pandas
pandas is an open source data analysis library built on top of the Python programming language. The import pandas portion of the code tells Python to bring the pandas data analysis library into your current environment.

numpy
pandas is an open-source library built on top of numpy providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. It allows for fast analysis and data cleaning and preparation.

matplotlib
Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.

Seaborn
The seaborn package was developed based on the Matplotlib library. It is used to create more attractive and informative statistical graphics. While seaborn is a different package, it can also be used to develop the attractiveness of matplotlib graphics

world cloud
Word Clouds are visual displays of text data – simple text analysis. Word Clouds display the most prominent or frequent words in a body of text (such as a State of the Union Address). Typically, a Word Cloud will ignore the most common words in the language (“a”, “an”, “the” etc).

Counter
Counter is a sub-class that is used to count hashable objects. It implicitly creates a hash table of an iterable when invoked. elements() is one of the functions of Counter class, when invoked on the Counter object will return an itertool of all the known elements in the Counter object.

import nltk
The Natural Language Toolkit (NLTK) is a platform used for building Python programs that work with human language data for applying in statistical natural language processing (NLP). It contains text processing libraries for tokenization, parsing, classification, stemming, tagging and semantic reasoning.

nltk.tokensize
Tokenization breaks the raw text into words, sentences called tokens. These tokens help in understanding the context or developing the model for the NLP. The tokenization helps in interpreting the meaning of the text by analyzing the sequence of the words. Tokenization can be done to either separate words or sentences

nltk corpus
NLTK corpus readers. The modules in this package provide functions that can be used to read corpus files in a variety of formats. These functions can be used to read both the corpus files that are distributed in the NLTK corpus package, and corpus files that are part of external corpora.

sklearn

scikit-learn is probably the most useful library for machine learning in Python. The sklearn library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.