.....

—€–îìííìòíìîó–U –îìííìòíîðò–U –îìííìòíìðï–U –ì–U –ì–•z

“ z” —€–îìííìòíìîó–U –îìííìòíîðò–U –îì‡ìXíï‰ÔP—U –ì–U –ì–•

ì

š€ îÍW “ U”

“ z” l Ç À oµV

l ì ñ l Ç À î

ï “ [Ư [À oµ

] À “ –ï”–

“ ” “ ”

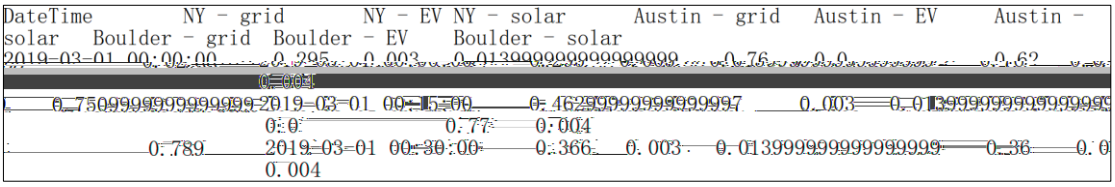
]

]]

î

>]vµÆ ^ÁP mšW lŒ}}šliµ‰Çš Œzv}š }}l
Zšš‰Wll š •Œ Xš]‰ uXv šW Œ]µ •}ou Œ s_ílíš XšÆš
'Œ] z•}o Œz sz >]vXµšZEl Œ}}šliµ‰Çš Œzv}š }}l

“ 'Œ] z•}o Œz s”z š XšÆš Ez µ•š]v }µo Œ



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	0.75	0.0	0.46
2019-03-01 00:15:00	0.77	0.004	0.013999999999999999	0.789	0.003	0.36	0.366	0.0	0.004

pš

Ez µ•š]v }µo Œ

šÆš

š d]u Ez “ írEzX’šÆµš•š]v
“ ír µ•š]v”XšÆµšo Œ “ ír }µo ŒXšÆš
]
^'Œ] z•}o Œz s_z š XšÆšv }]v[ÞÁ[V

]]

]]]

]À

v }]v[ÞÁ[U

“ ” “ 1”

]

]]

î

]		íì	íì	
]]		íñ	ñ ' pš[ñ ñ	
]]]		ïì	ñ íñ ñ íñ	
]À		ïì	šÆš íì ïì	
s		íñ	ï ï ï ï ï	

>]vµÆ ^ÁP mšW lŒ}}šliµ‰Çš Œzv}š }}l
Zšš‰Wll š •Œ Xš]‰ uXv šWlôíµlŒEµ}vìíví1_xšÆEöµíŒEí}îìíîXš>ÆEvšµÆ
lŒ}}šliµ‰Çš Œzv}š }}l

“ µŒ}îìíî”XšÆšîìíî ïñ
-pš•

~]•
^ µŒ}îìíî_XšÆš v }]v[Pxšví(ò[V

~]]•
V

~]]]•

~] À • '

'

[' [

[' [

“ ” “ ”

]

]]

]		íì	íì	
]]		íñ	ñ þš ñ	
]]]		îì	o v íì íì	
		ðì	ñ ' íì (}Œ íì íì ñ	
		íñ	ï ï ï ï ï	

>]vµÆ ^ÁP šW lŒ}}šliµ‰Çš Œzv}š }}l
Zšš‰Wll š •Œ Xš]‰ uXv šWÕÕŒl,µšvvPlu}XþšØŒšlš vPo X]švÆŒÆ
lŒ}}šliµ‰Çš Œzv}š }}l

Œ š vPo XšÆš

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

Z š vPo

P š]Œ µu(Œ v P š Œ

Œ š vPo XšÆš Œ]Œ µu(Œ v

Œ š vPo îXšÆš

~]• ^Z š vPo

Z š vPo V

o vPšZUÁ] šZ

P š Œ

P š]Œ µu(Œ v

~]]•

^Œ š vP_o XšÆš v }]v[µ[V

(

~]]]•

~]À•

^ _

Z š vPo Œ

Œ P š Œ ~• P š]Œ µu(Œ v ~•

^ _

~À•

“ Œ š vP_o îXšÆš

^Œ š vP_o îXšÆš v }]v[µ[V

“ ” “ ”

