In [ ]:
```
# DOWNLOAD https://www.kaggle.com/datasets/andrewsundberg/college-basketball-da
import polars as pl
import numpy as np
import matplotlib.pyplot as plt
```

In [ ]:
```
# # unzip archive.zip and put them into a folder called cbb
# !unzip archive.zip -d cbb
# # erase the zip file
# !rm archive.zip -y
```

In [ ]:
```
cbb = pl.read_csv('cbb/cbb.csv')
cbb.head()
```

Out[ ]: shape: (5, 24)

TEAM	CONF	G	W	ADJOE	ADJDE	BARTHAG	EFG_O	EFG_D	TOR	TORD	ORB	DR
str	str	i64	i64	f64	f64	f64	f64	f64	f64	f64	f64	f6
"North Carolina...	"ACC"	40	33	123.3	94.9	0.9531	52.6	48.1	15.4	18.2	40.7	30
"Wisconsin"	"B10"	40	36	129.1	93.6	0.9758	54.8	47.7	12.4	15.8	32.1	23
"Michigan"	"B10"	40	33	114.4	90.4	0.9375	53.9	47.7	14.0	19.5	25.5	24
"Texas Tech"	"B12"	38	31	115.2	85.2	0.9696	53.5	43.0	17.7	22.8	27.4	28
"Gonzaga"	"WCC"	39	37	117.8	86.3	0.9728	56.6	41.1	16.2	17.1	30.0	26

In [ ]:
```
cbb.describe()
```

Out[ ]: shape: (7, 25)

describe	TEAM	CONF	G	W	ADJOE	ADJDE	BARTHAG
str	str	str	f64	f64	f64	f64	f64
"count"	"2455"	"2455"	2455.0	2455.0	2455.0	2455.0	2455.0
"null_count"	"0"	"0"	0.0	0.0	0.0	0.0	0.0
"mean"	null	null	31.492464	16.284318	103.304481	103.304603	0.493957	49
"std"	null	null	2.657401	6.61096	7.376981	6.605318	0.256244	3
"min"	"Abilene Christ...	"A10"	15.0	0.0	76.6	84.0	0.005
"max"	"Youngstown St....	"ind"	40.0	38.0	129.1	124.0	0.9842
"median"	null	null	31.0	16.0	103.0	103.5	0.475

In [ ]:
```

```

In [ ]:
```
# compare the two
cbb_champions.describe()
```

```
cbb_2nd.describe()
```

Out[ ]: shape: (7, 25)

describe	TEAM	CONF	G	W	ADJOE	ADJDE	BARTHAG	EFG
str	str	str	f64	f64	f64	f64	f64	
"count"	"7"	"7"	7.0	7.0	7.0	7.0	7.0	
"null_count"	"0"	"0"	0.0	0.0	0.0	0.0	0.0	
"mean"	null	null	39.285714	32.714286	119.785714	91.471429	0.952457	5
"std"	null	null	0.95119	2.984085	5.207504	4.297951	0.024551	2.088
"min"	"Gonzaga"	"ACC"	38.0	29.0	114.4	85.2	0.9062	4
"max"	"Wisconsin"	"WCC"	40.0	37.0	129.1	96.2	0.9758	5
"median"	null	null	40.0	33.0	117.8	93.6	0.9531	5

In [ ]:
```
# who has qualified the most times, qualifiying means they're not in NA

cbb.filter(pl.col('POSTSEASON') != 'NA').select(['TEAM', 'YEAR']).sort('YEAR',
```

Out[ ]: shape: (188, 2)

TEAM	count
str	u32
"Gonzaga"	7
"Cincinnati"	7
"Duke"	7
"Villanova"	7
"Kansas"	7
"North Carolina...	7
"Michigan St."	7
"Kentucky"	6
"VCU"	6
"Wichita St."	6
"Wisconsin"	6
"Arizona"	6
...	...
"North Dakota"	1
"Stanford"	1
"UNC Greensboro...	1
"Nebraska"	1
"American"	1
"North Florida"	1
"Massachusetts"	1
"Marshall"	1
"Holy Cross"	1
"Western Kentuc...	1
"Lipscomb"	1
"Georgia"	1

