A2 Annie_Fan

October 9, 2020

1 Project 2: Text Analysis

[1]:

```
# import packages
import numpy as np
import pandas as pd
from nltk.sentiment.vader import SentimentIntensityAnalyzer
import requests
import re
from urllib.parse import urlparse
import urllib.robotparser
from bs4 import BeautifulSoup

# This code checks the robots.txt file
def canFetch(url):

    parsed_uri = urlparse(url)
    domain = '{uri.scheme}://{uri.netloc}/'.format(uri=parsed_uri)

    rp = urllib.robotparser.RobotFileParser()
    rp.set_url(domain + "/robots.txt")
    try:
        rp.read()
        canFetchBool = rp.can_fetch("*", url)
    except:
        canFetchBool = None

    return canFetchBool
```

2 Pride and Prejudice analysis:

Find out main characters through counting occurance of words.

[ ]:

```
#Read the text file, and remove some unuseful words to make the counting␣
↪process more efficient.

f=[x.strip('\n').strip(',') for x in open('pride and prejudice.txt',␣
↪encoding='utf8').readlines()]
```

```
str=''
for i in f:
    if i!='':
        str=str+i
str=str.replace('Chapter','').replace('and','').replace('the','').
 ↪replace('to','').replace('was','').\
    replace('that','').replace('not','').replace('with','').replace('but','').
 ↪replace('you','').replace(',','').\
    replace('.','').replace(';','').replace('!','')
str
```

[3]:
```
#function for counting occurance of words
def count(str, word):
    lst=str.split(" ")
    length=len(lst)
    count = 0
    for i in range(0, length):
        if (word == lst[i]):
           count = count + 1
    return count
```

[4]:
```
#Put the output of the function above into a dictionary and sort it
dict={}
for word in str.split(' '):
    if word not in dict:
        dict[word]=count(str, word)
dict_sorted=sorted(dict.items(), key=lambda x: x[1], reverse = True)
dict_sorted
```

