Kmeans_Kmedoids

September 29, 2023

1 Mini-projeto K-Means

- Camila Vieira
- José Vinicius

##Importações

[1]:
```
!pip install kneed
!pip install scikit-learn-extra
```

```
Collecting kneed
  Downloading kneed-0.8.5-py3-none-any.whl (10 kB)
Requirement already satisfied: numpy>=1.14.2 in /usr/local/lib/python3.10/dist-
packages (from kneed) (1.23.5)
Requirement already satisfied: scipy>=1.0.0 in /usr/local/lib/python3.10/dist-
packages (from kneed) (1.11.2)
Installing collected packages: kneed
Successfully installed kneed-0.8.5
Collecting scikit-learn-extra
  Downloading scikit_learn_extra-0.3.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl (2.0 MB)
                           2.0/2.0 MB
22.6 MB/s eta 0:00:00
Requirement already satisfied: numpy>=1.13.3 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn-extra) (1.23.5)
Requirement already satisfied: scipy>=0.19.1 in /usr/local/lib/python3.10/dist-
packages (from scikit-learn-extra) (1.11.2)
Requirement already satisfied: scikit-learn>=0.23.0 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn-extra) (1.2.2)
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.10/dist-
packages (from scikit-learn>=0.23.0->scikit-learn-extra) (1.3.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn>=0.23.0->scikit-
learn-extra) (3.2.0)
Installing collected packages: scikit-learn-extra
Successfully installed scikit-learn-extra-0.3.0
```

[2]:
```
import numpy as np
import pandas as pd
import seaborn as sns
from sklearn import metrics
from kneed import KneeLocator
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from scipy.spatial import distance
from sklearn.decomposition import PCA
from sklearn_extra.cluster import KMedoids
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import silhouette_score, davies_bouldin_score,␣
↪calinski_harabasz_score, pairwise_distances
```

##Preparação dos Dados

[101]:
```
df = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/SI/College.csv")
df.head()
```

[101]:

	Unnamed: 0	Private	Apps	Accept	Enroll	Top10perc	\
0	Abilene Christian University	Yes	1660	1232	721	23	
1	Adelphi University	Yes	2186	1924	512	16	
2	Adrian College	Yes	1428	1097	336	22	
3	Agnes Scott College	Yes	417	349	137	60	
4	Alaska Pacific University	Yes	193	146	55	16	

	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	\
0	52	2885	537	7440	3300	450	2200	
1	29	2683	1227	12280	6450	750	1500	
2	50	1036	99	11250	3750	400	1165	
3	89	510	63	12960	5450	450	875	
4	44	249	869	7560	4120	800	1500	

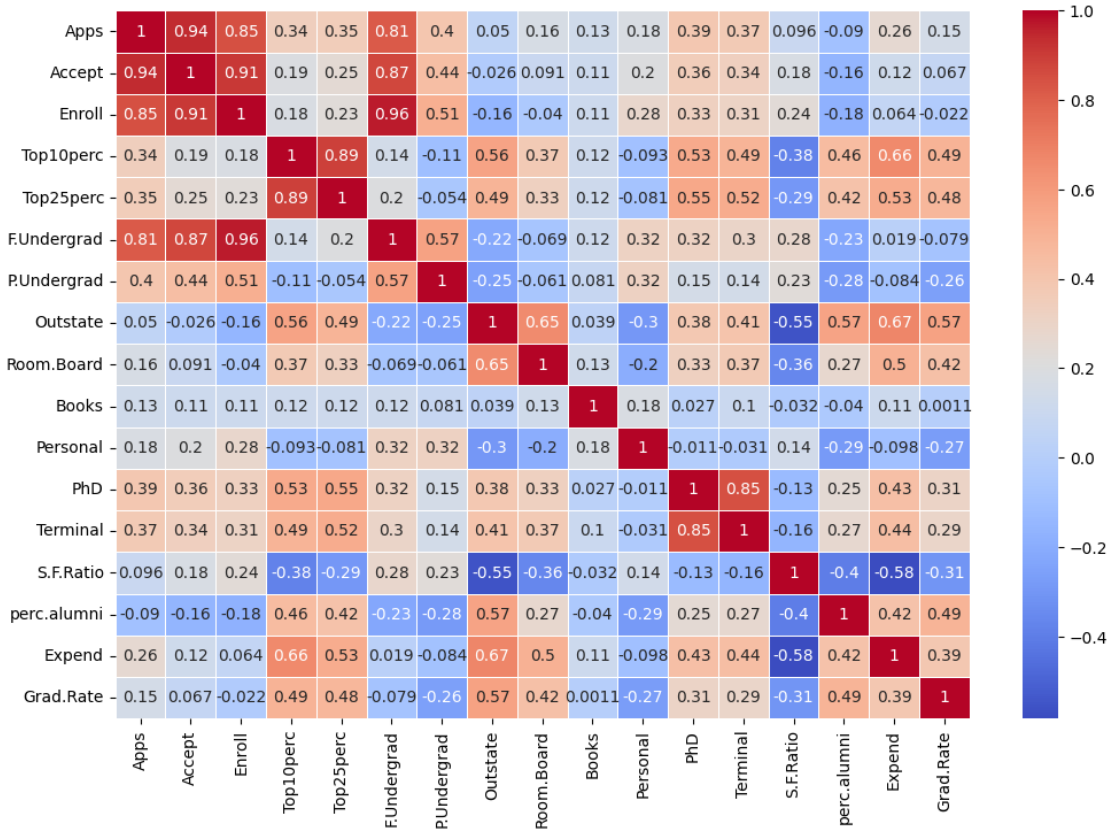
	PhD	Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate
0	70	78	18.1	12	7041	60
1	29	30	12.2	16	10527	56
2	53	66	12.9	30	8735	54
3	92	97	7.7	37	19016	59
4	76	72	11.9	2	10922	15

[102]:
```
df['Private'] = df['Private'].replace({'No': 0, 'Yes': 1})
binary_att = df[['Private']]
df = df.iloc[:, 2:]
df.info() #sem valores do tipo float ou categóricos
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 777 entries, 0 to 776
Data columns (total 17 columns):
```

```
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Apps         777 non-null    int64
 1   Accept       777 non-null    int64
 2   Enroll       777 non-null    int64
 3   Top10perc    777 non-null    int64
 4   Top25perc    777 non-null    int64
 5   F.Undergrad  777 non-null    int64
 6   P.Undergrad  777 non-null    int64
 7   Outstate     777 non-null    int64
 8   Room.Board   777 non-null    int64
 9   Books        777 non-null    int64
 10  Personal     777 non-null    int64
 11  PhD          777 non-null    int64
 12  Terminal     777 non-null    int64
 13  S.F.Ratio    777 non-null    float64
 14  perc.alumni  777 non-null    int64
 15  Expend       777 non-null    int64
 16  Grad.Rate    777 non-null    int64
dtypes: float64(1), int64(16)
memory usage: 103.3 KB
```

[103]:
```
correlacao = df.corr()
plt.figure(figsize=(12, 8))
sns.heatmap(correlacao, annot=True, cmap='coolwarm', linewidths=0.5)
plt.show()
```

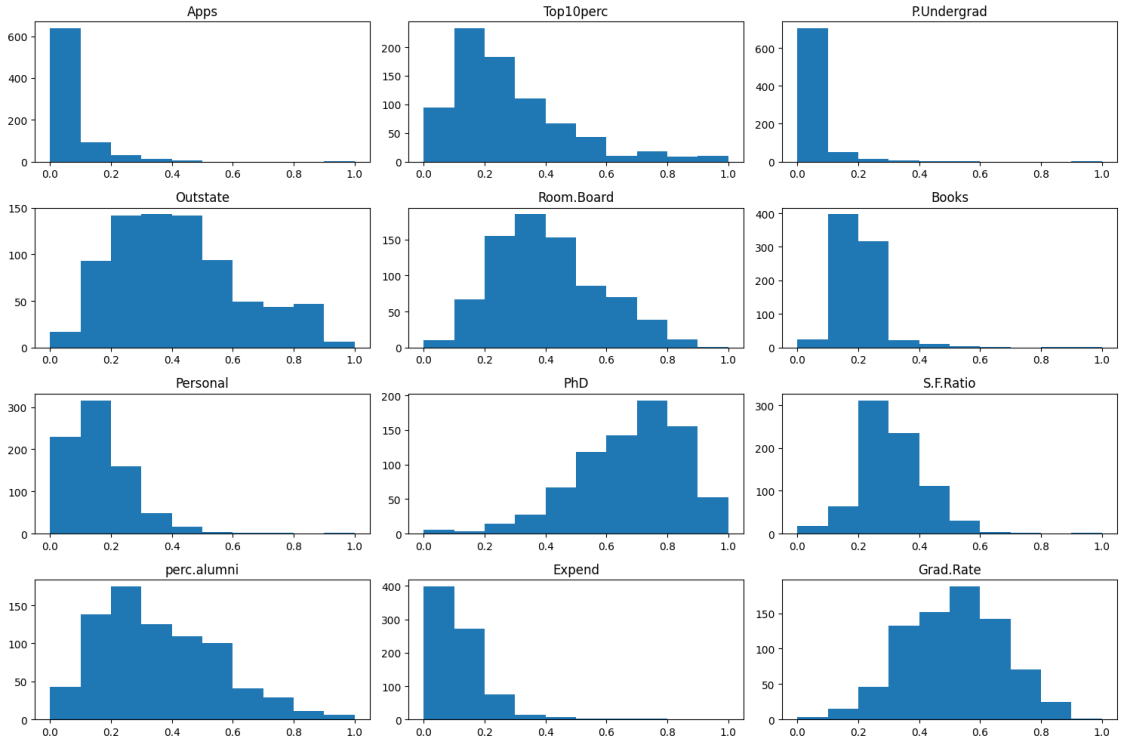


[104]:
```
#Remover as colunas com alta correlação para diminuir a dimensionalidade do modelo
colunas_para_remover = set()
for i in range(len(correlacao.columns)):
  for j in range(i):
    if abs(correlacao.iloc[i, j]) > 0.8:
      coluna = correlacao.columns[i]
      colunas_para_remover.add(coluna)
    if abs(correlacao.iloc[i, j]) < -0.8:
      coluna = correlacao.columns[i]
      colunas_para_remover.add(coluna)
print(colunas_para_remover)
df_sem_correlacao_alta = df.drop(columns=colunas_para_remover)
```

```
{'Terminal', 'F.Undergrad', 'Accept', 'Top25perc', 'Enroll'}
```

