Aleksey Kramer

Data Science 350

Final Project

The goal of the project is to explore how Longitude and Latitude affect selection of the restaurant inspection types within the city of Seattle.

**Prerequisites:**

1. R
2. RStudio
3. R packages: logging, nortest
4. Windows operating system

**Script/Datafile:**

Eatery.Analysis.R, restaurant.csv

**Data:**

Data downloaded from the <https://data.kingcounty.gov/> web site and saved in the same directory with the Eatery.Analysis.R script. If the data is already downloaded, local copy of the data will be used. However, the data file is attached as a part of the submission in .zip archive.

**To execute the script:**

1. Open Eatery.Analysis.R with RStudio
2. Changed the setwd() settings to match location of the script/data
3. Update a section within if statement to reflect correct operating system (if needed)
4. Highlight all the text in the Eatery.Analysis.R file
5. Click ‘Run’ button

**Procedure:**

1. Acquire data
2. Extract Seattle’s data only from the download
3. Remove the name of the City (all the data at this stage pertains to Seattle)
4. Subset missing values in Violation.Type, Violation.Desscription, and Viloation\_Record\_ID variables with the string ‘none’. This assures that there is no missing data anywhere in the data set (even though these variables are not used)
5. Run normality tests on Longitude and Latitude variables (data in both variables is normally distributed)
6. Run ANVOA tests to test effects of Longitude and Latitude on the selection of Inspection Type.

**Findings:**

Two ANOVA tests conducted to evaluate effect of Longitude and Latitude on the inspection types produced significance levels P < 0.05, which is significant. This means that there are restaurants in certain areas in Seattle that are subjected to certain types of inspections more than restaurants in other areas of Seattle.

Following up with TukeyHSD test, it is observed that only ***Return Inspection-Consultation/Education – Field*** and ***Routine Inspection/Field Review-Return Inspection*** pairs produce statistically significant impact. This can further explain that restaurants that committed violations are more likely to be inspected again with Return Inspection or Consultation purposes to prevent future violations.

Having multiple follow-up inspections in the same restaurants located at the same Longitude and Latitude further explains the results of ANOVA tests producing statistically significant p-values: return inspections at the same coordinates create an appearance of certain areas in Seattle having disproportionally more certain types of inspections.