

# Raport walidacyjny

## Location Intelligence Data Clustering

Autorzy:

Michał Matuszyk

Paweł Pozorski

Michał Pytel

## Contents

Streszczenie.....	3
Wstęp.....	3
Dane .....	3
Exploratory Data Analysis .....	4
Feature Engineering oraz wyniki walidacji.....	6
Walidacja.....	<b>Error! Bookmark not defined.</b>

## Streszczenie

Celem projektu było przeprowadzenie analizy klastrowania biznesów zawartych w Google Places Comprehensive Business Dataset, aby zidentyfikować znaczące grupy na podstawie ich cech charakterystycznych. Proces klastrowania miał na celu odkrycie trendów i wzorców, które mogą być wykorzystane do analiz rynkowych. Raport przedstawia szczegółowe kroki podjęte podczas procesu walidacji, metody zastosowane do klastrowania, wyniki uzyskane z analizy oraz ich interpretację. Wyniki projektu wskazują na istotne różnice i podobieństwa między grupami przedsiębiorstw, co może stanowić cenną wskazówkę dla dalszych badań rynkowych i strategii biznesowych.

## Wstęp

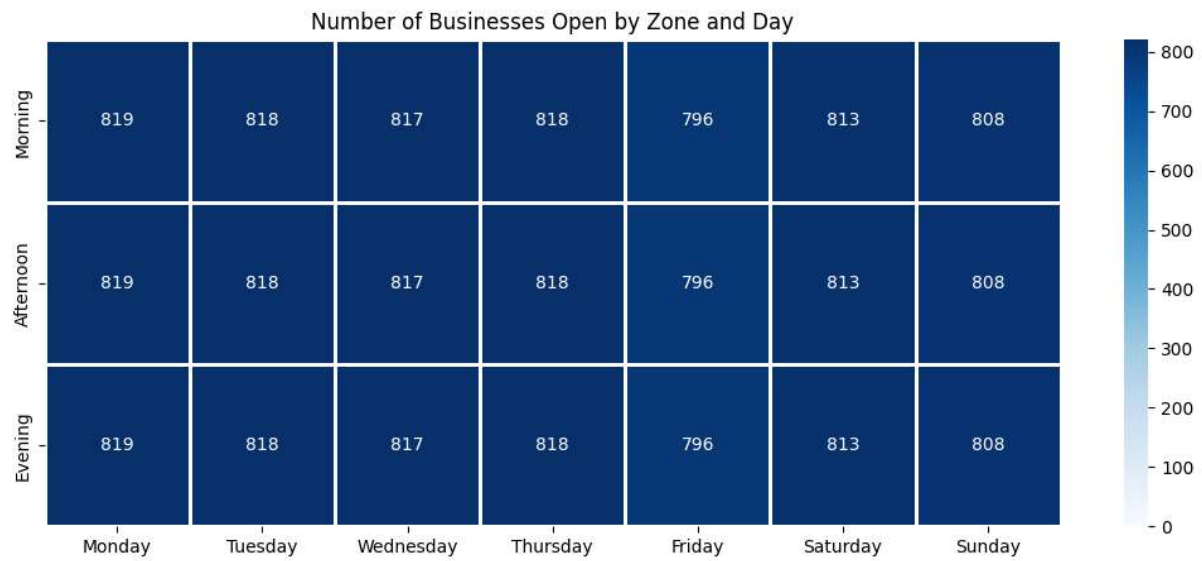
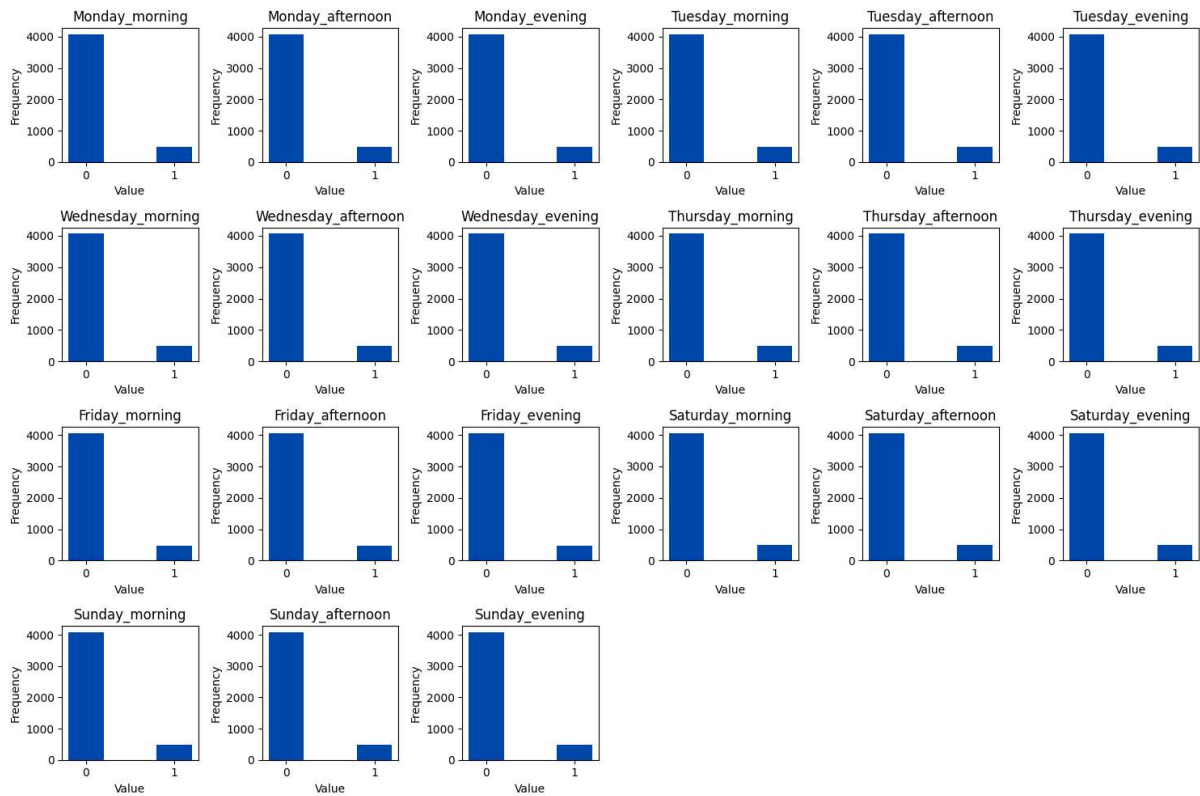
Raport walidacyjny dotyczy projektu, którego celem było wykorzystanie szczegółowych danych zawartych w Google Places Comprehensive Business Dataset do analizy i grupowania biznesów na podstawie ich cech charakterystycznych. Zbiór danych, precyzyjnie zebrany z Google Maps, zawiera rozległe informacje o przedsiębiorstwach z różnych krajów. Każdy wpis w bazie danych dostarcza szczegółowych informacji o działalności firm, specyfikacji lokalizacji, interakcjach z klientami i wielu innych aspektach. Dzięki temu zbiór ten stanowi nieocenione źródło dla analityków danych oraz naukowców zajmujących się analizą trendów biznesowych, analizą danych geograficznych czy badaniami nad wzorcami zachowań konsumentów.

