

**TRƯỜNG ĐẠI HỌC MỞ THÀNH PHỐ HỒ CHÍ MINH**  
**KHOA CÔNG NGHỆ THÔNG TIN**

---- . ----



## **BÀI TẬP LỚN MÔN HỌC**

### **KHAI PHÁ DỮ LIỆU**

**ĐỀ TÀI:**

**Phân tích khai phá dữ liệu tập dữ liệu doanh nghiệp**

**Sinh viên thực hiện:**

**Đặng Minh Phúc**

**Lê Trung Hậu**

**Hồ Sỹ Quang Trung**

**Lớp:**

**ITEC3417**

**Giảng viên hướng dẫn:**

**Nguyễn Trung Hậu**

**Tháng 11 năm 2023**

---

---

## Mục lục

<b>Lời nói đầu.....</b>	<b>2</b>
<b>1. Bảng Phân Công.....</b>	<b>3</b>
<b>2.Giới thiệu về nguồn dữ liệu, phương pháp thu thập.....</b>	<b>3</b>
<b>3. Tiền xử lý dữ liệu:.....</b>	<b>3</b>
3.1 Phân tích tập dữ liệu.....	3
3.2. Xử lý dữ liệu.....	4
3.3 Biến đổi dữ liệu.....	4
<b>4.Mô tả tập dữ liệu sau khi xử lý.....</b>	<b>4</b>
<b>5. Phương pháp khai phá dữ liệu.....</b>	<b>5</b>
5.1 Trực quan hóa, biểu diễn dữ liệu.....	5
5.2 Mô hình hóa.....	5
<b>6. Kết quả thực nghiệm của mô hình và đối sánh với các lập luận đã đưa ra.....</b>	<b>5</b>
6.1 Trực quan hóa, biểu diễn dữ liệu.....	5
6.2 Mô hình hóa.....	7
<b>7. Kết luận.....</b>	<b>9</b>
<b>8.Colab.....</b>	
..... <a href="#">link</a>	

## Lời nói đầu

Khai phá dữ liệu là một phần quan trọng đối với sự thành công của bất kỳ sáng kiến phân tích nào. Các doanh nghiệp có thể sử dụng quy trình khai phá kiến thức để tăng niềm tin của khách hàng, tìm kiếm nguồn doanh thu mới và thu hút khách hàng quay lại. Nhận thấy tầm quan trọng của việc khai phá nên nhóm 7 đã chọn chủ đề “Khai phá dữ liệu trong doanh nghiệp”.

Việc khai thác thông tin từ dữ liệu này mang lại cơ hội phong phú để nghiên cứu và phát triển kiến thức về nền kinh tế và doanh nghiệp. Một trong những khía cạnh quan trọng của dữ liệu doanh nghiệp là khả năng phân tích xu hướng ngành. Bằng cách xem xét sự phát triển của các ngành công nghiệp trong thời gian, chúng ta có thể nhận biết các xu hướng quan trọng và dự đoán những thách thức và cơ hội trong tương lai. Thông qua việc phân tích lịch sử phát triển của doanh nghiệp trong ngành, ta có thể xác định những chiến lược thành công và những bài học quý báu từ những trường hợp không thành công. Ngoài ra, việc tìm hiểu về các yếu tố đằng sau sự phát triển của doanh nghiệp có thể là một hướng nghiên cứu quan trọng. Điều này có thể bao gồm các yếu tố như chiến lược kinh doanh, quản lý tài nguyên, đổi mới công nghệ, và tương tác với môi trường kinh doanh. Bằng cách hiểu rõ các yếu tố này, chúng ta có thể rút ra những kinh nghiệm quý báu cho các doanh nghiệp khác và đề xuất các chiến lược cụ thể để cải thiện hiệu suất kinh doanh.

Cuối cùng, việc nghiên cứu về dữ liệu doanh nghiệp không chỉ giúp chúng ta hiểu rõ hơn về doanh nghiệp và ngành công nghiệp, mà còn có thể hỗ trợ trong việc đưa ra quyết định chiến lược và dự báo kịch bản tương lai. Việc này là quan trọng trong bối cảnh môi trường kinh doanh đang thay đổi nhanh chóng và đầy thách thức.

