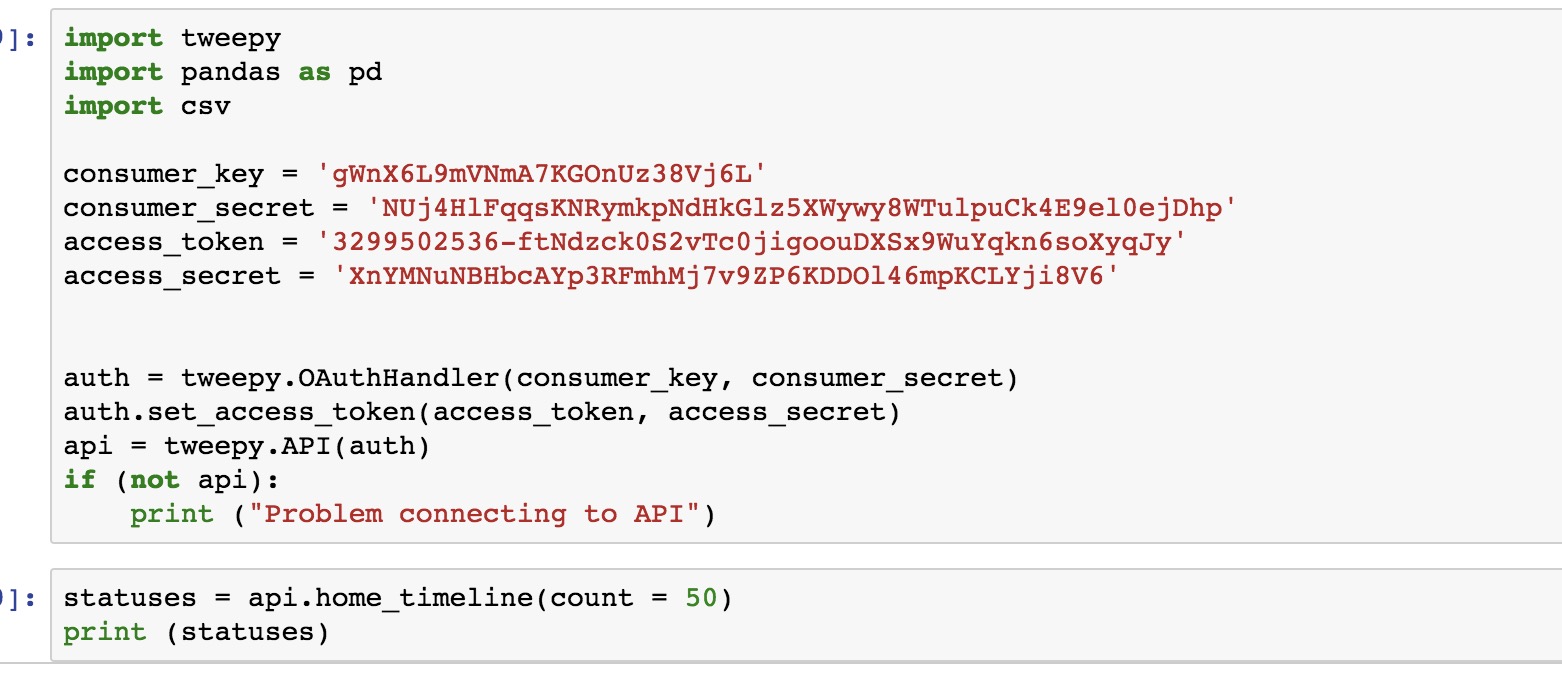
**Xinying Shi 001819931**

**Wendi Yu 001825852**

## README

In this assignment, we populate our database with tag data to the database on male tennis topic. This domain have entities that represent people (players, coaches), places(courts) and things(tournaments, match scores, tournaments and match statistics).

