

## Answers

### § Task: Answer the following questions:

1. Which were the 5 most frequent words in all\_words.csv were most frequent?
2. Which were the 5 most frequent words in all\_uppercase\_words.csv.
3. Compare and contrast these top 5. Explain in 2-3 sentences what you observe about the similarities and differences.
4. In your own words, what were the most surprising parts of each list?

### Solutions

Below are the top 5 most frequent words in **all\_words.csv** and **all\_uppercase\_words.csv**

5 Most Frequent Words in 'all\_words.csv':

	Word	Frequency
3266	the	2504
2408	of	1859
2049	is	1316
3307	to	1232
621	and	986

5 Most Frequent Words in 'all\_uppercase\_words.csv':

	Word	Frequency
48	I	151
0	A	34
10	B	21
80	OF	20
18	CHAPTER	15

### Comparison

#### Similarities

Both lists contain common English words like the, of and to among the most frequent words, which suggests that these words are frequently used in the text.

#### Differences

The all\_uppercase\_words.csv list contains uppercase words, which may indicate the presence of proper nouns or significant terms in uppercase.

all\_words.csv includes lowercase words like is and and, which are not present in all\_uppercase\_words.csv.

The frequencies of the words in `all_words.csv` are relatively higher compared to `all_uppercase_words.csv`, indicating that common lowercase words are used more frequently in the text.

## Observations

The similarities between the two lists show that certain common words, both in uppercase and lowercase, are prevalent in the text.

The differences suggest that `all_uppercase_words.csv` likely contains proper nouns, acronyms, or terms emphasized in uppercase.

Additionally, the higher frequencies in `all_words.csv` indicate the importance of common lowercase words in the text.

## Surprising Parts

It's surprising that `I` is the most frequent word in `all_uppercase_words.csv`. This suggests that the text may contain a significant amount of first-person narrative or dialogue.

The presence of `CHAPTER` in `all_uppercase_words.csv` is unexpected, indicating that chapters or sections of the text are marked in uppercase. This is common in literary works but may not be seen in all types of text.

These surprising elements provide insights into the specific characteristics of the text being analyzed and hint at its content and structure.

## Analysis answers for all-ns-words-with-probability

1. How many unique non-stop words are in the text?

### Solution

2853

2. Which is the least probable word? (if there is a tie, please state the tie words)

### Solution (check `least_probable_words.csv`)

**Least Probable Word(s):** ACQUAINTANCE, ACTUAL, AGREEMENT, APPEARANCE, Acquaintance, Additional, Author, Awareness, BIBLIOGRAPHICAL, Bannerman, Books, Bradley, British, Broadly, CONSEQUENTIAL, CONTRACT, Campbell, Cantor, City, Compliance, Considered, Contact, Continental, Contributions, Copyright, Creating, DAMAGE, DESCRIPTION, DIRECT, DISCLAIMER, DISTRIBUTE, DISTRIBUTOR, DONATIONS, December, Defect, Defects, Degrees, Domestic, EIN, ERROR, EXISTENCE, EXPRESS, Earth, East, Email, England, English, Euclid, Europe, FALSEHOOD, FITNESS, FOUNDATION, Food, Future, Georg, George, Gilbert, Gordon, Granted, Greeks, Hart, Henry, Human, IDEALISM, III, IMPLIED, INCIDENTAL, INCLUDING, INDEMNITY, INDIRECT, INDUCTION, INTUITIVE, IRS, Idealists, Ides, Immanuel, Instances, Internal, International, Judgements, Judging, June, Kantian, Keener, Keynes, Königsberg, LIABILITY,

