

2000 年至今的全球空难事故数据分析

在网站“[Database index \(planecrashinfo.com\)](#)”中，我们可以找到从 1920 年至 2024 年的全球空难事故数据统计，其中包括 Date, Time, Airlin, Route 等航班信息。



网站截图 1

2012			
Date	Location / Operator	Aircraft Type / Registration	Fatalities
30 Jan 2012	Namoya, Congo TRACEP Congo Aviation	Antonov AN-28 9Q-CUN	3/5(0)
15 Mar 2012	Near San Juan, Puerto Rico Jet One Express	Convair CV-340 N153JR	2/2(0)
16 Mar 2012	Kabul, Afghanistan Military - Turkish Army (NATO)	Sukorsky S-70A-28 10981	12/12(4)
02 Apr 2012	Tyumen, Russia UTAir	ATR 72-201 VP-BYZ	33/43(0)
20 Apr 2012	Islamabad, Pakistan Bhoja Airlines	Boeing 737-236 AP-BRC	127/127(0)
30 Apr 2012	Sabanagrande, Colombia Military - Fuerza Aerea Colombiana	Boeing 212 FAC 4020	13/13(0)
09 May 2012	Near Cidahu, Indonesia Sukhoi	Sukhoi Superjet 100-95 RA-97004	45/45(0)
14 May 2012	Jomsom, Nepal Agni Air	Dornier 228-212 9N-AIG	15/21(0)
02 Jun 2012	Accra, Ghana Allied Air	Boeing B-727-200 5N-BJN	0/4(12)
03 Jun 2012	Lagos, Nigeria Dana Air	McDonnell Douglas MD-83 5N-RAM	153/153(10)
06 Jun 2012	Huaila Huaila, Mount Mamarosa, Peru HeliCusco	Sikorsky S-58ET OB-1840-P	14/14(0)
11 Jun 2012	Near Sierov, Russia Illegal flight	Antonov An-2R RA-40312	13/13(0)
21 Jun 2012	Jakarta, Indonesia Military - Indonesian Air Force	Fokker F-27 Friendship 400M A-2708	7/7(4)
18 Aug 2012	Off Masbate Island, Philippines Aviation's Fly'n Inc.	PA-34-200 Seneca I RP-C4431	3/4(0)
19 Aug 2012	Talodi, Sudan Alfa Airlines	Antonov An-26-100 ST-ARL	32/32(0)
12 Sep 2012	Near Palana, Russia	Antonov An-28	10/14(0)

网站截图 2

ACCIDENT DETAILS	
Date:	January 30, 2012
Time:	0845
Location:	Namoya, Congo
Operator:	TRACEP Congo Aviation
Flight #:	?
Route:	Bukavu - Namoya
AC Type:	Antonov AN-28
Registration:	9Q-CUN
cn / ln:	1AJ006-11
Aboard:	5 (passengers:3 crew:3)
Fatalities:	3 (passengers:1 crew:2)
Ground:	0
Summary:	The cargo plane was on approach to land when it crashed in a forest 10km from Namoya. The airworthiness of this aircraft had expired on June 12, 1993.

网站截图 3

从中挑选出具有分析价值的几项，利用如下代码对其进行爬虫获取。

```
import requests
from bs4 import BeautifulSoup
import csv

for year in range(2000,2025):
    response = requests.get(f'https://www.planecrashinfo.com/{year}/{year}.htm')
    trs = BeautifulSoup(response.text, features='html.parser').find_all('tr')
    for tr in trs:
        url_content = tr.find('td').find('a')
        if url_content != None:
            url = (f'https://www.planecrashinfo.com/{year}/{url_content.get("href")}')
            response_detail = requests.get(url)
            trs_detail = BeautifulSoup(response_detail.text, features='html.parser').find_all('tr')
            result = []
            for tr_detail in trs_detail:
                row_title = (tr_detail.find('td').text.replace('\n',' '))
                if ('Date' in row_title) | ('Operator' in row_title) | ('Type' in row_title) | ('Fatalities' in row_title) | ('Summary' in row_title):
                    result.append(tr_detail.find_all('td')[-1].text.replace('\n',' '))
            print(result + [url])
    with open(f'plane.csv', encoding='utf-8-sig', mode='a', newline='') as f:
        csv_writer = csv.writer(f)
        csv_writer.writerow(result + [url])
```

