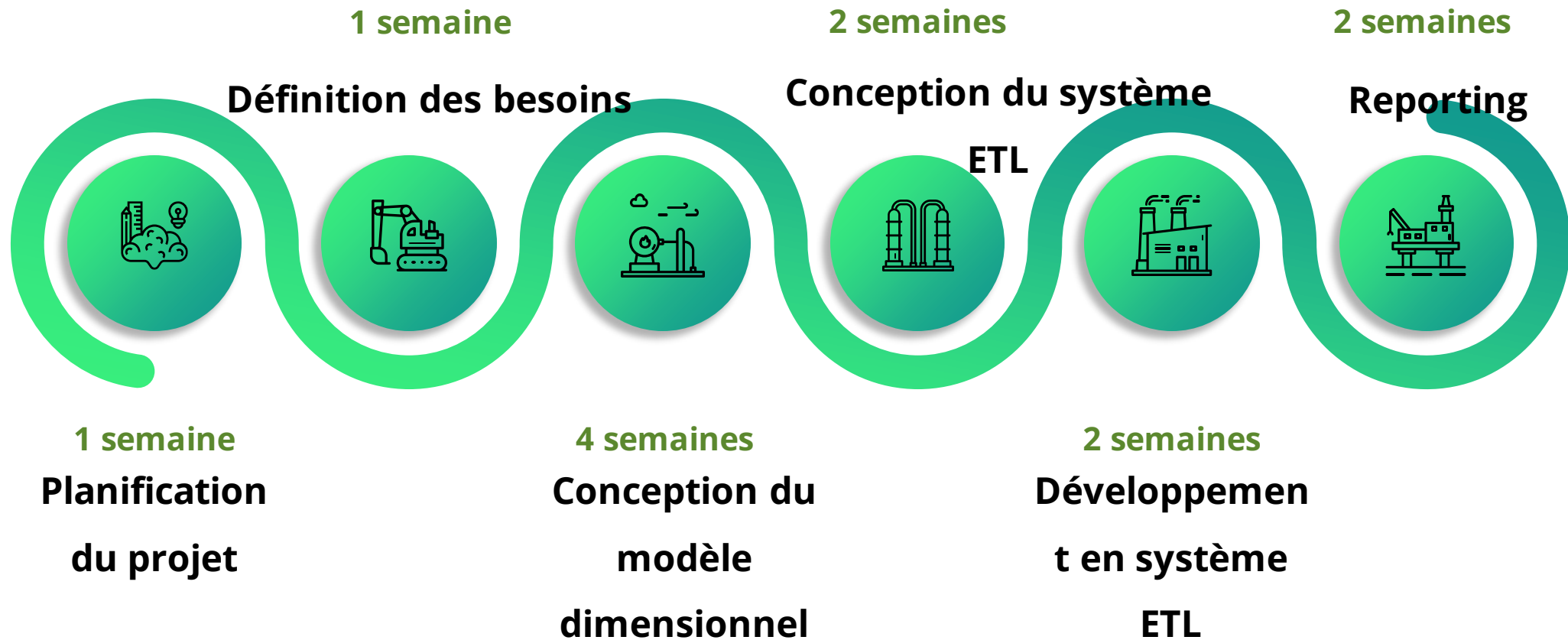
L'entrepôt de données concentre pour nous toutes les informations pour tous les thèmes de l'entrepôt. Dans une structure décisionnelle classique, les magasins sont thématiques.

Méthodologie

Notre étude se compose de données provenant du passé et qui ne changeront pas pour le moment où on mène cette étude. Cela nous permet de suivre l'approche **Kimball**.

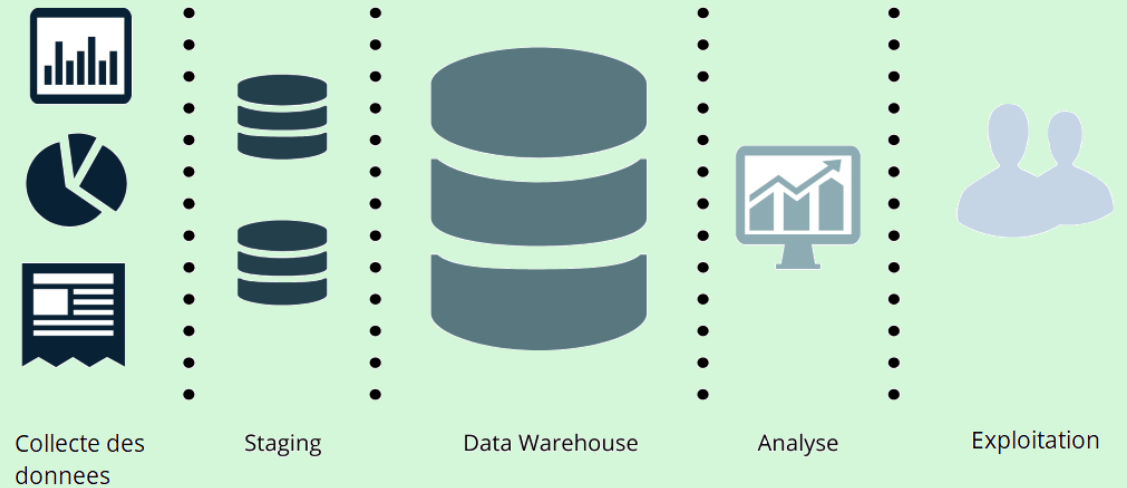
PLANNING PRÉVISIONNEL

*Ces durées étaient altérées pas mal de fois.



