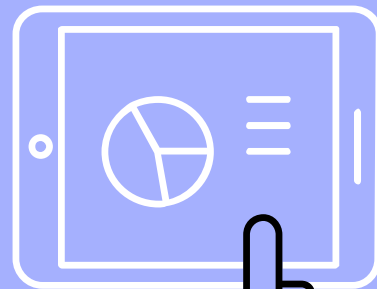
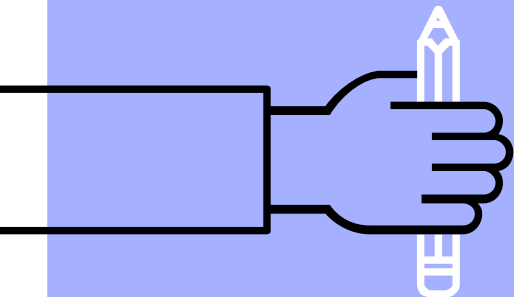
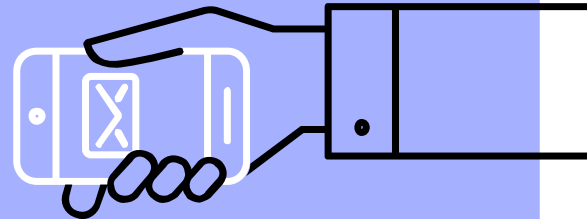
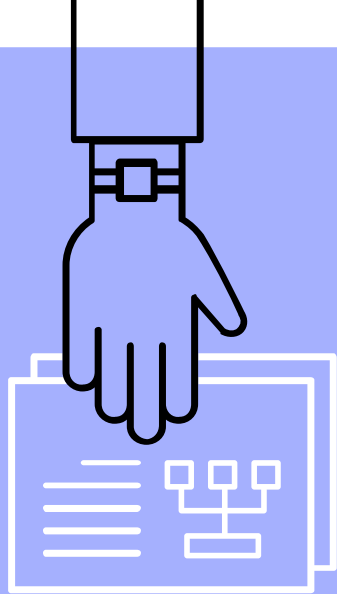


ANÁLISE: MULTINACIONAL DE VAREJO

Silas Genário



Objetivos

- ▶ Sugerir um grupo de municípios para a entrada de uma multinacional varejista no Brasil
 - Caracterizar os municípios em grupos
 - Classificar novos municípios entre os grupos



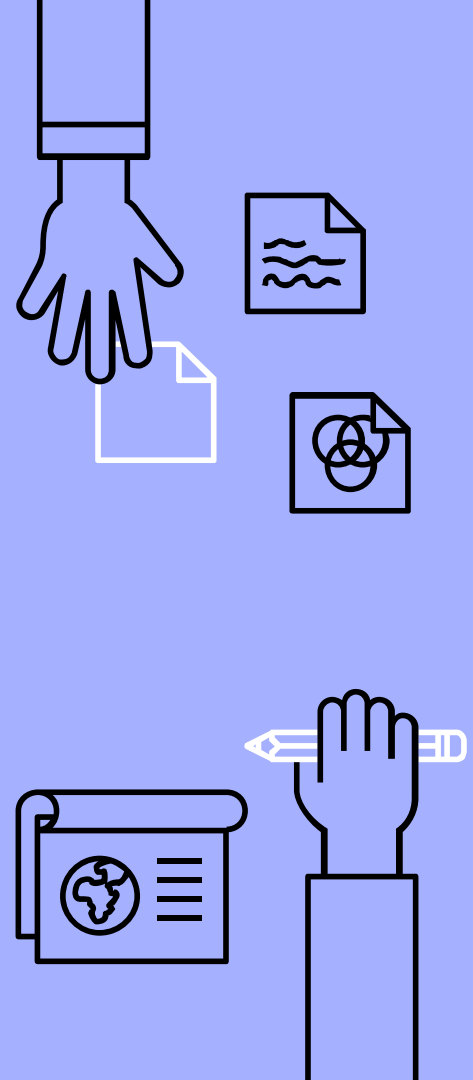
O que busca uma empresa de varejo?