[4]:
```
[('', 78887),
 ('of', 3681),
 ('her', 2079),
 ('in', 2000),
 ('a', 1937),
 ('I', 1730),
 ('she', 1358),
 ('be', 1242),
 ('it', 1221),
 ('his', 1172),
 ('had', 1143),
 ('as', 1122),
 ('he', 1082),
 ('for', 1027),
 ('is', 838),
 ('have', 826),
 ('Mr', 766),
 ('at', 752),
 ('him', 702),
 ('on', 683),
 ('by', 633),
 ('all', 595),
 ('Elizabeth', 584),
 ('my', 580),
 ('or', 564),
 ('so', 561),
 ('were', 558),
 ('which', 537),
 ('could', 513),
 ('been', 512),
 ('from', 476),
 ('y', 472),
 ('very', 468),
 ('would', 463),
 ('no', 437),
 ('ir', 406),
 ('m', 406),
 ('r', 399),
 ('will', 398),
 ('me', 388),
 ('this', 386),
 ('what', 385),
 ('said', 377),
 ('such', 370),
 ('an', 348),
 ('Darcy', 346),
 ('Mrs', 338),
 ('are', 334),
 ('can', 320),
 ('more', 319),
 ('much', 315),
 ('must', 311),
 ('am', 311),
 ('She', 304),
 ('do', 302),
 ('"I', 295),
 ('any', 293),
 ('out', 289),
 ('when', 287),
 ('than', 281),
 ('Bennet', 276),
 ('who', 273),
 ('re', 272),
 ('Miss', 270),
 ('The', 267),
 ('But', 262),
 ('Jane', 254),
 ('if', 247),
 ('should', 245),
 ('one', 242),
 ('did', 242),
 ('Bingley', 240),
 ('know', 231),
 ('herself', 221),
 ('before', 217),
 ('has', 213),
 ('soon', 212),
 ('never', 212),
 ('He', 211),
 ('only', 208),
 ('think', 203),
 ('though', 203),
 ('It', 200),
 ('some', 200),
 ('time', 198),
 ('might', 198),
 ('may', 197),
 ('we', 191),
 ('most', 184),
 ('now', 183),
 ('every', 183),
 ('little', 182),
 ('own', 179),
 ('being', 174),
 ('sister', 170),
 ('far', 167),
 ('well', 166),
 ('good', 165),
 ('make', 164),
 ('how', 163),
 ('after', 162),
 ('hing', 162),
 ('again', 162),
 ('shall', 158),
 ('Wickham', 155),
 ('n', 154),
 ('see', 149),
 ('Collins', 149),
 ('say', 145),
 ('first', 142),
 ('family', 142),
 ('great', 140),
 ('o', 138),
 ('dear', 136),
 ('man', 133),
 ('Lady', 130),
 ('however', 127),
 ('ng', 126),
 ('two', 126),
 ('You', 125),
 ('made', 125),
 ('give', 124),
 ('ever', 123),
 ('Lydia', 121),
 ('up', 121),
 ('day', 120),
 ('"', 120),
 ('about', 118),
 ('always', 117),
 ('hope', 117),
 ('himself', 116),
 ('last', 115),
 ('They', 114),
 ('many', 114),
 ('away', 113),
 ('us', 113),
 ('thought', 111),
 ('room', 110),
 ('letter', 110),
 ('Carine', 108),
 ('mor', 105),
 ('its', 104),
 ('saw', 102),
 ('enough', 101),
 ('way', 101),
 ('felt', 99),
 ('friend', 98),
 ('replied', 97),
 ('go', 97),
 ('wish', 97),
 ('long', 97),
 ('quite', 96),
 ('house', 95),
 ('sure', 92),
 ('cried', 91),
 ('And', 91),
 ('came', 90),
 ('having', 90),
 ('often', 89),
 ('manner', 89),
 ('pleasure', 88),
 ('over', 87),
 ('till', 86),
 ('really', 86),
 ('better', 85),
 ('whom', 85),
 ('Longbourn', 85),
 ('feelings', 84),
 ('heard', 84),
 ('where', 84),
 ('done', 84),
 ('love', 83),
 ('"You', 83),
 ('Her', 83),
 ('believe', 82),
 ('myself', 81),
 ('come', 80),
 ('subject', 80),
 ('Gardiner', 80),
 ('even', 79),
 ('whole', 78),
 ('take', 77),
 ('less', 77),
 ('In', 77),
 ('anything', 76),
 ('morning', 75),
 ('aunt', 75),
 ('like', 74),
