

Analiza rozkładu słów w wybranych książkach kucharskich przy wykorzystaniu Prawa Zipfa i Rozkładu Pareto

inż. Bazan Michał
inż. Pawłoś Łukasz



Cele projektu

- Przeprowadzenie analizy statystycznej rozkładu słów w wybranych książkach w celu zrozumienia, czy rozkład ten jest zgodny z modelem Prawa Zipfa oraz Rozkładu Pareto.
- Zbadanie, czy istnieje zależność między częstością występowania słów a ich pozycją w rankingu popularności.

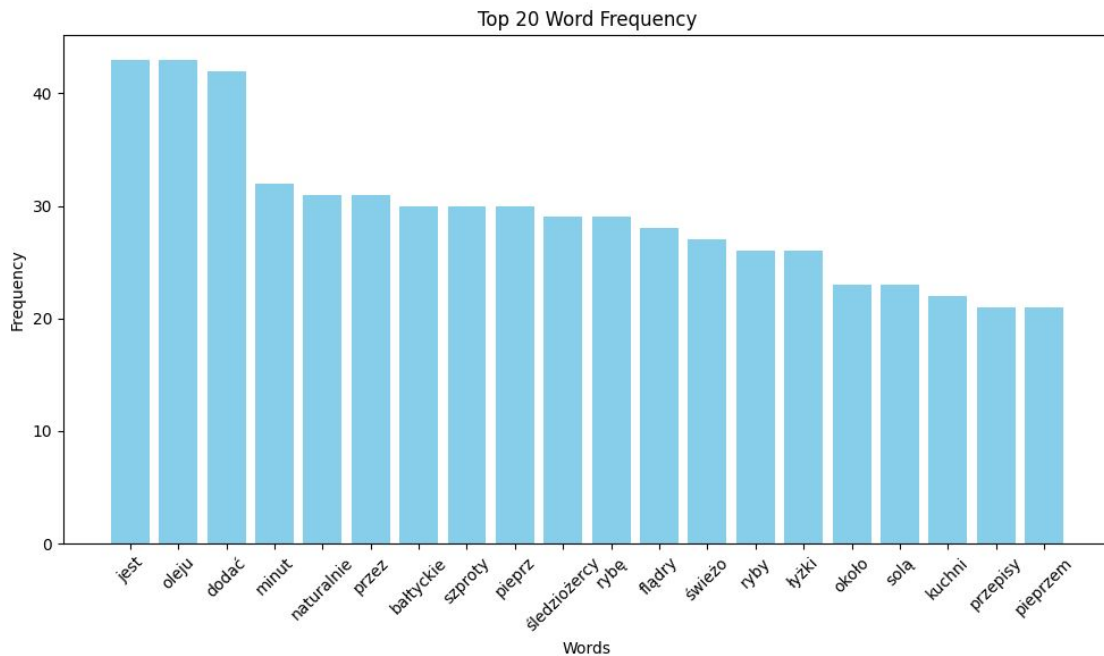


Obiekt badań

- *Smaki bałtyckie naturalnie* – Kołobrzaska Grupa Producentów Ryb sp. z o.o. ul. Węgorzowa 8, 78-100 Kołobrzeg © Kołobrzaska Grupa Producentów Ryb 2020
- *Kucharz Polski* – Marja Śleżańska, Księgarnia J. Leitgebera i S-ki, 1932 Poznań
- *Wędrówki kulinarne Roberta Makłowicza* – Robert Makłowicz