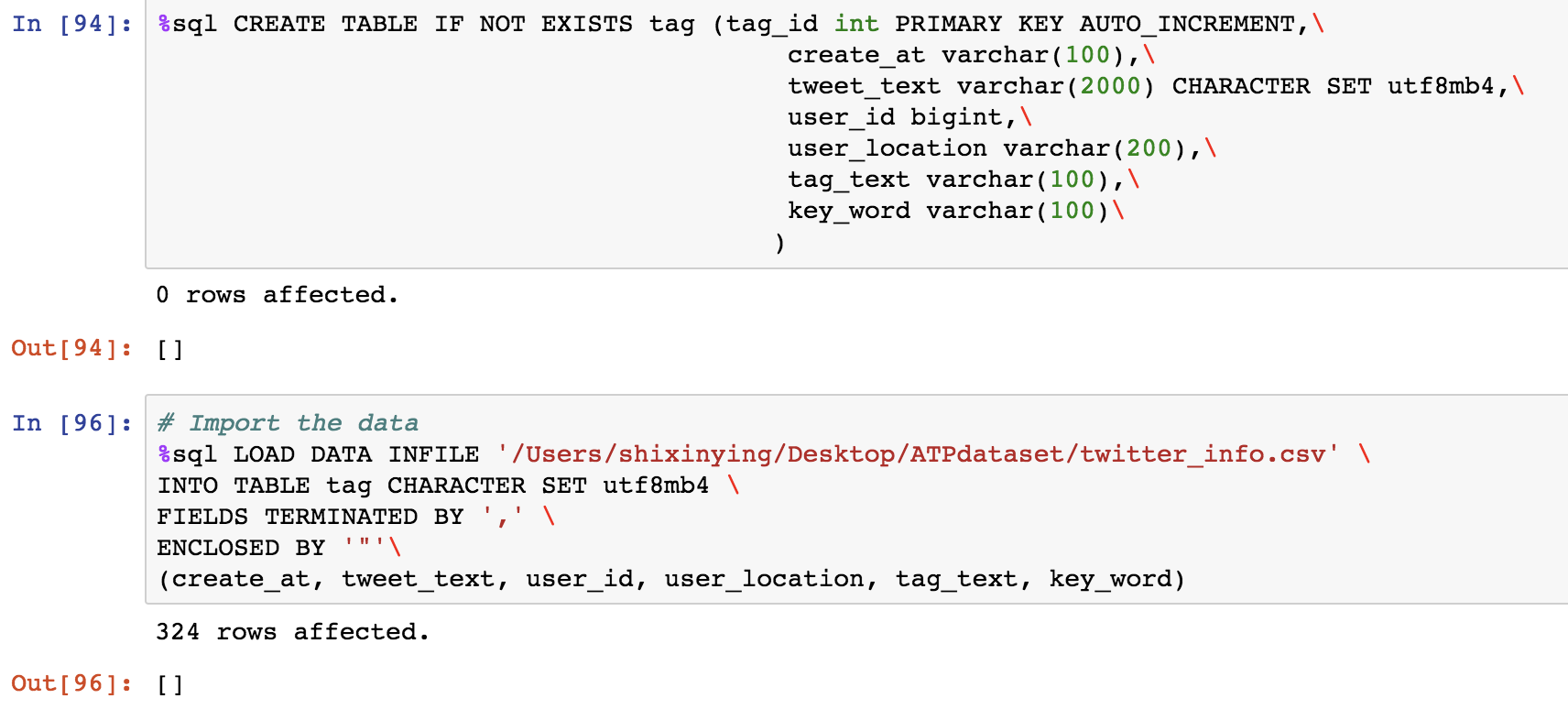
1. Firstly, we scrape some data from [http://www.atpworldtour.com](http://www.atpworldtour.com/) and some data from github.
2. Secondly, we cleansed ATP tennis dataset.
3. Then, we set up database and create tables.
4. The forth step is to scrape tags data from Twitter
5. Authorizing an application to access Twitter account data