LIABLE, LIMITS, Lake, Language, Latin, Leaving, Library, Light, Locke, MERCHANTABILITY, March, Memories, Michael, Mission, Mississippi, Moore, Murray, NATURE, NEGLIGENCE, NOTE, NOTICE, Neglecting, Newton, North, OPINION, OWNER, Opposition, PARAGRAPH, PGLAF, PHILOSOPHICAL, POSSIBILITY, PREFACE, PRINCIPLES, PROBABLE, PUNITIVE, PURPOSE, Philosophical, Principles, Produced, Pronouns, Prussia, READ, REALITY, REFUND, REMEDIES, REPLACEMENT, Rationalists, Redistributing, Redistribution, Reflection, Refund, Release, Replacement, Returning, Revenue, Revolution, Roman, SEND, STRICT, Salt, Sceptics, Sections, Service, Sir, Space, Speaking, Special, Starting, Struldbugs, Swift, TRADEMARK, TRUTH, Times, Title, Truth, Truths, UT, Uneducated, Universals, Universe, University, Updated, VIII, Volunteers, War, Waterloo, West, Whitehead, XII, XIII, XIV, XV, York, abide, abstain, abstracted, abstractions, abstractly, absurdities, acceptance, accepts, accessed, accessible, accident, accidental, accidents, accompanied, accord, accounting, accurate, accusative, achieve, achieves, acquiesce, acquiescence, acquisition, activity, acute, adapts, adding, address, addresses, adduced, adequacy, adequately, admirable, admitting, admixture, advances, adversaries, advocate, advocates, advocating, aesthetic, aether, affairs, affections, affirmative, afforded, agent, agreeably, aim, aimed, akin, alien, allowed, allowing, alteration, altered, alternate, alternatives, alters, ambitious, amend, amply, analogous, analyse, analyses, anatomist, animal, announcement, anticipate, anticipating, anticipation, appearances, appearing, appetite, application, applied, apportions, apprehends, approaches, approaching, arduous, argues, arguing, arose, arrange, array, arrived, arriving, arrogant, artificial, ascertained, assassinated, assented, assigns, assimilate, assist, assumptions, assurance, assure, assured, astray, ate, atoms, attached, attack, attain, attainable, attested, attitude, attributing, avoided, avoiding, babies, baldness, banging, bar, barrier, bed, beg, begging, beginner, begs, begun, behave, behaviour, beleagured, bell, bewildering, bid, bigger, binary, binds, blank, bolder, bone, bored, borne, bounds, brick, briefest, brighter, brightness, broken, build, builder, buildings, calculate, calculated, calm, camps, candid, card, careful, carelessly, carry, catalogue, catalogued, catching, categories, causing, centuries, century, chance, changeable, changed, charitable, charities, checks, child, chimaera, choice, chosen, citizens, citizenship, city, clash, clashes, clean, clock, closed, closer, closes, clouds, coalesce, codes, cohere, coin, coins, collective, colony, combination, combine, coming, commercial, commit, communicate, community, comparative, compare, compared, comparing, competent, compilation, complement, completed, compressed, conceivable, conceivably, concentrate, concluded, concourse, condemns, conduct, confess, confine, confining, confirmation, confirmed, confirms, conflict, conformity, confused, confusing, confusions, confute, conjecture, connecting, connects, consent, consequences, consisted, constituting, construct, constructed, constructing, constructions, constructive, contemplated, contemplating, contemplative, contemptuously, contended, contention, contentions, continued, continuing, continuous, continuously, contradict, contradictory, contrast, contributed, contribution, contributions, convenient, converged, conversation, convert, convincing, cool, corporation, correlative, corresponded, corrupt, counterparts, countless, countries, cranny, creates, creation, criterion, criticisms, criticized, crudely, current, custom, customary, cut, cutting, damage, damaged, danger, dates, debate, debated, deceitful, deceive, deceptive, decides, deciding, decision, declared, deducing, deductible, deeds, deeply, defeated, defensible, defined, definiteness, definitions, deletions, deliberate, delightful, delusion, demanding, demonstrably, demonstrative, demonstratively, denies, dentist, dependence, depending, descend, describes, describing, descriptive, desert, deserts, deserves, deserving, designate, desirability, desirable, desiring, desirous, destroyed, destroys, destructive, detach, determines, development, developments, diary, dictionary, die, diminish, diminished, diminishes, dined, directions, directs, disappear, disappoint, disappointing, disclaim, disclaimers, discontinue, discuss, disease, disengaged, disk, dismiss, dispassionately, display, displayed, disputable, disputants, dissociated, distinctly, distinguishes, distort, distorts, diverge, divest, divide, divided, divisibility, dogmas, dogmatic, dogmatism, domestic, dominion, donation, donors, doubtless, downloading, drawn, dried, drifted, drive, drives, duly, easiest, eclipse, edition, educated, educational, effected, elapse, elapsed,

