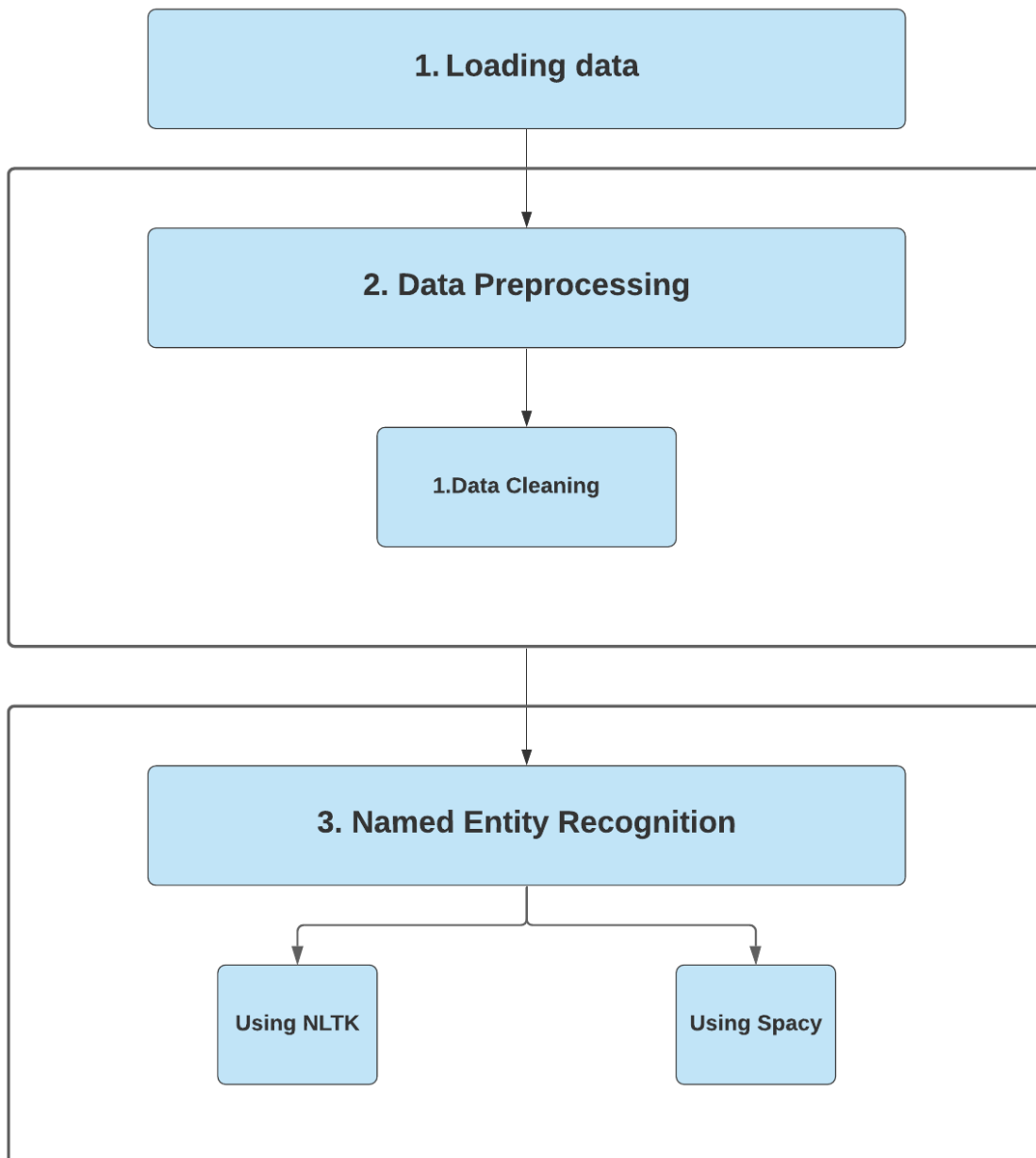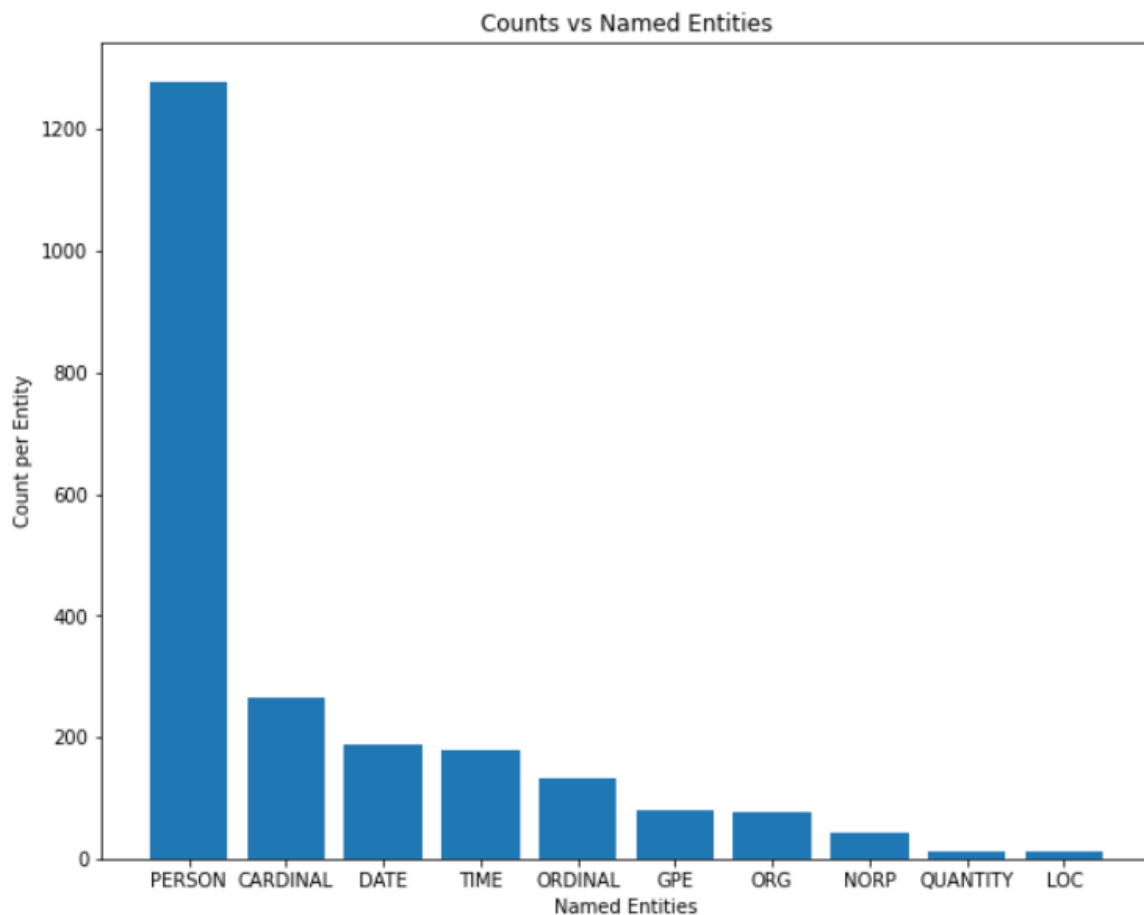# Named Entity Recognition(NER)