]

]]

]		ï ì	ï ì í ì í ì í ì	
]]		í ì	í ì	
]]]		í ñ	ñ pš ñ	
		î ì	í ì (} Œ ñ ñ	
À		í ì	ñ ñ	
À		í ñ	ï ï ï ï ï	

>] v µ Æ ^Á P nšW l Œ } } š l i µ ‰ Ç š Œ z v } š } } l
Z š š ‰ W l l š • Œ X š] ‰ u X v š W ñ ô Z l] ‰ v } š v o l u X } š µ Æ Œ Z š] í ‰ d } š o X š v Æ µ š Æ E
l Œ } } š l i µ ‰ } Ç š š } Œ E l z

Z] ‰ } š o X š Æ š

~] •

^ Z] ‰ } X š š o Æ š v }] v [Ø ð Ø [

~]] •

~]]] •

~] •

~•

V

“ ” “ ”

]
]]

]]]		ï ì	ñ í ñ ñ í ñ	
] À		ï ì	š Æ š í ì ï ì	
s		í ñ	ï ï ï ï ï	

>] v µ Æ ^Á P ŠW l Œ } } š l i µ ‰ Ç š Œ z v } š } } l
Z š š ‰ W l l š • Œ X š] ‰ u X v š W ó ô } l z µ] v (y _ X u š } Æ µ š ò ý] l í v (} X š > Æ v š µ Æ
l Œ } } š l i µ ‰ Ç š Œ z v } š } } l

i } z] v (} X š Æ š
p š

, – – W ð ì ì U – – W í ì ì ì U – – W ð ì ì ì U – – W í ì ì ƒ
~] •
^i } z] v (} _ X š Æ š v }] v [p ø Ý (ò [V

~]] •

~]]] •

~] À •
]]]
~ À •

~ À] •

“ ” W W

‘ ’

W W

“ ” “ ”

]

]]

]		ï ì	](l o]•(í ì	
]]		î ì	í ì í ì	
]]]		î ñ	ñ ñ í ì ñ	
]À		í ì	ñ Á Œ] š ñ	
s		í ñ	ï ï ï ï ï ï	

>] v µ Æ ^Á P ‰W l Œ } } š l i µ ‰ Ç š Œ z v } š } } l
Z š š ‰ W l l š • Œ X š] ‰ u X v š W ð î í l X š v Æ š l } X š o Æ í] v µ Æ
l Œ } } š l i µ ‰ Ç š Œ z v } š } } l

“ > X š Æ š

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

]

^> X š Æ š v }] v [í (ò [V

]]

]]]

U

] À

uzîìíñ

uzîìíò

îìíñ îìíò uzîìíñ

uzîìíò

À

“ ” “ ”

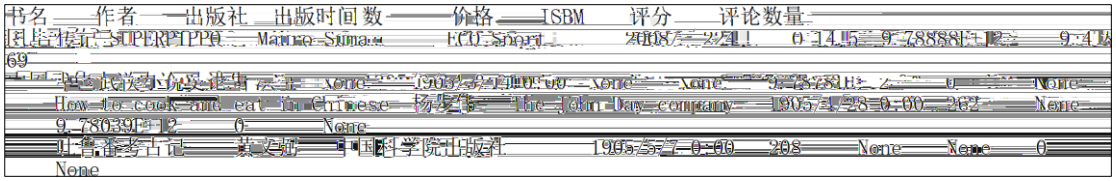
]

]]

]		íì	íì	
]]		íñ	ñ pš ñ ñ	
]]]		îñ	ñ íì ñ ñ	
] À		îñ	](íñ (}Œ ñ ñ	
À		íì	ñ ñ	
s]		íñ	ï ï ï ï ï	

