In [172…
```
# Import Library
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import statsmodels.api as sm
from scipy import stats
```

In [194…
```
aqi = pd.read_csv(r'C:\Users\HP\Desktop\Advance Data Analyst\4. The Power of Stats\3. Module 3\3. Work with sampling distribution\Files\c4_epa_air_quality.csv')
aqi.head(10)
```

Out[194…

	Unnamed: 0	date_local	state_name	county_name	city_name	local_site_name	parameter_name	units_of_measure	arithmetic_mean	aqi
0	0	2018-01-01	Arizona	Maricopa	Buckeye	BUCKEYE	Carbon monoxide	Parts per million	0.473684	7
1	1	2018-01-01	Ohio	Belmont	Shadyside	Shadyside	Carbon monoxide	Parts per million	0.263158	5
2	2	2018-01-01	Wyoming	Teton	Not in a city	Yellowstone National Park - Old Faithful Snow ...	Carbon monoxide	Parts per million	0.111111	2
3	3	2018-01-01	Pennsylvania	Philadelphia	Philadelphia	North East Waste (NEW)	Carbon monoxide	Parts per million	0.300000	3
4	4	2018-01-01	Iowa	Polk	Des Moines	CARPENTER	Carbon monoxide	Parts per million	0.215789	3
5	5	2018-01-01	Hawaii	Honolulu	Not in a city	Kapolei	Carbon monoxide	Parts per million	0.994737	14
6	6	2018-01-01	Hawaii	Honolulu	Not in a city	Kapolei	Carbon monoxide	Parts per million	0.200000	2
7	7	2018-01-01	Pennsylvania	Erie	Erie	NaN	Carbon monoxide	Parts per million	0.200000	2
8	8	2018-01-01	Hawaii	Honolulu	Honolulu	Honolulu	Carbon monoxide	Parts per million	0.400000	5
9	9	2018-01-01	Colorado	Larimer	Fort Collins	Fort Collins - CSU - S. Mason	Carbon monoxide	Parts per million	0.300000	6

In [196…
```
# Explore your dataframe `aqi` here:

print("Use head() to show a sample of data")
print(aqi.head())

print("Use describe() to summarize AQI")
print(aqi.describe(include='all'))

print("For a more thorough examination of observations by state use values_counts()")
print(aqi['state_name'].value_counts())

print('for a more')
```

```
Use head() to show a sample of data
   Unnamed: 0  date_local    state_name   county_name      city_name  \
0           0  2018-01-01       Arizona      Maricopa        Buckeye
1           1  2018-01-01          Ohio       Belmont      Shadyside
2           2  2018-01-01       Wyoming         Teton  Not in a city
3           3  2018-01-01  Pennsylvania  Philadelphia   Philadelphia
4           4  2018-01-01          Iowa          Polk     Des Moines

                                local_site_name   parameter_name  \
0                                     BUCKEYE  Carbon monoxide
1                                   Shadyside  Carbon monoxide
2  Yellowstone National Park - Old Faithful Snow ...  Carbon monoxide
3                           North East Waste (NEW)  Carbon monoxide
4                                   CARPENTER  Carbon monoxide

    units_of_measure  arithmetic_mean  aqi
0  Parts per million         0.473684    7
1  Parts per million         0.263158    5
2  Parts per million         0.111111    2
3  Parts per million         0.300000    3
4  Parts per million         0.215789    3
Use describe() to summarize AQI
        Unnamed: 0  date_local  state_name  county_name      city_name  \
count   260.000000         260         260          260            260
unique         NaN           1          52          149            190
top            NaN  2018-01-01  California  Los Angeles  Not in a city
freq           NaN         260          66           14             21
mean    129.500000         NaN         NaN          NaN            NaN
std      75.199734         NaN         NaN          NaN            NaN
min       0.000000         NaN         NaN          NaN            NaN
25%      64.750000         NaN         NaN          NaN            NaN
50%     129.500000         NaN         NaN          NaN            NaN
75%     194.250000         NaN         NaN          NaN            NaN
max     259.000000         NaN         NaN          NaN            NaN

       local_site_name   parameter_name   units_of_measure  arithmetic_mean  \
count              257              260                260       260.000000
unique             253                1                  1              NaN
top            Kapolei  Carbon monoxide  Parts per million              NaN
freq                 2              260                260              NaN
mean               NaN              NaN                NaN         0.403169
std                NaN              NaN                NaN         0.317902
min                NaN              NaN                NaN         0.000000
25%                NaN              NaN                NaN         0.200000
50%                NaN              NaN                NaN         0.276315
75%                NaN              NaN                NaN         0.516009
max                NaN              NaN                NaN         1.921053

              aqi
count   260.000000
unique         NaN
top            NaN
freq           NaN
mean      6.757692
std       7.061707
min       0.000000
25%       2.000000
50%       5.000000
75%       9.000000
max      50.000000
For a more thorough examination of observations by state use values_counts()
state_name
California              66
Arizona                 14
Ohio                    12
Florida                 12
Texas                   10
New York                10
Pennsylvania            10
Michigan                 9
Colorado                 9
Minnesota                7
New Jersey               6
Indiana                  5
North Carolina           4
Massachusetts            4
Maryland                 4
Oklahoma                 4
Virginia                 4
Nevada                   4
Connecticut              4
Kentucky                 3
Missouri                 3
Wyoming                  3
Iowa                     3
Hawaii                   3
Utah                     3
Vermont                  3
Illinois                 3
New Hampshire            2
District Of Columbia     2
New Mexico               2
Montana                  2
Oregon                   2
Alaska                   2
Georgia                  2
Washington               2
Idaho                    2
Nebraska                 2
Rhode Island             2
Tennessee                2
Maine                    2
South Carolina           1
Puerto Rico              1
Arkansas                 1
Kansas                   1
Mississippi              1
Alabama                  1
Louisiana                1
Delaware                 1
South Dakota             1
West Virginia            1
North Dakota             1
Wisconsin                1
Name: count, dtype: int64
for a more
```

