|  |
| --- |
| Food Violations Report |
| TSZ MING LAU2810ICT Software TechnologiesOctober 5, 2019 |

# Abstract

In this report, use python code to interact with the database and extract and visualize the data with excel files, numpy and matplotlib. The amount of data provided is relatively large. Extract the code and description of the offending product by creating a statement in the table and sql and add it to the excel file. Create a stereo array with numpy. Matplotlib is to visualize the data. The methods provided in this report make the data more straightforward and meet the requirements of data extraction. The methods covered in this article are valid and can show the desired results.

# Introduction

This report is based on the big data provided to use python code and database, Excel, numpy to analyze the data. Matplotlib is able to visualize all data generation data graphs. Through the raw data provided by itself, the code can create a table and the data can be inserted into the table. The purpose of this report is to be able to simplify big data and classify and count the data. Classify the same kind of data into one category. Task1 is implemented by creating a python script to pour the data in the two excel files provided by itself into the newly created database through python code. Task2 is to find all the violating enterprises and sort them by sql statemen. Task3 is to find out all the types of violations and create a new excel file to dump all the data. The last task is to visualize the data.

## Database Structure

# Create table inspection:

# activity\_data DATE,

# employee\_id CHAR(9),

# facility\_address VARCHAR(50),

# facility\_city VARCHAR(20),

# facility\_id CHAR(9),

# facility\_name VARCHAR(50),

# facility\_state CHAR(2),

# facility\_zip VARCHAR(10),

# grade CHAR(1),

# owner\_id CHAR(9),

# owner\_name VARCHAR(30),

# pe\_description VARCHAR(50),

# program\_element\_pe INTEGER(4),

# program\_name VARCHAR(30),

# program\_status VARCHAR(8),

# record\_id CHAR(9),

# score INTEGER(2),

# serial\_number VARCHAR(30) not null primary key,

# service\_code INTEGER(5),

# service\_description VARCHAR(30)

create table Violations:

# points INTEGER(1),

# serial\_number CHAR(9),

violation\_code CHAR(4),

violation\_description VARCHAR(50),

violation\_status VARCHAR(30),

FOREIGN KEY (serial\_number) REFERENCES inspections(serial\_number)

Create table PreviousViolations:

﻿name VARCHAR(30),

address VARCHAR(30),

zip\_code INTEGER(15),

city VARCHAR(20)

# 

# 

# 

# 

# **Violation counts**

# In task 3, the purpose of task 3 is to sort according to the code of the violation. Classify products that belong to the same code and description into one category and count the number of times. In task3, you first need to create a new excel file to prepare for the subsequent insert data. At the same time, it needs to apply to the statement of sql. The categorization of data based on the big data that has been provided is to make the data more concise. Use sql's statement to query the code in the database that violates the product and find a description of the violation. Count by the standard of group by in sql. Task3 is counted based on the violation code. Finally, the data is extracted from the database and then inserted into the previously created new excel file. The first column of the excel file shows the violation code. The second column shows the violation description. The third column shows the number of violation code violations.

# **Violations over time**

Based on Task 4 of your assignment, describe the results of your analysis (in addition to the graph – you may need to screenshot or otherwise export from your iPython console). Make any comments about the data you see fit.