 ('daughter', 74),
 ('does', 74),
 ('happy', 74),
 ('down', 73),
 ('refore', 73),
 ('ladies', 73),
 ('His', 73),
 ('looked', 73),
 ('seen', 73),
 ('just', 71),
 ('still', 71),
 ('few', 71),
 ('If', 71),
 ('our', 70),
 ('evening', 70),
 ('Nerfield', 69),
 ('sisters', 69),
 ('same', 69),
 ('ld', 68),
 ('When', 68),
 ('something', 68),
 ('hear', 68),
 ('attention', 68),
 ('found', 68),
 ('happiness', 68),
 ('next', 67),
 ('My', 67),
 ('place', 66),
 ('Lizzy', 66),
 ('wards', 66),
 ('present', 66),
 ('received', 66),
 ('both', 66),
 ('Charlotte', 66),
 ('went', 66),
 ('tell', 65),
 ('upon', 65),
 ('opinion', 65),
 ('Kitty', 65),
 ('left', 65),
 ('speak', 65),
 ('seemed', 65),
 ('nor', 64),
 ('marriage', 64),
 ('least', 64),
 ('world', 63),
 ('part', 63),
 ('added', 63),
 ('each', 63),
 ('answer', 62),
 ('"Oh', 62),
 ('Lucas', 62),
 ('while', 62),
 ('Colonel', 62),
 ('"And', 62),
 ('certainly', 61),
 ('passed', 61),
 ('yet', 61),
 ('home', 61),
 ('ger', 61),
 ('Project', 60),
 ('almost', 60),
 ('gone', 60),
 ('moment', 60),
 ('off', 60),
 ('seeing', 60),
 ('leave', 60),
 ('immediately', 59),
 ('indeed', 59),
 ('se', 59),
 ('whose', 59),
 ('wn', 59),
 ('between', 59),
 ('conversation', 59),
 ('kind', 59),
 ('work', 59),
 ('means', 58),
 ('known', 57),
 ('find', 57),
 ('began', 57),
 ('__', 57),
 ('eir', 57),
 ('going', 57),
 ('coming', 57),
 ('married', 56),
 ('knew', 56),
 ('given', 56),
 ('look', 56),
 ('because', 56),
 ('affection', 56),
 ('character', 55),
 ('once', 55),
 ('behaviour', 55),
 ('rar', 55),
 ('bror', 55),
 ('three', 55),
 ('uncle', 55),
 ('"', 54),
 ('gave', 54),
 ('"But', 53),
 ('perhaps', 53),
 ('those', 53),
 ('anor', 53),
 ('ok', 53),
 ('able', 53),
 ('get', 52),
 ('since', 52),
 ('mind', 52),
 ('party', 52),
 ('As', 52),
 ('visit', 51),
 ('against', 51),
 ('reason', 51),
 ('life', 50),
 ('return', 50),
 ('certain', 50),
 ('wher', 50),
 ('Meryn', 50),
 ('walk', 50),
 ('side', 50),
 ('woman', 49),
 ('general', 49),
 ('Bingley’s', 49),
 ('idea', 49),
 ('London', 49),
 ('person', 49),
 ('eyes', 49),
 ('There', 49),
 ('possible', 49),
 ('Pemberley', 49),
 ('daughters', 48),
 ('here', 48),
 ('"It', 48),
 ('Sir', 48),
 ('ors', 48),
 ('friends', 48),
 ('acquaintance', 47),
 ('object', 47),
 ('suppose', 47),
 ('Rosings', 47),
 ('"My', 46),
 ('This', 46),
 ('What', 46),
 ('regard', 46),
 ('hardly', 46),
 ('course', 46),
 ('let', 45),
 ('returned', 45),
 ('word', 45),
 ('half', 45),
 ('continued', 45),
 ('people', 45),
 ('perfectly', 45),
 ('lady', 44),
 ('settled', 44),
 ('walked', 44),
 ('help', 44),
 ('We', 44),
 ('want', 43),
 ('ought', 43),
 ('business', 43),
 ('manners', 43),
 ('turned', 43),
 ('Darcy’s', 43),
 ('short', 43),
 ('scarcely', 43),
 ('point', 43),
 ('honour', 42),
 ('spoke', 42),
 ('To', 42),
 ('rself', 42),
 ('power', 42),
 ('"Yes', 42),
 ('carriage', 42),
 ('wife', 41),
 ('girls', 41),
 ('back', 41),
 ('read', 41),
 ('called', 41),
 ('At', 41),
 ('husb', 41),
 ('heart', 41),
 ('ill', 41),
 ('through', 41),
