**Security and Privacy in Online Social Network**

**Lab 01 - OSN Analytics**

**Due 6:00pm on 5/12/2022**

**POLICIES:**

1. **Coverage**

OSN analytics: data collection and analysis

1. **Grade**

Lab 01 accounts for 20% of the final grade

1. **Individual or Group**

Individual based, but group discussion is allowed and encouraged

1. **Academic Honesty**

Violation of academic honesty may result in a penalty more severe than zero credit for an assignment, a test, and/or an exam.

1. **Submission**

Submit your *code, collected data (in .sql format), and a report* including data collection methodology, data analysis, findings, and discussion, to course.zju.edu.cn

1. **Late Submission**

Deduction ceases upon zero;

Late submissions after 6:30pm on 5/12/2022 will NOT be graded.

**PREPARATION:**

1. **Lab Goal**

Lab 01 aims to master techniques and tools for collecting real-world data from OSN and conducting analytics.

1. **Software recommended to install:**
2. Install Visual Studio Community 2019 (for Python programming)

[Downloads - Visual Studio Subscriptions Portal](https://my.visualstudio.com/Downloads?q=visual%20studio%202019&wt.mc_id=o~msft~vscom~older-downloads)

1. Install MySQL Installer

[MySQL :: Download MySQL Installer](https://dev.mysql.com/downloads/installer/)

1. Install Python

[Download Python | Python.org](https://www.python.org/downloads/)

1. Install Sublime or Notepad++ (***optional***)

[Download - Sublime Text](https://www.sublimetext.com/download)

1. Install a browser and the corresponding WebDriver (Chrome as an example here)

Install Chrome

Install Chromedriver

<https://chromedriver.chromium.org/downloads>

Stop the automatic update option <https://www.makeuseof.com/tag/stop-automatic-chrome-updates-windows/>

1. Install the Mysqlclient:

a fork of the MySQL-python interface for the MySQL database.

[Python Extension Packages for Windows - Christoph Gohlke (uci.edu)](https://www.lfd.uci.edu/~gohlke/pythonlibs/#mysqlclient)

1. Install the following package in CMD (do not forget to set up the environment for python):

pip install -U selenium

pip install mysqlclient‑1xxx.whl

pip install requests

pip install bs4

Common error and solution:

Error message: \_*mysql\_exceptions.OperationalError: (1251, 'Client does not support authentication protocol ...*

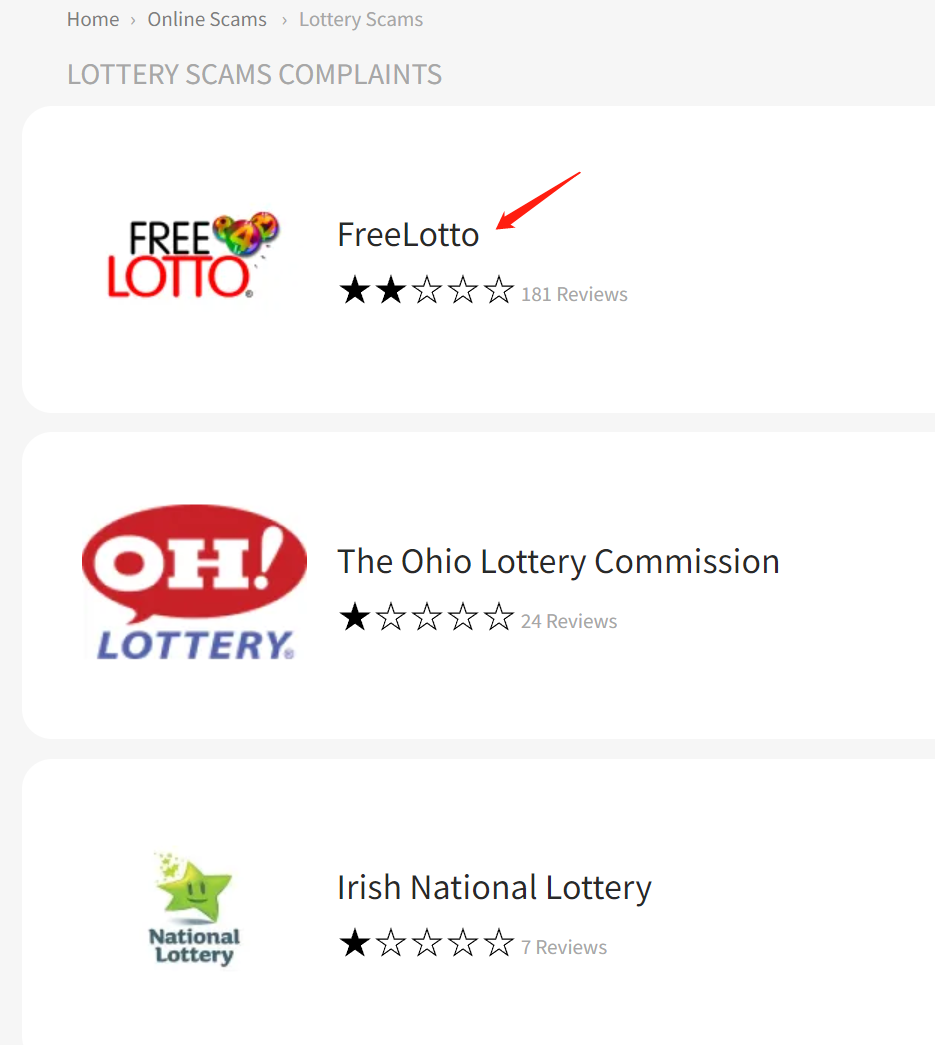
Solution: *https://www.jianshu.com/p/3c5e6264f849*