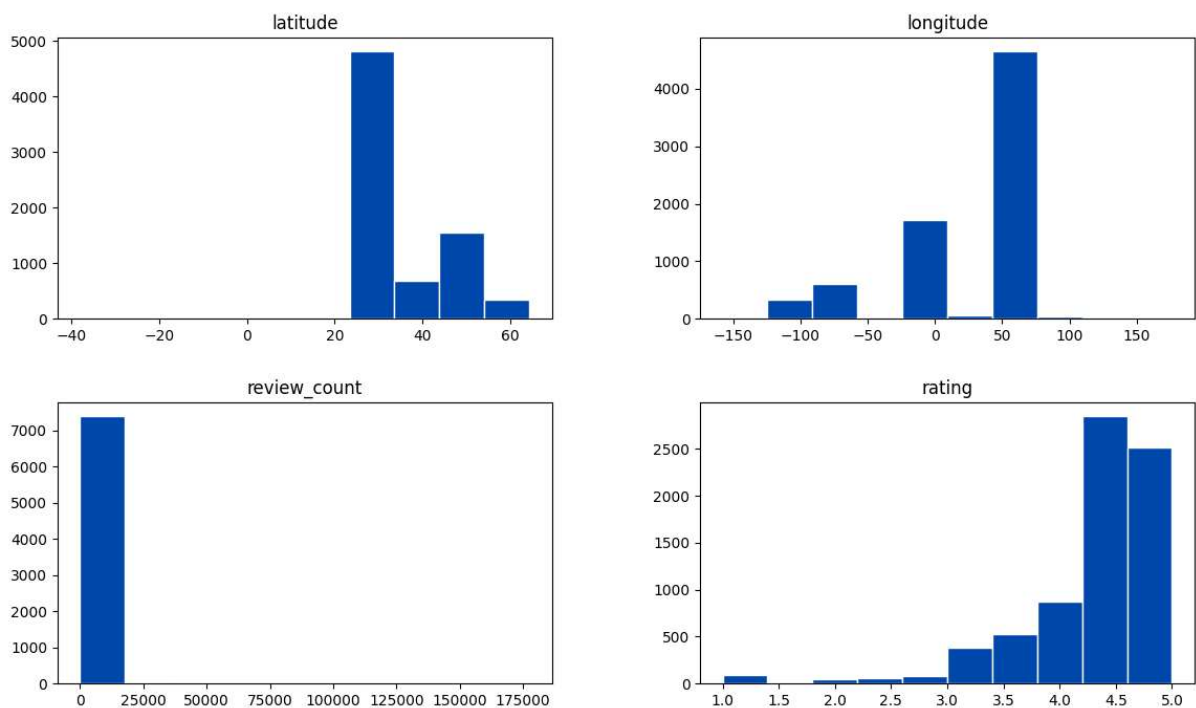
## Dane

Dane zespołu zawierały następujące kolumny:

Column Name	Description
business_id	A unique Google Places identifier for each business, ensuring distinct entries.
phone_number	The contact number associated with the business. It provides a direct means of communication.
name	The official name of the business as listed on Google Maps.
full_address	The complete postal address of the business, including locality and geographic details.
latitude	The geographic latitude coordinate of the business location, useful for mapping and spatial analysis.
longitude	The geographic longitude coordinate of the business location.
review_count	The total number of reviews the business has received on Google Maps.
rating	The average user rating out of 5 for the business, reflecting customer satisfaction.
timezone	The world timezone the business is located in, important for temporal analysis.
website	The official website URL of the business, providing further information and contact options.
place_id	A unique Google Place ID
place_link	A link to the business on Google Maps
Friday	The opening hours on Friday
Saturday	The opening hours on Saturday
Sunday	The opening hours on Sunday
Monday	The opening hours on Monday
Tuesday	The opening hours on Tuesday
Wednesday	The opening hours on Wednesday
Thursday	The opening hours on Thursday
city	The city where the business is located



