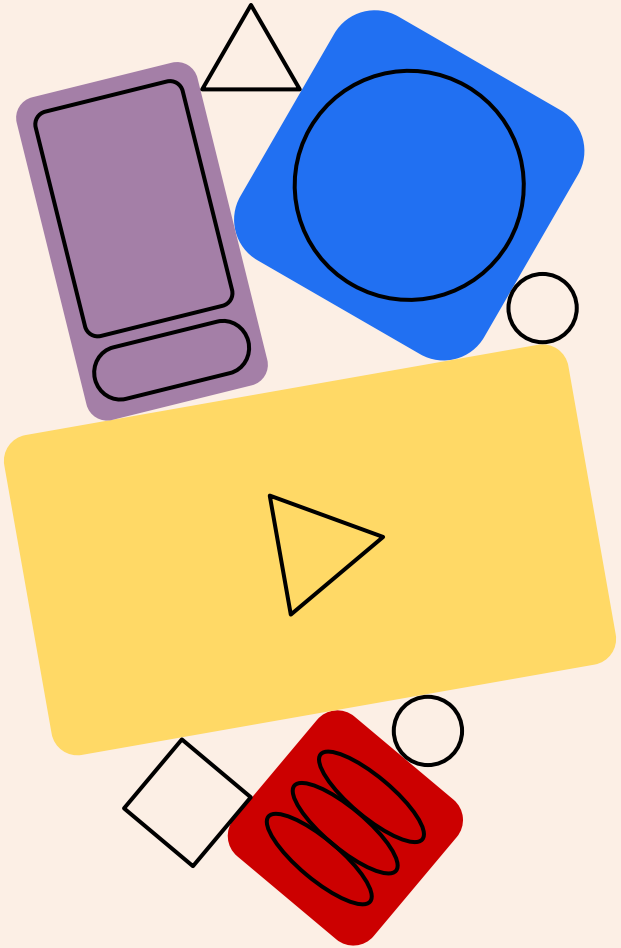
Where at?

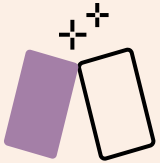
Business analysis by ratings and locations



Paula BD.

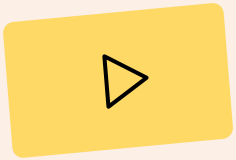
Intro

The goal of the project is to analyse businesses based on:



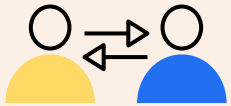
Ratings

1 to 5 stars



Type

Business type



Location

Where in the world is it at?

2019 dataset of 15203 businesses and 46 features.

https://www.kaggle.com/datasets/azharsaleem/location-intelligence-data-from-google-map

✅ Ratings vs Type analysis

🌎 Locations- sub-grouping

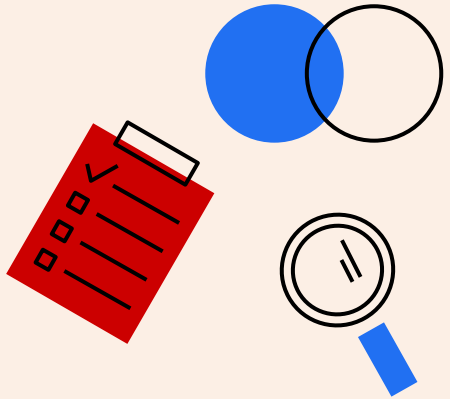
I want to open a restaurant in city X, where is the best place for it?