Named Entity Recognition is a sub-task of information extraction to classify named entities - places, people, organizations, dates, etc. in a text.
Following is the methodology adopted for this project.

```
┌─────────────────────────────────────┐
│        1. Loading data              │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────────┐
│  ┌─────────────────────────────────────┐ │
│  │      2. Data Preprocessing          │ │
│  └─────────────────────────────────────┘ │
│                  │                        │
│                  ▼                        │
│         ┌──────────────────┐              │
│         │  1.Data Cleaning  │             │
│         └──────────────────┘              │
└─────────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────────┐
│  ┌─────────────────────────────────────┐ │
│  │   3. Named Entity Recognition       │ │
│  └─────────────────────────────────────┘ │
│            │                 │            │
│            ▼                 ▼            │
│   ┌──────────────┐   ┌──────────────┐    │
│   │  Using NLTK  │   │ Using Spacy  │    │
│   └──────────────┘   └──────────────┘    │
└─────────────────────────────────────────┘
```

Results

We basically tried two techniques, one using the NLTK library and another with spacy. Below some sample results for both are given. It is quite evident after checking a few samples that the spacy library does an impressive job of identifying the named entities in the given text. The results from NER are used further in many tasks like finding which issues/products are being named in the reviews and companies can dig deep down on such products to find any bottlenecks. Using sentiment analysis and NER to better understand the stock markets. For eg. Below is a snap of the top 10 entities in our data.

Spacy Results

	Entities	Labels
1064	(one)	CARDINAL
922	(five)	CARDINAL
1237	(ben, ben)	PERSON
2051	(monica)	PERSON
818	(morning)	TIME
757	(one)	CARDINAL
438	(one)	CARDINAL
2024	(richard, dorfman)	PERSON
18	(eighteenth, century)	DATE
1919	(first)	ORDINAL
1417	(Emily)	PERSON
1175	(ross)	PERSON
1305	(Five)	CARDINAL
2188	(ross)	PERSON
1155	(one)	CARDINAL
1631	(ross)	PERSON
54	(ross)	PERSON

Nltk results

	Entities	Labels
97	Lift	PERSON
772	God All	PERSON
134	Hey Hey Well	PERSON
51	Disneyland	GPE
635	Did	PERSON
522	Bye Bye Bye	PERSON
742	Heh Y	PERSON
547	Jen Jen	PERSON
556	Freemont	GPE
359	Andy	PERSON
566	Hmmm	PERSON
696	Green	GPE
691	Ross Yeah Hey	PERSON
637	Robert Oh	PERSON
645	Okay Look	PERSON
712	Huh Well	PERSON
288	Ehh Oh	PERSON