



江西财经大学
JIANGXI UNIVERSITY OF FINANCE AND ECONOMICS

课程名称: Python语言与数据分析

课 程 报 告

项目名称 中国大学排行榜及数据分析

班 级 金融202

学 号 0204807

姓 名 宋冰凌

任课教师 肖 泉

开课学期: 2020 至 2021 学年 第 二 学期

完成时间: 2021 年 7 月 2 日

《中国大学排行榜及数据分析》数据分析报告

目 录

1 概述

2 数据描述

3 数据分析内容

4 数据分析图表

5 数据分析结果

6 总结

附录-数据分析代码

1 概述

近年来，我国参与高考的人数越来越多，填报志愿选择自己心仪的大学也越来越难，竞争越来越激烈。

省/市	2020年	2019年	2018年	2017年	2016年	2015年	2014年	2013年	2012年
各地高考报名人数详情请点击人数了解									
全国	1071万↑	1031万↑	975万↑	940万	940万↓	942万↑	939万↑	912万↓	915万↓
河南	115.8万↑	100万+↑	98.3万↑	86.3万↑	82万↑	77.2万↑	72.4万↓	71.63万↓	80.5万↑
广东	78.8万↑	76.8万↑	75.8万↑	75.7万↑	73.3万↓	75.4万↓	75.6万↑	72.7万↑	69.2万↑
四川	67万↑	65万↑	62万↑	58.3万↑	57.13万	超57万	57.17万↑	54万↑	53.8万↑
河北	62.48万↑	55.96万↑	48.6万↑	43.6万↑	42.31万↑	40.48万↓	41.82万↓	44.98万↓	45.93万↓
安徽	52.38万↑	51.3万↑	49.9万	49.9万↓	50.99万↓	54.6万↑	52.7万↑	51.1万↑	50.6万↓
湖南	53.7万↑	50万↑	45.2万↑	41.1万↑	40.16万↑	38.99万↑	37.8万↑	37.3万↑	35.2万↓
山东	53万↓	55.99万↑	59.2万↑	58.3万↓	60.2万↑	69.61万↑	55.8万↑	50万↓	51万↓
广西	50.7万↑	46万↑	40万↑	36.5万↑	33万余↑	近31万↓	31.5万↑	29.8万↑	28.5万↓
贵州	47万↑	45.8万↑	44.1万↑	41.2万↑	37.38万↑	33.05万↑	29.27万↑	24.78万↓	24.8万↑
江西	38.94万↓	42.1万↑	38万↑	36.5万↑	36.06万↑	35.46万↑	32.59万↑	27.43万↑	26.9万↓
湖北	39.48万↑	38.4万↑	37.4万↑	36.2万↑	36.14万↓	36.84万↓	40.27万↓	43.8万↓	45.7万↓
江苏	34.89万↑	33.9万↑	33万	33万↓	36.04万↓	39.29万↓	42.57万↓	45.1万↓	47.4万↓

中国大学排行榜是对中国的各大高校通过教学实力，录取分数，毕业后就业薪酬等综合起来进行排名的一个榜单，就院校地形，学科实力，就业去向，毕业薪酬等因素，在数据分析下，将可视化的数据展现在大家的面前，让更多学子可以在数据分析中，选择自己喜欢的学校。

2 数据描述

数据来源：网络爬虫

<https://www.shanghairanking.cn/rankings/bcur/2021>

和鲸平台（[Heywhale 和鲸（官网）](#) - 数据科学协同创新平台）：

引用数据中国大学综合排名 2021.csv

```
import pandas as pd
from pandas import DataFrame, Series
df = pd.read_csv('中国大学综合排名2021.csv', encoding='cp936')
df.head()
```

查看数据类型：df.dtypes

```
排名                int64
升/降              float64
学校名称            object
英文名称            object
省市                object
类型                object
总分              float64
办学层次            float64
学科水平            float64
办学资源            float64
师资规模与结构      float64
人才培养            float64
科学研究            float64
服务社会            float64
学术人才            float64
重大项目与成果      float64
国际竞争力          float64
dtype: object
```

