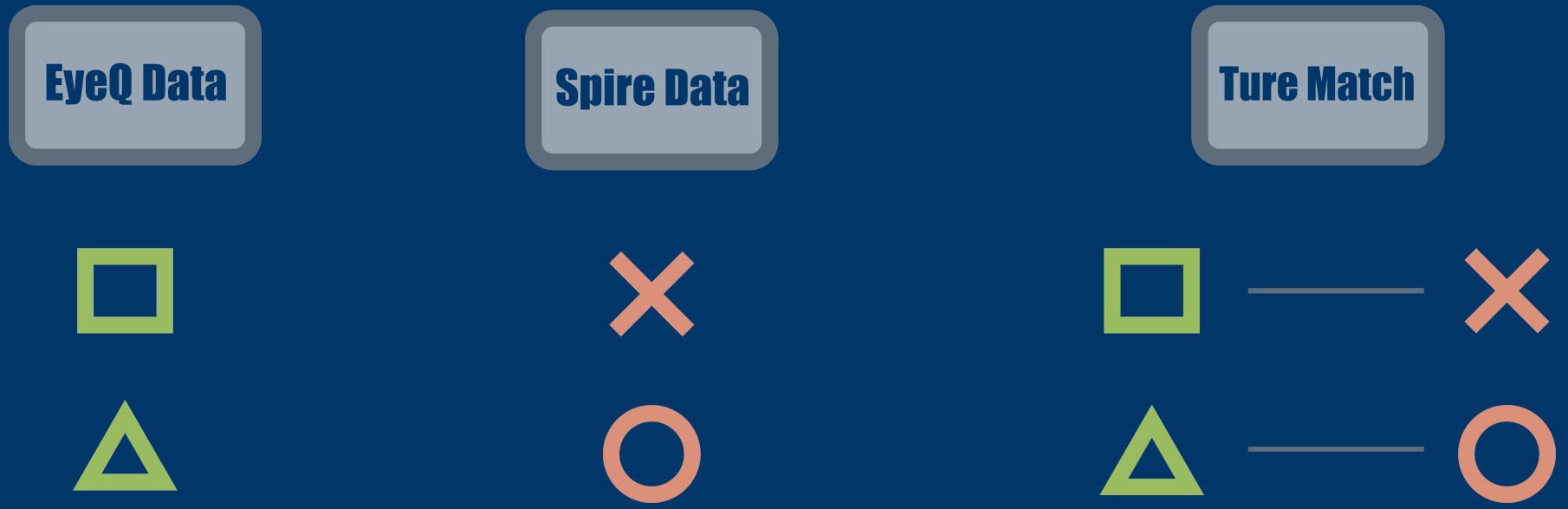
Put a Good Tittle Here

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

Time Shift

Merge Data

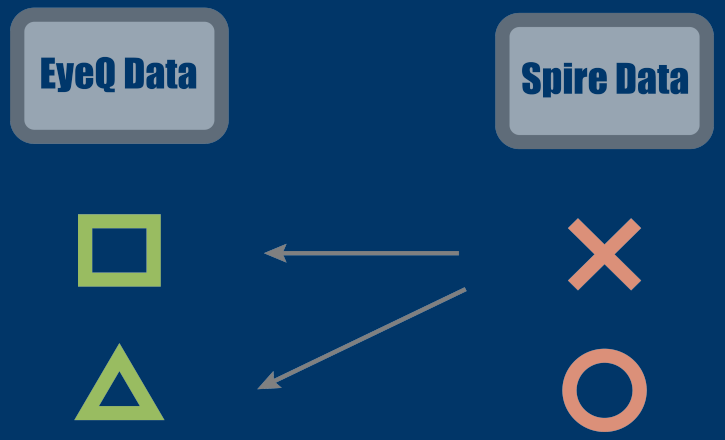
Problem we had when use pd.merge_asof():