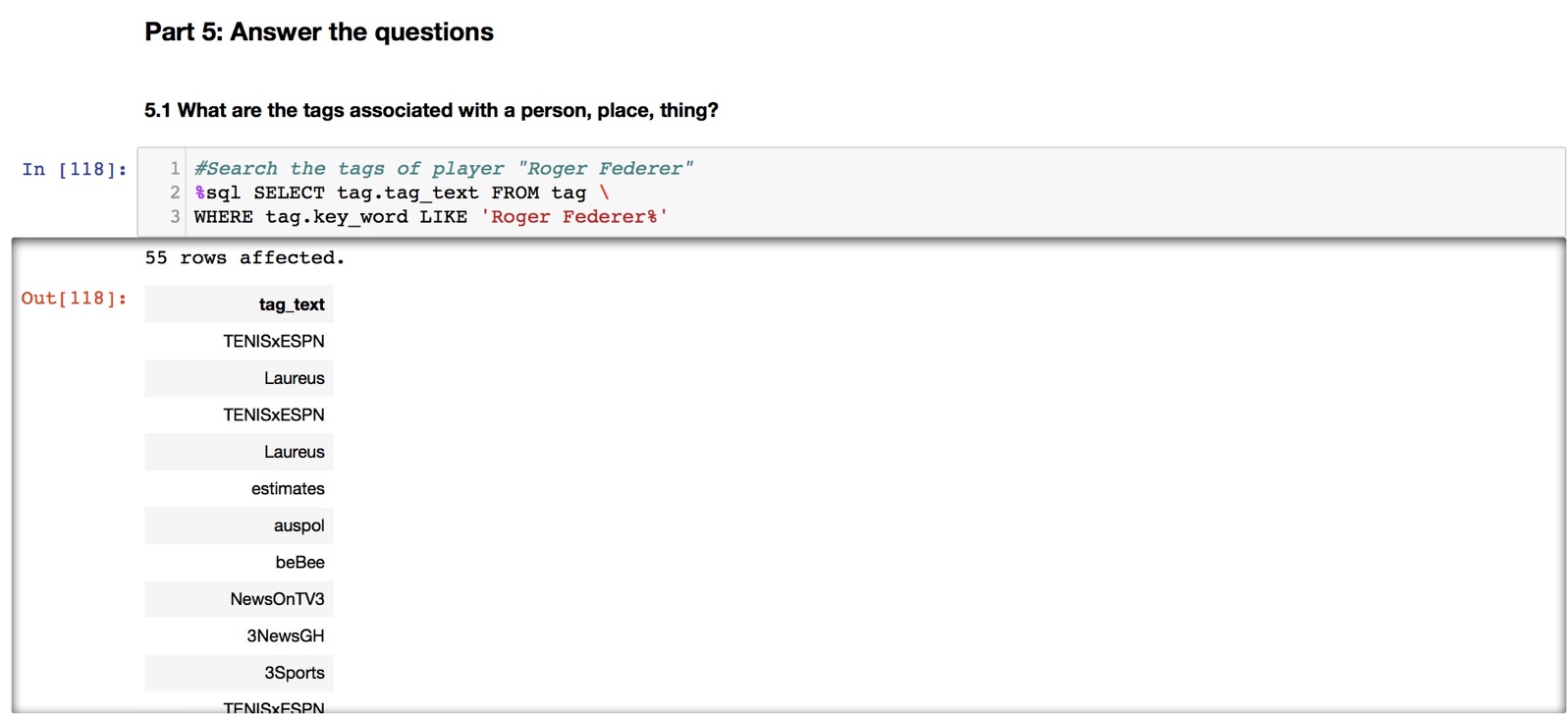
1. Extracting create time, text, user ids, user location, hashtags and key words from tweets