elect, elected, electric, elementary, elements, elucidation, email, embark, emerged, emerges, emitted, emitting, emphasized, emphatically, employee, employees, employs, engagement, engendered, enlarged, enlarges, enormously, enrich, ensuring, entails, entangled, enter, entertaining, enumerated, enumeration, enunciated, equal, equivocation, ergo, estimated, eternal, eternally, everyday, exact, examines, exceedingly, excluding, exclusion, exemplify, existent, expend, expense, explicitly, explored, exploring, exponents, exporting, expressions, extend, extends, extensive, extrinsic, fabric, facility, faculty, failing, fainter, faintness, fallacies, fallacy, fallible, falls, falsified, family, fancy, fast, fatal, favouritism, fear, features, fed, feeds, fetters, feverish, file, files, financial, fine, fixed, flash, foes, foolish, football, force, forces, forget, forgetting, forgo, forgotten, forks, formal, formally, formats, fortress, fortuitous, fortune, forwards, fostered, fourth, fragment, fragments, framework, frightened, fulfilment, fulfils, gain, garrison, generality, geography, geology, glance, glasses, globe, goals, goodwill, gradations, grain, grapple, grasp, gratefully, gratuitous, gravely, greens, groundless, grow, growing, guarded, guide, habits, habitual, habitually, halves, handbooks, happiness, hardnesses, harm, harmless, harmony, hates, hatred, hatreds, heartbeats, heaven, heavens, helpful, heretofore, hesitatingly, hides, hills, hinder, historian, historic, historically, honour, hoofs, hoped, hopeless, horrors, host, hot, hundreds, hypertext, identification, ignorance, ignorant, illumination, illusoriness, illusory, illustrated, illustrations, immutable, impaired, impairs, impartial, impartially, impenetrable, impersonal, impose, impression, imprisoned, inaccurate, inclination, includes, incompatible, inconsistencies, inconsistency, increase, increasing, incredulous, indefinitely, indemnify, independence, indestructible, indicating, inevitable, inexplicable, infallibility, infancy, infected, inferring, infinitesimal, inflected, inflections, influence, infringed, infringement, inhabitants, inhabited, inherent, innate, innocent, inoperative, inquire, inquired, insistent, insoluble, inspiration, instants, instincts, instrument, insubstantial, intelligent, intended, intently, interact, interferes, intermediary, internal, interpret, interpreted, interrelation, interrupted, introduce, intuitively, invalidity, invented, inventing, inventions, investigations, invites, irrefutable, isolation, joint, justly, killed, knits, knives, knowable, lapse, larger, launched, lay, learnt, leaving, legally, legitimately, level, liberating, liberation, liberator, liberty, library, licensed, lies, limitations, limiting, linked, lips, live, lived, locations, logician, looked, loose, loosen, lot, loud, lowest, luminously, lying, magnifies, maintain, maintaining, majority, manager, marching, marriage, match, maximum, meanings, melt, memories, mention, mercilessly, merest, merges, merits, metaphorically, midst, miles, mineralogist, minor, miracle, misery, misleadingness, mistake, mistakes, misunderstanding, mitigate, mode, modification, modifications, modify, momentary, month, mortality, moved, movements, multiplication, multiplicity, musical, mutually, mystic, mystical, mysticism, nation, native, nearest, nearness, negligible, network, newness, news, newsletter, newspapers, noises, nominative, nonproprietary, nook, noteworthy, noticing, notifies, notions, nouns, novelist, nowadays, nowhen, numerous, objectionable, objective, oblivious, obscure, observe, observing, obsolete, obstacle, obstacles, obstinate, obtuse, occupant, occupying, odd, office, oftener, omission, oneself, opaque, operated, operative, opportunities, opportunity, opposites, orator, orderly, ordnance, organized, organizing, organs, origin, originally, originator, outcome, outdated, outlines, outward, oval, owed, painfully, palpable, papers, paperwork, paradoxes, paradoxical, partially, party, passage, passes, pausing, pay, peace, peculiar, peculiarly, perfection, performances, periodic, permanence, perpetual, persons, perspective, persuaded, persuading, perturbed, phantasmagoria, phenomenon, philosophize, physics, physiological, physiology, pink, placing, plane, planet, plausibility, playing, poisonous, positively, possess, possessed, possession, poverty, powerlessness, powers, practice, practised, precisely, preferable, preparing, prescribe, presence, preservation, pressing, pressure, pressures, presumption, presupposed, pretty, prevent, primitive, proceeded, proceeds, processes, processing, producing, production, products, profess, profit, profitable, profited, profits, profoundest, progress, progressively, prohibition, prolonged, promote, promotion, pronounce, pronouns, proofread, proposed, proprietary, provision, provisionally, provisions, prudent, psychologically, psychology, pursue, pushed, puzzling, quibbling, range, ranging, rapping,