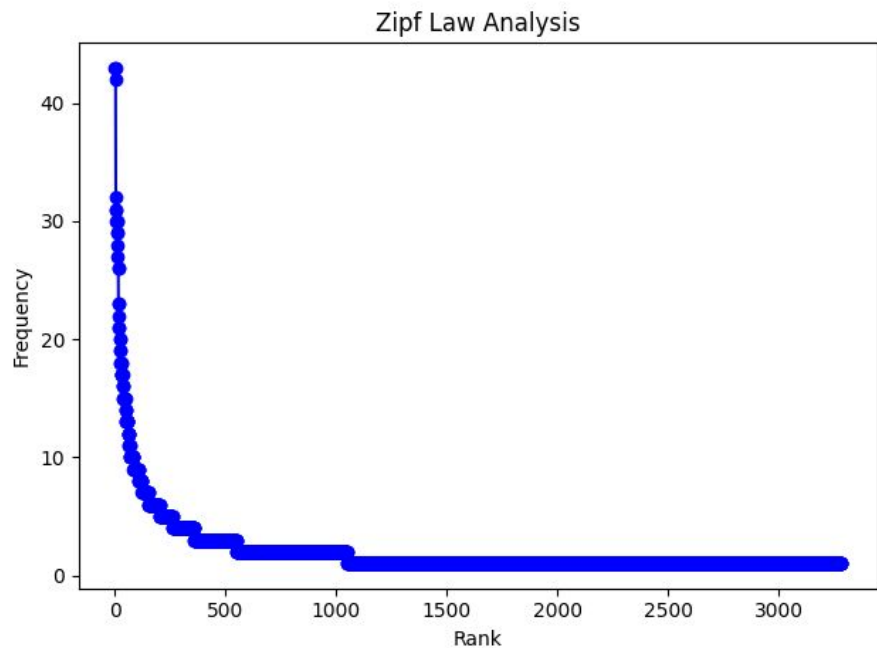


Smaki bałtyckie naturalnie – badania



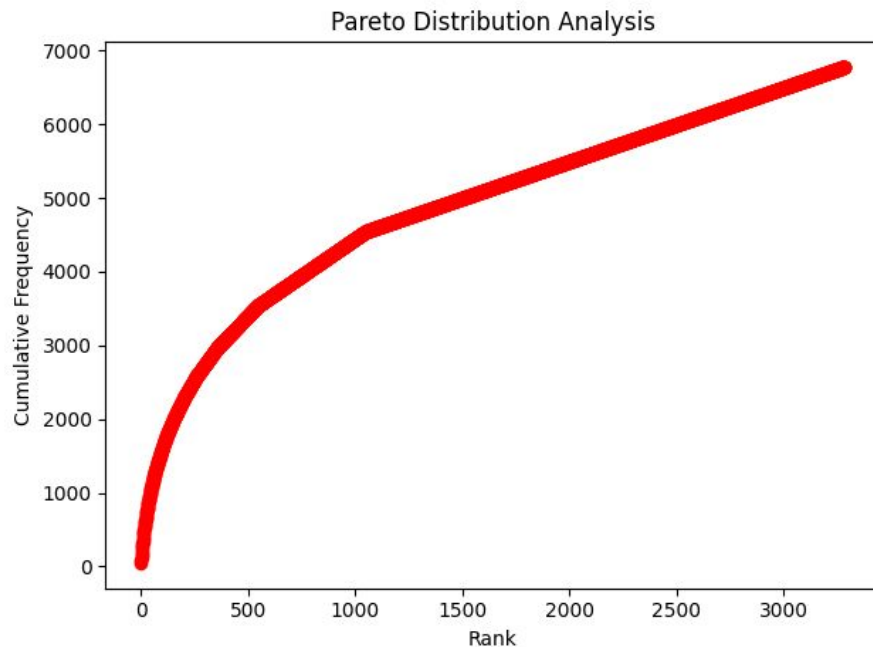


Smaki bałtyckie naturalnie – badania



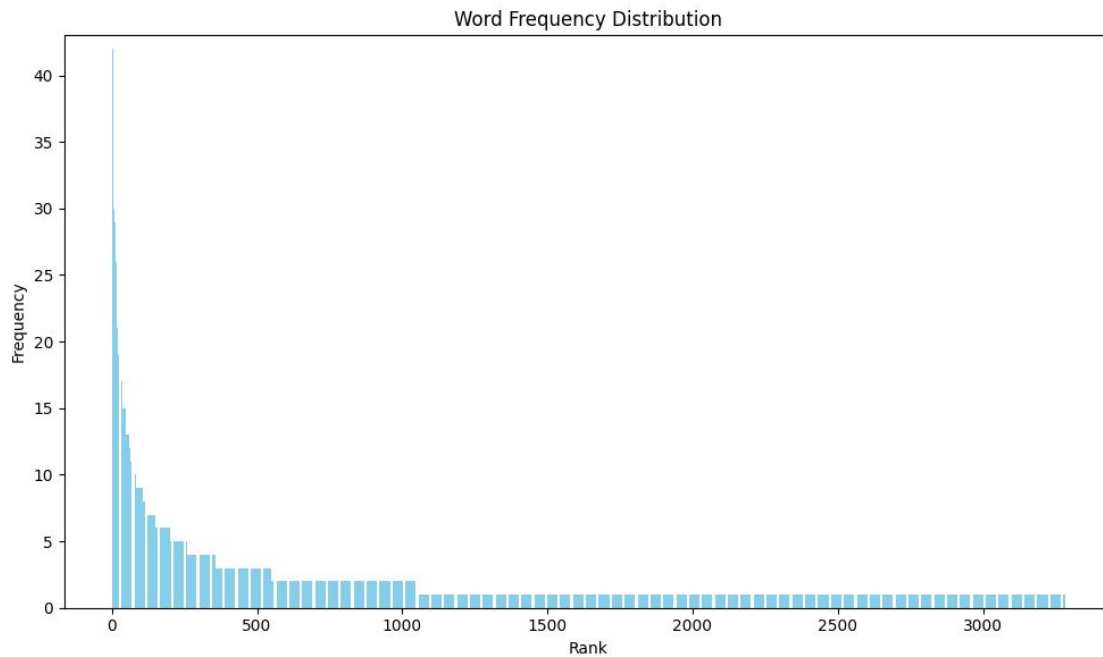


Smaki bałtyckie naturalnie – badania



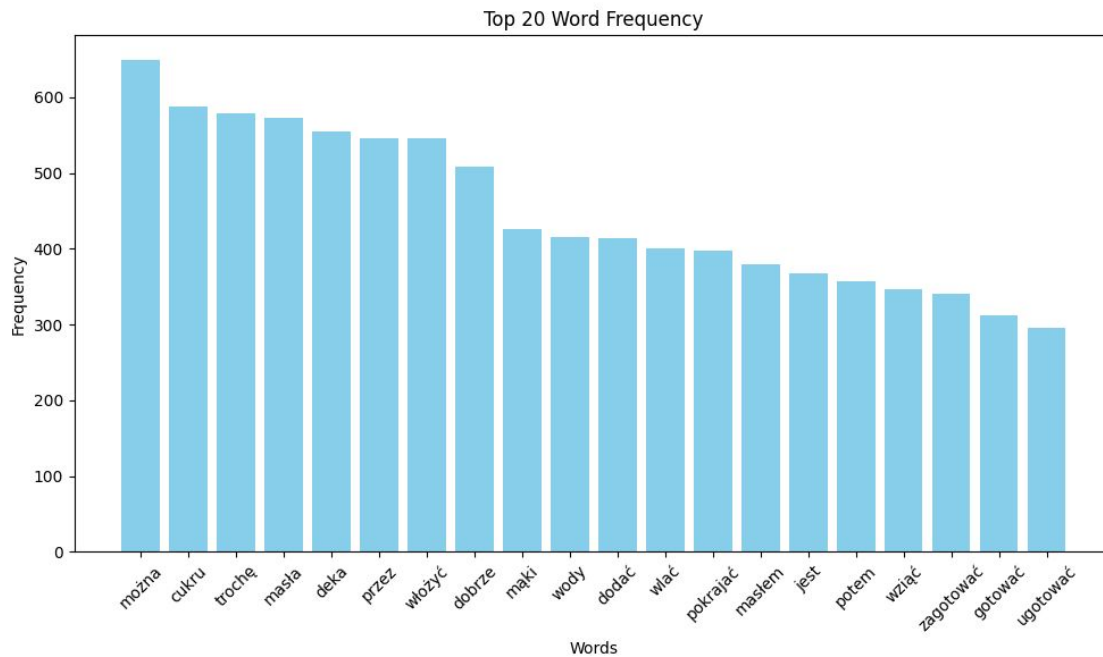


Smaki bałtyckie naturalnie – badania



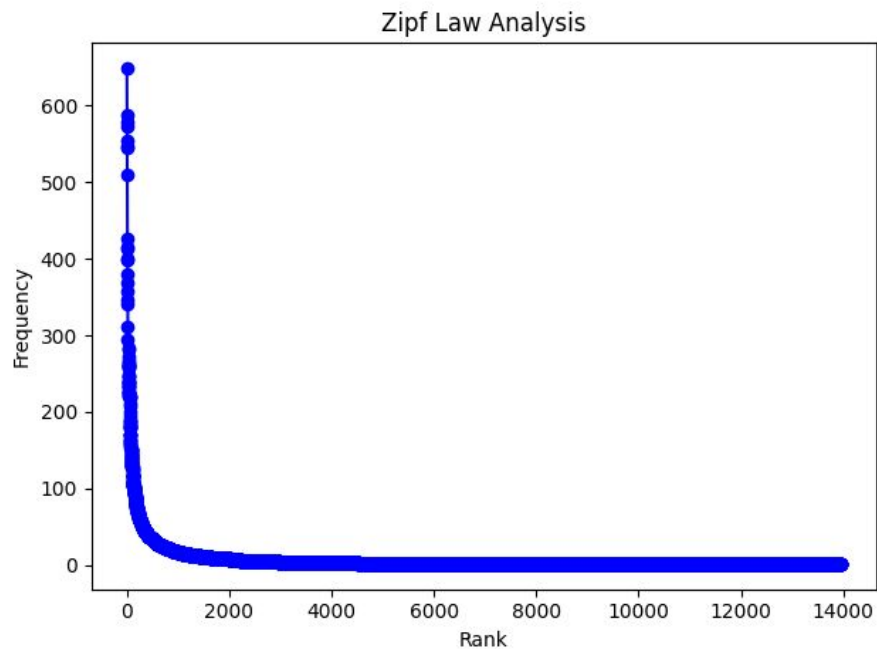


Kucharz Polski – badania