查看是否有缺失值：df.isnull().sum()

排名	0
升/降	16
学校名称	0
英文名称	0
省市	0
类型	0
总分	0
办学层次	482
学科水平	482
办学资源	482
师资规模与结构	482
人才培养	482
科学研究	482
服务社会	482
学术人才	482
重大项目与成果	482
国际竞争力	482

3 数据分析内容

1、中国大学总分排行榜及院校（前 100 名）:

In [9]: runfile('C:/Users/赤井秀亿/.spyder-py3/temp.py', wdir='C:/Users/赤井秀亿/.spyder-py3')

排名	学校名称	总分	类型
1	清华大学	852.5	综合
2	北京大学	746.7	综合
3	浙江大学	649.2	综合
4	上海交通大学	625.9	综合
5	南京大学	566.1	综合
6	复旦大学	556.7	综合
7	中国科学技术大学	526.4	理工
8	华中科技大学	497.7	综合
9	武汉大学	488	综合
10	中山大学	457.2	综合
11	西安交通大学	452.5	综合
12	哈尔滨工业大学	450.2	理工
13	北京航空航天大学	445.1	理工
14	北京师范大学	440.9	师范
15	同济大学	439	理工
16	四川大学	435.7	综合
17	东南大学	432.7	综合
18	中国人民大学	409.7	综合
19	南开大学	402.1	综合
20	北京理工大学	395.6	理工
21	天津大学	390.3	理工
22	山东大学	387.9	综合
23	厦门大学	383.3	综合
24	吉林大学	379.5	综合
25	华南理工大学	379.4	理工
26	中南大学	378.6	综合
27	大连理工大学	365.1	理工

28	西北工业大学	359.6	理工
29	华东师范大学	358	师范
30	中国农业大学	351.5	农业
31	湖南大学	348.3	综合
32	电子科技大学	334.8	理工
33	北京科技大学	321.8	理工
34	重庆大学	320.9	综合
35	南京航空航天大学	319.5	理工
36	南京理工大学	317.1	理工
37	东北大学	314	理工
38	苏州大学	306.8	综合
39	华中农业大学	300.5	农业
40	兰州大学	300.4	综合
41	西安电子科技大学	299.4	理工
42	华东理工大学	294.6	理工
43	北京交通大学	293.6	理工
44	华中师范大学	292.4	师范
45	上海大学	290	综合
46	南方科技大学	289	综合
47	南京农业大学	283.2	农业
48	暨南大学	282.6	综合
49	中国海洋大学	281.2	综合
50	南京师范大学	279.9	师范
51	哈尔滨工程大学	279.4	理工
52	武汉理工大学	277.6	理工
53	西南交通大学	277.3	理工
54	中国地质大学（武汉）	274.9	理工
55	北京邮电大学	273.7	理工

56	东北师范大学	272.5	师范
57	中国石油大学（北京）	272.4	理工
58	河海大学	271.8	理工
59	江南大学	271.5	综合
60	中国矿业大学	264.9	理工
61	西北大学	262.4	综合
62	上海科技大学	261.5	综合
63	北京工业大学	260.5	理工
64	北京化工大学	259.7	理工
65	深圳大学	259.1	综合
66	西南大学	257.5	综合
67	华南师范大学	254.4	师范
68	西北农林科技大学	251.1	农业
69	扬州大学	250.5	综合
70	中国地质大学（北京）	248.8	理工
71	中国矿业大学（北京）	247	理工
72	江苏大学	245.9	综合
73	东华大学	245.2	理工
74	中国石油大学（华东）	244.8	理工
75	宁波大学	244.3	综合
76	福州大学	243.8	理工
77	北京林业大学	242.3	林业
78	合肥工业大学	240.5	理工
79	陕西师范大学	239.9	师范
80	华北电力大学	235.7	理工
81	浙江工业大学	235.5	理工
82	南京工业大学	235.4	理工
83	郑州大学	234.9	综合
84	云南大学	234.1	综合
85	湖南师范大学	233.4	师范
85	首都师范大学	233.4	师范
87	南昌大学	228.4	综合
88	华南农业大学	227.1	农业
89	上海理工大学	221.4	理工
90	安徽大学	219.2	综合
91	长安大学	218.9	理工
92	燕山大学	216.7	理工
93	南京信息工程大学	216.6	理工
94	湖北大学	214.5	综合
95	广东工业大学	214.2	理工
96	南京邮电大学	213.9	综合
97	河南大学	212.9	综合
98	上海师范大学	212.8	师范
99	杭州电子科技大学	211.6	理工
100	广州大学	211.1	综合