“spider.py” 截图

得到的结果储存在 “plane.csv” 中，获取的数据包括 Date, Time, Location, Operator, Route, AC type, Aboard, Fatalities, Summary, url

DATE	Time	Location	Operator	Route	AC Type	Aboard	Fatalities	Summary	url
5-Jan-00	?	Abuja, Ni	Skypower	Lagos - A	Embraer 116	(pa:1	(pas:1	(pas:During a	https://www.planecrashinfo.com/2000/2000-1.htm
10-Jan-00		1754 Niederhas	Crossair	Zürich - I	Saab 340B 10	(pa:10	(pa:10	(pa:The aircr	https://www.planecrashinfo.com/2000/2000-2.htm
13-Jan-00		1238 Off Marsa	Avisto	Tripoli I	Shorts 3641	(pa:22	(pa:22	(pa:The aircr	https://www.planecrashinfo.com/2000/2000-3.htm
15-Jan-00		1310 San Jose, TACA		San Jose	Let 410UV120	(pa:5	(pas:5	(pas:The aircr	https://www.planecrashinfo.com/2000/2000-4.htm
30-Jan-00		2108 Off Abidj	Kenya Air	Abidjan - A	irbusA311179	(p:169	(p:169	(p:The aircr	https://www.planecrashinfo.com/2000/2000-5.htm
31-Jan-00		1621 Off Point	Alaska Air	Puerto Va	McDonnell 88	(pa:88	(pa:88	(pa:The aircr	https://www.planecrashinfo.com/2000/2000-6.htm
2-Feb-00	?	Tehran, I	Military Training	Lockheed	8	(pas:8	(pas:8	(pas:Crashed o	https://www.planecrashinfo.com/2000/2000-7.htm
16-Feb-00		1950 Sacramento	Emery Wor	Sacramento	McDonnell 3	(pas:3	(pas:3	(pas:The cargo	https://www.planecrashinfo.com/2000/2000-8.htm
5-Mar-00		1811 Burbank, R	Southwest	Las Vegas	Boeing B-142	(p:0	(pas:0	(pas:While att	https://www.planecrashinfo.com/2000/2000-9.htm
9-Mar-00		840 Moscow, R	Vologodsk	Moscow - I	Yakovlev 9	(pas:9	(pas:9	(pas:The aircr	https://www.planecrashinfo.com/2000/2000-10.htm
17-Mar-00		930 Near Kuna	Aeroperla	Panama Ci	de Havill:10	(pa:10	(pa:10	(pa:Crashed i	https://www.planecrashinfo.com/2000/2000-11.htm
22-Mar-00		1231 Ennadai	L.Points No	Points No	Douglas C 2	(pas:2	(pas:2	(pas:After lif	https://www.planecrashinfo.com/2000/2000-12.htm
22-Mar-00		1745 Herreira, Military		Sevilla - C	AESA 212-17	(pas:7	(pas:7	(pas:Crashed w	https://www.planecrashinfo.com/2000/2000-13.htm
24-Mar-00		1804 Kadirana, Sky	Cabs	Bangkok - A	Antonov A/8	(pas:6	(pas:6	(pas:The pilot	https://www.planecrashinfo.com/2000/2000-14.htm
25-Mar-00	?	Huambo, A	Uralex	Huambo - A	Antonov A/33	(pa:3	(pas:3	(pas:The aircr	https://www.planecrashinfo.com/2000/2000-15.htm
30-Mar-00	?	Near Anur	Military	Jaffna/Pa	Antonov A/40	(pa:40	(pa:40	(pa:The crew	https://www.planecrashinfo.com/2000/2000-16.htm
5-Apr-00		930 Marianna, Bankair		Opa-Locka	Learjet 3.3	(pas:3	(pas:3	(pas:While on	https://www.planecrashinfo.com/2000/2000-17.htm
19-Apr-00		702 Samad Isl	Air Phili	Manila - B	Boeing B-131	(p:131	(p:131	(p:The aircr	https://www.planecrashinfo.com/2000/2000-18.htm
19-Apr-00	?	Pepa, Con	Military	Pepa, Con	Antonov A/24	(pa:24	(pa:24	(pa:The milit	https://www.planecrashinfo.com/2000/2000-19.htm
29-Apr-00	?	Bapi, Papi	North Coa	?	Britten-N/4	(pas:4	(pas:4	(pas:While tak	https://www.planecrashinfo.com/2000/2000-20.htm
2-May-00		1439 Lyon Sato	Northern	?	Learjet 3.5	(pas:2	(pas:2	(pas:During a	https://www.planecrashinfo.com/2000/2000-21.htm
3-May-00	?	Kurupung, Trans	Guy?		Britten-N/1	(pas:1	(pas:1	(pas:The cargo	https://www.planecrashinfo.com/2000/2000-22.htm
10-May-00		2030 Kaunakaka	Price Air	Kahului - R	ockwell 6	(pas:6	(pas:6	(pas:While on	https://www.planecrashinfo.com/2000/2000-23.htm
17-May-00		947 Moanda, G	Avirex	Librevill	Beechcraf 10	(pa:3	(pas:3	(pas:The chart	https://www.planecrashinfo.com/2000/2000-24.htm
21-May-00		1128 Near Wani	Executive	Atlantic	BAe Jetst:19	(pa:19	(pa:19	(pa:The plane	https://www.planecrashinfo.com/2000/2000-25.htm
25-May-00	?	Near Mani	Philippin	Davao - M	Airbus A:298	(p:1	(pas:1	(pas:A hijack	https://www.planecrashinfo.com/2000/2000-26.htm
31-May-00		1905 Off Whyal	Whyalla A	Adelaide	Piper PA-8	(pas:8	(pas:8	(pas:En route	https://www.planecrashinfo.com/2000/2000-27.htm
5-Jun-00		1135 Accra, Gh	Ghana Air	Tamale - F	okker F-52	(pa:7	(pas:7	(pas:The plane	https://www.planecrashinfo.com/2000/2000-28.htm
29-Jun-00		116 Chis	SW	Chis	SW	144	(pas:44	(pas:The plane	https://www.planecrashinfo.com/2000/2000-29.htm

