Flow of how our program is running:

1. Set up Kafka zookeeper services and producer
2. Once we have a producer, it makes API calls to publish data to the topics in brokers
3. For the scope of this project we are using Microsoft and Amazon stock pricing data
4. Hence, we make two topics named microsoft and amazon to which the data is published.
5. The microsoft data is retrieved from IEX cloud using an API key (P544V2LKQW6KH46R)in JSON format
6. The amazon data is retrieved from Alpha Vantage using API key (pk_045f8defa0e347ccb7109baaeacf3339) in CSV format
7. The data retrieved for each stock is of at least 6000 rows as it dates back to the year 2000
8. Once we have published data to the producer, the consumer streams this data using spark
9. The spark session reads data in real time, processes and cleans it to ensure it is ready for the LSTM
10. Then we pass our data to the LSTM which makes predictions based on that data

This achieves variety, volume, velocity