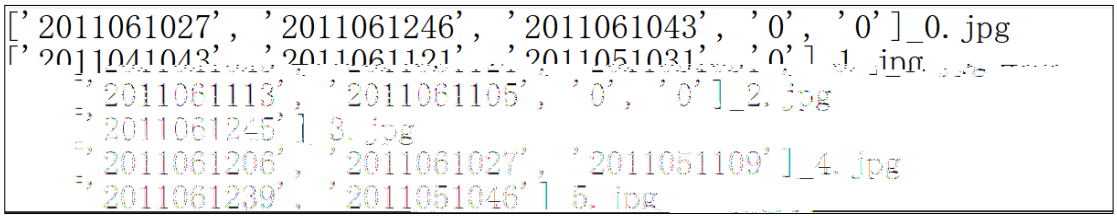
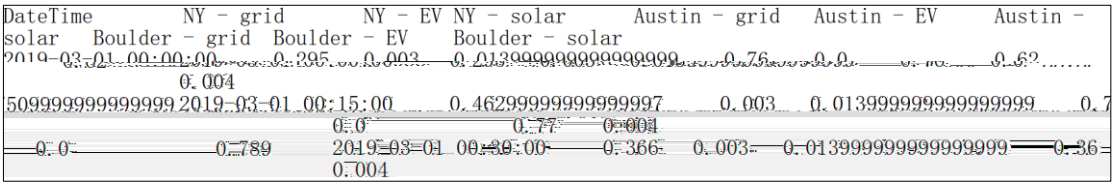
G


['2011061027', '2011061246', '2011061043', '0', '0']_0.jpg
'2011061113', '2011061105', '0', '0']_2.jpg
'2011061245']_3.jpg
'2011061206', '2011061027', '2011051109']_4.jpg
'2011061239', '2011051046']_5.jpg

3.6 Python

jupyter notebook pycharm

G



DateTime	NY - grid	NY - EV	NY - solar	Austin - grid	Austin - EV	Austin -
solar	Boulder - grid	Boulder - EV	Boulder - solar			
2019-03-01 00:00:00	0.295	0.003	0.013999999999999999	0.76	0.0	0.62
0.004						
5099999999999999	2019-03-01 00:15:00		0.46299999999999997	0.003	0.013999999999999999	0.7
		0.0	0.77	0.004		
0.0	0.789	2019-03-01 00:30:00	0.366	0.003	0.013999999999999999	0.36
		0.004				

word

\+ + +11111+1-

\data txt

3.6 Python

jupyter notebook pycharm

G

3.6 Python

jupyter notebook pycharm

2

			5	
iv	G			
v				

G

length	width
3	4
5	6
7	8
14	16
5	9
15	20
6	8

3.6 Python

jupyter notebook pycharm

2

			5	
iv				
i				

G

v

v

3.6 Python

jupyter notebook pycharm

2

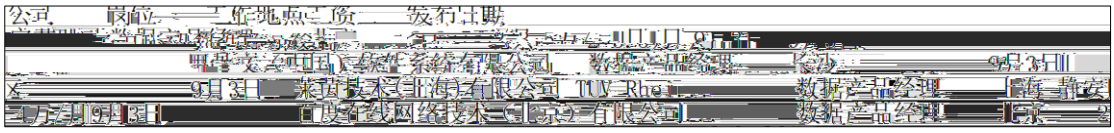
country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol	continent
Afghanistan	0	0	0	0	AS
Albania	89	132	54	4.9	EU
Algeria	25	0	14	0.7	AF
Andorra	245	138	312	12.4	EU

3.6 Python

jupyter notebook pycharm

2

G



3.6 Python

jupyter notebook pycharm

2

G

G

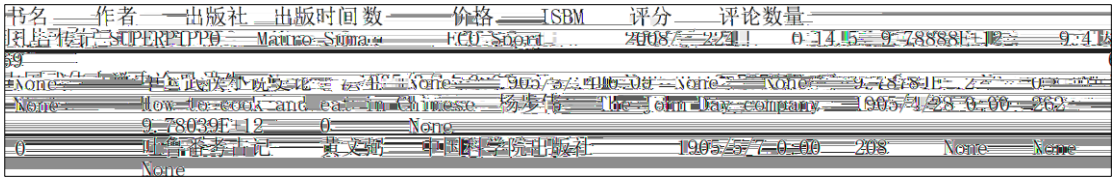
ListingId		借款金额	借款期限	借款利率	借款成功日期	初始评级
126541	18000	12	18	2015/5/4	C	
133291	9453	12	20	2015/3/16	D	
142421	27000	24	20	2016/4/26	E	
149711	25000	12	18	2015/3/30	C	
152141	20000	6	16	2015/1/22	C	