[105]:
```
#Normalização do modelo
scaler = MinMaxScaler()
df_normalized = pd.DataFrame(scaler.fit_transform(df_sem_correlacao_alta),columns=df_sem_correlacao_alta.columns)
```

```
plt.figure(figsize=(15, 10))
for i, coluna in enumerate(df_normalized.columns, 1):
    plt.subplot(4, 3, i)
    plt.hist(df_normalized[coluna])
    plt.title(coluna)
plt.tight_layout()
plt.show()
```



[106]:
```
data = df_normalized
```

##Valor de k

[107]:
```
def bic_score(kmeans, X):
  centers = [kmeans.cluster_centers_]
  labels = kmeans.labels_
  m = kmeans.n_clusters
  n = np.bincount(labels)
  N, d = X.shape

  cl_var = (1.0 / (N - m) / d) * sum([sum(distance.cdist(X.iloc[labels == i],␣
↪[centers[0][i]], 'euclidean')**2) for i in range(m)])
  const_term = 0.5 * m * np.log(N) * (d+1)
```

```
  BIC = np.sum([n[i] * np.log(n[i]) - n[i] * np.log(N) - ((n[i] * d) / 2) * np.
↪log(2*np.pi*cl_var) - ((n[i] - 1) * d / 2) for i in range(m)]) - const_term

  return (BIC)
```

[108]:
```
bic_scores = []
inertia_values = []
silhouette_scores = []
davies_bouldin_scores = []
calinski_harabasz_scores = []
min_clusters = 2
max_clusters = 10

kmeans = KMeans(n_clusters=1, random_state=42).fit(data)
inertia_values.append(kmeans.inertia_)

for n_clusters in range(min_clusters, max_clusters):
  kmeans = KMeans(n_clusters=n_clusters, random_state=42)
  kmeans.fit(data)

  inertia_values.append(kmeans.inertia_)
  bic_scores.append(bic_score(kmeans, data))
  silhouette_scores.append(silhouette_score(data, kmeans.labels_))
  davies_bouldin_scores.append(davies_bouldin_score(data, kmeans.labels_))
  calinski_harabasz_scores.append(calinski_harabasz_score(data, kmeans.labels_))
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
```

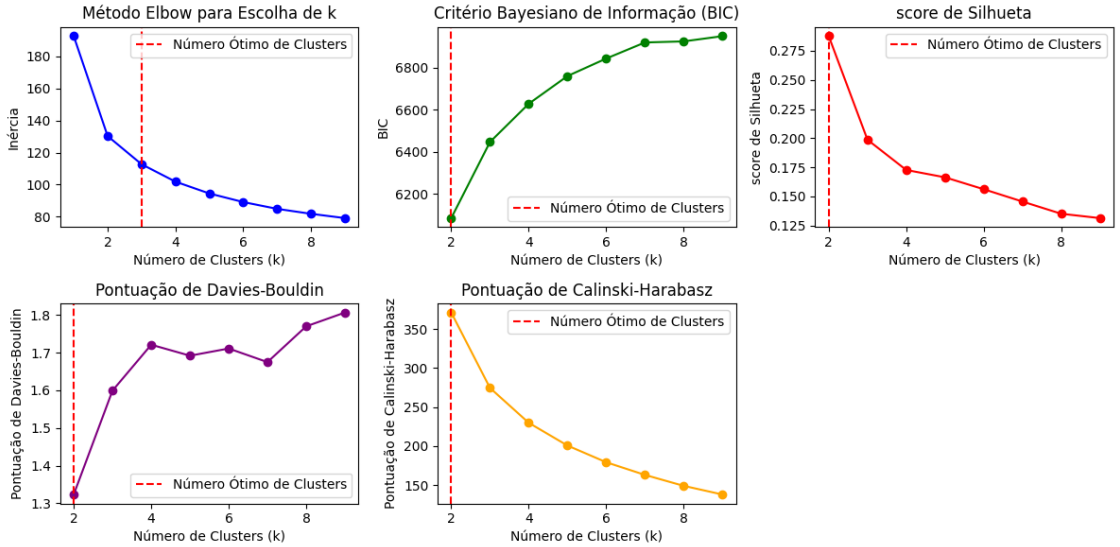
```
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
```

[109]:

```
knee_inertia = KneeLocator(range(min_clusters-1, max_clusters), inertia_values,␣
 ↪curve='convex', direction='decreasing')
optimal_clusters_inertia = knee_inertia.elbow

optimal_clusters_bic = np.argmin(bic_scores) + min_clusters
optimal_clusters_silhouette = np.argmax(silhouette_scores) + min_clusters
optimal_clusters_davies_bouldin = np.argmin(davies_bouldin_scores) +␣
 ↪min_clusters
optimal_clusters_calinski_harabasz = np.argmax(calinski_harabasz_scores) +␣
 ↪min_clusters

# Plot da inércia
plt.figure(figsize=(12, 6))
plt.subplot(2, 3, 1)
plt.plot(range(min_clusters-1, max_clusters), inertia_values, marker='o',␣
 ↪linestyle='-', color='b')
plt.axvline(x=optimal_clusters_inertia, color='red', linestyle='--',␣
 ↪label='Número Ótimo de Clusters')
plt.xlabel('Número de Clusters (k)')
plt.ylabel('Inércia')
plt.title('Método Elbow para Escolha de k')
plt.legend()

# Plot do BIC
plt.subplot(2, 3, 2)
plt.plot(range(min_clusters, max_clusters), bic_scores, marker='o',␣
 ↪linestyle='-', color='g')
plt.axvline(x=optimal_clusters_bic, color='red', linestyle='--', label='Número␣
 ↪Ótimo de Clusters')
plt.xlabel('Número de Clusters (k)')
plt.ylabel('BIC')
```

```
plt.title('Critério Bayesiano de Informação (BIC)')
plt.legend()

# Plot do escore de silhueta
plt.subplot(2, 3, 3)
plt.plot(range(min_clusters, max_clusters), silhouette_scores, marker='o',␣
 ↪linestyle='-', color='r')
plt.axvline(x=optimal_clusters_silhouette, color='red', linestyle='--',␣
 ↪label='Número Ótimo de Clusters')
plt.xlabel('Número de Clusters (k)')
plt.ylabel('score de Silhueta')
plt.title('score de Silhueta')
plt.legend()

# Plot da pontuação de Davies-Bouldin
plt.subplot(2, 3, 4)
plt.plot(range(min_clusters, max_clusters), davies_bouldin_scores, marker='o',␣
 ↪linestyle='-', color='purple')
plt.axvline(x=optimal_clusters_davies_bouldin, color='red', linestyle='--',␣
 ↪label='Número Ótimo de Clusters')
plt.xlabel('Número de Clusters (k)')
plt.ylabel('Pontuação de Davies-Bouldin')
plt.title('Pontuação de Davies-Bouldin')
plt.legend()

# Plot da pontuação de Calinski-Harabasz
plt.subplot(2, 3, 5)
plt.plot(range(min_clusters, max_clusters), calinski_harabasz_scores,␣
 ↪marker='o', linestyle='-', color='orange')
plt.axvline(x=optimal_clusters_calinski_harabasz, color='red', linestyle='--',␣
 ↪label='Número Ótimo de Clusters')
plt.xlabel('Número de Clusters (k)')
plt.ylabel('Pontuação de Calinski-Harabasz')
plt.title('Pontuação de Calinski-Harabasz')
plt.legend()

# Exibir os gráficos
plt.tight_layout()
plt.show()

print(f"Número ótimo de clusters (Elbow Method - Inércia):␣
 ↪{optimal_clusters_inertia}")
print(f"Número ótimo de clusters (BIC): {optimal_clusters_bic}")
print(f"Número ótimo de clusters (Escore de Silhueta):␣
 ↪{optimal_clusters_silhouette}")
```

```
print(f"Número ótimo de clusters (Davies-Bouldin):␣
  ↪{optimal_clusters_davies_bouldin}")
print(f"Número ótimo de clusters (Calinski-Harabasz):␣
  ↪{optimal_clusters_calinski_harabasz}")
```




```
Número ótimo de clusters (Elbow Method - Inércia): 3
Número ótimo de clusters (BIC): 2
Número ótimo de clusters (Escore de Silhueta): 2
Número ótimo de clusters (Davies-Bouldin): 2
Número ótimo de clusters (Calinski-Harabasz): 2
```