## 1. Bảng Phân Công

MSSV	Tên	Nhiệm vụ
2151010290	Đặng Minh Phúc	- Tiền xử lý dữ liệu thống kê, báo cáo
2151013020	Lê Trung Hậu	- Chọn mô hình khai phá phù hợp - Tổng hợp dữ liệu - Phương pháp đánh giá mô hình
1951052217	Hồ Sỹ Quang Trung	- Tìm kiếm, gộp, làm sạch data và mô tả tập dữ liệu sau khi xử lý.

## 2. Giới thiệu về nguồn dữ liệu, phương

---

## pháp thu thập

Bắt đầu hành trình nghiên cứu, nhóm của chúng tôi đã thu thập dữ liệu từ các nguồn tập dữ liệu mở như Kaggle, các cơ sở dữ liệu chính phủ ( GOV ) và Google dataset... hướng tới mục tiêu chính là thu thập thông tin đa chiều liên quan đến mọi khía cạnh của doanh nghiệp. Bằng cách này, chúng tôi mong muốn xây dựng nên một tập dữ liệu (dataset) để tiến hành khai phá dữ liệu một cách dễ dàng. Thông qua nhiều nguồn dữ liệu khác nhau, chúng tôi đã cố gắng đảm bảo tính chính xác và đồng nhất của dữ liệu, tạo ra một cơ sở dữ liệu chuyên sâu về thế giới doanh nghiệp.

Qua quá trình này, nhóm đã nắm bắt được sự đa dạng và đầy đủ về mặt thông tin, giúp chúng tôi có cái nhìn tổng thể về thị trường kinh doanh và những yếu tố đang ảnh hưởng.

Trong quá trình này, nhóm đã tiếp cận và tổng hợp một tập dữ liệu đa dạng và đầy đủ về các khía cạnh của doanh nghiệp. Các tập dữ liệu thu thập bao gồm:

**7+ Million Company Dataset:** Bộ dữ liệu này chứa thông tin về hơn 7 triệu doanh nghiệp, cung cấp cái nhìn toàn diện về thị trường doanh nghiệp và sự đa dạng trong các ngành công nghiệp.

**Continent List for 2021 Olympics in Tokyo Dataset:** Tập dữ liệu này cung cấp thông tin về các lực địa tham gia trong Thế vận hội Tokyo năm 2021. Điều này có thể hữu ích để đánh giá tầm ảnh hưởng quốc tế của các doanh nghiệp và nhóm nghiên cứu thị trường.

Bằng cách kết hợp cả hai nguồn dữ liệu trên, nhóm hy vọng sẽ có được cái nhìn đa chiều và chi tiết về hình ảnh toàn cảnh của doanh nghiệp.

Tập dữ liệu này sẽ là nguồn cơ sở cho quá trình khai phá dữ liệu và phân tích chi tiết về các khía cạnh kinh doanh quan trọng. Nhóm sẽ tiếp tục quá trình nghiên cứu và phát triển để đảm bảo rằng thông tin thu thập là chính xác và có ý nghĩa cho mục tiêu nghiên cứu của chúng tôi.

## 3. Tiền xử lý dữ liệu:

### 3.1 Phân tích tập dữ liệu

- Sau khi hoàn thành quá trình thu thập dữ liệu, bước quan trọng tiếp theo là xác định và giải quyết các vấn đề liên quan đến tính chất của dữ liệu. Việc làm này đặc biệt quan trọng vì nếu dữ liệu không được làm sạch, có thể ảnh hưởng đến độ tin cậy và chất lượng của dữ liệu, dẫn đến những quyết định không chính xác và khó khăn trong quá trình khai phá thông tin. Dưới đây là những vấn đề cụ thể mà nhóm của chúng tôi đã phải đối mặt khi chưa làm sạch dữ liệu trong tập dữ liệu "7+ Million Company" và "Continent List for 2021 Olympics in Tokyo Dataset," cùng với những giải pháp đề xuất.

