

A decorative background on the left side of the slide features a large, stylized Python logo icon in gold and blue, with many smaller, semi-transparent Python logo icons scattered around it. Below this, a cartoon-style panda bear is shown hanging from a vertical dark grey pillar.

Industrial Land Joining Exercise

CP101 Lab 1

Merge, Join and
Concatenate DataFrames

Merging Dataframe

```
pd.merge(df1, df2, how='')
```

	key	value1
0	a	0
1	b	1
2	c	2
3	d	3
4	f	4

+

	key	value2
0	a	0
1	b	1
2	f	2

how = "inner"

	key	value1	value2
0	a	0	0
1	b	1	1
2	f	4	2

how = "left"

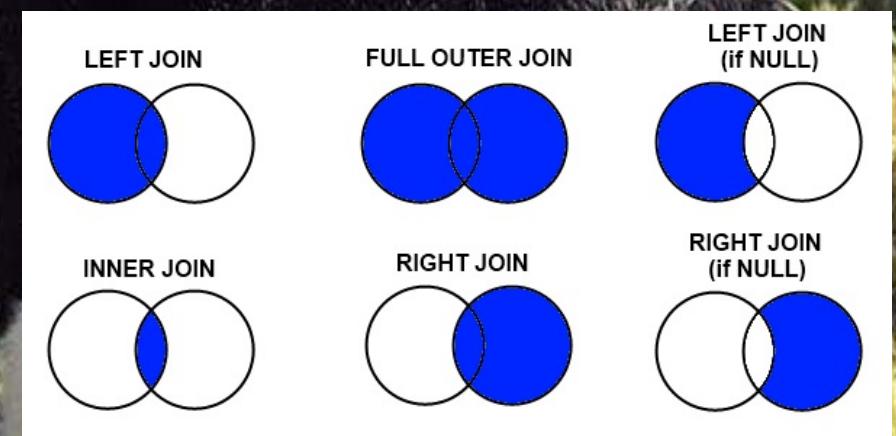
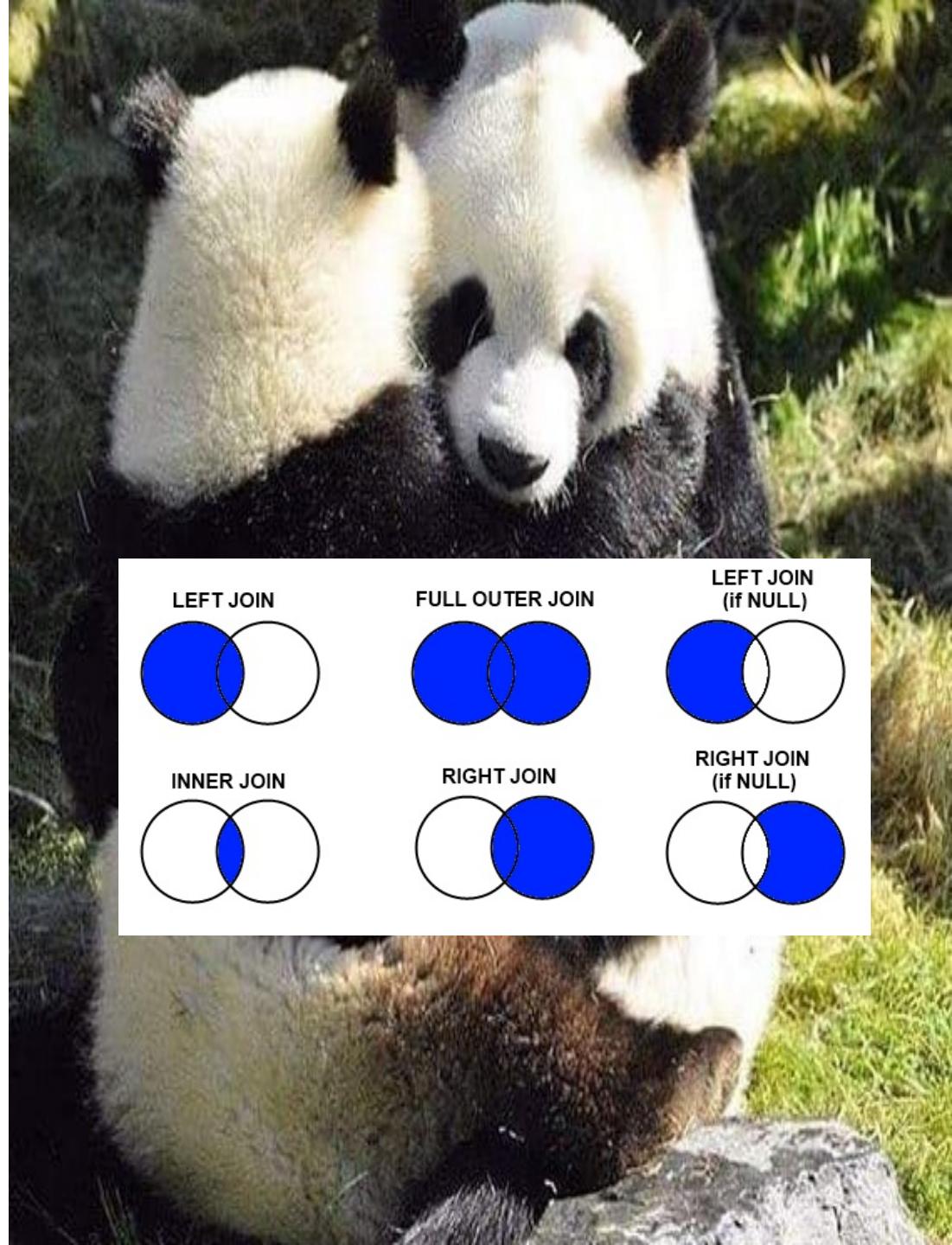
	key	value1	value2
0	a	0	0.0
1	b	1	1.0
2	c	2	NaN
3	d	3	NaN
4	f	4	2.0