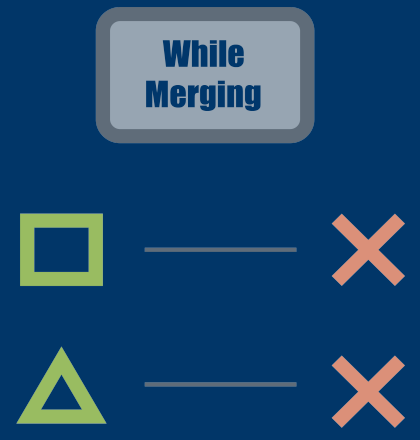


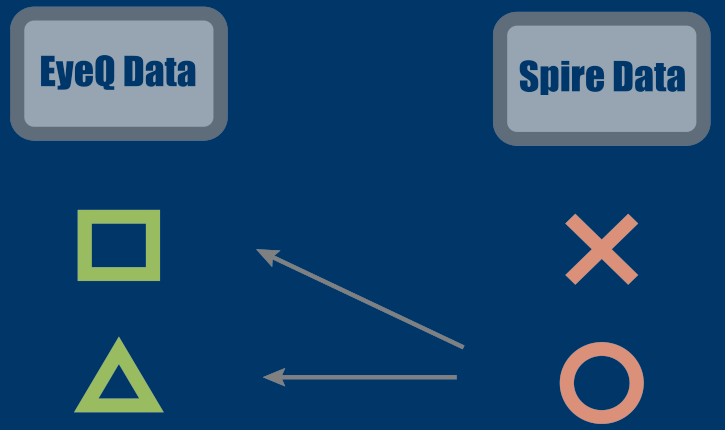
Suppose that □ △ are different event EyeQ data and ✕ ○ are different event in Spire data

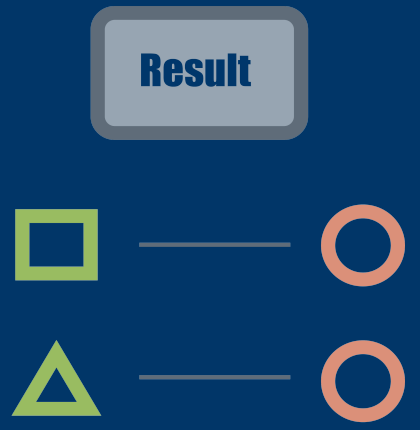
where □ ✕ and △ ○ are true match. But those two matches are very close to each other.

Merge Data









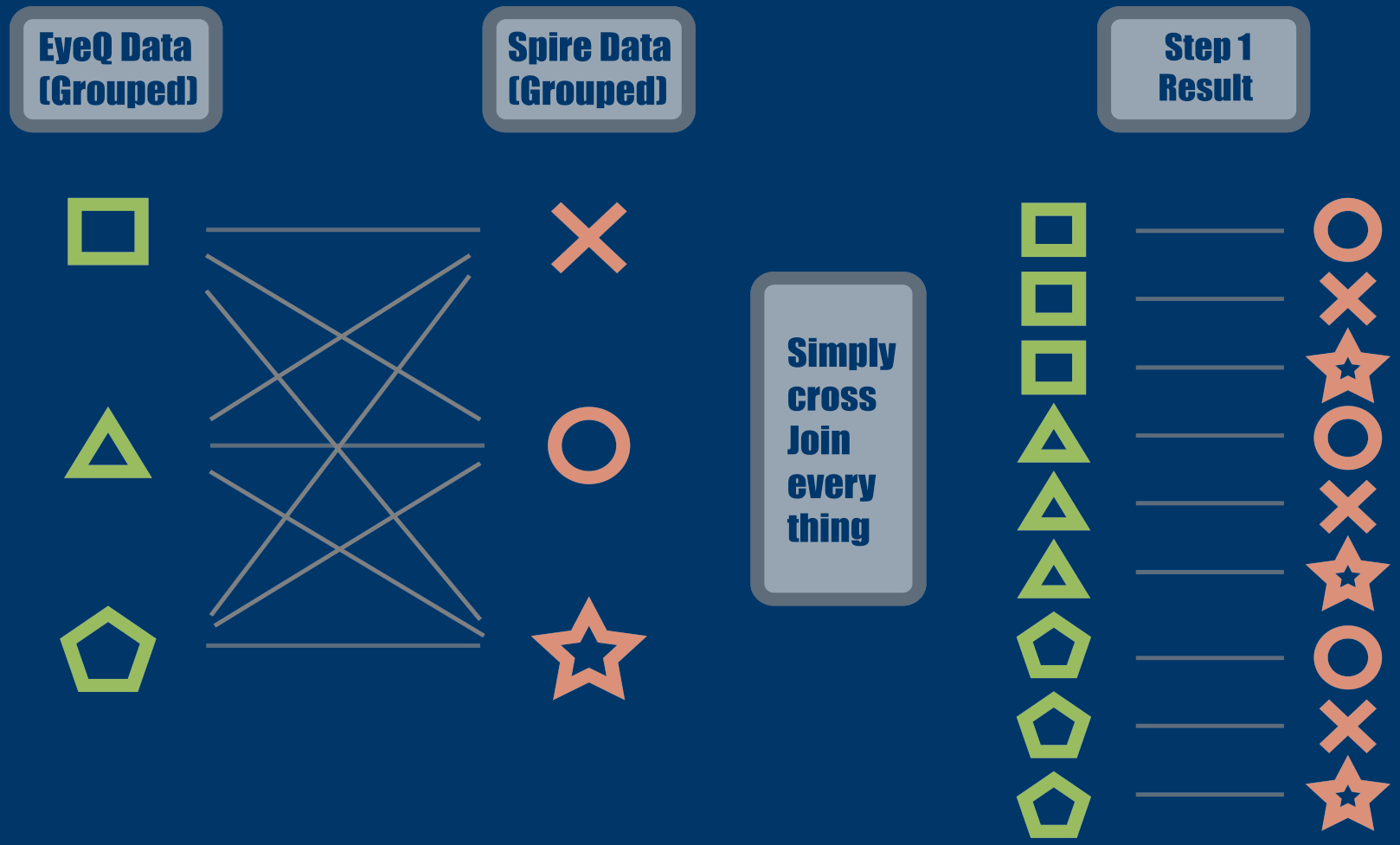
While merging, merge_asof() function will iterate each spire event and merge to a EyeQ event if within time tolerance. It will overwrite the former merge if the spire event can also merge to same EyeQ event.