- **Tập Dữ Liệu "7+ Million Company"**

Vấn Đề 1: Dữ liệu không hoàn chỉnh (incomplete):

Số liệu Mất Mát:

year\_founded: 50.11%

industry: 3.96%

country: 34.84%

locality: 32.62%

Giải Pháp: Tiến hành lọc từng cụm dữ liệu: Xóa các đối tượng thiếu giá trị, lọc các giá trị nhiễu. Cân nhắc sử dụng các kỹ thuật như điền giá trị trung bình, trung vị hoặc sử dụng mô

---

hình dự đoán để điền các giá trị thiếu.

- **Vấn Đề 2: Nhiễu/Lỗi (noise/error):**

Các giá trị year\_founded lớn hơn năm hiện tại: 7 trường hợp (2029, 2027, 2103, 2025, 2025, 2025, 2025).

Giải Pháp: Kiểm tra và xác minh lại nguồn dữ liệu để xác định xem liệu những giá trị này là do sai sót nhập liệu hay có ý nghĩa cụ thể. Nếu là sai sót, thực hiện việc sửa lỗi hoặc loại bỏ những đối tượng có giá trị không hợp lý.

- **Tập Dữ Liệu "Continent List for 2021 Olympics in Tokyo Dataset"**

Vấn Đề: Dữ liệu không trùng khớp (mismatched)

Một số giá trị với cùng tính chất nhưng khác tên, gây thiếu sót trong quá trình gộp dữ liệu bên cạnh một số thuộc tính thiếu so với tập dữ liệu cần gộp. Giải Pháp: Chuẩn hóa tên các thuộc tính để đảm bảo tính nhất quán. Bổ sung các giá trị còn thiếu bằng cách tham khảo các nguồn dữ liệu đáng tin cậy khác hoặc sử dụng các phương pháp điền giá trị như đã mô tả ở trên.

Bằng cách giải quyết những vấn đề này, chúng tôi hy vọng tập dữ liệu của mình sẽ trở nên đáng tin cậy hơn, từ đó tạo nền tảng cho các quá trình phân tích và khai phá thông tin chất lượng và đáng tin cậy

### 3.2. Xử lý dữ liệu

- **Đối với tập dữ liệu '7+ Million Company':**

+ Tiến hành xóa trống các đối tượng chứa thuộc tính nhiễu, thiếu nhưng không làm thay đổi tính chất, sự cân bằng của tập dữ liệu gốc.

- **Đối với tập dữ liệu "Continent List for 2021 Olympics in Tokyo Dataset":**

+ Tiến hành đối chiếu các thuộc tính không trùng giữa tập dữ liệu này và '7+ Million Company', tiến hành tinh chỉnh tên giá trị, thêm các giá trị bị thiếu, loại bỏ 1 số giá trị không mang nhiều ý nghĩa để giảm nhẹ chi phí tính toán mô hình mà không làm ảnh hưởng dữ liệu chung.

### 3.3 Biến đổi dữ liệu

Biến đổi dữ liệu là việc chuyển toàn bộ tập giá trị của một thuộc tính sang một tập giá trị thay thế sao cho giá trị cũ tương ứng giá trị mới.

Trong bài tập lớn này nhóm chúng tôi sử dụng các phương pháp biến đổi sau:

-Làm trơn (smoothing): Loại bỏ nhiễu/lỗi khỏi tập dữ liệu.

-Chuẩn hóa (normalization): Đưa các giá trị về một khoảng được chỉ định.

## 4.Mô tả tập dữ liệu sau khi xử lý

## Overview

Dataset Statistics		Dataset Insights	
Number of Variables	8	<code>year_fo_</code> is skewed	Skewed
Number of Rows	7.0046×10 <sup>06</sup>	<code>current_</code> is skewed	Skewed
Missing Cells	0	<code>total_e_</code> is skewed	Skewed
Missing Cells (%)	0.0%	<code>name</code> has a high cardinality: 7004634 distinct values	High Cardinality
Duplicate Rows	0	<code>industry</code> has a high cardinality: 148 distinct values	High Cardinality
Duplicate Rows (%)	0.0%	<code>locality</code> has a high cardinality: 95756 distinct values	High Cardinality
Total Size in Memory	2.7 GB	<code>country</code> has a high cardinality: 236 distinct values	High Cardinality
Average Row Size in Memory	411.3 B	<code>name</code> has all distinct values	Unique
Variable Types	Categorical: 4 Numerical: 3 GeoGraphy: 1	<code>current_</code> has 1414862 (20.2%) zeros	Zeros