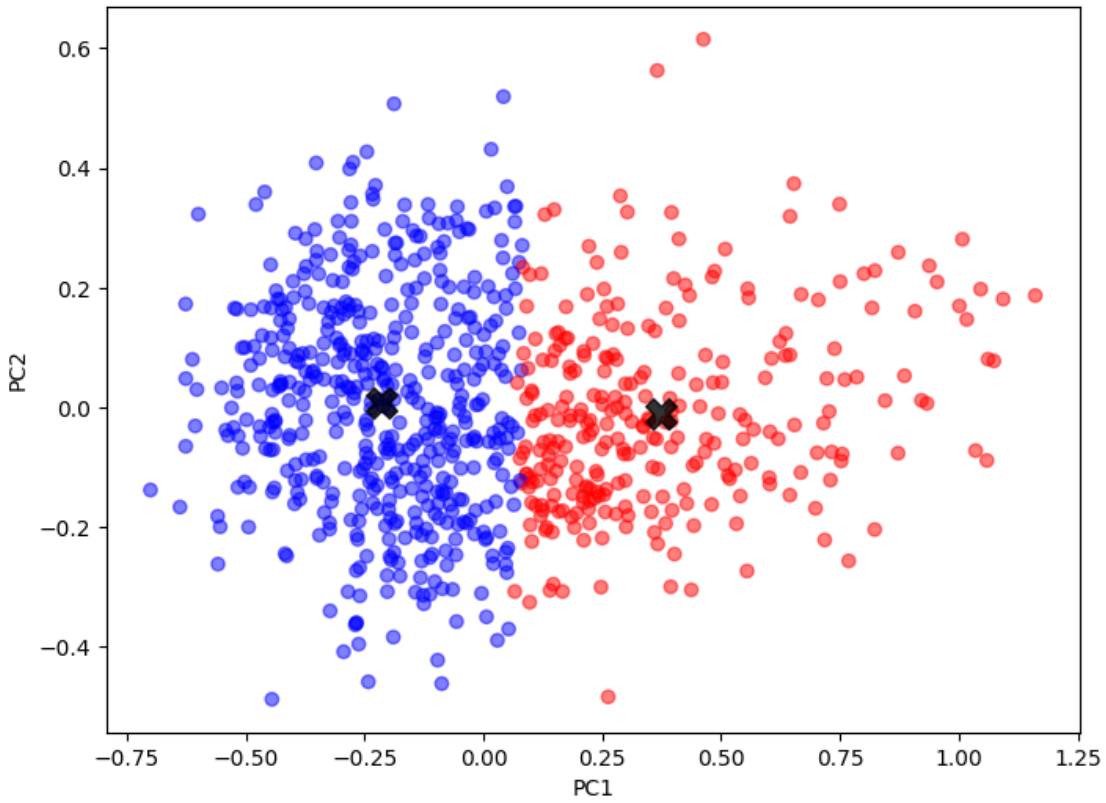
Melhor K com base nas métricas acima é 2.

1.1 K-Means

[110]:
```
#Treino com o melhor K
optimal_clusters = 2
kmeans_optimal = KMeans(n_clusters=optimal_clusters)
kmeans_optimal.fit(data)

pca = PCA(n_components=2)
X_pca = pca.fit_transform(data.values)

fig, ax = plt.subplots(figsize=(8, 6))
colors = np.array(['red', 'blue'])
ax.scatter(X_pca[:, 0], X_pca[:, 1], c=colors[kmeans_optimal.labels_], alpha=0.
  ↪5)
ax.scatter(pca.transform(kmeans_optimal.cluster_centers_)[:, 0], pca.
  ↪transform(kmeans_optimal.cluster_centers_)[:, 1],
```

```
           c='black', s=200, alpha=0.8, marker='X')
ax.set_xlabel('PC1')
ax.set_ylabel('PC2')
plt.show()
```

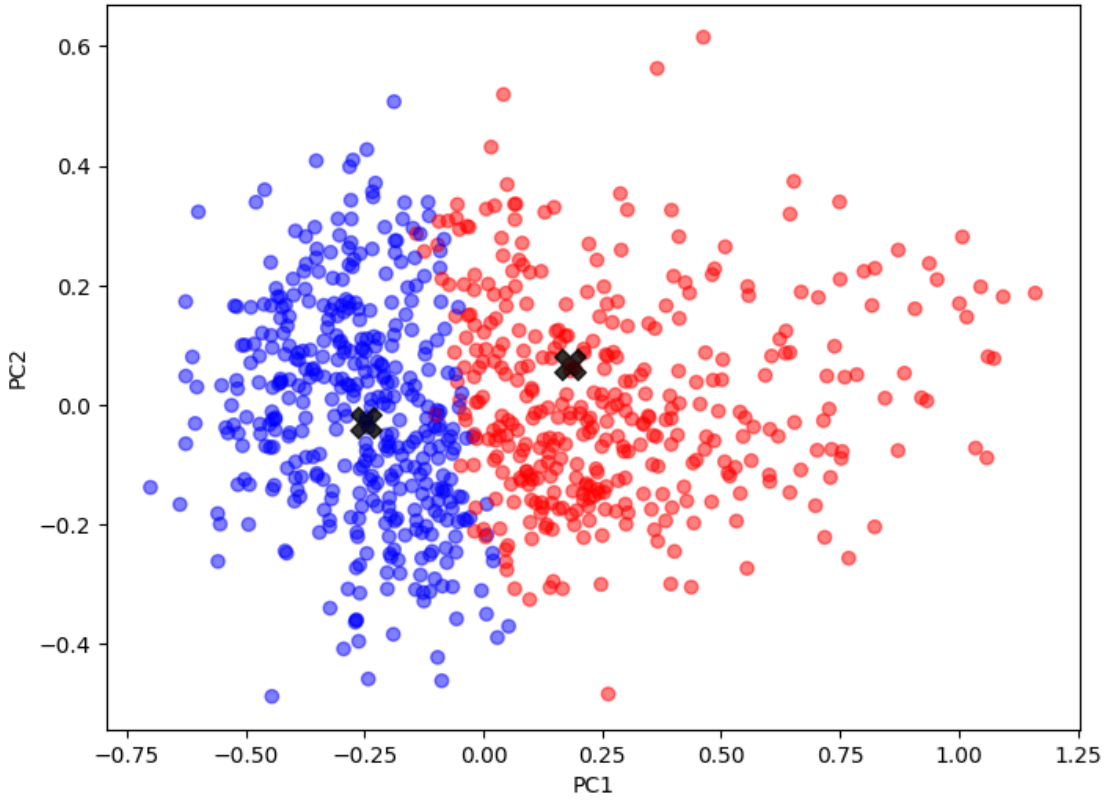
```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
```



1.2 K-Medoids

[111]:
```
#Treino com o melhor K
kmedoids_optimal = KMedoids(n_clusters=optimal_clusters)
kmedoids_optimal.fit(data)

pca = PCA(n_components=2)
X_pca = pca.fit_transform(data.values)

fig, ax = plt.subplots(figsize=(8, 6))
colors = np.array(['red', 'blue'])
```

```
ax.scatter(X_pca[:, 0], X_pca[:, 1], c=colors[kmedoids_optimal.labels_],␣
 ↪alpha=0.5)
ax.scatter(pca.transform(kmedoids_optimal.cluster_centers_)[:, 0], pca.
 ↪transform(kmedoids_optimal.cluster_centers_)[:, 1],
           c='black', s=200, alpha=0.8, marker='X')
ax.set_xlabel('PC1')
ax.set_ylabel('PC2')
plt.show()
```



##Comparação KMeans e KMedoids

[112]:
```
distances_kmeans = pairwise_distances(data, kmeans_optimal.cluster_centers_)
bic_value_kmeans = bic_score(kmeans, data)
silhouette_optimal_kmeans = silhouette_score(data, kmeans_optimal.labels_)
davies_bouldin_optimal_kmeans = davies_bouldin_score(data, kmeans_optimal.
 ↪labels_)
calinski_harabasz_optimal_kmeans = calinski_harabasz_score(data, kmeans_optimal.
 ↪labels_)

distances_kmedoids = pairwise_distances(data, kmedoids_optimal.cluster_centers_)
bic_value_kmedoids = bic_score(kmedoids_optimal, data)
```

```
silhouette_optimal_kmedoids = silhouette_score(data, kmedoids_optimal.labels_)
davies_bouldin_optimal_kmedoids = davies_bouldin_score(data, kmedoids_optimal.
 ↪labels_)
calinski_harabasz_optimal_kmedoids = calinski_harabasz_score(data,␣
 ↪kmedoids_optimal.labels_)
```

[113]:
```
print(f"Número ótimo de clusters: {optimal_clusters}\n\n")
print(f"Distâncias intra-clusters (KMeans):\n{distances_kmeans}")
print(f"Índice BIC (KMeans): {bic_value_kmeans:.2f}")
print(f"Silhouette Score (KMeans): {silhouette_optimal_kmeans:.2f}")
print(f"Davies-Bouldin Index (KMeans): {davies_bouldin_optimal_kmeans:.2f}")
print(f"Calinski-Harabasz Index (KMeans): {calinski_harabasz_optimal_kmeans:.
 ↪2f}\n")
print(f"Distâncias intra-clusters (KMedoid):\n{distances_kmedoids}")
print(f"Índice BIC (KMedoid): {bic_value_kmedoids:.2f}")
print(f"Silhouette Score (KMedoid): {silhouette_optimal_kmedoids:.2f}")
print(f"Davies-Bouldin Index (KMedoid): {davies_bouldin_optimal_kmedoids:.2f}")
print(f"Calinski-Harabasz Index (KMedoid): {calinski_harabasz_optimal_kmedoids:.
 ↪2f}")
```

```
Número ótimo de clusters: 2


Distâncias intra-clusters (KMeans):
[[0.693909   0.19951616]
 [0.75341966 0.62507088]
 [0.53535684 0.32267232]
 …
 [0.58644479 0.22774915]
 [0.94906651 1.4613338 ]
 [0.66106124 0.47515781]]
Índice BIC (KMeans): 6948.44
Silhouette Score (KMeans): 0.29
Davies-Bouldin Index (KMeans): 1.32
Calinski-Harabasz Index (KMeans): 371.26

Distâncias intra-clusters (KMedoid):
[[0.49694331 0.19556843]
 [0.73002531 0.65125503]
 [0.45556497 0.33021115]
 …
 [0.42880201 0.27033151]
 [1.11586012 1.48080371]
 [0.54176952 0.49450849]]
Índice BIC (KMedoid): 5396.08
Silhouette Score (KMedoid): 0.25
Davies-Bouldin Index (KMedoid): 1.42
```

```
Calinski-Harabasz Index (KMedoid): 337.91
```

Ao avaliar as métricas, o KMeans obteve um desempenho um pouco melhor, apesar de ter um BIC pior quando comparado com o KMedoids. Por isso, utilizaremos o KMeans para realizar as próximas comparações.

