

Assignment 5

Working with CSV, JSON, SQL in Python

Applied Data Science

Instructions

This assignment requires you to work with three (3) different files of CSV, JSON and SQL formats. Use Python to answer the questions. Be sure to include a complete explanation on how to run/execute your code, and the complete Python source code. Document your code using block and inline comments accordingly.

You will be provided with data from the agencies that reported one or more hate crime incidents that occurred in their respective jurisdictions during one or more quarters in 2013. The data are distributed across three files, one each of CSV(HateCrimes-csv.csv), JSON(HateCrimes-json.json) and SQL file format. This task involves integrating the data from these files in order to answer specific questions about the data. For further clarity, refer to the data description file provided in Canvas(HateCrimes-data-defn.txt)

For this assignment you will be required to submit three files. A .pdf file containing answers the specific questions with your name and email in the header (see the naming conventions for the course), single-spaced and font no-larger than 12pt. Complete python code used for the analysis can be attached in the appendix or submitted as a python script/ python notebook. For Qn.1. submit the CSV file containing data from all three sources as your answer. Assignment will be graded based on the logic used for arriving at results and correctness of output values.

Question 1

Combine the data from the three files and write the output to a CSV file.

Question 2

Find the percentage of the Annual crime happened in each of the four Quarters when aggregated at Agency type level.

Question 3

Find the top 5 states where hate crimes committed against ethnic groups happened most frequently.

Question 4

Find each type of hate crimes' percentage happened in Cities.(Agency_type is Cities.)