BDA Mini Project
on
Alcohol and Happiness

Group Members :

1. Saurabh Rajendra Jadhav (24)
2. Sanket Chandrashekhar Harvande(19)
3. Shreya Chandrakant Malavade(35)

INTRODUCTION

- The workplace provides several opportunities for implementing prevention strategies to reduce the harm done by alcohol, since the majority of adults are employed and spend a significant proportion of their time at work. The workplace can also be a risk factor for harmful alcohol use.
- Many studies have found significant associations between stress in the workplace and elevated levels of alcohol consumption, an increased risk of problem drinking and alcohol dependence. Evidence has found that alcohol, and in particular heavy drinking, increases the risk of unemployment and, for those in work, absenteeism.
- Alcohol, especially episodic heavy drinking, has also been found to increase the risk of arriving late at work and leaving early or disciplinary suspension, resulting in loss of productivity; a higher turnover due to premature death
- We have analysed the happiness level on the basis of their alcohol consumption.

PROBLEM STATEMENT AND OBJECTIVE

Problem Statement :

- To identify the happiness level of the particular region based on the dataset

Objective :

- Each region has different levels of alcohol intake capacity and their percentages. The dataset contains percentages of different levels of alcohol intakes and we have to identify the happiness levels based on their intakes and calculate the accuracy of their levels of happiness.

DATA SET DETAILS

- Total alcohol per capita consumption is defined as the total (sum of recorded and unrecorded alcohol) amount of alcohol consumed per person (15 years of age or older) over a calendar year, in litres of pure alcohol, adjusted for tourist consumption.
- Statistical concept and methodology: The estimates for the total alcohol consumption are produced by summing up the 3-year average per capita (15+) recorded alcohol consumption and an estimate of per capita (15+) unrecorded alcohol consumption for a calendar year. Tourist consumption takes into account tourists visiting the country and inhabitants visiting other countries.
- Variable time span 2000 – 2018
- Link: https://ourworldindata.org/alcohol-consumption

ALGORITHMS AND LIBRARIES USED

- **Clustering Algorithm :**

Clustering or cluster analysis is an unsupervised learning problem. It is often used as a data analysis technique for discovering interesting patterns in data, such as groups of customers based on their behaviour. There are many clustering algorithms to choose from and no single best clustering algorithm for all cases.

- **Pandas :**

Pandas is a popular Python library for data analysis. It is not directly related to Machine Learning. As we know that the dataset must be prepared before training. In this case, Pandas comes handy as it was developed specifically for data extraction and preparation. It provides high-level data structures and a wide variety of tools for data analysis. It provides many inbuilt methods for grouping, combining and filtering data.

