

Performance Analysis and Strategic Insights for a Dental Clinic Chain in France

A Data-Driven approach to understand and improve center performance

Mohammed El Barhichi
AI Engineering Student
04/12/2024



CONTENT



1 Contextualization



2 Exploratory Data Analysis



3 Data Cleaning



4 Feature Importance and Performance Explainability



5 Clients Segmentation



6 Concurrential Analysis



7 Summary and Recommendations

1. Contextualization

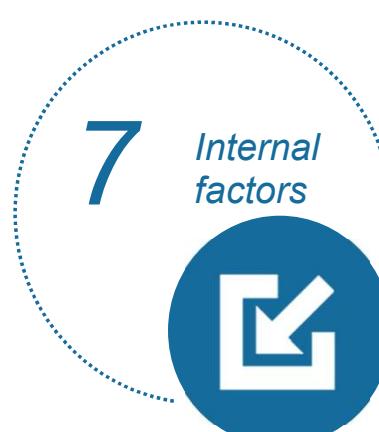
Problem statement :

The client, a private dental clinic chain, seeks to understand what drives its centers' performance and to what extent it is influenced by operational (**internal**) factors versus **external** contextual factors.

Objectifs :

- Understand Key Drivers of Performance
- Provide Actionable Insights
- Evaluate Competitive Landscape
- Support Decision-Making with Data

Dataset overview :



2. Exploratory Data Analysis

- Internal Factors:

Latitude and Longitude: Geographic coordinates of the center.

Revenue [21; EUR m]: Revenue of the center in 2021 (in millions of Euros).

Margin rate [21; %]: Profit margin rate of the center in 2021 (as a percentage).

Dentist FTE [#]: Full-time equivalent (FTE) of dentists working at the center.

Number of reviews on Google [#]: The number of reviews the center received on Google.

Average Google rating: Average rating of the center based on Google reviews.

Opening year: The year when the center was opened.

- External Factors:

Rent in the area [EUR/sqm]: Rent in the area where the center is located, measured in Euros per square meter.

Patients in the catchment area [#]: Total number of potential patients in the area served by the center.

Patients' yearly growth in the catchment area [%]: Annual growth rate of the patient population in the catchment area.

Share of patients in the catchment area over 65 [%]: Percentage of patients in the catchment area aged over 65.

Median revenue in catchment area [EUR]: Median income of individuals in the catchment area.

Unemployment rate in the catchment area [%]: Unemployment rate in the catchment area.

Poverty rate in the catchment area [%]: Poverty rate in the catchment area.

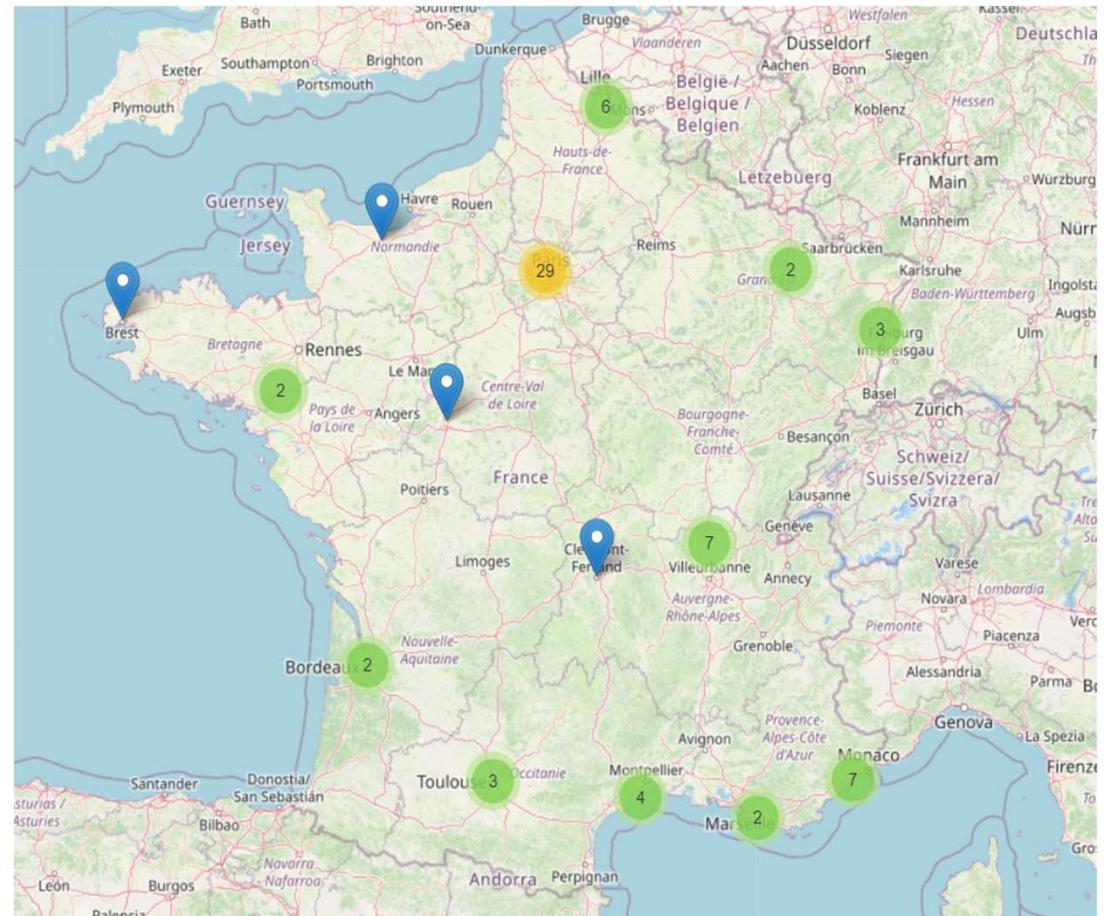
Density of doctors in the catchment area [#/100k patients]: Number of doctors per 100,000 patients in the catchment area.

Density of pharmacists in the catchment area [#/100k patients]: Number of pharmacists per 100,000 patients in the catchment area.

Density of dentists in the catchment area [#/100k]: Number of dentists per 100,000 patients in the catchment area.

2. Exploratory Data Analysis

- Internal Factors:
 - ✓ Latitude and Longitude
 - ✓ Revenue [21; EUR m]
 - Margin rate [21; %]
 - Dentist FTE [#]
 - ✓ Number of reviews on Google [#]
 - ✓ Average Google rating
 - Opening year
- External Factors:
 - ✓ Rent in the area [EUR/sqm]
 - Patients in the catchment area [#]
 - ✓ Patients' yearly growth in the catchment area [%]
 - ✓ Share of patients in the catchment area over 65 [%]
 - ✓ Median revenue in catchment area [EUR]
 - ✓ Unemployment rate in the catchment area [%]
 - ✓ Poverty rate in the catchment area [%]
 - ✓ Density of doctors in the catchment area [#/100k patients]
 - ✓ Density of pharmacists in the catchment area [#/100k patients]
 - ✓ Density of dentists in the catchment area [#/100k]



[clients_dental_centers.html](#)

2. Exploratory Data Analysis

	Latitude	Longitude	Revenue [21; EUR m]	Margin rate [21; %]	Dentist FTE [#]	Number of reviews on Google [#]	Average Google rating	Rent in the area [EUR/sqm]	Patients in the catchment area [#]	Patients' yearly growth in the catchment area [%]	Share of patients in the catchment area over 65 [%]	Median revenue in catchment area [EUR]	Unemployment rate in the catchment area [%]	Poverty rate in the catchment area [%]	Density of doctors in the catchment area [#/100k patients]
count	71.000000	71.000000	66.000000	66.000000	66.000000	68.000000	68.000000	71.000000	70.000000	71.000000	71.000000	71.000000	70.000000	70.000000	69.000000
mean	47.304452	3.228897	2.306061	0.143636	6.378788	132.029412	4.137079	5362.507042	79894.527810	0.148732	16.864059	23000.078755	7.835714	16.752048	169.869565
std	2.410416	2.459225	2.053206	0.079897	2.777742	96.637339	0.421768	3462.712499	65210.404188	0.891498	6.134942	4738.040804	1.659670	7.758953	67.395434
min	42.691889	-4.479540	0.000000	-0.050000	0.000000	4.000000	3.331897	1121.000000	1564.226438	-2.740000	9.059540	14317.058824	5.800000	2.370061	-237.000000
25%	45.312752	2.287864	1.350000	0.100000	4.250000	47.000000	3.852150	2700.500000	25911.164404	-0.490000	13.126878	20218.888499	6.400000	11.987072	150.000000
50%	48.692024	2.430776	2.150000	0.145000	6.000000	119.500000	4.126894	4200.000000	55924.496673	0.040000	15.303172	22262.121212	7.200000	15.454031	174.000000
75%	48.880986	4.854985	2.975000	0.197500	8.000000	190.500000	4.469948	7277.000000	124574.171986	0.710000	18.216715	24000.727969	8.425000	19.536252	186.000000
max	50.725329	7.750131	15.000000	0.340000	15.000000	359.000000	5.000000	13515.000000	246399.292251	4.130000	41.315940	40315.194805	12.500000	47.600799	248.000000

df.describe()

Scaling Issue :

Margin Rate

2. Exploratory Data Analysis

	Latitude	Longitude	Revenue [21; EUR m]	Margin rate [21; %]	Dentist FTE [#]	Number of reviews on Google [#]	Average Google rating	Rent in the area [EUR/sqm]	Patients in the catchment area [#]	Patients' yearly growth in the catchment area [%]	Share of patients in the catchment area over 65 [%]	Median revenue in catchment area [EUR]	Unemployment rate in the catchment area [%]	Poverty rate in the catchment area [%]	Density of doctors in the catchment area [#/100k patients]
count	71.000000	71.000000	66.000000	66.000000	66.000000	68.000000	68.000000	71.000000	70.000000	71.000000	71.000000	71.000000	70.000000	70.000000	69.000000
mean	47.304452	3.228897	2.306061	0.143636	6.378788	132.029412	4.137079	5362.507042	79894.527810	0.148732	16.864059	23000.078755	7.835714	16.752048	169.869565
std	2.410416	2.459225	2.053206	0.079897	2.777742	96.637339	0.421768	3462.712499	65210.404188	0.891498	6.134942	4738.040804	1.659670	7.758953	67.395434
min	42.691889	-4.479540	0.000000	0.050000	0.000000	4.000000	3.331897	1121.000000	1564.226438	-2.740000	9.059540	14317.058824	5.800000	2.370061	-237.000000
25%	45.312752	2.287864	1.350000	0.100000	4.250000	47.000000	3.852150	2700.500000	25911.164404	-0.490000	13.126878	20218.888499	6.400000	11.987072	150.000000
50%	48.692024	2.430776	2.150000	0.145000	6.000000	119.500000	4.126894	4200.000000	55924.496673	0.040000	15.303172	22262.121212	7.200000	15.454031	174.000000
75%	48.880986	4.854985	2.975000	0.197500	8.000000	190.500000	4.469948	7277.000000	124574.171986	0.710000	18.216715	24000.727969	8.425000	19.536252	186.000000
max	50.725329	7.750131	15.000000	0.340000	15.000000	359.000000	5.000000	13515.000000	246399.292251	4.130000	41.315940	40315.194805	12.500000	47.600799	248.000000

df.describe()