rationalist, rationality, readable, readers, realization, realized, reappear, receiving, recognizes, recombines, reconcile, reconstructed, recorded, redistribute, redistributing, redistribution, reduce, reducible, refined, reflecting, refutable, regress, regulating, rejection, relate, relational, reliable, religion, remarkably, remarks, remote, remoter, removal, removes, renamed, renounce, repeat, repeatedly, repetition, replaced, reported, reports, represent, representations, representative, representing, request, resemblances, residue, restate, resting, restricted, retain, retained, retort, returns, revealed, reversed, revert, review, rigid, rival, rob, robbing, roof, rooted, rotate, rotating, roundabout, rounded, rouse, roused, royalty, runs, samples, satisfaction, satisfying, save, scaffold, sceptic, sceptical, schoolmen, sea, searches, secures, seldom, sending, separated, sequel, serve, service, sets, seventeenth, severe, shadow, sharing, sharp, sheets, shiny, shock, short, shortly, simplify, situated, skeleton, sleeping, slowly, smelling, smells, sober, society, sold, solicitation, solid, solve, solves, solving, sophistry, soul, sour, south, specially, species, spectacles, speculative, speech, spoke, spoons, sprang, staff, stages, started, starting, starts, step, stone, stored, straightforward, strangeness, stranger, strangest, strike, strikes, striking, strives, striving, struck, structure, struggling, studies, sublime, subscribe, subsequent, subsequently, subsumed, subsumptions, success, successively, suffering, sufficed, suggestion, suggestions, sunset, supplied, surprised, surrender, surrounded, surveys, survive, surviving, suspended, swamp, sweet, synonymous, systematize, tablecloth, tactile, tap, task, tasting, taught, taxes, telegram, temper, temperaments, tempted, tending, tenet, term, texts, texture, theses, thesis, thralldom, throw, throws, thwarting, timeless, tincture, tolerably, tongue, toothache, topics, total, touches, traditional, trains, trammels, transcribe, transcription, transfer, transform, transitory, travels, treated, treating, treatment, triangularity, trifling, trotting, trouble, troubled, troubles, trusted, tuning, type, types, typical, tyranny, unaccustomed, unaffected, unattainable, uncertain, unchanged, unchanging, uncomfortable, uncommon, unconscious, undeniable, undeniably, underlie, underlies, undertake, undiminished, undue, unenforceability, unexpectedly, unexperienced, unimportant, unites, uniting, unlearn, unlink, unplausible, unprotected, unpublished, unreflectingly, unreflective, unsafe, unsolicited, unsolved, unsound, unsuspected, unusual, unwarrantable, unwise, updated, upset, upstairs, urges, usage, vagueness, vaguer, vain, valleys, variable, variations, variety, variously, varying, vast, veil, verification, version, viewing, vindicate, violates, violent, virtuous, virus, voice, voices, void, volume, volunteer, voyage, walk, walks, walled, walls, war, warranties, warrants, waste, wear, wearing, weather, weight, west, window, wise, wonderful, wooden, wore, worst, wrings, writers, yield

### **Minimum Probability: $8e-05$**

3. What observation can you make about the probabilities?

#### **Solution**

The probabilities of non-stop words vary, with some words being less probable than others. Some words may have very low probabilities.

4. Which sentence is more likely:

- a. If a belief is true, it can be deduced it is universal. (Probability: 0.00935 )
- b. Criticism of knowledge is counter to scientific results. (Probability: 0.02389 )

#### **Solution**

Sentence b is more likely because it has a higher cumulative probability of non-stop words.

## ANSWER THE FOLLOWING QUESTIONS AFTER LISTENING TO THE PODCAST

1. List 3 things that you learned from this podcast?

### Solution

**Python as a Dominant Language** - Python has become one of the most popular programming languages for data science and machine learning. Its simplicity and versatility make it accessible to data scientists, and it has a rich ecosystem of libraries and tools for various data-related tasks.

**Pandas for Data Manipulation** - Pandas is a powerful library in Python for data manipulation and analysis. It provides data structures like DataFrames and Series, which are essential for cleaning, transforming, and exploring data. Data scientists often use Pandas to prepare data for analysis.

**Scikit-Learn for Machine Learning** - Scikit-Learn is a popular machine learning library in Python. It offers a wide range of machine learning algorithms and tools for tasks such as classification, regression, clustering, and model evaluation. Data scientists use Scikit-Learn to build and evaluate predictive models.

**NumPy for Numerical Computing**: NumPy is a fundamental library for numerical computing in Python. It provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays efficiently.

**Editor likely used for data science** – Pycharm

**Notable API Package** – Transformers.

2. What is your reaction to the podcast? Pick at least one point brought up in the interview that you agree with and list your reason why.

### Solution

I agree with the point that Python's versatility and rich ecosystem makes it a powerful tool for data science.

Python's extensive libraries and packages make it easier to perform data analysis and machine learning tasks.

3. After listening to the podcast, do you think you are more informed about the importance of Python to Data Science? How? (Be brief – one sentence will suffice.)

### Solution

Yes, I am more informed about Python's importance to Data Science. The podcast highlighted Python's role in various aspects of data science, from data engineering to modeling and analysis.

4. List one surprising fact you learned from listening to this podcast.

**Solution**

Data scientists like working alone or in very small groups.