2、学校类型比例



3、排名下降的学校（前五十名）

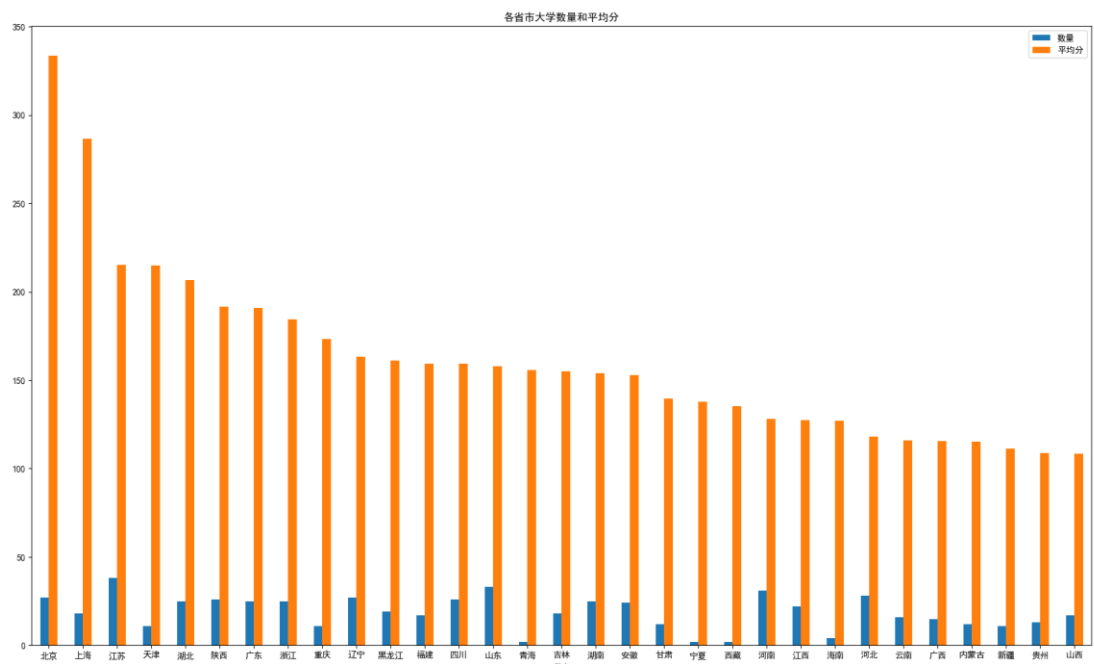
	排名	升/降	学校名称	...	学术人才	重大项目与成果	国际竞争力
11	12	-2.0	中山大学	...	37.8	31.4	33.3
14	15	-2.0	北京航空航天大学	...	26.7	54.7	23.8
15	16	-1.0	同济大学	...	29.1	35.0	22.1
19	20	-1.0	南开大学	...	29.1	27.3	24.9
21	22	-1.0	天津大学	...	22.7	23.4	21.8
25	26	-3.0	厦门大学	...	25.3	23.1	23.1
26	27	-2.0	华南理工大学	...	18.5	21.9	26.3
27	28	-1.0	大连理工大学	...	16.4	32.2	16.8
31	32	-1.0	湖南大学	...	16.8	22.6	26.4
36	37	-1.0	南京理工大学	...	8.3	18.9	16.2
41	42	-3.0	华中农业大学	...	14.5	17.1	16.0
43	44	-6.0	苏州大学	...	14.4	16.8	27.8
44	45	-3.0	华东理工大学	...	13.1	13.6	16.0
45	46	-2.0	华中师范大学	...	11.7	12.8	19.3

