# DAT 375 Project One Template

# Data Analysis Process Job Aid

# Ashley Littles

### Who should use this job aid?

This job aid would be useful for the newly hired data analysts at the data consulting firm that I work at. This job aid will help the new data analysts understand our particular data analysis process that this data consulting firm uses to assist clients.

### Introduction

We are creating a Storm and Crime Data Report (SCDR) requested by the police department. To create this Storm and Crime Data Report (SCDR) we will be reviewing historical data from the city of Miami dating from 10/1/2019 to 10/31/2019. According to the Miami police there is an apparent connection between the increase in crimes and storms. We are creating this report in hopes of giving the police department ample information that they can use to predict any possible timeframes of the crimes in the future.

### Section 1: Type of analysis

The problem within this situation that we are trying to figure out is if the crime rate truly increases with storms. A barrier surrounding this problem is that the data is from a short period of time. We are only looking at data from one month which will only contain a small number of results. Another barrier is the data is from October, which has different weather patterns than other months and seasons of the year.

Given the problem, the appropriate type of data analysis to use is descriptive analysis. Descriptive analysis is used to find what happened with past data and create a summary. This specific data analysis type is best for the problem of if the crime rates increase with the storms because it can help us detect any similarities between variables. So, we would be able to determine the similarities between the crime rates and the storms.

### Section 2: Define Parameters and collect data

To create the Storm and Crime Data Report (SCDR), we must go into the details of what the specific parameters are within the data set. The relevant weather data variables needed for the problem's analysis are the specific type of storm activity. The type of weather variables would be Coastal Flood, Strong Wind Rain, Lightning, Thunderstorm Wind, Tropical Storm, Hail, Storm Surge/Tide, and Flash Flood. The criminal data variables required for the analysis are the type of crime activity. These criminal data variables are Violent Crime, Murder and Non-negligent manslaughter, Motor Vehicle Theft, Aggravated Assault, Robbery, Property Crime, Larcency Theft, and Burglary.

In order to essentially solve the problem, we have to determine what kinds of data are needed. For both the storm activity data and crime activity data we need to look at the date from 10/1/2019 to 10/31/2019. We also need to identify what zone the activity occurred in and the city. The zones are Coastal Miami-Dade County, Inland Miami-Dade, Metropolitan Miami-Dade, and Miami-Dade Co. A few of the cities are South Miami, Biscayne Park, Miami Beach, Sweetwater, Miami Springs, Miami Gardens, Surfside, Aventura, Miami Shores, Hialeah Gardens, Homestead, Bal Harbour, Hialeah, Bay Harbor Islands etc.

### Section 3: Tool Selection

For a truly accurate data analysis we need to use the proper tools when performing the analysis. There are a few steps that will need to be taken to prepare the data for analysis. The first step would be to make sure the data set is cleaned through the MySQL platform. We then must figure out the best tool for this kind of analysis. For this analysis the tool that is the most suitable would-be python. Python will allow us to properly visualize the data with the use of models.

To demonstrate the projected outcome of the data analysis we must choose the right visualizations to use. For this problem it would be best to use a bar chart and histogram. A bar chart will compare the dates and numerical crime activity and numerical storm activity. A histogram will show the frequency of crime activity with each kind of storm.

### Section 4: Validation

To ensure that we are providing accurate results with the data we must go through the process of data validation. To validate the data using the python tool, we can use various techniques such as range validation to ensure that the data falls within the correct ranges and cross-field validation to ensure no inconsistencies. For this data set, we want to keep the necessary data for the problem and filter out the rest.

The data that we need to keep from this data set is the storm type, crime type, zones, and city. To filter data that is required from a larger data set we can use the SQL scripts:

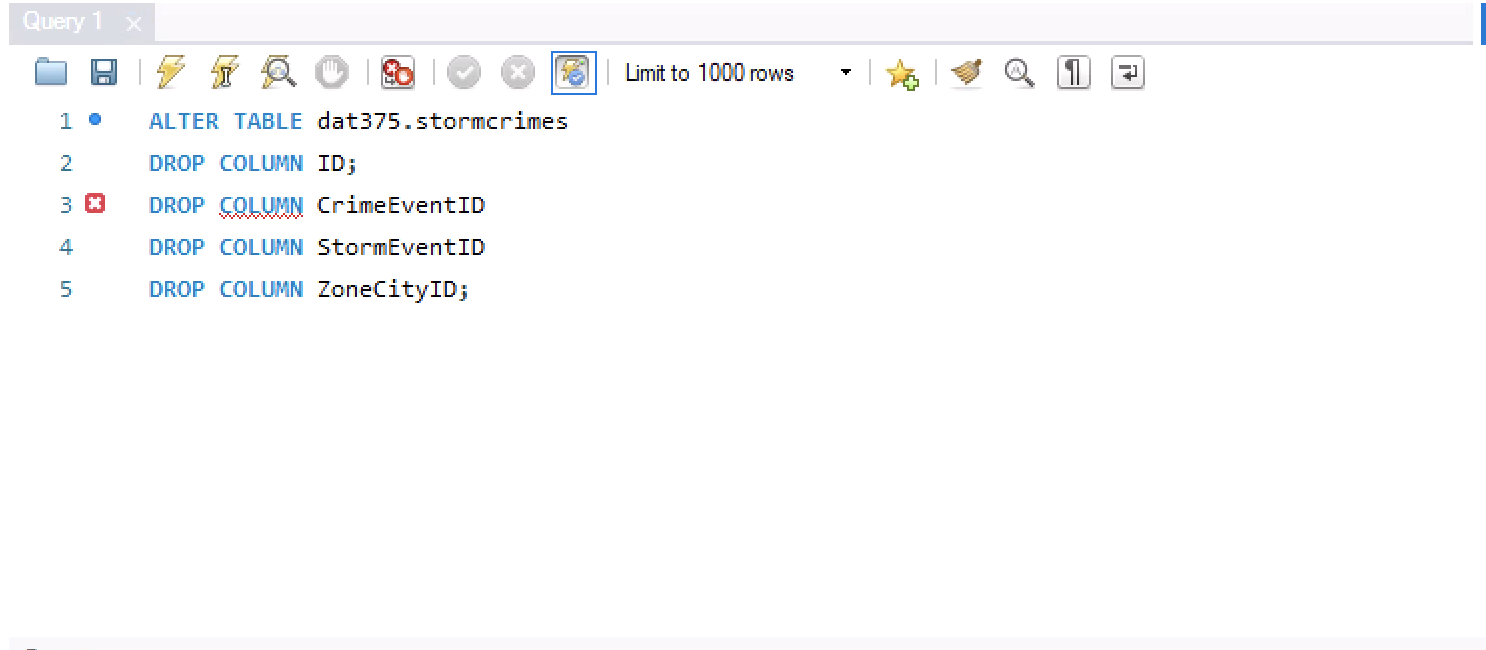
ALTER TABLE dat375.stormcrimes

DROP COLUMN ID;