46 columns to 10

- Business ID
- Latitude
- Longitude
- Review count
- Rating
- Timezone
- Business type
- City
- Country

Features

From 4130 business types went to 10 categories



10 Categories

From 4130 types to 10 categories

- Food & Beverage
- Tourism
- Retail & Shopping
- Finance
- Automotive
- Professional services
- Healthcare
- Education
- Real estate
- Entertainment

1. Food & Beverage: "restaurant|cafe|bakery|coffee shop|bar|pub|fast food|ice cream|pizzeria|deli| steakhouse|bistro|tea house"

2. Tourism: tourist attraction|museum|park|casino |amusement park|Zoo|aquarium|gallery|spa|resort| theme park"

3. Retail & Shopping:"convenience store|shopping mall|clothing store|supermarket|jewelry store|electronics store|gift shop|furniture store|bookstore|toy store|pharmacy|department store"

Ratings ⭐

With a total of 15,198 ratings overall

Going from 1 to 5 stars

Ratings overview

Count 15,198

Mean 4.29

Std 0.67

Min 1.00

25% 4.10

50% 4.40

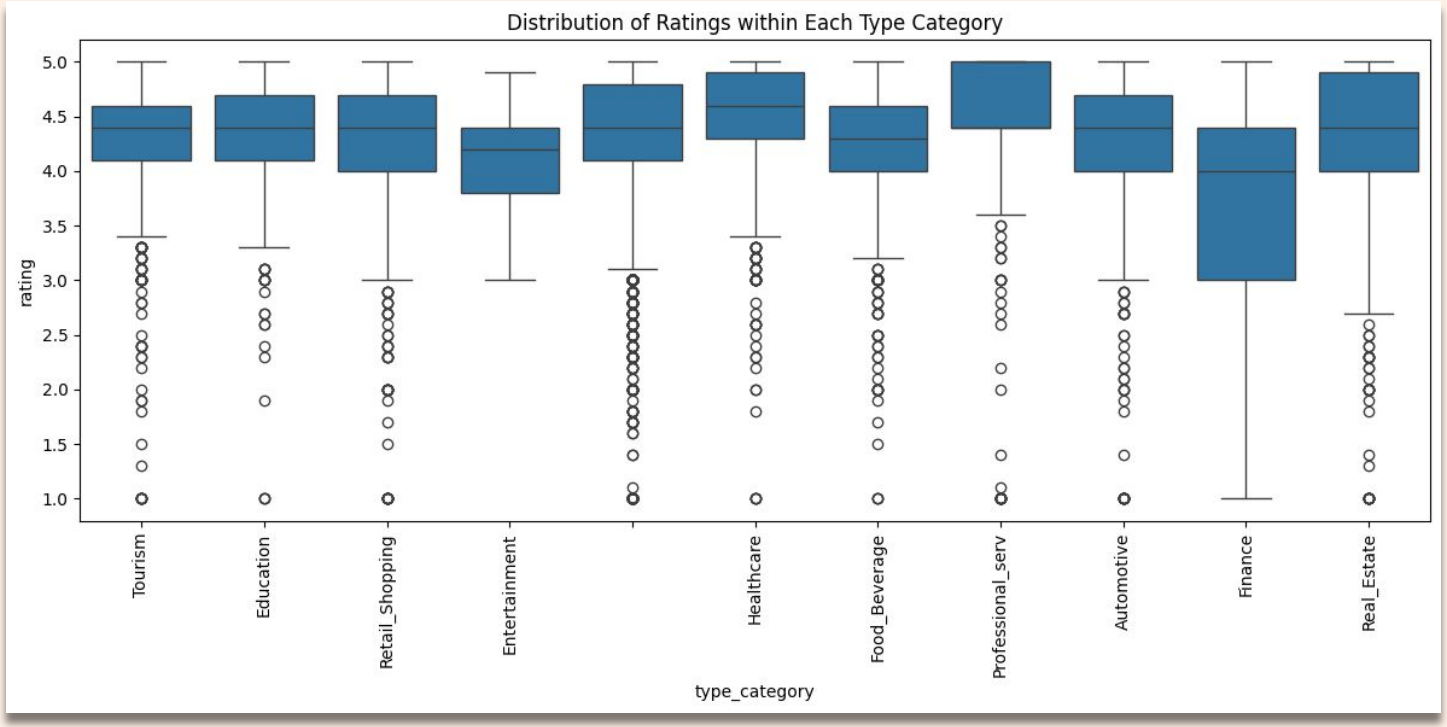
75% 4.70

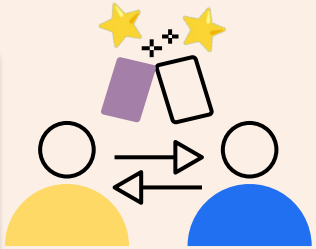
Max 5.00

Avg rating for the 10 categories

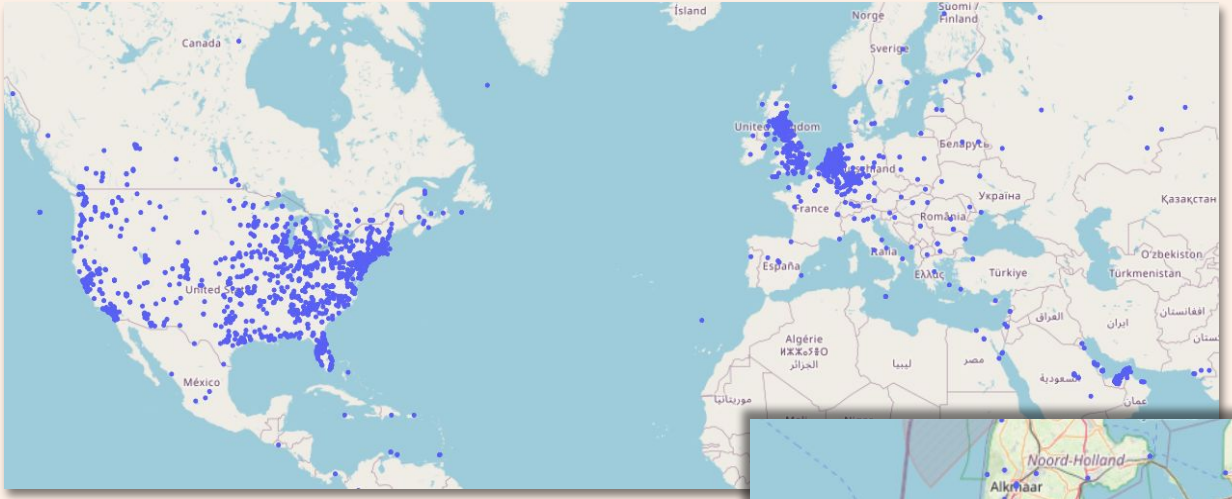
- ★ Healthcare 4.48
- ★ Professional_serv 4.45
- ★ Education 4.36
- ★ Tourism 4.33
- ★ Uncategorized 4.31
- ★ Retail_Shopping 4.29
- ★ Automotive 4.25
- ★ Food_Beverage 4.23
- ★ Real_Estate 4.22
- ★ Entertainment 4.10
- ★ Finance 3.67

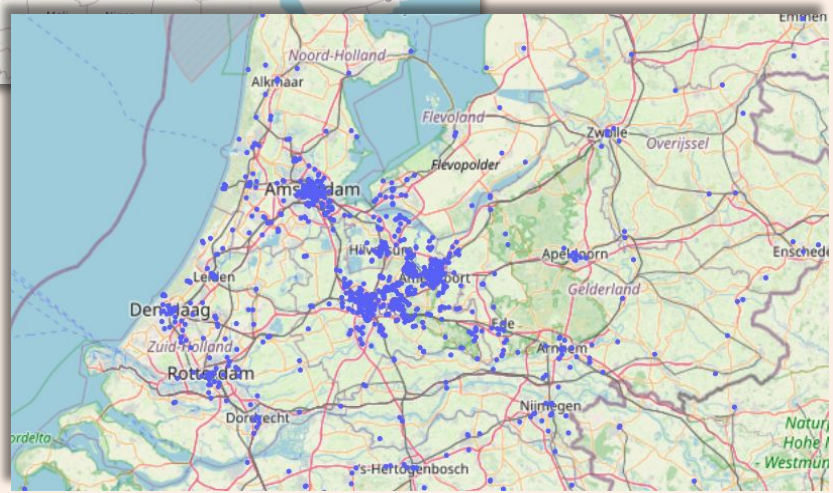
Boxplot coming next...





