) £

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

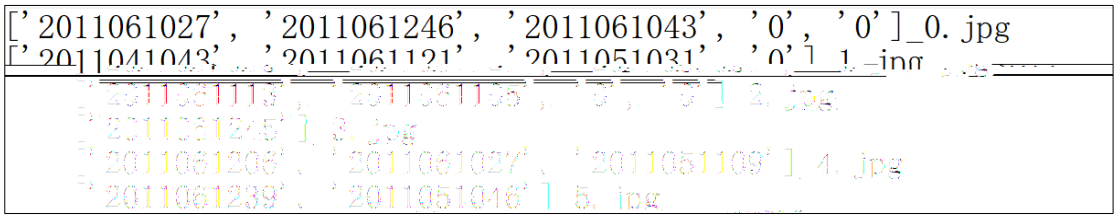
.....

.....

.....

.....

.....


['2011061027', '2011061246', '2011061043', '0', '0']_0.jpg
['2011061206', '2011061027', '2011051109']_4.jpg
['2011061239', '2011051016']_5.jpg

a c

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	0.789	0.004	0.0
2019-03-01 00:15:00	0.46299999999999997	0.003	0.013999999999999999	0.77	0.004	0.36	0.75099999999999999	0.004	0.0
2019-03-01 00:30:00	0.366	0.003	0.013999999999999999	0.36	0.0	0.0	0.789	0.004	0.0

a c

a c

			+	
–				

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

a c

			+	
—				
—				

a c

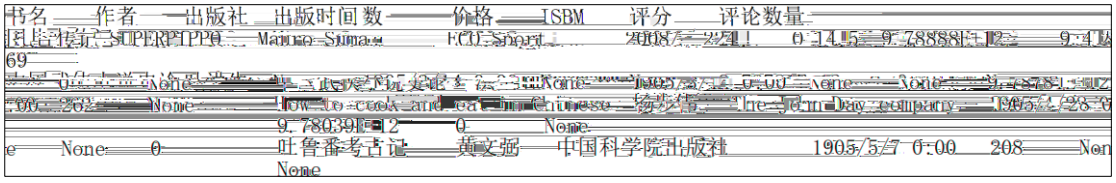
country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol	continent
Afghanistan	0	0	0	0	AS
Albania	89	132	54	4.9	EU
Algeria	25	0	14	0.7	AF
Andorra	245	138	312	12.4	EU

a c

a c

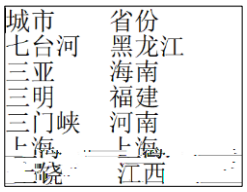
ListingId		借款金额	借款期限	借款利率	借款成功日期	初始评级
126541	18000	12	18	2015/5/4	C	
133291	9453	12	20	2015/3/16	D	
142421	27000	24	20	2016/4/26	E	
149711	25000	12	18	2015/3/30	C	
152141	20000	6	16	2015/1/22	C	

a c



书名	作者	出版社	出版时间	数	价格	ISBM	评分	评论数量
图片详情	SUPERPIPPO	Marino Sumaw	FCU Shard		4008/2/2	0	14.5	9
69								
	0	None						
0.00	302	None	How to cook and eat in Chinese		徐珍珍	The 36th Day Company	1905/2/28	0
			9	780391	2	0	None	
e	None	0	吐鲁番考古记		黄文弼	中国科学院出版社	1905/5/7 0:00	208
None								

a c

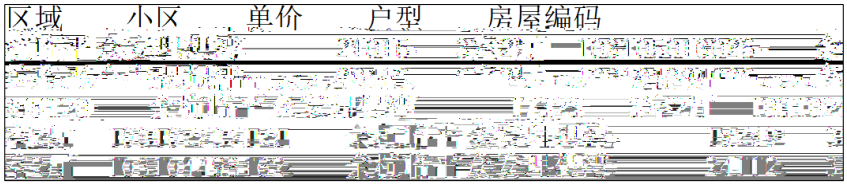


城市	省份
七台河	黑龙江
三亚	海南
三明	福建
三门峡	河南
上海	上海
上饶	江西

a c

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

a c



a c

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

Linux initnetwork

Linux wget -P /course/Hive/data
http://datasrc.tipdm.net:81/Hunan/module2/2-3/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

Hive

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

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

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

Hive

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

Hive

iii.

iv.

v.

vi.

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

Hive

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

导航条
首页
产品中心
实验室建设
培训认证
企业应用
校企合作
新闻中心
关于我们
大数据产品
人工智能产品
在线教育
教学资源
开源平台
其他
saas

etree

HTML

BeautifulSoup

BeautifulSoup BeautifulSoup

	bb_		bb_	

8 _\ b

——				

导航条
首页
产品中心
实验室建设
培训认证
企业应用
校企合作
新闻中心
关于我们
大数据产品
人工智能产品
在线教育
教学资源
开源平台
其他
saas

;

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

2017 6 5

x
x
y
zip
2017 6 5

(ii) “ ” “ ”

' '

(iii) “ ” “ ” “ ”

x
x
y

' '

(iii)

;

x
x
y

表 4-2 HR_comma_sep.csv 字段说明

' '

x
x
y
' '

_ a
_ c c b + + − c _ c _
_ a

表 4-3 movies.csv 字段说明

10

x
x
y
for

10

10

;

_ a

_ c c b + + c c _

c c _ _ a

表 4-4 mum_baby_trade_history.csv 字段说明

;

;

x
x
y

_ a
_ c c b + + a b a b _
a

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

;

x
x
y

_ c c b + + b b b b

_ a

(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

_ a

_ c c b + + b_ _ a b

b_ _ a b _ a

x

x

y

_ a

_ c c b + + c _ b c _ b

_ a

2000 data.iloc[0,5:]

2000

data.iloc[68, 2:5]

x
x
y

_ a

_ c c b + + bb_ \ b

bb_ \ b _ a

x

y