PDFs and Spreadsheets Puzzle Exercise

You will need to work with two files for this exercise and solve the following tasks:

- Task One: Grab the Google Drive link from the .csv file. (Hint: Its along the diagonal).
- Task Two: Download the PDF from the Google Drive link (we already downloaded it for you just in case you can't download from Google Drive) and find the phone number that is in the document. Note: There are different ways of formatting a phone number!

Task One: Grab the Google Drive Link from .csv File

In [ ]:
```
import csv
```

Grab all the lines of data.

In [12]:
```
data = open('Exercise_Files/find_the_link.csv',encoding="utf-8")
csv_data = csv.reader(data)
data_lines = list(csv_data)
```

We can see its along the diagonal, which means the values are at the index position that matches the row's number order. So the 1st letter is the 1st item in the 1st row, the 2nd letter is the 2nd item in the 2nd row, the 3rd item is the 3rd letter in the 3rd row and so on. We can use enumerate to track the row number and simply index off the data_lines.

Method One

In [13]:
```
link_list = []
for row_num,data in enumerate(data_lines):
    link_list.append(data[row_num])
```

In [14]:
```
''.join(link_list)
```

Out[14]:
```
'https://drive.google.com/open?id=1G6SEgg018UB4_4xsAJJ5TdzrhmXipr4Q'
```

Method Two

In [15]:
```
link_str = ''
for row_num,data in enumerate(data_lines):
    link_str+=data[row_num]
```

In [18]:
```
link_str
```

Out[18]:
```
'https://drive.google.com/open?id=1G6SEgg018UB4_4xsAJJ5TdzrhmXipr4Q'
```

Task Two: Download the PDF from the Google Drive link and find the phone number that is in the document.

In [19]:
```
import PyPDF2
```

In [20]:
```
f = open('Exercise_Files/Find_the_Phone_Number.pdf','rb')
```

In [21]:
```
pdf = PyPDF2.PdfFileReader(f)
```

In [22]:
```
pdf.numPages
```

Out[22]:
```
17
```

Phone Number Matching

Lot's of ways to do this, but you had to figure out the phone number was in format ###.###.####

Hint: https://stackoverflow.com/questions/4697882/how-can-i-find-all-matches-to-a-regular-expression-in-python

In [1]:
```
import re
```

In [2]:
```
pattern = r'\d{3}'
```

In [ ]:
```
all_text = ''

for n in range(pdf.numPages):

    page = pdf.getPage(n)
    page_text = page.extractText()

    all_text = all_text+' '+page_text
```

In [ ]:
```
for match in re.finditer(pattern,all_text):
    print(match)
```

Once you know the correct pattern:

In [23]:
```
import re
```

In [24]:
```
pattern = r'\d{3}.\d{3}.\d{4}'
```

In [26]:
```
for n in range(pdf.numPages):

    page  = pdf.getPage(n)
    page_text = page.extractText()
    match = re.search(pattern,page_text)

    if match:
        print(match.group())
```

```
505.503.4455
```

Great Job! Information on this phone number:

- https://www.businessinsider.com/better-call-saul-billboard-and-phone-number-2014-7
- https://www.reddit.com/r/betterCallSaul/comments/4awouf/heres_a_list_of_real_numbers_you_can_call_
- https://www.amc.com/shows/better-call-saul/talk/2020/03/saul-goodmans-phone-number-is-the-latest-breaking-bad-callback