**Assignment - SQL Statements**

Shaun Pritchard

Rasmussen College

QMB4000

Benjamin Tasker

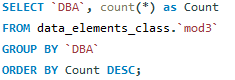
December 07, 2020

**Assignment - SQL Statements**

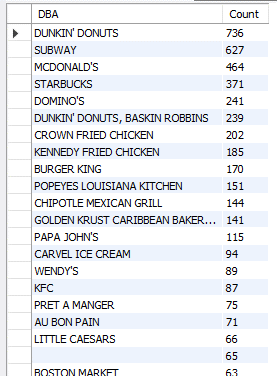
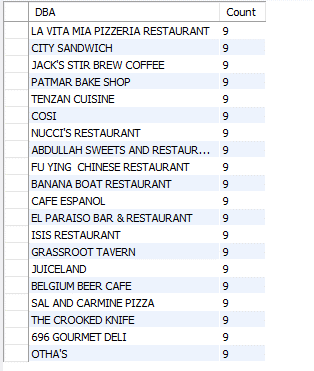
**Using Module 3 data I answered the following questions in MySQL**

* **Times is each restaurant was found in the dataset:**
  + There are 67218 restaurants in the module 3 data set that are all unique named restaurants.
  + Each restaurant appears numerous amounts of time with Starbucks showing up most frequent at 371 times.



Using SQL code, I was able to derive the following results in descending order. 

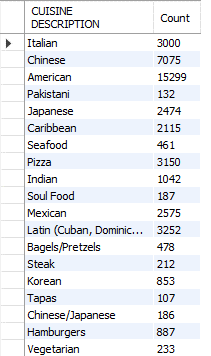
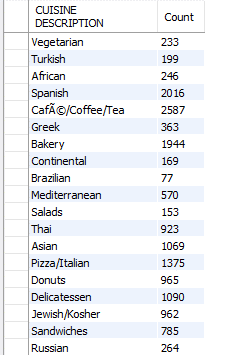
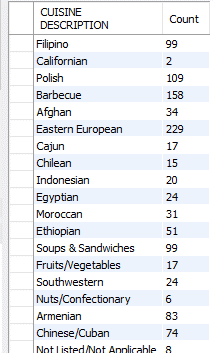
Due to the amount of data I only took screenshots of the maximum and minimum restaurants frequency.

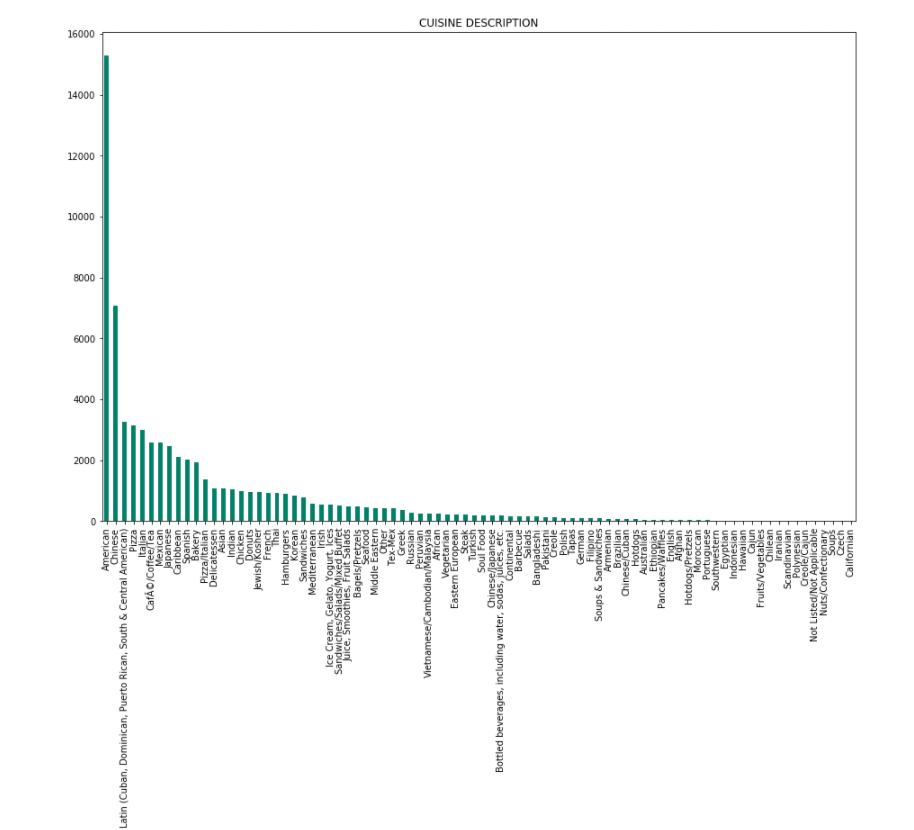
 

* **Number of times each cuisine in the dataset?**

Using SQL, I derived count data using this code (I renamed my table mod3 after running this query).