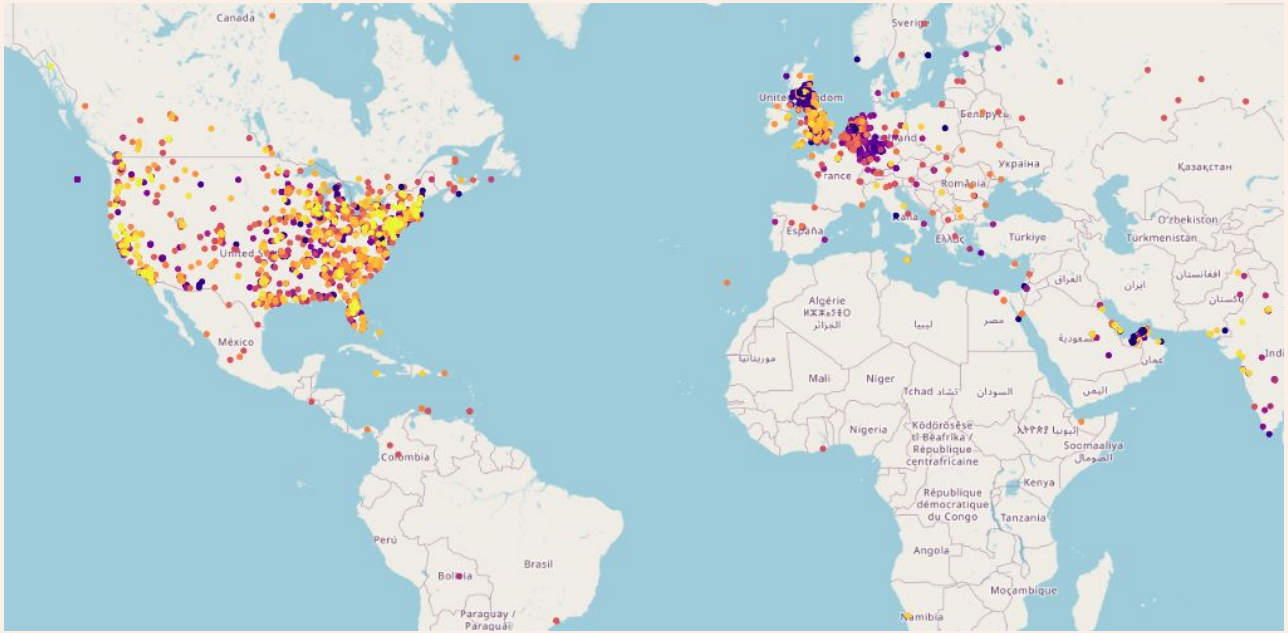
25–75% of the data
Mean
Outlier





Sub-grouping by location

World map WITH clusters!