In this example, we get 

instead of ture match 

Merge Data

Solution:
Manually merge two data set.
Firstly, cross join all the Rows from two datasets



Merge Data


Merged Data
Create new new column time_diff = EyeQ time - Spire time
EyeQ Time
Spire Time
Drop all the rows that time_diff is greater than the tolerance you set
12:20:05
12:20:08
00:00:03
12:20:05
12:20:20
00:00:15
12:20:05
22:33:05
10:13:00
12:20:15
12:20:08
00:00:07
12:20:15
12:20:20
00:00:05
12:20:15
22:33:05
10:12:50
19:50:05
12:20:08
07:29:57
19:50:05
12:20:20
07:29:45
19:50:05
22:33:05
02:43:00

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-dupilcated Spire event

For example:

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

Duplicated Spire events match with Non-dupilcated EyeQ event

For example:

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

Non-duplicated Spire event match with Dupilcated EyeQ event

For example:

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

Duplicated EyqQ event match with Dupilcated Spire event

For example:

Should consider both Spire and EyeQ

Merge Data

Duplicated
Spire events match with
Non-dupilcated
EyeQ event

For example:

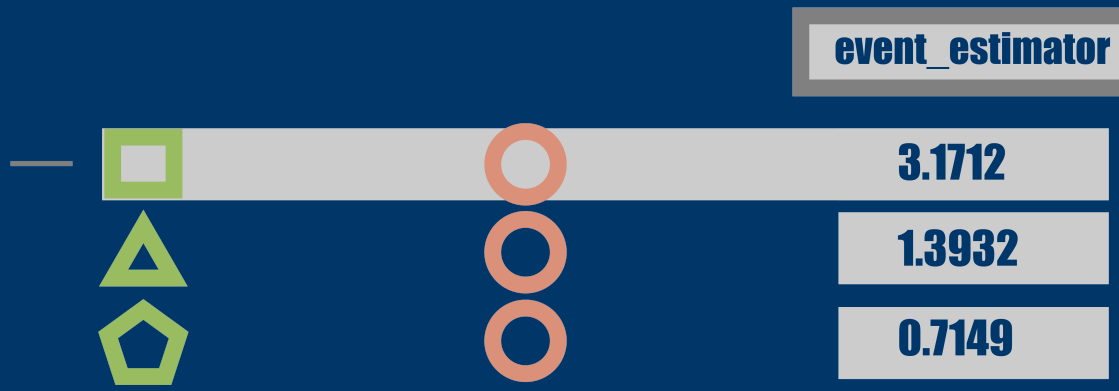
We should make choice which EyeQ event that match the Spire event

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



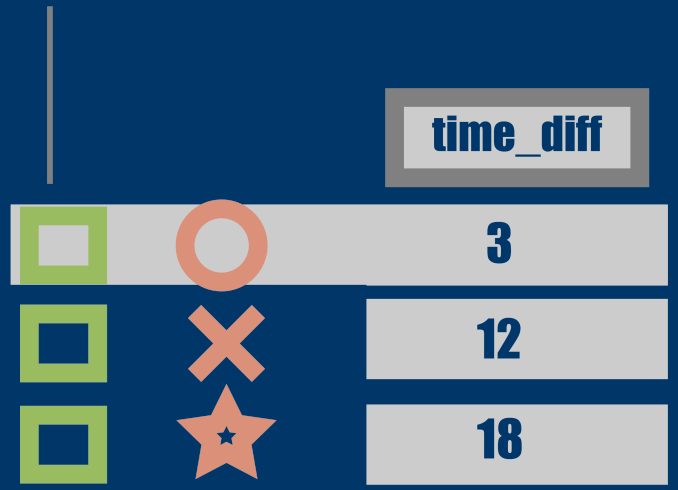
Merge Data

Non-duplicated Spire event match with Dupilcated 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



Duplicated EyqQ event match with Dupilcated 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 !!!