Instructions for Downloading and Processing 311 Service Requests Data for 2010

STEP 0: Setup

These instructions assume that you are familiar with working with spreadsheets. Download and Install RStudio and understand the basics of using the environment. A helpful tutorial by RStudio themselves can be found at- https://education.rstudio.com/learn/beginner/

Step 1: Downloading the Data

1. **Go to the NYC Open Data Portal:**
    - Visit the following link: NYC Open Data - 311 Service Requests.
    - OPTIONAL- Download the data catalog to look and understand the different variables.
2. **Filter Data by Year:**
    - Click on the "Filter" button.
    - In the filter panel, locate the "Created Date" condition and the “is between” condition from the drop down list.
    - Enter the date range for the year 2010:
        - **Start Date:** “01/01/2010 12:00:00 AM”
        - **End Date:** "12/31/2010 11:59:59 PM”

3. Export the Filtered Data:

- After applying the filter, click on "Export" in the top-right corner of the page.
- Select the "CSV" format to download the data.
- Save the file to your computer.

Step 2: Formatting the Data

1. Open RStudio.
2. Copy-Paste the script found below into a new R file in RStudio and save the R file as reading_csv_2010.R
3. Make the necessary changes to Replace `/path/to/your/downloaded/file.csv` with the actual path to the CSV file you downloaded.
4. Replace `/path/to/save/SRs_2010` with the directory where you want to save the RDS file.

```
# If these libraries are not already installed, RStudio will prompt you to
install them. Please go ahead and do so
library(data.table)
library(lubridate)

# Set file paths
csv_file <- "path/to/your/downloaded/file"

# Load the CSV file into a data.table
SRs_2010 <- fread(
   csv_file,
   check.names = TRUE, # Replaces spaces in column names with periods
   keepLeadingZeros = TRUE, # Keeps leading zeros(as they have information
value) in ZIP and BBL columns
   na.strings = c("NA", "N/A", "Unspecified", "", "null") # Define NA
strings
  )
# Formatting date columns since the dates are supplied in text format
date_cols <- c("Created.Date", "Closed.Date", "Due.Date",
"Resolution.Action.Updated.Date")
SRs_2010[, (date_cols) := lapply(.SD, function(x) mdy_hms(x, tz =
"America/New_York")), .SDcols = date_cols]
# Save the dataset as an RDS file called “SRs_2010.rds”
saveRDS(SRs_2010, "path/to/save/SRs_2010.rds")

# Clean up the environment
rm(SRs_2010)
```

Verify the RDS file’s integrity

1. You can verify the downloaded RDS file by looking at the unique key identifier for the first 3 rows and/or last 3 rows in the downloaded RDS file and verifying that the same exist on the 311 data on the NYC Open Data website.

STEP 3: Repeat to download Data from 2011- 2024

1. Go to NYC Open Data - 311 Service Requests.
2. Follow the instructions from Step 1 to download the data for any year between 2011 and 2024. Be sure to update the "Created Date" filter to match the desired year. For example:
    a. **Start Date:** "01/01/[Year] 12:00:00 AM"
    b. **End Date:** "12/31/[Year] 11:59:59 PM"

3. Reuse the script you saved as `reading_csv_2010.R` for processing data from other years.

   Update the paths in the script:

   a. **CSV File Path**: Update the path to point to the new CSV file you downloaded for the specific year.
   b. **RDS Save Path:** Ensure you save the output RDS file with a new name to avoid overwriting previous files (e.g., `SRs_2011.rds` for 2011 data).

4. Repeat this process for each subsequent year from 2011 to 2024, adjusting the file paths and the year in the script accordingly.