4、各省市大学数量和平均分（图中部分）

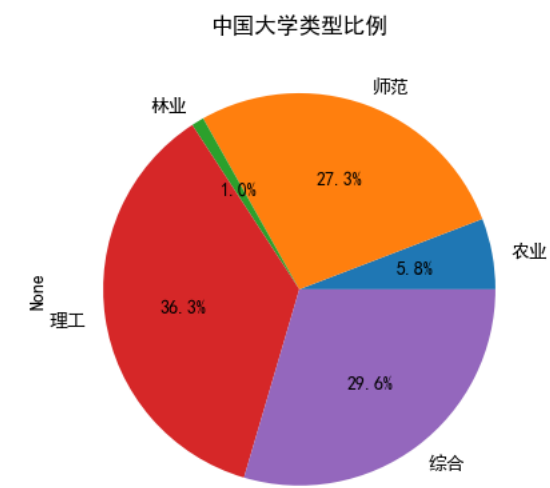
Index	数量	平均分
北京	27	333.61
上海	18	286.61
江苏	38	215.17
天津	11	214.79
湖北	25	206.57
陕西	26	191.44
广东	25	190.71
浙江	25	184.38
重庆	11	173.46
辽宁	27	163.2
黑龙江	19	161.19
福建	17	159.42
四川	26	159.22
山东	33	157.97
青海	2	155.85
吉林	18	155.02
湖南	25	153.84
安徽	24	152.82
甘肃	12	139.54
宁夏	2	137.7
西藏	2	135.4
河南	31	127.99
江西	22	127.24
海南	4	127.12
河北	28	118.24
云南	16	115.86
广西	15	115.59

4 数据分析图表

1、各省市大学数量和平均分柱状图

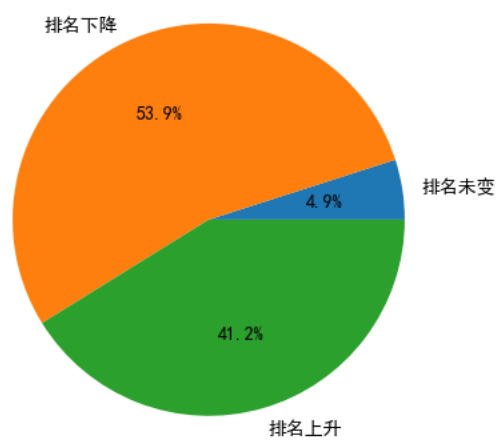


2、大学类型比例饼状图



3、中国大学变化升降情况

中国大学排名变化情况



5 数据分析结果

(1) 高校发展对于报考的参考数据

从分析得出的结果可以更清晰的看出中国大学的综合分数的排名,还有中国各大高校的排名的变化情况:清华大学蝉联榜单第一名,清华,北大,浙大等九所高校也稳居自己的排名没有变化;同济南开等大学有小幅度的下降,常州大学、川农等学校排名大幅度下降;西交哈工大等大学有小幅度的上升,郑州大学等有大幅度的上升。从排名的升降变化比较有参考性,可以让考生们在报考时了解该校最近的发展情况。

(2) 高校类型对于报考的参考

从分析出来的结果可以清晰的看出每个学校的类型分类比例,理工类的大学所占的比重最多,综合和师范的其次,还有一些如农业,林业等类型的学校,数量就比较少了。可以提供给自己心里有报考方向学生一个参考,也可以给还没有什么方向的同学参考可以报考哪些类型的学校。

(3) 高校所在地区对于报考的参考

从分析出来的结果可以清晰的看出每个地区有高校的数量,以及所有高校的综合评分的平均分,更易于对自己以后想去哪个城市发展的学生有一个参考,也可以对不清楚自己去哪个城市发展的学生提供应该去哪个城市的参考