Demanda Potencial

- ▶ Maior renda
- ▶ Maior densidade populacional
- ▶ Maior população urbana

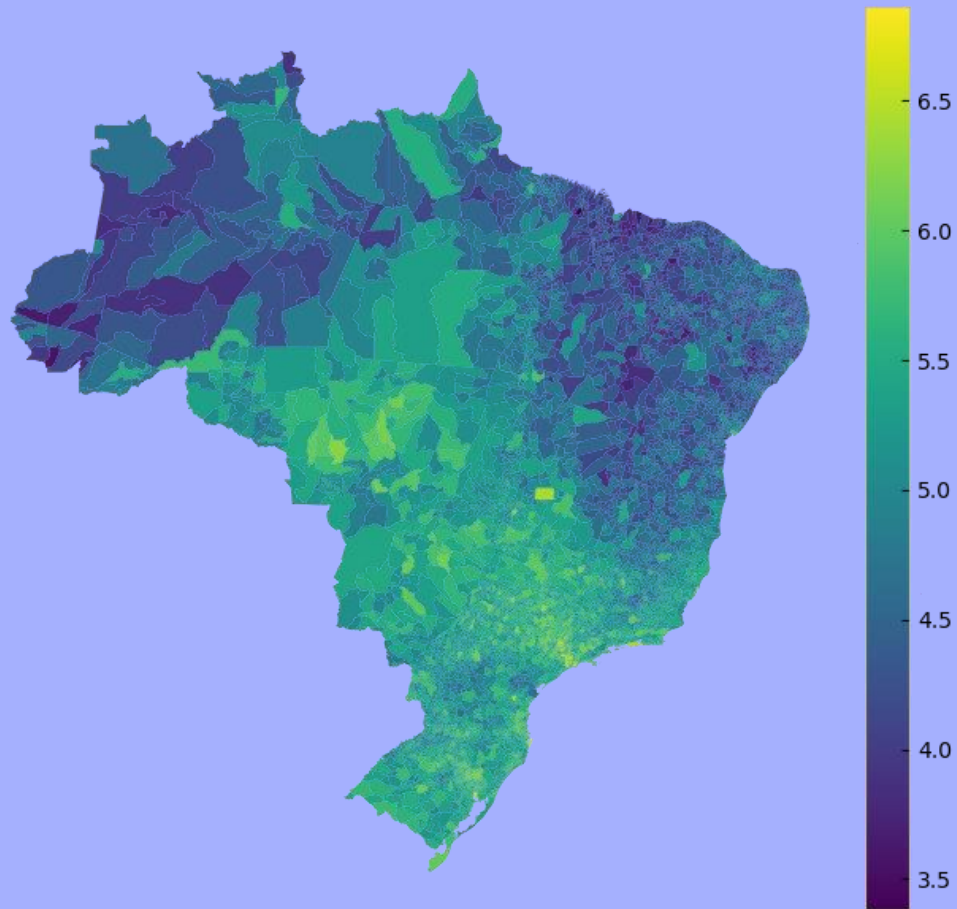
Baixos Custos

- ▶ Proximidade a centros de distribuição
- ▶ Proximidade à rodovias federais

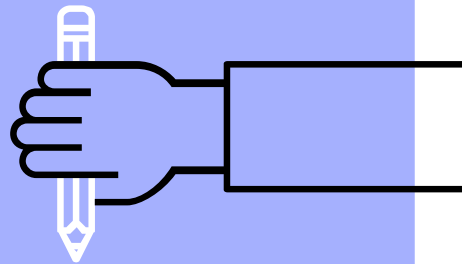


O problema

O Brasil é um país gigantesco. E a localização ideal nem sempre é simples de se decidir.

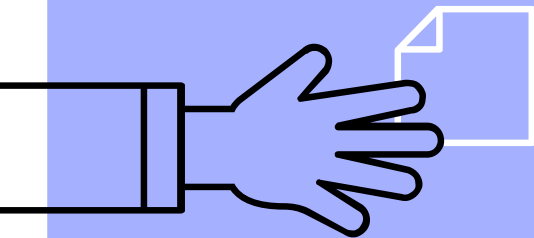


Renda Per Capita, 2000 (em Log)



