**Data Exploration - Making Sense of Google Search Data**

I've gone ahead and already added the import statements and created the four different DataFrames in this notebook. Your first step is to explore the data, by getting an understanding of what's actually in those .csv files for this project.

Start with df\_tesla, then have a look at df\_unemployment and finally, check out the two bitcoin DataFrames.

Challenge

Try to answer these questions about the DataFrames:

* What are the shapes of the DataFrames?
* How many rows & columns do they have?
* What are the column names?
* What is the largest number in the search data column? Try using the .describe() function.
* What is the periodicity of the time series data (daily, weekly, monthly)?

.

.

..

...

..

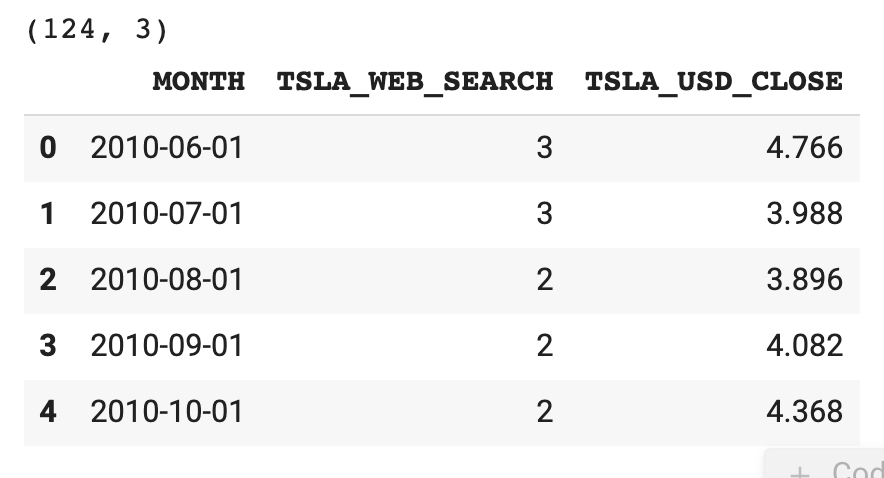
.

.

**Solution for Tesla**

The df\_tesla DataFrame has 124 rows and 3 columns: for the Month, the search popularity and the closing price of the Tesla stock.