Architecture

Notre modèle étant celui Kimball; Les sources de données primaires sont ensuite exploitées, et un outil d'extraction, de transformation et de chargement (ETL) est utilisé pour récupérer les données de plusieurs sources et les charger dans une zone de staging du serveur de base de données relationnelle.



Architecture (SUITE)



Collecte des données

La collecte a été faite depuis des datasets public sur Kaggle.



Staging

C'est là où on crée des bases qui vont contenir les données brutes provenant des systèmes sources et fournir un espace pour les tableaux transitoires créés au cours du processus de transformation.



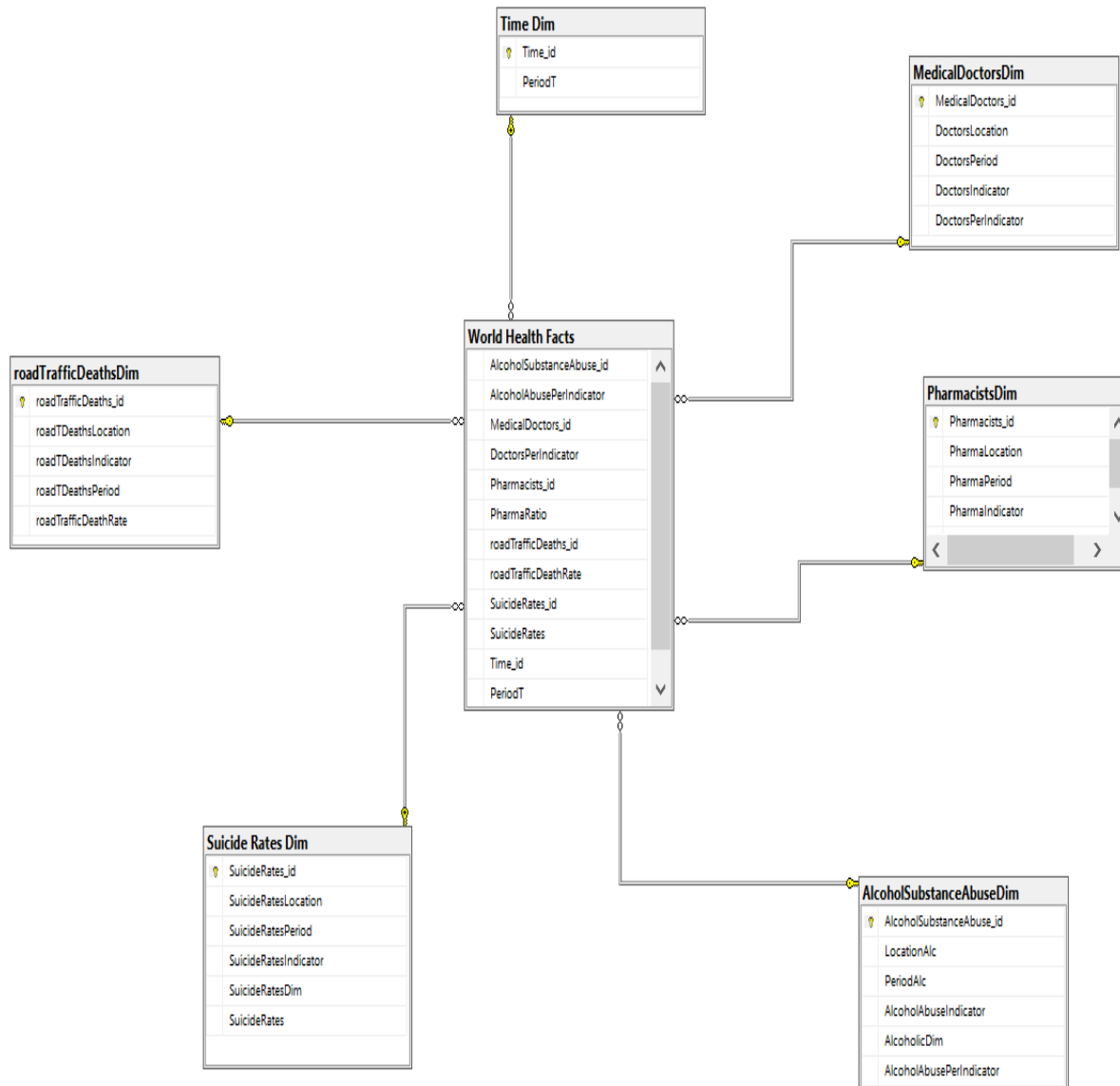
Datawarehouse

La phase de l'extraction, transformation et alimentation.



Analyse

Analyse ou reporting qui sera fait à travers Power BI.



MODELISATION

Table de fait : World Health Facts



Notre table de fait représente l'objet de l'analyse qui est l'usage des différents indicateurs de santé globale, notre mesure est le le taux d'abus d'alcool, de mortalité d'accidents de route , de suicide et le nombre de staff médical disponible.



SCHÉMA EN ÉTOILE

Il se compose du fait central (les indicateurs mentionnés dans le diapo 9) et de leurs dimensions. Dans ce schéma il existe une relation pour les faits et plusieurs pour les différentes dimensions autour de la relation centrale. La relation de faits contient les différentes mesures et une clé étrangère pour faire référence à chacune de leurs dimensions.