 ('longer', 41),
 ('How', 40),
 ('best', 40),
 ('expected', 40),
 ('agreeable', 40),
 ('met', 40),
 ('asked', 40),
 ('everything', 40),
 ('civility', 40),
 ('marry', 40),
 ('terms', 39),
 ('assure', 39),
 ('impossible', 39),
 ('else', 39),
 ('former', 39),
 ('right', 39),
 ('de', 39),
 ('thing', 38),
 ('likely', 38),
 ('William', 38),
 ('account', 38),
 ('dare', 38),
 ('rest', 38),
 ('sat', 38),
 ('Hertfordshire', 38),
 ('talked', 38),
 ('spirits', 38),
 ('convinced', 38),
 ('Elizabeth’s', 38),
 ('pride', 38),
 ('talking', 38),
 ('feel', 38),
 ('country', 38),
 ('sort', 38),
 ('sister’s', 38),
 ('write', 38),
 ('Lydia’s', 38),
 ('under', 37),
 ('set', 37),
 ('A', 37),
 ('wished', 37),
 ('afraid', 37),
 ('brought', 37),
 ('cousin', 37),
 ('different', 37),
 ('situation', 37),
 ('doubt', 37),
 ('fortune', 36),
 ('occasion', 36),
 ('glad', 36),
 ('pleased', 36),
 ('entered', 36),
 ('also', 36),
 ('feeling', 36),
 ('After', 36),
 ('stay', 36),
 ('believed', 36),
 ('Mary', 35),
 ('gentlemen', 35),
 ('near', 35),
 ('girl', 35),
 ('expect', 35),
 ('sense', 35),
 ('beyond', 35),
 ('Forster', 35),
 ('thous', 34),
 ('mselves', 34),
 ('years', 34),
 ('making', 34),
 ('amiable', 34),
 ('engaged', 34),
 ('silence', 34),
 ('wishes', 34),
 ('usual', 34),
 ('it"', 33),
 ('invitation', 33),
 ('hsome', 33),
 ('advantage', 33),
 ('door', 33),
 ('days', 33),
 ('care', 33),
 ('wonder', 33),
 ('night', 33),
 ('neir', 33),
 ('Gutenberg-tm', 33),
 ('determined', 32),
 ('merely', 32),
 ('him"', 32),
 ('sorry', 32),
 ('minutes', 32),
 ('afterwards', 32),
 ('particularly', 32),
 ('deal', 32),
 ('surprise', 32),
 ('sometimes', 32),
 ('On', 32),
 ('head', 32),
 ('length', 32),
 ('taken', 31),
 ('"What', 31),
 ('talk', 31),
 ('meet', 31),
 ('sensible', 31),
 ('dinner', 31),
 ('"If', 31),
 ('Hurst', 31),
 ('speaking', 31),
 ('Their', 31),
 ('satisfied', 31),
 ('appeared', 31),
 ('necessary', 31),
 ('real', 31),
 ('surprised', 31),
 ('instantly', 31),
 ('answered', 31),
 ('comfort', 31),
 ('name', 31),
 ('appear', 31),
 ('Wickham’s', 31),
 ('fine', 30),
 ('consider', 30),
 ('men', 30),
 ('ball', 30),
 ('ten', 30),
 ('countenance', 30),
 ('air', 30),
 ('put', 30),
 ('receive', 30),
 ('exactly', 30),
 ('resolved', 30),
 ('Bourgh', 30),
 ('ladyship', 30),
 ('Fitzwilliam', 30),
 ('four', 29),
 ('underst', 29),
 ('why', 29),
 ('asnishment', 29),
 ('dance', 29),
 ('Bennet’s', 29),
 ('obliged', 29),
 ('arrival', 29),
 ('gentleman', 29),
 ('particular', 29),
 ('consequence', 29),
 ('equal', 29),
 ('opportunity', 29),
 ('saying', 29),
 ('appearance', 29),
 ('lost', 29),
 ('full', 29),
 ('end', 28),
 ('respect', 28),
 ('mention', 28),
 ('kindness', 28),
 ('ask', 28),
 ('already', 28),
 ('early', 28),
 ('everybody', 28),
 ('looking', 28),
 ('mentioned', 28),
 ('alone', 28),
 ('offer', 28),
 ('nger', 28),
 ('case', 28),
 ('allow', 28),
 ('joined', 28),
 ('society', 28),
 ('whatever', 28),
 ('assured', 28),
 ('anyone', 27),
 ('five', 27),
 ('poor', 27),
 ('mean', 27),