” plane.csv” 截图

在表格中，Time 表示事故发生的当地时间，但由于部分事故的时间不明，以及 csv 表格的格式性，需要对数据进行二次处理以便后续的分析。利用

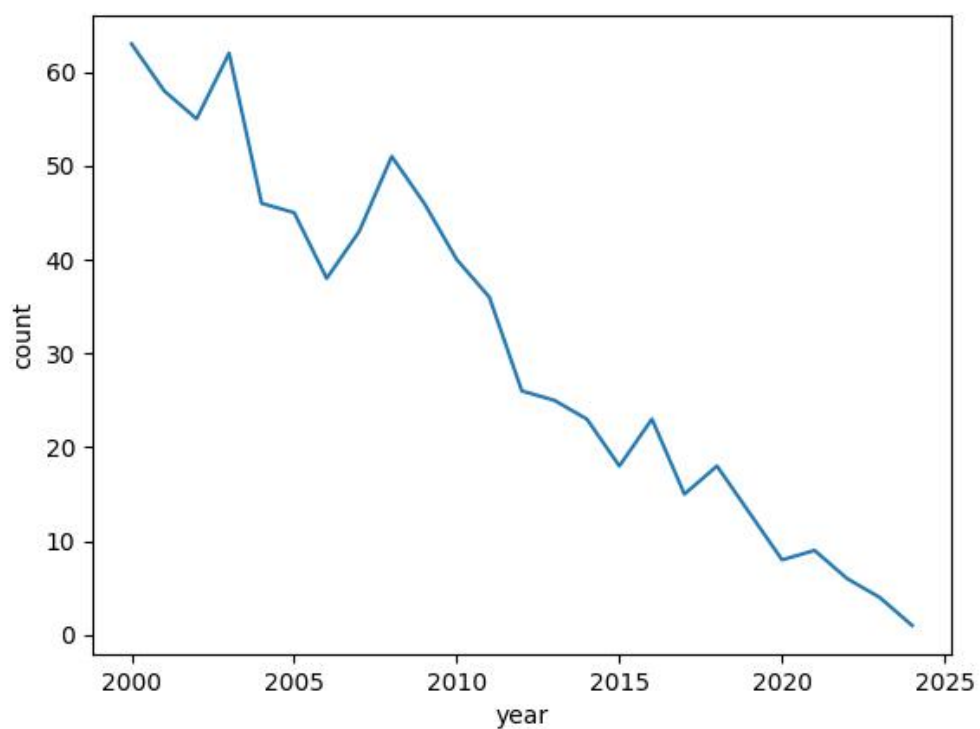
“timechange.py” 对 Time 的数据进行处理，将其全部转化为四位数字，不够则在前方补 0，忽略 “？”，并将结果储存在 “newtime.txt” 文档中（由于网站的格式不统一，部分数据还使用了 “：”，对该部分数据则是在 txt 文档中进行了手动删除）之后再只用 “timecount.py” 对前两位数字表示的时间进行统计。而对于 Date 中的年份和月份则统一在 csv 文档中进行处理。