Modelagem

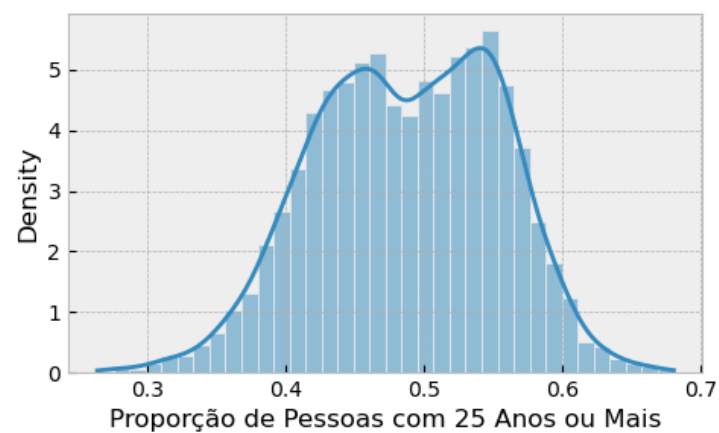
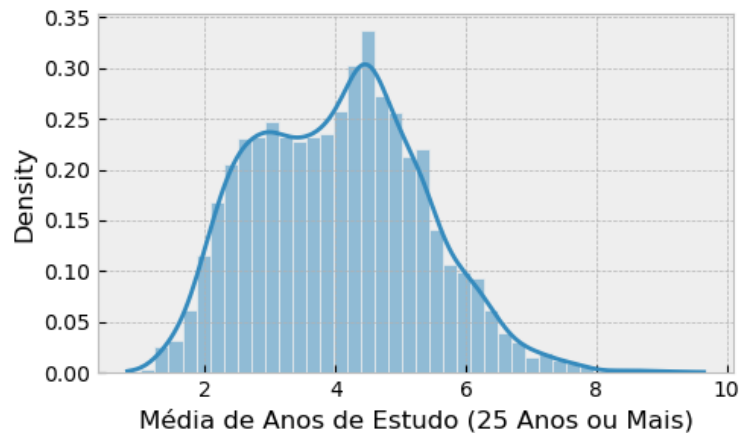
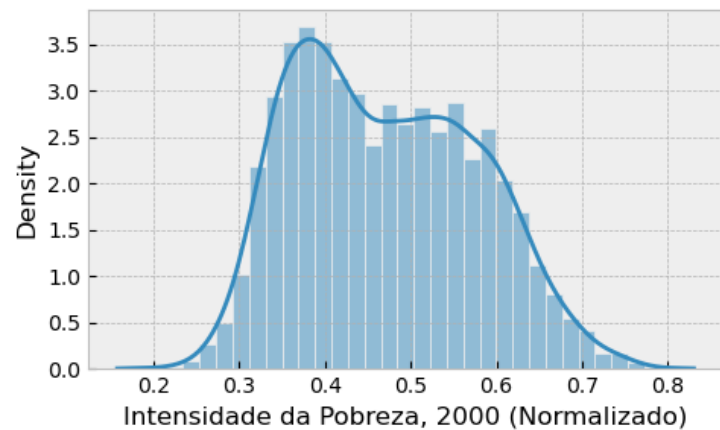
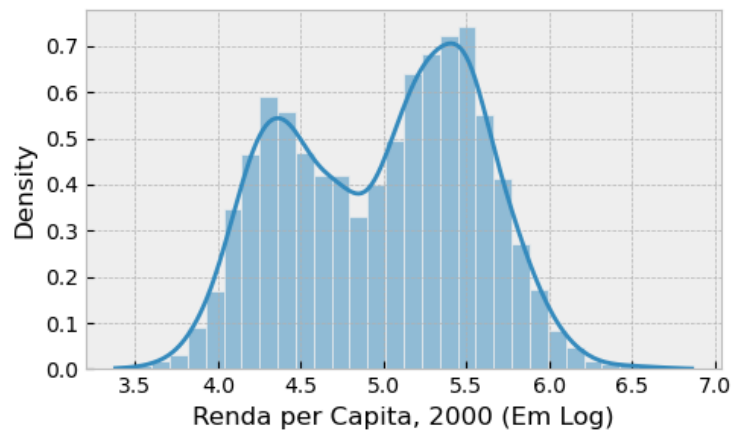
Parte I: Aprendizado Não-
Supervisionado



Objetivo

- ▶ Buscar agrupamento de cidades com potencial para ser porta de entrada da multinacional de supermercados





Método

- ▶ Existência de misturas de gaussianas;
- ▶ Gaussian Mixtures!