In [ ]:
```
# who has made it to the sweet 16 the most?
s16 = cbb.filter(pl.col('POSTSEASON') == 'S16').select(['TEAM', 'YEAR']).sort(
s16.sort('count', descending=True)
```

Out[ ]: shape: (39, 2)

TEAM	count
str	u32
"West Virginia"	3
"UCLA"	3
"Michigan"	2
"Indiana"	2
"Gonzaga"	2
"Texas A&M"	2
"Miami FL"	2
"Oregon"	2
"Arizona"	2
"Tennessee"	2
"Baylor"	2
"Wisconsin"	2
…	…
"Houston"	1
"Duke"	1
"Kentucky"	1
"North Carolina…	1
"Louisville"	1
"Oklahoma"	1
"Maryland"	1
"Florida St."	1
"Florida Gulf C…	1
"LSU"	1
"Nevada"	1
"Virginia Tech"	1

In [ ]:
```
cbb_2nd.select(['TEAM', 'YEAR']).sort('YEAR', descending=True).groupby("TEAM").
```

Out[ ]: shape: (6, 2)

TEAM	count
str	u32
"Gonzaga"	1
"Kentucky"	1
"North Carolina...	1
"Texas Tech"	1
"Michigan"	2
"Wisconsin"	1

In [ ]:
```
# who's been the most champions?

cbb_champions.select(['TEAM', 'YEAR']).sort('YEAR', descending=True).groupby("T
```

Out[ ]: shape: (6, 2)

TEAM	count
str	u32
"Louisville"	1
"Villanova"	2
"Connecticut"	1
"Virginia"	1
"North Carolina...	1
"Duke"	1

In [ ]:
```
# what kind of postseason information do we have?

cbb.groupby('POSTSEASON').count()
```

Out[ ]: shape: (9, 2)

POSTSEASON	count
str	u32
"R32"	112
"2ND"	7
"R64"	224
"R68"	28
"E8"	28
"Champions"	7
"NA"	1979
"S16"	56
"F4"	14

In [ ]:
```
# what about the Georgia teams like uGA (Georgia), Georgia Tech, Georgia State,

ga = cbb.filter(pl.col('TEAM').str.contains('Georgia'))
```

In [ ]:
```
ga
```

Out[ ]: shape: (28, 24)