Cela
sert à :

①

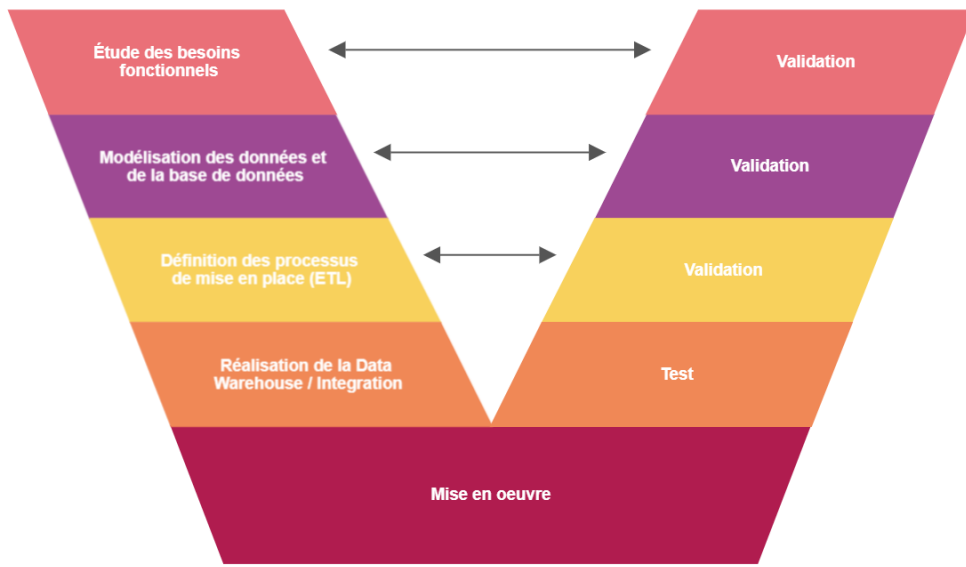
Faciliter la navigation

②

**Nombre de jointures
limité**

LE CYCLE DE DÉVELOPPEMENT

La conduite d'un projet informatique, tel que le développement d'un système d'information, fait appel à des méthodes formalisées.



Le cycle en
« V »

“

Après une étude exploratoire des méthodes de conduite de projet et pour répondre aux objectifs fixés en début, le cycle de développement en « V » s'est révélé le plus approprié pour ce travail.

”

ETL

Une solution ETL (Extract Transform Load) permet d'extraire les données d'une source de données et de les adapter pour les charger dans une autre.

Pandas

Pandas est une bibliothèque écrite pour le langage de programmation Python permettant la manipulation et l'analyse des données. Elle propose en particulier des structures de données et des opérations de manipulation de tableaux numériques et de séries temporelles.



ETL

Une solution ETL (Extract Transform Load) permet d'extraire les données d'une source de données et de les adapter pour les charger dans une autre.

Microsoft SQL Server

Microsoft SQL Server est un système de gestion de base de données relationnelle (SGBDR) qui prend en charge une grande variété d'applications de traitement des transactions, de veille stratégique et d'analyse dans les environnements informatiques des entreprises. Microsoft SQL Server est l'une des trois technologies de base de données leaders du marché, avec Oracle Database et DB2 d'IBM.



ETL (SUITE)

Le choix d'utiliser Python(Pandas) pour l'ETL s'est avéré plus pratique et fluide après plusieurs obstacles qui ont rendu la réalisation avec SSIS impossible et inutile d'y investir plus que le temps consacré

Loading the data to the DW

```
In [4]: listData = ['roadTrafficDeaths.csv', 'pharmacists.csv', 'alcoholSubstanceAbuse.csv', 'medicalDoctors.csv', 'crudeSuicideRates.csv']
for i in listData:
    myData = pd.read_csv(f"C:/Users/Mars/Desktop/ProjetBI/projectV3/{i}", sep=';', engine='python')
    #z.to_sql(i.split(".")[0].capitalize(), eng, if_exists='replace', index=False)

print("Adding records to the database is done successfully")
```