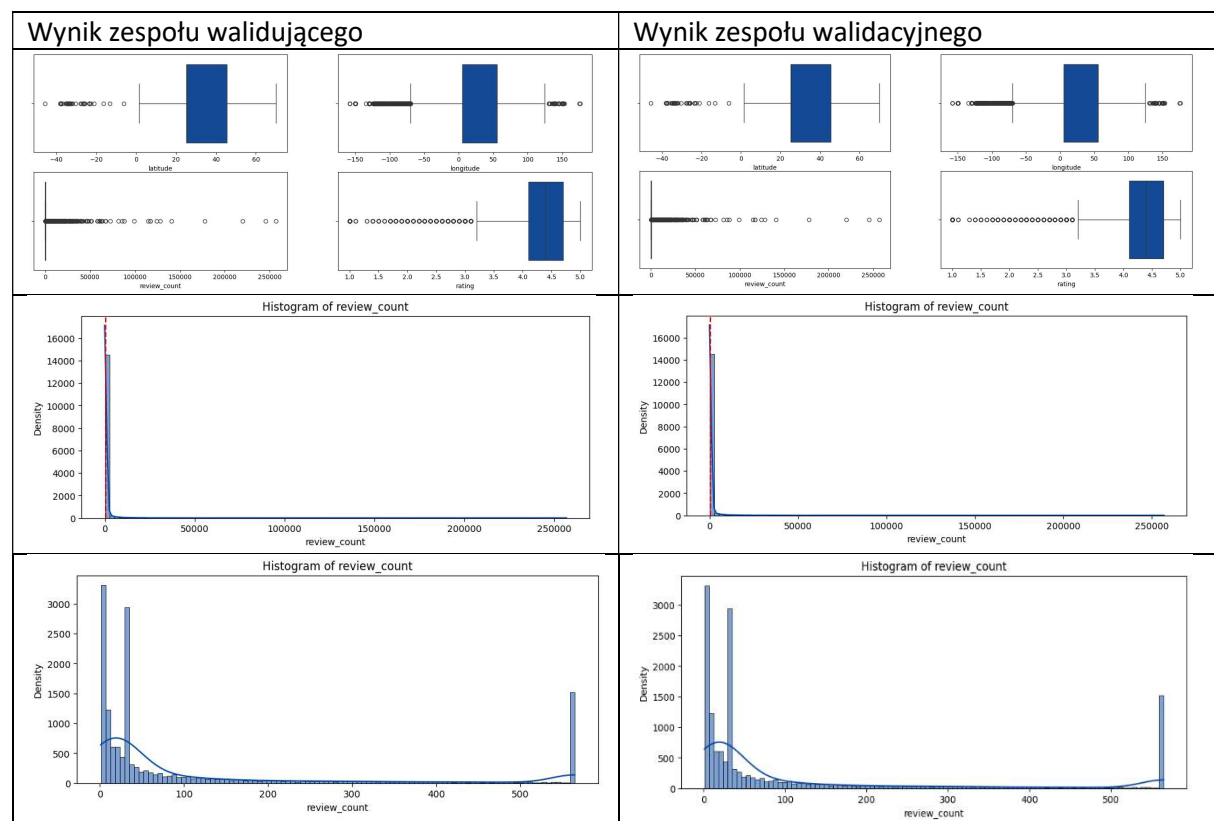


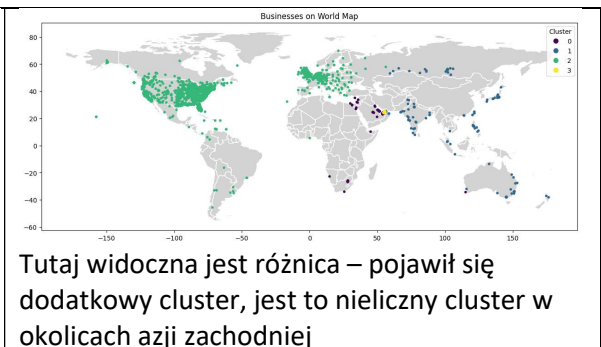
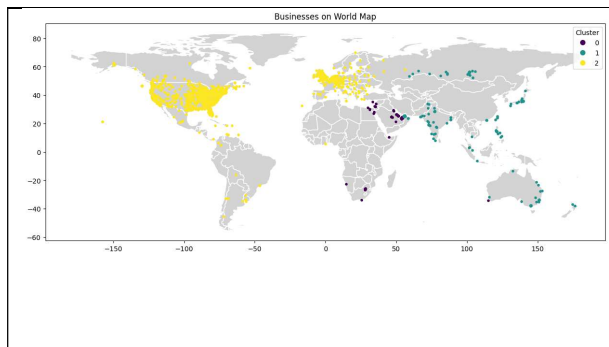


Zespół poprawnie rozpoznał dane oraz dobrał poprawne wykresy, tak aby poprawnie zilustrować zależności w danych.

