other salaries got the same salary as the first salary

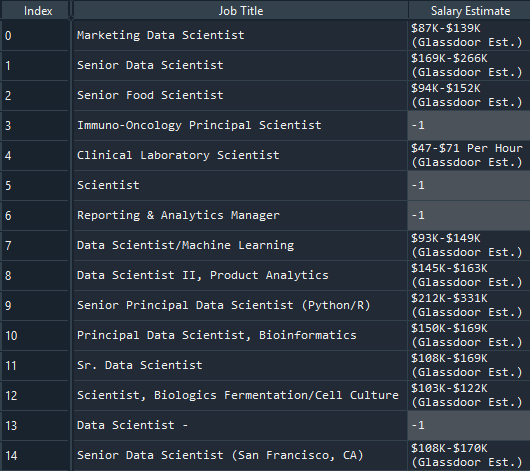
**code:**

salary\_estimate = driver.find\_element\_by\_xpath('.//span[@class="gray salary"]').text



**code:**

salary\_estimate = driver.find\_element\_by\_xpath('.//\*[@id="HeroHeaderModule"]/div[3]/div[1]/div[4]/span').text



Open GitBash

go to the file “ds\_salary\_proj” location

enter the following statements:

echo “# ds\_salary\_proj” >> README.md

git add . 🡪 add all the files in the folder

git commit -m "uploaded scraper and run code"

git remote add origin https://github.com/slothpunch/ds\_salary\_proj.git

git push -u origin master

I tried to scrap 500 salaries, but the scraper stopped. No idea why.

I deleted some code associated with Headquarter and Competitor as the website is no longer showing those labels, and switched my internet connection to Wi-Fi from LAN as even though I’m using LAN cable, it disconnects when trains pass… (I’m living by a train station)

Collecting 500 jobs was successful, but I overlooked the fact that there were not enough jobs as many as I wanted. Eventually, I got 500 jobs, but when the scraper finished the last page, the total number of jobs was 410. And as there was no next page, the scraper iterated the code in the last page, so it gathered the same jobs about 90 jobs from the last page until finished. I should have checked if there was enough data as much as I wanted.

The reason that we are making different versions and different branches is that we don’t have to conflict with the master. So, we can mess up, do whatever we want in this branch and the master branch will be perfectly fine. We can just revert back to that it we want to start over again. We can reconcile those changes when we merge it.

git add .

git commit -m ‘updating data and data collection script’

git push

git pull

git checkout -b data\_cleaning

3. Data Cleaning

I messed up with my .csv file and got this error, but after replacing new one, the error is gone.

Traceback (most recent call last):

File "pandas\\_libs\parsers.pyx", line 1130, in pandas.\_libs.parsers.TextReader.\_convert\_tokens

File "pandas\\_libs\parsers.pyx", line 1254, in pandas.\_libs.parsers.TextReader.\_convert\_with\_dtype

File "pandas\\_libs\parsers.pyx", line 1269, in pandas.\_libs.parsers.TextReader.\_string\_convert

File "pandas\\_libs\parsers.pyx", line 1459, in pandas.\_libs.parsers.\_string\_box\_utf8

UnicodeDecodeError: 'utf-8' codec can't decode byte 0x80 in position 15: invalid start byte

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "D:\Kaggle\0. Data\_Science\_Salaries\ds\_salary\_proj\data\_cleaning.py", line 11, in <module>

df = pd.read\_csv('glassdoor\_jobs.csv')

File "C:\Users\GIS\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 676, in parser\_f

return \_read(filepath\_or\_buffer, kwds)

File "C:\Users\GIS\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 454, in \_read

data = parser.read(nrows)

File "C:\Users\GIS\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 1133, in read

ret = self.\_engine.read(nrows)

File "C:\Users\GIS\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 2037, in read

data = self.\_reader.read(nrows)

File "pandas\\_libs\parsers.pyx", line 860, in pandas.\_libs.parsers.TextReader.read

File "pandas\\_libs\parsers.pyx", line 875, in pandas.\_libs.parsers.TextReader.\_read\_low\_memory

File "pandas\\_libs\parsers.pyx", line 952, in pandas.\_libs.parsers.TextReader.\_read\_rows

File "pandas\\_libs\parsers.pyx", line 1084, in pandas.\_libs.parsers.TextReader.\_convert\_column\_data

File "pandas\\_libs\parsers.pyx", line 1137, in pandas.\_libs.parsers.TextReader.\_convert\_tokens

File "pandas\\_libs\parsers.pyx", line 1254, in pandas.\_libs.parsers.TextReader.\_convert\_with\_dtype

File "pandas\\_libs\parsers.pyx", line 1269, in pandas.\_libs.parsers.TextReader.\_string\_convert

File "pandas\\_libs\parsers.pyx", line 1459, in pandas.\_libs.parsers.\_string\_box\_utf8

UnicodeDecodeError: 'utf-8' codec can't decode byte 0x80 in position 15: invalid start byte