

# Assignment 02 Report

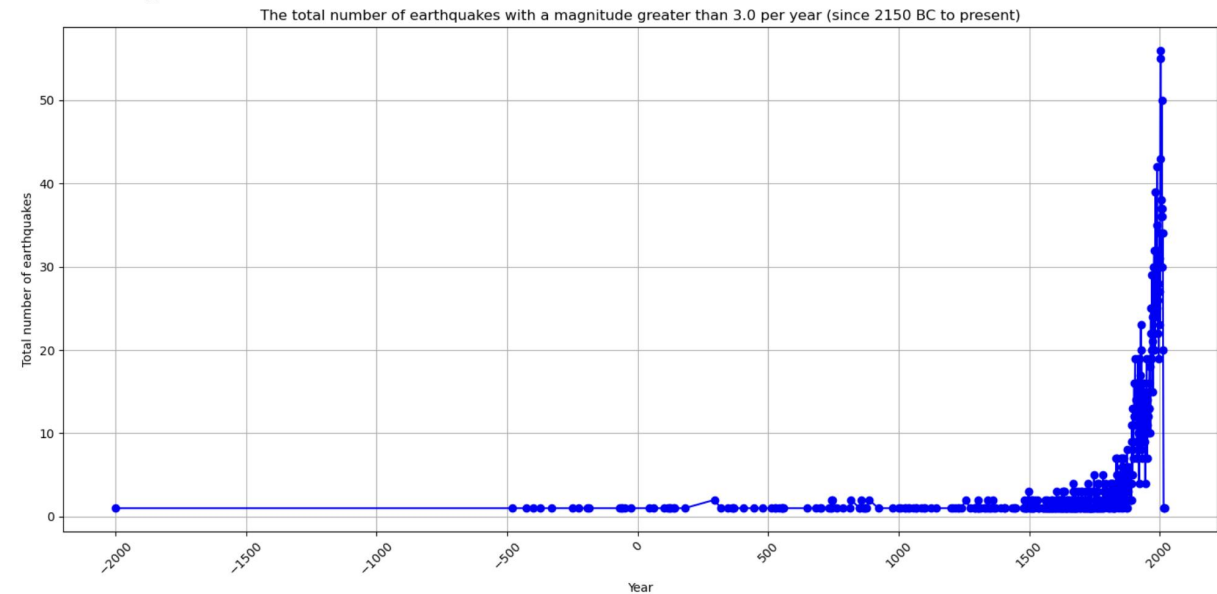
第一题：（上网搜索了关于缺失值的替换的问题：CSDN 社区的《05\_Pandas 删除，替换并提取其中的缺失值 NaN(dropna,fillna,isnull)》。第三小问的初始化一个变量来储存数值的操作上网查询了一下。）

（1）输出结果：

自公元前2150年以来，前20个国家因地震而死亡的总人数：

Country	
CHINA	2075947.0
TURKEY	1188881.0
IRAN	1011453.0
ITALY	498418.0
SYRIA	439224.0
HAITI	323478.0
AZERBAIJAN	317219.0
JAPAN	279607.0
ARMENIA	191890.0
PAKISTAN	145083.0
IRAQ	136200.0
ECUADOR	135496.0
TURKMENISTAN	117412.0
PERU	102169.0
ISRAEL	90388.0
PORTUGAL	83572.0
GREECE	80378.0
CHILE	64277.0
INDIA	63507.0
TAIWAN	57153.0

（2）输出结果：



观察到随着时间推移，发生的总数在不断升高。可能是 1.在过去由于技术限制和记录不完整，很多地震可能没有被记录下来。而随着时间的推移，地震数据的记录变得更加完整和准确，这也会使得地震总数看似在升高。2.人类活动也可能对地震的发生频率产生一定影响。例如，地下水抽取、矿产资源开采等活动可能改变地壳的应力状态，从而增加地震的风险。

(3) 输出结果:

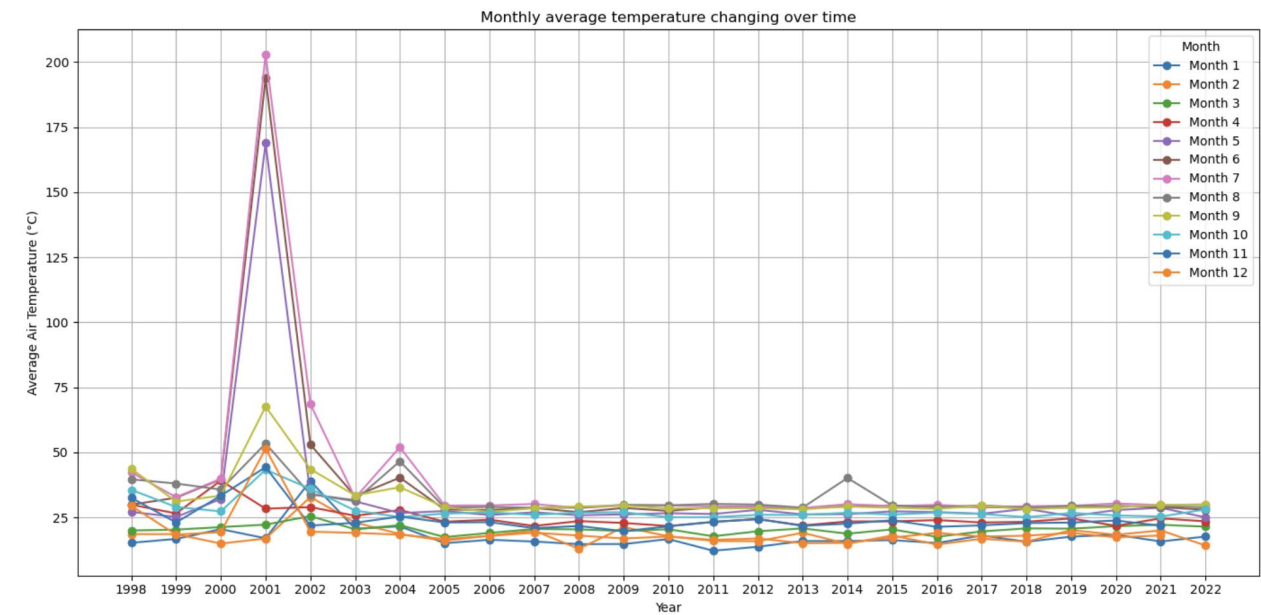
每个国家自公元前2150年以来发生的地震总数及最大地震的日期和地点（按地震总数降序）:

	Country	Total_Earthquakes	Largest_Eq_Date \
15	CHINA	623	1668.0
34	JAPAN	419	2011.0
71	INDONESIA	412	2004.0
8	IRAN	386	856.0
10	TURKEY	337	1939.0
..	...	...	...
120	KIRIBATI	1	1905.0
126	PALAU	1	1914.0
128	CENTRAL AFRICAN REPUBLIC	1	1921.0
135	LIBYA	1	1963.0
0	NaN	0	NaN

	Largest_Eq_Location
15	纬度: 35.3, 经度: 118.6
34	纬度: 38.297, 经度: 142.373
71	纬度: 3.295, 经度: 95.982
8	纬度: 36.2, 经度: 54.3
10	纬度: 39.907, 经度: 39.586
..	...
120	纬度: -1.0, 经度: -168.0
126	纬度: 6.0, 经度: 132.5
128	纬度: 3.8, 经度: 16.3
135	纬度: 32.6, 经度: 21.0
0	None

[158 rows x 4 columns]