综合上面三组元素，我们得到有关时间的数据，并将其储存在“newdata.csv”文档中

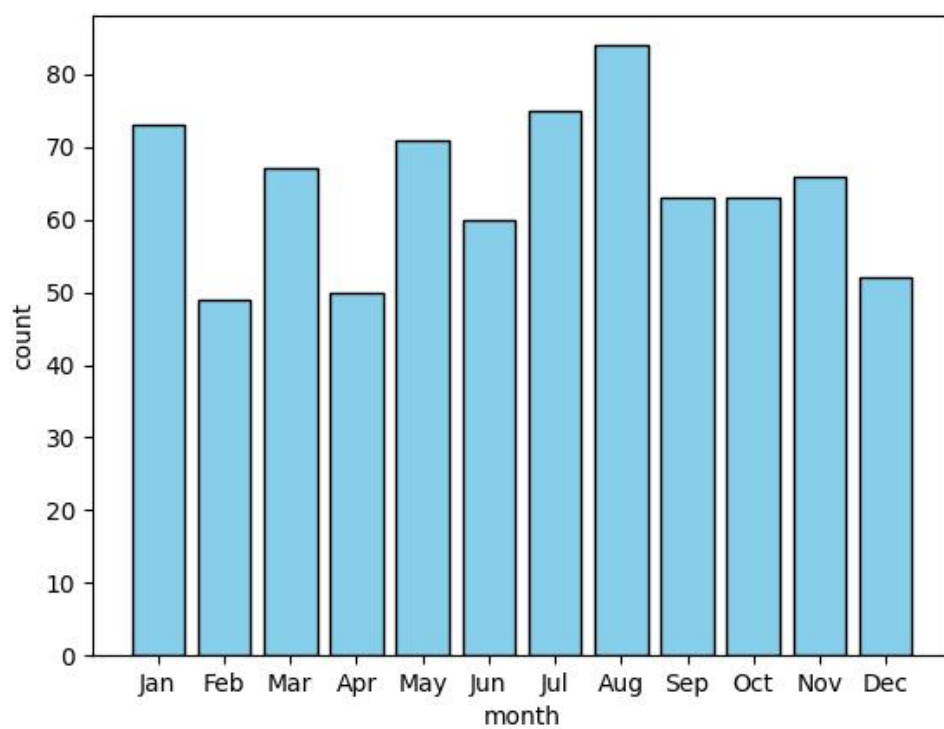
year	count	month	count	hour	count
2000	63	Jan	73	0	12
2001	58	Feb	49	1	15
2002	55	Mar	67	2	14
2003	62	Apr	50	3	9
2004	46	May	71	4	14
2005	45	Jun	60	5	7
2006	38	Jul	75	6	26
2007	43	Aug	84	7	40
2008	51	Sep	63	8	48
2009	46	Oct	63	9	47
2010	40	Nov	66	10	44
2011	36	Dec	52	11	48
2012	26			12	42
2013	25			13	35
2014	23			14	46
2015	18			15	50
2016	23			16	43
2017	15			17	28
2018	18			18	34
2019	13			19	32
2020	8			20	18
2021	9			21	10
2022	6			22	17
2023	4			23	15
2024	1				