Scaling Issue :

Margin Rate

Potential Outliers :

Revenue, Density of doctors,

2. Exploratory Data Analysis

	Latitude	Longitude	Revenue [21; EUR m]	Margin rate [21; %]	Dentist FTE [#]	Number of reviews on Google [#]	Average Google rating	Rent in the area [EUR/sqm]	Patients in the catchment area [#]	Patients' yearly growth in the catchment area [%]	Share of patients in the catchment area over 65 [%]	Median revenue in catchment area [EUR]	Unemployment rate in the catchment area [%]	Poverty rate in the catchment area [%]	Density of doctors in the catchment area [#/100k patients]
count	71.000000	71.000000	66.000000	66.000000	66.000000	68.000000	68.000000	71.000000	70.000000	71.000000	71.000000	71.000000	70.000000	70.000000	69.000000
mean	47.304452	3.228897	2.306061	0.143636	6.378788	132.029412	4.137079	5362.507042	79894.527810	0.148732	16.864059	23000.078755	7.835714	16.752048	169.869565
std	2.410416	2.459225	2.053206	0.079897	2.777742	96.637339	0.421768	3462.712499	65210.404188	0.891498	6.134942	4738.040804	1.659670	7.758953	67.395434
min	42.691889	-4.479540	0.000000	-0.050000	0.000000	4.000000	3.331897	1121.000000	1564.226438	-2.740000	9.059540	14317.058824	5.800000	2.370061	-237.000000
25%	45.312752	2.287864	1.350000	0.100000	4.250000	47.000000	3.852150	2700.500000	25911.164404	-0.490000	13.126878	20218.888499	6.400000	11.987072	150.000000
50%	48.692024	2.430776	2.150000	0.145000	6.000000	119.500000	4.126894	4200.000000	55924.496673	0.040000	15.303172	22262.121212	7.200000	15.454031	174.000000
75%	48.880986	4.854985	2.975000	0.197500	8.000000	190.500000	4.469948	7277.000000	124574.171986	0.710000	18.216715	24000.727969	8.425000	19.536252	186.000000
max	50.725329	7.750131	15.000000	0.340000	15.000000	359.000000	5.000000	13515.000000	246399.292251	4.130000	41.315940	40315.194805	12.500000	47.600799	248.000000

df.describe()

Scaling Issue :

Margin Rate

Potential Outliers :

Revenue, Density of doctors,...

Missing Values :

Revenue, Dentist FTE, Average Google Rating,...

3. Data Cleaning

Handling Scaling Issues :

The values appear to be stored as decimals instead of percentages (0.05 representing 5%)

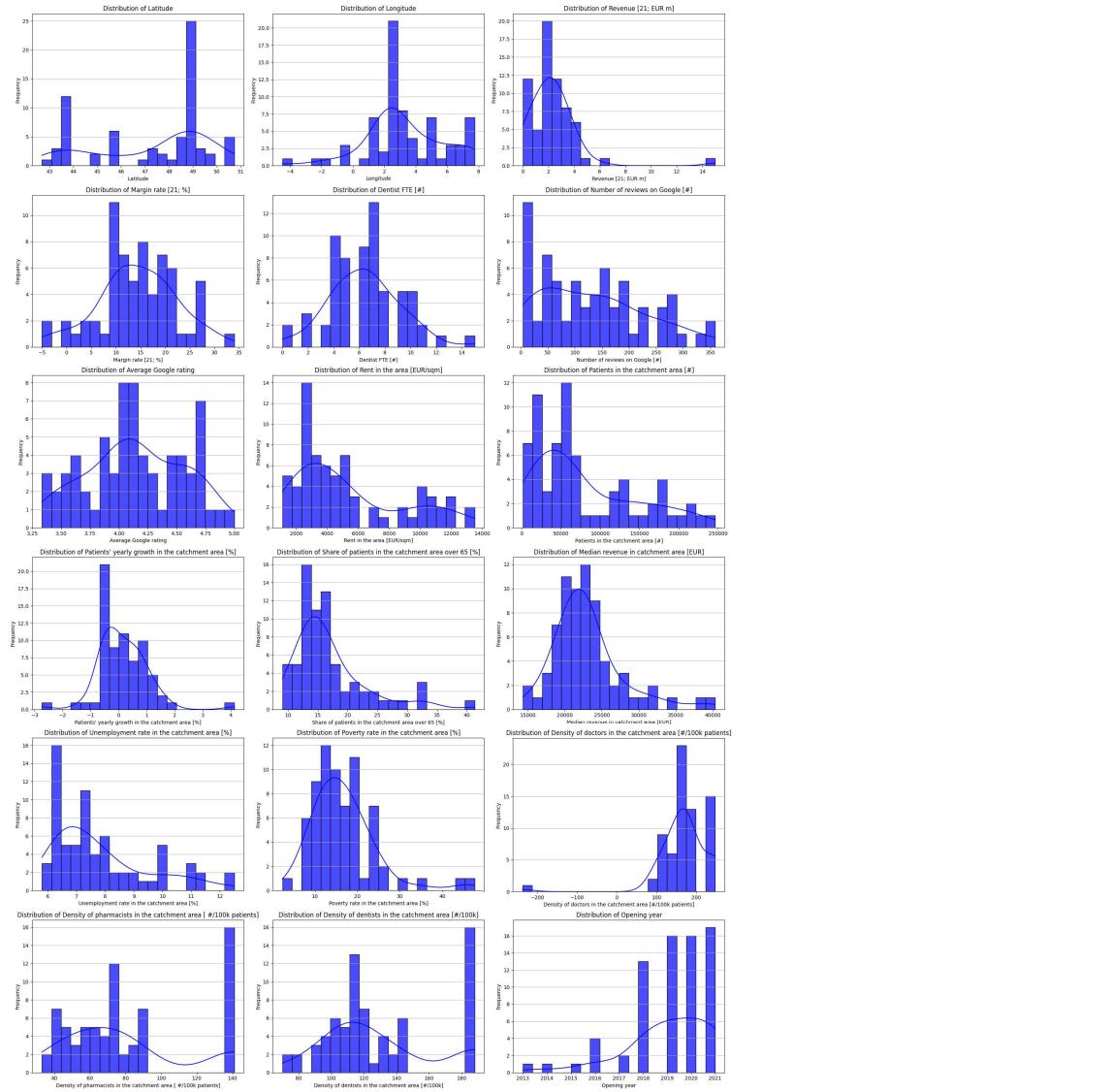
Margin rate [21; %]	
count	66.000000
mean	0.143636
std	0.079897
min	-0.050000
25%	0.100000
50%	0.145000
75%	0.197500
max	0.340000



Margin rate [21; %]	
count	66.000000
mean	14.363636
std	7.989679
min	-5.000000
25%	10.000000
50%	14.500000
75%	19.750000
max	34.000000

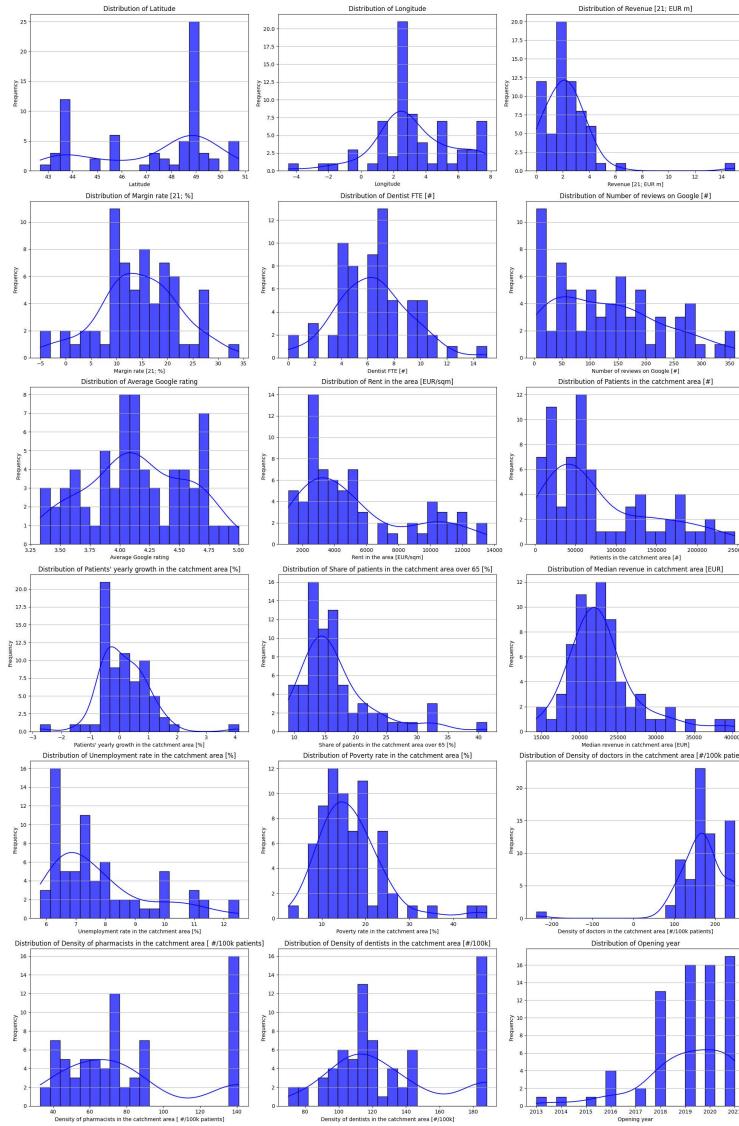
3. Data Cleaning

Handling Outliers:



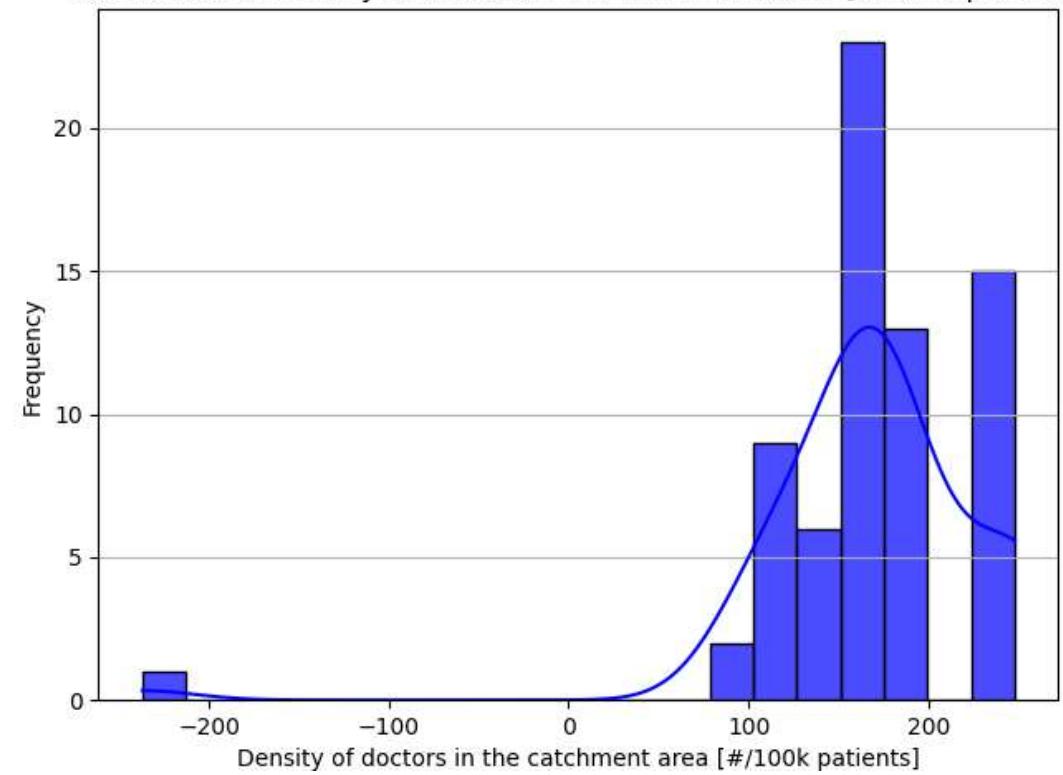
3. Data Cleaning

Handling Outliers:



- Density of doctors in the catchment area

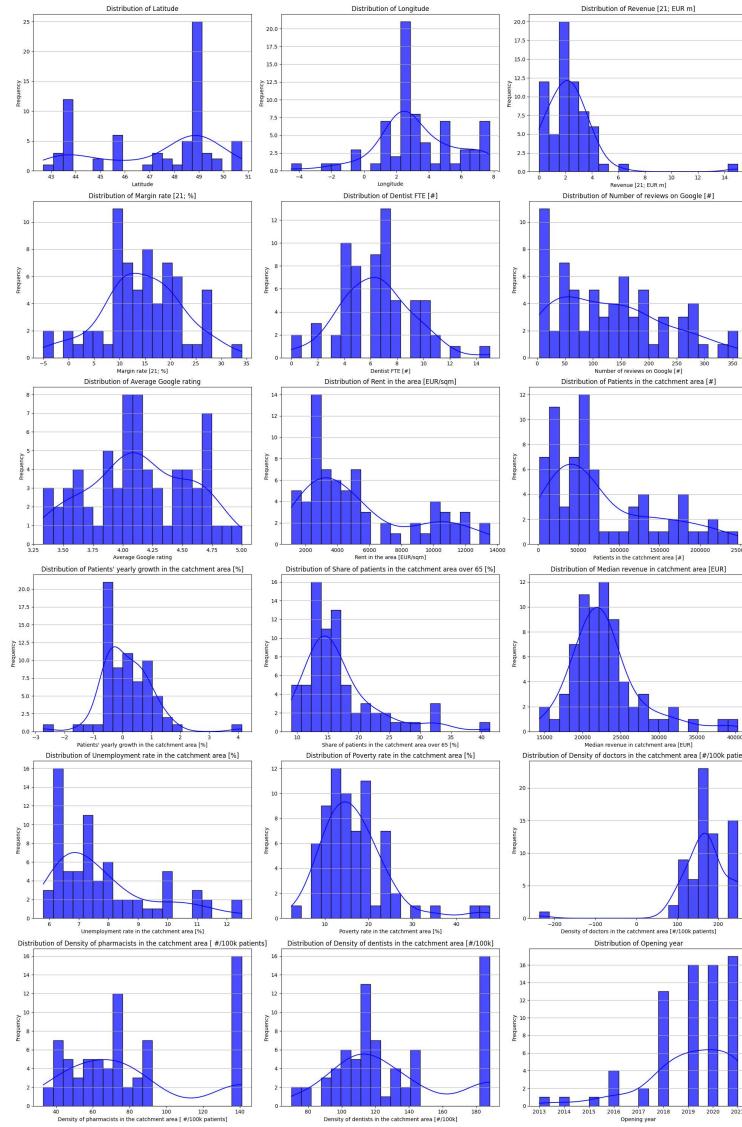
Distribution of Density of doctors in the catchment area [#/100k patients]



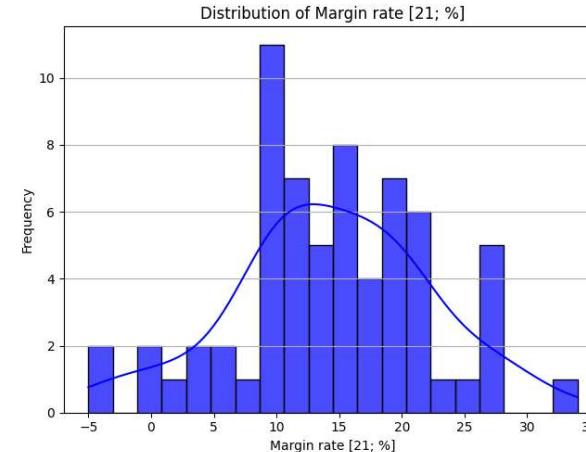
Delete value

3. Data Cleaning

Handling Outliers:

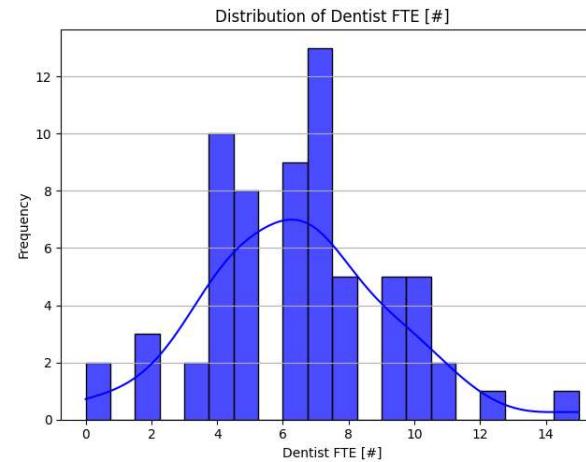


- Margin rate



The costs of these centers exceeded their revenues during the year of 2021

- Dentist FTE

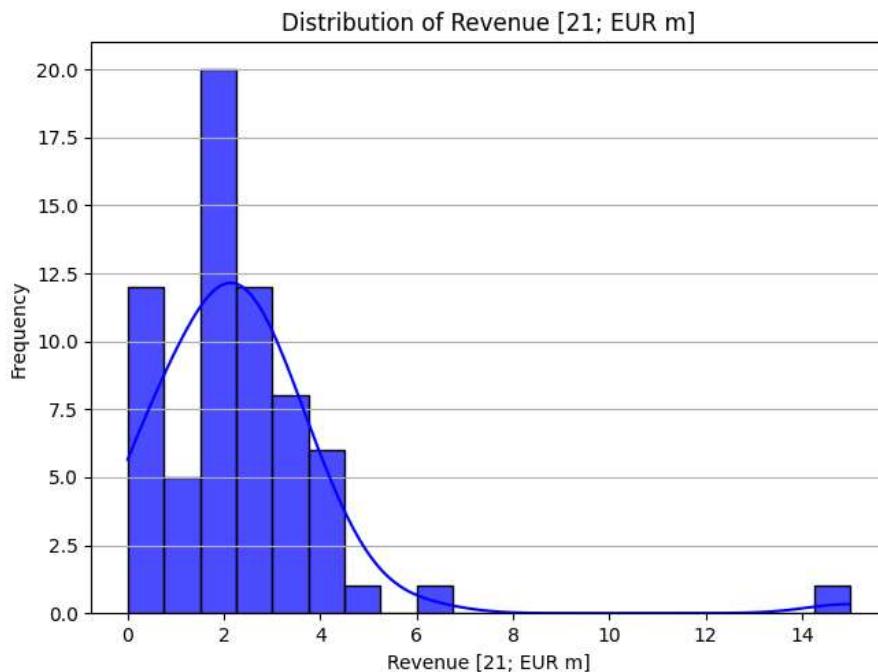


Non-operational or temporarily inactive centers

3. Data Cleaning

Handling Outliers:

- Revenue



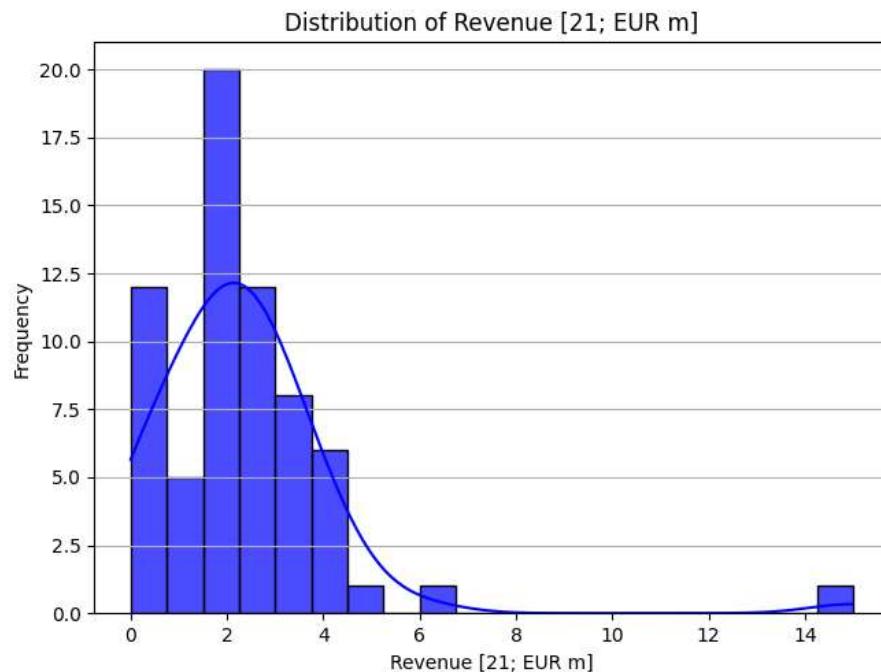
Two potential outliers :

- Revenue = 0
- Revenue = 15

3. Data Cleaning

Handling Outliers:

- Revenue



Two potential outliers :

- Revenue = 0
- Revenue = 15
- **Revenue = 0**

Rows with negative Margin Rate, null Dentist FTE, no data

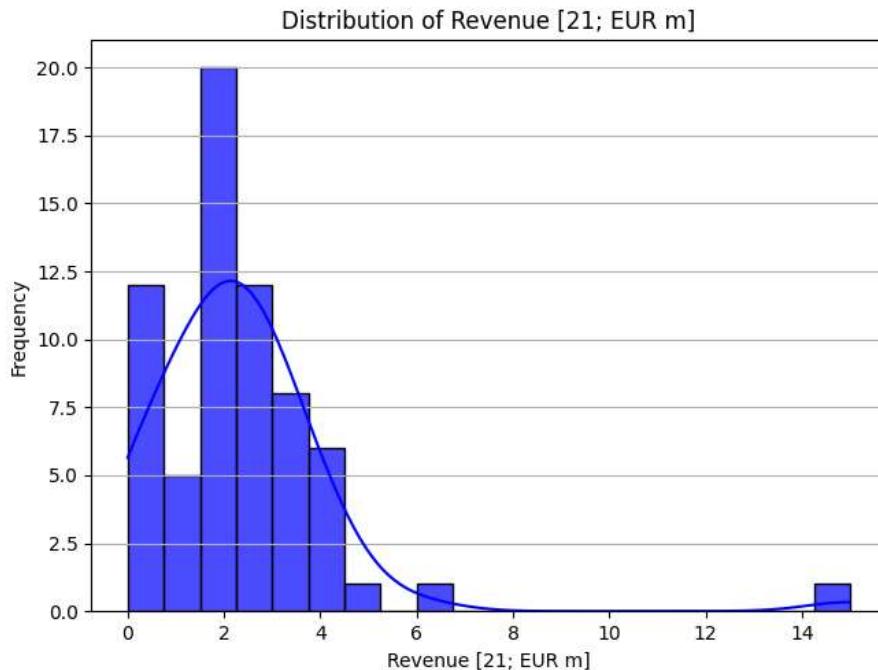
Google Rating

→ Non-operational or temporarily inactive centers

3. Data Cleaning

Handling Outliers:

- Revenue



Two potential outliers :

- Revenue = 0
- Revenue = 15
- Revenue = 0

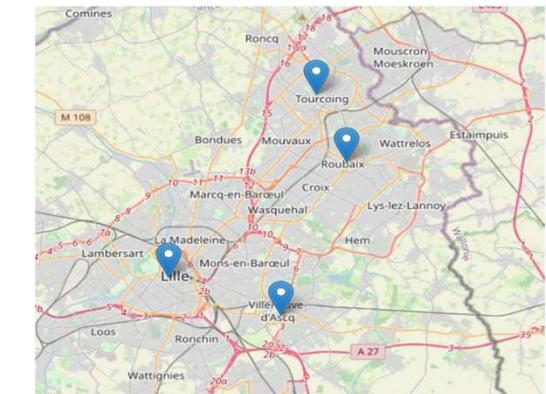
Rows with negative Margin Rate, null Dentist FTE, no data Google Rating

Non-operational or temporarily inactive centers

- Revenue = 15

- Geographical Analysis

ID	Latitude	Longitude	Revenue [21; EUR m]
70	50.616725	3.133466	15.0
23	50.635571	3.058779	1.8
58	50.692226	3.176593	1.6
67	50.725329	3.156516	3.0



Nearby centers within 15 km

- Operational Analysis

ID	Revenue [21; EUR m]	Dentist FTE [#]	Patients in the catchment area [#]	Average Google rating	Number of reviews on Google [#]
70	15.0	4.0	24675.533216	4.891304	46.0
10	2.2	6.0	27532.624233	4.316583	199.0
14	1.6	4.0	25030.600363	4.136986	73.0

Similar operational characteristics

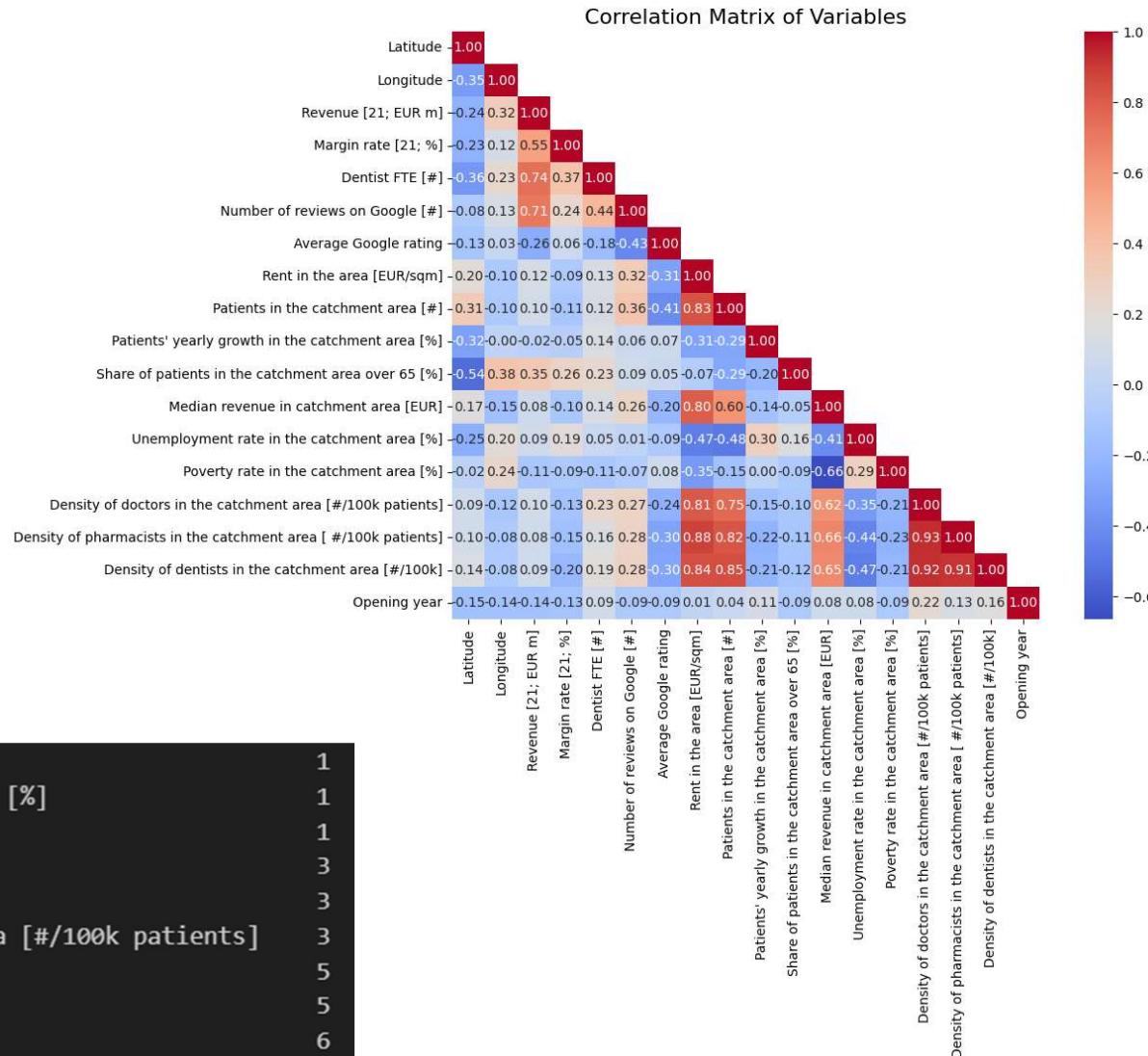


Delete value

3. Data Cleaning

Handling Missing Values:

- Correlation Matrix :



- Missing values :

Patients in the catchment area [#]	1
Unemployment rate in the catchment area [%]	1
Poverty rate in the catchment area [%]	1
Number of reviews on Google [#]	3
Average Google rating	3
Density of doctors in the catchment area [#/100k patients]	3
Margin rate [21; %]	5
Dentist FTE [#]	5
Revenue [21; EUR m]	6

3. Data Cleaning

Handling Missing Values:

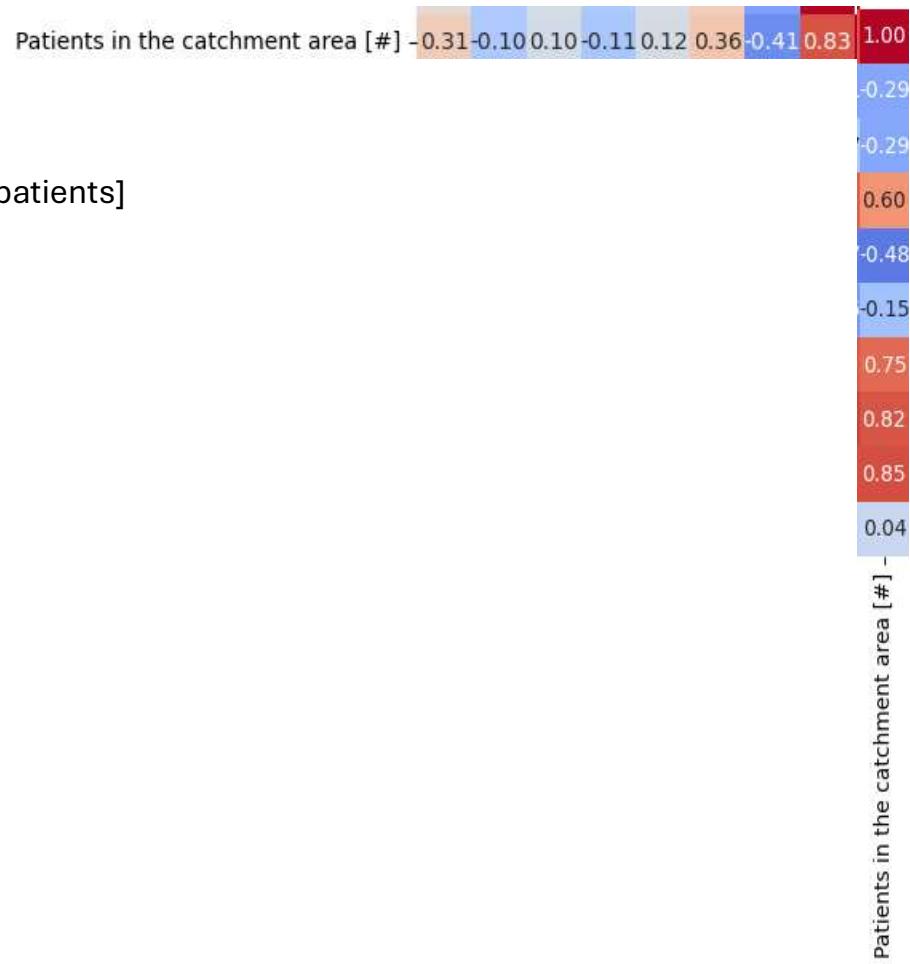
1. Patients in the Catchment Area :

Strong correlation with :

- Rent in the area [EUR/sqm]
 - Density of pharmacists in the catchment area [#/100k patients]
 - Density of dentists in the catchment area [#/100k]



Imputation using a Linear Regression model



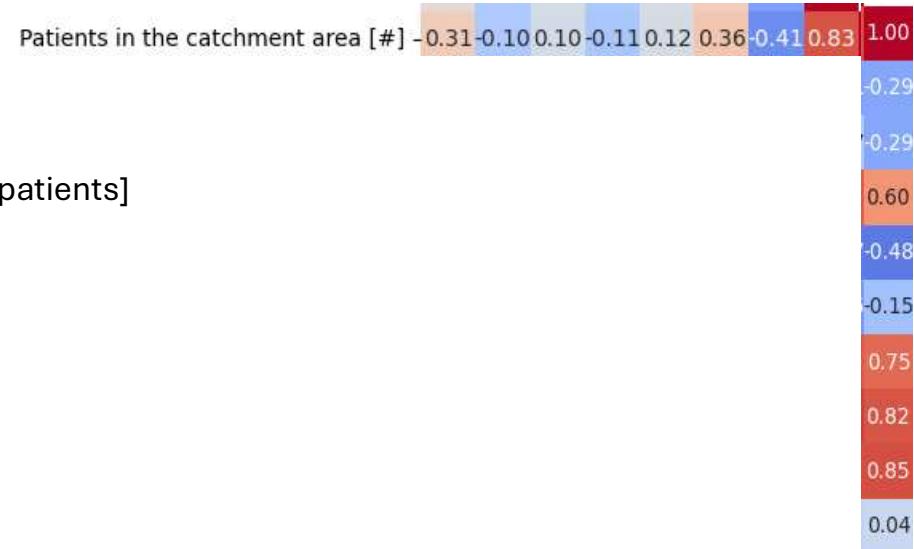
3. Data Cleaning

Handling Missing Values:

1. Patients in the Catchment Area :

Strong correlation with :

- Rent in the area [EUR/sqm]
- Density of pharmacists in the catchment area [#/100k patients]
- Density of dentists in the catchment area [#/100k]



→ Imputation using a Linear Regression model

2. Unemployment rate :

ID	Poverty rate in the catchment area [%]	Rent in the area [EUR/sqm]	Density of dentists in the catchment area [#/100k]	Unemployment rate in the catchment area [%]
6	13.676555	6007	121	7.4
25	13.144815	5117	144	7.2
26	13.212571	4806	144	7.2
28	12.317151	5503	144	7.2
31	15.734146	5089	130	Nan
36	20.218534	5142	134	7.0
62	18.587686	5293	117	8.5
66	15.468347	4941	116	8.1
71	14.731955	4711	144	7.2

→ Imputation using the mean of similar row in the most related features

3. Data Cleaning

Handling Missing Values:

3. Poverty Rate

	Unemployment rate in the catchment area [%]	Poverty rate in the catchment area [%]	Density of doctors in the catchment area [#/100k patients]	Density of pharmacists in the catchment area [#/100k patients]	Density of dentists in the catchment area [#/100k]
ID					
15	7.0	NaN	158.0	78	134
36	7.0	20.218534	158.0	78	134

The two centers are in the same catchment area



Imputation using the value of the other center

3. Data Cleaning

Handling Missing Values:

3. Poverty Rate :

	Unemployment rate in the catchment area [%]	Poverty rate in the catchment area [%]	Density of doctors in the catchment area [#/100k patients]	Density of pharmacists in the catchment area [#/100k patients]	Density of dentists in the catchment area [#/100k]
ID					
15	7.0	NaN	158.0	78	134
36	7.0	20.218534	158.0	78	134

The two centers are in the same catchment area

→ Imputation using the value of the other center

4. Number of reviews on Google and Average Google rating :

Strong correlation with :

- Patients in the catchment area [#]
- Density of dentists in the catchment area [#/100k]

→ Imputation using KNN_Imputer

3. Data Cleaning

Handling Missing Values:

3. Poverty Rate :

ID	Unemployment rate in the catchment area [%]	Poverty rate in the catchment area [%]	Density of doctors in the catchment area [#/100k patients]	Density of pharmacists in the catchment area [#/100k patients]	Density of dentists in the catchment area [#/100k]
15	7.0	NaN	158.0	78	134
36	7.0	20.218534	158.0	78	134

The two centers are in the same catchment area

→ Imputation using the value of the other center

4. Number of reviews on Google and Average Google rating :

Strong correlation with :

- Patients in the catchment area [#]
- Density of dentists in the catchment area [#/100k]

→ Imputation using KNN_Imputer

5. Density of Doctors in the Catchment Area :

16 centers are in the same Catchment Area

→ Imputation using the value of the other centers

ID	Patients' yearly growth in the catchment area [%]	Unemployment rate in the catchment area [%]	Poverty rate in the catchment area [%]	Density of doctors in the catchment area [#/100k patients]	Density of pharmacists in the catchment area [#/100k patients]	Density of dentists in the catchment area [#/100k]
20	-0.49	6.4	13.358924	248.0	141	188
40	-0.49	6.4	17.389824	248.0	141	188
41	-0.49	6.4	12.511394	248.0	141	188
42	-0.49	6.4	10.846958	248.0	141	188
43	-0.49	6.4	17.110562	248.0	141	188
44	-0.49	6.4	9.504357	248.0	141	188
45	-0.49	6.4	7.165592	248.0	141	188
46	-0.49	6.4	7.014793	248.0	141	188
47	-0.49	6.4	7.330330	248.0	141	188
48	-0.49	6.4	18.332756	248.0	141	188
49	-0.49	6.4	23.005147	248.0	141	188
50	-0.49	6.4	11.651570	248.0	141	188
51	-0.49	6.4	19.437242	248.0	141	188
52	-0.49	6.4	15.638180	248.0	141	188
53	-0.49	6.4	19.526397	248.0	141	188
54	-0.49	6.4	10.308393	248.0	141	188

3. Data Cleaning

Handling Missing Values:

6. Revenue, Marge Rate and Dentists FTE :

	Latitude	Longitude	Revenue [21; EUR m]	Margin rate [21; %]	Dentist FTE [#]
ID					
18	48.736834	1.365297	NaN	NaN	NaN
21	49.023062	1.148888	NaN	NaN	NaN
30	48.959236	2.882169	NaN	NaN	NaN
35	48.692024	6.181425	NaN	NaN	NaN
38	43.838278	4.363345	NaN	NaN	NaN
70	50.616725	3.133466	NaN	19.0	4.0

3. Data Cleaning

Handling Missing Values:

6. Revenue, Marge Rate and Dentists FTE :

	Latitude	Longitude	Revenue [21; EUR m]	Margin rate [21; %]	Dentist FTE [#]
ID					
18	48.736834	1.365297	NaN	NaN	NaN
21	49.023062	1.148888	NaN	NaN	NaN
30	48.959236	2.882169	NaN	NaN	NaN
35	48.692024	6.181425	NaN	NaN	NaN
38	43.838278	4.363345	NaN	NaN	NaN
70	50.616725	3.133466	NaN	19.0	4.0

The outlier with 15 as Revenue

- KNN Imputer : Value of 1,6

The original value of 15.0 was likely a misentry, where 1.5 was intended



Imputation using the value 1,5

3. Data Cleaning

Handling Missing Values:

6. Revenue, Marge Rate and Dentists FTE :

	Latitude	Longitude	Revenue [21; EUR m]	Margin rate [21; %]	Dentist FTE [#]
ID					
18	48.736834	1.365297	NaN	NaN	NaN
21	49.023062	1.148888	NaN	NaN	NaN
30	48.959236	2.882169	NaN	NaN	NaN
35	48.692024	6.181425	NaN	NaN	NaN
38	43.838278	4.363345	NaN	NaN	NaN
70	50.616725	3.133466	NaN	19.0	4.0

- Highly correlated columns
- Traditional imputation methods are not effective : dependency between variables

→ Imputation using advanced Iterative Imputer

The outlier with 15 as Revenue

- KNN Imputer : Value of 1,6

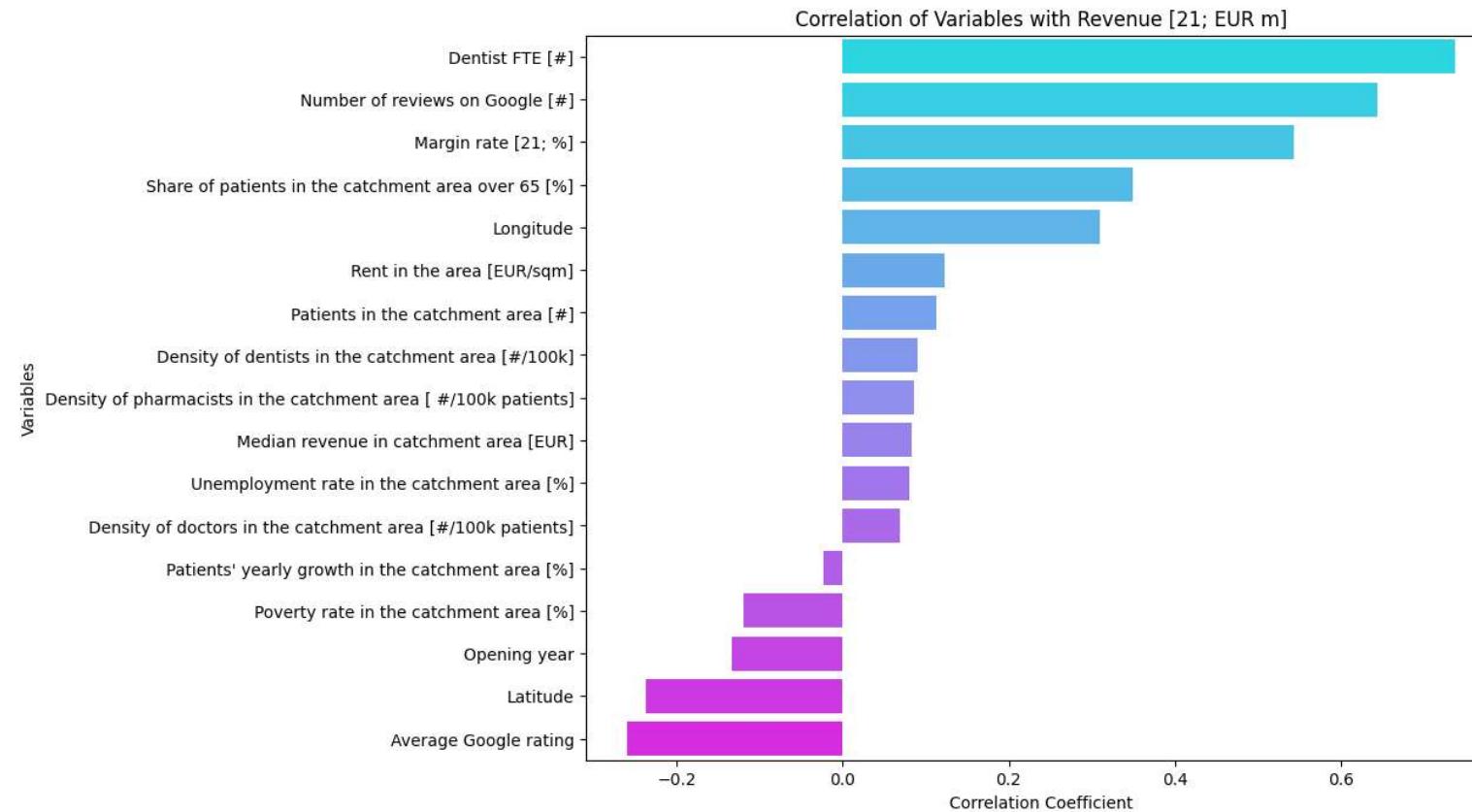
The original value of 15.0 was likely a misentry, where 1.5 was intended

→ Imputation using the value 1,5

4. Feature Importance and Performance Explainability

Target : Revenue

Correlation Analysis :