- Tổng quan tập dữ liệu sau tiền xử lý:
  - o 7046000 đối tượng
  - o 8 thuộc tính
  - o Số giá trị bị mất 0 (0%)
  - o Các cột giá trị bị lệch: ‘year founded’, ‘current\_employee\_estimate’, ‘total employee estimate’
  - o Các cột có giá trị 0:
    - current\_employee\_estimate’: 1414862 (20.2%)

## 5. Phương pháp khai phá dữ liệu

### 5.1 Trực quan hóa, biểu diễn dữ liệu

Chúng tôi sử dụng các biểu đồ cũng như các bảng để thể hiện các xu hướng dữ liệu, sự phân bố để tìm ra các xu hướng cũng như giải thích, chứng minh cho các xu hướng đó

### 5.2 Mô hình hóa

Chúng tôi sử dụng mô hình gom cụm bằng thuật toán K-means cho tập dữ liệu này vì đây là mô hình khá phổ biến cũng như dễ áp dụng để tìm ra các đặc trưng trong các cụm dữ liệu.

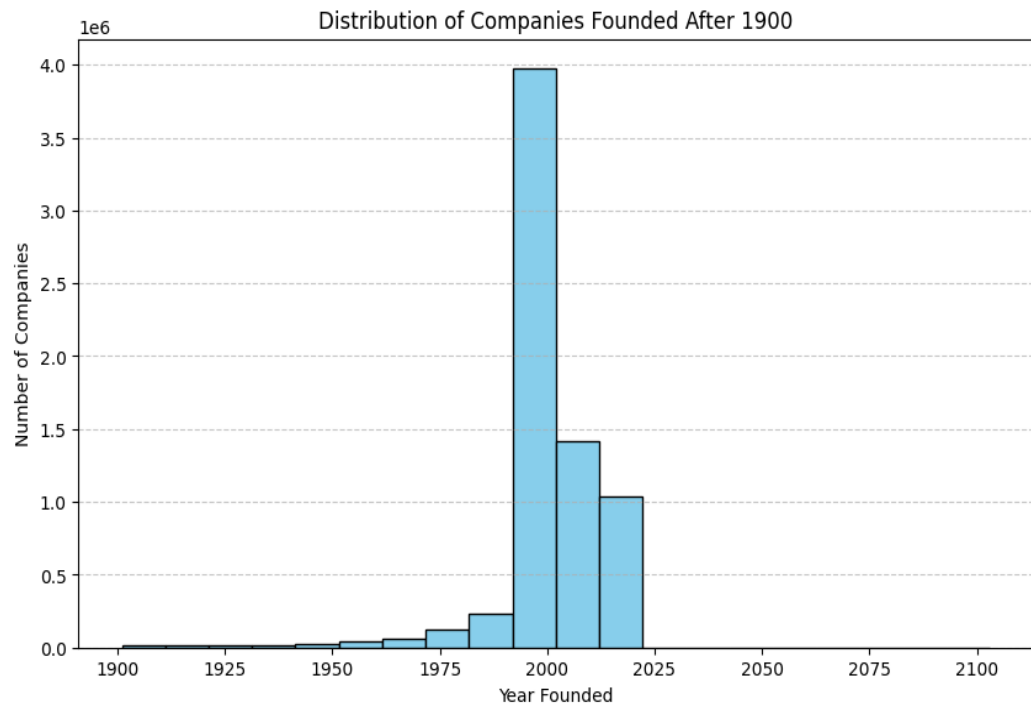
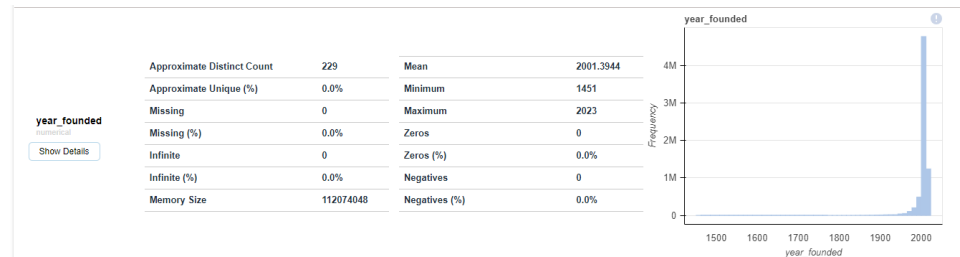
Vấn đề đặt ra cho việc gom cụm dữ liệu là chọn các k clusters phù hợp. Có nhiều cách tiếp cận để tìm k cluster như: Elbow method, AffinityMatrix, SilhouetteScore,... và dựa vào kinh nghiệm cá nhân. Trong tập dữ liệu này chúng tôi tiếp cận với Elbow method là một trong những cách phổ biến nhất để tối ưu k clusters. Chúng tôi lặp lại các giá trị của k từ 1 đến 11 và tính toán độ biến dạng hoặc quán tính cho từng giá trị của k trong phạm vi đã cho. Trong đó Quán tính là tổng bình phương khoảng cách của các mẫu đến tâm cụm gần nhất của chúng. Để xác định số cụm tối ưu, chúng ta phải chọn giá trị k tại “khuyết tay”, tức là điểm mà sau đó độ méo/quán tính bắt đầu giảm theo kiểu tuyến tính.