- **Scikit-learn (sklearn) :**

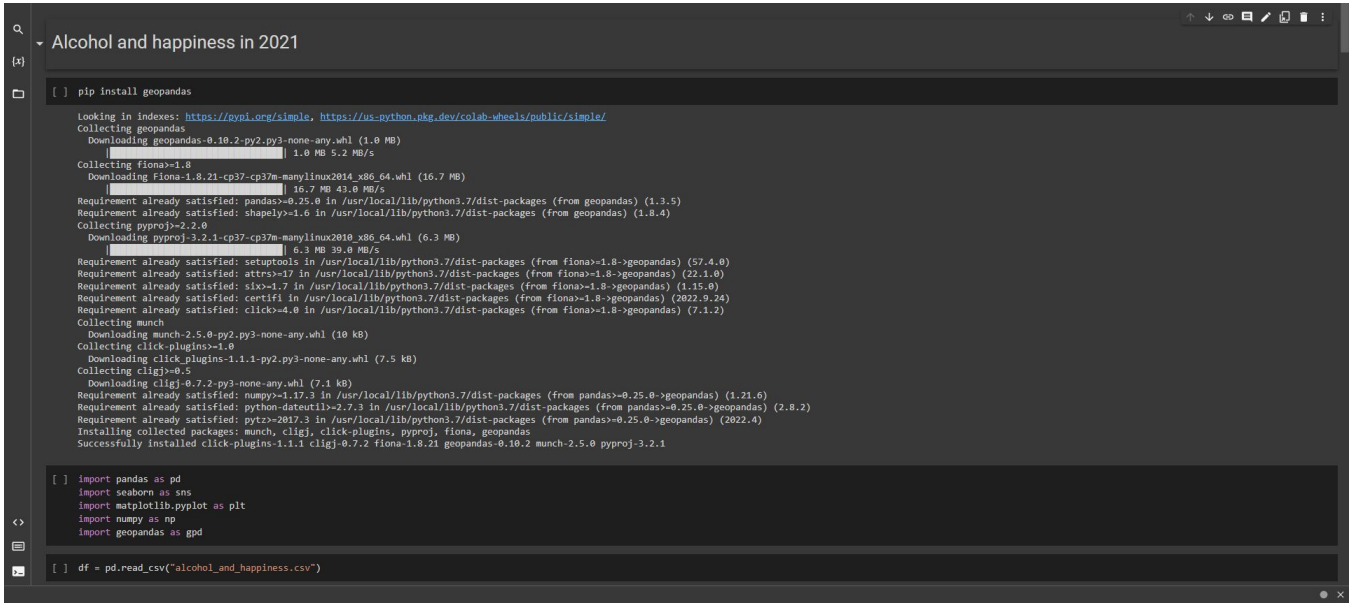
Scikit-learn is one of the most popular ML libraries for classical ML algorithms. It is built on top of two basic Python libraries, viz., NumPy and SciPy. Scikit-learn supports most of the supervised and unsupervised learning algorithms. Scikit-learn can also be used for data-mining and data-analysis, which makes it a great tool for those starting out with ML.

ALGORITHMS AND LIBRARIES USED

- **Matplotlib :**

Matplotlib is a very popular Python library for data visualisation. Like Pandas, it is not directly related to Machine Learning. It particularly comes in handy when a programmer wants to visualise the patterns in the data. It is a 2D plotting library used for creating 2D graphs and plots. A module named pyplot makes it easy for programmers for plotting as it provides features to control line styles, font properties, formatting axes, etc. It provides various kinds of graphs and plots for data visualisation, viz., histogram, error charts, bar charts, etc .

RESULTS



Alcohol and happiness in 2021

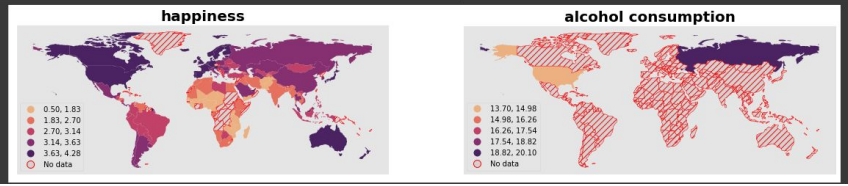
```
[ ] pip install geopandas

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting geopandas
  Downloading geopandas-0.10.2-py2.py3-none-any.whl (1.0 MB)
     |████████████████████████████████| 1.0 MB 5.2 MB/s
Collecting fiona>=1.8
  Downloading Fiona-1.8.21-cp37-cp37m-manylinux2014_x86_64.whl (16.7 MB)
     |████████████████████████████████| 16.7 MB 43.0 MB/s
Requirement already satisfied: pandas>=0.25.0 in /usr/local/lib/python3.7/dist-packages (from geopandas) (1.3.5)
Requirement already satisfied: shapely>=1.6 in /usr/local/lib/python3.7/dist-packages (from geopandas) (1.8.4)
Collecting pyproj>=2.2.0
  Downloading pyproj-3.2.1-cp37-cp37m-manylinux2010_x86_64.whl (6.3 MB)
     |████████████████████████████████| 6.3 MB 39.0 MB/s
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (from fiona>=1.8->geopandas) (57.4.0)
Requirement already satisfied: attrs>=17 in /usr/local/lib/python3.7/dist-packages (from fiona>=1.8->geopandas) (22.1.0)
Requirement already satisfied: six>=1.7 in /usr/local/lib/python3.7/dist-packages (from fiona>=1.8->geopandas) (1.15.0)
Requirement already satisfied: certifi in /usr/local/lib/python3.7/dist-packages (from fiona>=1.8->geopandas) (2022.9.24)
Requirement already satisfied: click>=4.0 in /usr/local/lib/python3.7/dist-packages (from fiona>=1.8->geopandas) (7.1.2)
Collecting munch
  Downloading munch-2.5.0-py2.py3-none-any.whl (10 kB)
Collecting click-plugins>=1.0
  Downloading click_plugins-1.1.1-py2.py3-none-any.whl (7.5 kB)
Collecting cligj>=0.5
  Downloading cligj-0.7.2-py3-none-any.whl (7.1 kB)
Requirement already satisfied: numpy>=1.17.3 in /usr/local/lib/python3.7/dist-packages (from pandas>=0.25.0->geopandas) (1.21.6)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-packages (from pandas>=0.25.0->geopandas) (2.8.2)
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-packages (from pandas>=0.25.0->geopandas) (2022.4)
Installing collected packages: munch, cligj, click-plugins, pyproj, fiona, geopandas
Successfully installed click-plugins-1.1.1 cligj-0.7.2 fiona-1.8.21 geopandas-0.10.2 munch-2.5.0 pyproj-3.2.1
```

```
[ ] import pandas as pd
    import seaborn as sns
    import matplotlib.pyplot as plt
    import numpy as np
    import geopandas as gpd
```

```
[ ] df = pd.read_csv("alcohol_and_happiness.csv")
```