3.6 Python

jupyter notebook pycharm

2

G

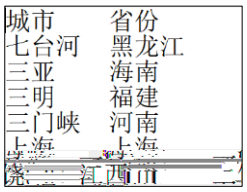


3.6 Python

jupyter notebook pycharm

2

G



城市	省份
七台河	黑龙江
三亚	海南
三明	福建
三门峡	河南
上海	上海
景德镇	江西省

3.6 Python

jupyter notebook pycharm

2

G G G

G

学校名称	录取批次
湘潭大学	本科一批
吉首大学	本科一批
湖南大学	本科一批
中南大学	本科一批

G

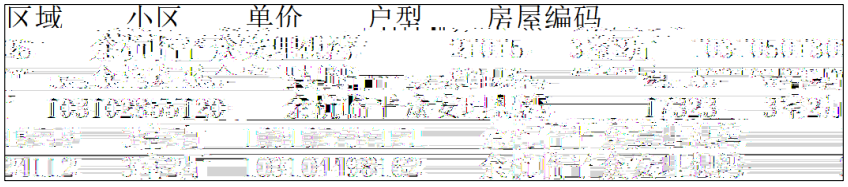
3.6 Python

jupyter notebook pycharm

2

G G G

G



G

3.6 Python

jupyter notebook pycharm

2

Linux wget -P /opt/software http://datasrc.tipdm.net:81/Hunan/module2/2-1/jdk-8u151-linux-x64.rpm JDK jdk-8u151-linux-x64.rpm Linux /opt/software

Linux wget -P /opt/software http://datasrc.tipdm.net:81/Hunan/module2/2-1/hadoop-2.6.4.tar.gz Hadoop hadoop-2.6.4.tar.gz Linux /opt/software

hadoop /usr/local

core-site.xml

/var/log/hadoop/tmp hadoop.tmp.dir value

hdfs://master:8020 fs.defaultFS value

hadoop-env.sh JAVA_HOME

cp mapred-site.xml.template mapred-site.xml

yarn mapreduce.framework.name value

yarn-site.xml ,

master yarn.resourcemanager.hostname value;

mapreduce_shuffle yarn.nodemanager.aux-services value

hdfs-site.xml

2 dfs.replication value

slaves slave1 slave2

iii. scp Hadoop

Hadoop

iv. /etc/profile JAVA_HOME HADOOP_HOME PATH source

v. NameNode

vi. Hadoop jps

vii.

i

ii

i.				
ii.	G		G	
iii.	G			
iv.	G			
v.				

vi. G

ii. G

iii.

iv.

v. [redacted]

i.
ii.
iii. G
iv.G

i.			G	
ii.				
iii.				
iv.				
v.				
vi.				

Linux initnetwork

Linux wget -P /course/Hive/data http://datasrc.tipdm.net:81/Hunan/module2/2-4/business_circle.csv business_circle.csv Linux /course/Hive/data hive Hadoop MySQL hive hive Linux $HADOOP_HOME/sbin/start-all.sh hadoop hadoop MySQL hive hive hive

business_circle.csv

Hive

i. G

ii. G

iii.

iv.

v.

iii.				
iv.				
v.				
vi.				

Linux initnetwork

Linux wget -P /course/Hive/data http://datasrc.tipdm.net:81/Hunan/module2/2-5/business_circle.csv business_circle.csv Linux /course/Hive/data hive Hadoop MySQL hive hive Linux $HADOOP_HOME/sbin/start-all.sh hadoop hadoop MySQL hive hive hive

business_circle.csv

Hive

i. G

ii. G

iii.

iv.

v.

i.
ii.
iii. G
iv.G

i.			G	
ii.				
iii.				
iv.				
v.				
vi.				

http://datasrc.tipdm.net:81/Hunan/module2/2-7/student.csv student.csv
Linux /course/Hive/data
hive Hadoop MySQL
hive hive Linux
$HADOOP_HOME/sbin/start-all.sh hadoop
hadoop MySQL hive hive
hive

			G

Hive

i.
G
ii. G G

iii.

iv.
v.

i.
ii.
iii. G
iv.G

i.			G	
ii.				

iii.				
iv.				
v.				
vi.				

Linux initnetwork

Linux wget -P /course/Hive/data http://datasrc.tipdm.net:81/Hunan/module2/2-8/emp.csv emp.csv Linux /course/Hive/data

hive Hadoop MySQL hive hive Linux $HADOOP_HOME/sbin/start-all.sh hadoop hadoop MySQL hive hive hive

	G			

Hive

i.

G

ii. G G

G

iii.

iv.

v.

vi.

(i)

(ii)

(iii) G

(iv)G

i.			G	
ii.				
iii.				
iv.				
v.				

vi.				
vii.				

