Praktice Data Science Assignment

Instructions

1. Do this assignment with your own skills.
2. Preferred programming language: Python

Your Tasks are

Task 1 (Data Collection, Cleaning and storage):

URL: https://www.internetworldstats.com/list2.htm#a-c

Task Details:

1. Scrape The country data in above url.
2. Clean above data, and store it in Mysql DB
   - You need store the data in proper format for it to be used in the task 2
   - You also need to connect all the tables and store it in a single mysql table.

Task 2 (Data Exploration)

Task Details:

1. Explore the data and find the insight that are useful for a software company.
2. Be creative and use of graphs to show your results (You may use any application, package and library you see fit).

Task 3 (Data Analysis using Mysql)

Task Details:

1. which country has the highest internet user per sq. Km

2. Count of countries where internet penetration exceeds 30% and population below 34,169,169.

3. Average area of the country with respect to the first character in the country name.

    Result Sample:

first character in the country name.	Average are of country
A	20
B	12
C	13
D	11
...	....

4. Find the country count with respect to the Internet Penetration% in the BIN size of 10% (Eg. 50 countries lie in the Internet Penetration range of 0-10%).

    Result Sample:

Internet Penetration	Country count
10%	20
20%	12
30%	13
40%	11
...	....

5. From the country and region name, extract all the region names.

6. Which country had the highest Internet users with respect to the date.

    Result Sample:

Internet Penetration	Country count
Mar/2017	Andorra

Dec/2017	Australia
…	….

7. List of country where second letter of the name is equal to the second letter of symbol

Task 4 (Making ML Model)

Task Details:

1. Use Kmean Clustering algorithm to cluster the countries together and explain the useful insight you obtained from it. You may use suitable graphs if necessary.

Assignment Submission

Submission Details:

1. Task 1: Send SQL database dump in sql and your code.
2. Task 2: Compile the insight, make a PDF document and send it to us.
3. Task 3: Send your sql query and sql result in excel sheet. (Limit the result to 10 if needed.)
4. Task 4: Export your model using pickle. Send you model, python code and Insight you obtained.