Adding records to the database is done successfully

Libraries Imports

Setting Up DB onnections and destinations files

#creating the tables

Extracting the dimensions , selecting only a limited number of rows due to the large size of the data and loading into the DW

```
In [1]: import pandas as pd
import pyodbc
from sqlalchemy import *
import matplotlib.pyplot as plt
svr_name = 'ANAS\YORUDEV'
#stagingDb= 'binatbit' and the datawarehouse = 'dw_v3'
# had to delete the staging db due to the disk usage and moved to working on the main db directly
db_name= 'dw_v3'

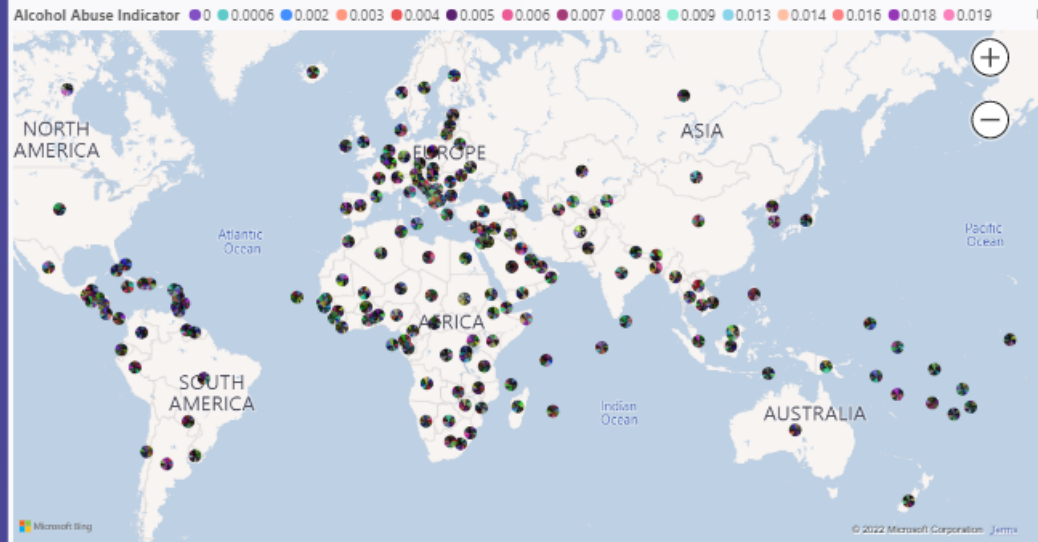
#Connect to SQL Server
conn = pyodbc.connect('Driver={SQL Server};'
                    'Server=ANAS\YORUDEV;'
                    'Database=dw_v3;'
                    'Trusted_Connection=yes;')

cursor = conn.cursor()
eng = create_engine("mssql+pyodbc://"+svr_name+"/"+db_name+"?driver=SQL+Server")
```

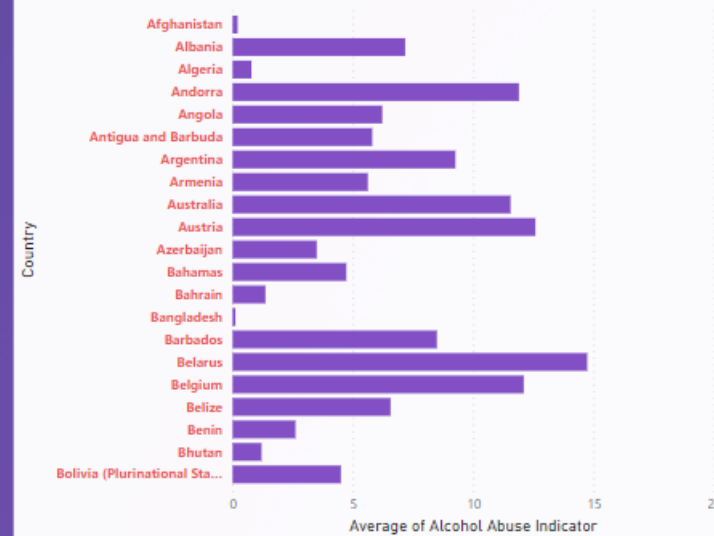
REPORTING

WHS Indicator

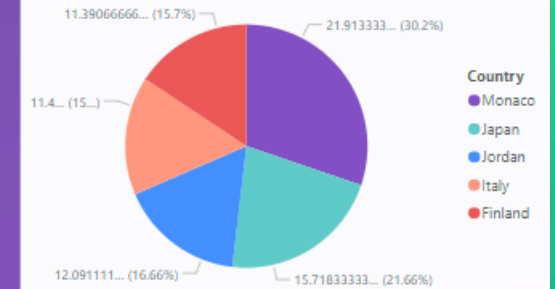
Average of Alcohol Abuse Indicator by Country and Year



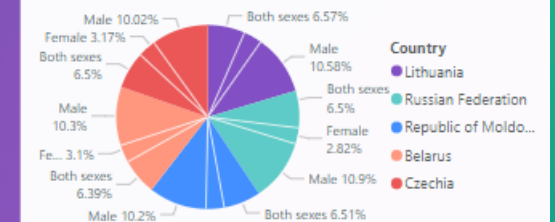
Average of Alcohol Abuse Indicator by Country



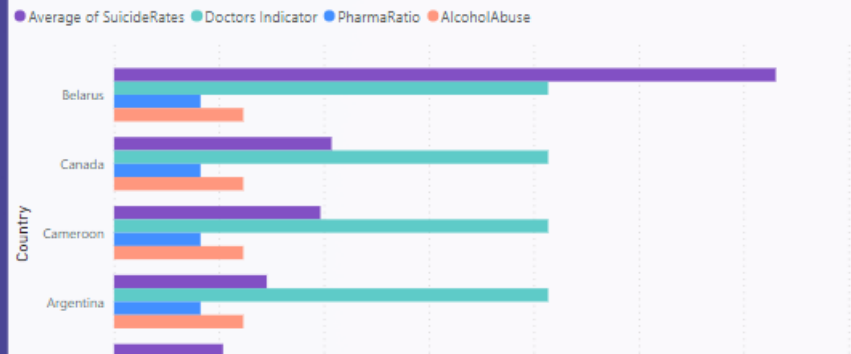
Average of Pharmacists by Country



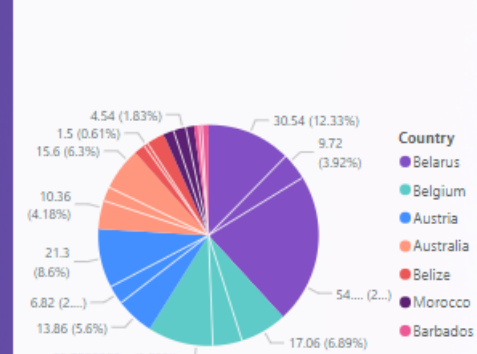
Average of Alcohol Abuse Indicator by Country and Gender



