入门

- 基本语法
- 库的运用
- 循环语句
- if-else逻辑语句
- 函数运用

In [74]:

```
# 调用os库
import os

def main():
    # print函数可以输出文字
    print('Hello,this is Python!')

    # 调用os.getcwd()函数
    print('当前工作目录是：' , os.getcwd())

    # 声明list01变量（是一个list列表）
    list01 = [-1,0,1]

    for r in list01:
        print(r)

    JudgeNum(list01[0])


def JudgeNum(x):
    if x < 0:
        print('负数')
    elif x > 0:
        print('正数')
    else:
        print('零')
    return x
```

In [75]:

```
main()
```

```
Hello,this is Python!
当前工作目录是： /Users/wdt/Desktop/py
-1
0
1
负数
```

python连接Excel

In [12]:

```
import pandas as pd
```

导入csv/xlsx文件

In [8]:

```
df1 = pd.read_excel('/Users/wdt/Desktop/tpy/raw_data_pool/现券市场交易情况总结/日报/现券市场交易情况总结日报_20190613.xls')

df2 = pd.read_excel('/Users/wdt/Desktop/tpy/raw_data_pool/二级成交/成交统计2021年3月15日.xlsx', header=1)
```

In [11]:

```
df2.iloc[:,:-2]
```

Out[11]:

	方向	类型	剩余期限	代码	简称	浮动利率	含权	价格	价格备注	成交久期	昨日平均	昨日中债估值	中债估值备注	中证估值	成交-中债(BP)	成交-中证(BP)	债券余额(亿)	时间
0	TKN	国债	9.68Y	200016.IB	20附息国债16	固定利率	不含权	3.2710	NaN	NaN	3.251	3.2575	NaN	3.2550	1.35	1.60	2481.0	19:20:25
1	GVN	国开	9.60Y	200215.IB	20国开15	固定利率	不含权	3.7075	到期	NaN	3.692	3.698	到期	3.6950	0.95	1.25	2736.6	17:51:44
2	GVN	国债	29.50Y	200012.IB	20附息国债12	固定利率	不含权	3.7675	NaN	NaN	3.758	3.7625	NaN	3.7600	0.50	0.75	2503.3	17:47:15
3	TKN	国开	8.18Y	190210.IB	19国开10	固定利率	不含权	3.7075	到期	NaN	3.699	3.7	到期	3.6934	0.75	1.41	2473.3	17:39:49
4	TKN	国开	1.17Y	190207.IB	19国开07	固定利率	不含权	2.8650	到期	NaN	2.865	2.865	到期	2.8650	0.00	0.00	1227.3	17:39:31
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3357	TKN	国开	9.60Y	200215.IB	20国开15	固定利率	不含权	3.6950	NaN	NaN	3.692	3.698	NaN	3.6950	-0.30	0.00	2736.6	08:21:06
3358	TKN	国开	9.60Y	200215.IB	20国开15	固定利率	不含权	3.6950	NaN	NaN	3.692	3.698	NaN	3.6950	-0.30	0.00	2736.6	08:20:35
3359	TKN	国开	9.60Y	200215.IB	20国开15	固定利率	不含权	3.6950	NaN	NaN	3.692	3.698	NaN	3.6950	-0.30	0.00	2736.6	08:19:23
3360	TRD	国开	9.60Y	200215.IB	20国开15	固定利率	不含权	3.6975	NaN	NaN	3.692	3.698	NaN	3.6950	-0.05	0.25	2736.6	08:19:07
3361	GVN	国开	9.60Y	200215.IB	20国开15	固定利率	不含权	3.6950	NaN	NaN	3.692	3.698	NaN	3.6950	-0.30	0.00	2736.6	08:18:37

3362 rows × 18 columns

导出为csv/xlsx文件

In [ ]:

```
df.to_csv('df01.csv')
df.to_excel('df02.xlsx')
```

python连接数据库

In [14]:

```
import pymysql
from sqlalchemy import create_engine
```

In [ ]:

```
conn = pymysql.connect(
          host = host,
          user = username,
          passwd = password,
          db = 'finance',
          port=port,
          charset = 'utf8'
      )
engine = create_engine(mysql+pymysql://user:password@host:port/finance?charset=utf8)
```

In [21]:

```
net_buy_bond = pd.read_sql('select * from Net_buy_bond', conn)
```

In [ ]:

```
# df.to_sql(..)
```

常用数据类型

- int, float
- list, np.array
- pd.series, pd.DataFrame

数字

In [58]:

```
a1 = 1
a2 = 1.22
print(type(a1), type(a2))
```

```
<class 'int'> <class 'float'>
```