>] v µ Æ ^Á P ŠW l Œ } } š l i µ ‰ Ç š Œ z v } š } } l
Z š š ‰ W l l š • Œ X š] ‰ u X v š W ˜ ðlí } } µ X š_ v Æ Eš } µ } } l X ŠíÆ > š] v µ Æ
l Œ } } š l i µ ‰ Ç š Œ z v } š } } l

“ } } l X š Æ š



pš

]
^ } } l X š_ Æ š v }] v [p ‚ Ø Ý (ò [V

]]

Œ } Á •

]]]

] À

Œ Á • ó

‰ Œ] v š

“ ” “ ”

]
]]

]		í ì	í ì	
]]		í ñ	ñ pš ñ ñ	
]]]		ï ì	í ì (} Œ í ì í ì	
] À		ï ì	(} Œ ñ u Æ• ~ ñ] (í ì í ì	
s		í ñ	ï ï ï ï ï	

>] v µ Æ ^Á P ‰W l Œ } } š l i µ ‰ Ç š Œ z v } š } } l
Z š š ‰ W l l š • Œ X š] ‰ u X v š W ô l í] šµÇ X š Æ š] ‰ Ç X š Æ š
l Œ } } š l i µ ‰ Ç š Œ z v } š } } l

“] š Ç X š Æ š

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

]
v µ u ‰ Ç
] š Ç X š Æ š o } š Æ š v }] v [P
]]

]]]

] À

l Ç À o µ

' W

“ ñ ô ò ï ì ð X î

]]] ' W

' W

' W

“ î ì í ì í ó ó õ ñ î ò X ð õ ð ” õ õ õ õ õ ï

“ ” “ ”

]

]]

]		î ì	v µ ‰ Ç ñ v µ u ‰ Ç X o → š Æ š í ñ	
]]	' W	ï ñ	(} Œ í ì ' W î ì ‰ Œ] v š ~ • ñ	
]]]	' W	ï ì	(} Œ í ì ' W í ñ ‰ Œ] v š ~ • ñ	
] À		í ñ	ï ï ï ï ï	

>] v µ Æ ^ Á P ‰ W l Œ } } š l i µ ‰ Ç š Œ z v } š } } l
Z š š ‰ W l l š • Œ X š] ‰ u X v š W ô l í , l µ µ v š Æ u š µ µ v í l š Æ š] v µ Æ
l Œ } } š l i µ ‰ Ç š Œ z v } š } } l

“ , µ v š ” Æ š

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

]

v µ u ‰ Ç

, µ v v X š Æ š o } š Æ š v }] v [µ ‰ Ç] [P

]]

]]]

] À

l Ç À o µ

“ ” “ ”

]

]]

]		î ì	v µ u ‰ Ç ñ v µ u ‰ Ç X o } š Æ š í ñ	
]]		î ì	ñ (} Œ ñ í ì	
]]]		î ì	î ì	
] À		î ñ	í ì } µ v š • ñ í ì	
s		í ñ	ï ï ï ï	

			ï	

>]vµÆ ^ÁP šW lŒ}}šliµ‰ÇšŒzv}š }}l
Zšš‰Wll š •Œ Xš]‰ uXv šW ðÍ], µP ìZ Æ o,íP ìZ Æ]vµÆ
lŒ}}šliµ‰ÇšŒzv}š }}l

“ , vPìZ ”Æš

区域	小区	单价	户型	房屋编码

]
vµu‰Ç
, vP•Z}µXšÆš o} šÆš v }]v]P
]]

]]]

]À
l Ç

“ ” “ ”

]
]]

]		îì	vµu‰Ç ñ vµu‰ÇXo} šÆš	

			í ñ	
]]		î ì	ñ (} Œ ñ í ì	
]]]		ï ì	í ì ñ (} Œ ñ í ì	
] À		í ñ	} Œ ñ í ì	
s		í ñ	ï ï ï ï ï	