 ('information', 27),
 ('highly', 27),
 ('credit', 27),
 ('With', 27),
 ('easily', 27),
 ('things', 27),
 ('words', 27),
 ('attachment', 27),
 ('face', 27),
 ('officers', 27),
 ('week', 27),
 ('knowing', 27),
 ('call', 27),
 ('pain', 27),
 ('satisfaction', 27),
 ('truth', 26),
 ('neighbourhood', 26),
 ('year', 26),
 ('"How', 26),
 ('news', 26),
 ('following', 26),
 ('second', 26),
 ('new', 26),
 ('walking', 26),
 ('Your', 26),
 ('Do', 26),
 ('Jane’s', 26),
 ('liked', 26),
 ('allowed', 26),
 ('hours', 26),
 ('attentions', 26),
 ('Phillips', 26),
 ('supposed', 26),
 ('table', 26),
 ('forward', 26),
 ('smile', 26),
 ('absence', 26),
 ('circumstances', 26),
 ('money', 26),
 ('beauty', 25),
 ('giving', 25),
 ('children', 25),
 ('delight', 25),
 ('reply', 25),
 ('keep', 25),
 ('meant', 25),
 ('followed', 25),
 ('easy', 25),
 ('admiration', 25),
 ('during', 25),
 ('disposition', 25),
 ('got', 25),
 ('expressed', 25),
 ('anxious', 25),
 ('interest', 25),
 ('turn', 25),
 ('thus', 25),
 ('compliment', 25),
 ('mor’s', 25),
 ('concern', 25),
 ('wholly', 25),
 ('bear', 25),
 ('temper', 24),
 ('among', 24),
 ('paid', 24),
 ('imagine', 24),
 ('drew', 24),
 ('Oh', 24),
 ('greater', 24),
 ('small', 24),
 ('observed', 24),
 ('Had', 24),
 ('entirely', 24),
 ('attended', 24),
 ('silent', 24),
 ('h', 24),
 ('serious', 24),
 ('wanted', 24),
 ('change', 24),
 ('ago', 24),
 ('Collins’s', 24),
 ('persuaded', 24),
 ('use', 23),
 ('considered', 23),
 ('hearing', 23),
 ('delighted', 23),
 ('send', 23),
 ('acquainted', 23),
 ('hopes', 23),
 ('fear', 23),
 ('several', 23),
 ('pretty', 23),
 ('said:', 23),
 ('me"', 23),
 ('remained', 23),
 ('curiosity', 23),
 ('No', 23),
 ('anybody', 23),
 ('spite', 23),
 ('nature', 23),
 ('That', 23),
 ('doing', 23),
 ('thoughts', 23),
 ('prevented', 23),
 ('inquiries', 23),
 ('probably', 23),
 ('conduct', 23),
 ('exceedingly', 23),
 ('especially', 23),
 ('state', 23),
 ('living', 23),
 ('match', 23),
 ('large', 22),
 ('"No', 22),
 ('depend', 22),
 ('sake', 22),
 ('matter', 22),
 ('spent', 22),
 ('Derbyshire', 22),
 ('dislike', 22),
 ('praise', 22),
 ('"He', 22),
 ('natural', 22),
 ('pounds', 22),
 ('used', 22),
 ('gratitude', 22),
 ('degree', 22),
 ('ice', 22),
 ('seems', 22),
 ('latter', 22),
 ('meeting', 22),
 ('frequently', 22),
 ('hour', 22),
 ('master', 22),
 ('fair', 22),
 ('Brighn', 22),
 ('works', 22),
 ('fixed', 21),
 ('marrying', 21),
 ('desire', 21),
 ('live', 21),
 ('intended', 21),
 ('promised', 21),
 ('above', 21),
 ('company', 21),
 ('event', 21),
 ('pleasing', 21),
 ('listened', 21),
 ('importance', 21),
 ('favour', 21),
 ('express', 21),
 ('"That', 21),
 ('whenever', 21),
 ('expression', 21),
 ('repeated', 21),
 ('"This', 21),
 ('directly', 21),
 ('relations', 21),
 ('voice', 21),
 ('sent', 21),
 ('except', 21),
 ('wrote', 21),
 ('inclination', 21),
 ('happened', 21),
 ('receiving', 21),
 ('acknowledged', 20),
 ('_her_', 20),
 ('humour', 20),
 ('understing', 20),
 ('knowledge', 20),
 ('pleasant', 20),