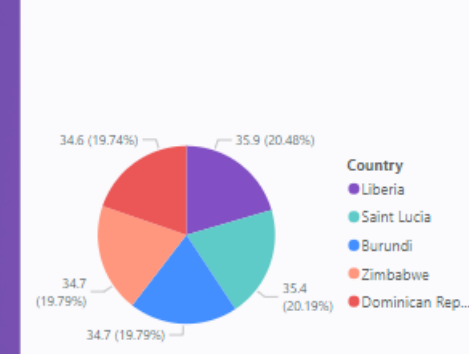
Average of SuicideRates, Doctors Indicator, PharmaRatio and AlcoholAbuse by Country



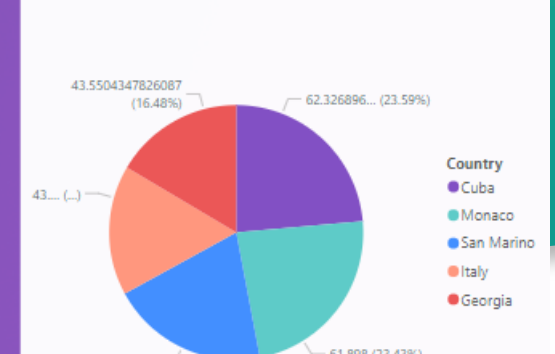
Average of SuicideRates by Country and Gender



