With the guidance of the tutorial from the given link, I've completed the following tasks:

- 1. Collect the needed data from that website with Scrapy package for Python
- 2. "Pipeline" the gathered data to a PostgreSQL database.
- 3. Export the database to an uncompressed SQL file (as attached), and it can be easily imported to anywhere else.

The entries I need are advertised products' title, description, link, category, location, original price and current price.

After setting up all the dependencies for Scrapy in my Windows 7 OS, I used command shell to create a new project named "my\_scraper" as follows:

## E:\>scrapy startproject my\_scraper

Subsequently, a hierarchy structure was created:

```
--my_scraper
--my_scraper
--spider
--_init_.py
--_init_.py, items.py, pipelines.py, settings.py
--scrapy.cfg
```

Next, I modified the codes as follows:

### Define the Item (items.py):

I added all the needed variable and set them to "Field()", which was imported from Scrapy.item

#### Define the spider (living\_social.py):

Here I created a file named "living\_social.py" in in the "spiders" directory and defined a spider called "livingsocialscraper". This file included spider's name, domain, start point and the parse function.

After these, I exported the data to a csv file to check the result:

```
E:\Scrapy\my_scraper>scrapy crawl livingsocial -o test.csv -t csv
```

#### **Export the results to a PostgreSQL database:**

I created a database in PostgreSQL and a database model in the file "model.py", and connected them in the "pipeline.py" file.

After setting up and configuring the scrapy and database, I executed the spider as follows:

# E:\Scrapy\my\_scraper>scrapy crawl livingsocialscraper\_

The code collected all needed entries and pipelined data to a postgreSQL database. Also, I used "pg\_dump" command to export the entire database to an uncompressed SQL file named "mydb.sql" (as attached).