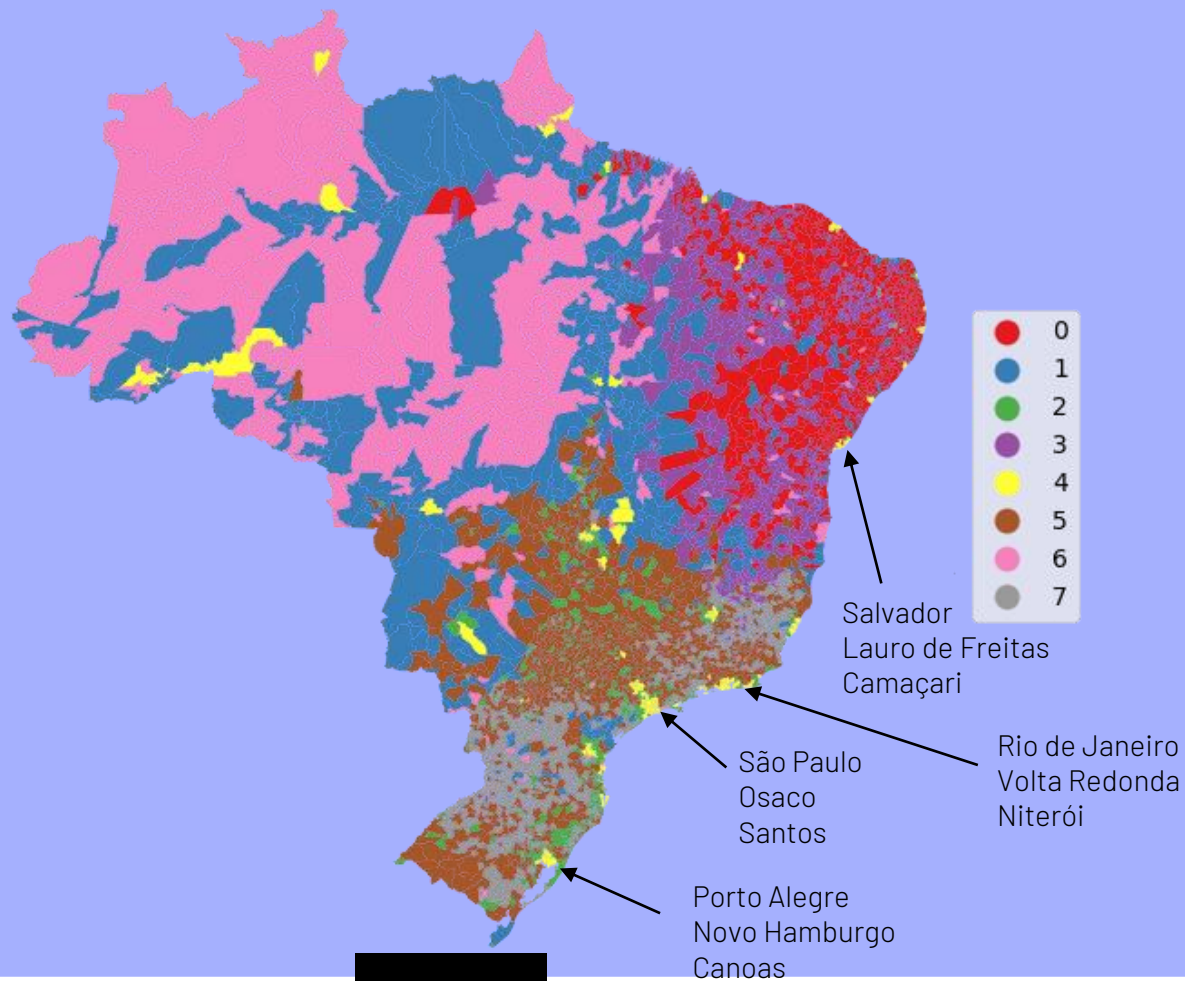


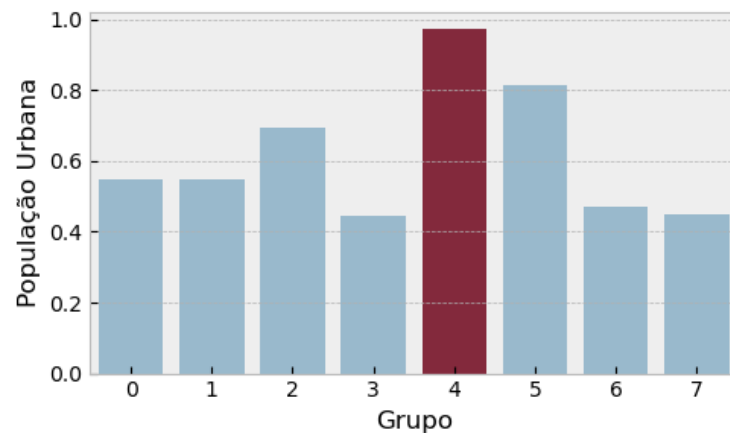
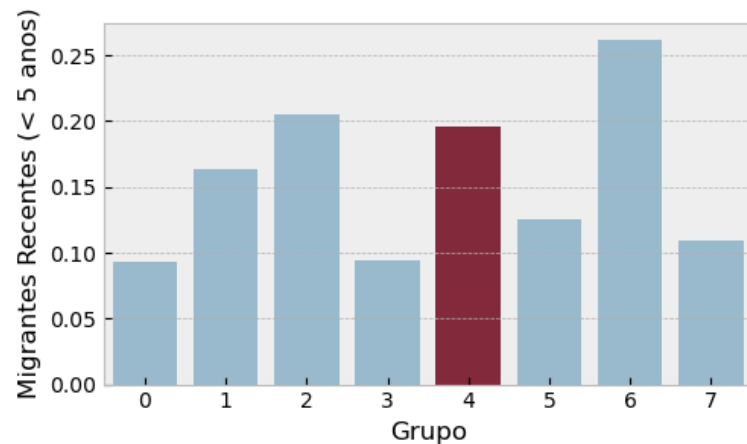
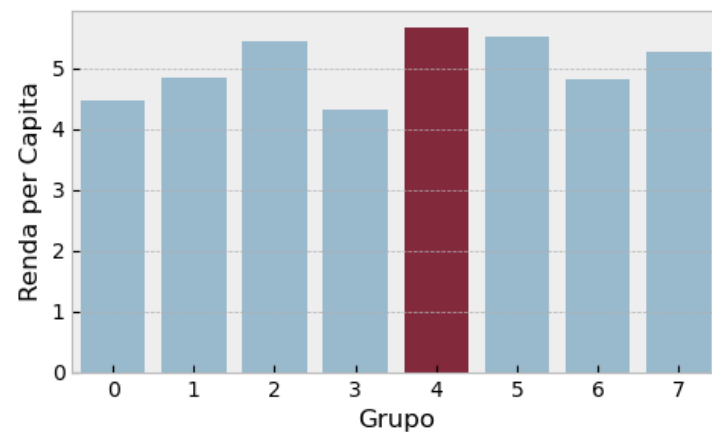
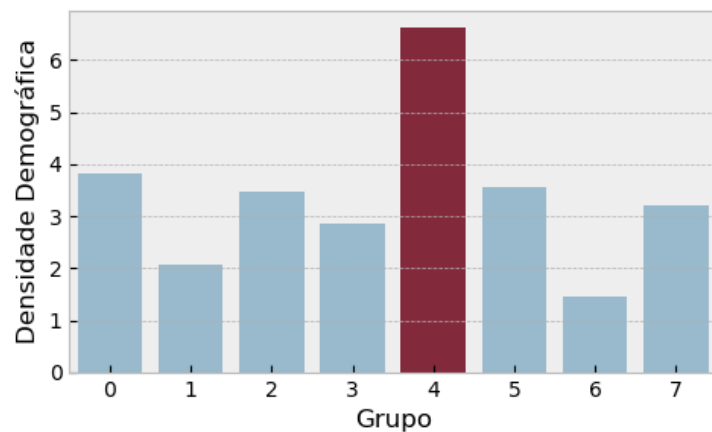
```
from sklearn.mixture import GaussianMixture

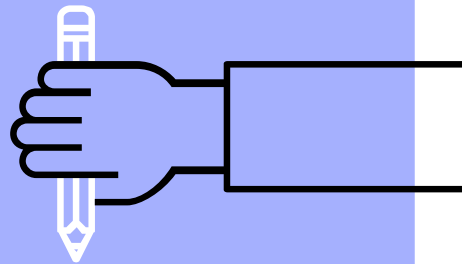
gm = GaussianMixture(n_components=8, random_state=42)
y_pred = gm.fit_predict(municipios_features)

municipios_com_geo["y_pred"] = y_pred
```

Cluster	Nº
0	895
1	562
2	472
3	921
4	172
5	1137
6	297
7	1051

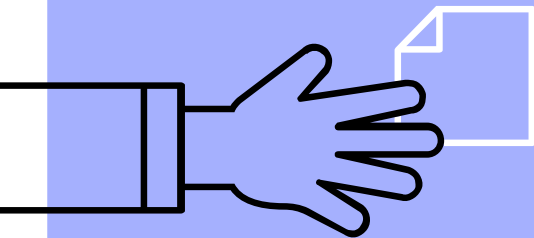






Modelagem

Parte II: Aprendizado
Supervisionado



Objetivo

- ▶ Classificar novos municípios entre os 8 grupos encontrados.



