 **📚 Lesson** |Cleaning & Organizing Data

— Cleaning and Interpreting Data

PROMPT: Let’s practice creating our own calculated fields in SQL. For this practice, we’ll be using several different schemas in our queries. Pay attention to the phrasing of each question so that you know which schema to use.

1. The **election\_candidates** table contains information about major candidates who have run for President of the United States. Write a query that extracts only each candidate’s last name from the **candidate** column.

|  |
| --- |
| SELECT  SUBSTR(candidate, 1, INSTR(candidate, ',') - 1) AS last\_name  FROM election\_candidates |

1. Using the lyft\_baywheels dataset, write a query that calculates the difference, in seconds, between the started\_at and ended\_at columns and call this column **duration\_in\_seconds**. Hint: Take a look at the .strftime documentation [here](https://www.sqlite.org/lang_datefunc.html).

|  |
| --- |
| select  strftime('%s', ended\_at) - strftime('%s', started\_at) AS duration\_in\_seconds  from lyft\_baywheels |

1. Can you edit your query above to create another column called **duration\_in\_minutes**? Bonus if you can round this to the nearest hundredth place!

|  |
| --- |
| select  strftime('%s', ended\_at) - strftime('%s', started\_at) AS duration\_in\_seconds,  round(cast((strftime('%s', ended\_at) - strftime('%s', started\_at)) as float)/60,2) AS duration\_in\_minutes  from lyft\_baywheels |

1. The life expectancy of a country is an important indicator that tells you a lot about the health, education level, and economic status of its citizens. Instead of using the raw number, you are tasked to put each country into “tiers” using a calculated field. Create a calculated field in SQL that captures any country that has a life expectancy above 80 as ‘high’, between 72 and 79 as ‘medium’, and less than 72 as ‘low’.

Use the **gapminder\_life\_expectancy** dataset to write a query that returns your new calculated field and the name of the country for the year 2019.

|  |
| --- |
| select  country,  case when life\_expectancy >= 80.0 then 'high'  when life\_expectancy between 72 and 79 then 'medium'  else 'low' end as life\_expectancy\_rating  from gapminder\_life\_expectancy  where year = 2019 |