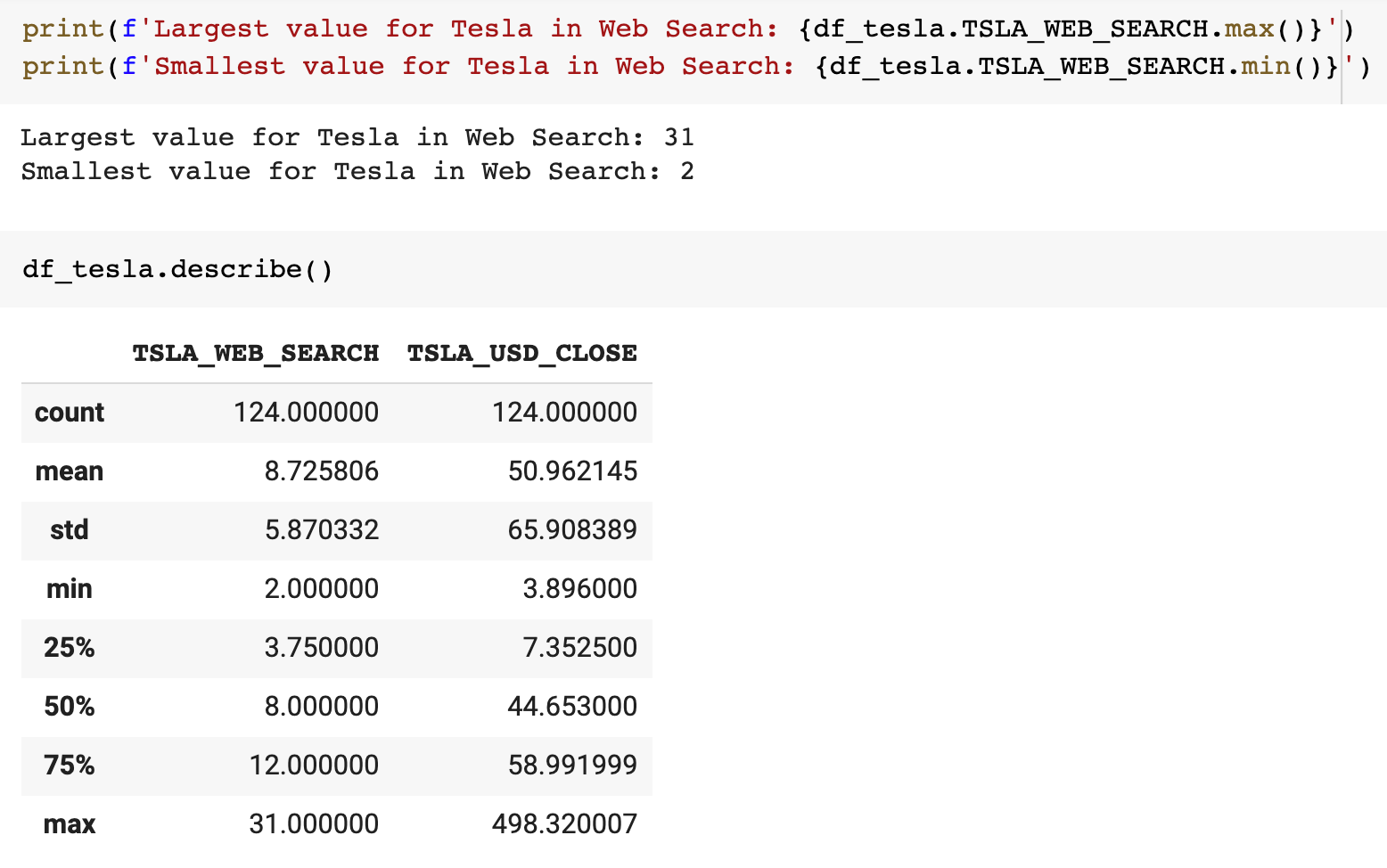
1. print(df\_tesla.shape)
2. df\_tesla.head()



You can use the max() and min() functions to see that the largest value in the search column is 31 and the smallest value is 2.