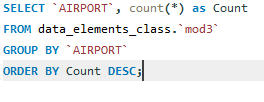
* +    
  + Here is the chart I created using Matplotlib in python with MySQL. Connect

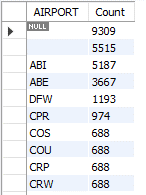


* **Airport the showed up the most:**

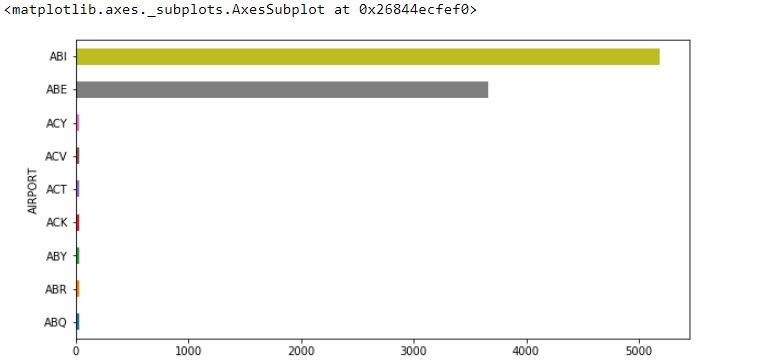
Using both MySQL and python connection to the server I was able to derive data on the airport which showed up most frequently pinpointing the top airports with most violations.

Using SQL Code to derive the results:



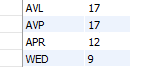


Python matplotlib visualization shows ABI Airport shows up the most aside from the Null values and missing data which is at the highest frequency.

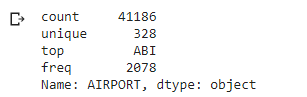


* **Airport that shows up the least amount of times:**

According to the data using the same SQL query the results show WED airport shows up the least. Meaning that accrued the least amount of violations.



* **What is the total number of violations?**
  + Total number of features in the dataset based on airports



Based on the total number of violations based on the SQL code there where 67283 total violations.





* **How Many Inspection Types exist? What are they?**

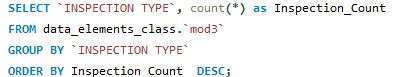
Using MySQL code below I was able to derive that there where 33 types of inspections.



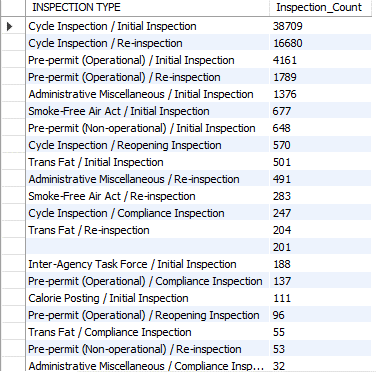


From the following SQL count statement, I was able to discern a distinct list of each type of inspection. Also, I was able to discern how many violations types where issued to restaurants in the airport.

**Code Used:**



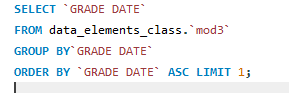
**Results:**

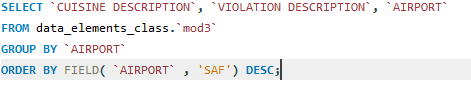
* **What is the earliest GRADE DATE?**
  + The earliest grade date for violations reports is 9/13/2013



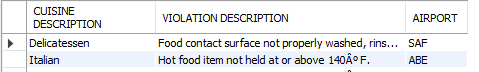
I had to use serval queries to find this and scan through the table manually here are the queries I used. Also, these queries returned the lowest date as null missing values.

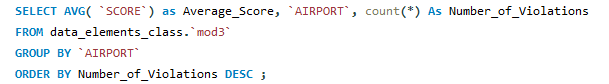
* **What Cuisine and violation description occur the most for the Airport SAF?**
  + Using the following code below I found the SAF had violations in Delicatessen cuisine with violations of food contact surfaces not properly washed.
  + **Code:**

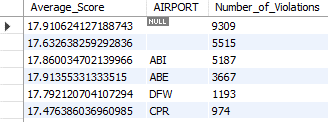


* + **Results:**

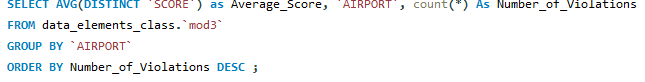


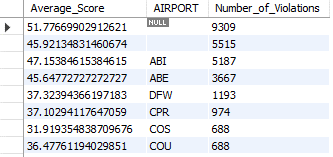
* **What is the AVERAGE score per airport, and what is the OVERALL AVERAGE**
  + The average score per airport based on all violations and inspections was calculated using this code.





* + The Average score of airports based on distinct values ignoring duplicate values is as follows.