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

k

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

	: l	í ì	: < ï ñ i l î	
	, } } ‰	ð ì	, } } ‰ ñ ñ ð ì	
	, } } ‰	í ì	 ò î	
	, } } ‰	í ì	î ô î	
	E u E }	ñ	ñ ñ	

, }}‰ ñ ð

i‰• ŏ ƨLJ

i‰•

			ï	

Linux “ initnetwork” “ ”

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

µ•]v ••z]Œ o X •À

/		/vš	ïòõì î ïòõì ï ïòõì Ÿð
d]u zÁ}Œl		/vš	óôUíððUõìñYUòõ
d]u zu] v]P		/vš	ñîíUòì ìUðñóY
d]u zÁ l v		/vš	òì îUñîíUðòôY
d]u z Ç		/vš	îôòïUîîðñUíìñð

Hive

,]À “ ”
µ•]v ••
>]vµÆ l }µŒ• l,]À l šµl•]v ••z]Œ o X •Àµ•]v ••
íì

îìì ì

u ÆÁ}Œlu]vÁ}Œl

îìì ì

“ ” “ ”

v ¥ ^ ò ï Ì >] v µ Æ
í Ø ò : <
, } } ‰ î X ò Ø ð
,] À

□

		í ñ	,] À í ñ í	
		í ñ	í ñ í ì ñ ï	
		í ñ	 í ì rñ	
		î ì	 í ì rñ ô î	
		î ì	} µ v Á Z Œ ô ð	
		í ñ	ï ï ï ï ï ï	

Linux “ initnetwork ” “ ”

Linux “ wget -P /course/Hive/data http://datasrc.tipdm.net:81/Hunan/module2/2.2business_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

/		/v š	ï ò õ ì î ï ò õ ì ï ï ò õ ì Yð
d] u z Á } Œ l		/ v š	ó ô U í ð ð U õ ì ñ Y U ò õ
d] u z u] v] P		/ v š	ñ î í U ò ì ì U ð ñ ó Y
d] u z Á l v		/v š	ò ì î U ñ î í U ð ò ô Y
d] u z Ç		/ v š	î ô ò ï U î î ð ñ U í ì ñ ð

Hive

,] À “ ”

µ •] v • •

>] v µ Æ l } µ Œ • l ,] À l š µ l •] v • • z] Œ o X • À µ •] v • •

í ì

î ì ì ì

í ì ì ñ ì ì

À P z Ç

À Ê P Á l v

			ñ ï	
		í ñ	í ì rñ	
		î ì	í ì rñ ô ð	
		î ì	í ì rñ ó í X ñ ï	
		í ñ	ï ï ï ï ï	

Linux “ initnetwork ” “ ”

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

/

v ß ^ò ì Ì >] v µ Æ

í Ì ø : <

, } } ‰ î X ò X ð

,] À

□

		í ì	,] À í ì	
			í	

í ñ í ñ

í ì

ñ

ï

			ñ ï	
		î ì	Á Z Œ ô ð	
		í ñ	Á Z Œ ñ	
		î ì	Á Z Œ ô ð	
		í ñ	ï ï ï ï ï	

Linux “ initnetwork ” “ ”

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

u ‰ À • À
u ‰ v } v u i } • o ‰ š v

u ‰ v }	v u	i }	• o	‰ š v }
ó ï ò õ	^ D / d ,	> Z <	ô ì ì	ï ì
ó ð õ õ	> > E	^ > ^ D E	í ò ì ì	î ì
ó ñ ò ò	: K E ^	D E ' Z	î õ ó ñ	ï ì
ó ò ñ ð	D Z d / E	^ > ^ D E	í î ñ ì	ï ì
ó ò õ ô	> <	D E ' Z	î ô ñ ì	í ì

ó ô ï õ	< / E '	W Z ^ / E d	ñ ì ì ì	î ì

Hive

Z] À

,] À " " u ‰] v

>] v µ Æ l } µ Œ • l ,] À l l š u ‰ • Ì À ,] À

,] À

iii.

iv.

v.

vi.