Método

- ▶ Modelo de Classificação: Random Forests





```
from sklearn.model_selection import train_test_split

X = municipios_com_geo[features]
y = municipios_com_geo["y_pred"]

X_train, X_test, y_train, y_test = train_test_split(
    X,
    y,
    test_size=0.2,
    random_state=42,
    stratify=y
)
```



```
from sklearn.ensemble import RandomForestClassifier

best_params = {'n_estimators': 183,
               'min_samples_split': 0.0006950940663198544,
               'min_samples_leaf': 6.84953650608236e-06,
               'max_depth': None,
               'max_features': 0.20117959445844186,
               'class_weight': None}

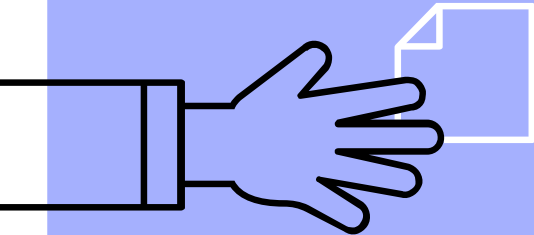
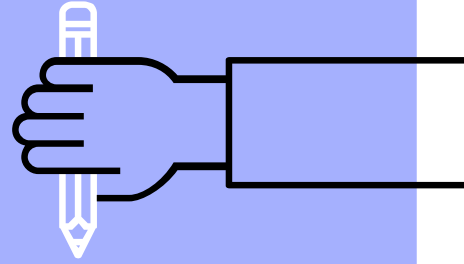
rf = RandomForestClassifier(random_state=42, **best_params)
rf.fit(X_train, y_train)
y_pred = rf.predict(X_test)
```


Métricas no Conjunto de Teste

- ▶ Acurácia: 0.809
- ▶ ROC AUC: 0.977
- ▶ Precision: 0.812
- ▶ Recall: 0.809
- ▶ F1 Score: 0.808

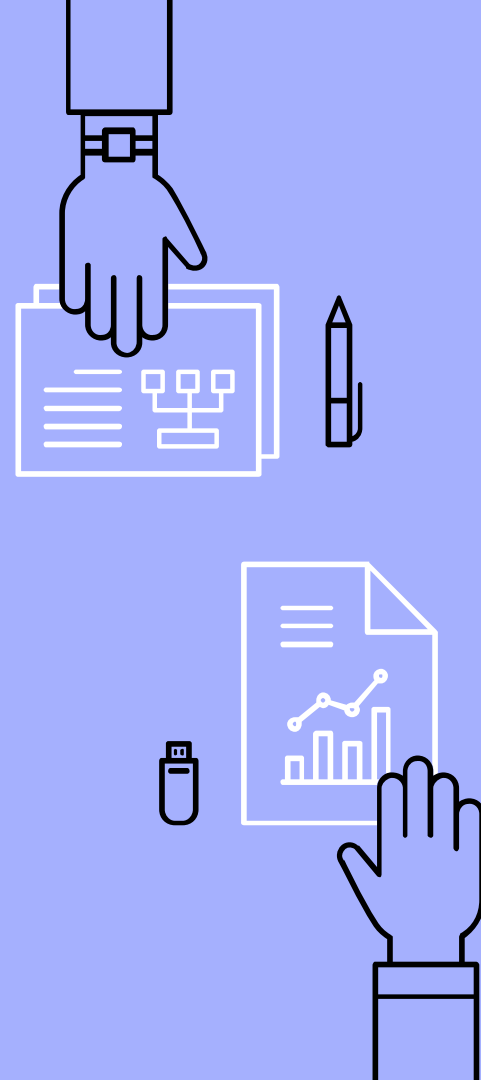


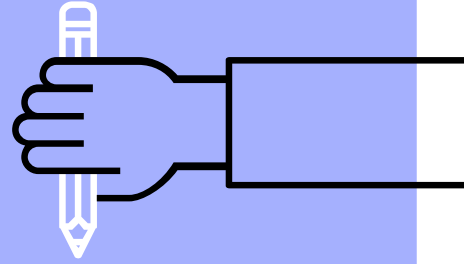
Conclusões



Qual grupo de municípios tem a melhor porta de entrada para a empresa?

- ▶ O grupo 4 fornece as melhores condições;
 - Demanda potencial
 - Menor custo





FIM

Repositório:

<https://github.com/silasge/case-plusoft/>