第二题：（日期转换的细节上网查询了一下相关操作，CSDN 社区的《Python 中日期和时间处理的详细教程》。温度数据的处理上网查询解决方案。）  
输出结果：



观察到大多数月份温度表现的平稳，2001 年五月份左右气温很高，这由于该月份数据中出现了很多+99999

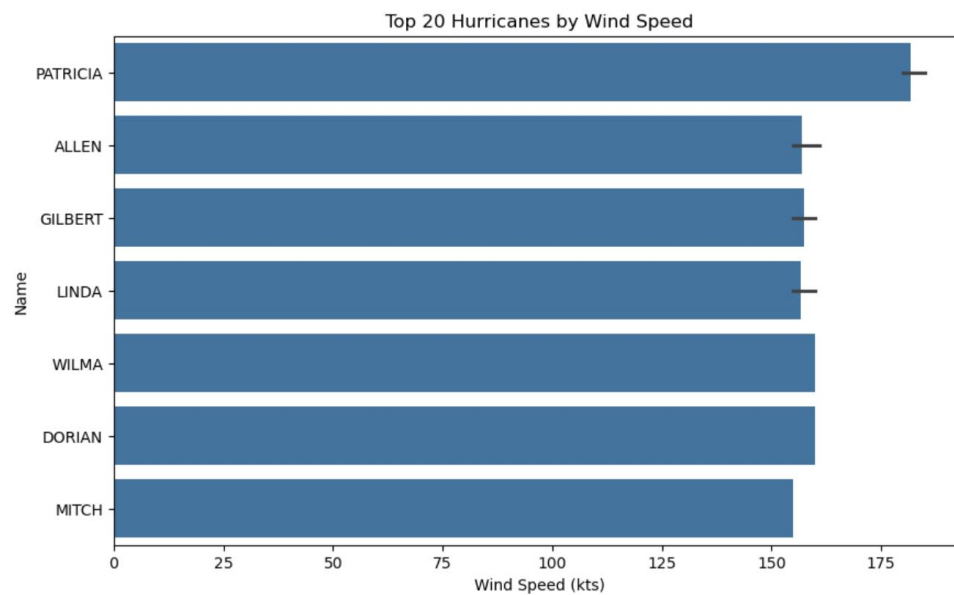
第三题：（转换数值类型的操作上网进行了查询。）

（1）输出结果：

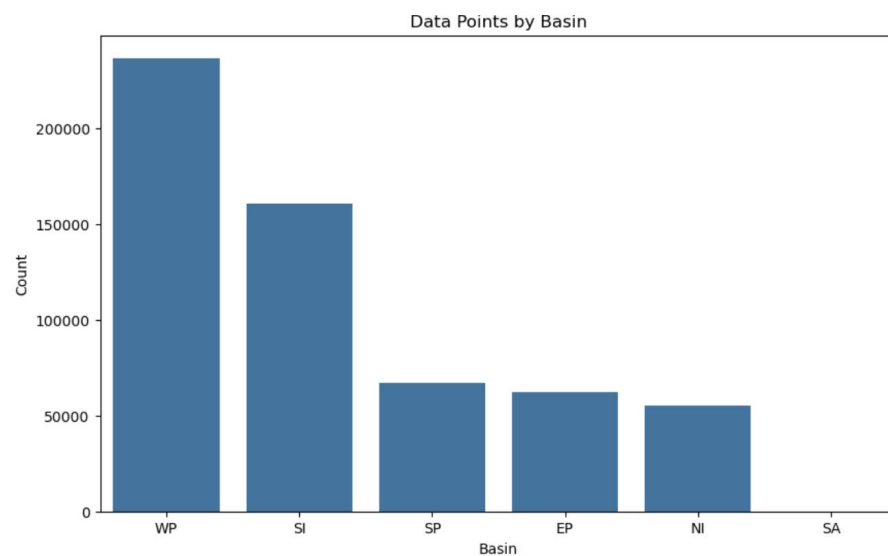
Top 10 hurricanes by wind speed:

	SID	NAME
665954	2015293N13266	PATRICIA
665952	2015293N13266	PATRICIA
665956	2015293N13266	PATRICIA
427636	1980214N11330	ALLEN
178209	1935241N23291	NaN
178210	1935241N23291	NaN
178212	1935241N23291	NaN
482074	1988253N12306	GILBERT
552459	1997253N12255	LINDA
605746	2005289N18282	WILMA