how = "right"

	key	value1	value2
0	a	0	0
1	b	1	1
2	f	4	2

how = "outer"

	key	value1	value2
0	a	0	0.0
1	b	1	1.0
2	c	2	NaN
3	d	3	NaN
4	f	4	2.0





Many-to-one

	key	value1
0	b	0
1	b	1
2	a	2
3	c	3
4	a	4
5	a	5
6	b	6

+

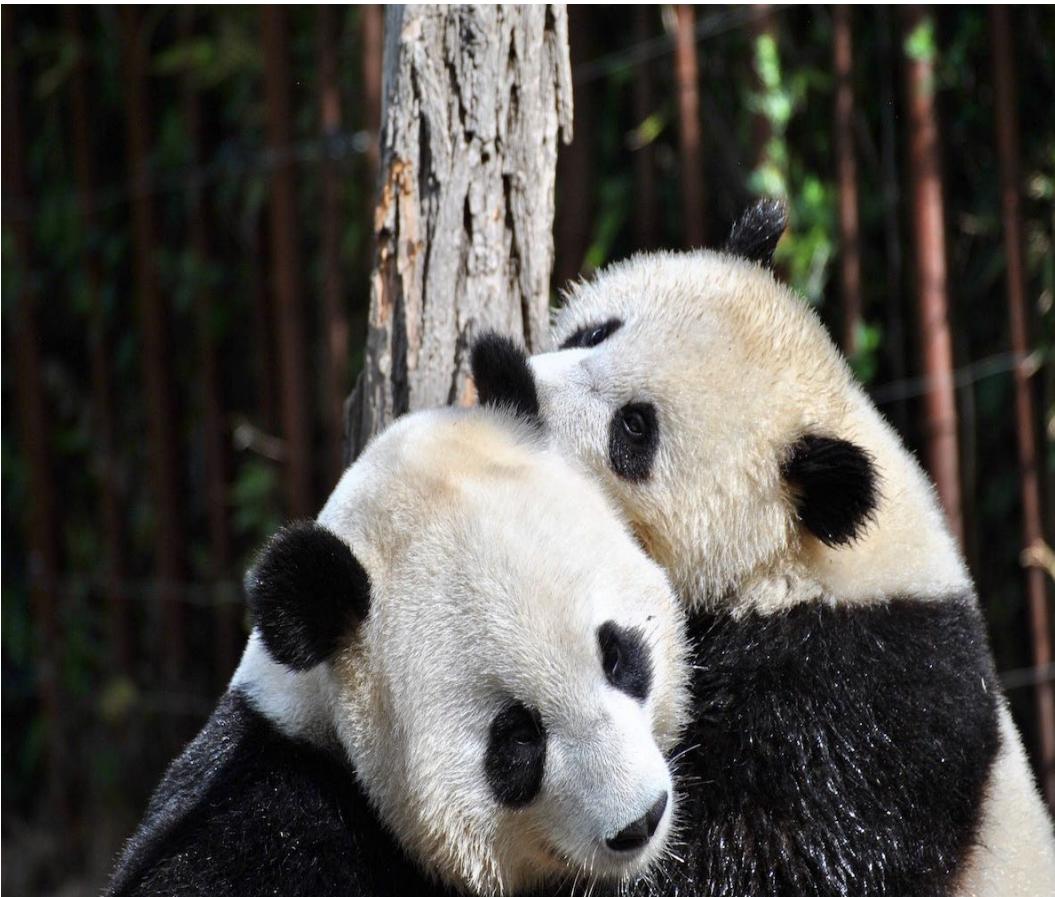
	key	value2
0	a	0
1	b	1
2	d	2

=

	key	value1	value2
0	b	0	1
1	b	1	1
2	b	6	1
3	a	2	0
4	a	4	0
5	a	5	0

pd.merge(df1, df2, on='key')

