Using Python for Research Homework: Week 4, Case Study 2

In this case study, we will continue taking a look at patterns of flight for each of the three birds in our dataset.

In [10]:
```
# DO NOT EDIT THIS CODE
import pandas as pd
import numpy as np
birddata = pd.read_csv("https://courses.edx.org/asset-v1:HarvardX+PH526x+2T2019+type@as
birddata.head()
```

Out[10]:

	altitude	date_time	device_info_serial	direction	latitude	longitude	speed_2d	bird_name
0	71	2013-08-15 00:18:08+00	851	-150.469753	49.419860	2.120733	0.150000	Eric
1	68	2013-08-15 00:48:07+00	851	-136.151141	49.419880	2.120746	2.438360	Eric
2	68	2013-08-15 01:17:58+00	851	160.797477	49.420310	2.120885	0.596657	Eric
3	73	2013-08-15 01:47:51+00	851	32.769360	49.420359	2.120859	0.310161	Eric
4	69	2013-08-15 02:17:42+00	851	45.191230	49.420331	2.120887	0.193132	Eric

Exercise 1

In this case study, we will continue taking a look at patterns of flight for each of the three birds in our dataset. We will group the flight patterns by bird and date, and plot the mean altitude for these groupings.

`pandas` makes it easy to perform basic operations on groups within a dataframe without needing to loop through each value in the dataframe. In this exercise, we will group the dataframe by `birdname` and then find the average `speed_2d` for each bird.

Instructions

- Fill in the code to find the mean altitudes of each bird using the pre-loaded `birddata` dataframe.

In [14]:
```
# First, use `groupby()` to group the data by "bird_name".
grouped_birds = birddata.groupby(by='bird_name')

# Now calculate the mean of `speed_2d` using the `mean()` function.
mean_speeds = grouped_birds.speed_2d.mean()

# Find the mean `altitude` for each bird.
mean_altitudes = grouped_birds.altitude.mean()
```

In [47]:
```
mean_speeds
```

Out[47]:
```
bird_name
Eric     2.300545
Nico     2.908726
Sanne    2.450434
Name: speed_2d, dtype: float64
```

Exercise 2

In this exercise, we will group the flight times by date and calculate the mean altitude within that day.

Instructions

- Convert `birddata.date_time` to the `pd.datetime` format, and store as `birddata["date"]`.
- Fill in the code to find the mean altitudes for each day.

In [12]:
```
# Convert birddata.date_time to the `pd.datetime` format.
birddata.date_time = pd.to_datetime(birddata.date_time)

# Create a new column of day of observation
birddata["date"] = birddata.date_time.dt.date

# Use `groupby()` to group the data by date.
grouped_bydates = birddata.groupby(by='date')

# # Find the mean `altitude` for each date.
mean_altitudes_perday = grouped_bydates.altitude.mean()
```

Exercise 3

In this exercise, we will group the flight times by both bird and date, and calculate the mean altitude for each.

Instructions

- `birddata` already contains the `date` column. To find the average speed for each bird and day, create a new grouped dataframe called `grouped_birdday` that groups the data by both `bird_name` and date.

In [53]:
```
# Use `groupby()` to group the data by bird and date.
grouped_birdday = birddata.groupby(by=["bird_name","date"])

# Find the mean `altitude` for each bird and date.
mean_altitudes_perday = grouped_birdday.altitude.mean()
```

In [54]:
```
mean_altitudes_perday
```

Out[54]:
```
bird_name  date
Eric       2013-08-15     74.988095
           2013-08-16    127.773810
           2013-08-17    125.890244
           2013-08-18    121.353659
```

```
            2013-08-19    134.928571
                             ...
Sanne       2014-04-26     17.116667
            2014-04-27     17.391892
            2014-04-28     58.876712
            2014-04-29     30.530120
            2014-04-30      4.361111
Name: altitude, Length: 770, dtype: float64
```

Exercise 4

Great! Now find the average speed for each bird and day.

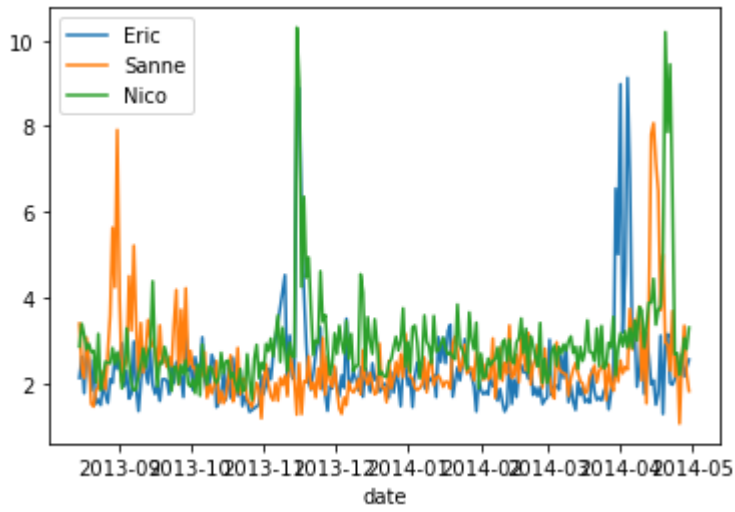
Instructions

- Store these are three `pandas` `Series` objects, one for each bird.
- Use the plotting code provided to plot the average speeds for each bird.

In [77]:
```
import matplotlib.pyplot as plt

mean_speed_perday = grouped_birdday.speed_2d.mean()
eric_daily_speed  = mean_speed_perday.get("Eric")
sanne_daily_speed = mean_speed_perday.get("Sanne")
nico_daily_speed  = mean_speed_perday.get("Nico")

eric_daily_speed.plot(label="Eric")
sanne_daily_speed.plot(label="Sanne")
nico_daily_speed.plot(label="Nico")
plt.legend(loc="upper left")
plt.show()
```



In [88]:
```
nico_daily_speed.tail(30)
```

Out[88]:
```
date
2014-04-01     3.163723
2014-04-02     2.861222
2014-04-03     3.212099
2014-04-04     2.832465
2014-04-05     3.283842
2014-04-06     2.824700
2014-04-07     3.455989
2014-04-08     2.995421
2014-04-09     3.780186
```

```
2014-04-10     3.703409
2014-04-11     2.829536
2014-04-12     3.341111
2014-04-13     3.878121
2014-04-14     3.882314
2014-04-15     4.437659
2014-04-16     3.366451
2014-04-17     3.713230
2014-04-18     3.798646
2014-04-19     5.061530
2014-04-20    10.196981
2014-04-21     7.861385
2014-04-22     9.445087
2014-04-23     6.384096
2014-04-24     2.674536
2014-04-25     2.705160
2014-04-26     2.192028
2014-04-27     2.582072
2014-04-28     3.055051
2014-04-29     2.793232
2014-04-30     3.297032
Name: speed_2d, dtype: float64
```