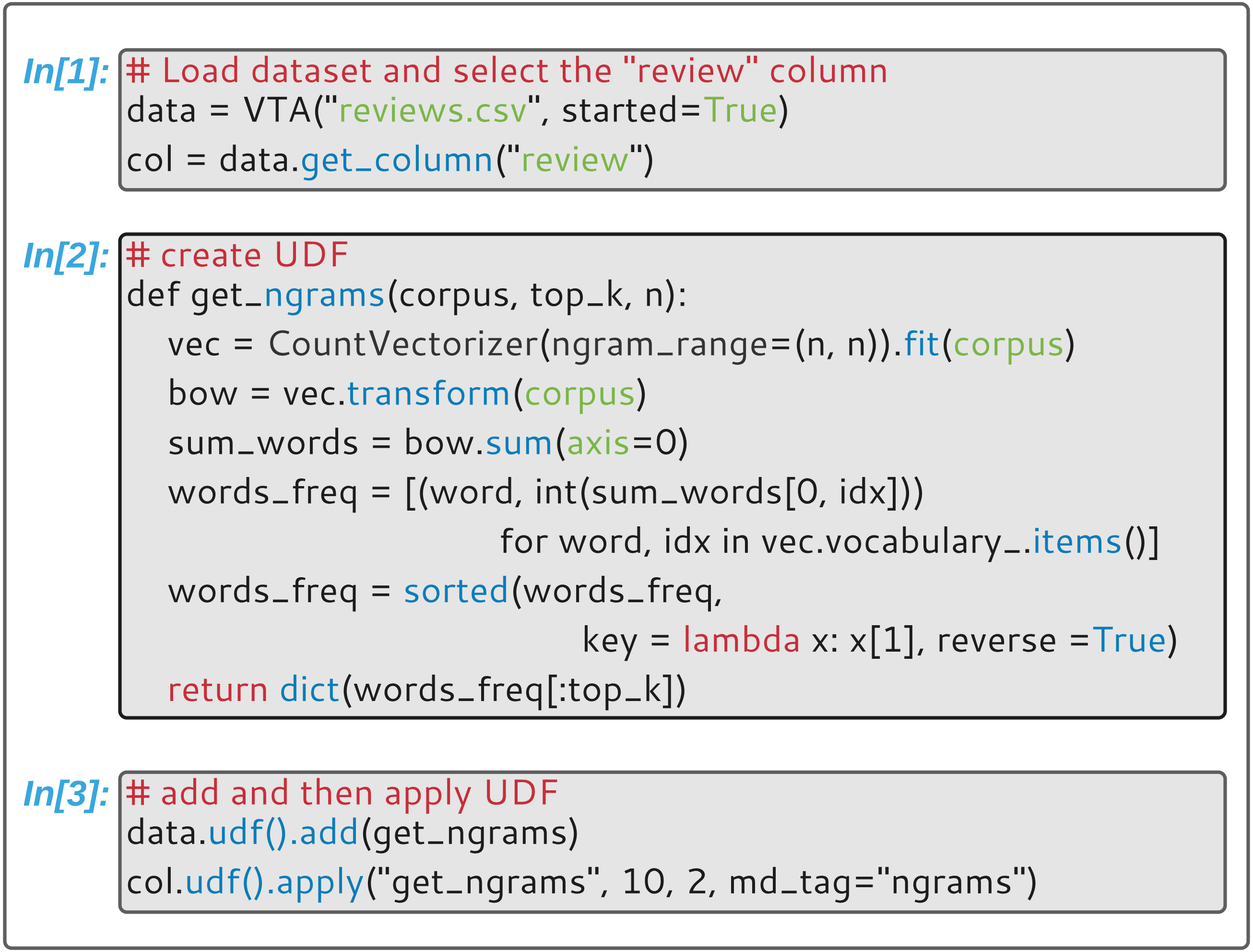

```
In[1]:
# Load dataset and select the "review" column
data = VTA("reviews.csv", started=True)
col = data.get_column("review")
```

```
In[2]:
# create UDF
def get_ngrams(corpus, top_k, n):
    vec = CountVectorizer(ngram_range=(n, n)).fit(corpus)
    bow = vec.transform(corpus)
    sum_words = bow.sum(axis=0)
    words_freq = [(word, int(sum_words[0, idx]))
                            for word, idx in vec.vocabulary_.items()]
    words_freq = sorted(words_freq,
                             key = lambda x: x[1], reverse =True)
    return dict(words_freq[:top_k])
```

```
In[3]:
# add and then apply UDF
data.udf().add(get_ngrams)
col.udf().apply("get_ngrams", 10, 2, md_tag="ngrams")
```