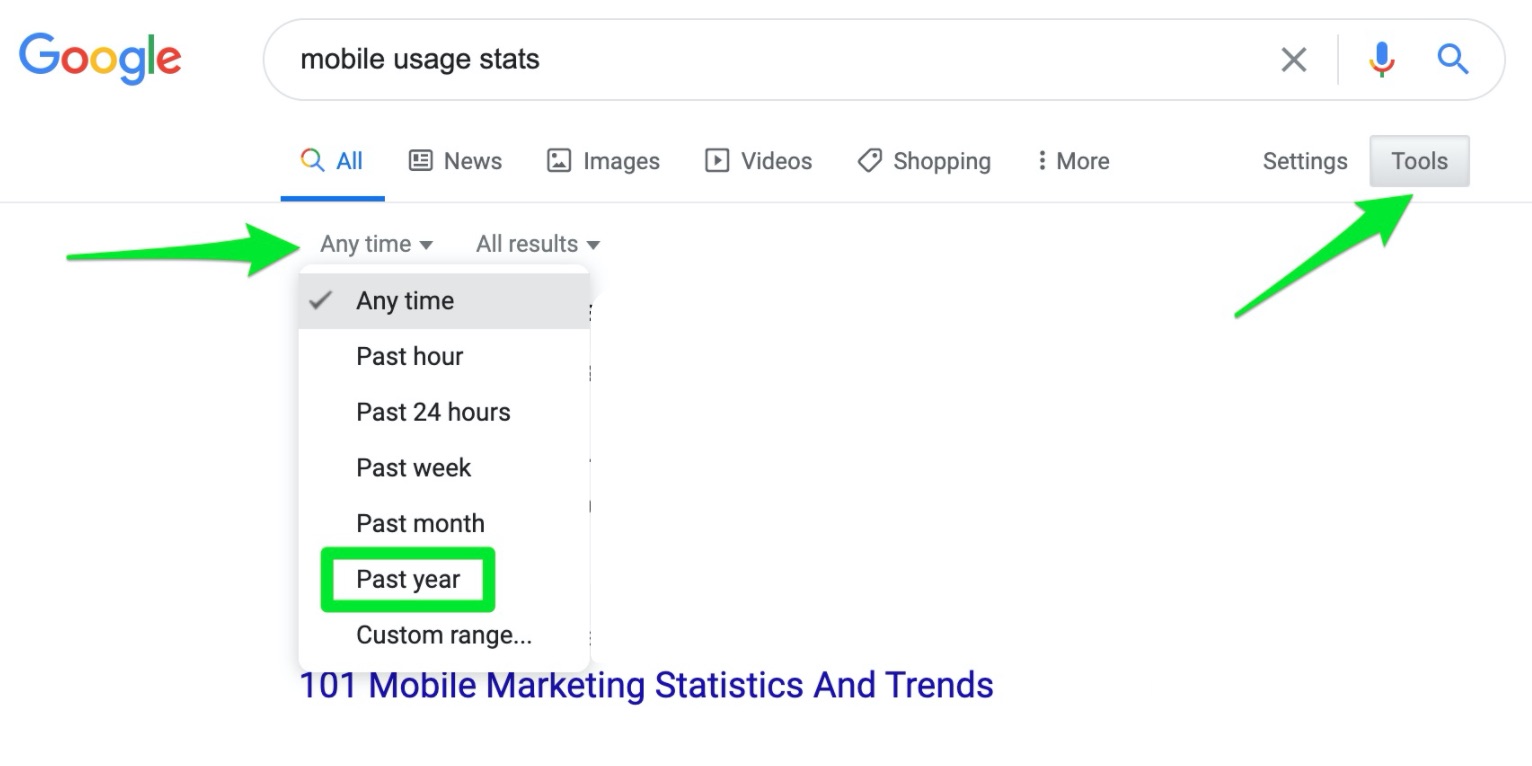
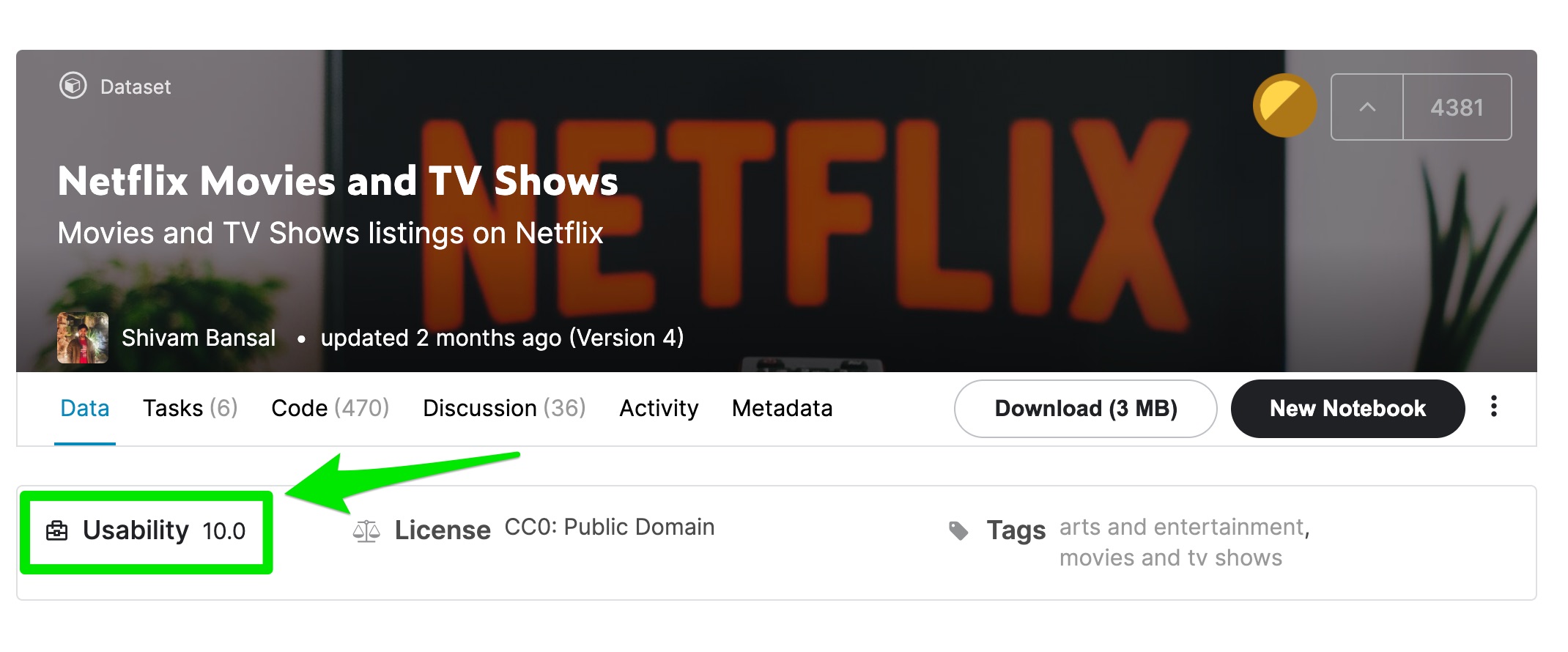
## Step 2: Consider Data