" " " "

v ‰ Ì ^ ò ì Ì >] v µ Æ

í Ì ð : <

, } } ‰ î X ò Ì X ð

,] À

□

		í ì	,] À í ì í	
		í ñ	í ñ í ì ñ ï	
		í ì	í ì ı ñ • } Œ Ş ñ ñ	
		í ì	í ì ı ñ • } Œ Ş ñ ñ o] u] š ñ ñ	
		î ì	u Æ u] v ô	

			ð	
		î ì	À P • µ u ô ð	
		í ñ	ï ï ï ï ï	

Linux “ initnetwork” “ ”

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

“ ”

u ‰ v }	v u	i }	• o	‰ š v }
ó ï ò õ	^ D / d ,	> Z <	ô ì ì	ï ì
ó ð õ õ	> > E	^ > ^ D E	í ò ì ì	î ì
ó ñ ò ò	: K E ^	D E ' Z	î õ ó ñ	ï ì
ó ò ñ ð	D Z d / E	^ > ^ D E	í î ñ ì	ï ì
ó ò õ ô	> <	D E ' Z	î ô ñ ì	í ì
ó ô ï õ	< / E '	W Z ^ / E d	ñ ì ì ì	î ì

Hive

,] À “ ”

u ‰

>] v µ Æ l } µ Œ • l ,] À l š u l ‰ • À u ‰

u ‰

î ì ì ì ï ì ì ì

î ì ï ì ì ì

ï ì ô ì ì

‰ š v } À P• z o z ‰ š v }

“ ” “ ”

v ¥ ^ ò ï ì >] v µ Æ
í ï ò : <
, } } ‰ î X ò ï ð
,] À

□

		í ì	,] À í ì í	
		í ñ	í ñ í ì ñ ï	
		í ì	í ì ı ñ Á Z Œ v ñ	
		í ì	í ì ı ñ Á Z Œ v ñ	
		î ì	Á Z Œ ô ð	
		î ì	P Œ } µ Ç ô õ ï	
		í ñ	ï ï ï	

/		/v š	ï ò õ ì î ï ò õ ì ï ï ò õ ì Ÿð
d] u z Á } Œ l		/ v š	ó ô U í ð ð U õ ì ñ Y U ò õ U
d] u z u] v] P		/ v š	ñ î í U ò ì ì U ð ñ ó Y
d] u z Á l v		/v š	ò ì î U ñ î í U ð ò ô Y
d] u z Ç		/v š	î ô ò ï U î î ð ñ U í ì ñ ð Y

Hive

,] À “ ”

µ •] v • •

>] v µ Æ l } µ Œ • l ,] À l š µ l •] v • • z] Œ o X • À µ •] v • •

í ì

ñ ì ì

í ñ ì ì

“ ” “ ”

v ¼ ^ ò í Ì >] v µ Æ

í Ì : <

, } } ‰ î X ò Ì ð

,] À

□

		í ñ	,] À í ñ í	
		í ñ	í ñ í ì ñ ï	
		í ñ	í ì ı ñ Á Z Œ ñ ñ	

î ì

í ñrí ì Á Z Œ

ñ ñ

ò ì ì

í ì ì ñì ì

À PÁz} Œ Ç À ŽPÁ l v

“ ” “ ”

v ‰ ^ò ïXì >] v µ Æ
íXõ : <
, }} ‰ î X ò X ð
,] À

		í ñ	,] À í ñ í	
		í ñ	í ñ í ì ñ ï	
		í ñ	í ì ѫñ	
		î ì	Á Z Œ ô ð	
		î ì	À P ô î	
		í ñ	ï ï ï ï ï	

Œ ‹ µ •š• , d d W Œ ‹ µ •š• P š•
Æ ‰ š Z i• } v “ š X i• } v
, — — W € — U Y • U — W € — U Y • U Y ƒ
Á } Œ i • } v

¬ • Œ ‹ µ •š• , d d W
Œ ‹ µ •š• “ Z š š ‰ W l l š] ‰ u X } u l , d d W
~]] • y ‰ š Z
o Æ u o š Œ
, d D > , d D > y W š Z

etree

HTML

“ ”