**Demonstration of What to Collect:**

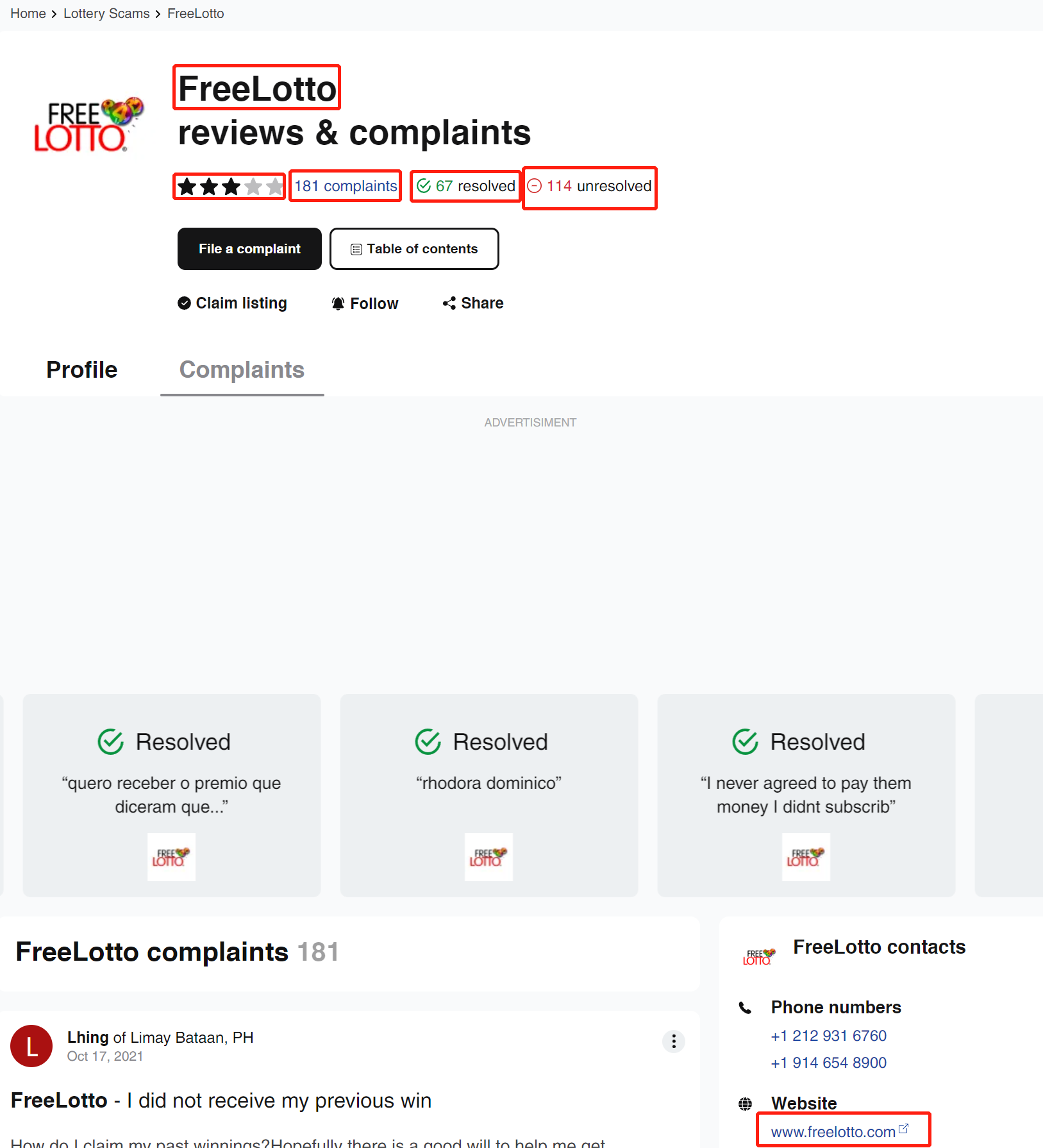
On the first layer of the destined URL

Example:

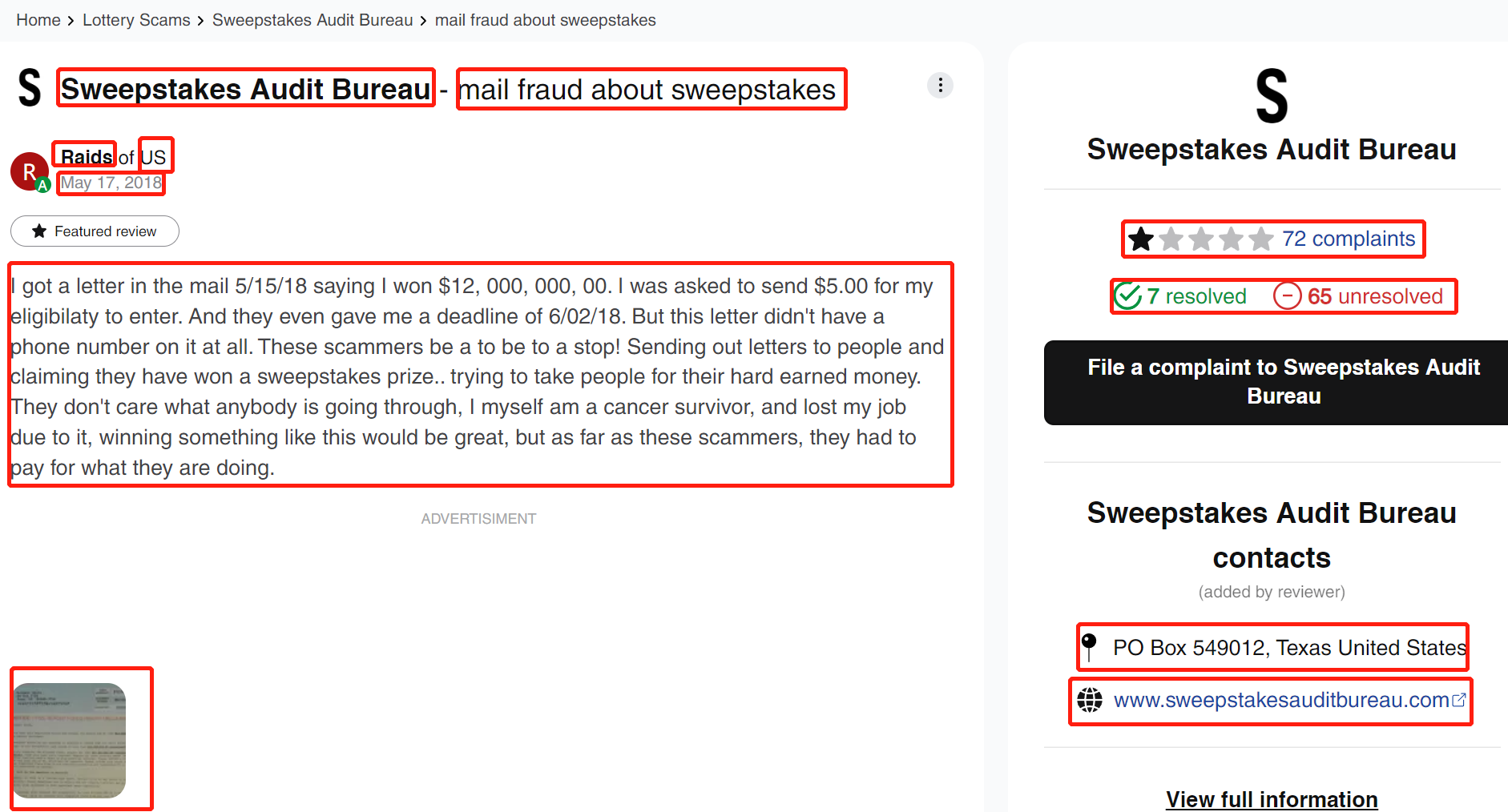
First-level page: <https://www.complaintsboard.com/bysubcategory/lottery-scams>



