**PROIECT BUSINESS INTELLIGENCE**

**STUDENT:**

**Valcov Bianca**

**Obiectivul** acestui proiect este acela de a analiza producția și consumul diferitelor surse de energie din întreaga lume,care reprezintă un factor mare de influență asupra schimbărilor climatice, precum și asupra sănătății umane. Acest proiect se concentrează pe defalcarea surselor de energie: cum variază acestea în lume și cum se schimbă acest lucru în timp.

Tabelele utilizate reprezintă o imagine clară a surselor de energie care au într-adevăr un impact negativ asupra planetei și ce alte surse am putea folosi în schimb.

La nivel global, obținem cea mai mare cantitate de energie din petrol, urmat de cărbune, gaze, apoi putere hidroelectrică. Mixul energetic global este încă dominat de combustibilii fosili. Acestea reprezintă mai mult de 80% din consumul de energie.

În respectivul proiect sunt analizate cantitatea de emisii de gaze cu efect de seră pe care o produce fiecare sursă de energie în lume, rata mortalității pe care o generează fiecare sursă de energie în lume, cantitatea de energie consumată pe lume, după sursă, prin diferite grafice.

De asemenea am creat și un tabel cu Python Script care afișează consumul de energie primară din combustibili fosili, nuclear și din sursele regenerabile.

**Sursele folosite:**

1 .Energy consumption (energy mix)

<https://ourworldindata.org/energy-mix>

2.Safest and cleanest sources of energy

<https://ourworldindata.org/safest-sources-of-energy>

3.Cheap renewables energy sources´s growth

<https://ourworldindata.org/cheap-renewables-growth>

4. Pollution, death rates, and cost by energy sources

<https://www.kaggle.com/diegoandresm/pollution-and-death-rates-by-energy-sources>

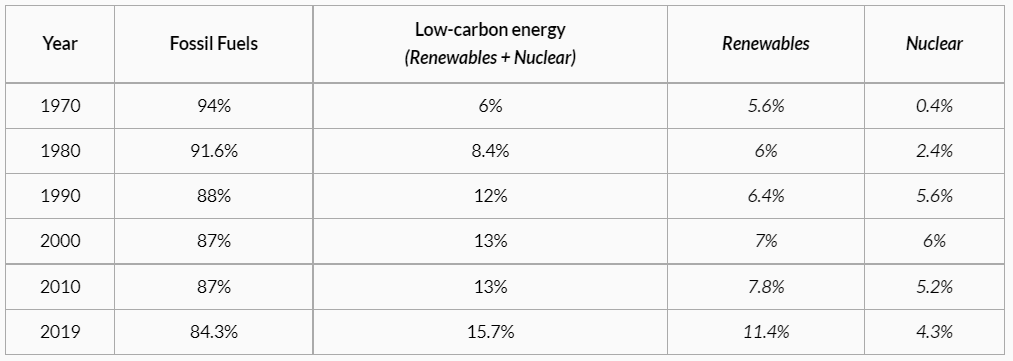
**Crearea unui tabel cu Python Script**

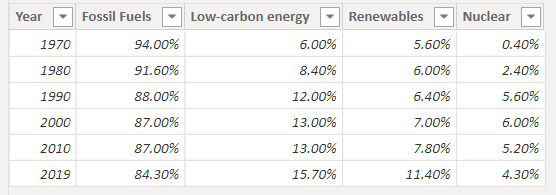
import pandas as pd

set\_date = [[1970, '94%', '6%', '5.6%', '0.4%'], [1980, '91.6%', '8.4%', '6%', '2.4%'], [1990, '88%', '12%', '6.4%', '5.6%'], [2000, '87%', '13%', '7%', '6%'], [2010, '87%', '13%', '7.8%', '5.2%'], [2019, '84.3%', '15.7%', '11.4%', '4.3%']]

Primary\_energy\_consumption\_from\_fossil\_fuels\_nuclear\_and\_renewables = pd.DataFrame (set\_date, columns = ['Year', 'Fossil Fuels', 'Low-carbon energy', 'Renewables', 'Nuclear'])

print (Primary\_energy\_consumption\_from\_fossil\_fuels\_nuclear\_and\_renewables)





**Cod clusterizare pentru tabelul energy\_consumption\_by\_source\_and\_region**

from pycaret.clustering import \*

dataset = get\_clusters(dataset, num\_clusters=4, ignore\_features =['Entity','Code'] )

**Calcul coloana Total\_consumption**

Total\_consumption = [Oil Consumption - EJ]+[Gas Consumption - EJ]+[Coal Consumption - EJ]+[Solar Consumption - EJ]+[Hydro Consumption - EJ]+[Nuclear Consumption - EJ]+[Wind Consumption - EJ]+[Geo Biomass Other - EJ]+[Biofuels (TWh)]

**Coloana cu calcul Stare\_cost pentru anul 2019**

Stare\_cost = if([Cost of MWh production for electricity on 2019 ($)]<100,"COST SCAZUT",if([Cost of MWh production for electricity on 2019 ($)]>100,"COST RIDICAT","COST MEDIU"))

**Măsuri care calculează target-ul pentru consumul de gas și diferența dintre consumul de gas și target în funcție de fiecare țară**

M\_target\_gas\_consum = 10000

M\_dif\_target = SUM(energy\_consumption\_by\_source\_and\_region[Gas Consumption - EJ])-[M\_target\_gas\_consum]

**Măsură care filtrează țările ale căror consum total depășește 1000:**

M\_total\_consum = CALCULATE(COUNT(energy\_consumption\_by\_source\_and\_region[Entity]),FILTER('energy\_consumption\_by\_source\_and\_region', (energy\_consumption\_by\_source\_and\_region[Total\_consumption])>=1000))

**Măsuri cu SUM și MAX**

M1 = SUM([percentual amount of global energy production])

M2 = MAX([percentual amount of global energy production])

M3 = SUM([tonnes of greenhouse gas emissions per gigawatt-hour of electricity production])

M4 = SUM([Death rate from accidents and air pollution per terawatt-hour of energy production])

M\_Total = SUM([percentual amount of global energy production])+SUM([tonnes of greenhouse gas emissions per gigawatt-hour of electricity production])+SUM([Death rate from accidents and air pollution per terawatt-hour of energy production])

**Secvente Python**

import matplotlib.pyplot as plt

ax = plt.gca()

dataset.plot(kind='line',x='source of energy',y='Cost of MWh production for electricity on 2010 ($)',ax=ax)

dataset.plot(kind='line',x='source of energy',y='Cost of MWh production for electricity on 2019 ($)', color='red', ax=ax)

plt.show()

import matplotlib.pyplot as plt

bar1 = dataset.plot(kind='area',x='Entity',y='Deaths per TWh')

plt.show()

**disertatie**

**from pycaret.clustering import \***

**dataset = get\_clusters(dataset, num\_clusters=11, ignore\_features =['Last\_Update','Actualizare\_cursuri'] )**