6 总结

在这个大数据的年代，也是一个高考人数越来越多的年代，高校的建设越来越好，高校的发展也越来越好，大学同时也有各种各样不同的学科。为了让更多的高中面临对院校的选择和未来学习专业方向选择有更多可参考的数据，就总结了所有的中国大学排行榜前几百位的大学进行数据分析和数据可视化。在将数据整理在一起后，并将其化作更易理解的可视化的饼图和柱状图后，给了同学们更容易看出如何对院校进行筛选等。在一个个数据可视化的图上，可以比散落的数据信息更容易看出自己对院校实力的选择，院校类型的选择和院校地区分落的选择。这就是数据整理和数据可视化的优势，这也是这个时代科技带给我们的好处和优势。在这个信息化的时代，我们就应该利用一些数据类的思维，将数据拼凑在一起，更方便于我们平时对零星数据的处理，也节约了更多我们日常人为去整理这些琐碎的数据的事件。这是人类的进步，也是时代的进步，我们就应该顺应时代的变化，好好利用数据带给我们生活的便利。

附录-数据分析代码

```
g = df.groupby('省市')
df_counts = g.count()['排名']
df0 = df_counts.copy()
df0.sort_values(ascending=False, inplace=True)
df_means0 = g.mean()['总分']
df_means = df_means0.round(2)
df1 = pd.concat([df_counts, df_means], join='outer', axis=1)
df1.columns = ['数量', '平均分']
df1.sort_values(by=['平均分'], ascending=False, inplace=True)
df1.plot(kind='bar', rot=0, label=('各省市大学数量', '各省市大学平均分'), title='各省市大学数量
和平均分')
```

```
dp1=df.groupby('类型').size()
dp1.plot(kind='pie',title='中国大学类型比例',autopct='%.1f%%')
```

```
dp2=df[df['升/降']==0]
a=dp2.iloc[:,2].count()
dp3=df.loc[df['升/降']<0]
b=dp3.iloc[:,2].count()
dp4=df.loc[(df['升/降']>0)]
c=dp4.iloc[:,2].count()
rate=[a,b,c]
labels=['排名未变','排名下降','排名上升']
plt.pie(rate,labels=labels,autopct='%.1f%%')
plt.title('中国大学排名变化情况')
```

爬虫代码（查询到的）

```
import json
import requests
```

```
def getHTMLText(url):
    """从网络上获取大学排名网页内容"""
    try:
        r = requests.get(url, timeout=40)
        ##如果状态不是 200, 就会引发 HTTPError 异常
        r.raise_for_status()
        r.encoding = r.apparent_encoding
        return r.text
    except:
        return ""
```



```

def printUnivList(ulist, html,num):
    """提取 html 网页内容中 前 num 名大学信息到 ulist 列表中 """
    data = json.loads(html) #对数据进行解码
    # 提取 数据 rankings 包含的内容
    content = data['data']['rankings']

    # 把 学校的相关信息放到 ulist 里面
    for i in range(num):
        index = content[i]['rankOverall']
        name = content[i]['univNameCn']
        score = content[i]['score']
        category = content[i]['univCategory']
        ulist.append([index, name, score,category])

    # 打印前 num 名的大学
    tplt = "{0:^10}\t{1:{3}^10}\t{2:^10}\t{4:^10}" # {1:{3}^10} 中的 {3} 代表取第三个
参数
    print(tplt.format("排名 ", "学校名称", "总分", chr(12288), "类型")) # chr(12288) 代表中文空
格
    for i in range(num):
        u=ulist[i]
        print(tplt.format(u[0],u[1],u[2],chr(12288),u[3])) # chr(12288) 代表中文空格

def main():
    uinfo = []
    url = 'https://www.shanghairanking.cn/api/pub/v1/bcur?bcur_type=11&year=2020'
    html = getHTMLText(url)# 获取大学排名内容
    printUnivList(uinfo, html, 100)#输出 排名前 100 的大学内容

main()

```