一维

In [ ]:

```
list01 = [1,2,3]
array01 = np.array([1,2,3])
```

多维

In [63]:

```
df['DR007']
```

Out[63]:

```
0          NaN
1          NaN
2          NaN
3          NaN
4          NaN
         ...  
4850    2.1516
4851    2.2294
4852    2.1893
4853    2.1664
4854    2.2039
Name: DR007, Length: 4855, dtype: float64
```

In [67]:

```
cash_cost.iloc[-5:,:]
```

Out[67]:

	DR001	DR007	GC007	shibor_3m	R007	date
4994	2.0012	2.1516	2.138	2.434	2.1131	2021-06-11
4995	2.0971	2.2294	2.250	2.434	2.2556	2021-06-15
4996	1.9960	2.1893	2.216	2.439	2.2518	2021-06-16
4997	1.8733	2.1664	2.217	2.441	2.2255	2021-06-17
4998	2.0290	2.2039	2.350	2.444	2.2142	2021-06-18

数据处理常用方法

- 数据框（DataFrame）的清洗
  - 数据框的分组
  - 不同数据框的组合
  - 数据框变量（列）的衍生
- 数据框的内部筛选

数据清洗

In [26]:

```
df = pd.read_sql('select * from Net_buy_bond', conn)
```

In [27]:

```
df['date']
df['机构名称']
df['期限']
```

Out[27]:

```
0        1年及1年以下
1             1-3年
2             3-5年
3             5-7年
4            7-10年
           ...   
72115       10-15年
72116       15-20年
72117       20-30年
72118        30年以上
72119           合计
Name: 期限, Length: 72120, dtype: object
```

In [31]:

```
df_jjgs = df.loc[df['机构名称']=='基金公司及产品']
```

In [34]:

```
df_jjgs.groupby(by='期限').sum()
```

Out[34]:

期限	国债-新债	国债-老债	政策性金融债-新债	政策性金融债-老债	中期票据	短期/超短期融资券	企业债	地方政府债	同业存单	资产支持证券	其他	合计
1-3年	275.828220	-458.641110	805.80000	8656.359600	1970.430580	0.0000	127.216970	148.001186	0.000	37.131806	3184.117870	14746.295200
10-15年	-0.520000	-33.550000	-2.94000	-7.810000	62.980000	0.0000	-4.220000	0.110000	0.000	3.880000	16.555268	34.445268
15-20年	-0.800000	-0.060000	-4.06000	-11.060277	5.717242	0.0000	2.110000	-2.580000	0.000	10.390000	58.150000	57.836965
1年及1年以下	1661.689170	1886.457040	1073.16641	5619.848700	326.366740	23214.7035	26.786626	166.320000	-1934.429	27.520000	2180.660600	34248.779000
20-30年	406.946850	-342.101690	0.98000	0.000000	76.407776	0.0000	0.500000	-13.460000	0.000	-1.290000	1.420000	129.482930
3-5年	198.679438	-283.676270	1115.92412	4496.065900	542.142118	0.0000	269.355258	319.268970	0.000	-4.570000	1308.896300	7962.185800
30年以上	74.920000	-71.870689	0.00000	0.000000	0.000000	0.0000	0.000000	-1.050000	0.000	0.000000	0.000000	1.979311
5-7年	99.105130	-1181.452566	601.01000	4408.524580	20.185514	0.0000	70.870800	323.156400	0.000	15.811970	7.319340	4364.531200
7-10年	605.303870	-289.487280	2365.81000	1616.192400	433.009441	0.0000	40.400000	-8.336705	0.000	-26.351016	1591.346000	6290.416700
合计	3321.222700	-774.372576	5955.58053	24778.361000	3463.499410	23214.7035	533.089650	931.069860	-1934.429	62.532760	8348.475000	67762.963000