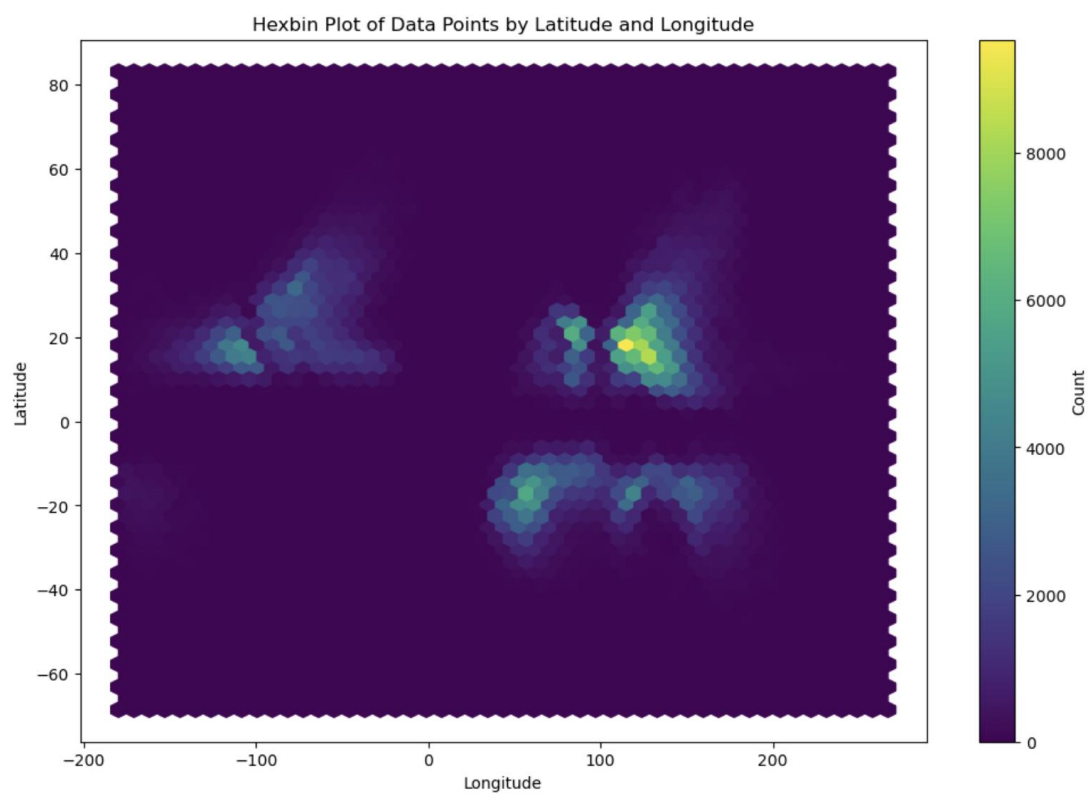
（2）输出结果：



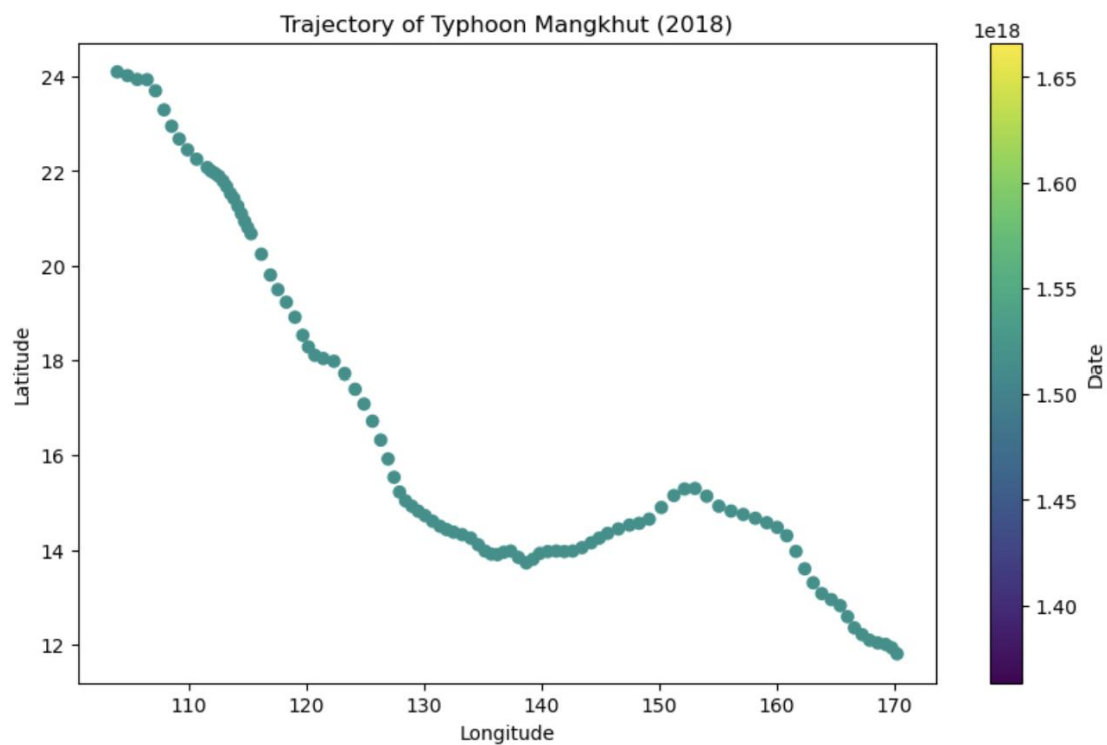
（3）输出结果：