“ ” “ ”

□

]	Œ ‹µ •š• ,ddW	îì	Œ ‹µ •š• ,ddW íì ^îìì_ íì	
]]	y‰ šZ	òì	ñ ñ Zšuo ñ Zšuo ñ ïñ íì	
]]]	i•}v	ñ	i•}v ñ	
]À		íñ	ï ï ï ï ï	

µŒooo] ï ,ddW Œ ‹µ •š• P š•
µš](µo•}µ‰ i•}v “ š Xi•}v
,— —W€— —UY•U —W€— —UY•UYƒ
Á}Œ i•}v

“ ”

BeautifulSoup

BeautifulSoup BeautifulSoup

“ ”

“ ” “ ”

]		î ì	“ ”	
]]		ò ì	ñ	

		ñ	i • } v ñ	
] À		í ñ	ï ï ï ï ï	

“ ”

“ ”

– D o] o • • A — — E D Z Œ (A — Z š š ‰ W l l š] ‰ u X } D W E i D U K Z] E ¥ X / ^ E E X i Z Ù u o — š Œ P š A –

;

“ ”

“ ” “ ”

]	Œ ‹µ •š•	î ì	Œ ‹µ •š• ,ddW 7... g	
	,ddW			

“ ”

urls

4 urls

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

requests BeautifulSoup

;

urls

requests

BeautifulSoup BeautifulSoup

BeautifulSoup

“ ”

“ ” “ ”

□

]		í ñ	ñ í ì	
]]		ò ñ	ñ	

“ ”

urls

4 urls

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

urllib3 etree

;

urls

urllib3

etree.HTML get html

html

“ ”

“ ” “ ”

□

]		í ñ	ñ í ì	
]]		ò ñ	ñ í ì ñ í ì í ì í ì í ì ñ	
]]	i • } v	ñ	i • } v ñ	
]À		í ñ	ï ï ï ï ï	

“ ”

“ ”

urls

4 urls

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

requests re

;

urls

requests

'<h1><ahref="http://tipdm.com:80/.*.jhtml" target="_blank">( )</h1>'

' ( ) '

“ ”

“ ” “ ”

□

]		í ñ	ñ í ì	
]]		ò ñ	ñ ñ ñ í ì í ì í ì í ì í ì	
]]	i • } v	ñ	i • } v ñ	
]À		í ñ	ï ï ï ï ï	

“ ”

“ ”

□ □

□ □

urls

4 urls

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

requests BeautifulSoup

;

urls

requests

BeautifulSoup BeautifulSoup

BeautifulSoup

data

“ ”

“ ” “ ”

□

]		í ñ	ñ í ì	
]]		ò ñ	ñ í ì ñ í ì í ì í ì í ì ñ	
]]	i • } v	ñ	i • } v ñ	
]À		í ñ	ï ï ï ï ï	

“ ”

“ ”

□ □

□ □

urls

4 urls

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

urllib3 etree

;

urls

urllib3

etreeHTML html

html

data

“ ”

“ ” “ ”

□

]		í ñ	ñ í ì	
]]		ò ñ	ñ í ì ñ í ì í ì í ì í ì ñ	
]]	i • } v	ñ	i • } v ñ	
]À		í ñ	ï ï ï ï ï	

“ ”

“ ”

urls

4 urls

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

requests BeautifulSoup

;

urls

url

BeautifulSoup BeautifulSoup

BeautifulSoup

data

“ ”

“ ” “ ”

□

]		í ñ	ñ í ì	
]]		ò ñ	ñ í ì ñ í ì í ì í ì í ì ñ	
]]	i • } v	ñ	i • } v ñ	
]À		í ñ	ï ï ï ï ï	

“ ”

“ ”

urls

4 urls

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

requests re

;

urls

url requests 5Q EMC /Span <</MCID 1156 Tf 1 0 0 1 128.9 352.85 Tm08871 0

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

' ( ) '