Average of roadTrafficDeathRate by Country and Year



Average of Doctors Indicator by Country





Medical Staff

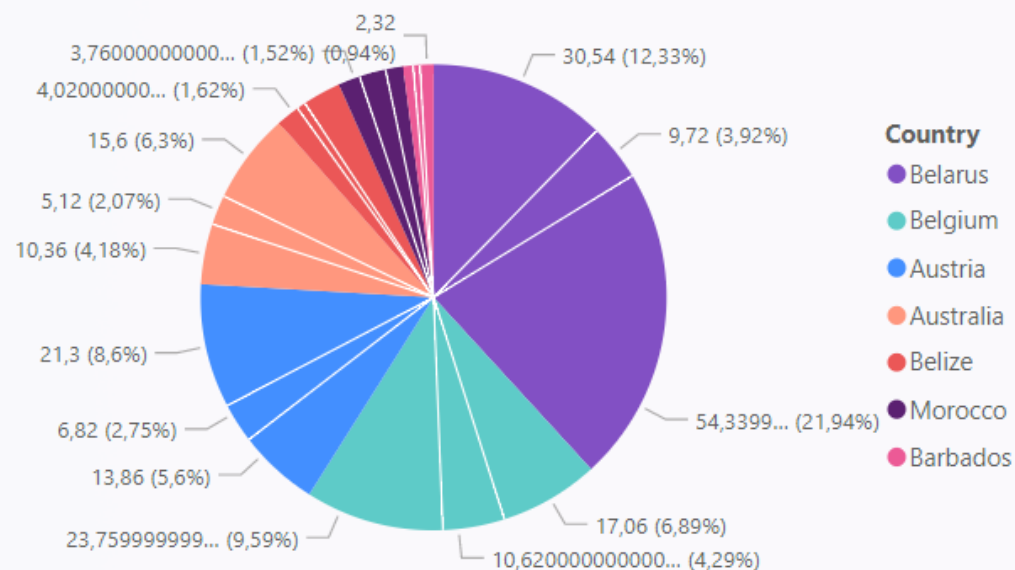
1795

Count of Pharmacists

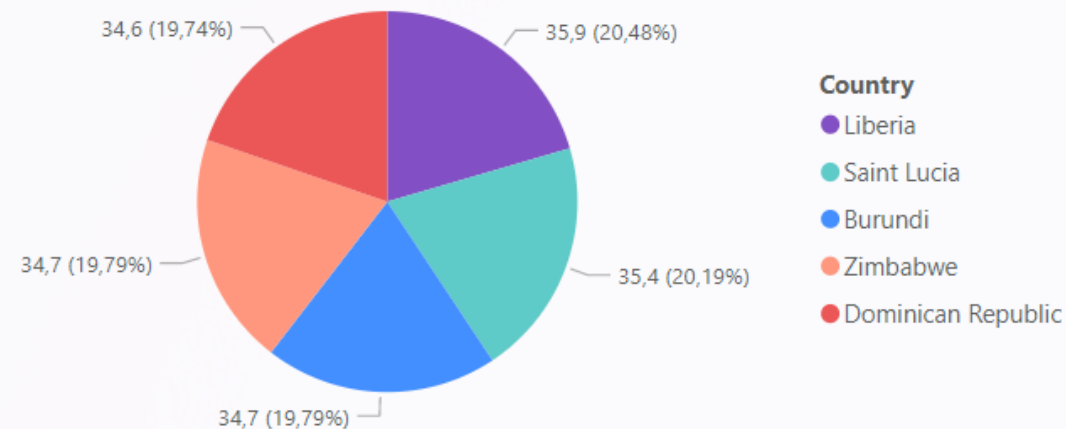
2506

Count of Medical Doctors

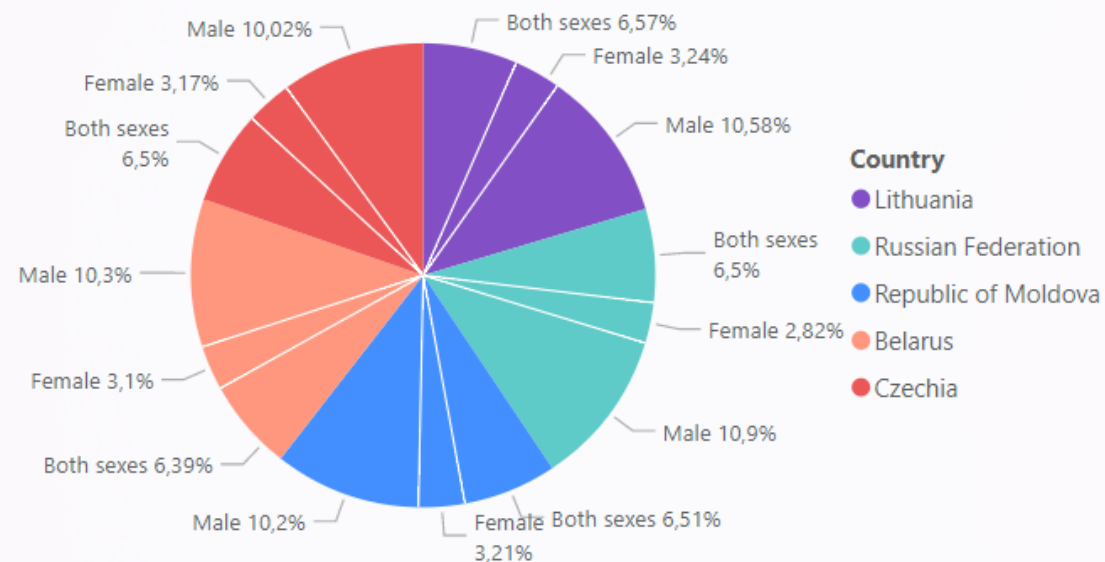
Average of SuicideRates by Country and Gender



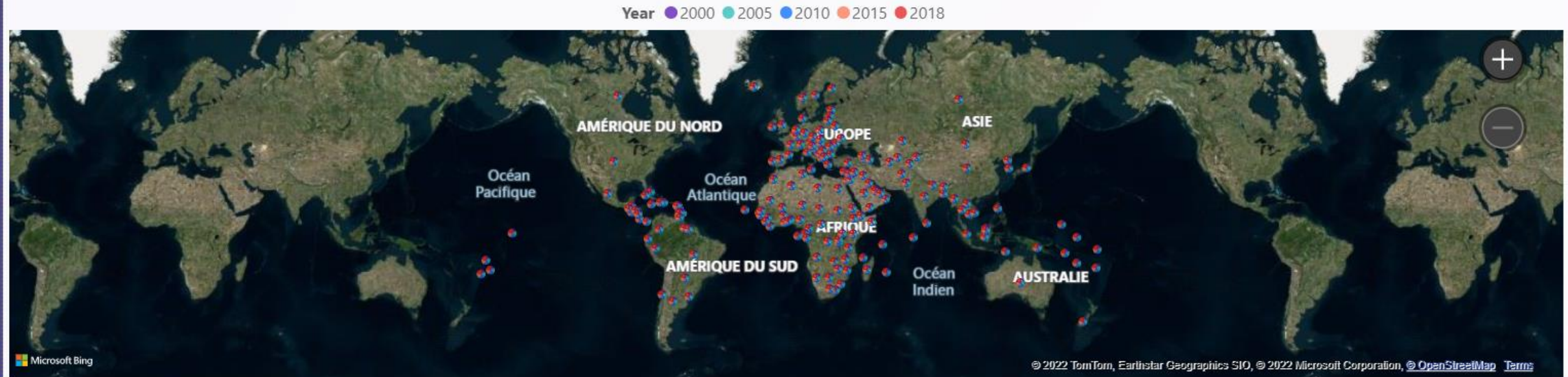
Average of roadTrafficDeathRate by Country and Year



