Przetwarzanie języka naturalnego (8 punktów)

Zadanie:

Przeprowadź analizę recenzji napisanych przez pracowników czołowych firm technologicznych w USA i dotyczących ich pracodawców. Dane do analizy zawarte są w pliku *employee-reviews.csv*.

Program musi zawierać następujące elementy:

1. Zmianę sposobu zapisu daty w kolumnie *dates* i ograniczenie się jedynie do roku, w którym recenzja została napisana. Jeśli zmiana spowoduje wystąpienie braków w danych powinny one zostać usunięte (1 punkt).
2. Obliczenie, ile recenzji ma każda firma z osobna i pokazanie wyniku na wykresie (słupkowym lub kołowym) (1 punkt).
3. Wizualizację średniej ocen (*overall-rating*) dla każdej firmy w latach (wykres liniowy) (1 punkt).
4. Określenie, ile recenzji zostało napisanych przez obecnych (*Current Employee*) a ile przez byłych pracowników (*Former Employee*) oraz wizualizację uzyskanego wyniku na wykresie kołowym (dla każdej firmy z osobna) (1 punkt).
5. Identyfikację najpopularniejszych słów kluczowych w komentarzach *pros* i *cons*. Należy ograniczyć się do 20 najpopularniejszych słów. (2 punkty).
6. Stworzenie chmury 20 najpopularniejszych słów dla każdej z firm (osobno dla komentarzy *pros* i *cons*) (2 punkty).

UWAGA: Program musi zawierać komentarze objaśniające kod.

Wskazówki

1. Aby zidentyfikować najpopularniejsze słowa kluczowe z tekstu komentarzy należy usunąć najpierw tzw. przerywniki (*stopwords*), czyli słowa, które nie niosą ze sobą znaczenia, ale są konieczne do stworzenia zdania (*a, and, the, will,* itp.). Można do tego wykorzystać klasę *stopwords* z modułu *nltk.* Z tekstu powinny być również usunięte znaki interpunkcyjne. Słowa powinny zostać poddane lematyzacji, czyli sprowadzone do swojej formy podstawowej (słownikowej) przy użyciu klasy *WordNetLemmatizer* z modułu *nltk*.
2. Stworzenie chmury słów jest możliwe z wykorzystaniem klasy *WordCloud* z modułu *nltk*.