##Perguntas

Nesta seção é exibida as características estatíscas dos clusters formados pelo KMeans e também dos clusters feito a partir do atributo binário "Private". Apesar desta coluna ter sido retirada para o treinamento, ela aqui retorna para efeito comparativo. Apesar do KMeans ter divido em dois grupos o dataset, são grupos muito diferentes do intuito incial (Faculdade Privada ou Pública). Ao analisar medidas como a média e desvio padrão, a distribuição por variável e posteriormente a plotagem dos cluster em relação a "Private", torna-se claro as diferenças e possibilita uma análise mais profunda. Em outras palavras, a partir dos dados abaixos, podemos responder a média, a distribuição e consequentemente a importância de cada variável no cluster.

[114]:
```
data_cluster_0 = data[kmeans.labels_ == 0]
data_cluster_0.describe()
```

[114]:

	Apps	Top10perc	P.Undergrad	Outstate	Room.Board	Books	\
count	58.000000	58.000000	58.000000	58.000000	58.000000	58.000000	
mean	0.018577	0.152632	0.024759	0.386712	0.418430	0.214073	
std	0.021040	0.076955	0.021927	0.085563	0.130392	0.138013	
min	0.000396	0.000000	0.000504	0.200930	0.162673	0.057487	
25%	0.006733	0.094737	0.007900	0.324251	0.313288	0.149398	
50%	0.010456	0.152632	0.016922	0.384039	0.410230	0.180036	
75%	0.018958	0.207895	0.032665	0.443569	0.489833	0.224599	
max	0.106346	0.368421	0.100527	0.597107	0.766551	1.000000	

	Personal	PhD	S.F.Ratio	perc.alumni	Expend	Grad.Rate
count	58.000000	58.000000	58.000000	58.000000	58.000000	58.000000
mean	0.156670	0.349909	0.298650	0.222791	0.088318	0.475575
std	0.123343	0.132186	0.107519	0.107601	0.043457	0.150672
min	0.022901	0.000000	0.000000	0.031250	0.021415	0.129630
25%	0.074046	0.294737	0.257373	0.140625	0.063076	0.388889
50%	0.129771	0.368421	0.284182	0.234375	0.085075	0.453704
75%	0.190840	0.421053	0.345845	0.292969	0.102409	0.537037
max	0.711908	0.568421	0.662198	0.468750	0.276585	1.000000

[115]:
```
data_cluster_1 = data[kmeans.labels_ == 1]
data_cluster_1.describe()
```

[115]:

	Apps	Top10perc	P.Undergrad	Outstate	Room.Board	Books	\
count	96.000000	96.000000	96.000000	96.000000	96.000000	96.000000	
mean	0.037923	0.402961	0.005057	0.650935	0.446355	0.185453	
std	0.026848	0.117372	0.006960	0.129121	0.106837	0.047593	
min	0.002728	0.200000	0.000000	0.322211	0.237232	0.090909	
25%	0.019802	0.305263	0.001225	0.563481	0.371611	0.143828	

```
50%     0.032356   0.378947     0.002267   0.647469    0.421658   0.180036
75%     0.046373   0.484211     0.006549   0.734039    0.511901   0.224599
max     0.146544   0.705263     0.038425   1.000000    0.692623   0.313725

        Personal         PhD  S.F.Ratio  perc.alumni     Expend  Grad.Rate
count  96.000000   96.000000  96.000000    96.000000  96.000000  96.000000
mean    0.097055    0.798684   0.255502     0.597656   0.182150   0.612461
std     0.053238    0.092542   0.054923     0.126146   0.082082   0.093918
min     0.007634    0.284211   0.107239     0.359375   0.075952   0.351852
25%     0.057443    0.744737   0.224531     0.500000   0.134654   0.553241
50%     0.086336    0.800000   0.257373     0.578125   0.168483   0.625000
75%     0.130344    0.873684   0.281501     0.656250   0.206840   0.675926
max     0.297710    0.957895   0.442359     0.937500   0.749147   0.833333
```

[122]:
```
data['Private'] = binary_att
data_public_colleges = data[data['Private'] == 0]
data_private_colleges = data[data['Private'] == 1]
```

[117]:
```
data_public_colleges.describe()
```

[117]:
```
              Apps   Top10perc  P.Undergrad    Outstate  Room.Board  \
count  212.000000  212.000000   212.000000  212.000000  212.000000
mean     0.117654    0.229841     0.090551    0.231065    0.310252
std      0.111859    0.170320     0.106299    0.110808    0.135268
min      0.003166    0.000000     0.000366    0.012397    0.000000
25%      0.043941    0.115789     0.027433    0.156302    0.211460
50%      0.088018    0.189474     0.062926    0.220506    0.303909
75%      0.159155    0.278947     0.114232    0.284298    0.406999
max      1.000000    0.989474     1.000000    0.691736    0.750315

             Books    Personal         PhD   S.F.Ratio  perc.alumni  \
count  212.000000  212.000000  212.000000  212.000000   212.000000
mean     0.204268    0.217860    0.724578    0.392471     0.224351
std      0.060486    0.103438    0.129658    0.091637     0.117483
min      0.000000    0.022901    0.263158    0.112601     0.000000
25%      0.180036    0.145038    0.663158    0.337802     0.140625
50%      0.202317    0.213588    0.742105    0.395442     0.210938
75%      0.229947    0.275000    0.821053    0.451072     0.296875
max      0.458556    0.616489    1.000000    0.705094     0.750000

            Expend   Grad.Rate  Private
count  212.000000  212.000000    212.0
mean     0.080538    0.426319      0.0
std      0.050814    0.135032      0.0
min      0.007899    0.000000      0.0
25%      0.047675    0.333333      0.0
50%      0.066554    0.416667      0.0
```

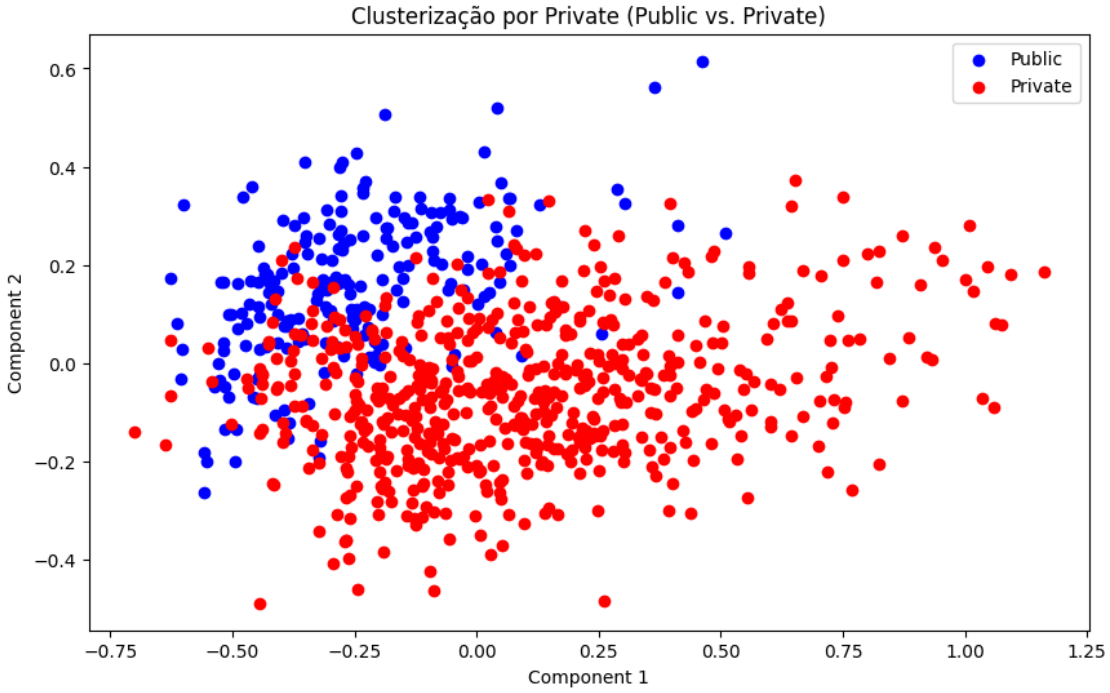
```
75%      0.101500    0.509259      0.0
max      0.251494    0.833333      0.0
```

