

Pre-ass. Session

12.

```
len (df[(df.comp_wat_hpn == "" or
         df.comp_wat_hpn == " " ]])
```

```
24 pos_tag = []
    nouns = []
```

```
for complaint in tqdm(dftr_clean.lemmatized_complaint):
```

```
    pos = ''
```

```
    noun = ''
```

```
    proc_complaint = nlp(complaint)
```

```
    for tok in proc_complaint:
```

```
        pos += tok.pos + ' '
```

```
        if tok.pos_ in str(('NOUN', 'PROPN')):
```

```
            noun += (tok.lemma_.lower()) + ' '
```

```
    pos_tags.append(pos.strip())
```

```
    nouns.append(noun.strip())
```

34

```
vectorizer = CV(ngram_range=(1,1), analyzer='word')  
word_vect = vectorizer.fit_tr(df_for_clean[comp_name_only])  
freqs = sum(word_vect).toarray()[0]
```

```
un_df = pd.DataFrame(freqs, index=vectorizer.get_feature_names(),  
                      cols=['freq'])
```

```
35. un_df.s_v(by='freq', as=F).head(30)
```

_____ X _____