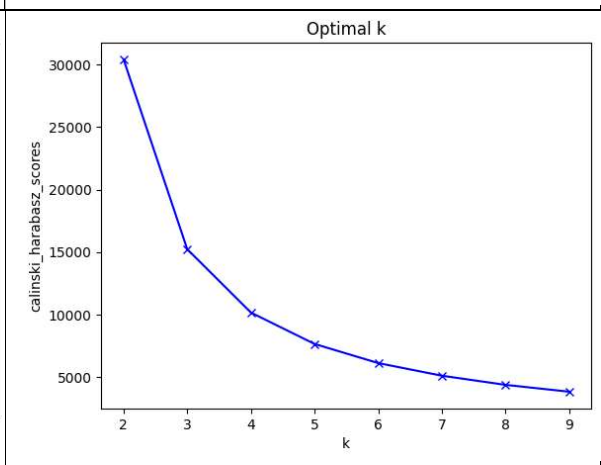
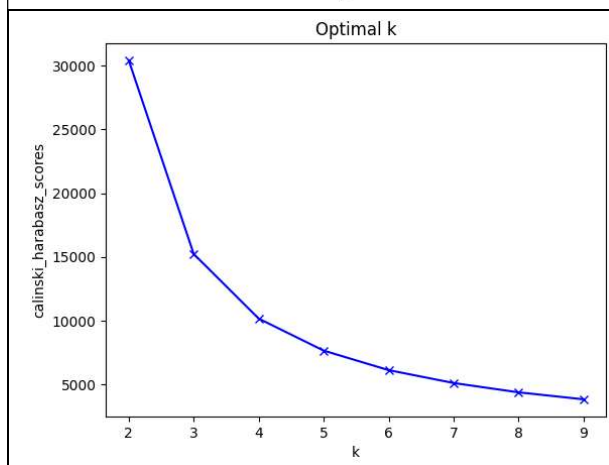
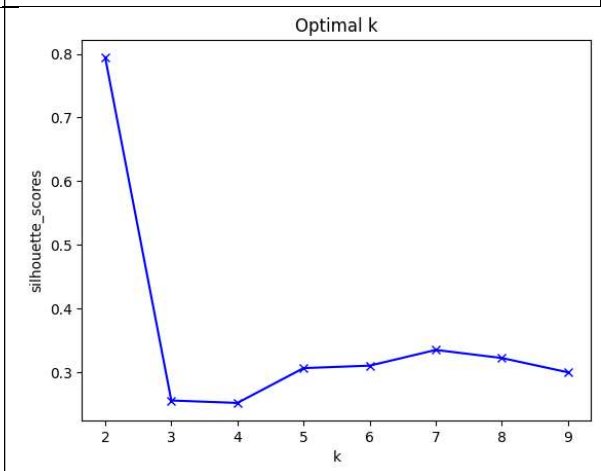
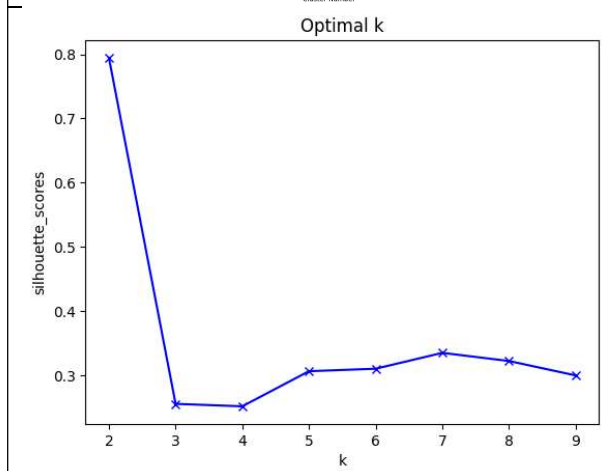
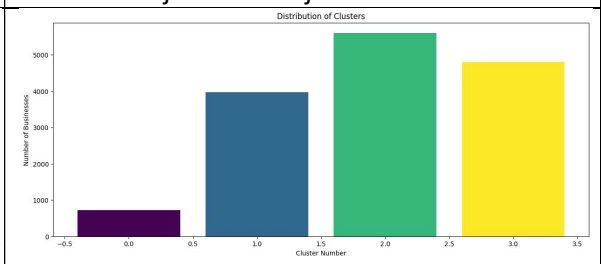
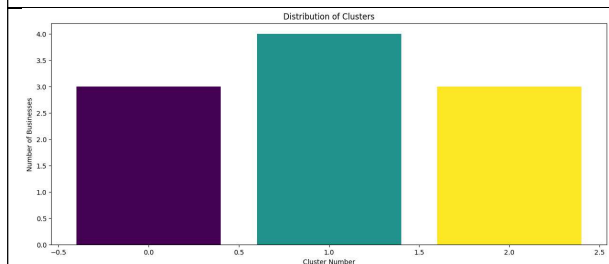
## Feature Engineering oraz wyniki walidacji