Average of Alcohol Abuse Indicator by Country and Gender



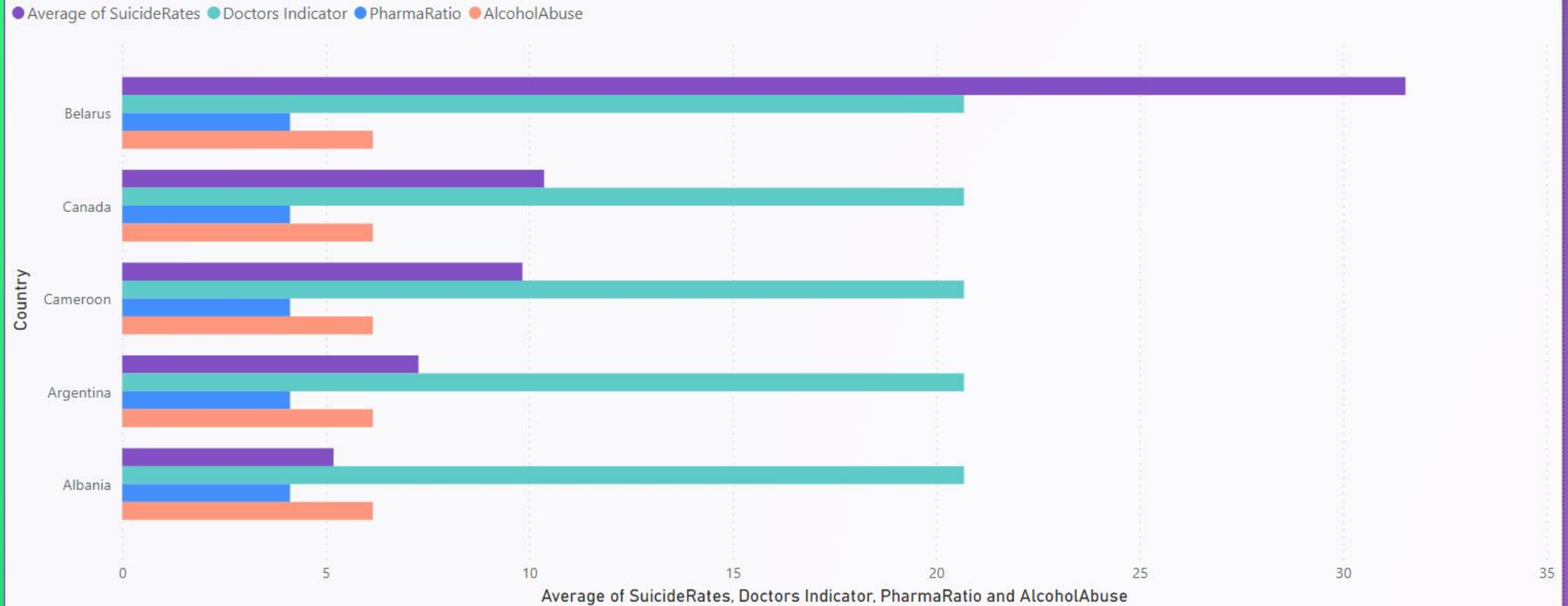
Average of Alcohol Abuse Indicator by Country and Year



Average of Alcohol Abuse Indicator by Country



Average of SuicideRates, Doctors Indicator, PharmaRatio and AlcoholAbuse by Country



At 31,53, Belarus had the highest Average of SuicideRates and was 507,19% higher than Albania, which had the lowest Average of SuicideRates at 5,19.

Across all 5 SuicideRatesLocation, Average of SuicideRates ranged from 5,19 to 31,53, Average of DoctorsPerIndicator ranged from 20,69 to 20,69, and Average of PharmaRatio ranged from 4,12 to