 ('equally', 20),
 ('library', 20),
 ('proud', 20),
 ('sitting', 20),
 ('_I_', 20),
 ('address', 20),
 ('greatest', 20),
 ('endeavour', 20),
 ('forced', 20),
 ('proper', 20),
 ('attempt', 20),
 ('ready', 20),
 ('breakfast', 20),
 ('favourite', 20),
 ('Let', 20),
 ('particulars', 20),
 ('furr', 20),
 ('need', 20),
 ('written', 20),
 ('Hunsford', 20),
 ('connection', 20),
 ('cause', 20),
 ('journey', 20),
 ('thinking', 19),
 ('design', 19),
 ('times', 19),
 ('dancing', 19),
 ('instead', 19),
 ('eldest', 19),
 ('estate', 19),
 ('hoped', 19),
 ('join', 19),
 ('fancy', 19),
 ('common', 19),
 ('circumstance', 19),
 ('public', 19),
 ('danger', 19),
 ('resolution', 19),
 ('affected', 19),
 ('late', 19),
 ('strong', 19),
 ('agreement', 19),
 ('trouble', 19),
 ('far’s', 19),
 ('regret', 19),
 ('niece', 19),
 ('electronic', 19),
 ('none', 18),
 ('fortnight', 18),
 ('chance', 18),
 ('persuade', 18),
 ('Well', 18),
 ('excellent', 18),
 ('intelligence', 18),
 ('delightful', 18),
 ('unless', 18),
 ('"She', 18),
 ('admire', 18),
 ('Maria', 18),
 ('charming', 18),
 ('mistaken', 18),
 ('superior', 18),
 ('smiled', 18),
 ('removed', 18),
 ('civil', 18),
 ('purpose', 18),
 ('vain', 18),
 ('evident', 18),
 ('slight', 18),
 ('question', 18),
 ('remember', 18),
 ('light', 18),
 ('play', 18),
 ('reached', 18),
 ('pause', 18),
 ('avoid', 18),
 ('taking', 18),
 ('regiment', 18),
 ('accepted', 18),
 ('anxiety', 18),
 ('leaving', 18),
 ('comprehend', 18),
 ('health', 18),
 ('engagement', 18),
 ('concerned', 18),
 ('future', 18),
 ('led', 18),
 ('capable', 18),
 ('nephew', 18),
 ('m"', 17),
 ('high', 17),
 ('addressed', 17),
 ('Not', 17),
 ('accept', 17),
 ('bring', 17),
 ('women', 17),
 ('resentment', 17),
 ('round', 17),
 ('ease', 17),
 ('perfect', 17),
 ('absolutely', 17),
 ('beg', 17),
 ('Eliza', 17),
 ('"Your', 17),
 ('open', 17),
 ('connections', 17),
 ('cold', 17),
 ('drawing-room', 17),
 ('wait', 17),
 ('proved', 17),
 ('choose', 17),
 ('loss', 17),
 ('kept', 17),
 ('summer', 17),
 ('son', 17),
 ('affectionate', 17),
 ('expectation', 17),
 ('prevent', 17),
 ('From', 17),
 ('Charlotte’s', 17),
 ('extraordinary', 16),
 ('all"', 16),
 ('suddenly', 16),
 ('forget', 16),
 ('For', 16),
 ('report', 16),
 ('fond', 16),
 ('step', 16),
 ('sight', 16),
 ('_me_', 16),
 ('likewise', 16),
 ('chose', 16),
 ('seem', 16),
 ('vanity', 16),
 ('form', 16),
 ('generally', 16),
 ('composure', 16),
 ('success', 16),
 ('pass', 16),
 ('companion', 16),
 ('indifference', 16),
 ('opened', 16),
 ('dine', 16),
 ('view', 16),
 ('politeness', 16),
 ('readily', 16),
 ('_she_', 16),
 ('spoken', 16),
 ('ne', 16),
 ('formed', 16),
 ('true', 16),
 ('conviction', 16),
 ('tried', 16),
 ('Carine’s', 16),
 ('became', 16),
 ('distress', 16),
 ('scheme', 16),
 ('scene', 16),
 ('months', 16),
 ('stairs', 16),
 ('ashamed', 16),
 ('Foundation', 16),
 ('Gutenberg', 15),
 ('share', 15),
 ('ignorant', 15),
 ('"Mr', 15),
 ('extremely', 15),
 ('lively', 15),
 ('window', 15),
 ('declared', 15),