In [35]:

```
df_jjgs.groupby(by='date').sum()
```

Out[35]:

date	国债-新债	国债-老债	政策性金融债-新债	政策性金融债-老债	中期票据	短期/超短期融资券	企业债	地方政府债	同业存单	资产支持证券	其他	合计
2019-06-13	-3.93	-0.41	85.96	37.86	-19.62	22.18	-2.86	-0.62	530.42	0.00	-6.71	642.28
2019-06-14	1.80	-4.28	44.48	-3.63	-6.74	75.32	-13.60	-0.26	101.90	0.00	-20.00	175.01
2019-06-15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2019-06-16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2019-06-17	23.70	18.02	46.77	47.20	22.61	31.84	-6.23	-0.40	157.96	0.00	-7.68	333.80
...	...	...	...	...	...	...	...	...	...	...	...	...
2021-06-08	-9.55	-24.91	31.27	-247.62	24.54	24.98	0.75	1.32	-28.22	-12.42	-42.00	-281.90
2021-06-09	0.34	-22.69	3.00	70.59	-40.61	10.76	7.01	4.10	-32.00	-8.85	7.74	-0.59
2021-06-10	-4.86	-41.91	-9.20	-48.01	-18.71	57.12	3.84	-14.12	38.24	-1.56	26.49	-12.69
2021-06-17	18.26	-8.46	34.63	73.80	-29.86	91.26	-1.62	-0.86	88.18	-18.38	40.62	287.60
2021-06-18	13.52	-26.85	12.36	249.47	58.86	35.74	-5.98	0.19	82.26	-2.66	-63.44	353.49

601 rows × 12 columns

In [36]:

```
rates = pd.read_sql('select * from rates', conn)
cash_cost = pd.read_sql('select * from cash_cost', conn)
rates_us = pd.read_sql('select * from rates_us', conn)
```

In [44]:

```
df = pd.merge(rates,cash_cost)
df
# pd.merge(rates, rates_us, left_on='date', right_on='date')
```

Out[44]:

	国债1年	国债3年	国债5年	国债7年	国债10年	地方1年	地方3年	地方5年	地方7年	地方10年	...	中票_AA-_5y	农发10年	口行10年	国债30年	date	DR001	DR007	GC007	shibor_3m	R007
0	2.5850	2.7271	2.8674	3.0057	3.2096	NaN	NaN	NaN	NaN	NaN	...	NaN	3.6348	3.6348	4.4600	2002-01-04	NaN	NaN	NaN	NaN	2.1389
1	2.6009	2.7380	2.8728	3.0055	3.2003	NaN	NaN	NaN	NaN	NaN	...	NaN	3.6906	3.6906	4.3724	2002-01-07	NaN	NaN	NaN	NaN	2.1379
2	1.9156	2.3842	2.7890	3.1302	3.5225	NaN	NaN	NaN	NaN	NaN	...	NaN	3.7607	3.7607	2.4752	2002-01-08	NaN	NaN	NaN	NaN	2.1388
3	1.9040	2.4036	2.8317	3.1884	3.5896	NaN	NaN	NaN	NaN	NaN	...	NaN	3.7644	3.7644	2.1569	2002-01-09	NaN	NaN	NaN	NaN	2.1382
4	1.8987	2.3954	2.8216	3.1772	3.5784	NaN	NaN	NaN	NaN	NaN	...	NaN	3.7702	3.7702	2.1966	2002-01-10	NaN	NaN	NaN	NaN	2.1374
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4850	2.4459	2.8485	3.0085	3.1229	3.1276	2.5203	3.0238	3.2108	3.3401	3.4262	...	6.7004	3.5575	3.5650	3.6625	2021-06-11	2.0012	2.1516	2.138	2.434	2.1131
4851	2.4699	2.8466	3.0044	3.1200	3.1176	2.5310	3.0498	3.2139	3.3401	3.4402	...	6.7002	3.5551	3.5625	3.6601	2021-06-15	2.0971	2.2294	2.250	2.434	2.2556
4852	2.4945	2.8490	3.0101	3.1300	3.1301	2.5299	3.0576	3.2271	3.3408	3.4402	...	6.6902	3.5794	3.5869	3.6830	2021-06-16	1.9960	2.1893	2.216	2.439	2.2518
4853	2.5272	2.8548	3.0236	3.1400	3.1401	2.5318	3.0710	3.2314	3.3435	3.4506	...	6.7002	3.5651	3.5725	3.7305	2021-06-17	1.8733	2.1664	2.217	2.441	2.2255
4854	2.5274	2.8406	2.9867	3.1109	3.1202	2.5318	3.0690	3.2307	3.3539	3.4554	...	6.7038	3.5656	3.5731	3.6948	2021-06-18	2.0290	2.2039	2.350	2.444	2.2142

4855 rows × 48 columns

In [47]:

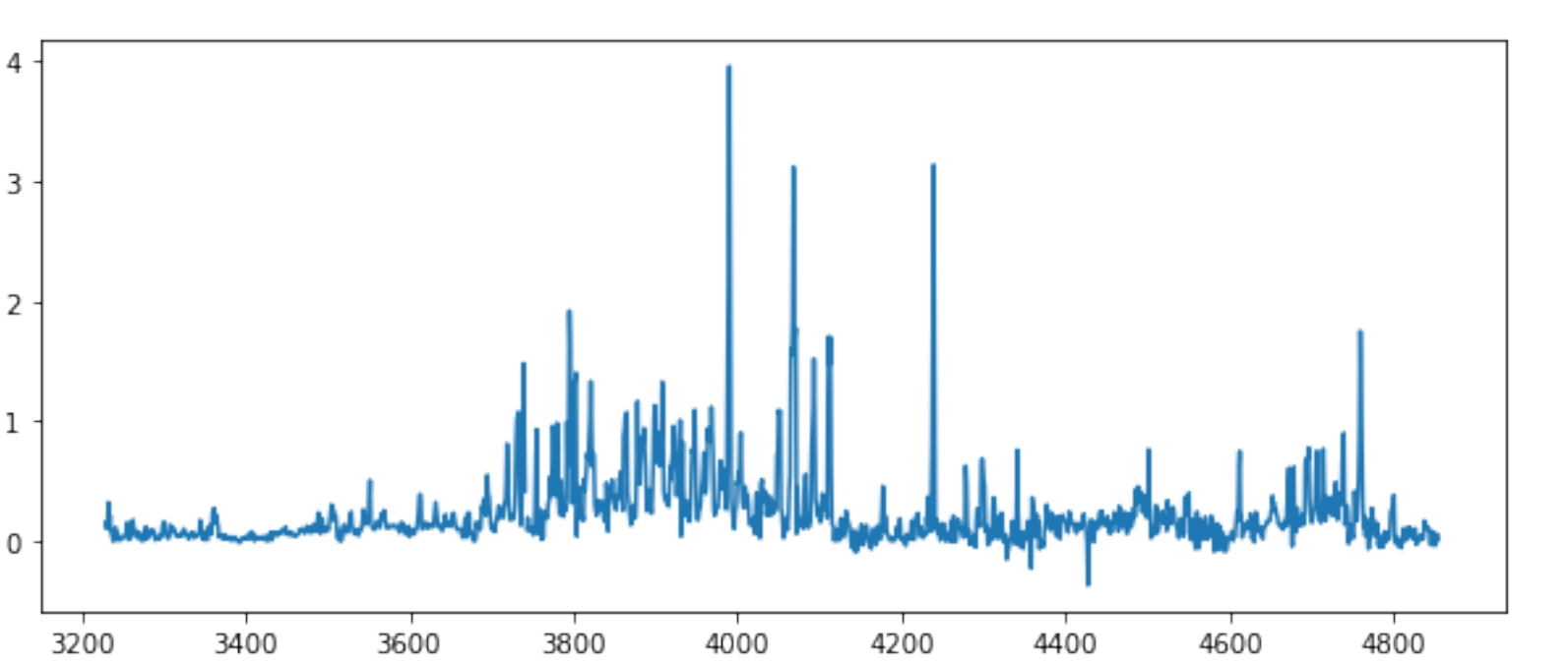
```
df['R_DR利差'] = df['R007'] - df['DR007']
```

In [52]:

```
df['R_DR利差'].plot(figsize=(10,4))
```

Out[52]:

```
<AxesSubplot:>
```



以二级利率债成交数据

In [80]:

```
data = pd.read_sql('select * from secondary_rate_sec where date="2021-06-01";', engine)
```

In [ ]:

```
# 查看数据的前五行
data.head()
```

In [88]:

```
data_gk = data.loc[ data['类型']=='国开' ]
```

In [89]:

```
data_gk.head()
```

Out[89]:

	方向	类型	剩余期限	代码	简称	浮动利率	含权	价格	价格备注	成交久期	昨日平均	昨日中债估值	中债估值备注	中证估值	成交-中债(BP)	成交-中证(BP)	债券余额(亿)	时间	date
0	TRD	国开	9.75Y	210205.IB	21国开05	固定利率	不含权	3.4875	None	None	3.487	3.48	None	3.480	0.75	0.75	2690.0	19:16:22	2021-06-01
1	TKN	国开	9.75Y	210205.IB	21国开05	固定利率	不含权	3.4875	到期	None	3.487	3.48	到期	3.480	0.75	0.75	2690.0	18:25:37	2021-06-01
2	GVN	国开	9.75Y	210205.IB	21国开05	固定利率	不含权	3.4875	None	None	3.487	3.48	None	3.480	0.75	0.75	2690.0	18:21:41	2021-06-01
3	GVN	国开	9.75Y	210205.IB	21国开05	固定利率	不含权	3.4875	None	None	3.487	3.48	None	3.480	0.75	0.75	2690.0	18:19:51	2021-06-01
6	TKN	国开	9.39Y	200215.IB	20国开15	固定利率	不含权	3.5075	到期	None	3.510	3.5075	到期	3.505	0.00	0.25	2736.6	18:05:59	2021-06-01

In [90]:

```
data_gk['代码'].value_counts()
```

Out[90]:

```
210205.IB      1170
200215.IB       307
210203.IB       229
210202.IB       107
200212.IB        74
210206.IB        72
180210.IB        35
200202.IB        35
150218.IB        34
200210.IB        29
190203.IB        27
200205.IB        24
200203.IB        20
190210.IB        19
150210.IB        18
190207.IB        17
180211.IB        15
190202.IB        14
200207.IB        14
210201.IB        13
160210.IB        12
190205.IB        11
190215.IB        11
170210.IB        10
200208.IB        10
190214.IB        10
150205.IB         8
180212.IB         8
210212.IB         7
170206.IB         7
190208.IB         6
200216.IB         5
217706.IB         4
130240.IB         4
200204.IB         3
160218.IB         3
210204.IB         2
170212.IB         2
160213.IB         2
180205.IB         2
010216.IB         1
150221.IB         1
180204.IB         1
2102001QF.IB      1
190216.IB         1
190204.IB         1
130231.IB         1
217702.IB         1
Name: 代码, dtype: int64
```

In [93]:

```
data_gk.loc[data_gk['代码'] == '210205.IB']
```

Out[93]:

	方向	类型	剩余期限	代码	简称	浮动利率	含权	价格	价格备注	成交久期	昨日平均	昨日中债估值	中债估值备注	中证估值	成交-中债(BP)	成交-中证(BP)	债券余额(亿)	时间	date
0	TRD	国开	9.75Y	210205.IB	21国开05	固定利率	不含权	3.4875	None	None	3.487	3.48	None	3.48	0.75	0.75	2690.0	19:16:22	2021-06-01
1	TKN	国开	9.75Y	210205.IB	21国开05	固定利率	不含权	3.4875	到期	None	3.487	3.48	到期	3.48	0.75	0.75	2690.0	18:25:37	2021-06-01
2	GVN	国开	9.75Y	210205.IB	21国开05	固定利率	不含权	3.4875	None	None	3.487	3.48	None	3.48	0.75	0.75	2690.0	18:21:41	2021-06-01
3	GVN	国开	9.75Y	210205.IB	21国开05	固定利率	不含权	3.4875	None	None	3.487	3.48	None	3.48	0.75	0.75	2690.0	18:19:51	2021-06-01
10	TKN	国开	9.75Y	210205.IB	21国开05	固定利率	不含权	3.4875	到期	None	3.487	3.48	到期	3.48	0.75	0.75	2690.0	18:01:58	2021-06-01
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4090	TKN	国开	9.75Y	210205.IB	21国开05	固定利率	不含权	3.4775	到期	None	3.487	3.48	到期	3.48	-0.25	-0.25	2690.0	08:38:12	2021-06-01
4091	TKN	国开	9.75Y	210205.IB	21国开05	固定利率	不含权	3.4775	到期	None	3.487	3.48	到期	3.48	-0.25	-0.25	2690.0	08:38:12	2021-06-01
4092	TKN	国开	9.75Y	210205.IB	21国开05	固定利率	不含权	3.4750	到期	None	3.487	3.48	到期	3.48	-0.50	-0.50	2690.0	08:38:06	2021-06-01
4093	TKN	国开	9.75Y	210205.IB	21国开05	固定利率	不含权	3.4750	到期	None	3.487	3.48	到期	3.48	-0.50	-0.50	2690.0	08:38:06	2021-06-01
4094	GVN	国开	9.75Y	210205.IB	21国开05	固定利率	不含权	3.4825	到期	None	3.487	3.48	到期	3.48	0.25	0.25	2690.0	08:20:53	2021-06-01

1170 rows × 19 columns