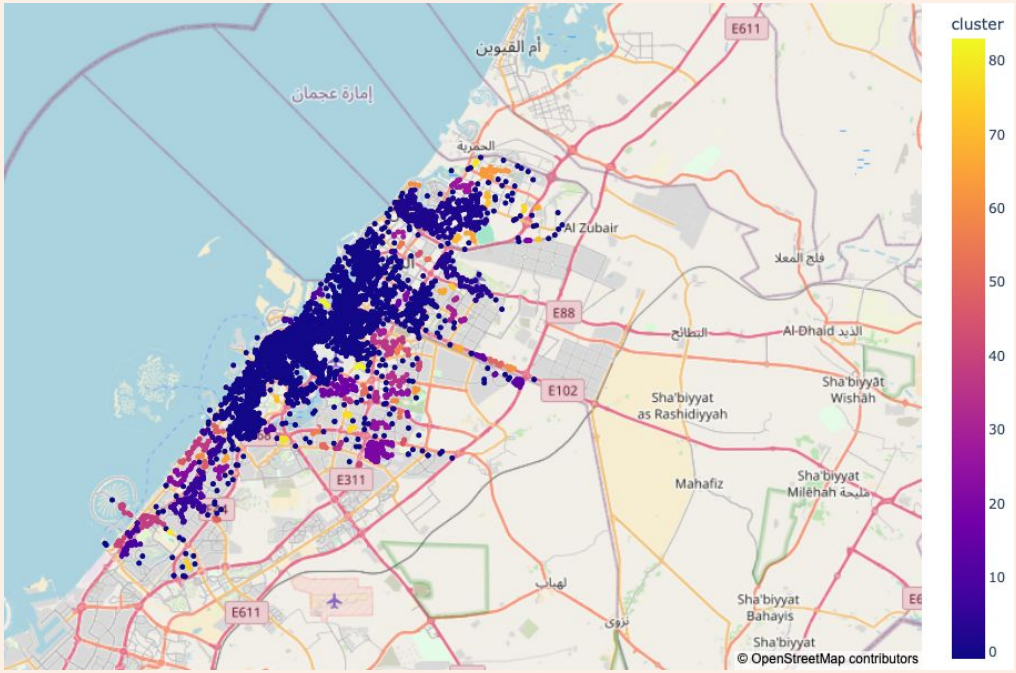


Total of 2985 clusters

Sweet! Now what?...

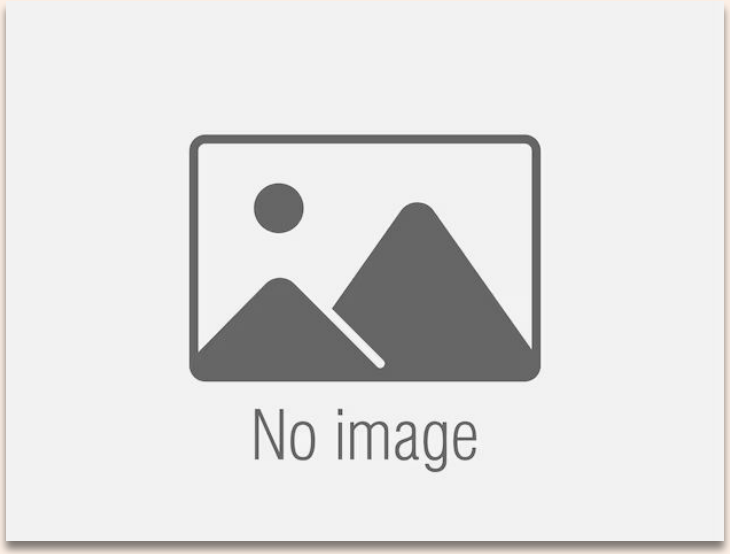
Let's say a business owner wants to open their business in Dubai!

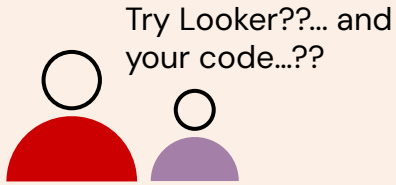
Cluster 87 with 8,123 businesses 🤑 = 85 sub-groups



Location, ratings, and business category analysis

🌎 Clusters in 87 💼 Type category ⭐ AVG rating

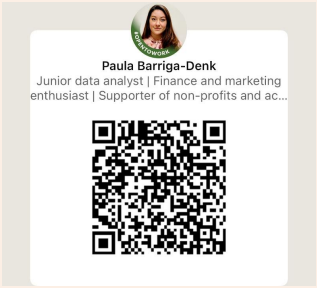




Conclusion

- Updated data
- Defined business type
- Defined city

I could advise a business owner where to open a new business.


Paula Barriga-Denk
Junior data analyst | Finance and marketing enthusiast | Supporter of non-profits and ac...