## 6. Kết quả thực nghiệm của mô hình và đối sánh với các lập luận đã đưa ra.

### 6.1 Trực quan hóa, biểu diễn dữ liệu

Tổng quan:

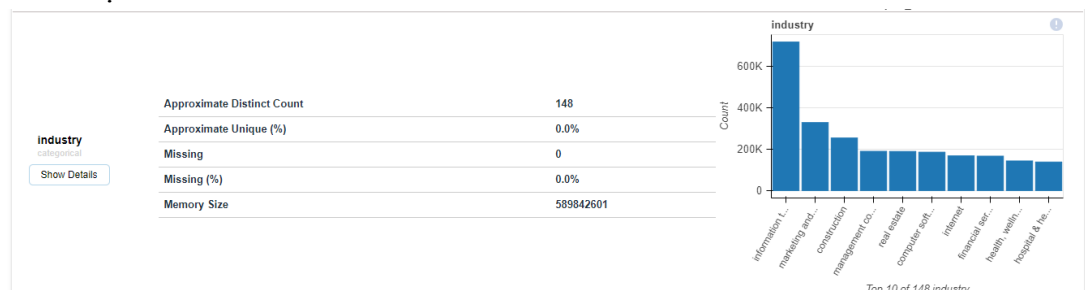
- Thuộc tính năm thành lập:

