

```
In [1]: # Practica 1: Analisis de grandes volúmenes de datos

# Autor: Alfonso Reyna Cerón

import matplotlib.pyplot as plt

# Abrir el archivo de texto en modo lectura
text_file = open("palabras.txt", "r")
```

Leo el archivo de texto y asigno su contenido a la variable text

```
In [2]: text = text_file.read()
```

Creo un arreglo de todas las palabras contenidas en el archivo

```
In [3]: word_array = text.split('\n')
```

Cantidad total de palabras

```
In [4]: # Print count of words
total_count = len(word_array)

print(total_count)
```

113810

Esta función determina si una palabra tiene o no la letra e incluida y de ser así aumenta el contador de palabras con e y si no aumenta el contador de las palabras que no lo tiene

Para cada palabra también evalúa si su longitud es mayor que 20

```
In [5]: # Two variables control variables
count_words_that_contain_e = 0
count_words_that_does_not_contain_e = 0
words_bigger_than_20 = 0

for word in word_array:
    if "e" in word:
        count_words_that_contain_e += 1
    else:
        count_words_that_does_not_contain_e += 1
    if len(word) > 20:
        words_bigger_than_20 += 1

print("Palabras que contienen e")
print(count_words_that_contain_e)
print("Palabras que no contienen e")
print(count_words_that_does_not_contain_e)
print("Palabras con longitud mayor a 20")
print(words_bigger_than_20)
```

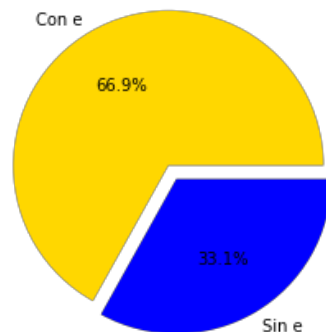
```
Palabras que contienen e
76168
Palabras que no contienen e
37642
Palabras con longitud mayor a 20
3
```

Creo dos labels para la grafica y le doy el total de contadores a plt como parametro

```
In [6]: # Data to plot
labels = 'Con e', 'Sin e'
sizes = [count_words_that_contain_e, count_words_that_does_not_contain_e]
colors = ['gold', 'blue']
explode = (0.1, 0) # explode 1st slice

# Plot
plt.pie(sizes, explode=explode, labels=labels, colors=colors, autopct='%1.1f%%')

plt.axis('equal')
plt.show()
```



Creo un diccionario de longitudes a las cuales se le asigna el contador de elementos para cada una, si la llave existe en el diccionario aumento el contador y si no creo la llave y le asigno el valor 1

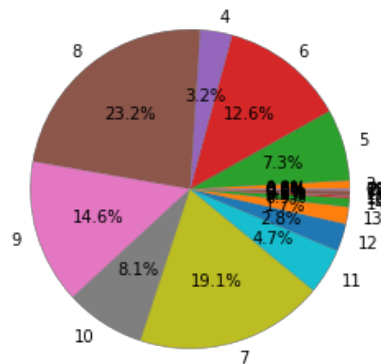
```
In [7]: # Graficar y agrupar palabras por longitud
length_array = {}

for word in word_array:
    length = len(word)
    if length in length_array:
        length_array[length] += 1
    else:
        length_array[length] = 1
```

Uso las llaves como labels y los valores como datos para generar otra grafica de pie

```
In [18]: labels = list(length_array.keys())
        sizes = list(length_array.values())

        plt.pie(sizes, labels=labels, autopct='%1.1f%%')
        plt.axis('equal')
        plt.show()
```



```
In [ ]:
```