Thank you! ✌️

Icon Library

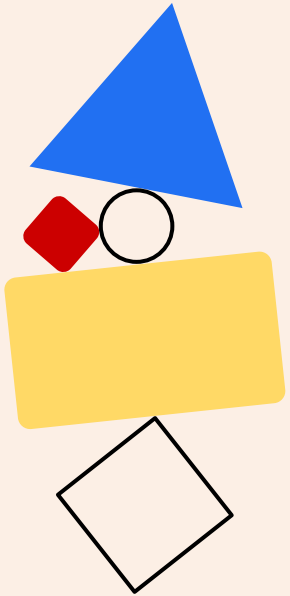
Sub-grouping by location

When using actual geographical coordinates, it is important to make sure that the Transformation is set to Raw. This is to avoid distortions in the distance measure that may result from the transformations.

Why DBSCAN?

- DBSCAN can form clusters of any shape, as it groups points that are density-connected.
- DBSCAN automatically determines the number of clusters
- DBSCAN classifies isolated points as noise

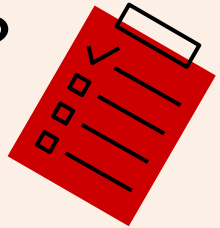
By applying DBSCAN, to latitude and longitude data, one can identify clusters of businesses that are geographically close to each other.

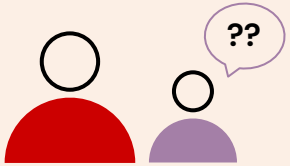


- The Ball-Tree algorithm works by organizing data points into a hierarchical structure, where each node in the tree represents a ball (or hypersphere) containing a subset of the data points.

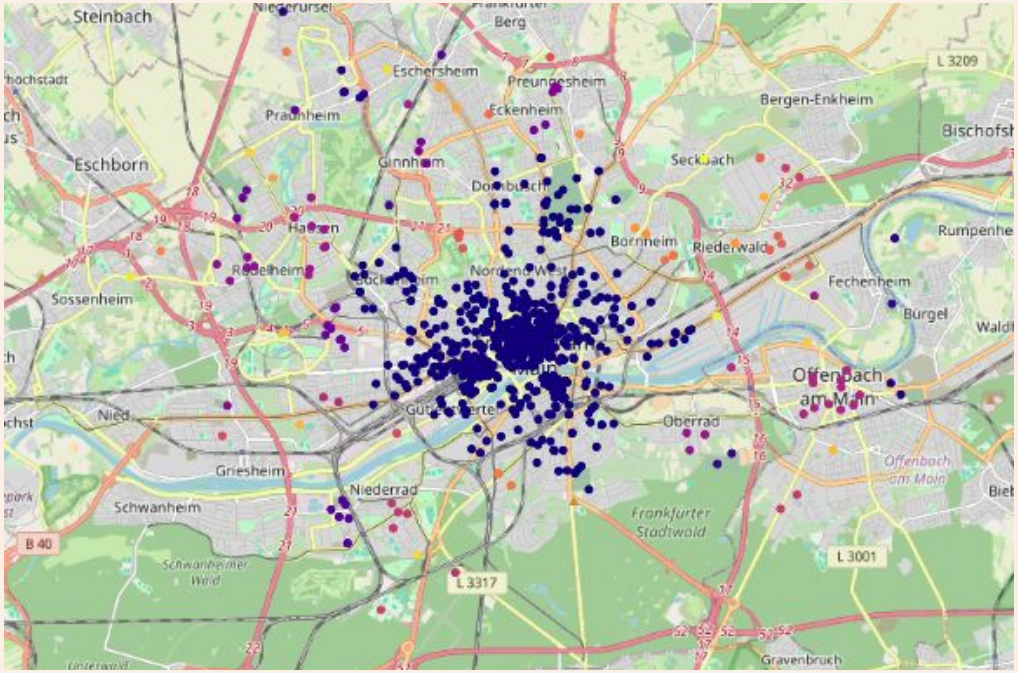
- The haversine formula determines the great-circle distance between two points on a sphere given their longitudes and latitudes.

Scaling our data?





Cluster 369 with 915 businesses 🤑 = 60 sub-groups



Cluster 87 with Ratings = 78 clusters