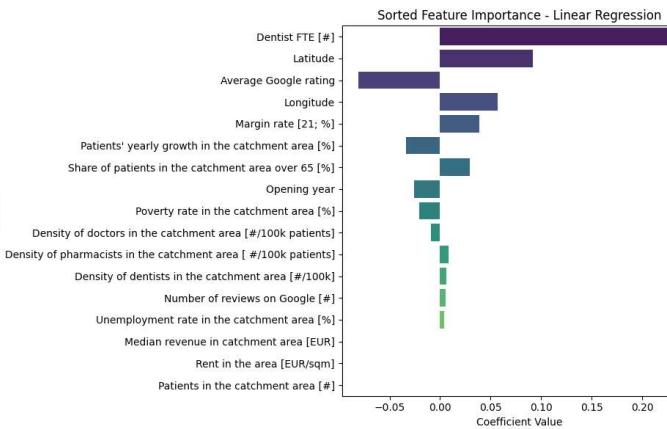
- Strong direct correlation between many **internal** factors and the target
- Negative correlation with Average Google rating : Non-Linear relationships

4. Feature Importance and Performance Explainability

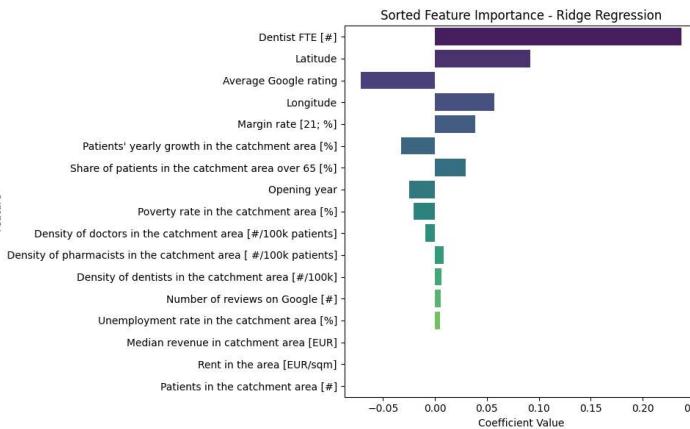
Simple Models :

Model	Mean Squared Error (MSE)	Mean Absolute Error (MAE)	R-squared (R^2)
0 Linear Regression	0.279084	0.400619	0.821116
1 Ridge Regression	0.279096	0.400702	0.821108
2 Lasso Regression	0.292815	0.418032	0.812315

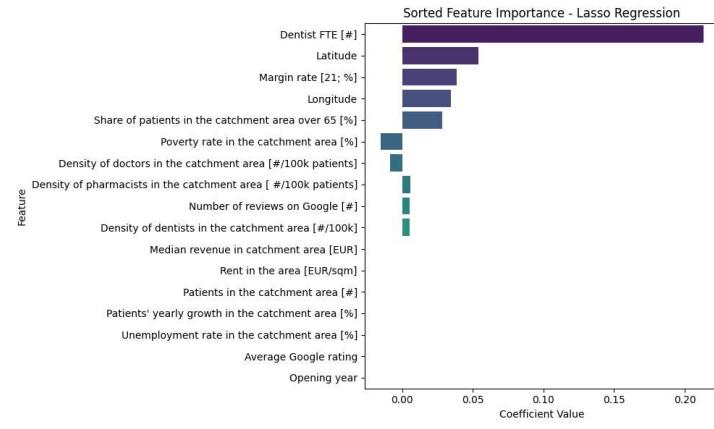
Feature



Feature



Feature



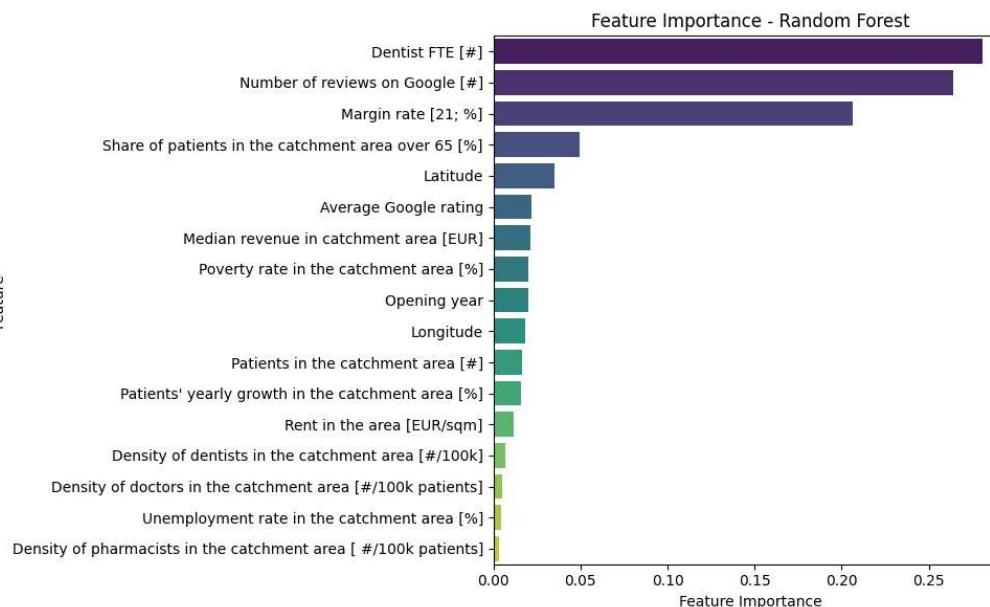
- **Internal** factors have significant positive impacts on revenue across all models
- **Average Google rating** show negative contributions, its relationships with revenue might be more complex or involve interactions with other factors.

4. Feature Importance and Performance Explainability

Non-Linear Models :

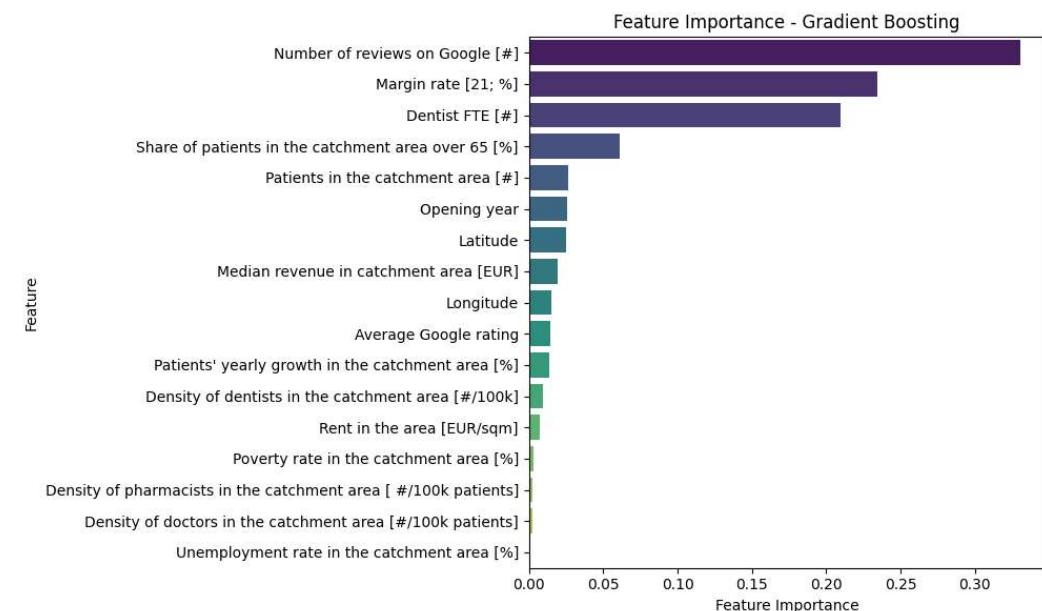
- Random Forest

Model	Mean Squared Error (MSE)	Mean Absolute Error (MAE)	R-squared (R ²)
0 Random Forest	0.08422	0.218299	0.946018



- Gradient Boosting :

Model	Mean Squared Error (MSE)	Mean Absolute Error (MAE)	R-squared (R ²)
0 Gradient Boosting	0.001635	0.034075	0.998952



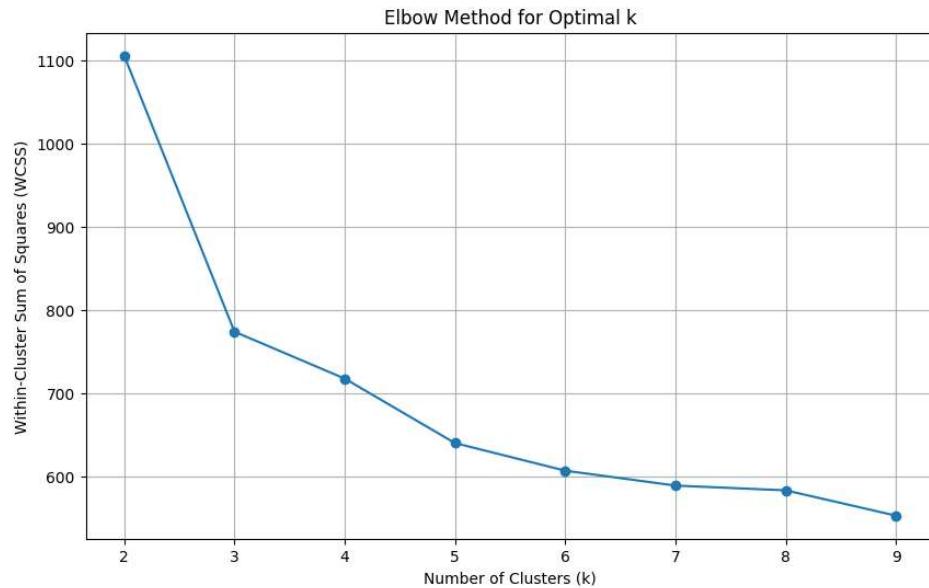
- `Number of reviews on Google [#]` emerges as the most influential feature, accounting for 33% of the importance.
- `Margin rate [21; %]` and `Dentist FTE [#]` remain critical, aligning with the findings from linear models.
- `Average Google rating`, while still less impactful than other internal factors, shows a slight increase in importance compared to linear models.

-
- Internal factors consistently emerge as the **most critical** drivers of revenue across all model like `Dentist FTE [#]`, `Number of reviews on Google [#]`, and `Margin rate [21; %]` that are the top contributors
 - External factors play a **secondary** role in influencing revenue with demographic variables like `Share of patients in the catchment area over 65 [%]` and `Patients in the catchment area [#]` that are moderately important.