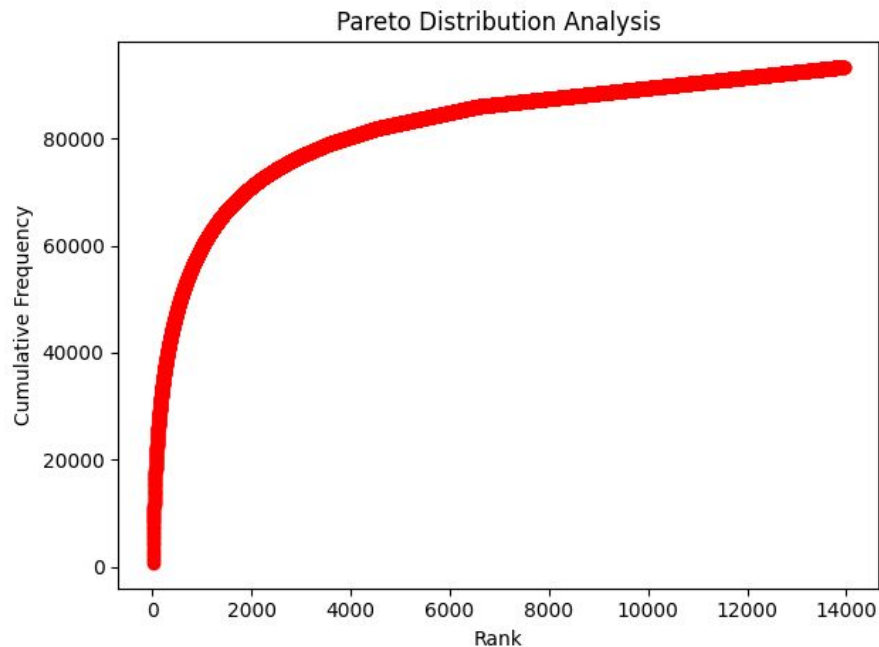


Kucharz Polski – badania



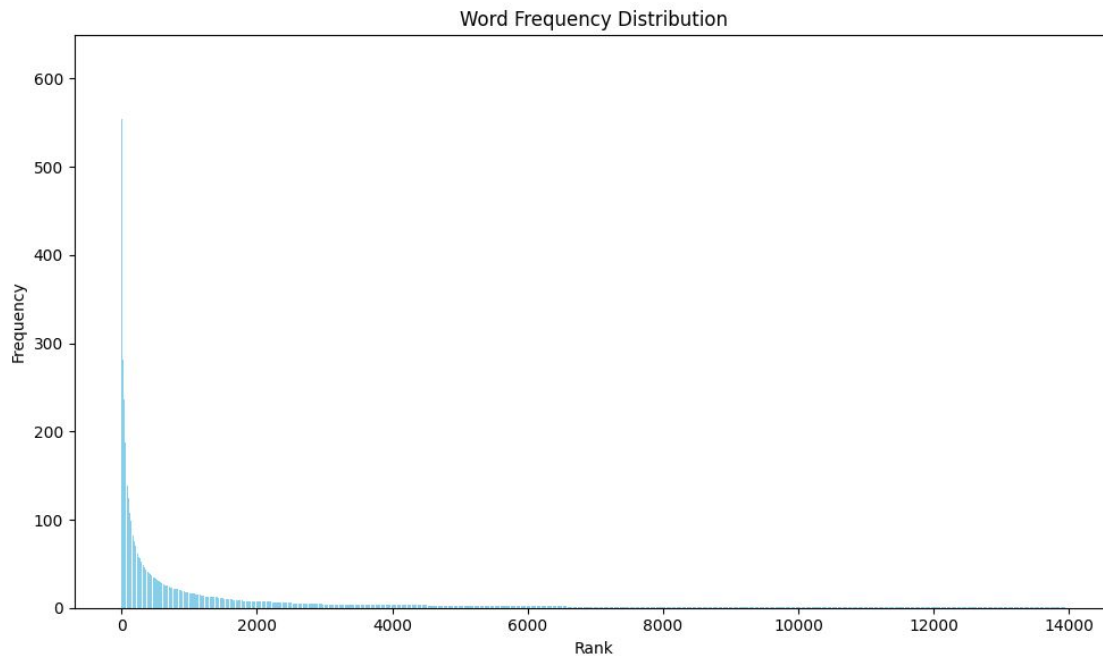


Kucharz Polski – badania



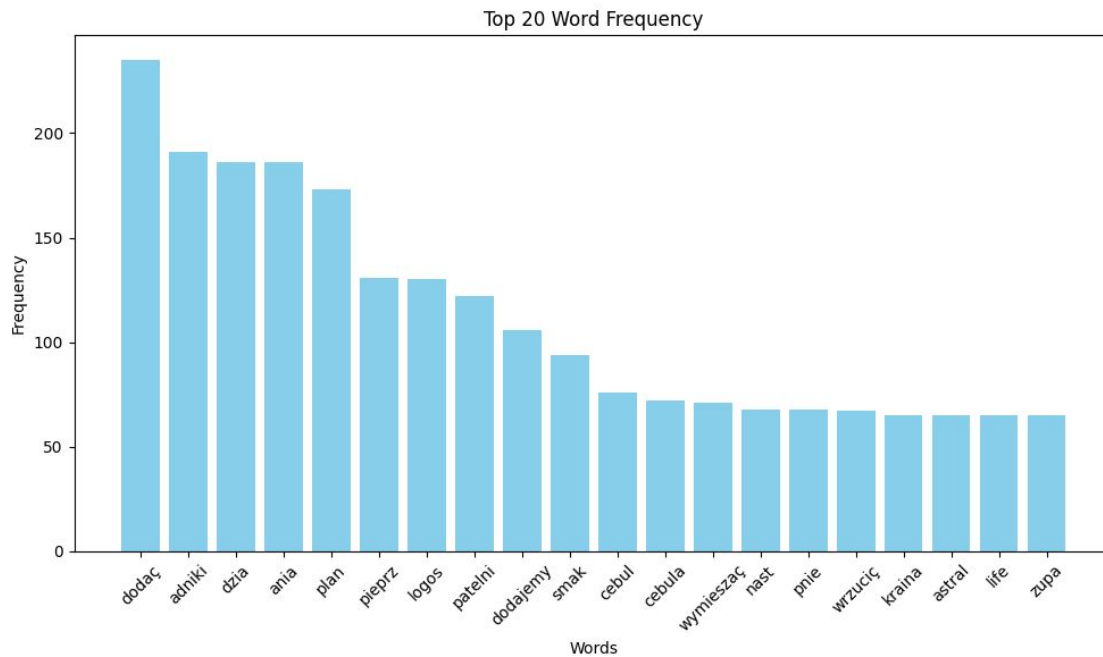


Kucharz Polski – badania



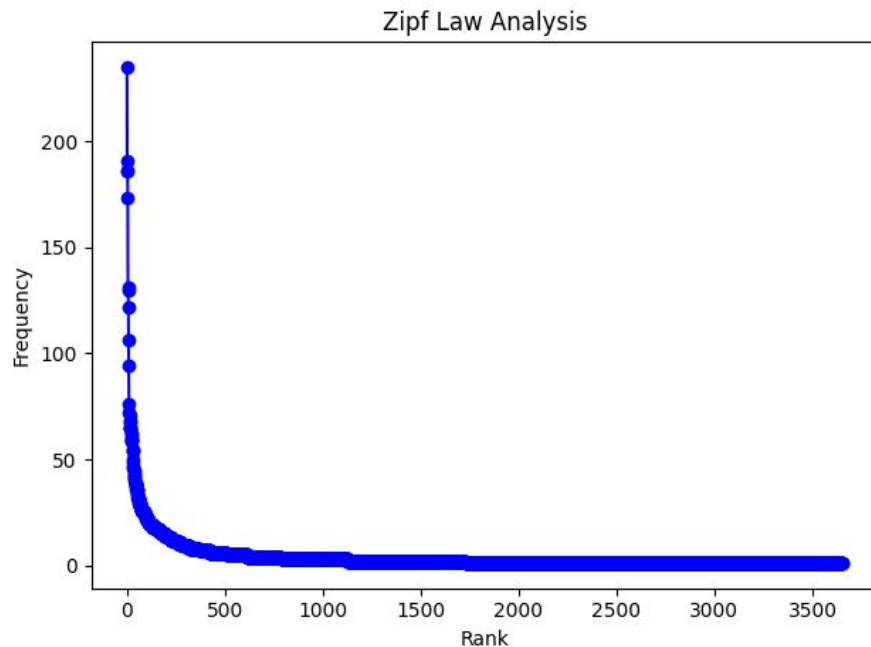


Wędrówki kulinarne Roberta Makłowicza – badania



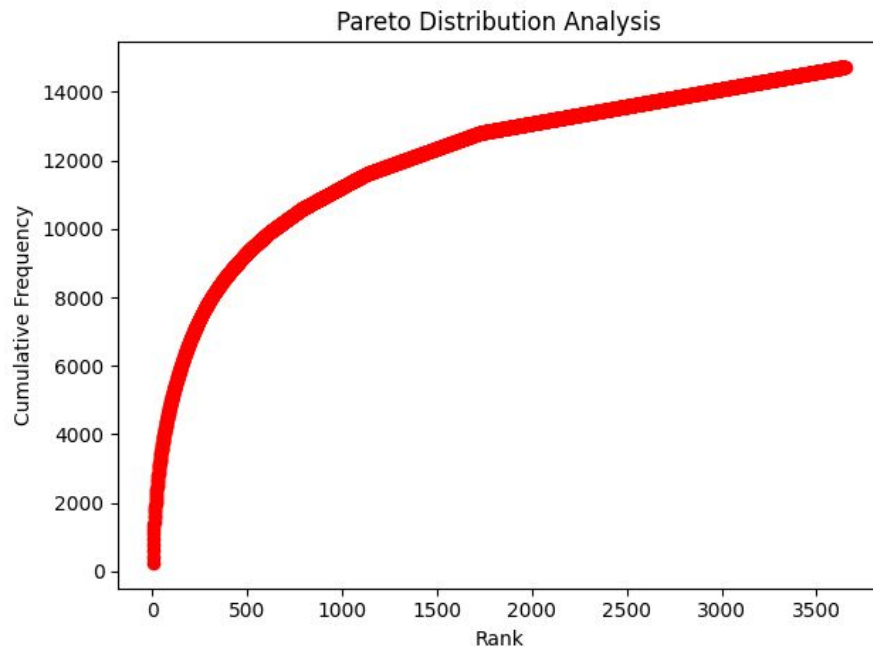


Wędrówki kulinarne Roberta Makłowicza – badania



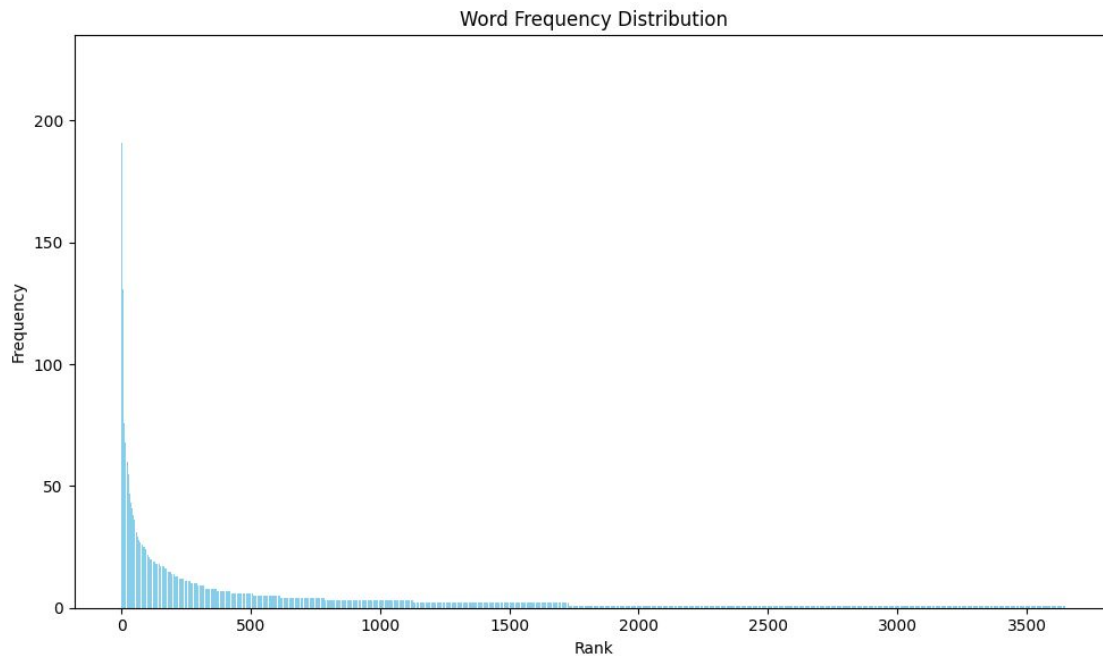


