Text Analysis

For this project I chose to analyze *The Picture of Dorian Gray* by Oscar Wilde. I built a TextAnalyzer class using Python and used it to find some interesting facts about the book. I implemented all of the methods outlined in the prompt for this project. For details on how the class methods work please see documents.txt. The following are various types of data which were drawn from the book using the TextAnalyzer class.

Number of words in the book: 79,088

Number of unique words: 7,702

Most frequent words: [*frequency, word*]
[[3351, 'the'], [2140, 'of'], [2082, 'and'], [2010, 'to'], [1688, 'I'], [1563, 'a'], [1284, 'that'], [1147, 'in'], [1085, 'you'], [1059, 'was'], [1040, 'he'], [958, 'it'], [935, 'his'], [893, 'is'], [819, 'had'], [659, 'him'], [649, 'with'], [552, 'have'], [550, 'for'], [531, 'at']]

Most Frequent words (ignoring the 100 most common English words): [*frequency, word*]
[[522, 'me'], [408, 'Dorian'], [243, 'Lord'], [229, 'am'], [223, 'Henry'], [215, 'life'], [194, "don't"], [186, 'Gray'], [171, 'Harry'], [170, 'man'], [163, 'never'], [159, 'into'], [153, 'Basil'], [153, 'must'], [130, 'own'], [125, 'think'], [123, 'things'], [121, 'us'], [119, 'seemed'], [119, 'very']]

Most Frequent words (ignoring the 300 most common English words): [*frequency, word*]
 [[408, 'Dorian'], [243, 'Lord'], [229, 'am'], [223, 'Henry'], [186, 'Gray'], [171, 'Harry'], [159, 'into'], [153, 'Basil'], [123, 'things'], [119, 'seemed'], [117, 'something'], [115, 'its'], [108, 'eyes'], [105, 'himself'], [105, 'love'], [101, 'quite'], [100, 'really'], [99, 'away'], [94, 'cried'], [94, 'looked']]

20 of the least frequent words: [*frequency, word*]
[[1, 'unable'], [1, 'winged'], [1, 'hair;'], [1, 'primarily'], [1, 'lap'], [1, 'waggons'], [1, 'luck;'], [1, 'mint'], [1, 'fateful'], [1, 'snatched'], [1, 'fulfilled'], [1, 'fight'], [1, 'retreat'], [1, 'articles'], [1, 'narrow'], [1, 'household'], [1, 'waxy'], [1, 'Trunks'], [1, 'heart-broken'], [1, 'rules']]

Chapter frequency of the word "Dorian":
[0, 22, 43, 9, 24, 1, 26, 25, 30, 30, 14, 10, 23, 15, 25, 22, 22, 12, 25, 28, 2]
The word 'Dorian' appears in chapter 0 0 times, in chapter 1 22 times, in chapter 2 43 times, etc.

Chapter(s) of quote "Dorian Gray": [*all of the chapters the quote appears in*]
 [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]
Chapter(s) of quote "Dorian Gray is": [1, 6]

Chapter(s) of quote "Dorian Gray is to me simply a motive in art.": [1]

A sentence generated with words from the book:
The duchess with words had seen any attention to begin with which it is sheer brute force but don't leave.

Closest matching quote for "eternal pathos human tragedy": [*quote, chapter*]
['eternal pathos of the human tragedy that it sought to s', 11]

All of the sentences that begin with "Basil": [*sentence, chapter*]
[["Basil would have helped him to resist Lord Henry's influence, and the still more poisonous influences that came from his own temperament ", 10], ['Basil would be sure to do that ', 8], ['Basil Hallward had left the house at eleven ', 13], ['Basil Hallward leaped to his feet and began to applaud ', 7], ['Basil had been with him part of the time, and had gone wild over Tintoret ', 14], ['Basil had painted the portrait that had marred his life ', 20], ['Basil had said things to him that were unbearable, and that he had yet borne with patience ', 20], ['Basil was really rather dull ', 19], ["Basil Hallward's compliments had seemed to him to be merely the charming exaggeration of friendship ", 2]]

All of the sentences that begin with "Basil had": [*sentence, chapter*]
[['Basil had been with him part of the time, and had gone wild over Tintoret ', 14], ['Basil had painted the portrait that had marred his life ', 20], ['Basil had said things to him that were unbearable, and that he had yet borne with patience ', 20]]

All of the sentences that begin with "Basil had painted": [*sentence, chapter*]
[['Basil had painted the portrait that had marred his life ', 20]]

TextAnalyzer can also be used to create a plot of the number of times that a word is used in each chapter throughout the book. In this example we can see that the word 'love' is used frequently at the beginning of the book and infrequently at the end, while the use of the word 'death' has the opposite trend.

Frequency of Words