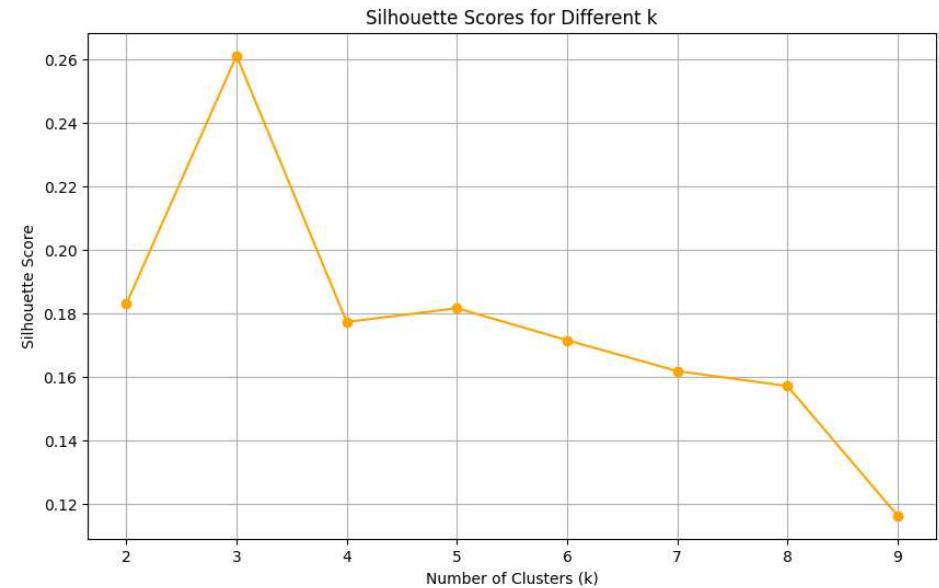
5. Clients Segmentation

Unsupervised method : K-Means

- Elbow Method :



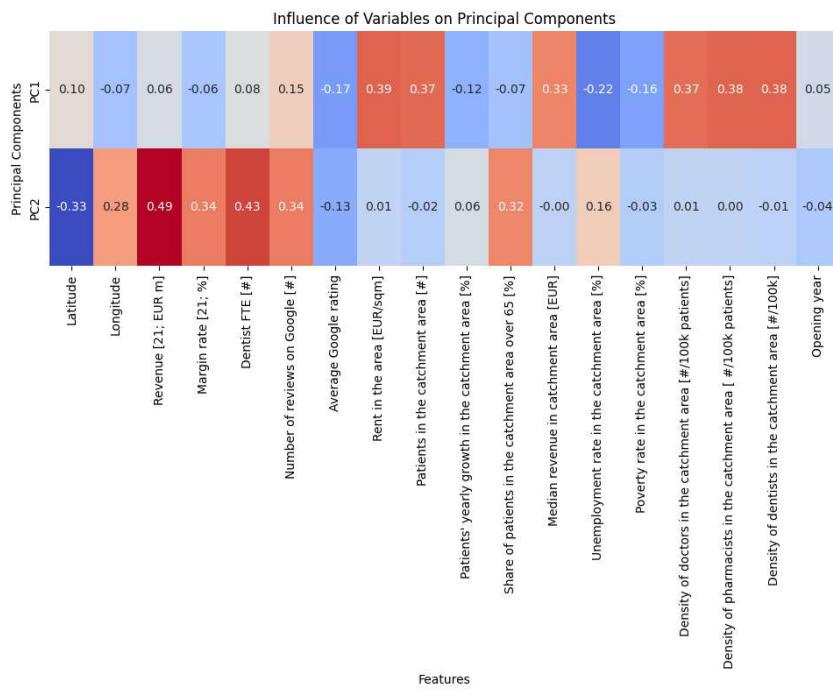
- Silhouette Score :



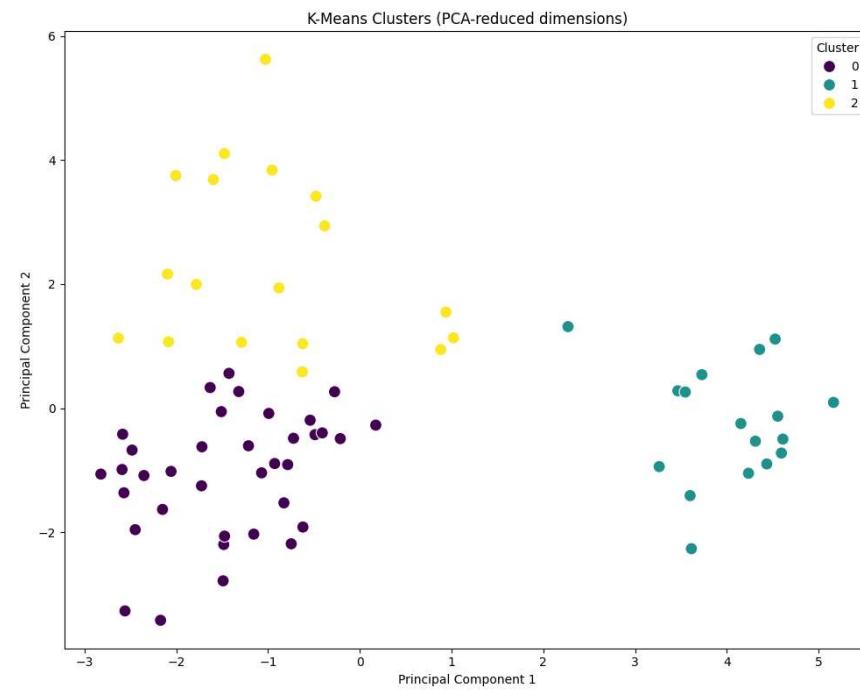
5. Clients Segmentation

Unsupervised method : K-Means

- PCA Analysis :



- Clusters Visualization:



5. Clients Segmentation

Unsupervised method : K-Means

- Cluster Summary :

	Latitude	Longitude	Revenue [21; EUR m]	Margin rate [21; %]	Dentist FTE [#]	Number of reviews on Google [#]	Average Google rating	Rent in the area [EUR/sqm]	Patients in the catchment area [#]	Patients' yearly growth in the catchment area [%]	Share of patients in the catchment area over 65 [%]	Median revenue in catchment area [EUR]	Unemployment rate in the catchment area [%]	Poverty rate in the catchment area [%]	Density of doctors in the catchment area [#/100k patients]	De phar cat p.	
Cluster	0	47.924617	2.609731	1.358502	12.901515	5.069865	78.127778	4.254754	3339.500000	48013.791137	0.215556	14.436781	20699.107190	8.163889	18.967887	152.338889	61
	1	48.858972	2.347932	2.341176	12.529412	6.764706	183.294118	3.878503	10885.823529	178821.971007	-0.415882	15.462680	28829.160072	6.411765	13.379118	242.647059	137
	2	44.595964	5.299251	3.361195	19.020202	8.632155	193.433333	4.159762	4192.055556	48383.600105	0.548333	23.042139	22096.778420	8.504167	15.698496	161.166667	67

- Cluster 0: Moderate Performers:**

Centers with moderate revenue and profitability.

Moderate staffing levels and reviews.

Located in regions with lower patient density and higher elderly population.

Recommendation: Focus on operational optimizations and building stronger customer engagement to boost Google ratings and profitability.

- Cluster 1: High-Demand Urban Centers:**

High revenue centers operating in high-cost, high-density urban areas.

Relatively lower Google ratings despite having many reviews.

Recommendation: Improve customer satisfaction (like patient experience, Google ratings) and optimize costs in these high-rent areas.

- Cluster 2: Top Performers:**

Centers with the highest revenue and profitability, well-staffed with good Google ratings.

Operate in areas with affluent demographics but moderate patient density.

Recommendation: Maintain operational excellence while exploring premium services or expanding in similar regions.



6. Concurrential Analysis

6.1. Identifying Competitors

- Data source :

 data.gouv.fr

Se connecter S'enregistrer

Recherche 

Données API Réutilisations Organisations Commencer sur data.gouv.fr Actualités Nous contacter + Publier sur data.gouv.fr

Ajouter aux favoris

Accueil > Jeux de données > FINESS Extraction du Fichier des établissements

FINESS Extraction du Fichier des établissements

Description

Liste des établissements du domaine sanitaire et social.

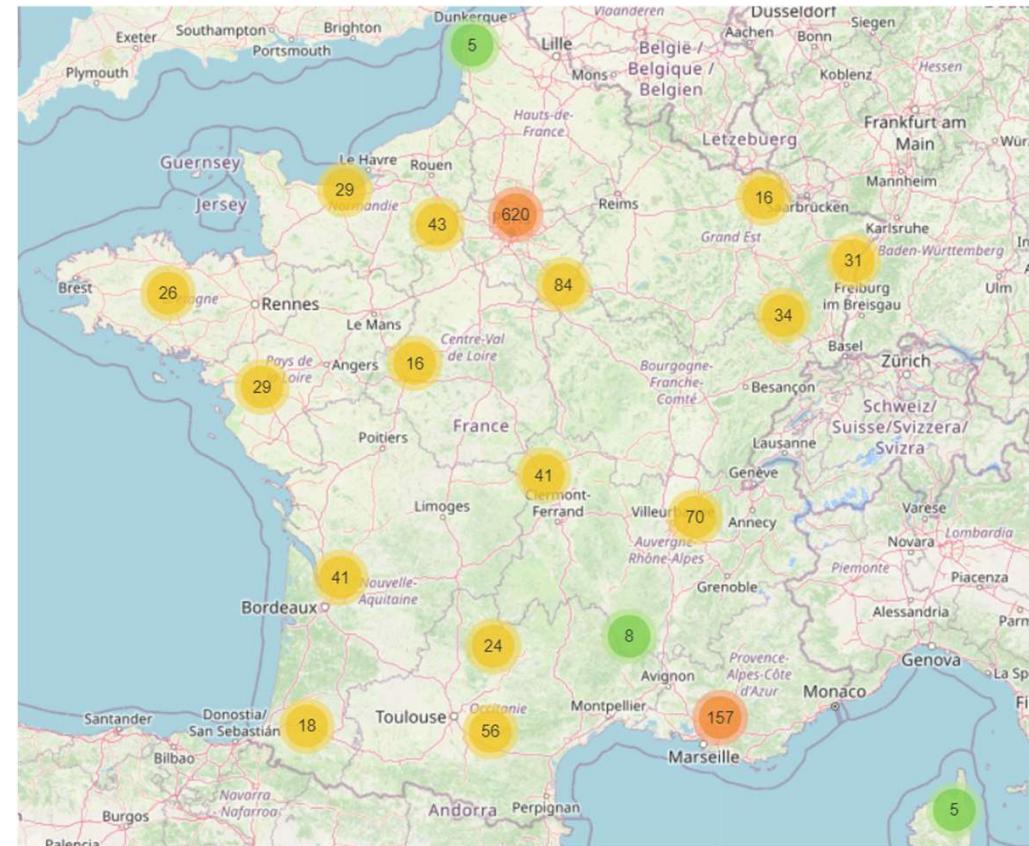
Informations sur la géolocalisation :

Tous les deux mois, le projet mutualisé des Ars – Atlasanté – géocode les adresses de toutes les structures Finess à partir d'une extraction réalisée par l'Agence du Numérique en Santé (ANS).

Jusqu'en septembre 2024, le géocodage s'appuyait sur le produit BD-ADRESSE v2.2 de l'IGN (qui n'était plus mis à jour depuis 2017-2018). Une nouvelle procédure de géocodage a donc été mise en place pour intégrer la Base d'Adresses Nationale (BAN), le référentiel d'adresses officiellement reconnu par l'administration.