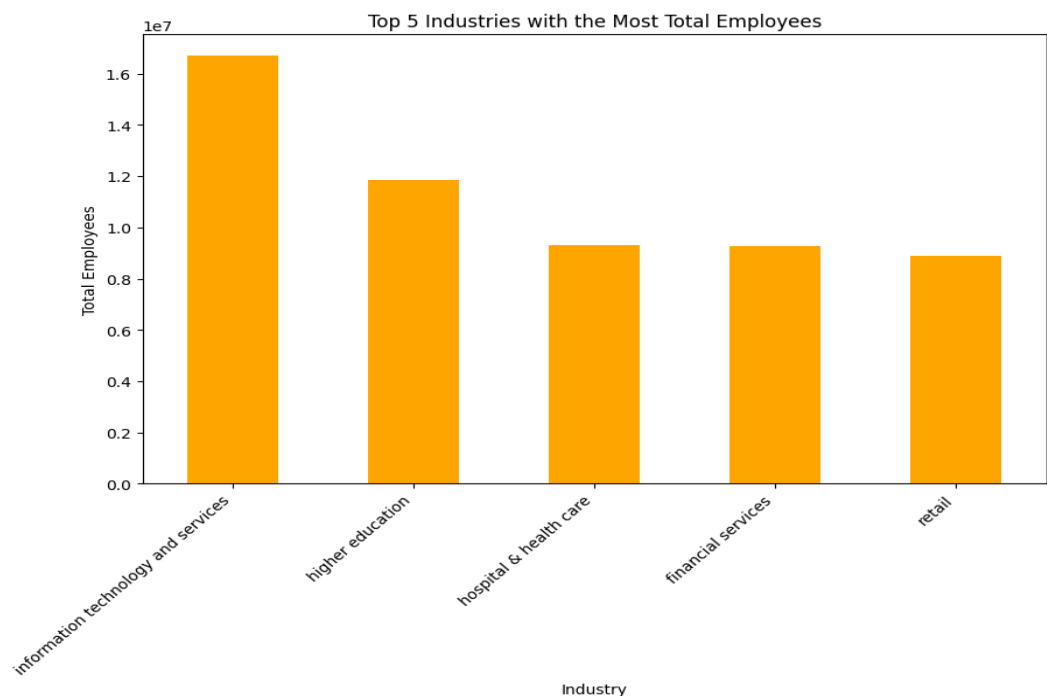
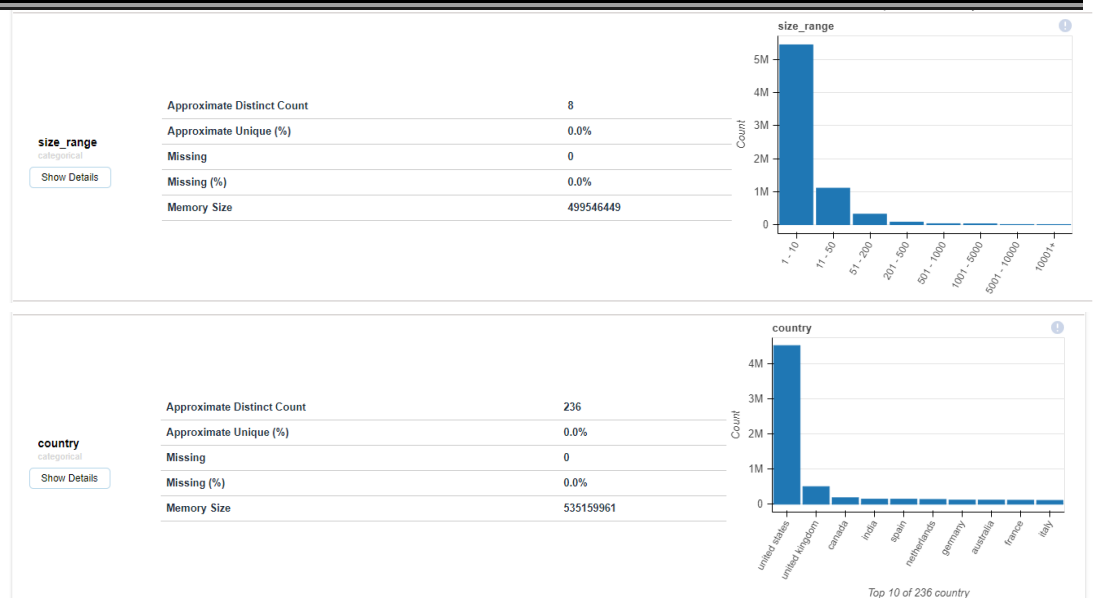


Biểu đồ phân bố năm thành lập sau 1900

⇒ Phần lớn các công ty thành lập vào những năm 2000 với giá trị Mean rơi vào giá trị 2001. Trong đó công ty thành lập cũ nhất vào năm 1451 và mới nhất vào năm 2023

- Các thuộc tính khác:





Ngành công nghệ thông tin chiếm đa số, phần lớn công ty thành lập có quy mô nhỏ và Mỹ là quốc gia với nhiều công ty nhất.

