

## CSE 158 HW 4

Q1:

```
(4587, 'with a'),  
(2595, 'in the'),  
(2245, 'of the'),  
(2056, 'is a'),  
(2033, 'on the')
```

nombre of unique bigrams: 182246

Q2:

MSE: 0.343153014061

Q3:

MSE 0.289047333034

Q4:

Negative: ['sort of', 'water', 'corn', 'the background', 'straw']

Positive: ['sort', 'a bad', 'of these', 'not bad', 'the best']

Q5:

IDF:

```
idf foam : -1.13786862069
idf smell : -0.537901618865
idf banana : -1.67778070527
idf lactic : -2.92081875395
idf tart : -1.80687540165
```

TF-IDF in first review

```
'foam': -2.2757372413739256,
'smell': -0.5379016188648442,
'banana': -3.355561410532161,
'lactic': -5.84163750790475,
'tart': -1.8068754016455382
```

Q6:

```
cosine similariy: 0.106130241679
```

Q7:

```
beer review most similar to first :
beerId: 52211
profileName: Heatwave33
```

Q8:

```
MSE: 0.278759560078
```