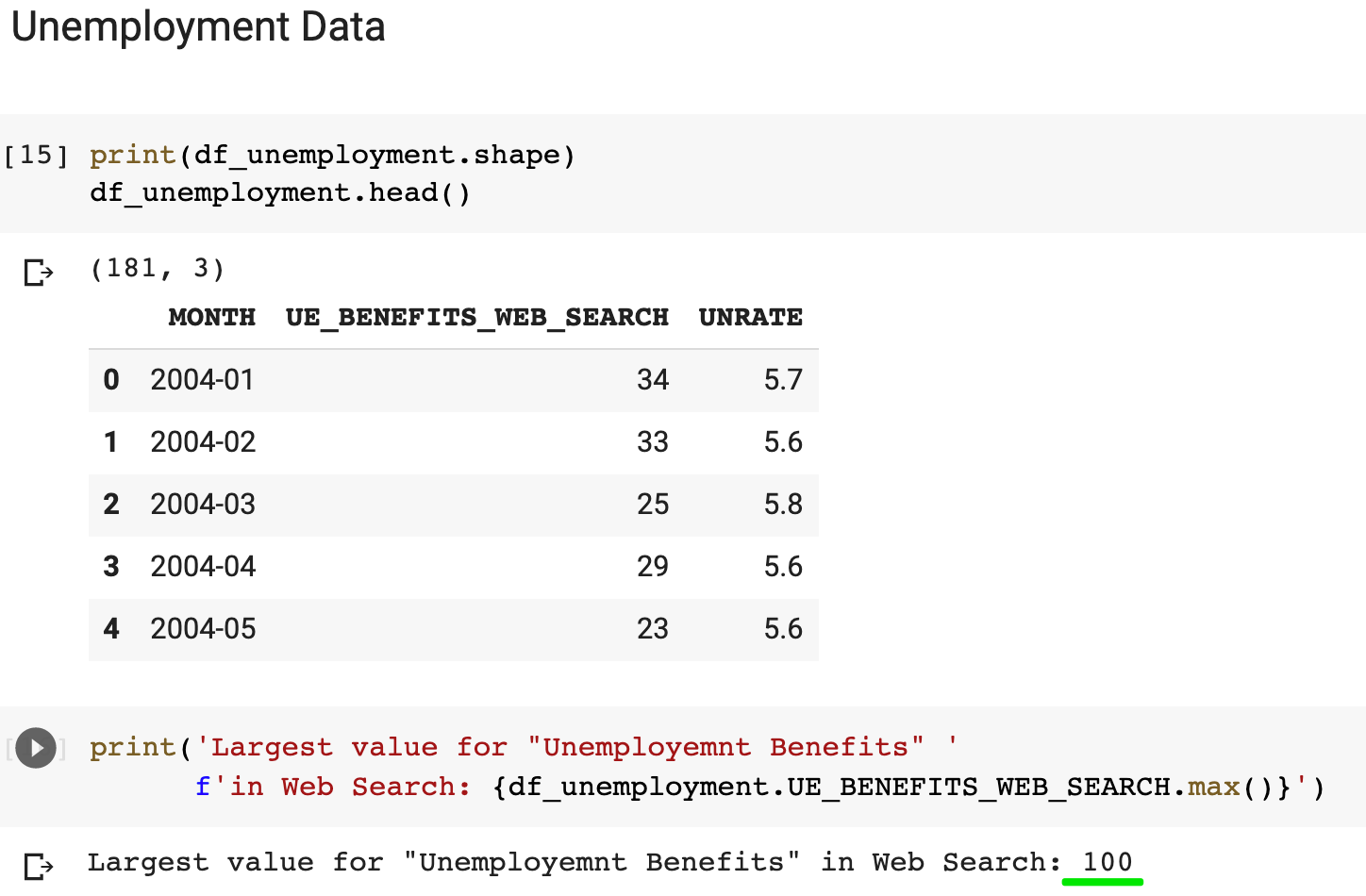
1. print(f'Largest value for Tesla in Web Search: {df\_tesla.TSLA\_WEB\_SEARCH.max()}')
2. print(f'Smallest value for Tesla in Web Search: {df\_tesla.TSLA\_WEB\_SEARCH.min()}')

One of my favourite functions to run on DataFrames is .describe(). If you use df\_tesla.describe(), you get a whole bunch of descriptive statistics. Right off the bat.



**Solution for Unemployment**

The unemployment DataFrame has 181 rows and 3 columns. As with Tesla, we have monthly data from 2004 onwards, organised in rows. Interestingly here, the largest value in the search column is 100.



**Solution for Bitcoin**

With the Bitcoin data we see that we have two different .csv files. One of them has the day-by-day closing price and the trade volume of Bitcoin across 2204 rows. The other has the monthly search volume from Google Trends.



What do the Search Numbers mean?

We can see from our DataFrames that Google's search interest ranges between 0 and 100. But what does that mean? Google defines the values of search interest as:

Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term.

Basically, the actual search volume of a term is not publicly available. Google only offers a scaled number. Each data point is divided by the total searches of the geography and time range it represents to compare relative popularity.

For each word in your search, Google finds how much search volume in each region and time period your term had relative to all the searches in that region and time period. It then combines all of these measures into a single measure of popularity, and then it scales the values across your topics, so the largest measure is set to 100. In short: Google Trends doesn’t exactly tell you how many searches occurred for your topic, but it does give you a nice proxy.

Here are the Google Trends Search Parameters that I used to generate the .csv data:

* "Tesla", Worldwide, Web Search
* "Bitcoin", Worldwide, News Search
* "Unemployment Benefits", United States, Web Search

