# Nashville Traffic Accidents

For this exercise, you have been provided a dataset of traffic accidents that occurred in Davidson County which was retrieved from https://data.nashville.gov/datasets/1df26ca05f3f407e95ab5a4fc013eab7\_0/explore.

This spreadsheet has been divided into the following sheets:

\* Accidents: Contains the raw data for each accident. In order to familiarize yourself with what is contained in each column, take some time to review the data dictionary, contained in Traffic-Accidents-Metadata.pdf.

\* Analysis: This sheet will hold most of your answers.

\* Zero Car Crashes: This sheet will be used in question 3.

\* Collision Types: Contains the text description of each collision type. You will fill in your answers for question 5 here.

\* Weather Types, Illumination Types, and Harmful Types: Contain descriptions for the other three codes. Not used in this exercise.

Use formulas to answer each question. Unless otherwise stated, fill in your answers in the "Analysis" tab.

1. How many total accidents are contained in this dataset?

2. What are the earliest and latest records that appear in this dataset?

3. a. Create a new column to the right of the "Number of Motor Vehicles" column called "Single or Multiple". This column should contain "Single" if the number of vehicles is 1 and "Multiple" if it involved more than one vehicle.

* What about if “Number of Motor Vehicles” is 0? Will it be categorized in “Multiple?”[ST1]

b. Are there any rows that involved zero vehicles? How many? Make sure that your formula accounts for these cases.

c. Investigate the rows that have zero vehicles using the FILTER function in the "Zero Car Crashes" sheet. What do you find?

* All “Multiple”
* Most accidents b/t 2021-2022

d. What percentage of crashes are single-car?

4. How many accidents occurred which are hit and run and had at least one injury?

5. a. What is the overall average number of injuries?

b. Go to the "Collision Types" sheet and fill in the table to find the total number of crashes, average number of injuries, and total number of injuries per collision type. For each calculation, write a single formula and copy it down the table. What do you find? (Hint: If you're not sure how to answer this question, revisit the "Conditional functions and lookups" chapter of [Data Analysis in Spreadsheets](https://app.datacamp.com/learn/courses/data-analysis-in-spreadsheets).)

6. Add four new columns, Month, Year, Hour, and Weekday to the right of the Date and Time column. Use the [TEXT function](https://support.microsoft.com/en-us/office/text-function-20d5ac4d-7b94-49fd-bb38-93d29371225c) to extract out the Month, Year, Hour, and Weekday from the "Date and Time" column.

7. Fill in the Hour table in the Analysis spreadsheet to find the number of accidents that occurred for each hour of the day. Again, write a single formula and copy it down the table. Do you see anything unusual? What might be the explanation for this?

8. Do the same for the year and day of the week. What stands out?

9. Add a column to the right of the Collision Type Code called "Collision Type". Use the table contained in the Collision Types sheet to fill in this column.

10. Which interstate has the most accidents between I-24, I-40, I-65, and I-440? Answer this by counting the number of accidents that contain the strings "I 24", "I 40", "I 65", or "I 440" in their Street Address. Hint: You may need to make use of [wildcards]([https://support.microsoft.com/en-us/office/using-wildcard-characters-in-searches-ef94362e-9999-4350-ad74-4d2371110adb](https://www.datacamp.com/cheat-sheet/getting-started-with-excel-cheat-sheet)) in combination with the CONCAT function to answer this. Then do the same but search for "I24", "I40", "I65", and "I440" and then "I-24", "I-40", "I-65", and "I-440". Sum the results to get a total count.