 ('principal', 15),
 ('angry', 15),
 ('Such', 15),
 ('sit', 15),
 ('partner', 15),
 ('beautiful', 15),
 ('advice', 15),
 ('book', 15),
 ('related', 15),
 ('worth', 15),
 ('_r_', 15),
 ('disposed', 15),
 ('period', 15),
 ('promise', 15),
 ('felicity', 15),
 ('intention', 15),
 ('sir', 15),
 ('duty', 15),
 ('"Well', 15),
 ('motive', 15),
 ('miles', 15),
 ('chief', 15),
 ('trust', 15),
 ('assurance', 15),
 ('prevailed', 15),
 ('writing', 15),
 ('seated', 15),
 ('letters', 15),
 ('Saturday', 15),
 ('earnest', 15),
 ('weeks', 15),
 ('affair', 15),
 ('confess', 15),
 ('effect', 15),
 ('blame', 15),
 ('marked', 15),
 ('belief', 15),
 ('reasonable', 15),
 ('Long', 14),
 ('"Do', 14),
 ('objection', 14),
 ('says', 14),
 ('comes', 14),
 ('old', 14),
 ('consideration', 14),
 ('"We', 14),
 ('unable', 14),
 ('fortunate', 14),
 ('arrived', 14),
 ('Darcy"', 14),
 ('creature', 14),
 ('behind', 14),
 ('eye', 14),
 ('lerable', 14),
 ('nobody', 14),
 ('struck', 14),
 ('elegant', 14),
 ('asking', 14),
 ('thanks', 14),
 ('please', 14),
 ('greatly', 14),
 ('excuse', 14),
 ('influence', 14),
 ('lerably', 14),
 ('secure', 14),
 ('laugh', 14),
 ('"There', 14),
 ('eagerly', 14),
 ('sod', 14),
 ('Every', 14),
 ('wrong', 14),
 ('wishing', 14),
 ('distance', 14),
 ('winter', 14),
 ('interesting', 14),
 ('servant', 14),
 ('parted', 14),
 ('reading', 14),
 ('orwise', 14),
 ('housekeeper', 14),
 ('prospect', 14),
 ('"The', 14),
 ('_my_', 14),
 ('indifferent', 14),
 ('connected', 14),
 ('park', 14),
 ('dearest', 14),
 ('Parsonage', 14),
 ('visirs', 14),
 ('occurred', 14),
 ('disappointment', 14),
 ('misery', 14),
 ('sentence', 14),
 ('farr', 14),
 ('charge', 14),
 ('agreed', 13),
 ('compassion', 13),
 ('her"', 13),
 ('One', 13),
 ('meaning', 13),
 ('ngest', 13),
 ('assistance', 13),
 ('questions', 13),
 ('sting', 13),
 ('turning', 13),
 ('twice', 13),
 ('gratified', 13),
 ('learnt', 13),
 ('lived', 13),
 ('_his_', 13),
 ('All', 13),
 ('observation', 13),
 ('fact', 13),
 ('respectable', 13),
 …]
```

```
[13]:
```

```
#Even though some words are deleted, still there are a lot of uninteresting␣
 ↪words. This cell extracts some names.
names=['Elizabeth', 'Darcy', 'Bennet', 'Jane', 'Bingley', 'Wickham', 'Collins',␣
 ↪'Lydia', 'Carine', 'Longbourn', 'Gardiner']

result=[]
for i in dict_sorted:
    if i[0] in names:
        result.append(i)
result
```

[13]:

```
[('Elizabeth', 584),
 ('Darcy', 346),
 ('Bennet', 276),
 ('Jane', 254),
 ('Bingley', 240),
 ('Wickham', 155),
 ('Collins', 149),
 ('Lydia', 121),
 ('Carine', 108),
 ('Longbourn', 85),
 ('Gardiner', 80)]
```

According to the result above, Elizabeth is the name that appears the most (238 times more than the second most frequent name), and thus Elizabeth is highly likely to be the main character of the book. Darcy, the second most frequent name, is the next most important character. The names listed above are mostly main characters in the book, with some exception of names of places, such as Longbourn.

[ ]:

```

```