[118]:
```
data_private_colleges.describe()
```

[118]:
```
              Apps   Top10perc  P.Undergrad    Outstate  Room.Board  \
count  565.000000  565.000000   565.000000  565.000000  565.000000
mean     0.039509    0.298221     0.019829    0.488724    0.442330
std      0.050889    0.187909     0.033083    0.191502    0.171768
min      0.000000    0.000000     0.000000    0.000000    0.093001
25%      0.011205    0.168421     0.002839    0.349174    0.308323
50%      0.021911    0.252632     0.009434    0.457645    0.412989
75%      0.043842    0.368421     0.024731    0.600723    0.570618
max      0.418866    1.000000     0.468056    1.000000    1.000000

             Books    Personal         PhD   S.F.Ratio  perc.alumni  \
count  565.000000  565.000000  565.000000  565.000000   565.000000
mean     0.201206    0.147243    0.664145    0.280040     0.404535
std      0.077956    0.096623    0.182641    0.094332     0.193762
min      0.068627    0.000000    0.000000    0.000000     0.031250
25%      0.157754    0.083969    0.547368    0.230563     0.250000
50%      0.180036    0.129771    0.684211    0.273458     0.390625
75%      0.224599    0.190840    0.810526    0.321716     0.531250
max      1.000000    1.000000    0.968421    1.000000     1.000000

           Expend   Grad.Rate  Private
count  565.000000  565.000000    565.0
mean     0.137620    0.546280      1.0
std      0.107123    0.155088      0.0
min      0.000000    0.046296      1.0
25%      0.080891    0.444444      1.0
50%      0.108734    0.546296      1.0
75%      0.159085    0.657407      1.0
max      1.000000    1.000000      1.0
```

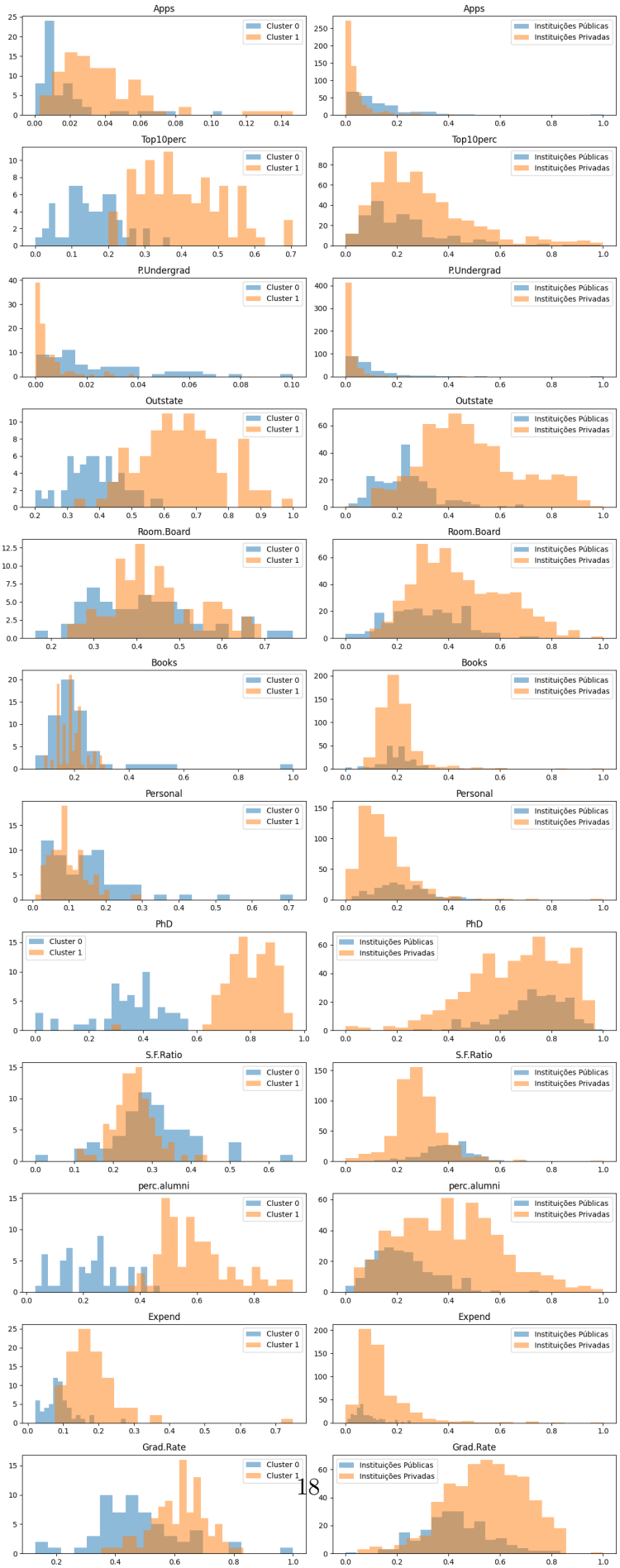
[119]:
```
pca = PCA(n_components=2)
data_pca = pca.fit_transform(data.iloc[:, :-1])  # Exclua a coluna 'Private'␣
↪para a redução de dimensionalidade
data_pca_with_private = pd.DataFrame(data_pca, columns=['Component 1',␣
↪'Component 2'])
data_pca_with_private['Private'] = data['Private']
data_public_colleges_pca =␣
↪data_pca_with_private[data_pca_with_private['Private'] == 0]
data_private_colleges_pca =␣
↪data_pca_with_private[data_pca_with_private['Private'] == 1]

plt.figure(figsize=(10, 6))
```

```
plt.scatter(data_public_colleges_pca['Component 1'],␣
 ↪data_public_colleges_pca['Component 2'], c='blue', label='Public')
plt.scatter(data_private_colleges_pca['Component 1'],␣
 ↪data_private_colleges_pca['Component 2'], c='red', label='Private')
plt.xlabel('Component 1')
plt.ylabel('Component 2')
plt.legend()
plt.title('Clusterização por Private (Public vs. Private)')
plt.show()
```



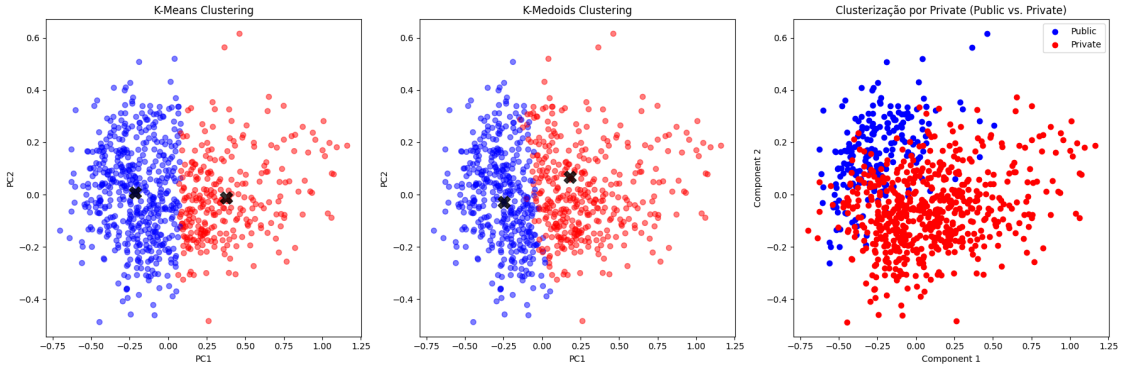
[126]:
```
columns_to_compare = data_cluster_0.columns
num_rows = len(columns_to_compare)
num_cols = 2

plt.figure(figsize=(12, 30))

for i, col in enumerate(columns_to_compare):
  plt.subplot(num_rows, num_cols, i * num_cols + 1)
  plt.hist(data_cluster_0[col], alpha=0.5, label='Cluster 0', bins=20)
  plt.hist(data_cluster_1[col], alpha=0.5, label='Cluster 1', bins=20)
  plt.title(col)
  plt.legend()

  plt.subplot(num_rows, num_cols, i * num_cols + 2)
```

```
  plt.hist(data_public_colleges[col], alpha=0.5, label='Instituições Públicas',␣
 ↪bins=20)
  plt.hist(data_private_colleges[col], alpha=0.5, label='Instituições␣
 ↪Privadas', bins=20)
  plt.title(col)
  plt.legend()

plt.tight_layout()
plt.show()
```


Apps
Cluster 0
Cluster 1
Apps
Instituições Públicas
Instituições Privadas
Top10perc
Cluster 0
Cluster 1
Top10perc
Instituições Públicas
Instituições Privadas
P.Undergrad
Cluster 0
Cluster 1
P.Undergrad
Instituições Públicas
Instituições Privadas
Outstate
Cluster 0
Cluster 1
Outstate
Instituições Públicas
Instituições Privadas
Room.Board
Cluster 0
Cluster 1
Room.Board
Instituições Públicas
Instituições Privadas
Books
Cluster 0
Cluster 1
Books
Instituições Públicas
Instituições Privadas
Personal
Cluster 0
Cluster 1
Personal
Instituições Públicas
Instituições Privadas
PhD
Cluster 0
Cluster 1
PhD
Instituições Públicas
Instituições Privadas
S.F.Ratio
Cluster 0
Cluster 1
S.F.Ratio
Instituições Públicas
Instituições Privadas
perc.alumni
Cluster 0
Cluster 1
perc.alumni
Instituições Públicas
Instituições Privadas
Expend
Cluster 0
Cluster 1
Expend
Instituições Públicas
Instituições Privadas
Grad.Rate
Cluster 0
Cluster 1
Grad.Rate
Instituições Públicas
Instituições Privadas