DROP COLUMN CrimeEventID

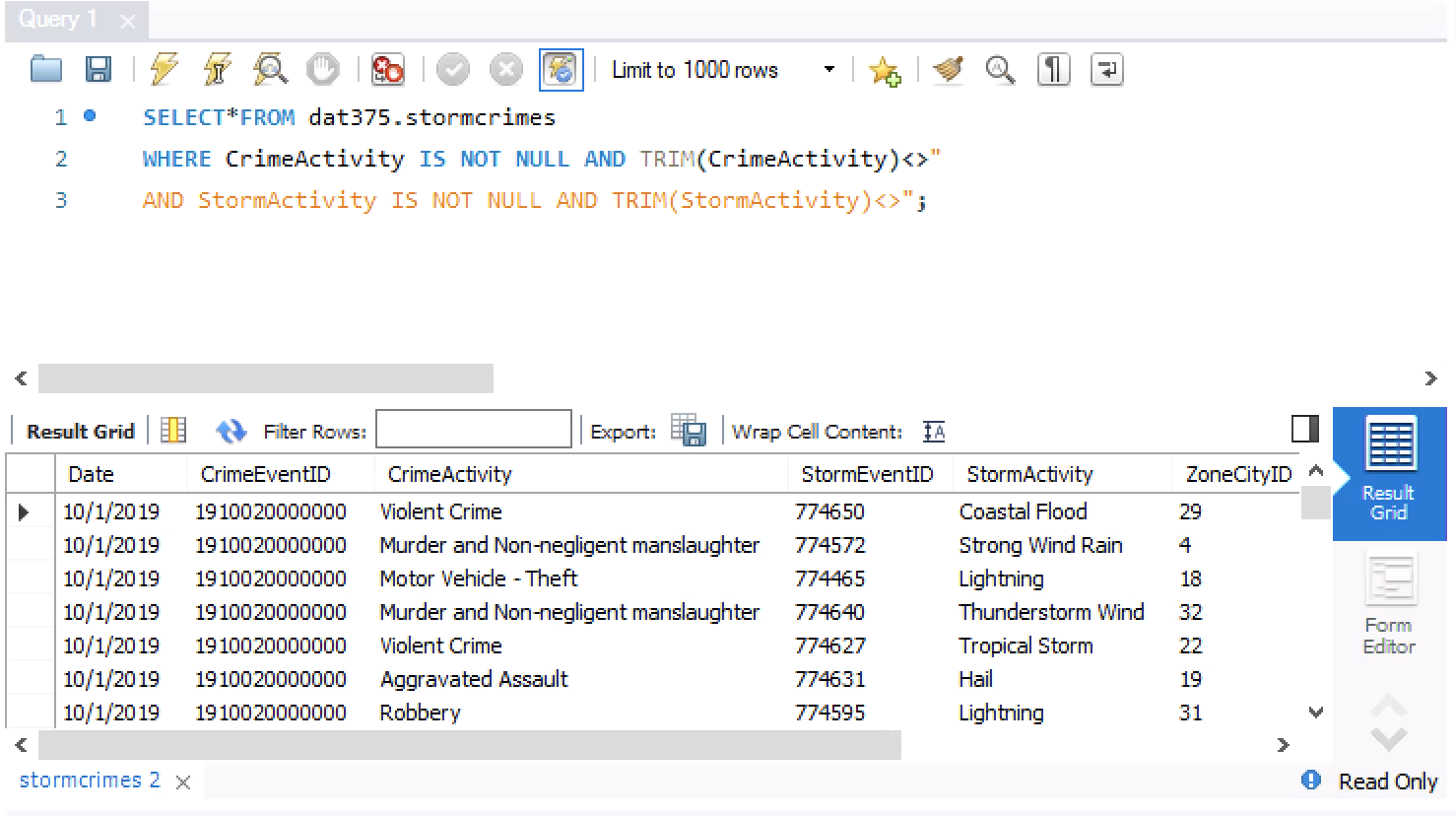
DROP COLUMN StormEventID

DROP COLUMN ZoneCityID;



SELECT\*FROM dat375.stormcrimes

WHERE CrimeActivity IS NOT NULL AND TRIM(CrimeActivity)<>”

AND StormActivity IS NOT NULL AND TRIM(StormActivity)<>”;

Separating this information allows us to present the data to the stakeholders in a much neater way. We also only need to provide the necessary data for the problem so as a part of the data cleaning process we remove the unnecessary data.

References

Southern New Hampshire University. (n.d.) *DAT 375 Project One Tutorial.* <https://learn.snhu.edu/content/enforced/1461927-DAT-375-T3045-OL-TRAD-UG.24EW3/course_documents/DAT%20375%20Project%20One%20Tutorial.pdf?isCourseFile=true&ou=1461927>

Gibson, P. (n.d.) *Types of Data Analysis.* Chartio. <https://chartio.com/learn/data-analytics/types-of-data-analysis/>

Rawat, S, A. (Mar 31, 2021) *An Overview of Descriptive Analysis.* analyticssteps. <https://www.analyticssteps.com/blogs/overview-descriptive-analysis>

Tableau (n.d.) *What Is Data Visualization? Definition, Examples, And Learning Resources.* <https://www.tableau.com/learn/articles/data-visualization>

ProjectPro (Jan 05, 2024) *What is Data Validation in Python?* <https://www.projectpro.io/recipes/perform-data-validation-python-by-processing-only-matched-columns#mcetoc_1hjc8dae2m>