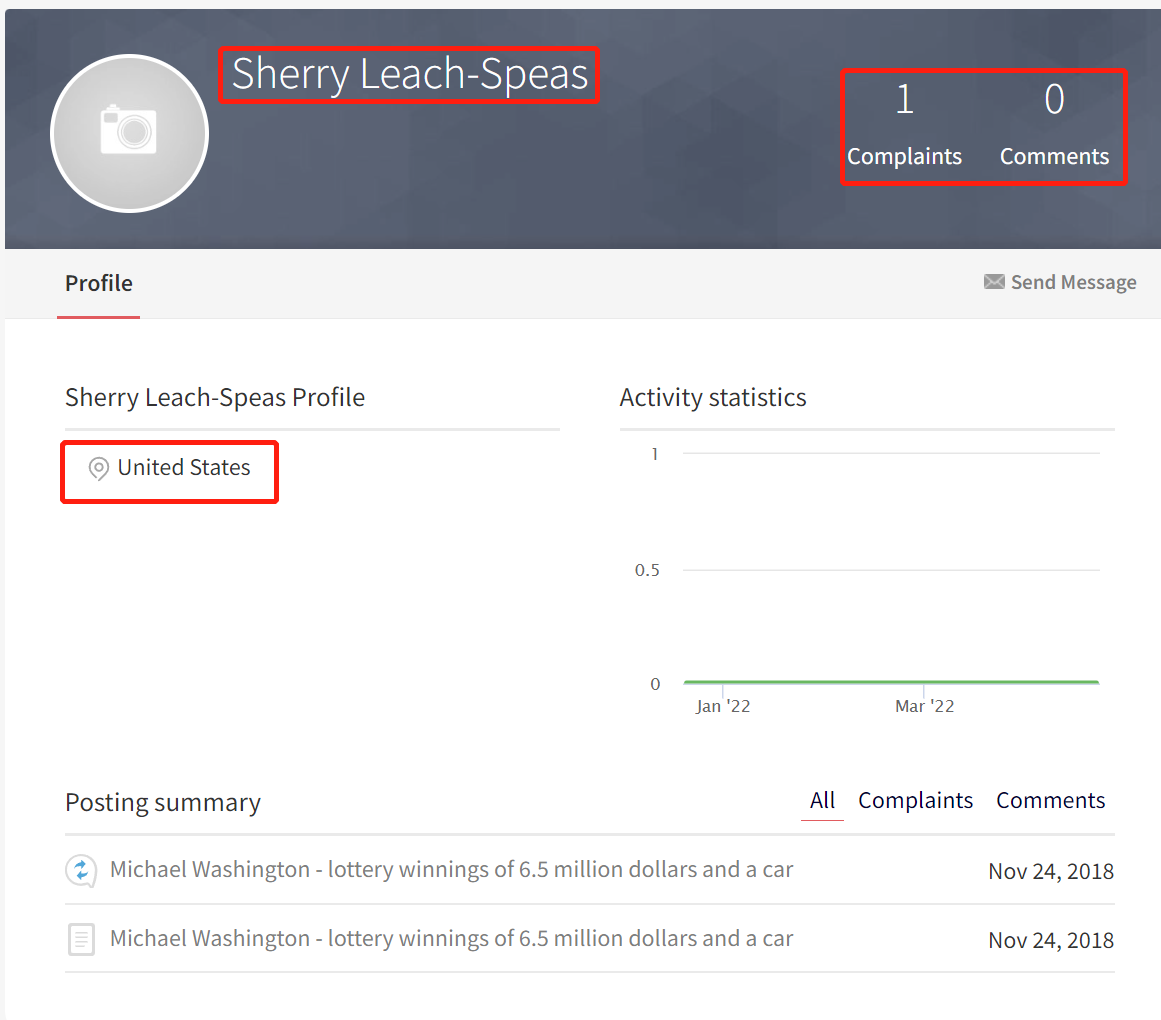
1. Collect the data about those Top companies (click into each company and collect data)
   1. Company name, rating star, total # of complaints, #solved, #unresolved, website URL



1. Collect the data about each specific complaint (click into each complaint and collect data)
   1. Company name, title, user name, country/city, date, detailed description, image if any



* 1. Profile info about the user by clicking into the user name link



1. Navigate to the next page, and continue the steps 1) and 2), till the last page.