You may already have a dataset in mind to use for your project or you may still be searching and considering which dataset to use.

Data collection is a very important step in the data science life cycle because all findings, conclusions, and decisions are made using this data! If the data used is not good or reliable, then the solution will not be good or reliable as well.

In this item, you will list the source(s) used for your data, followed by a 2-3 sentences explanation of why the data is reliable and can be trusted.

**How to Find Reliable Statistics Sources**

1. Typically, official government and educational sources are reliable. These sites will have a web address that ends in .gov or .edu. To search for these sites only, use the following search trick:  
   * Type in site:gov OR site:edu along with your search topic.  
     + *Example*: site:gov OR site:edu mobile usage stats
2. You should also check to see if the data you are using is current. If using Google, you can search for current search results by clicking on “Tools” and then the “Any time” dropdown and choosing “Past year”.  
   
3. If using [kaggle.com](https://www.kaggle.com/datasets), there is a usability score that corresponds with each dataset. This is very helpful! The highest usability score is a 10.  
   

Possible Data Sources to Use:

* <https://www.kaggle.com/>
* <https://www.gapminder.org/data/>
* <https://www.census.gov/data.html>
* <http://data.un.org/>
* <https://www.healthdata.gov/>
* <https://www.cdc.gov/datastatistics/>

1. List the source(s) used for your data.

2. Provide a 2-3 sentence explanation of why the data is reliable and can be trusted.