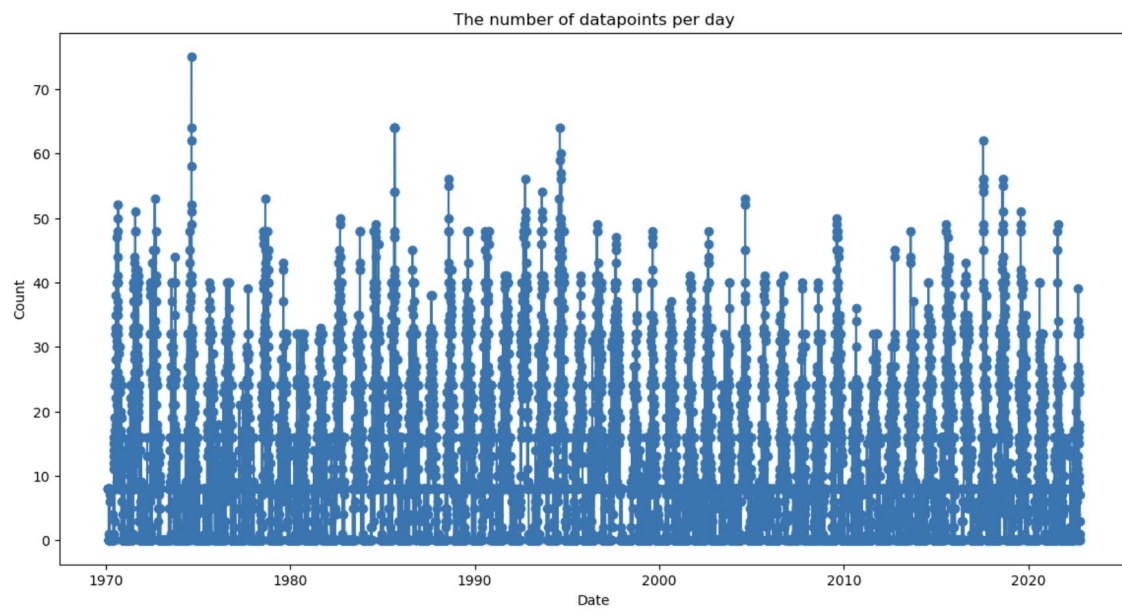
(4) 输出结果:



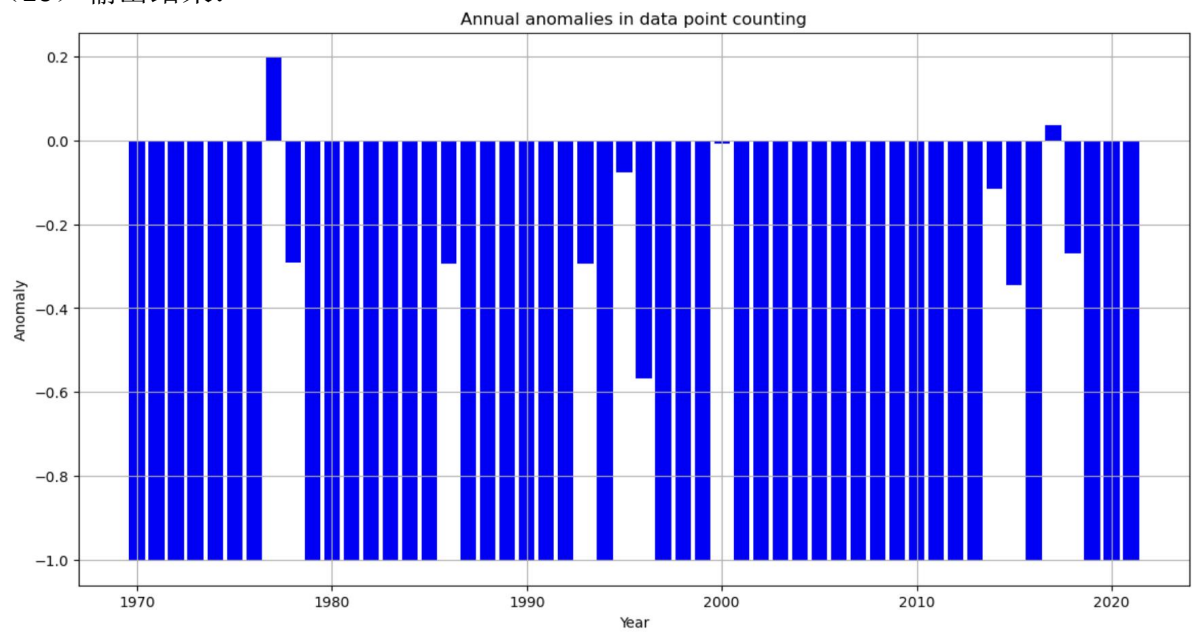
(5) 输出结果:



(7) 输出结果:



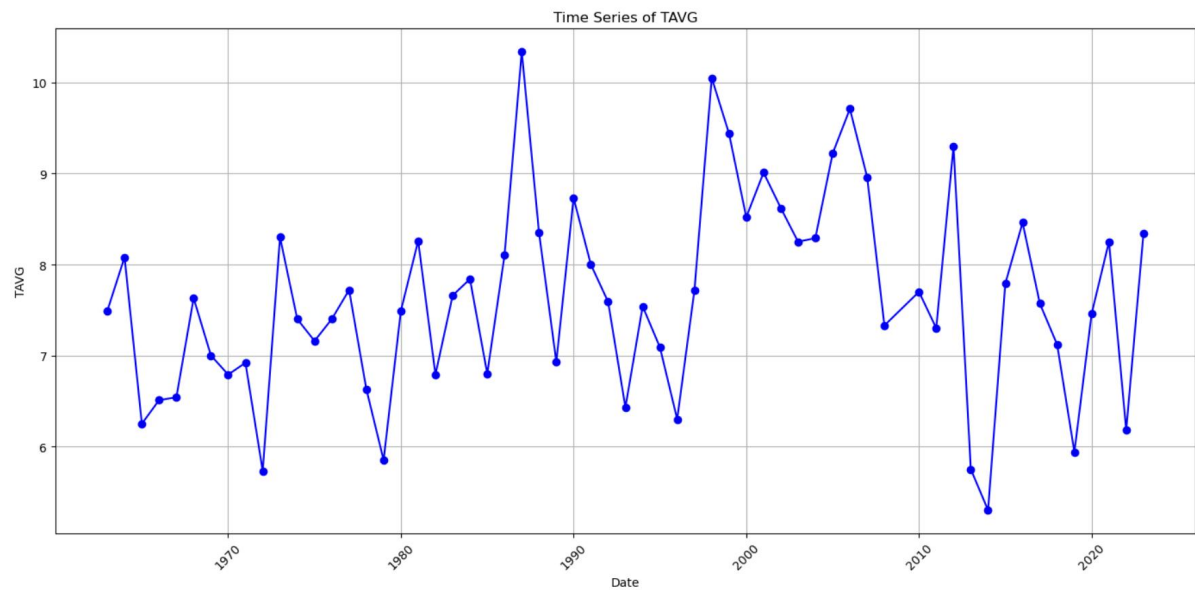
(10) 输出结果:



```
Years with anomalous hurricane activity:
Index([1970, 1971, 1972, 1973, 1974, 1975, 1976, 1979, 1980, 1981, 1982, 1983,
       1984, 1985, 1987, 1988, 1989, 1990, 1991, 1992, 1994, 1996, 1997, 1998,
       1999, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011,
       2012, 2013, 2016, 2019, 2020, 2021],
      dtype='int32', name='ISO_TIME')
```

第四题：（第三问中的统计方法的 python 语句进行了上网查询）

（2）输出结果：



（3）输出结果：

TAVG的统计信息：

```
count    60.000000
mean      7.620333
std       1.097682
min       5.300000
25%      6.890000
50%      7.580000
75%      8.292500
max      10.340000
Name: TAVG, dtype: float64
TAVG的偏度: 0.2297484589902748
TAVG的峰度: -0.09534896674444715
TAVG的缺失值数量: 3
TAVG的异常值数量: 0
TAVG与TMAX的相关性: 0.9532118124676898
```