



BDA

Assignment 3

Aakash Tanwar 2016215

Priya Rajpurohit 2015073





Learnings



What did you learn?

- ❑ Using Streaming platforms like Tweepy, Apache Kafka
- ❑ Real time Streaming

How much time did it take?

- ❑ Few days to understand and write the code/program as per the question
- ❑ 10 hrs to run the code



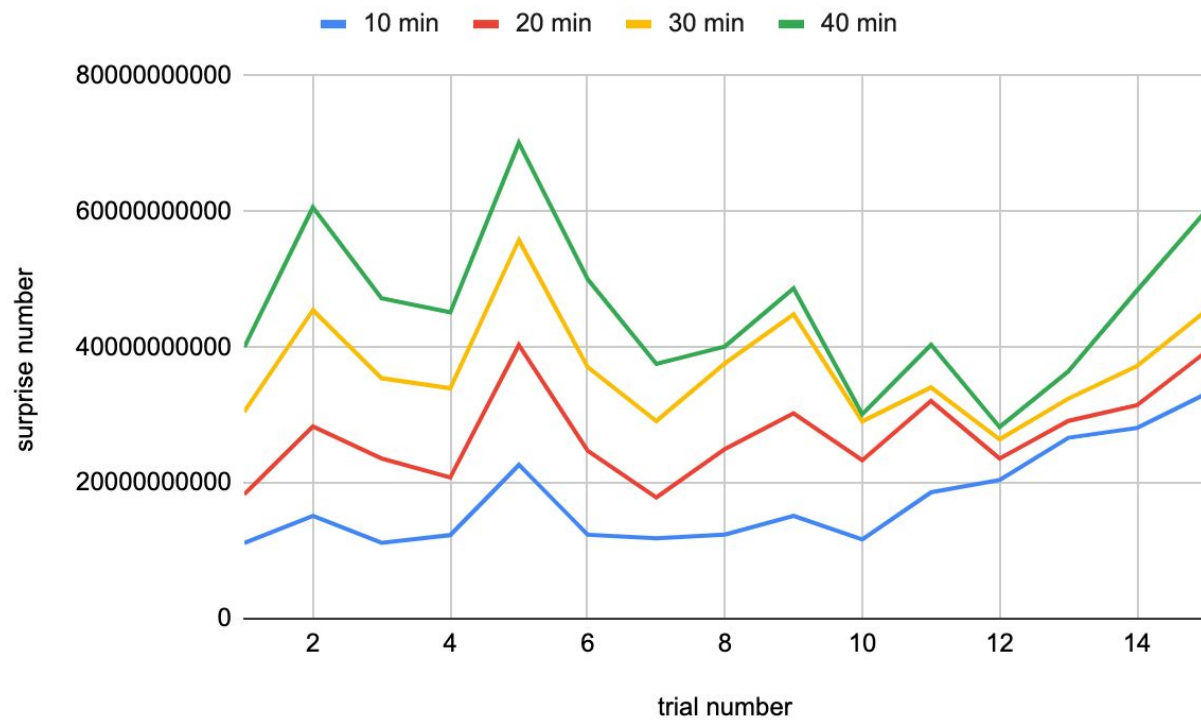
Challenges faced

- ❑ Connection timeout every 3 hours
- ❑ Kafka was timing out in 10 mins
 - ❑ Network issues
- ❑ Running kafka before zookeeper

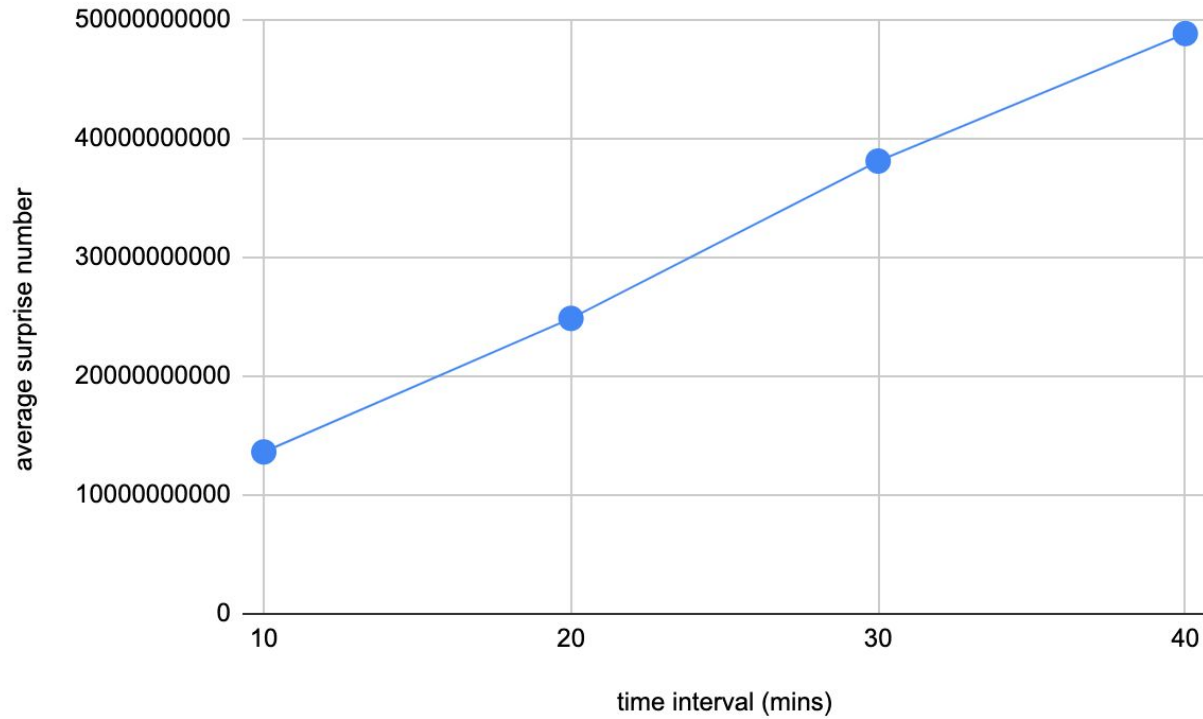


Results

Graph




Graph





Average of surprise numbers

- ❑ For 10 mins: 13632258722
- ❑ For 20 mins: 24876960473
- ❑ For 30 mins: 38125276663
- ❑ For 40 mins: 48862243314



Could not find an API for
selecting the samples
with uniform probability as
computation of 2nd moment



Part 3

Resources Used

1. Kafka Documentation
2. Kafka-python documentation
3. Twitter Developers documentation
4. Tweepy documentation