

# Put a Good Tittle Here

By: Micheal, Ben, Shantanu, Manasa, Vishwa  
Josh, Yufeng, Joshuwa, Ning, Tianyi

# Merge Data

Problem we had when use pd.merge\_asof():

**EyeQ Data**



**Spire Data**



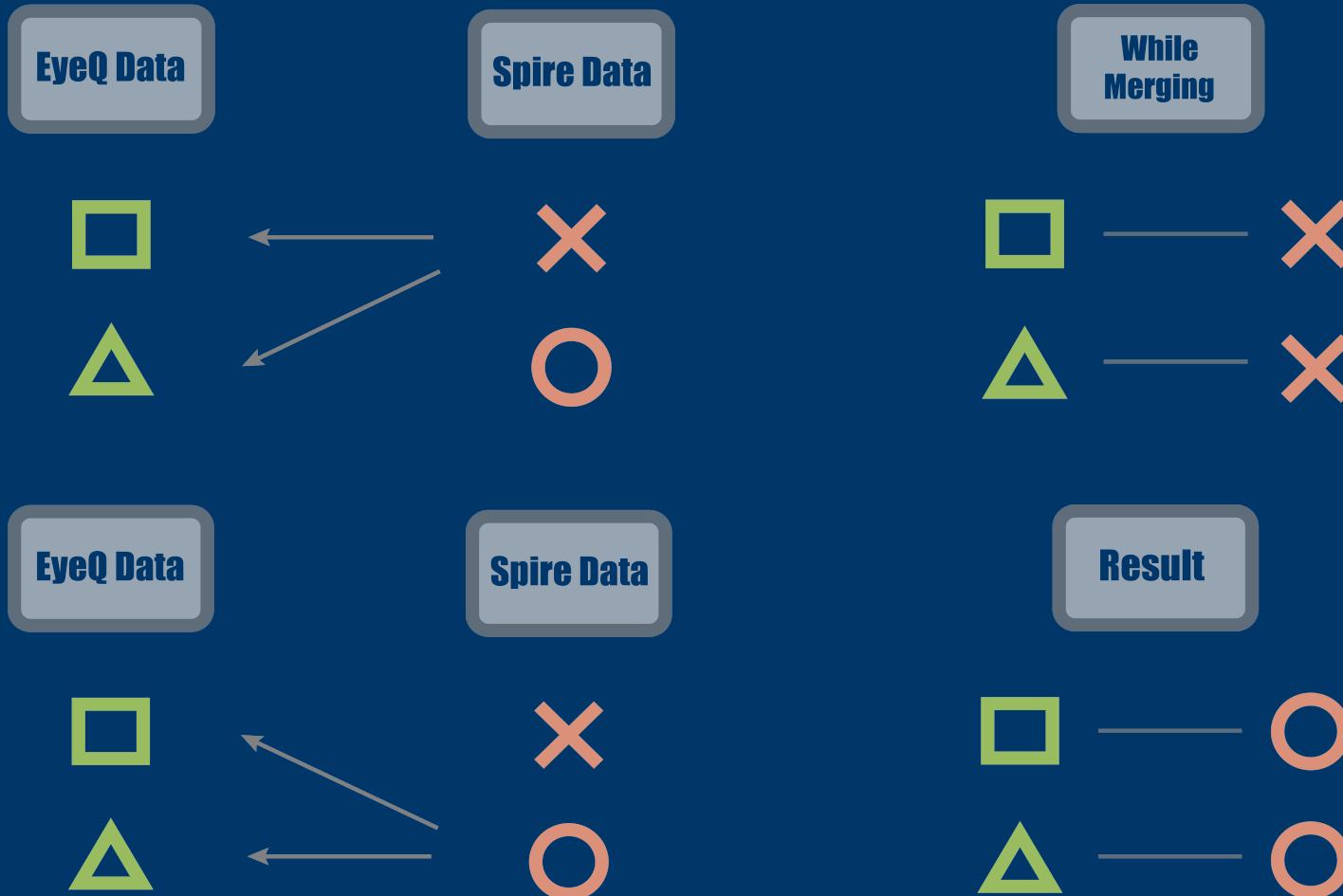
**True Match**



Suppose that are different event EyeQ data and are different event in Spire data

where and are true match. But those two matches are very close to each other.

# Merge Data



While merging, `merge_asof()` function will iterate each spire event and merge to a EyeQ event if within time tolerance. It will overwrite the former merge if the spire event can also merge to same EyeQ event.