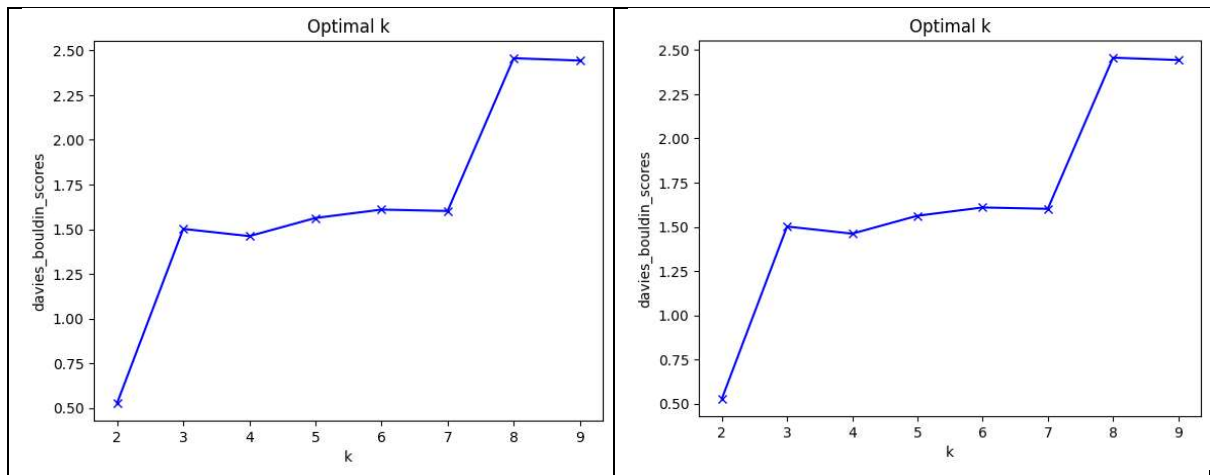
Zespół trenujący udostępnił swoją ramkę danych, przez co, jako zespół walidacyjny mogliśmy porównać otrzymane wyniki.





Tutaj widoczna jest różnica – pojawił się dodatkowy cluster, jest to nieliczny cluster w okolicach azji zachodniej







## Rekomendacje dla zespołu

Rekomendacja	Czy zaimplementowano
Dodanie pliku requirements.txt Dobra praktyka + znacznie ułatwia weryfikację innym osobom	TAK
Drobne poprawki estetyczne kodu	TAK
Dodanie sekcji w notebook'u – poprawia czytelność kodu	TAK

## Podsumowanie

Raport walidacyjny dotyczący projektu "Location Intelligence Data Clustering" szczegółowo opisuje proces analizy i grupowania biznesów zawartych w Google Places Comprehensive Business Dataset. Celem projektu było zidentyfikowanie znaczących grup przedsiębiorstw na podstawie ich cech charakterystycznych, co pozwoliło na odkrycie trendów i wzorców mogących być użytecznymi w analizach rynkowych.

W ramach projektu przeprowadzono analizę eksploracyjną danych (EDA), inżynierię cech oraz klastrowanie. Analiza eksploracyjna wykazała istotne zależności w danych, a dobrze dobrane wykresy i komentarze w kodzie ułatwiły zrozumienie struktury zbioru danych. Proces podziału danych na zestawy walidacyjne, testowe i treningowe został wykonany zgodnie z dobrymi praktykami.

Inżynieria cech i klastrowanie pozwoliły na odkrycie dodatkowego klastra, który zidentyfikowano w regionie Azji Zachodniej. Wyniki te są cenne dla dalszych badań rynkowych i strategii biznesowych, ponieważ pozwalają na lepsze zrozumienie różnic i podobieństw między grupami przedsiębiorstw.

Na podstawie przeprowadzonej walidacji, zespół walidacyjny sugerował kilka rekomendacji, takich jak dodanie pliku requirements.txt oraz drobne poprawki estetyczne kodu, które zostały sprawnie zaimplementowane przez zespół trenujący.