```
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import Select
from selenium.common.exceptions import NoSuchElementException
import unittest, time, re
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.action_chains import ActionChains
import csv
import os
import random
from random import randint
import numpy


def hover(s):
    element = driver.find_element_by_id(s)
    hov = ActionChains(driver).move_to_element(element)
    hov.perform()

def hover_by_name(s):
    element = driver.find_element_by_name(s)
    hov = ActionChains(driver).move_to_element(element)
    hov.perform()


input_file = "/Users/NilMas/Documents/SamuelsDocument/linkedinInputv6.csv"
input = open(input_file,"rU")
reader = csv.reader(input)

output_file = "/Users/NilMas/Documents/SamuelsDocument/linkedinOutput.csv"
output = open(output_file, 'wb')
writer = csv.writer(output, delimiter=",")

#driver = webdriver.Chrome()

chromedriver = "/Users/NilMas/Downloads/chromedriver"
os.environ["webdriver.chrome.driver"] = chromedriver
driver = webdriver.Chrome(chromedriver)


driver.get("https://linkedin.com")
driver.find_element_by_id("login-email").clear()
driver.find_element_by_id("login-email").send_keys("xxx@gg.com")
driver.find_element_by_id("login-password").clear()
driver.find_element_by_id("login-password").send_keys("xxxyyy")
#hover("signin")
driver.find_element_by_name("submit").click()

time.sleep(30)




counter = 0
for row in reader:
    if counter == 0:
        writer.writerow(["Name", "PageURL", "HeaderPosition", "Location", "Industry",
"HeaderEducation"])

    if counter >= 1 : #starting value

        if row[0] != 'Name':

            CandidateName = unicode(row[0], errors = "ignore")
            CandidateName1 = unicode(row[1], errors = "ignore")
            CandidateEmail = unicode(row[2], errors = "ignore")
            ClickName = unicode(row[3], errors = "ignore")
            driver.get("https://linkedin.com")
            driver.find_element_by_id("main-search-box").clear()
            hover("main-search-box")
            time.sleep(5)
            driver.find_element_by_id("main-search-box").click()
            driver.find_element_by_id("main-search-box").send_keys(CandidateName)
            time.sleep(3)
            hover("main-search-box")
            driver.find_element_by_name("search").click()

            driver.implicitly_wait(5)

            try:
                driver.implicitly_wait(6)
                driver.find_element_by_partial_link_text(ClickName).click()

            except:
```

```
            driver.find_element_by_id("main-search-box").clear()
            hover("main-search-box")
            time.sleep(5)
            driver.find_element_by_id("main-search-box").click()
            driver.find_element_by_id("main-search-box").send_keys(CandidateName1)
            time.sleep(3)
            hover("main-search-box")
            try:
                driver.find_element_by_name("search").click()
                driver.find_element_by_partial_link_text(ClickName).click()
            except:
                print "ClickNameError000"
                driver.find_element_by_xpath("//h3/a[@class='title main-headline' and
contains(@href, 'www.linkedin.com/profile')]").click()
                CandidateName = "000" + CandidateName

        ##random wait time
        s = numpy.random.poisson(lam=27, size=1)
        print "AlumName: " + CandidateName
        print ("random waiting time: ", s)
        time.sleep(s) #this is the random wait
        driver.execute_script("window.scrollTo(0, document.body.scrollHeight);")
        time.sleep(s/2)
        driver.execute_script("window.scrollTo(0, document.body.scrollHeight/3);")
        time.sleep(s/3)

        pageURL = driver.current_url



        try:
            position_at_company = driver.find_element_by_id("headline").text
        except:
            position_at_company = "not found"
        try:
            industry = driver.find_element_by_name("industry").text
        except:
            industry = "not found"
        try:
            currentcity = driver.find_element_by_name("location").text
        except:
            currentcity = "not found"
        try:
            education_list = driver.find_elements_by_xpath("//a[@title='More details for this school']")
            counter = 0
            for eachcontent in education_list:
                if counter == 0:
                    education = eachcontent.text
                    counter = counter + 1
        except:
            education = ""



        try:
            jtime = driver.find_elements_by_xpath("//span[@class='experience-date-locale']")
        except: #don't handle
            pass
        try:
            jtitle = driver.find_elements_by_xpath("//h4//a[@title='Learn more about this title']")
        except: #don't handle
            jtitle=[]
        try:
            jcomp = driver.find_elements_by_xpath("//a[contains(@href, 'exp-company-name') and @dir='auto']")

        except: #don't handle
            jcomp=[]

        try:
            writeContent  = []
            print(CandidateName + " " + position_at_company + " " + currentcity + " " + industry)
            writeContent = [CandidateName , CandidateEmail, pageURL, position_at_company , currentcity , industry , education]




        except:
            print "error occured"

            writer.writerow(["python error occured"])

        #we need to figure out the array length
```

```
num = 1
ttime = []
print "time: "
for eachcontent in jtime:
    print num
    print eachcontent.text
    ttime.append(eachcontent.text)
    num = num + 1


num = 1
ttitle = []
print "title: "
for eachcontent in jtitle:
    print num
    print eachcontent.text
    ttitle.append(eachcontent.text)
    num = num + 1

num = 1
tcomp = []
print "company: "
for eachcontent in jcomp:
    print num
    print eachcontent.text
    tcomp.append(eachcontent.text)
    num = num + 1

comp_length = len(tcomp)
print "comp_length: "
print comp_length
title_length = len(ttitle)
print "title_length: "
print title_length


for ii in range(max(comp_length, title_length)): # 1 to length

    try:
        if ii < len(ttitle):
```

```
                        writeContent.append(ttitle[ii])
                except:
                    writeContent.append("out of range")
                try:
                    if ii < len(ttitle):
                        #writeContent.append(tcomp[ii + comp_length - title_length])
                        writeContent.append(tcomp[ii])
                except:
                    writeContent.append("out of range")

                try:
                    if ii < len(ttime):
                        writeContent.append(ttime[ii])
                except:
                    writeContent.append("out of range")


            Content =[]
            for x in writeContent:
                Content.append(x.encode('ascii','ignore'))
                #print x


            writer.writerow(Content)
    counter  = counter + 1
    if counter == 68:
        break



        #


input.close()
output.close()
```