TEAM	CONF	G	W	ADJOE	ADJDE	BARTHAG	EFG_O	EFG_D	TOR	TORD	ORB	DRB
str	str	i64	i64	f64	f64	f64	f64	f64	f64	f64	f64	f64
"Georgia Tech"	"ACC"	31	12	102.1	92.9	0.7463	44.4	49.2	19.6	18.0	38.7	25.8
"Georgia Tech"	"ACC"	36	21	113.9	98.2	0.845	49.8	49.1	16.7	14.6	37.5	27.2
"Georgia Tech"	"ACC"	36	21	99.4	90.6	0.743	47.4	46.0	19.3	18.8	27.7	30.8
"Georgia Tech"	"ACC"	32	13	103.4	98.5	0.6385	46.9	49.2	19.2	18.9	30.8	30.6
"Georgia Tech"	"ACC"	32	14	99.9	95.7	0.6205	49.2	45.7	21.0	18.7	25.3	31.6
"Georgia Southe...	"SB"	28	19	98.2	97.5	0.5203	47.3	46.2	18.3	22.0	32.6	32.2
"Georgia Southe...	"SB"	29	12	103.9	108.0	0.3913	47.5	51.8	16.2	20.0	29.7	34.5
"Georgia St."	"SB"	28	14	99.7	99.8	0.4947	48.5	47.9	18.0	21.2	24.5	31.0
"Georgia Southe...	"SB"	31	18	104.4	106.5	0.4424	51.5	51.3	17.0	18.7	25.2	33.0
"Georgia St."	"SB"	31	20	103.0	99.8	0.5883	53.0	49.0	20.0	19.8	29.9	29.1
"Georgia Southe...	"SB"	31	21	104.8	104.0	0.5225	49.2	49.7	16.1	20.3	30.3	31.9
"Georgia Southe...	"SB"	31	21	107.3	100.3	0.6826	54.3	51.4	18.6	20.0	27.4	28.7
...	...	...	...	...	...	...	...	...	...	...	...	...
"Georgia"	"SEC"	34	20	107.9	99.0	0.7275	48.9	45.9	20.0	16.0	35.3	29.8
"Georgia Southe...	"SC"	31	12	97.9	110.1	0.2072	48.8	51.8	17.9	18.1	30.3	30.3
"Georgia St."	"SB"	32	23	114.0	104.3	0.736	53.1	49.6	11.9	20.1	24.9	34.1
"Georgia Tech"	"ACC"	33	16	105.7	97.9	0.7082	47.8	47.4	18.4	15.4	32.7	28.5
"Georgia St."	"CAA"	30	14	103.6	104.6	0.4716	50.0	51.1	18.0	20.9	29.4	32.7
"Georgia"	"SEC"	32	15	100.9	94.4	0.6817	46.9	45.0	23.0	18.5	32.1	30.5
"Georgia Tech"	"ACC"	31	16	99.6	91.1	0.7369	46.8	45.4	19.3	19.0	32.6	28.3
"Georgia Southe...	"SC"	31	12	94.6	106.3	0.2071	45.9	49.7	20.1	20.5	30.9	34.4
"Georgia St."	"SB"	34	24	109.5	98.0	0.7802	52.0	45.8	16.3	23.2	30.0	34.8

TEAM	CONF	G	W	ADJOE	ADJDE	BARTHAG	EFG_O	EFG_D	TOR	TORD	ORB	DRB
str	str	i64	i64	f64	f64	f64	f64	f64	f64	f64	f64	f64
"Georgia St."	"SB"	33	24	108.8	101.5	0.6894	53.9	47.1	16.7	20.6	25.6	33.4
"Georgia St."	"SB"	33	24	106.4	101.1	0.6418	53.7	50.3	16.4	21.0	23.0	33.8
"Georgia"	"SEC"	33	21	107.6	92.6	0.8482	48.5	44.4	19.2	16.7	32.2	28.5

In [ ]:
```
good_ranking = ['S16', 'E8', 'F4', '2ND', 'Champions']
# do we have any teams that have been in any of these good rankings?

for r in good_ranking:
    print(r)
    print(ga.filter(pl.col("POSTSEASON") == r).groupby("TEAM").count())

#hope.
```

S16
shape: (0, 2)

TEAM	count
str	u32

E8
shape: (0, 2)

TEAM	count
str	u32

F4
shape: (0, 2)

TEAM	count
str	u32

2ND
shape: (0, 2)

TEAM	count
str	u32

Champions
shape: (0, 2)

TEAM	count
str	u32

In [ ]:
```
#nope, no hope for Georgia teams. what's the best we did?

ga.filter(pl.col("TEAM").str.contains("Georgia")).groupby("POSTSEASON").count()
```

Out[ ]: shape: (3, 2)

POSTSEASON	count
str	u32
"NA"	24
"R64"	3
"R32"	1

In [ ]:
```
# who did R32?
```

```
ga.filter(pl.col("POSTSEASON") == 'R32').groupby("TEAM").count()
```

Out[ ]: shape: (1, 2)

TEAM	count
str	u32
"Georgia St."	1

In [ ]:
```
#oh! good job panthers! what year was that?
ga.filter(pl.col("POSTSEASON") == 'R32').groupby("YEAR").count() # 2015? cool!
```

Out[ ]: shape: (1, 2)

YEAR	count
i64	u32
2015	1

In [ ]:
```
# what did gatech do?

ga.filter(pl.col("TEAM") == "Georgia").groupby("POSTSEASON").count()
```

Out[ ]: shape: (2, 2)

POSTSEASON	count
str	u32
"NA"	6
"R64"	1

In [ ]:
```

```