Wędrówki kulinarne Roberta Makłowicza – badania





Wędrówki kulinarne Roberta Makłowicza – badania





Podsumowanie

Analiza rozkładu słów w wybranych książkach kucharskich przy wykorzystaniu Prawa Zipfa i Rozkładu Pareto ujawniła, że wszystkie przypadki spełniają te prawa. Prawo Zipfa, które mówi o relacji między częstością występowania słów a ich rangą, oraz Rozkład Pareto, który opisuje zjawisko ciągłości w częstości występowania różnych zdarzeń, zostały potwierdzone przez analizę tekstu kucharskiego.

Prawidłowe spełnienie Prawa Zipfa oznacza, że częstość występowania słów maleje w sposób odwrotnie proporcjonalny do ich rangi. W kontekście książek kucharskich, gdzie niektóre słowa, takie jak "jajka", "mąka" czy "przepis", mogą być często powtarzane, a inne, takie jak specyficzne nazwy potraw, mogą być rzadsze, analiza ta potwierdza podstawowe założenie Prawa Zipfa.

Rozkład Pareto, natomiast, jest obserwowany w przypadku, gdy pewne zdarzenia są znacznie bardziej częste niż inne, co prowadzi do asymetrycznego rozkładu. W kontekście książek kucharskich, gdzie kilka podstawowych składników lub technik gotowania może dominować nad innymi, takie zjawisko również może zostać potwierdzone przez analizę danych.

Spełnienie obu tych praw w przypadku analizy tekstu kucharskiego sugeruje, że istnieje pewna regularność w użyciu słów, co może odzwierciedlać strukturę języka kuchennego oraz preferencje i nawyki autorów tego typu literatury.



Pliki projektowe

https://github.com/DevxMike/books_wordcount

- kod
- dane w formacie CSV (ilość wystąpień wszystkich słów)
- wykresy
- logi z wyjścia standardowego



Dziękujemy za uwagę

inż. Bazan Michał
inż. Pawłoś Łukasz