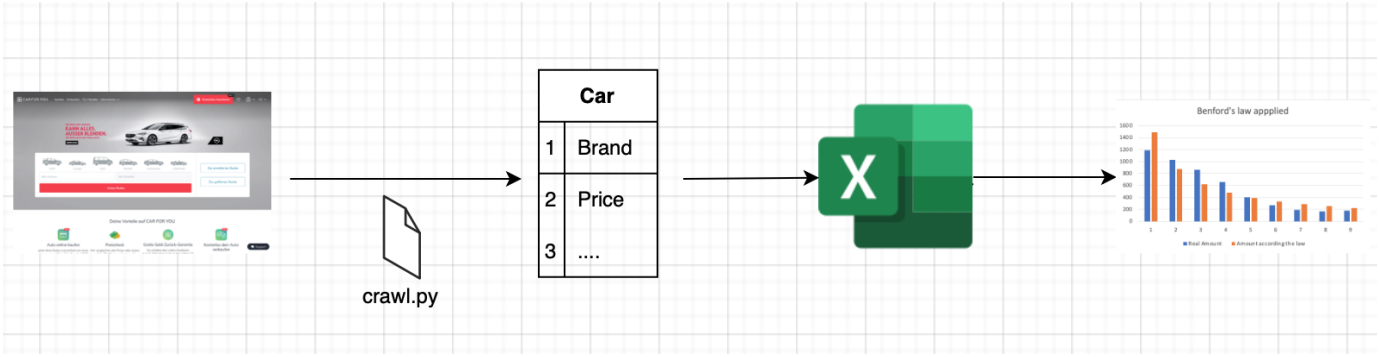
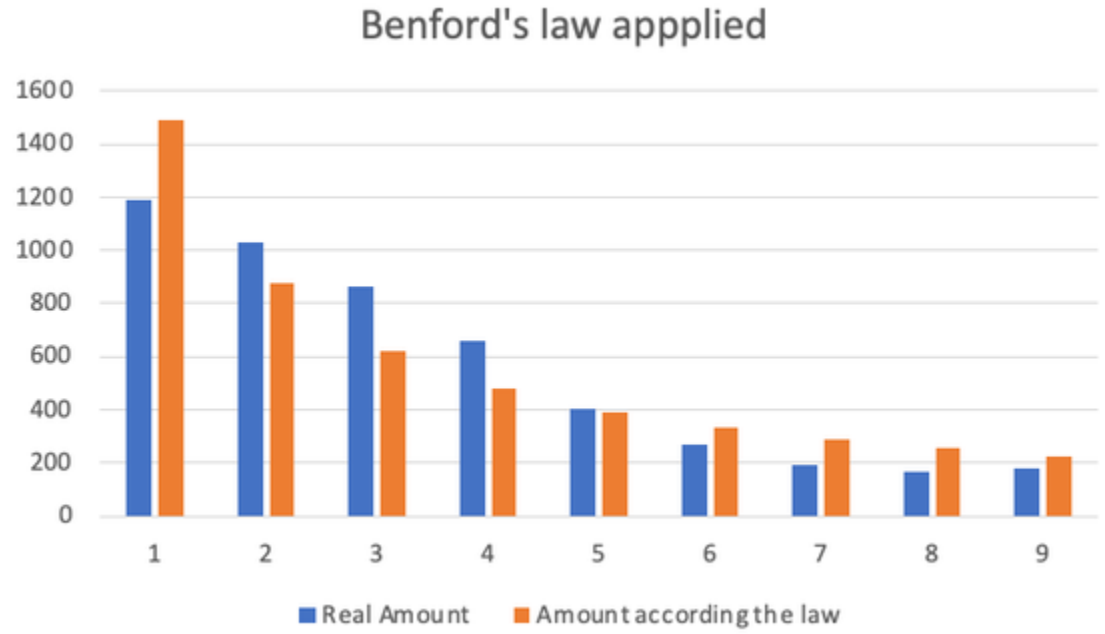
Benford's Law applied

The following paragraph digs deeper and makes a real life proof of concept of the Benford's law. Real data is needed. Therefore we have constructed a python crawler that visits the website https://www.carforyou.ch. A website where private or business users can offer their car to potential buyers. In a nutshell, some people offer a car others buy.

The diagram below shows the flow of the data. At first a python crawl.py program visits the car website downloads the entire page and search for the relevant attributes. The price of a car seams a reasonable attribute to run a Benford's law analysis over it. The outcome of the crawler program is a list with around 5000 cars (depending on the crawler settings). This list is fed to Excel and the occurrence of each first digit is counted.



Furthermore a bar chart is drawn which consist of two colored bars. Orange bars indicate the distribution predicted by Benford's law. Whereas blue bars represents real life occurrence. The total amount of cars crawled is n=4959.



Although the data basis is pretty small (n=4959) it seems that the frequency distribution of leading digits is very close the predicted one by Benford's law. The trend that higher digits appears less often can be observed, but digits as 1,6,7,8,9 are rather less often in the sample dataset as Benford's law would suggest. The distribution bests fits the digit 5.

The analysis was done by using the files attached:

benfords_law_applied.xlsx

cars_crawl.pdf