[130]:
```
# Primeiro gráfico: K-Means
pca = PCA(n_components=2)
X_pca = pca.fit_transform(data.iloc[:, :-1].values)

# Segundo gráfico: K-Medoids
pca_kmedoids = PCA(n_components=2)
X_pca_kmedoids = pca_kmedoids.fit_transform(data.iloc[:, :-1].values)

# Terceiro gráfico: Clusterização por Private
pca_private = PCA(n_components=2)
data_pca = pca_private.fit_transform(data.iloc[:, :-1])
data_pca_with_private = pd.DataFrame(data_pca, columns=['Component 1',␣
 ↪'Component 2'])
data_pca_with_private['Private'] = data['Private']
data_public_colleges_pca =␣
 ↪data_pca_with_private[data_pca_with_private['Private'] == 0]
data_private_colleges_pca =␣
 ↪data_pca_with_private[data_pca_with_private['Private'] == 1]


plt.figure(figsize=(18, 6))

# Subplot 1: K-Means
plt.subplot(1, 3, 1)
colors = np.array(['red', 'blue'])
plt.scatter(X_pca[:, 0], X_pca[:, 1], c=colors[kmeans_optimal.labels_], alpha=0.
 ↪5)
plt.scatter(pca.transform(kmeans_optimal.cluster_centers_)[:, 0], pca.
 ↪transform(kmeans_optimal.cluster_centers_)[:, 1],
            c='black', s=200, alpha=0.8, marker='X')
plt.xlabel('PC1')
plt.ylabel('PC2')
plt.title('K-Means Clustering')

# Subplot 2: K-Medoids
plt.subplot(1, 3, 2)
colors = np.array(['red', 'blue'])
plt.scatter(X_pca_kmedoids[:, 0], X_pca_kmedoids[:, 1],␣
 ↪c=colors[kmedoids_optimal.labels_], alpha=0.5)
plt.scatter(pca_kmedoids.transform(kmedoids_optimal.cluster_centers_)[:, 0],␣
 ↪pca_kmedoids.transform(kmedoids_optimal.cluster_centers_)[:, 1],
            c='black', s=200, alpha=0.8, marker='X')
plt.xlabel('PC1')
plt.ylabel('PC2')
```

```
plt.title('K-Medoids Clustering')

# Subplot 3: Clusterização por Private
plt.subplot(1, 3, 3)
plt.scatter(data_public_colleges_pca['Component 1'],␣
 ↪data_public_colleges_pca['Component 2'], c='blue', label='Public')
plt.scatter(data_private_colleges_pca['Component 1'],␣
 ↪data_private_colleges_pca['Component 2'], c='red', label='Private')
plt.xlabel('Component 1')
plt.ylabel('Component 2')
plt.legend()
plt.title('Clusterização por Private (Public vs. Private)')

plt.tight_layout()
plt.show()
```



