# Nashville City Cemetery

The city of Nashville provides a dataset of known burials in city cemeteries from 1846 through 1979. This dataset holds factual information, but it also offers a fascinating glimpse into historical trends in medicine, literacy, racial equality, and more.

The Nashville City Cemetery Association has approached you asking to analyze some of their data. This analysis will be used to compare the attributes of The Nashville City Cemetery with other historic cemeteries throughout the country. Note that much of this data comes from handwritten record books, so watch out for inconsistencies, typographical errors, etc. They realize their data is messy and there will be times during your analysis in which you will need to decide how data cleaning will be handled. Be sure to explain your cleaning process and consider the pros and cons of your chosen method. The association realizes that completely cleaning the data would be too time intensive and they are willing to sacrifice some accuracy for speed.

Before diving into the data, take some time to orient yourself to what each column contains here: [Nashville City Cemetery Metadata](https://data.nashville.gov/api/views/diwx-fgj7/files/963c111e-2e2a-4133-a807-2becd590feb2?download=true&filename=Historic-Nashville-City-Cemetery-Interments-1846-1979-Metadata-v2.pdf).

The questions that follow provide you with some guidance for the initial analysis of the data. As you are working, create a new sheet in your Excel file for each pivot table or analysis work done to create a visualization, and give each sheet a descriptive/meaningful name.

After working through these questions, you will shift gears to create marketing materials that have a data focus. Think of an interesting question that you can ask of the dataset, then report your findings and one or two visualizations that go along with your story in an engaging format. The Nashville City Cemetery has said that they are flexible with the marketing medium (social media post, marketing email, flyer, etc.), but they state it should be something short, focused, and easily disseminated. You will share your final product in a "show-and-tell" style presentation. (Note, your presentation should NOT include the work you completed in answering the original analysis questions. You will only share what you created that the organization can use for marketing.)

![top 10 causes](assets/top\_10\_death\_causes.png)

1. Use a pivot table to find the 10 most common (known) recorded causes of death, and evaluate the counts of each type. Once you have your metrics, plot these in a \*\*bar chart\*\*. In the analysis of the top 10 causes of death, you may see spelling mistakes that are affecting your counts. For example, you can assume Cholera and Cholrea are the same cause of death. Create a new column in the original dataset to update spelling errors to make your count of the top 10 causes more accurate. You will need to \*\*update and refresh your pivot table\*\* to see changes applied. (For the new column you create, copy the column with `Cause of Death/Burial` and correct spelling errors only for causes that appear in your top 10 list).

2. Create a line chart showing the number of burials per year. How would you describe the overall trend in number of burials per year? In what years were there the most burials? Can you think of any plausible reason why?

3. Create a pivot table that shows the number of burials per year per month. Your rows should be years and columns should be months.

a. Apply conditional formatting to change the background color based on the number of burials. You can read about using conditional formatting here: https://support.microsoft.com/en-us/office/use-conditional-formatting-to-highlight-information-fed60dfa-1d3f-4e13-9ecb-f1951ff89d7f

b. Drill down into any months that stand out to investigate. What do you find?

4. Create a "Decade" column based off of the Burial Year column. Use a pivot table to examine burials for each decade. Look at the total number of burials and the proportion (aka percentage) of male burials and female burials. Create a clustered bar chart to show how male and female burials have changed over time. If you could use a refresher on percentages and proportions, here are some resources:

\* https://www.wikihow.com/Calculate-Percentages

\* https://www.wikihow.com/Solve-Proportions

5. Next look at how age at the time of death has changed over time. Add a column to the original dataset to classify each row to one of the following categories (0-18, 19-25, 26-40, 41-64, and 65+). Do this by creating a lookup table with VLOOKUP. Be sure to think about a strategy to deal with missing values. Also, create a "Year Group" dividing the records up based on burial year using a table and a VLOOKUP. Use the following groups: Before 1880, 1880-1900, 1901-1920, After 1920. Use a grouped bar chart to show how the proportion of burials per age group has changed over time. How would you describe the trends in terms of proportion of burials per age group over time?

6. Create a new column titled Last Name. Extract the last name from the Name column by subsetting to all characters to the left of the comma (see the DataCamp exercise titled "String Information - LEN, SEARCH" from the Data Analysis with Spreadsheets if you need help with this). This will result in many errors for rows missing commas.

a. Drill down to those rows without a comma - what do you notice?

Descriptors used, presumably due to missing name data 🡪 Infants mainly

b. What are the most common last names of people buried in this cemetery? Smith, Brown, Johnson

c. There was a particularly famous person buried in this cemetery. Can you find that person? Ex President James K. Polk

7. The Civil War ended in April, 1865, with the last full year of fighting occurring in 1864.

a. Create an additional sheet using FILTER() to pull only the burials that occurred in 1864. Name this new sheet "Burials\_1864". In this new sheet, create a new column called age\_category. If the individual was age 0 put "Infant". If the individual was not an infant but under age 18, put "Child". Everyone aged 18 and above should be listed as "Adult".

b. Find the Mean() age for those classified as "Child" and then again for those classified as "Adult".

c. Create a burial\_date column using the DATE() function to combine the burial year, month, and day. What do you notice when you apply this function to your data?

d. Find the last infant to be buried in 1864. Do this by first finding the maximum date that an infant was buried using the MAXIFS function. Then use FILTER to find the infants that were buried on that date. You'll need to apply multiple conditions in your filter. This can be accomplished by multiplying the desired conditions. See [https://www.spreadsheetclass.com/excel-filter-function/](https://www.spreadsheetclass.com/excel-filter-function/%20) for more information about applying complex conditions inside the filter function.