ƛ

□ƛ

ƛ

]		í ñ	ñ í ì	
]]		ò ñ	ñ ñ ñ í ì í ì í ì í ì í ì	
]]	]•}v	ñ	i•}v ñ	
]À		í ñ	ƛ ï ƛ ï ï	

ƛ

•µŒÀ]À	ìA íA
v u	
• Æ	ìA íA
P	
•] •‰	
‰ Œ Z	
š] l š	
(Œ	
(Œ z• š]}v	
]v	
u Œl	A Z Œ }µŒP YAYµ v•š}^ÁA^}µšZ u‰š}v
} š	
} Ç	
Z}u X •š	

“ ”

“ ”

“ ”

x
x “ ”
y “ ”
“ ”

“ ” “ ”

“ ”

“ ”

“ ”

5

x
x “ ”
y “ ”
“ 5 ”

5

“ ” “ ”

□

]	W Ç š Z } v	î ñ	‰ v •ñ u š ‰ o } š o] Xñ‰ Ç ‰ o } š í ì ñ	
]]		ï ì	í ñ í ì ñ	
]]]	ñ	ï ì	í ì	

			í ì “ ” ñ ñ	
] À		í ñ	ï ï ï ï ï	

>] v µ Æ “ Á P ‰ W Œ } } š l i µ ‰ Ç š Œ z v } š } } l
Z š š ‰ W l l š • Œ X š] ‰ u X v š W ô À l , v µ] v P l z u u } µ Z o] ’ v ñ Ì ñ Æ o • Æ
À v] v P z u Z] v >] X v Æ µ ø • l Æ Œ } } š l i µ ‰ Ç š Œ z v } š } } l

□ □
□ □ □

¬] • ‰ v • u š ‰ o } š o] X ‰ Ç ‰ o } š
‰ o š X Œ Z Ù Y v G Œ X ¤ • v Œ [•] (Á v Z Y µ v z] • [v ‰] þ š X Œ Z Ù Y v G Œ X ¤ • v Œ [•] (A
Z ^] u , []

“ ” “ ” □

2017 6 5

x
x “ ”
y “ ”
zip
“ 2017 6 5 ”

¬]] • “ ” “ ”

“ ”

“ ” “ ”

] W Ç š Z }

>]vµÆ “ÁP ŠW lŒ}}šliµ‰ÇšŒzv}š }}l
Zšš‰Wll š •Œ Xš]‰ uXv šWŏ•µ‰vŒuuŒµšž•lño •XÆo•Æ
•µ‰ Œu Œl šz•>]vµXÆŒ}Æšliµ‰ÇšŒzv}š }}l

□ □

] ‰ v • u š‰o}šo] X‰Ç‰o}š
‰ošXŒ ZWvŒEXu••vŒE[•](A vZYµ vz] • v ‰]ošXŒ ZWvŒEXu••vŒE[•](A
Z^]u,[]

(ii) “ ” “ ”

'

'

(iii) “ ” “ ” “ ”

x
x
y

“ ” “ ”

□

]	W Ç š Z } v	î ñ	‰ v •ñ u š ‰ o } š o] Xñ‰ Ç ‰ o } š í ì ñ	
]]		ï ì	í ì	

í ì

ñ

ñ

Z] u , []

“ ”

' '

(iii) “ ”

;

x
x “ ”
y “ ”

“ ”

“ ” “ ”

□

]	W Ç š Z } v	î ñ	‰ v •ñ u š ‰ o } š o] Xñ‰ Ç ‰ o } š í ì ñ	
]]		ï ì	í ñ ñ ñ ñ	

]]]		ï ì	í ì í ì “ ” ñ ñ	
]À		í ñ	ï ï ï ï ï	

>]vµÆ “ÁP nšW lŒ}}šliµ‰ÇšŒzv}š }}l
Zšš‰Wll š •Œ Xš]‰ uXv šWòôí,Z,zµv}uvulu}•µ‰XñlñÀ
,Zz }uu z• ‰X>]•vÀµÆlŒ}}šliµ‰ÇšŒzv}š }}l