##Salvar em PDF

[ ]:
```
!apt-get install texlive texlive-xetex texlive-latex-extra pandoc
!pip install pypandoc
```

```
Reading package lists… Done
Building dependency tree… Done
Reading state information… Done
pandoc is already the newest version (2.9.2.1-3ubuntu2).
pandoc set to manually installed.
The following additional packages will be installed:
  dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono
  fonts-texgyre fonts-urw-base35 libapache-pom-java libcommons-logging-java
  libcommons-parent-java libfontbox-java libfontenc1 libgs9 libgs9-common
  libidn12 libijs-0.35 libjbig2dec0 libkpathsea6 libpdfbox-java libptexenc1
  libruby3.0 libsynctex2 libteckit0 libtexlua53 libtexluajit2 libwoff1
  libzzip-0-13 lmodern poppler-data preview-latex-style rake ruby
  ruby-net-telnet ruby-rubygems ruby-webrick ruby-xmlrpc ruby3.0
```

```
  rubygems-integration t1utils teckit tex-common tex-gyre texlive-base
  texlive-binaries texlive-fonts-recommended texlive-latex-base
  texlive-latex-recommended texlive-pictures texlive-plain-generic tipa
  xfonts-encodings xfonts-utils
Suggested packages:
  fonts-noto fonts-freefont-otf | fonts-freefont-ttf libavalon-framework-java
  libcommons-logging-java-doc libexcalibur-logkit-java liblog4j1.2-java
  poppler-utils ghostscript fonts-japanese-mincho | fonts-ipafont-mincho
  fonts-japanese-gothic | fonts-ipafont-gothic fonts-arphic-ukai
  fonts-arphic-uming fonts-nanum ri ruby-dev bundler debhelper gv
  | postscript-viewer perl-tk xpdf | pdf-viewer xzdec
  texlive-fonts-recommended-doc texlive-latex-base-doc python3-pygments
  icc-profiles libfile-which-perl libspreadsheet-parseexcel-perl
  texlive-latex-extra-doc texlive-latex-recommended-doc texlive-luatex
  texlive-pstricks dot2tex prerex texlive-pictures-doc vprerex
  default-jre-headless tipa-doc
The following NEW packages will be installed:
  dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono
  fonts-texgyre fonts-urw-base35 libapache-pom-java libcommons-logging-java
  libcommons-parent-java libfontbox-java libfontenc1 libgs9 libgs9-common
  libidn12 libijs-0.35 libjbig2dec0 libkpathsea6 libpdfbox-java libptexenc1
  libruby3.0 libsynctex2 libteckit0 libtexlua53 libtexluajit2 libwoff1
  libzzip-0-13 lmodern poppler-data preview-latex-style rake ruby
  ruby-net-telnet ruby-rubygems ruby-webrick ruby-xmlrpc ruby3.0
  rubygems-integration t1utils teckit tex-common tex-gyre texlive texlive-base
  texlive-binaries texlive-fonts-recommended texlive-latex-base
  texlive-latex-extra texlive-latex-recommended texlive-pictures
  texlive-plain-generic texlive-xetex tipa xfonts-encodings xfonts-utils
0 upgraded, 55 newly installed, 0 to remove and 18 not upgraded.
Need to get 182 MB of archives.
After this operation, 572 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-droid-fallback all
1:6.0.1r16-1.1build1 [1,805 kB]
Get:2 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-lato all 2.0-2.1
[2,696 kB]
Get:3 http://archive.ubuntu.com/ubuntu jammy/main amd64 poppler-data all
0.4.11-1 [2,171 kB]
Get:4 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-common all 6.17
[33.7 kB]
Get:5 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-urw-base35 all
20200910-1 [6,367 kB]
Get:6 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libgs9-common
all 9.55.0~dfsg1-0ubuntu5.4 [752 kB]
Get:7 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libidn12 amd64
1.38-4ubuntu1 [60.0 kB]
Get:8 http://archive.ubuntu.com/ubuntu jammy/main amd64 libijs-0.35 amd64
0.35-15build2 [16.5 kB]
Get:9 http://archive.ubuntu.com/ubuntu jammy/main amd64 libjbig2dec0 amd64
```

```
0.19-3build2 [64.7 kB]
Get:10 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libgs9 amd64
9.55.0~dfsg1-0ubuntu5.4 [5,032 kB]
Get:11 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libkpathsea6
amd64 2021.20210626.59705-1ubuntu0.1 [60.3 kB]
Get:12 http://archive.ubuntu.com/ubuntu jammy/main amd64 libwoff1 amd64
1.0.2-1build4 [45.2 kB]
Get:13 http://archive.ubuntu.com/ubuntu jammy/universe amd64 dvisvgm amd64
2.13.1-1 [1,221 kB]
Get:14 http://archive.ubuntu.com/ubuntu jammy/universe amd64 fonts-lmodern all
2.004.5-6.1 [4,532 kB]
Get:15 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-noto-mono all
20201225-1build1 [397 kB]
Get:16 http://archive.ubuntu.com/ubuntu jammy/universe amd64 fonts-texgyre all
20180621-3.1 [10.2 MB]
Get:17 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libapache-pom-java
all 18-1 [4,720 B]
Get:18 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libcommons-parent-
java all 43-1 [10.8 kB]
Get:19 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libcommons-logging-
java all 1.2-2 [60.3 kB]
Get:20 http://archive.ubuntu.com/ubuntu jammy/main amd64 libfontenc1 amd64
1:1.1.4-1build3 [14.7 kB]
Get:21 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libptexenc1
amd64 2021.20210626.59705-1ubuntu0.1 [39.1 kB]
Get:22 http://archive.ubuntu.com/ubuntu jammy/main amd64 rubygems-integration
all 1.18 [5,336 B]
Get:23 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 ruby3.0 amd64
3.0.2-7ubuntu2.4 [50.1 kB]
Get:24 http://archive.ubuntu.com/ubuntu jammy/main amd64 ruby-rubygems all
3.3.5-2 [228 kB]
Get:25 http://archive.ubuntu.com/ubuntu jammy/main amd64 ruby amd64 1:3.0~exp1
[5,100 B]
Get:26 http://archive.ubuntu.com/ubuntu jammy/main amd64 rake all 13.0.6-2 [61.7
kB]
Get:27 http://archive.ubuntu.com/ubuntu jammy/main amd64 ruby-net-telnet all
0.1.1-2 [12.6 kB]
Get:28 http://archive.ubuntu.com/ubuntu jammy/universe amd64 ruby-webrick all
1.7.0-3 [51.8 kB]
Get:29 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 ruby-xmlrpc all
0.3.2-1ubuntu0.1 [24.9 kB]
Get:30 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libruby3.0
amd64 3.0.2-7ubuntu2.4 [5,113 kB]
Get:31 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libsynctex2
amd64 2021.20210626.59705-1ubuntu0.1 [55.5 kB]
Get:32 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libteckit0 amd64
2.5.11+ds1-1 [421 kB]
Get:33 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libtexlua53
```

```
amd64 2021.20210626.59705-1ubuntu0.1 [120 kB]
Get:34 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libtexluajit2
amd64 2021.20210626.59705-1ubuntu0.1 [267 kB]
Get:35 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libzzip-0-13 amd64
0.13.72+dfsg.1-1.1 [27.0 kB]
Get:36 http://archive.ubuntu.com/ubuntu jammy/main amd64 xfonts-encodings all
1:1.0.5-0ubuntu2 [578 kB]
Get:37 http://archive.ubuntu.com/ubuntu jammy/main amd64 xfonts-utils amd64
1:7.7+6build2 [94.6 kB]
Get:38 http://archive.ubuntu.com/ubuntu jammy/universe amd64 lmodern all
2.004.5-6.1 [9,471 kB]
Get:39 http://archive.ubuntu.com/ubuntu jammy/universe amd64 preview-latex-style
all 12.2-1ubuntu1 [185 kB]
Get:40 http://archive.ubuntu.com/ubuntu jammy/main amd64 t1utils amd64
1.41-4build2 [61.3 kB]
Get:41 http://archive.ubuntu.com/ubuntu jammy/universe amd64 teckit amd64
2.5.11+ds1-1 [699 kB]
Get:42 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-gyre all
20180621-3.1 [6,209 kB]
Get:43 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 texlive-
binaries amd64 2021.20210626.59705-1ubuntu0.1 [9,848 kB]
Get:44 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-base all
2021.20220204-1 [21.0 MB]
Get:45 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-fonts-
recommended all 2021.20220204-1 [4,972 kB]
Get:46 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-base
all 2021.20220204-1 [1,128 kB]
Get:47 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-
recommended all 2021.20220204-1 [14.4 MB]
Get:48 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive all
2021.20220204-1 [14.3 kB]
Get:49 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libfontbox-java all
1:1.8.16-2 [207 kB]
Get:50 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libpdfbox-java all
1:1.8.16-2 [5,199 kB]
Get:51 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-pictures
all 2021.20220204-1 [8,720 kB]
Get:52 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-extra
all 2021.20220204-1 [13.9 MB]
Get:53 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-plain-
generic all 2021.20220204-1 [27.5 MB]
Get:54 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tipa all 2:1.3-21
[2,967 kB]
Get:55 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-xetex all
2021.20220204-1 [12.4 MB]
Fetched 182 MB in 14s (13.1 MB/s)
Extracting templates from packages: 100%
Preconfiguring packages …
```

```
Selecting previously unselected package fonts-droid-fallback.
(Reading database … 120895 files and directories currently installed.)
Preparing to unpack …/00-fonts-droid-fallback_1%3a6.0.1r16-1.1build1_all.deb
…
Unpacking fonts-droid-fallback (1:6.0.1r16-1.1build1) …
Selecting previously unselected package fonts-lato.
Preparing to unpack …/01-fonts-lato_2.0-2.1_all.deb …
Unpacking fonts-lato (2.0-2.1) …
Selecting previously unselected package poppler-data.
Preparing to unpack …/02-poppler-data_0.4.11-1_all.deb …
Unpacking poppler-data (0.4.11-1) …
Selecting previously unselected package tex-common.
Preparing to unpack …/03-tex-common_6.17_all.deb …
Unpacking tex-common (6.17) …
Selecting previously unselected package fonts-urw-base35.
Preparing to unpack …/04-fonts-urw-base35_20200910-1_all.deb …
Unpacking fonts-urw-base35 (20200910-1) …
Selecting previously unselected package libgs9-common.
Preparing to unpack …/05-libgs9-common_9.55.0~dfsg1-0ubuntu5.4_all.deb …
Unpacking libgs9-common (9.55.0~dfsg1-0ubuntu5.4) …
Selecting previously unselected package libidn12:amd64.
Preparing to unpack …/06-libidn12_1.38-4ubuntu1_amd64.deb …
Unpacking libidn12:amd64 (1.38-4ubuntu1) …
Selecting previously unselected package libijs-0.35:amd64.
Preparing to unpack …/07-libijs-0.35_0.35-15build2_amd64.deb …
Unpacking libijs-0.35:amd64 (0.35-15build2) …
Selecting previously unselected package libjbig2dec0:amd64.
Preparing to unpack …/08-libjbig2dec0_0.19-3build2_amd64.deb …
Unpacking libjbig2dec0:amd64 (0.19-3build2) …
Selecting previously unselected package libgs9:amd64.
Preparing to unpack …/09-libgs9_9.55.0~dfsg1-0ubuntu5.4_amd64.deb …
Unpacking libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.4) …
Selecting previously unselected package libkpathsea6:amd64.
Preparing to unpack …/10-libkpathsea6_2021.20210626.59705-1ubuntu0.1_amd64.deb
…
Unpacking libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.1) …
Selecting previously unselected package libwoff1:amd64.
Preparing to unpack …/11-libwoff1_1.0.2-1build4_amd64.deb …
Unpacking libwoff1:amd64 (1.0.2-1build4) …
Selecting previously unselected package dvisvgm.
Preparing to unpack …/12-dvisvgm_2.13.1-1_amd64.deb …
Unpacking dvisvgm (2.13.1-1) …
Selecting previously unselected package fonts-lmodern.
Preparing to unpack …/13-fonts-lmodern_2.004.5-6.1_all.deb …
Unpacking fonts-lmodern (2.004.5-6.1) …
