

```
#MARK ROSATO - HW3
```

```
#IMPORT PACKAGES
```

```
import urllib
```

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
from itertools import groupby
```

```
#IMPORT SCRABBLE DICTIONARY
```

```
data = urllib.urlopen("http://www.puzzlers.org/pub/wordlists/ospd.txt")
```

```
#CONVERT IT TO A LIST
```

```
scrabble = list(set(line.strip() for line in data))
```

```
#QUESTION 1
```

```
#GROUP ANAGRAMS
```

```
anagrams = [list(group) for key,group in groupby(sorted(scrabble,key=sorted),sorted)]
```

```
#NUMBER OF ANAGRAMS
```

```
scrabble_anagrams = [s for s in anagrams if len(s) >= 2]
```

```
#PRINT NUMBER OF ANAGRAMS
```

```
print len(scrabble_anagrams)
```

```
#QUESTION 2
```

```
#FIND LARGEST ANAGRAM SET
```

```
print max(anagrams, key=len)
```

```
#FIND SIZE OF LARGEST ANAGRAM SET
```

```
print len(max(anagrams, key=len))
```

```
#QUESTION 3
```

```
#CONVERT LIST FROM WORDS TO FREQUENCIES
```

```
a5 = [len(list(group)) for key, group in groupby(sorted(scrabble,key=sorted),sorted)]
```

```
#SORT FREQUENCIES
```

```
a5.sort()
```

```
#MAKE HISTOGRAM
```

```
n_groups = 12
```

```
frequencies = [len(list(group)) for key, group in groupby(a5)]
```

```
fig, ax = plt.subplots()
```

```
index = np.arange(n_groups)
```

```
bar_width = 0.5
```

```
opacity = 0.4
```

```
error_config = {'ecolor': '0.3'}
```

```
rects = plt.bar(index, frequencies, bar_width,  
                alpha=opacity,
```

```

        color='b',
        error_kw=error_config,
        label='Frequencies')

def autolabel(rects):
    # attach some text labels
    for ii,rect in enumerate(rects):
        height = rect.get_height()
        plt.text(rect.get_x()+rect.get_width()/2., 1.02*height, '%d'%int(height),
                 ha='center', va='bottom')
    autolabel(rects)

plt.xlabel('Anagram Size')
plt.ylabel('Frequency')
plt.title('Anagram Size with Frequency')
plt.xticks(index + bar_width, ('1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12'))
plt.legend()

plt.tight_layout()
plt.show()

#QUESTION 4
#UNIQUE ANAGRAM IDENTIFIER
unique_anagram_identifier = [x[0] for x in anagrams]
number_words_in_anagram = [len(list(group)) for key, group in
groupby(sorted(scrabble,key=sorted),sorted)]
actual_words = [' '.join([str(c) for c in lst]) for lst in anagrams]

#COMBINE INTO A LIST OF TUPLES
combine = zip(unique_anagram_identifier,number_words_in_anagram,actual_words)

#SQL UPLOAD AND COMMIT
import sqlite3
conn = sqlite3.connect('anagrams2.db')
c = conn.cursor()
c.execute("DROP TABLE Q4")
c.execute("CREATE TABLE Q4 (unique_anagram_identifier TEXT PRIMARY KEY,
number_words_in_anagram INT, actual_words TEXT)")
c.executemany('INSERT INTO Q4 VALUES (?,?)', combine)
conn.commit()

#FIND NUMBER OF UNIQUE ANAGRAMS IN THE WORD LIST
for row in c.execute("SELECT DISTINCT count(*) from Q4
                     where number_words_in_anagram &gt;= 2"):
    print row
</pre

```

#QUESTION 5

#FIND NUMBER OF WORDS IN LARGEST ANAGRAM

```
for row in c.execute("select distinct max(number_words_in_anagram), count(*) FROM Q4
                    group by number_words_in_anagram"):
    print row
```

#FIND ANAGRAM WITH LARGEST NUMBER OF WORDS IN THE WORD LIST

```
for row in c.execute("select distinct unique_anagram_identifier FROM Q4
                    where number_words_in_anagram = 12"):
    print row
```

#QUESTION 6

#IMPORT MW DICTIONARY AND CONVERT IT TO A LIST

```
lines = [line.strip() for line in open('/Users/mrosato/Dropbox (2U)/SMUDS/Database
Management/wordsnew.txt')]
mwmw = list(set(line.strip() for line in lines))
```

#FIND ANAGRAMS

```
mw_anagrams = [list(group) for key,group in groupby(sorted(mwmw,key=sorted),sorted)]
```

#DATABASE COLUMNS

```
mw_unique_anagram_identifier = [x[0] for x in mw_anagrams]
mw_number_of_words_in_anagram = [len(list(group)) for key,group in
groupby(sorted(mwmw,key=sorted),sorted)]
mw_actual_words = [' '.join([str(d) for d in lst]) for lst in mw_anagrams]
```

#COMBINE INTO A LIST OF TUPLES

```
mw_combine =
zip(mw_unique_anagram_identifier,mw_number_of_words_in_anagram,mw_actual_words)
```

#SQL UPLOAD AND COMMIT

```
import sqlite3
conn = sqlite3.connect('mw_anagrams2.db')
c = conn.cursor()
c.execute("DROP TABLE MW")
c.execute("CREATE TABLE MW (mw_unique_anagram_identifier TEXT PRIMARY KEY,
mw_number_words_in_anagram INT, mw_actual_words TEXT)")
c.executemany('INSERT INTO MW VALUES (?,?,?)', mw_combine)
conn.commit()
```

#FIND NUMBER OF UNIQUE ANAGRAMS IN THE WORD LIST

```
for row in c.execute("SELECT DISTINCT count(*) from MW
                    where mw_number_words_in_anagram >= 2"):
    print row
```

#FIND NUMBER OF WORDS IN LARGEST ANAGRAM

```
for row in c.execute("select distinct max(mw_number_words_in_anagram), count(*) FROM
```

MW

```
        group by mw_number_words_in_anagram"):
    print row
```

#FIND ANAGRAM WITH LARGEST NUMBER OF WORDS IN THE WORD LIST

```
for row in c.execute("select distinct mw_unique_anagram_identifier FROM MW
                    where mw_number_words_in_anagram = 15"):
    print row
```