Joining Dataframe



join_df1

	Ohio	Nevada
a	1	2
c	3	4
e	5	6

join_df2

	Missouri	Alabama
b	7	8
c	9	10
d	11	12
e	13	14

join_df3

	New York	Oregon
a	11	12
c	13	14
e	15	16

`join_df1.join(join_df2, how='inner')`

	Ohio	Nevada	Missouri	Alabama
a	1	2	NaN	NaN
c	3	4	9.0	10.0
e	5	6	13.0	14.0

`join_df2.join([join_df1, join_df3], how='outer')`

	Missouri	Alabama	Ohio	Nevada	New York	Oregon
b	7.0	8.0	NaN	NaN	NaN	NaN
c	9.0	10.0	3.0	4.0	13.0	14.0
d	11.0	12.0	NaN	NaN	NaN	NaN
e	13.0	14.0	5.0	6.0	15.0	16.0

Concatenating Dataframe

`pd.concat([df1, df2], axis=)`

	key	value1
0	a	0
1	b	1
2	c	2
3	d	3
4	f	4

+

	key	value2
0	a	0
1	b	1
2	f	2

axis = 0
(row by row)

axis = 1
(column by column)

	key	value1	value2
0	a	0.0	NaN
1	b	1.0	NaN
2	c	2.0	NaN
3	d	3.0	NaN
4	f	4.0	NaN

	key	value1	key	value2
0	a	0	a	0.0
1	b	1	b	1.0
2	c	2	f	2.0
3	d	3	NaN	NaN
4	f	4	NaN	NaN