RESULTS

```
[ ]  fig,(ax,ax2) = plt.subplots(ncols = 2,figsize=(20, 10))
     plt.style.use("ggplot")


     ax.set_xticks([]) #  removing ticks
     ax.set_yticks([])

     ax.set_title('happiness', weight = 'bold', fontsize  = 20)
     world.plot('total_happiness', legend = True, ax = ax,
                missing_kwds= {'color': 'lightgrey', 'edgecolor': 'red', 'hatch': '///', 'label': 'No data'},
                cmap='flare', scheme = 'quantiles',legend_kwds={'loc': 'lower left'})

     ax2.set_xticks([])
     ax2.set_yticks([])

     ax2.set_title('alcohol consumption', fontsize = 20, weight = 'bold')
     world.plot('total_literes_of_pure_alcohol', ax= ax2 , legend = True,
                missing_kwds = {'color': 'lightgrey', 'edgecolor': 'red', 'hatch': '///', 'label': 'No data'},
                cmap='flare',scheme = 'quantiles', legend_kwds={'loc': 'lower left'})

     plt.show()
```



CONCLUSION

As more and more data is generated and collected, data analysis requires scalable, flexible, and high performing tools to provide insights in a timely fashion. Thus we have implemented the analysis of the dataset which contains the happiness percentages of the regions and their intake capacities.

REFERENCES

[1] B. Xiao, "Mobile Internet and Residents' Happiness-Quantitative Analysis Based on CFPS," 2021 2nd International Conference on Big Data Economy and Information Management (BDEIM), 2021, pp. 93-96, doi: 10.1109/BDEIM55082.2021.00027.

[2] *Y. -Y. Ou, A. -C. Tsai, T. -W. Kuan, J. -F. Wang and J. -H. Tian, "Happiness understanding system based on human action recognition," 2016 International Conference on Orange Technologies(ICOT), 2016, pp. 56-59, doi: 10.1109/ICOT.2016.8278978.*

[3] Y. -Y. Ou, D. -R. Yeh, C. -C. Kung, P. -C. Lin, T. -W. Kuan and J. -F. Wang, "Design and implementation of happiness database for fMRI study," 2015 International Conference on Orange Technologies (ICOT), 2015, pp. 127-130, doi: 10.1109/ICOT.2015.7498493.

[4] *D. -Y. Liu, C. -T. Lin, K. -C. Wang and M. -Y. Chen, "Real and Virtual Happiness Prediction Model Based on Multiple Regression analysis by Instagram Social Network," 2021 IEEE International Conference on Social Sciences and Intelligent Management (SSIM), 2021, pp. 1-8, doi: 10.1109/SSIM49526.2021.9555216.*