1. Create table 'tag' and import the data into it

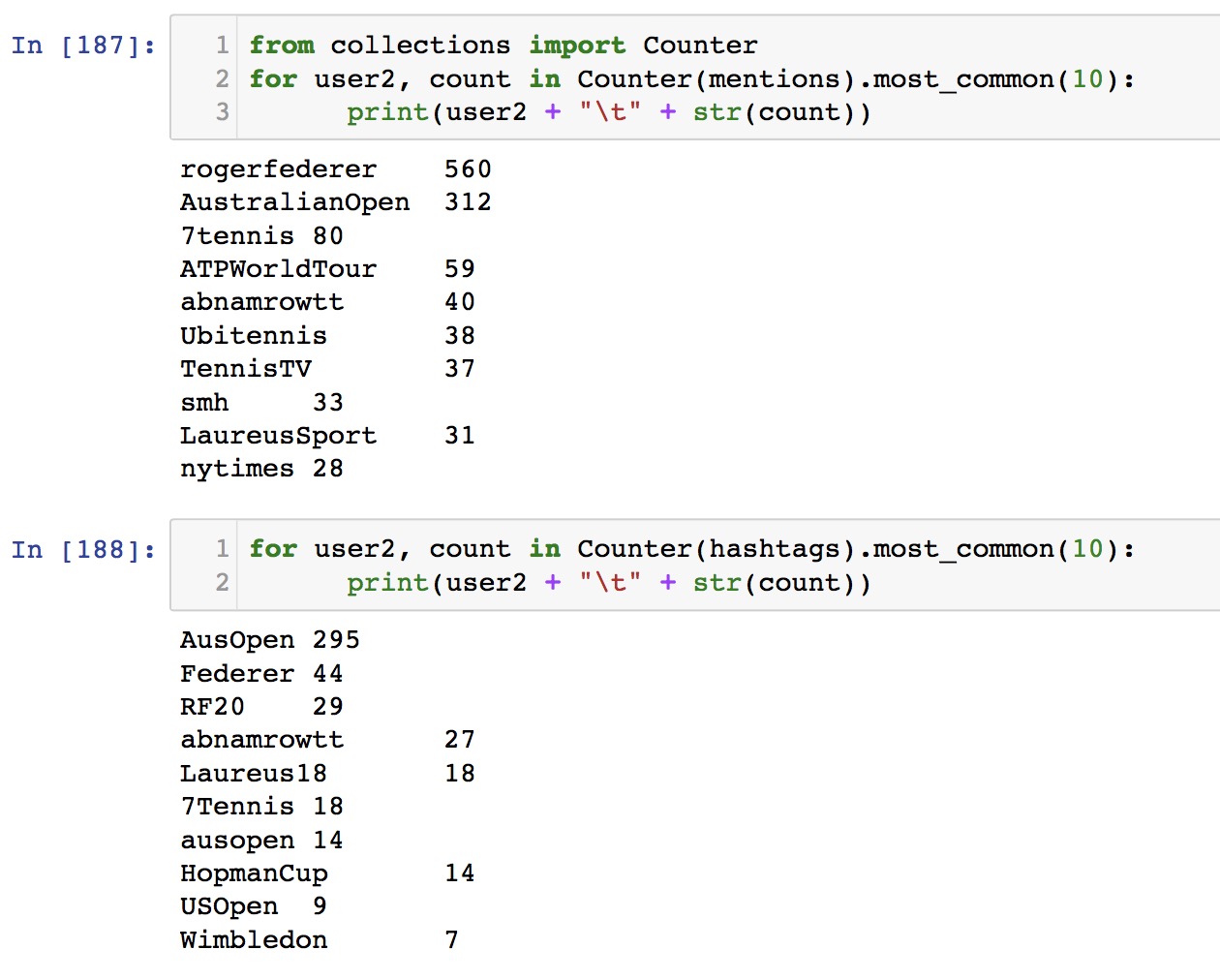
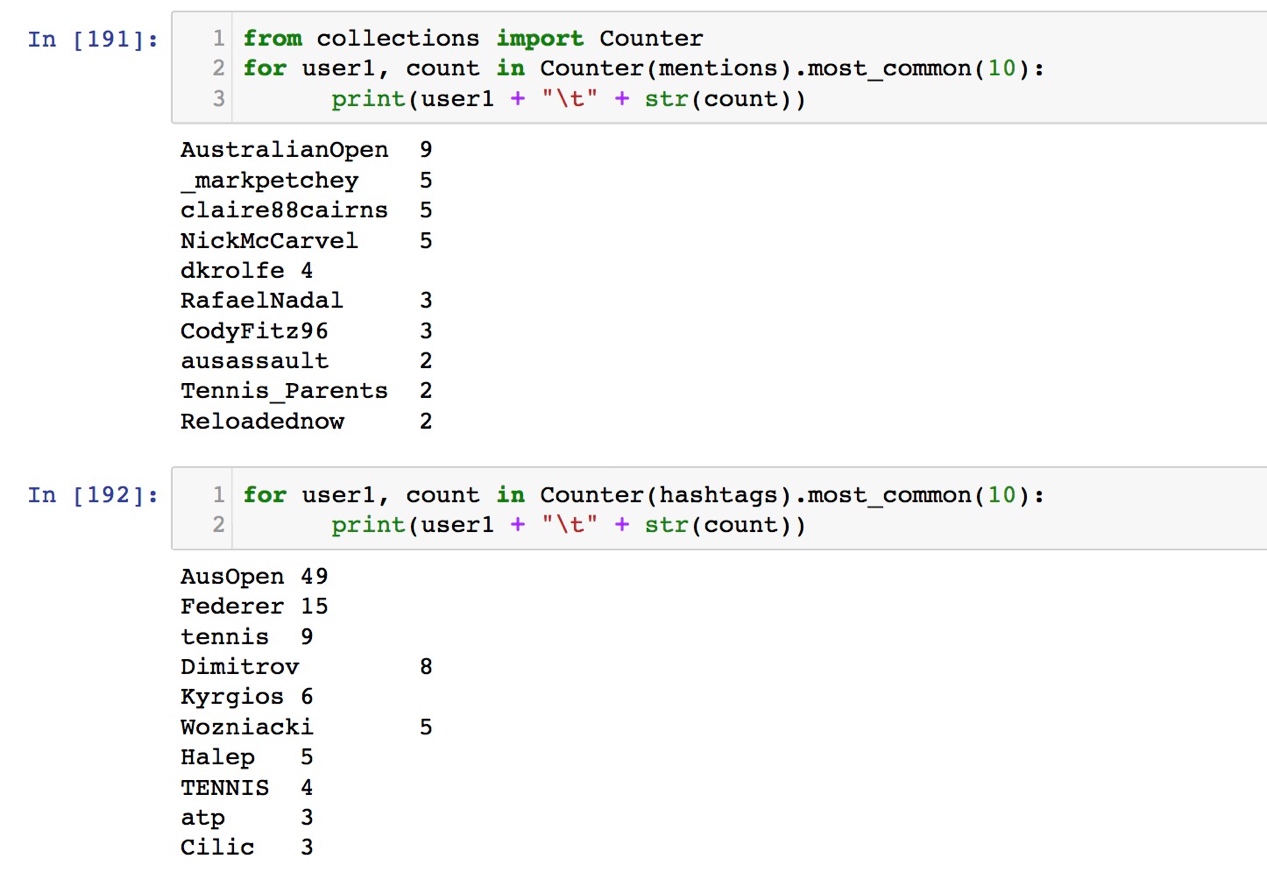