In this example, we get instead of true match

# Merge Data

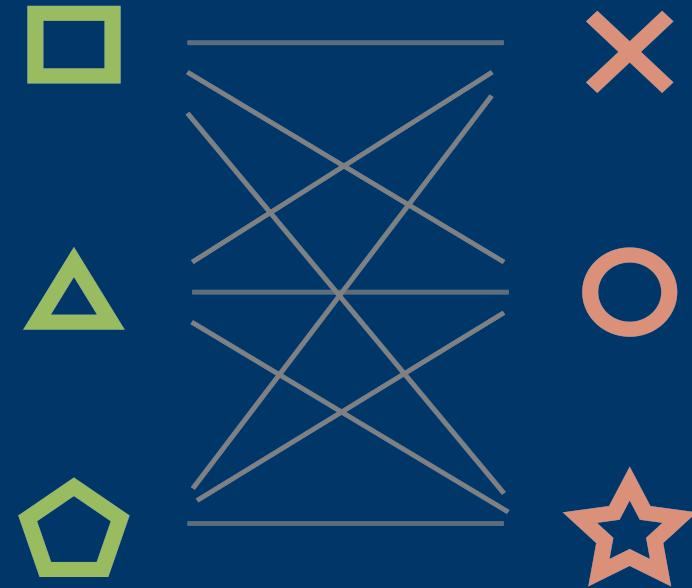
Solution:  
Manually merge two  
data set.

Firstly, cross join all the  
Rows from two datasets

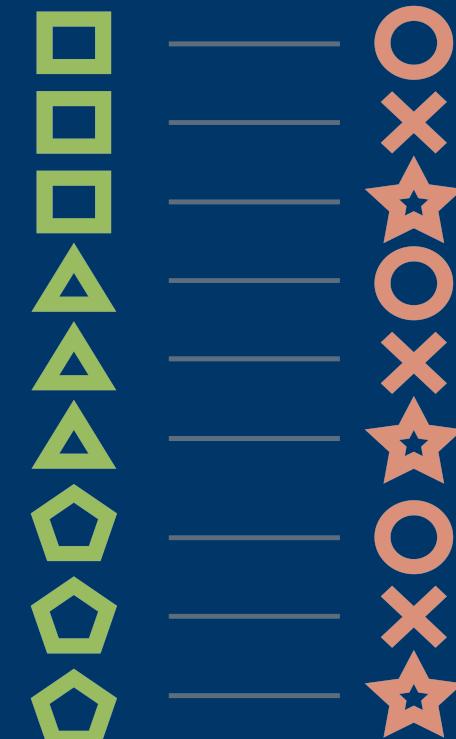
**EyeQ Data  
(Grouped)**

**Spire Data  
(Grouped)**

**Step 1  
Result**



**Simply  
cross  
Join  
every  
thing**



# Merge Data

## Merged Data

Create new  
new column  
time\_diff =  
EyeQ time -  
Spire time

	EyeQ Time	Spire Time	
□	12:20:05	○	12:20:08
□	12:20:05	✗	12:20:20
□	12:20:05	★	22:33:05
△	12:20:15	○	12:20:08
△	12:20:15	✗	12:20:20
△	12:20:15	★	22:33:05
○	19:50:05	○	12:20:08
○	19:50:05	✗	12:20:20
○	19:50:05	★	22:33:05

Drop all the rows that time\_diff is greater than the tolerance you set

# Merge Data

Merged data after drop rows  
(It can be regarded as four different parts according to duplicate events)

Non-duplicated  
EyqQ event match  
with  
Non-duplicated  
Spire event

For example:



Witch is perfect  
match, will be  
treated as part  
of final merge

Duplicated  
Spire events match  
with  
Non-duplicated  
EyeQ event

For example:



We should make  
choice which  
EyeQ event that  
match the Spire  
event

Non-duplicated  
Spire event match  
with  
Duplicated  
EyeQ event

For example:



We should make  
choice which  
Spire event that  
match the EyeQ  
event

Duplicated  
EyqQ event match  
with  
Duplicated  
Spire event

For example:



Should consider  
both Spire and  
EyeQ

# Merge Data

Duplicated  
Spire events match  
with  
Non-duplicated  
EyeQ event

For example:



We should make choice which EyeQ event that match the Spire event

Create a new flag column  
event\_estimator

=

Face\_size on scale of 0 to 1  
(the large the face\_size, the large the scale)  
+  
x (x-axis of face) on scale of 0 to 1  
(the close to 50 the x, the large the scale)  
+  
Time\_diff on scale of 0 to 2  
(the small the time\_diff, the large the scale)

Select the row has largest event\_estimator from Rows that have same spire event, concat to final dataset



event\_estimator

3.1712

1.3932

0.7149

# Merge Data

**Non-duplicated Spire event match with Duplicated EyeQ event**

For example:



We should make choice which Spire event that match the EyeQ event

Since we are selecting Spire event, event\_estimator is useless.  
Simply select the row has smallest time\_dff and concat it to the final dataset

	time_diff
□ ○	3
□ ✗	12
□ ★	18

**Duplicated EyeQ event match with Duplicated Spire event**

For example:



Should consider both Spire and EyeQ

Simply use event\_estimator (same way as handling duplicated spire event). Each time after concat a selected row to final dataset, drop any rows in the rest dataset which have identical EyeQ or Spire event number to the row just added.

**Then we get our final merged dataset !!!**