表 4-2 HR_comma_sep.csv 字段说明

• š]•(š]}vzo À	ìr ír
o •šz À oµ š]}v	
vµu Œz‰Œ}i š	
À Œ P zu}všoÇz	ZŒ
š]u z•‰ v z }u‰	
t}ŒlzÇ] vš	ìr ír
o (š	ìr ír
‰Œ}u}š]}vzo •šz	ñ ìr ír
• o •	
• o ŒÇ	

' '

x
x
y
' '

“ ” “ ”

□

			í ì í ì	
] À		í ñ	ï ï ï ï ï	

“

”

表 4-3 movies.csv 字段说明

]	
] u z]	/ D
‰ } ‰ µ o Œ] š Ç	
µ P š	
Œ À v µ	
} Œ] P] v o z š] š o	
• š	
] Œ š } Œ	
} À Œ À] Á	
Œ µ v š] u	
P v Œ •	
Œ o • z š	
À } š z } µ v š	

À } š z À Œ P	
Œ o • z Ç Œ	
‰ Œ } ( ] š	

10

x
x
y
for

10

10

;

“ ” “ ”

□

]	W Ç š Z } v	î ñ	‰ v •ñ u š ‰ o } š o] Xñ‰ Ç ‰ o } š í ì ñ	
]]		ï ì	í ì	

			í ì Œ u ñ ñ	
]]]	í ì	ï ì	í ñ ñ X : µ Œ • •] t } Œ o ñ ñ	
] À		í ñ	ï ï ï ï ï	

“

”

□ □

□

□ □ □

□ □ □ □

□ □

表 4-4 mum_baby_trade_history.csv 字段说明

µ • Œ z]	]
µ š] } v z]	
š z]	]
š í	

µ Ç z u } µ v š	
Ç	

;

“ ”

;

x
x “ ”
y “ ”
“ ”

“ ” “ ”

□

]	W Ç š Z } v	î ñ	‰ v •ñ u š ‰ o } š o] Xñ‰ Ç ‰ o } š í ì ñ	
]]		ï ì	í ñ ñ	

			“ □ ” ñ ñ	
]]]		ï ì	í ì í ì “ ï ô” í ì	
] À		í ñ	ï ï ï ï ï ï	

“

”

□

□ □

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

“ ”

;

x

x

y

“ ”

“ ”

“ ” “ ”

”

“ ”

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

“ ”

“ ”

“ ”

;

x
x
y

10

“ ” “ ”

□

]	W Ç š Z } v	î ñ	‰ v •ñ u š ‰ o } š o] Xñ‰ Ç ‰ o } š í ì	

ñ

“ ”

x
x
y

“ ” “ ”

□

]	W Ç š Z } v	î ñ	‰ v •ñ u š ‰ o } š o] Xñ‰ Ç ‰ o } š í ì ñ	
]]		ï ì	í ì í ì ñ ñ	
]]]		ï ì	í ì í ì ñ ñ	

] À		í ñ	ï ï ï ï ï	

“

”

“ ”

x
x
y

“ ”

“ ” “ ”

□

]	W Ç š Z } v	î ñ	‰ v ¶ñ u š ‰ o } š o] Xñ‰ Ç ‰ o } š í ì ñ	
]]		ï ì	í ì í ì “] } Œ ” ñ ñ	
]]]		ï ì	í ì í ì ñ ñ	
] À		í ñ	ï ï ï ï ï	

“

”

□ □

□

□

2000 data.iloc[0,5:]

“ 2000 ”

□

data.iloc[68, 2:5]

x
x
y

□

“ ” “ ”

□

]	W Ç š Z } v	î ñ	‰ v •ñ u š ‰ o } š o] Xñ‰ Ç ‰ o } š í ì ñ	
]]	î ì í ó	ï ì	í ì í ì ñ ñ í ì	

]]]	î ì ì ì	ï ì	í ì í ì ñ ñ	
] À		í ñ	ï ï ï ï ï	

“

”

“ ” “ ”

x “ ”
y “ ”
“ ”

“ ” “ ”

“ ” “ ”