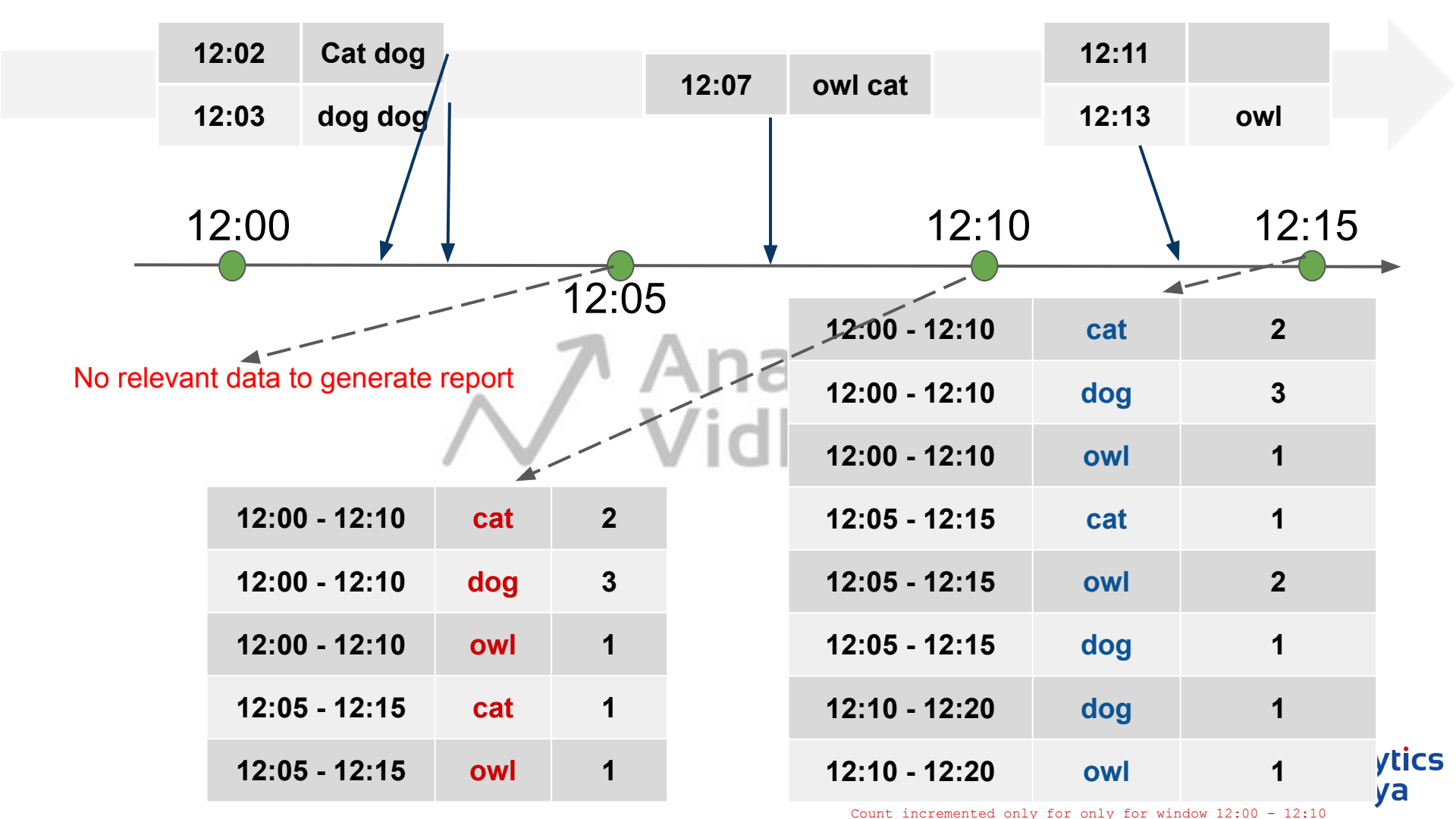
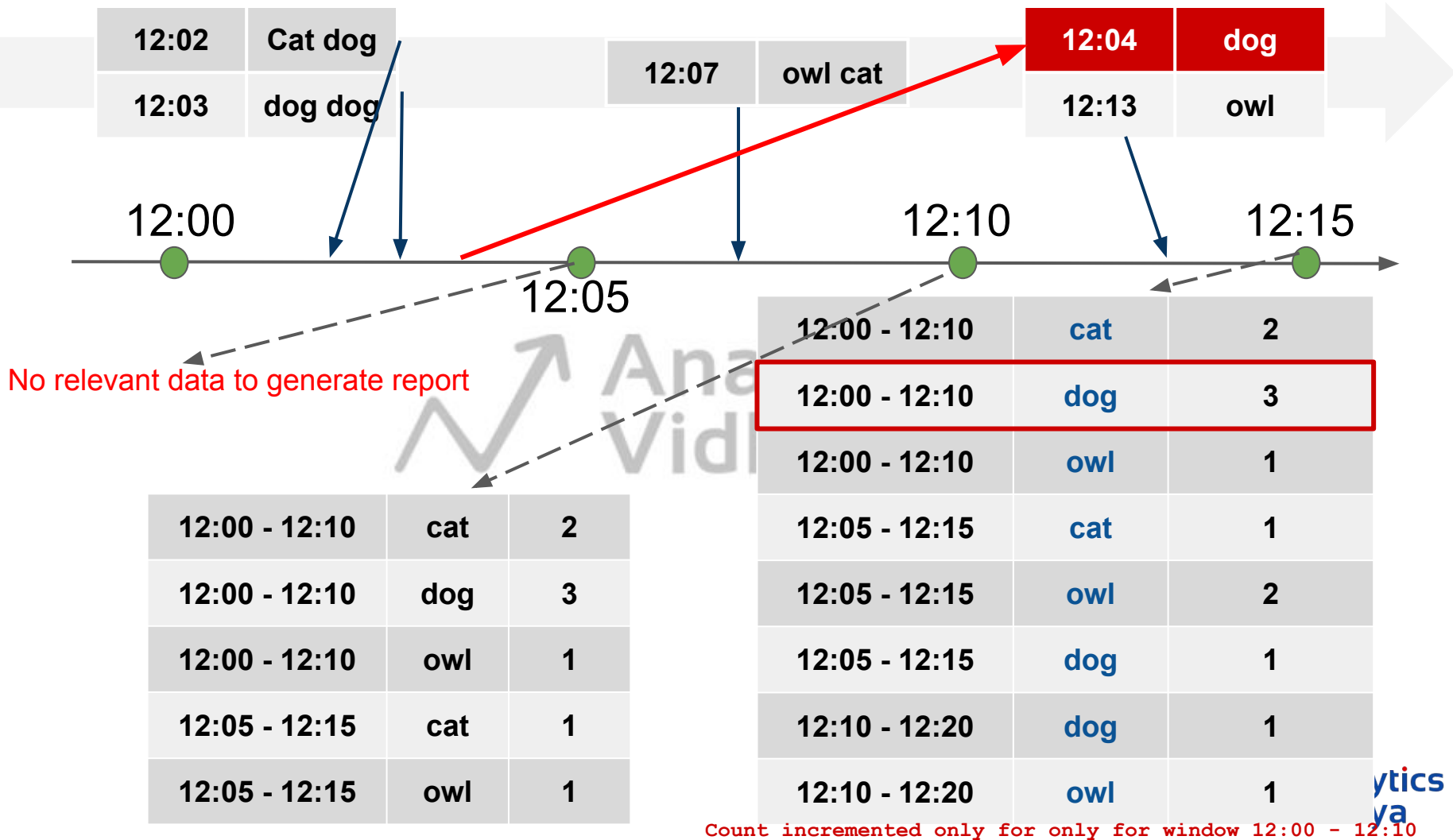
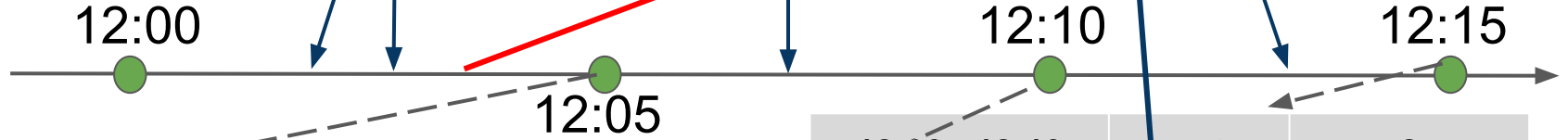


Watermarking





12:02	Cat dog		12:07	owl cat		12:04	dog
12:03	dog dog					12:13	owl



12:00 - 12:10	cat	1
12:00 - 12:10	dog	3

12:00 - 12:10	cat	2
12:00 - 12:10	dog	3
12:00 - 12:10	owl	1
12:05 - 12:15	cat	1
12:05 - 12:15	owl	1

12:00 - 12:10	cat	2
12:00 - 12:10	dog	4
12:00 - 12:10	owl	1
12:05 - 12:15	cat	1
12:05 - 12:15	owl	2
12:10 - 12:20	owl	1

Count incremented only for only
for window 12:00 - 12:10

Watermarking Implementation - Example

```
words = ... # streaming DataFrame of schema {timestamp: Timestamp, word: String}

# Group the data by window and word and compute the count of each group
windowedCounts = words \
    .withWatermark("timestamp", "10 minutes") \
    .groupBy(
        window(words.timestamp, "10 minutes", "5 minutes"),
        words.word) \
    .count()
```

 Thank You
Analytics
Vidhya