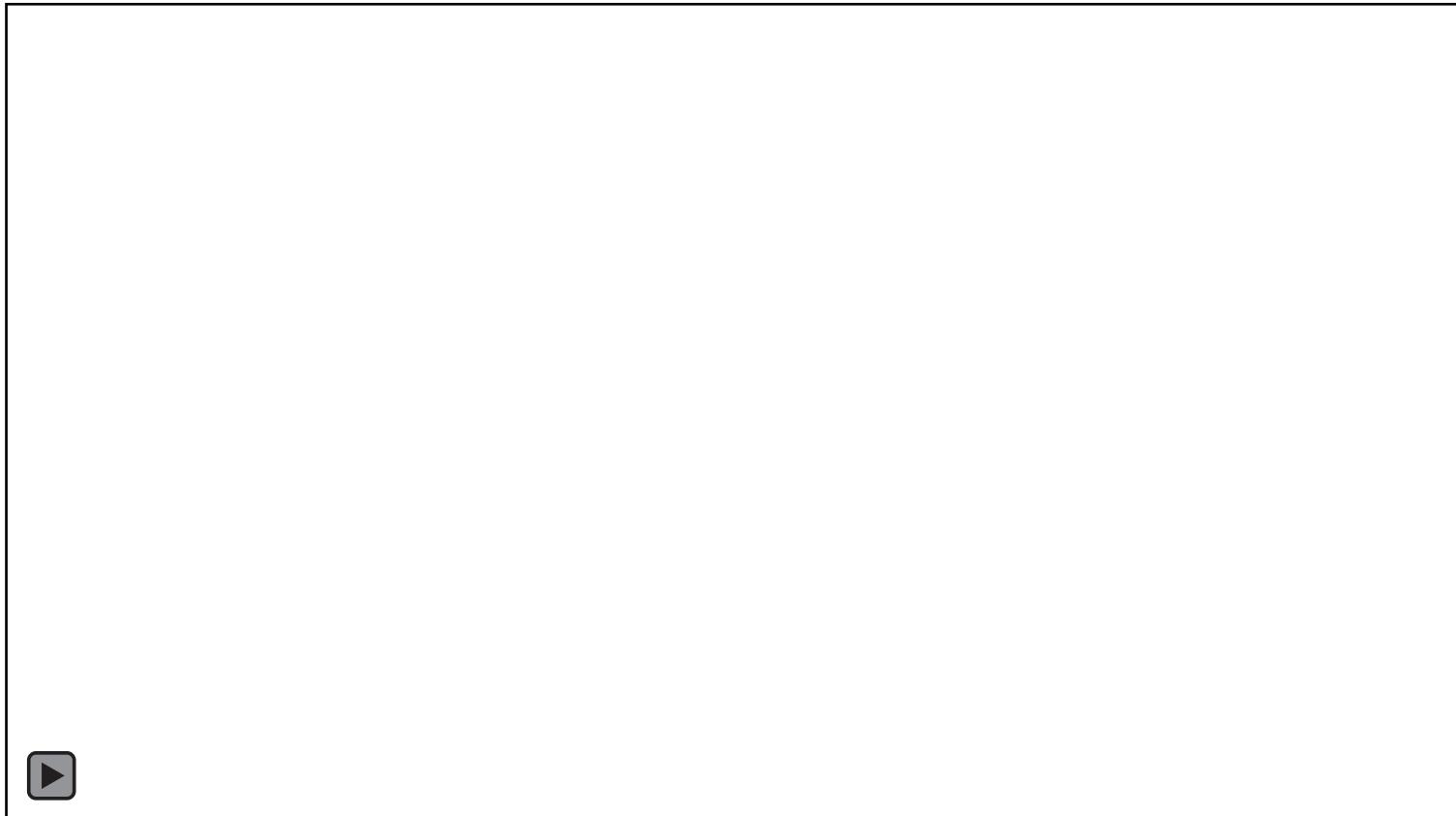


Reshape & Pivot Dataframe



This pivot() from data8 in the video is called **pivot_table()** in pandas
If this video doesn't play, watch it [here](#).

Groupby & Aggregate Dataframe



This group() from data8 in the video is called **groupby()** in pandas. Together, **df.groupby(['col_name']).mean()** or another agg function. If this video doesn't play, watch it [here](#).

Attribute & Spatial Join – Geopandas, ArcGIS, QGIS

Attribute

```
GeoDataFrame.merge(DataFame, on='unique key')
```

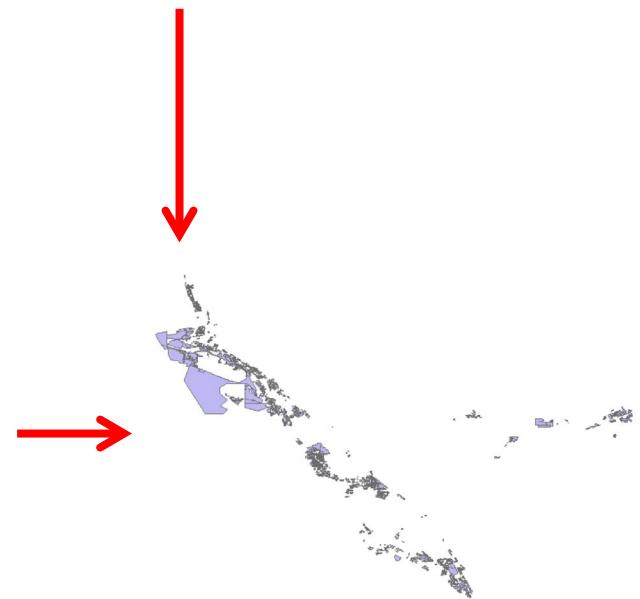
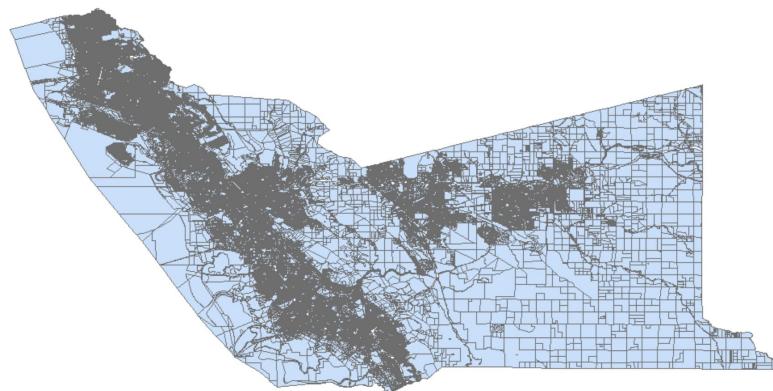
Spatial

```
GeoDataFrame.sjoin(GeoDataFrame,  
                   op='intersect, within or contains',  
                   how='left, right or inner')
```

Linked Tutorials



ArcMap



More Pandas Tutorial



References

- https://pandas.pydata.org/pandas-docs/stable/user_guide/merging.html
- <http://geopandas.org/mergingdata.html>
- http://geopandas.org/data_structures.html
- <https://www.shanelynn.ie/summarising-aggregation-and-grouping-data-in-python-pandas/>
- <https://towardsdatascience.com/transform-reality-with-pandas-96f061628030>
- Python for Data Analysis by Wes McKinney (Book)