In [198…
```
# Create dataframes for each sample being compared in your test

ca_la = aqi[aqi['county_name']=='Los Angeles']
ca_other = aqi[(aqi['state_name']=='California') & (aqi['county_name']!='Los Angeles')]
```

In [200…
```
# For this analysis, the significance level is 5%

significance_level = 0.05
significance_level
```

Out[200… 0.05

In [202…
```
stats.ttest_ind(a=ca_la['aqi'], b=ca_other['aqi'], equal_var=False)
```

Out[202… TtestResult(statistic=2.1107010796372014, pvalue=0.049839056842410995, df=17.08246830361151)

In [204…
```
# Create dataframes for each sample being compared in your test

ny = aqi[aqi['state_name']=='New York']
ohio = aqi[aqi['state_name']=='Ohio']
```

In [206…
```
# Compute your p-value here

tstat, pvalue = stats.ttest_ind(a=ny['aqi'], b=ohio['aqi'], alternative='less', equal_var=False)
print(tstat)
print(pvalue)
```

```
-2.025951038880333
0.03044650269193468
```

In [208…
```
# Create dataframes for each sample being compared in your test

michigan = aqi[aqi['state_name']=='Michigan']

# Compute your p-value here

tstat, pvalue = stats.ttest_1samp(michigan['aqi'], 10, alternative='greater')
print(tstat)
print(pvalue)
```

```
-1.7395913343286131
0.9399405193140109
```

In [ ]:
```
# key takeaways

# Even with small sample sizes, the variation within the data is enough to allow you to make statistically significant conclusions.
# You identified at the 5% significance level that the Los Angeles mean AQI was stastitically different from the rest of California, and that New York does have a lower mean AQI than Ohio.
# However, you were unable to conclude at the 5% significance level that Michigan's mean AQI was greater than 10.

# What would you consider presenting to your manager as part of your findings?

# For each test, you would present the null and alternative hypothesis, then describe your conclusion and the resulting p-value that drove that conclusion.
# As the setup of t-test's have a few key configurations that dictate how you interpret the result, you would specify the type of test you chose, whether that tail was one-tail or two-tailed, and how you performed the t-test from stat
```

```
# What would you convey to external stakeholders?

# In answer to the research questions posed, you would convey the level of significance (5%) and your conclusion. Additionally, providing the sample statistics being compared in each case will likely provide important context for stak
```