1. What are tags are associated with player:



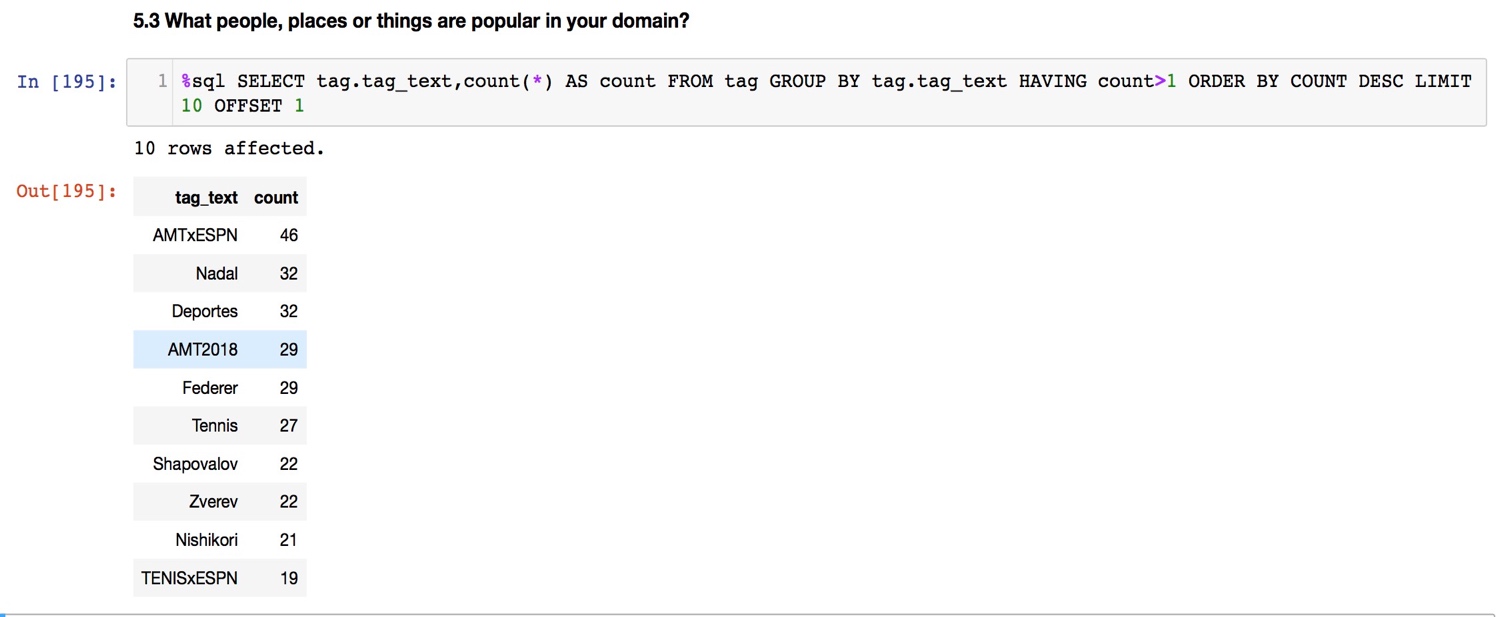
1. What social media users are like other social media users in your domain:

We search people who come from same country (Australia) and also posted some tweets about tennis in the past, then we compare tags in their tweets in past 100 days, we find the most common tags for two of them are both about Australian Open and Roger Federer, so they are similar user.



1. What people, places or things are popular in your domain:

We order the top 10 popular tags when users post tweets about some tennis players.



1. What topics about tennis players are trending in your domain? (A trend is popularity over a day.)

