Project 5.22. Web scraping of lianjia.com

Link: https://sz.lianjia.com/ershoufang/

Requirement: submit a code to realize the following function:

1. Find secondary housing trade information in Shenzhen
2. Go over the 100 pages of housing sale posting
3. Return a csv file with the following information:
   1. Name of housing
   2. Height of the building (number of floors)
   3. Year of construction
   4. Location of apartment in the city
   5. Number of rooms (living rooms and bedrooms)
   6. Total area of apartment
   7. Orientation of apartment
   8. Total price of apartment
   9. Unit price of apartment (price per square meter)
4. Bonus: can be applied to other cities (e.g. Beijing, Shanghai with minimal adjustments)

Codes will be graded based on:

1. The format of the codes
2. The organization of the csv file
3. The speed of the codes
4. The extendibility of the codes

Estimated completion time: 60 hours. Time consumption depends on the codes’ quality.

Suggestive points:

Deadline:

Bonus points:

1. Work completion one day before the deadline will generate an extra 20% bonus.
2. Slides to explain the codes will give 50 points/page

Penalty:

1. Late delivery/missing deadline will incur -10% of finally determined points of the whole project per missing day.
2. Serious mistakes that made the submitted results not useful will incur -120% of the whole project.