“newdata.csv” 截图

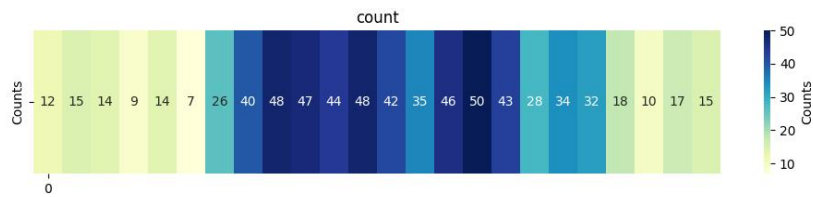
利用与时间相关的数据，我们可以得出历年来所发生的空难事故的数量变化，各个月份数量分布以及一天中空难数量最多的时段，利用“time.py”进行画图。



空难数量随年份变化的折线图



空难数量与月份关系的直方图



空难数量与一天中时刻关系的热力图

由上述三图可以看出，随着时间的推移，空难数量在大体上呈下降趋势，这与飞机制造商的技术进步有关，与飞行员的素质提升有关，也与全球各国对航空相关制度的完善有关。而在一年之中，空难数量的分布较为平均，个别突出的月份可能与假期来临，全球出行人数上升有关。在一天之中，空难数量主要集中在午间至傍晚的时段中，这也与此时段航班数量居多有关。

之后再对“Location”中所描述的空难的发生地点进行统计，这里使用“place.py”进行词频统计，得到“place.csv”

Keyword	Frequency
near	212
russia	50
indonesia	36
congo	35
new	29
brazil	25
colombia	23
canada	21
iran	20
alaska	19
sudan	18
mexico	17
nepal	17
republic	17
venezuela	16
south	16
afghanistan	15
island	14
democratic	13
york	13
philippines	12
nigeria	11
papua	10
guinea	10
pakistan	10
iraq	10
san	9
california	9

“place.csv” 截图

事故多发机型词云图

其中居多的为“Boeing”和“Antonov”的飞机，但这并不能完全表明两家公司的飞机质量较低，还需综合考量二者在飞机制造产业中居于领头地位，出产数量极大的因素。

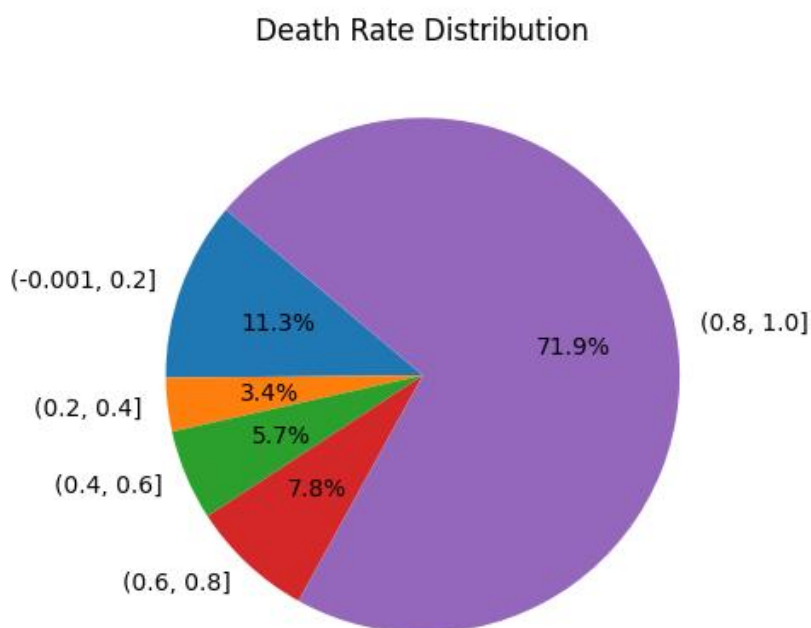
同理，对 Summary 中所描述的事故原因进行词频统计，并绘制词云图



事故原因词云图

可以看出，天气、落地、山脉等成为了空难发生的主要原因

最后，对于 Aboard, Fatalities 中描述的上机人数和死亡人数，使用“deathrate.py”对死亡率进行了计算，并绘制了饼图



死亡率分布饼图

可以看出空难事故发生后的死亡率普遍较高，在“deathrate.csv”中还能更清晰的看出，处于 0.8-1.0 的也大都为接近于 100%的死亡率。

通过对 2000 年以来的空难事故数据的分析，能够看出空难与各个相关内容间的关系，比如事故多发地、事故发生原因等等，有利于人们对于空难的全貌进行了解，通过对更广大数据的观测与分析，并及时做出应对，也能有利于航空事业向更安全的方向发展。