Linux initnetwork

Linux wget -P /course/Hive/data http://datasrc.tipdm.net:81/Hunan/module2/2-9/emp.csv emp.csv Linux /course/Hive/data

hive Hadoop MySQL hive hive Linux $HADOOP_HOME/sbin/start-all.sh hadoop hadoop MySQL hive hive hive

	G			

Hive

i. G

ii. G

iii.

iv.

v.

vi.

i.
ii.
iii. G
iv.G

i.			G	
ii.				
iii.				
iv.				
v.				
vi.				
vii.				

i.				
ii.				
iii.				
iv.				
v.				
vi.				

Linux initnetwork

Linux wget -P /course/Hive/data
http://datasrc.tipdm.net:81/Hunan/module2/2-11/business_circle.csv
business_circle.csv Linux /course/Hive/data
hive Hadoop MySQL
hive hive Linux
$HADOOP_HOME/sbin/start-all.sh hadoop
hadoop MySQL hive hive
hive

business_circle.csv

Hive

i. G

ii. G

iii.

iv.

v.

i.
ii.
iii. G
iv.G

i.			G	
ii.				
iii.				

iii.

iv.

v.

i.
ii.
iii. G
iv.G

i.			G	
ii.				
iii.				
iv.				
v.				
vi.				

G

G

G

G G

etree

HTML

	G		G	

G

BeautifulSoup

BeautifulSoup BeautifulSoup

	urllib3 HTTP BeautifulSoup		urllib3 HTTP 10 200 10	

iii	json			

;

G

G

urls

4 urls

http://tipdm.com/qyyy/index_1.jhtml

requests BeautifulSoup

;

urls

requests

BeautifulSoup BeautifulSoup

BeautifulSoup

urls

4 urls

http://tipdm.com/qyyy/index_1.jhtml

urllib3 etree

;

urls

urllib3

etree.HTML get html

html

urls

4 urls

http://tipdm.com/qyyy/index_1.jhtml

requests re

;

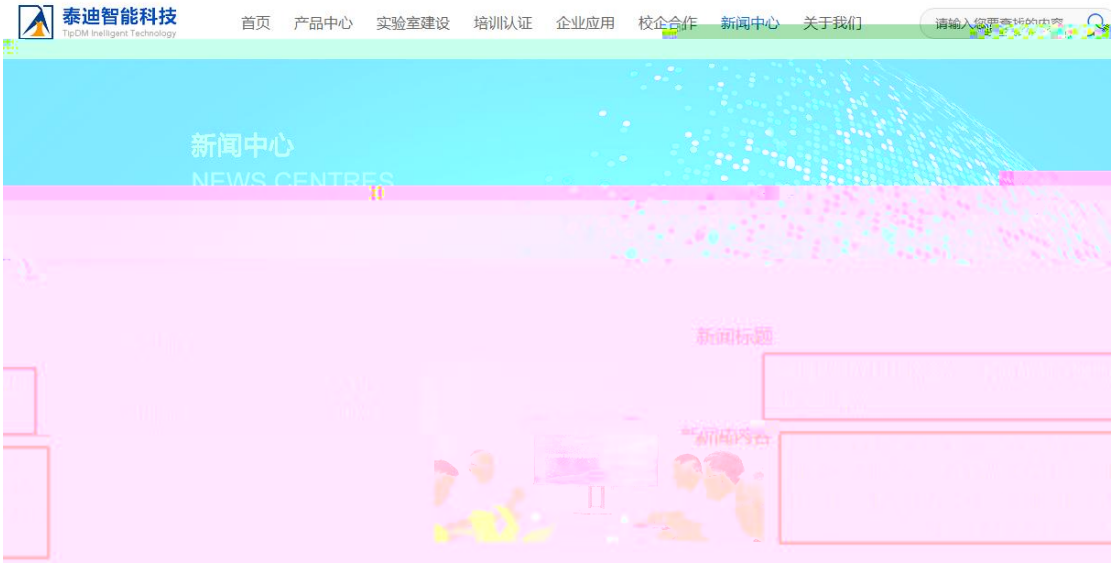
urls

requests

'<h1>(

)</h1>'

' ( ) '



□ □

□ □

urls

4 urls

http://tipdm.com/xwzx/index_1.jhtml

requests BeautifulSoup

;

urls

requests

BeautifulSoup BeautifulSoup

BeautifulSoup

data

□ □

□ □

urls

4 urls

http://tipdm.com/xwzx/index_1.jhtml

urllib3 etree

;