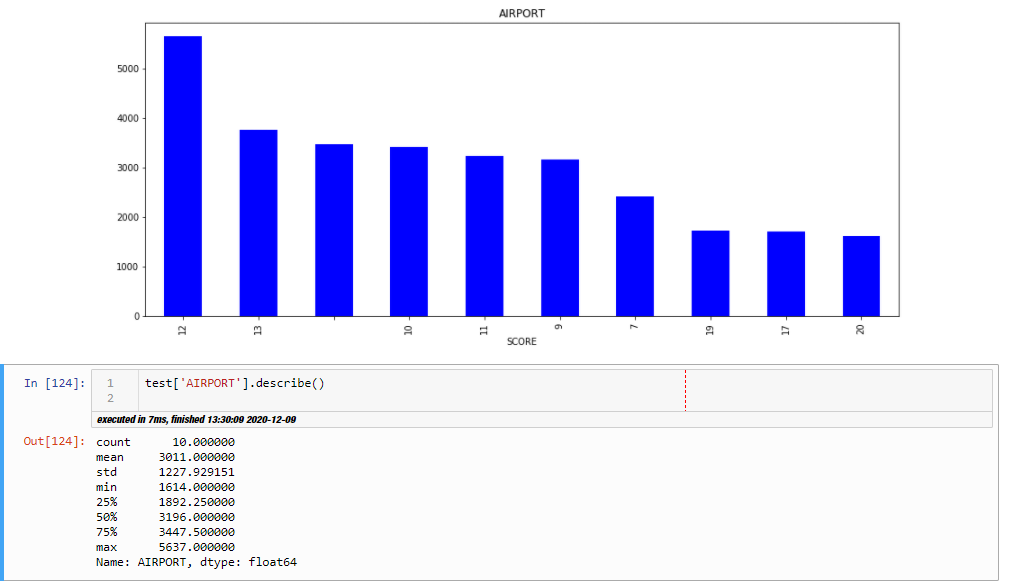


* + The overall average score for the entire score column is as follows.



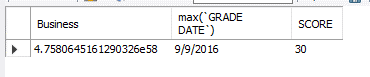
 

Data infeences for **SCORE** average From Python:

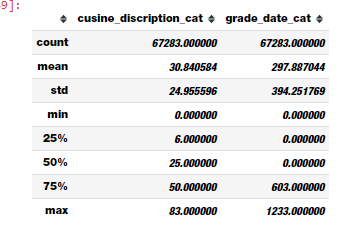


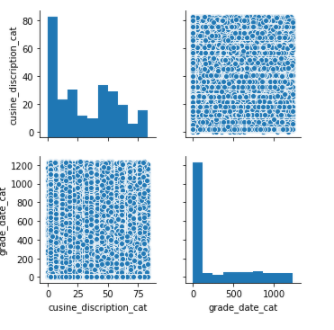
* **2 other custom queries**
  + The average score for each business based on inspections is 30 occurring on 9/92016. Which tells us this may have been a problematic year.

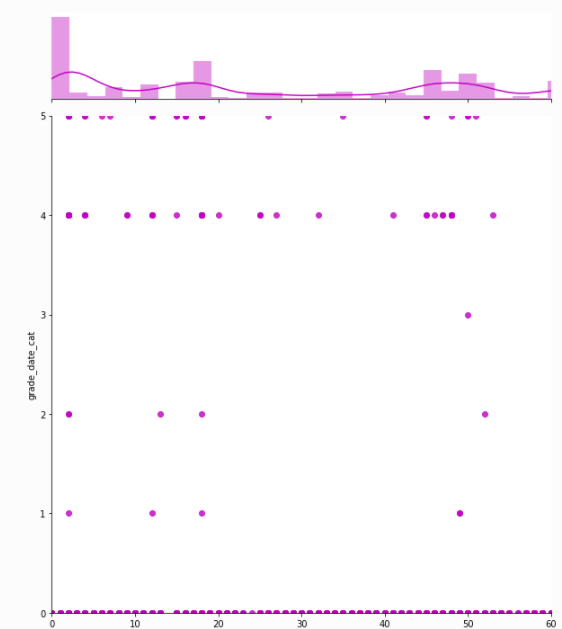




* + This show s the frequency of inspections and violations based on the type of cuisine correlated to the years from 2013 -2017 in which violations occurred per cuisine. (This I had to categorize my data and use cat.codes accessor labeling for columns in data Frame.







From the data we can see that more violations occurred during 2013, 2014, and 2015 with a steady decline which started to peak late 2016 and 2017 which corelates with other data showing 2016 as the year with trending violations.

* **What else have you noticed about the data, is it duplicative?**
  + I have noticed that there are many **NULL** values and the data needs heavy preprocessing.
  + I notice that big brand-named restaurants (i.e. Dunkin Donuts…) have more inspections than other smaller brand-named restaurants, but they have more restaurants in more airports.
  + I notice Chinese and American food are the cuisine types with most violations
  + I noticed WED airport has the least violations per every airport and maybe the CEO needs to take note on how they operate.
  + ABI restaurant has a the most issues ranging in the thousands of violations and should be the first airport to consider infrastructure and process changes to.