

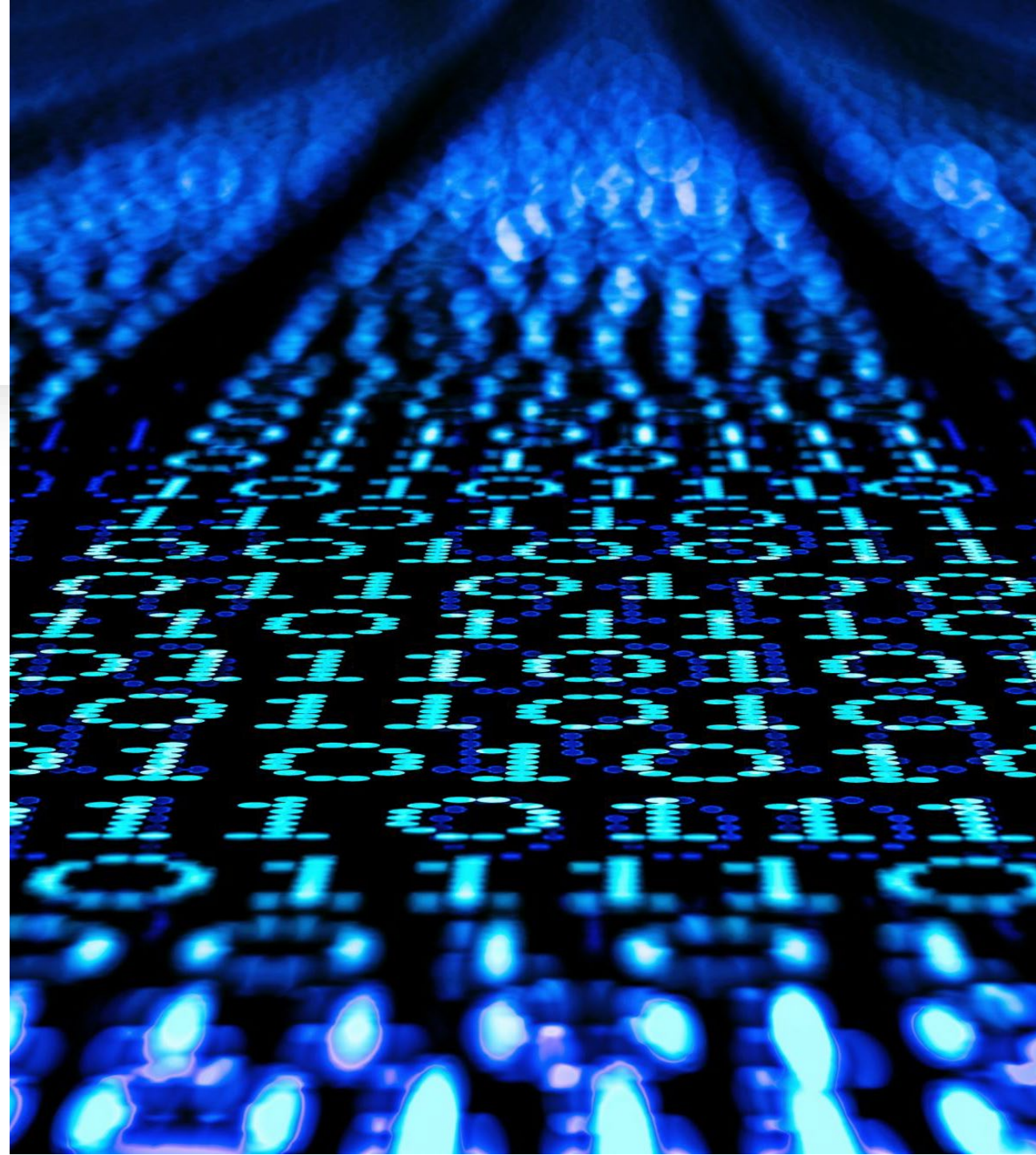
ACCT5943- Assignment Code

Download files from moodle and put them to a folder, you need

- student_scraper_new.ipynb
- filtered_sp-500.csv

You also need:

- group_number



Dependencies and packages

```
%pip install lxml
```

```
%pip install html5lib
```

```
%pip install selenium
```

```
%pip install pandas
```

```
%pip install numpy
```

```
%pip install bs4
```

```
from selenium import webdriver
```

```
from selenium.webdriver.common.by import By
```

```
from selenium.webdriver.support.ui import WebDriverWait
```

```
from selenium.webdriver.support import expected_conditions as EC
```

```
from selenium.common.exceptions import TimeoutException
```

```
from bs4 import BeautifulSoup
```

```
import pandas as pd
```

```
import numpy as np
```

```
import os
```

```
import time
```

Code structure

- `select_companies` (select xx companies from xx industries)
- `scrape_company_data` (automated process to scrape data)
- Main (where you run the functions)

Steps:

1. Update group number
2. Run all
3. During your assignment, you may want to do following things:
 - Get a list of new companies
 - Close the notebook entirely
 - Delete the txt file
 - Give your code to someone else to run

Notes

- When you first time run the code, the notebook creates a file called “selected_companies_group_{your group number}.txt”, if you want others to run the code afterwards, you will need to give this file to others so that they get the same set of companies for this assignment.
- Please make sure that Google Chrome is installed on your machine, the provided code runs on Google Chrome only
- If you encounter any technical problem, please come to the workshop seminar, if you can't make it, please post your questions on Moodle