Selecting previously unselected package fonts-noto-mono.
Preparing to unpack …/14-fonts-noto-mono_20201225-1build1_all.deb …
Unpacking fonts-noto-mono (20201225-1build1) …
```

```
Selecting previously unselected package fonts-texgyre.
Preparing to unpack …/15-fonts-texgyre_20180621-3.1_all.deb …
Unpacking fonts-texgyre (20180621-3.1) …
Selecting previously unselected package libapache-pom-java.
Preparing to unpack …/16-libapache-pom-java_18-1_all.deb …
Unpacking libapache-pom-java (18-1) …
Selecting previously unselected package libcommons-parent-java.
Preparing to unpack …/17-libcommons-parent-java_43-1_all.deb …
Unpacking libcommons-parent-java (43-1) …
Selecting previously unselected package libcommons-logging-java.
Preparing to unpack …/18-libcommons-logging-java_1.2-2_all.deb …
Unpacking libcommons-logging-java (1.2-2) …
Selecting previously unselected package libfontenc1:amd64.
Preparing to unpack …/19-libfontenc1_1%3a1.1.4-1build3_amd64.deb …
Unpacking libfontenc1:amd64 (1:1.1.4-1build3) …
Selecting previously unselected package libptexenc1:amd64.
Preparing to unpack …/20-libptexenc1_2021.20210626.59705-1ubuntu0.1_amd64.deb
…
Unpacking libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.1) …
Selecting previously unselected package rubygems-integration.
Preparing to unpack …/21-rubygems-integration_1.18_all.deb …
Unpacking rubygems-integration (1.18) …
Selecting previously unselected package ruby3.0.
Preparing to unpack …/22-ruby3.0_3.0.2-7ubuntu2.4_amd64.deb …
Unpacking ruby3.0 (3.0.2-7ubuntu2.4) …
Selecting previously unselected package ruby-rubygems.
Preparing to unpack …/23-ruby-rubygems_3.3.5-2_all.deb …
Unpacking ruby-rubygems (3.3.5-2) …
Selecting previously unselected package ruby.
Preparing to unpack …/24-ruby_1%3a3.0~exp1_amd64.deb …
Unpacking ruby (1:3.0~exp1) …
Selecting previously unselected package rake.
Preparing to unpack …/25-rake_13.0.6-2_all.deb …
Unpacking rake (13.0.6-2) …
Selecting previously unselected package ruby-net-telnet.
Preparing to unpack …/26-ruby-net-telnet_0.1.1-2_all.deb …
Unpacking ruby-net-telnet (0.1.1-2) …
Selecting previously unselected package ruby-webrick.
Preparing to unpack …/27-ruby-webrick_1.7.0-3_all.deb …
Unpacking ruby-webrick (1.7.0-3) …
Selecting previously unselected package ruby-xmlrpc.
Preparing to unpack …/28-ruby-xmlrpc_0.3.2-1ubuntu0.1_all.deb …
Unpacking ruby-xmlrpc (0.3.2-1ubuntu0.1) …
Selecting previously unselected package libruby3.0:amd64.
Preparing to unpack …/29-libruby3.0_3.0.2-7ubuntu2.4_amd64.deb …
Unpacking libruby3.0:amd64 (3.0.2-7ubuntu2.4) …
Selecting previously unselected package libsynctex2:amd64.
Preparing to unpack …/30-libsynctex2_2021.20210626.59705-1ubuntu0.1_amd64.deb
```

```
…
Unpacking libsynctex2:amd64 (2021.20210626.59705-1ubuntu0.1) …
Selecting previously unselected package libteckit0:amd64.
Preparing to unpack …/31-libteckit0_2.5.11+ds1-1_amd64.deb …
Unpacking libteckit0:amd64 (2.5.11+ds1-1) …
Selecting previously unselected package libtexlua53:amd64.
Preparing to unpack …/32-libtexlua53_2021.20210626.59705-1ubuntu0.1_amd64.deb
…
Unpacking libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.1) …
Selecting previously unselected package libtexluajit2:amd64.
Preparing to unpack
…/33-libtexluajit2_2021.20210626.59705-1ubuntu0.1_amd64.deb …
Unpacking libtexluajit2:amd64 (2021.20210626.59705-1ubuntu0.1) …
Selecting previously unselected package libzzip-0-13:amd64.
Preparing to unpack …/34-libzzip-0-13_0.13.72+dfsg.1-1.1_amd64.deb …
Unpacking libzzip-0-13:amd64 (0.13.72+dfsg.1-1.1) …
Selecting previously unselected package xfonts-encodings.
Preparing to unpack …/35-xfonts-encodings_1%3a1.0.5-0ubuntu2_all.deb …
Unpacking xfonts-encodings (1:1.0.5-0ubuntu2) …
Selecting previously unselected package xfonts-utils.
Preparing to unpack …/36-xfonts-utils_1%3a7.7+6build2_amd64.deb …
Unpacking xfonts-utils (1:7.7+6build2) …
Selecting previously unselected package lmodern.
Preparing to unpack …/37-lmodern_2.004.5-6.1_all.deb …
Unpacking lmodern (2.004.5-6.1) …
Selecting previously unselected package preview-latex-style.
Preparing to unpack …/38-preview-latex-style_12.2-1ubuntu1_all.deb …
Unpacking preview-latex-style (12.2-1ubuntu1) …
Selecting previously unselected package t1utils.
Preparing to unpack …/39-t1utils_1.41-4build2_amd64.deb …
Unpacking t1utils (1.41-4build2) …
Selecting previously unselected package teckit.
Preparing to unpack …/40-teckit_2.5.11+ds1-1_amd64.deb …
Unpacking teckit (2.5.11+ds1-1) …
Selecting previously unselected package tex-gyre.
Preparing to unpack …/41-tex-gyre_20180621-3.1_all.deb …
Unpacking tex-gyre (20180621-3.1) …
Selecting previously unselected package texlive-binaries.
Preparing to unpack …/42-texlive-
binaries_2021.20210626.59705-1ubuntu0.1_amd64.deb …
Unpacking texlive-binaries (2021.20210626.59705-1ubuntu0.1) …
Selecting previously unselected package texlive-base.
Preparing to unpack …/43-texlive-base_2021.20220204-1_all.deb …
Unpacking texlive-base (2021.20220204-1) …
Selecting previously unselected package texlive-fonts-recommended.
Preparing to unpack …/44-texlive-fonts-recommended_2021.20220204-1_all.deb …
Unpacking texlive-fonts-recommended (2021.20220204-1) …
Selecting previously unselected package texlive-latex-base.
```

```
Preparing to unpack …/45-texlive-latex-base_2021.20220204-1_all.deb …
Unpacking texlive-latex-base (2021.20220204-1) …
Selecting previously unselected package texlive-latex-recommended.
Preparing to unpack …/46-texlive-latex-recommended_2021.20220204-1_all.deb …
Unpacking texlive-latex-recommended (2021.20220204-1) …
Selecting previously unselected package texlive.
Preparing to unpack …/47-texlive_2021.20220204-1_all.deb …
Unpacking texlive (2021.20220204-1) …
Selecting previously unselected package libfontbox-java.
Preparing to unpack …/48-libfontbox-java_1%3a1.8.16-2_all.deb …
Unpacking libfontbox-java (1:1.8.16-2) …
Selecting previously unselected package libpdfbox-java.
Preparing to unpack …/49-libpdfbox-java_1%3a1.8.16-2_all.deb …
Unpacking libpdfbox-java (1:1.8.16-2) …
Selecting previously unselected package texlive-pictures.
Preparing to unpack …/50-texlive-pictures_2021.20220204-1_all.deb …
Unpacking texlive-pictures (2021.20220204-1) …
Selecting previously unselected package texlive-latex-extra.
Preparing to unpack …/51-texlive-latex-extra_2021.20220204-1_all.deb …
Unpacking texlive-latex-extra (2021.20220204-1) …
Selecting previously unselected package texlive-plain-generic.
Preparing to unpack …/52-texlive-plain-generic_2021.20220204-1_all.deb …
Unpacking texlive-plain-generic (2021.20220204-1) …
Selecting previously unselected package tipa.
Preparing to unpack …/53-tipa_2%3a1.3-21_all.deb …
Unpacking tipa (2:1.3-21) …
Selecting previously unselected package texlive-xetex.
Preparing to unpack …/54-texlive-xetex_2021.20220204-1_all.deb …
Unpacking texlive-xetex (2021.20220204-1) …
Setting up fonts-lato (2.0-2.1) …
Setting up fonts-noto-mono (20201225-1build1) …
Setting up libwoff1:amd64 (1.0.2-1build4) …
Setting up libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.1) …
Setting up libijs-0.35:amd64 (0.35-15build2) …
Setting up libtexluajit2:amd64 (2021.20210626.59705-1ubuntu0.1) …
Setting up libfontbox-java (1:1.8.16-2) …
Setting up rubygems-integration (1.18) …
Setting up libzzip-0-13:amd64 (0.13.72+dfsg.1-1.1) …
Setting up fonts-urw-base35 (20200910-1) …
Setting up poppler-data (0.4.11-1) …
Setting up tex-common (6.17) …
update-language: texlive-base not installed and configured, doing nothing!
Setting up libfontenc1:amd64 (1:1.1.4-1build3) …
Setting up libjbig2dec0:amd64 (0.19-3build2) …
Setting up libteckit0:amd64 (2.5.11+ds1-1) …
Setting up libapache-pom-java (18-1) …
Setting up ruby-net-telnet (0.1.1-2) …
Setting up xfonts-encodings (1:1.0.5-0ubuntu2) …
```

```
Setting up t1utils (1.41-4build2) …
Setting up libidn12:amd64 (1.38-4ubuntu1) …
Setting up fonts-texgyre (20180621-3.1) …
Setting up libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.1) …
Setting up ruby-webrick (1.7.0-3) …
Setting up fonts-lmodern (2.004.5-6.1) …
Setting up fonts-droid-fallback (1:6.0.1r16-1.1build1) …
Setting up ruby-xmlrpc (0.3.2-1ubuntu0.1) …
Setting up libsynctex2:amd64 (2021.20210626.59705-1ubuntu0.1) …
Setting up libgs9-common (9.55.0~dfsg1-0ubuntu5.4) …
Setting up teckit (2.5.11+ds1-1) …
Setting up libpdfbox-java (1:1.8.16-2) …
Setting up libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.4) …
Setting up preview-latex-style (12.2-1ubuntu1) …
Setting up libcommons-parent-java (43-1) …
Setting up dvisvgm (2.13.1-1) …
Setting up libcommons-logging-java (1.2-2) …
Setting up xfonts-utils (1:7.7+6build2) …
Setting up libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.1) …
Setting up texlive-binaries (2021.20210626.59705-1ubuntu0.1) …
update-alternatives: using /usr/bin/xdvi-xaw to provide /usr/bin/xdvi.bin
(xdvi.bin) in auto mode
update-alternatives: using /usr/bin/bibtex.original to provide /usr/bin/bibtex
(bibtex) in auto mode
Setting up lmodern (2.004.5-6.1) …
Setting up texlive-base (2021.20220204-1) …
/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
mktexlsr: Updating /var/lib/texmf/ls-R-TEXLIVEDIST…
mktexlsr: Updating /var/lib/texmf/ls-R-TEXMFMAIN…
mktexlsr: Updating /var/lib/texmf/ls-R…
mktexlsr: Done.
tl-paper: setting paper size for dvips to a4:
/var/lib/texmf/dvips/config/config-paper.ps
tl-paper: setting paper size for dvipdfmx to a4:
/var/lib/texmf/dvipdfmx/dvipdfmx-paper.cfg
tl-paper: setting paper size for xdvi to a4: /var/lib/texmf/xdvi/XDvi-paper
tl-paper: setting paper size for pdftex to a4: /var/lib/texmf/tex/generic/tex-
ini-files/pdftexconfig.tex
Setting up tex-gyre (20180621-3.1) …
Setting up texlive-plain-generic (2021.20220204-1) …
Setting up texlive-latex-base (2021.20220204-1) …
Setting up texlive-latex-recommended (2021.20220204-1) …
Setting up texlive-pictures (2021.20220204-1) …
Setting up texlive-fonts-recommended (2021.20220204-1) …
Setting up tipa (2:1.3-21) …
```

```
Setting up texlive (2021.20220204-1) …
Setting up texlive-latex-extra (2021.20220204-1) …
Setting up texlive-xetex (2021.20220204-1) …
Setting up rake (13.0.6-2) …
Setting up libruby3.0:amd64 (3.0.2-7ubuntu2.4) …
Setting up ruby3.0 (3.0.2-7ubuntu2.4) …
Setting up ruby (1:3.0~exp1) …
Setting up ruby-rubygems (3.3.5-2) …
Processing triggers for man-db (2.10.2-1) …
Processing triggers for fontconfig (2.13.1-4.2ubuntu5) …
Processing triggers for libc-bin (2.35-0ubuntu3.1) …
/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbb.so.12 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc.so.2 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc_proxy.so.2 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_5.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind.so.3 is not a symbolic link
```

[ ]:
```
!jupyter nbconvert --to PDF "/content/drive/MyDrive/Colab Notebooks/SI/
↪Kmeans_Kmedoids.ipynb"
```