Consider a text document. (may be txt file / .doc file / .docx file}

A. Split the data linewise. Split further wordwise. Determine the frequency of the each word.

B. Remove the stop words.

C. Store the new text processed into a new file.

D. Find the frequency of the words.

E. Compare the size of the text file and plot using matplotlib.

**Self learning:** matplotlib, file operations, library support

**Consider the following Text data:**

**Sport** includes all forms of [competitive](https://en.wikipedia.org/wiki/Competition) [physical activity](https://en.wikipedia.org/wiki/Physical_activity) or [games](https://en.wikipedia.org/wiki/Game) which,[[1]](https://en.wikipedia.org/wiki/Sport#cite_note-sportaccord-1) through casual or organised participation, aim to use, maintain or improve physical ability and skills while providing enjoyment to participants, and in some cases, [entertainment](https://en.wikipedia.org/wiki/Entertainment) for spectators.[[2]](https://en.wikipedia.org/wiki/Sport#cite_note-council-2) Hundreds of sports exist, from those between single contestants, through to those with hundreds of simultaneous participants, either in [teams](https://en.wikipedia.org/wiki/Team) or competing as individuals. In certain sports such as [racing](https://en.wikipedia.org/wiki/Racing), many contestants may compete, simultaneously or consecutively, with one winner; in others, the contest (a *match*) is between two sides, each attempting to exceed the other. Some sports allow a "tie" or "draw", in which there is no single winner; others provide [tie-breaking methods](https://en.wikipedia.org/wiki/Tie-breaking_methods) to ensure one winner and one loser. A number of contests may be arranged in a [tournament](https://en.wikipedia.org/wiki/Tournament) producing a [champion](https://en.wikipedia.org/wiki/Champion). Many [sports leagues](https://en.wikipedia.org/wiki/Sports_league) make an annual champion by arranging games in a regular [sports season](https://en.wikipedia.org/wiki/Sports_season), followed in some cases by [playoffs](https://en.wikipedia.org/wiki/Playoffs).

Sport is generally recognised as system of activities which are based in physical [athleticism](https://en.wikipedia.org/wiki/Athleticism) or physical [dexterity](https://en.wikipedia.org/wiki/Dexterity), with the largest major competitions such as the [Olympic Games](https://en.wikipedia.org/wiki/Olympic_Games) admitting only sports meeting this definition,[[3]](https://en.wikipedia.org/wiki/Sport#cite_note-3) and other organisations such as the [Council of Europe](https://en.wikipedia.org/wiki/Council_of_Europe) using definitions precluding activities without a physical element from classification as sports.[[2]](https://en.wikipedia.org/wiki/Sport#cite_note-council-2) However, a number of competitive, but non-physical, activities claim recognition as [mind sports](https://en.wikipedia.org/wiki/Mind_sport). The International Olympic Committee (through [ARISF](https://en.wikipedia.org/wiki/Association_of_IOC_Recognised_International_Sports_Federations)) recognises both [chess](https://en.wikipedia.org/wiki/Chess) and [bridge](https://en.wikipedia.org/wiki/Contract_bridge) as *bona fide* sports, and [SportAccord](https://en.wikipedia.org/wiki/SportAccord" \o "SportAccord), the international sports federation association, recognises five non-physical sports: bridge, chess, [draughts](https://en.wikipedia.org/wiki/Draughts) (checkers), [Go](https://en.wikipedia.org/wiki/Go_(game)) and [xiangqi](https://en.wikipedia.org/wiki/Xiangqi" \o "Xiangqi),[[4]](https://en.wikipedia.org/wiki/Sport#cite_note-4)[[5]](https://en.wikipedia.org/wiki/Sport#cite_note-5) and limits the number of mind games which can be admitted as sports.[[1]](https://en.wikipedia.org/wiki/Sport#cite_note-sportaccord-1)

Sport is usually governed by a set of [rules](https://en.wikipedia.org/wiki/Regulation_of_sport) or [customs](https://en.wikipedia.org/wiki/Tradition), which serve to ensure fair competition, and allow consistent adjudication of the winner. Winning can be determined by physical events such as scoring [goals](https://en.wikipedia.org/wiki/Goal_(sport)) or crossing a line first. It can also be determined by judges who are scoring elements of the sporting performance, including objective or subjective measures such as technical performance or artistic impression.

Records of performance are often kept, and for popular sports, this information may be widely announced or reported in [sport news](https://en.wikipedia.org/wiki/Sports_journalism). Sport is also a major source of entertainment for non-participants, with [spectator sport](https://en.wikipedia.org/wiki/Spectator_sport) drawing large crowds to [sport venues](https://en.wikipedia.org/wiki/Sport_venue), and reaching wider audiences through [broadcasting](https://en.wikipedia.org/wiki/Broadcasting_of_sports_events). [Sport betting](https://en.wikipedia.org/wiki/Sport_betting) is in some cases severely regulated, and in some cases is central to the sport.

According to A.T. Kearney, a consultancy, the global sporting industry is worth up to $620 billion as of 2013.[[6]](https://en.wikipedia.org/wiki/Sport#cite_note-6) The world's most accessible and practised sport is [running](https://en.wikipedia.org/wiki/Running), while [association football](https://en.wikipedia.org/wiki/Association_football) is the most popular spectator sport.[[7]](https://en.wikipedia.org/wiki/Sport#cite_note-wa-7)