- Relevant columns :

	Raison sociale	Coordonnées X	Coordonnées Y	Latitude	Longitude
0	CENTRE DE SANTE FERNEY-VOLTAIRE	939643.2	6577821.3	46.257763	6.111825
1	CENTRE DE SANTE DENTAIRE CLEMENCEAU	904204.7	6576314.4	46.255845	5.651367
2	CENTRE DE SANTE DENTAIRE D'OYONNAX	904155.7	6576568.8	46.258150	5.650842
3	CENTRE DE SANTE DE SAINT GENIS POUILLY	933186.6	6577271.7	46.255077	6.027794
4	CENTRE DE SANTE MEDICO-DENT VALSERHONE	918339.4	6560346.3	46.107727	5.827317
...
1348	CDS CENTRE MEDICO DENTAIRE LA FRETTE	639966.1	6875985.0	48.980995	2.179834
1349	CDS DENTAIRE D AUVERS SUR OISE	639666.0	6885963.2	49.070684	2.174309
1350	CDS CENTRE DENTAIRE CERGY GARE	627855.6	6883750.5	49.049571	2.013073
1351	CDS DENTAIRE FUTURSMILE OSNY	632655.1	6886864.4	49.078089	2.078227
1352	CDS CENTRE DENTAIRE SAINT GRATIEN	647553.9	6874141.9	48.965086	2.283719



all_competition_dental_centers.html

6. Concurrential Analysis

6.2. Identifying Direct Competitors

- Adding a new column :

Competitive Intensity : Number of competitors in the same catchment area (30m to 500m)

- Visualization :

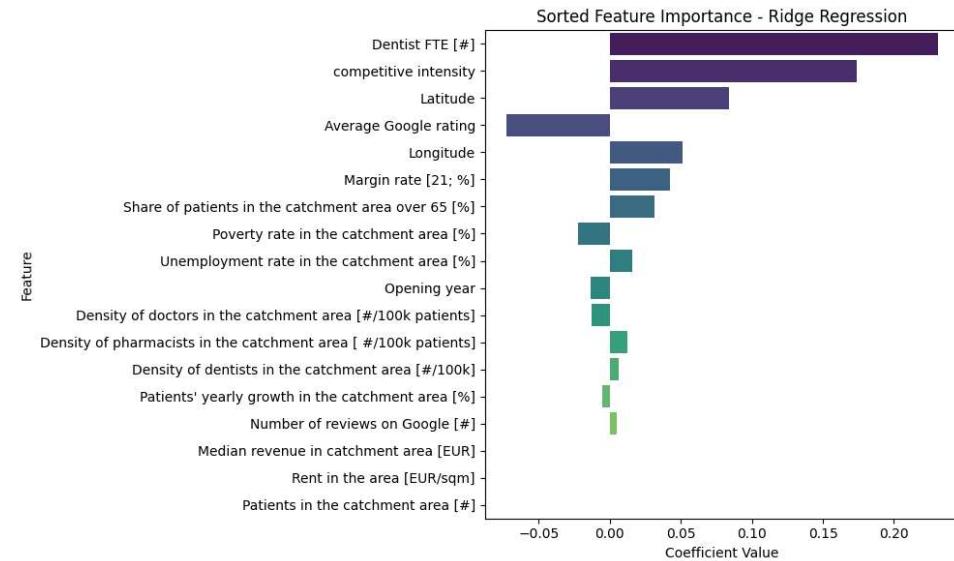
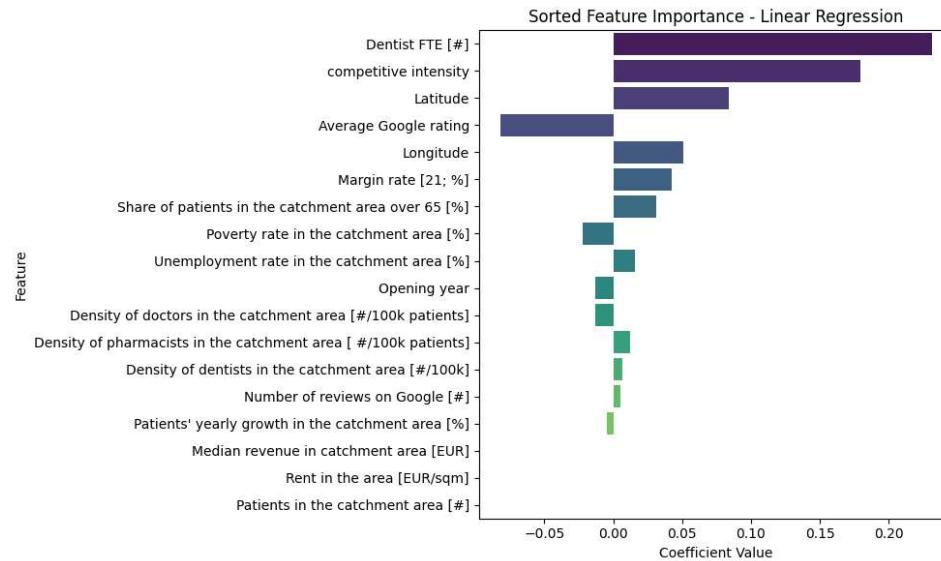


final_competitive_intensity_map.html

6. Concurrential Analysis

6.3. Reassessment with Competitive Intensity

- Feature Importance :



Observation:

Interestingly, we found a positive correlation between competitive intensity and revenue. Initially, it was expected that more competitors would result in lower revenues, as competition typically divides market share. However, the results reveal a different pattern.

Marketing Insight:

This finding aligns with a known marketing principle: the presence of competitors in close proximity can create a market concentration effect. When multiple centers exist in the same area, they collectively attract more clients, as potential customers are more likely to choose the area when they have multiple options. The decision is no longer whether to visit the area but rather which center to choose once there.

Balanced Competition:

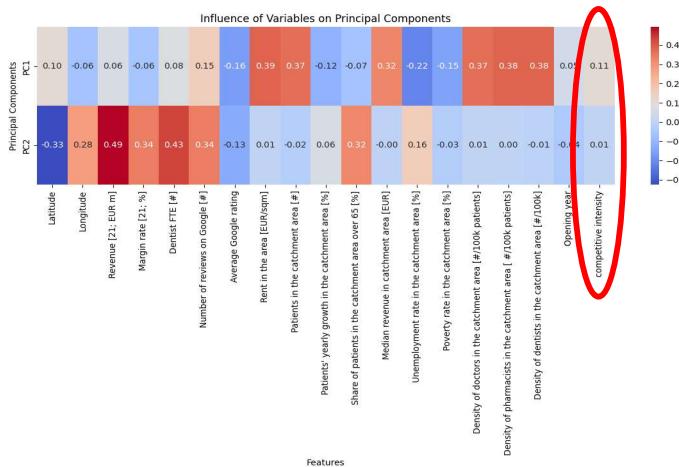
However, the positive correlation is not infinite. The analysis suggests there is an optimal level of competition:

- Having too few competitors may fail to create the market concentration effect, resulting in lower revenues.
- Having too many competitors may oversaturate the market, causing revenue to diminish.

6. Concurrential Analysis

6.3. Reassessment with Competitive Intensity

- Client Segmentation:
- PC Analysis :



- Cluster Summary :

Cluster	Revenue [21; EUR m]	Patients in the catchment area [#]	Patients' yearly growth in the catchment area [%]	competitive intensity
0	1.358502	48013.791137	0.215556	0.583333
1	2.341176	178821.971007	-0.415882	1.000000
2	3.361195	48383.600105	0.548333	0.611111

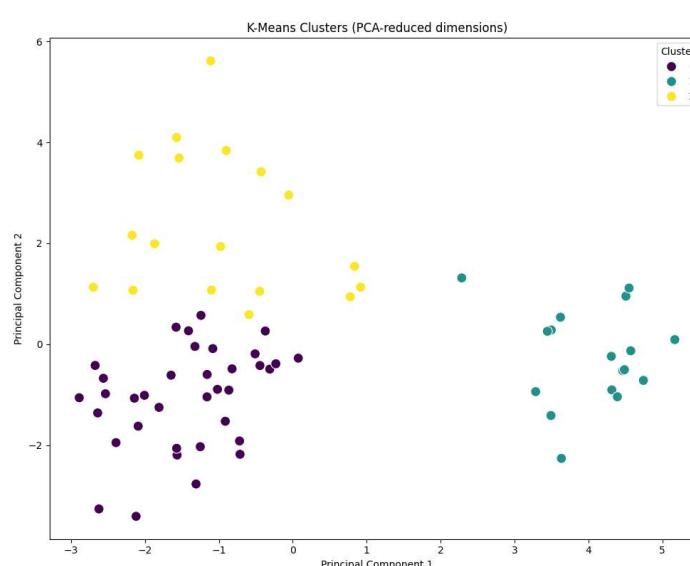
- **Low Competition (Cluster 0):** Fails to attract enough clients due to a lack of market concentration.
- **High Competition (Cluster 1):** Leads to diminishing returns as the market becomes oversaturated.
- **Moderate Competition (Cluster 2):** Strikes the right balance, creating enough market pull while keeping oversaturation in check.

Strategic Placement:

When opening new centers, aim for areas with moderate competitive intensity. These areas are likely to yield the best revenues, as they attract sufficient demand without being oversaturated.

Performance Optimization:

- Centers in areas with low competition may benefit from marketing strategies to increase local visibility and draw clients from neighboring areas.
- Centers in areas with high competition could focus on differentiation strategies (e.g., better customer service, specialized treatments) to capture a larger share of the local market.



7. Summary and Recommendations

Internal Factors as the Primary Drivers:

Dentist FTE (#), Number of Google reviews, and Margin rate (%) are the strongest predictors of revenue, emphasizing the importance of operational efficiency and customer engagement and centers with higher staffing levels and stronger online presence tend to perform better.

Role of External Factors:

External factors like Patients in the catchment area and Share of patients show moderate importance, indicating that while demographics matter, they are less influential than internal factors.

Competitive Intensity:

Interestingly, competitive intensity shows a positive correlation with revenue, but only up to a point. Moderate levels of competition enhance performance by creating market concentration, while oversaturation diminishes returns.

Cluster 0: Moderate Performers:

Centers with low revenue and low competitive intensity. Typically located in regions with lower patient density and higher elderly populations.

Recommendation: Focus on operational improvements and marketing strategies to boost customer engagement and visibility.

Cluster 1: High-Demand Urban Centers:

Centers with moderate revenue, operating in high-cost, high-density urban areas. These centers face intense competition but tend to underperform in customer satisfaction metrics (e.g., Google ratings).

Recommendation: Improve patient experience and differentiate services to gain a competitive edge in saturated markets.

Cluster 2: Top Performers:

Centers with the highest revenue and profitability, operating in areas with moderate competitive intensity and affluent demographics.

Recommendation: Maintain operational excellence while exploring premium service offerings or expanding into similar markets.

- **Strategic Expansion:** When opening new centers, prioritize areas with moderate competitive intensity and favorable demographics to maximize revenue potential and avoid markets with either low competition (insufficient demand) or high competition (oversaturated).
- **Operational Excellence:** Invest in staffing levels and ensure high-quality service delivery to boost operational efficiency and strengthen online presence by encouraging customer reviews and maintaining high Google ratings.
- **Localized Strategies:** For low-performing centers in areas with low competition, focus on increasing visibility through targeted marketing campaigns. In highly competitive areas, differentiate services through innovation and enhanced patient experiences.
- **Monitor Competitive Intensity:** Regularly assess competitive intensity and its impact on center performance to identify emerging opportunities or threats.

THANK YOU