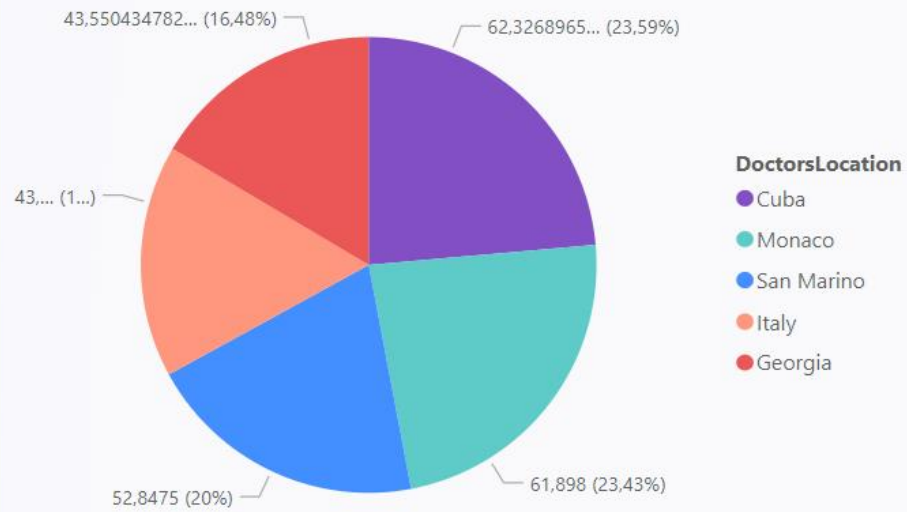
Doctor's Countries

All

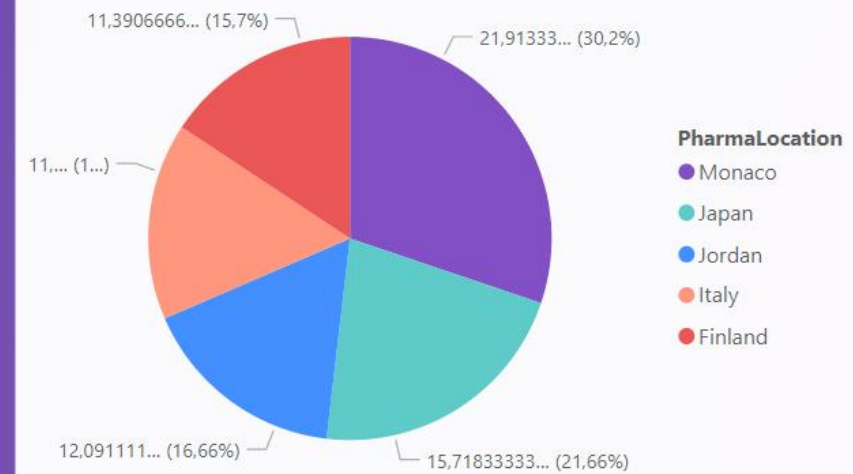
Pharmacist's Countries

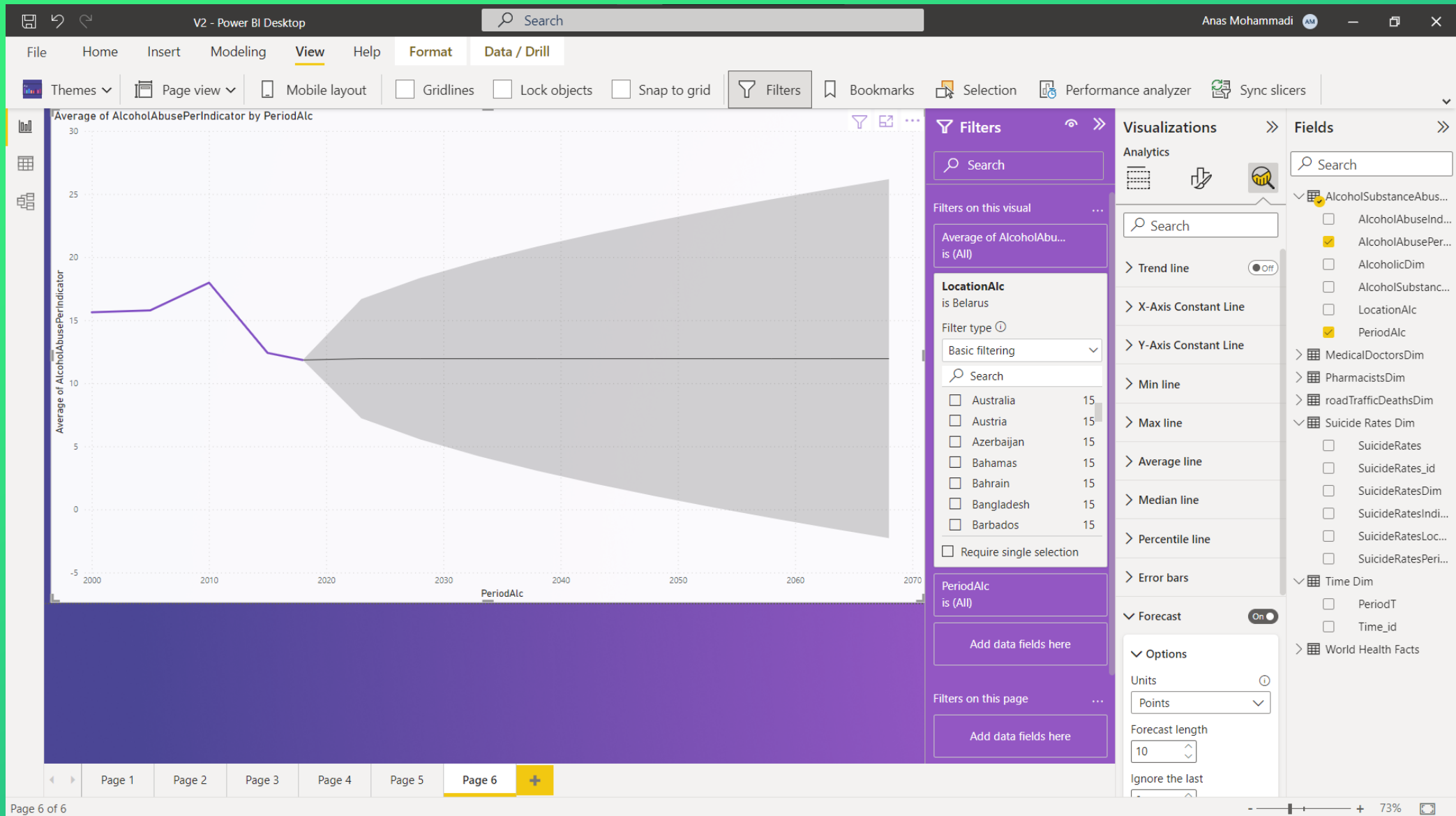
All

Average of DoctorsPerIndicator by DoctorsLocation



Average of PharmaRatio by PharmaLocation





Conclusion

Il existe de nombreux moyens sur internet qui permettent de présenter ces statistiques. Cependant, il n'est pas facile de quantifier ces indicateurs et rendre l'information facile et accessible au grand public. En espérant que cette étude sera à la hauteur des attentes attendues.



Perspectives :

- Mise en place d'un système de prediction (algorithme ML)
- Enrichir les données (plus d'indicateurs , dimensions)
- Raffiner les visualisations : Dashboard ...etc

Webographie

Dataset :

<https://www.kaggle.com/datasets/utkarshxy/who-worldhealth-statistics-2020-complete?select=WHOregionLifeExpectancyAtBirth.csv>

Pandas:

<https://pandas.pydata.org/docs/index.html>

Jupyter Notebook :

<https://jupyter.org/>

Partie ETL :

<https://datatofish.com/import-csv-sql-server-python/>

Traduction :

<https://en.pons.com/text-translation>

Github :

https://github.com/AnasM21/World_Health_Stats

MERCI POUR VOTRE
ATTENTION