Pandas Challenge 4

November 14, 2022

Weekly Pandas Challenge 4 - Data In Motion

Name : Mohammed Wasim R D

Linkedin : www.linkedin.com/in/mdwasimrd

Dataset : https://raw.githubusercontent.com/justmarkham/DAT8/master/data/drinks.csv

Challenge Questions

1) See the first 10 entries.
2) Delete the columns 'Unnamed: 0' and 'Id'.
3) Group the dataset by name, assign to a variable called names, and sort the dataset by highest to lowest count.
4) How many different names exist in the dataset?
5) What is the name with most occurrences?
6) What is the standard deviation of count of names?
7) Get a summary of the dataset with the mean, min, max, std and quartiles.

[59]:
```
import pandas as pd
baby_names = pd.read_csv('https://raw.githubusercontent.com/guipsamora/
↪pandas_exercises/master/06_Stats/US_Baby_Names/US_Baby_Names_right.csv')
```

1)

[61]:
```
baby_names.head(10)
```

[61]:

	Unnamed: 0	Id	Name	Year	Gender	State	Count
0	11349	11350	Emma	2004	F	AK	62
1	11350	11351	Madison	2004	F	AK	48
2	11351	11352	Hannah	2004	F	AK	46
3	11352	11353	Grace	2004	F	AK	44
4	11353	11354	Emily	2004	F	AK	41
5	11354	11355	Abigail	2004	F	AK	37
6	11355	11356	Olivia	2004	F	AK	33
7	11356	11357	Isabella	2004	F	AK	30
8	11357	11358	Alyssa	2004	F	AK	29
9	11358	11359	Sophia	2004	F	AK	28

2)

[15]:
```
drop_columns=baby_names.drop(["Unnamed: 0","Id"], axis = 1)
drop_columns.head(10)
```

[15]:

	Name	Year	Gender	State	Count
0	Emma	2004	F	AK	62
1	Madison	2004	F	AK	48
2	Hannah	2004	F	AK	46
3	Grace	2004	F	AK	44
4	Emily	2004	F	AK	41
5	Abigail	2004	F	AK	37
6	Olivia	2004	F	AK	33
7	Isabella	2004	F	AK	30
8	Alyssa	2004	F	AK	29
9	Sophia	2004	F	AK	28

3)

[53]:
```
names = drop_columns.groupby("Name").sum()
names.sort_values("Count", ascending = False).head()
```

[53]:

Name	Year	Count
Jacob	1141099	242874
Emma	1137085	214852
Michael	1161152	214405
Ethan	1139091	209277
Isabella	1137090	204798

4)

[41]:
```
diff_names = baby_names.Name.value_counts().count()
diff_names
```

[41]:
```
17632
```

5)

[54]:
```
names.Count.idxmax()
```

[54]:
```
'Jacob'
```

6)

[55]:
```
names.Count.std()
```

[55]:
```
11006.069467891111
```

7)

[56]:
```
names.describe()
```

[56]:

	Year	Count
count	1.763200e+04	17632.000000
mean	1.158117e+05	2008.932169
std	2.451618e+05	11006.069468
min	2.004000e+03	5.000000
25%	4.017000e+03	11.000000
50%	1.606100e+04	49.000000
75%	7.846425e+04	337.000000
max	2.233993e+06	242874.000000

[57]:
```
del names['Year']
```

[58]:
```
names.describe()
```

[58]:

	Count
count	17632.000000
mean	2008.932169
std	11006.069468
min	5.000000
25%	11.000000
50%	49.000000
75%	337.000000
max	242874.000000

[ ]:
```

```