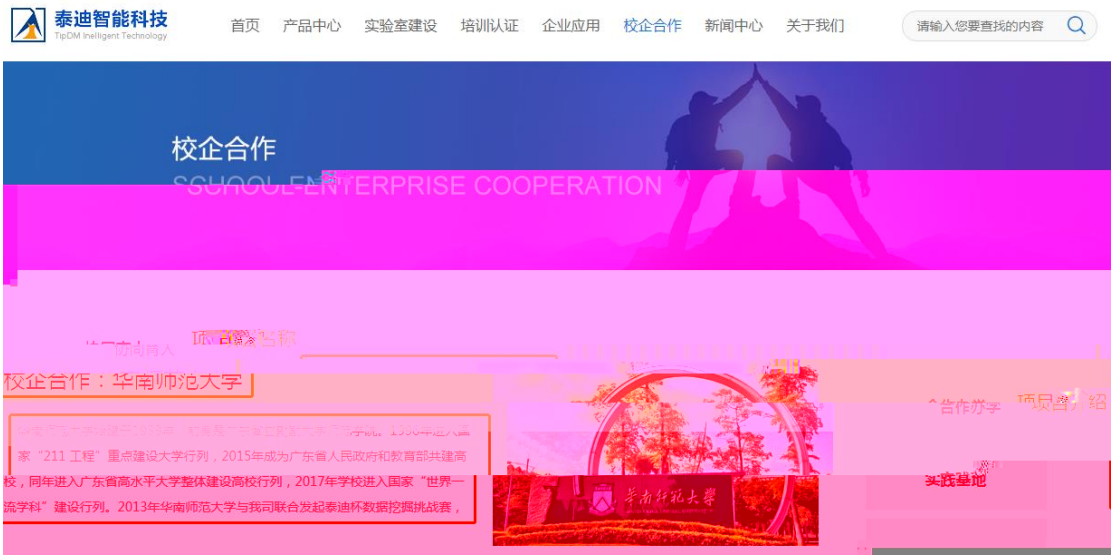
urls

urllib3

etree.HTML html

html

data



urls

4 urls

http://tipdm.com/xqhz/index_1.jhtml

requests BeautifulSoup

;

urls

url

BeautifulSoup BeautifulSoup

BeautifulSoup

data

urls

4 urls

http://tipdm.com/xqhz/index_1.jhtml

requests re

;

urls

url requests

'target="_blank">( )</h1'

' ( ) '

x

x

y

5

x
x
y

5

5

G

□ □ □

□

G

G

□

2017 6 5

x

x

y

zip

2017 6 5

G

□ □ □

G

G

(ii) “ ” “ ”

' '

(iii) “ ” “ ” “ ”

x

x

y

G

' '

(iii)

;

x
x
y

G G

G

表 4-2 HR_comma_sep.csv 字段说明

' '

x
x
y
' '

Linux wget -P /root/jupyter_notebook http://datasrc.tipdm.net:81/Hunan/module5/5-7/movies.csv movies.csv Linux /root/jupyter_notebook

□

□ □

表 4-3 movies.csv 字段说明

Linux wget -P /root/jupyter_notebook
http://datasrc.tipdm.net:81/Hunan/module5/5-8/mum_baby_trade_history.csv
mum_baby_trade_history.csv Linux /root/jupyter_notebook

□ □

□

□

□ □

□ □

表 4-4 mum_baby_trade_history.csv 字段说明

;

;

x
x
y

Linux wget -P /root/jupyter_notebook http://datasrc.tipdm.net:81/Hunan/module5/5-9/book.xlsx book.xlsx Linux /root/jupyter_notebook

□□

□ □□

(i) pandas matplotlib.pyplot plt.rcParams['font.sans-serif'] = 'WEnQuanYi Zen Hei' plt.rcParams['font.sans-serif'] = 'SimHei'

□

;

x

x

y

http://datasrc.tipdm.net:81/Hunan/module5/5-10/travel.xlsx travel.xlsx Linux /root/jupyter_notebook

□□□

(i) pandas matplotlib.pyplot plt.rcParams['font.sans-serif'] = 'WEnQuanYi Zen Hei' plt.rcParams['font.sans-serif'] = 'SimHei'

□

;

x

x

y

10

x

x

y

Linux wget -P /root/jupyter_notebook
http://datasrc.tipdm.net:81/Hunan/module5/5-12/lipsticks.xlsx
lipsticks.xlsx Linux /root/jupyter_notebook

x
x
y

Linux wget -P /root/jupyter_notebook http://datasrc.tipdm.net:81/Hunan/module5/5-13/guomin.xlsx guomin.xlsx Linux /root/jupyter_notebook

□□□

□

□□□

2000 data.iloc[0,5:]

2000

□

data.iloc[68, 2:5]

x

x

y

□

Linux wget -P /root/jupyter_notebook http://datasrc.tipdm.net:81/Hunan/module5/5-14/car_selling_fact.xlsx car_selling_fact.xlsx Linux /root/jupyter_notebook

x

y