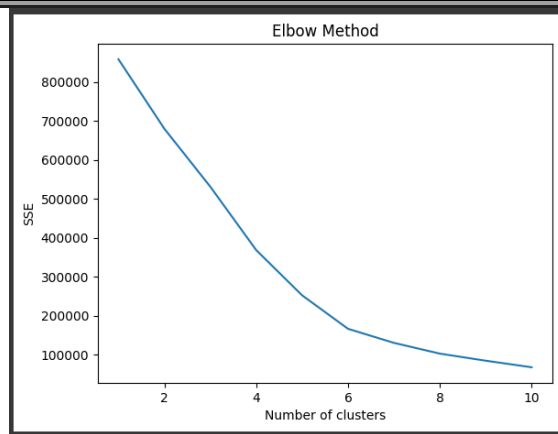
- Phần lớn các công ty CNTT ra đời vào những năm 2000 cùng với sự phát triển nhanh của mạng internet.

⇒ Xu hướng phát triển các công ty đang theo hướng ngành CNTT trong thời đại phát triển nhanh của các công nghệ hiện đại

## 6.2 Mô hình hóa

- Sau khi thực thi thuật toán chúng tôi tìm ra số cụm tối ưu cho tập dữ liệu này là 6





- Sau đó chúng tôi tiến hành gom cụm bằng thuật toán K-means  
Kết quả sau khi gom cụm như sau:

```
Cluster 1  
['501 - 1000']  
  
Cluster 2  
['201 - 500']  
  
Cluster 3  
['1001 - 5000']  
  
Cluster 4  
['10001+']  
  
Cluster 5  
['5001 - 10000']  
  
Cluster 6  
['10001+']
```

```
Cluster 1  
['north america' 'oceania' 'europe' 'south america' 'asia' 'africa' nan]  
  
Cluster 2  
['north america' 'oceania' 'europe' 'asia' 'south america' 'africa' nan]  
  
Cluster 3  
['north america' 'europe' 'oceania' 'asia' 'south america' 'africa' nan]  
  
Cluster 4  
['north america' 'europe' 'asia' 'south america' 'oceania' 'africa']  
  
Cluster 5  
['north america' 'oceania' 'europe' 'asia' 'africa' 'south america']  
  
Cluster 6  
['north america' 'asia' 'europe']
```

```

Cluster 1
[500 499 498 497 496 495 494 493 492 491 490 489 488 487 486 485 484 483
482 481 480 479 478 477 476 475 474 473 472 471 470 469 468 467 466 465
464 463 462 461 460 459 458 457 456 455 454 453 452 451 450 449 448 447
446 445 444 443 442 441 440 439 438 437 436 435 434 433 432 431 430 429
428 427 426 425 424 423 422 421 420 419 418 417 416 415 414 413 412 411
410 409 408 407 406 405 404 403 402 401 400 399 398 397 396 395 394 393
392 391 390 389 388 387 386 385 384 383 382 381 380 379 378 377 376 375
374 373 372 371 370 369 368 367 366 365 364 363 362 361 360 359 358 357
356 355 354 353 352 351 350 349 348 347 346 345 344 343 342 341 340 339
338 337 336 335 334 333 332 331 330 329 328 327 326 325 324 323 322 321
320 319 318 317 316 315 314 313 312 311 310 309 308 307 306 305 304 303
302 301 300 299 298 297 296 295 294 293 292 291 290 289 288 287 286 285
284 283 282 281 280 279 278 277 276 275 274 273 272 271 270 269 268 267
266 265 264 263 262 261 260 259 258 257 256 255 254 253 252 251 250 249
248 247 246 245 244 243 242 241 240 239 238 237 236 235 234 233 232 231
230 229 228 227 226 225 224 223 222 221 220 219 218 217 216 215 214 213
212 211 210 209 208 207 206 205 204 203 202 201 200 199 198 197 196 195
194 193 192 191 190 189 188 187 186 185 184 183 182 181 180 179 178 177
176 175 174 173 172 171 170 169 168 167]

Cluster 2
[250 249 248 247 246 245 244 243 242 241 240 239 238 237 236 235 234 233
232 231 230 229 228 227 226 225 224 223 222 221 220 219 218 217 216 215
214 213 212 211 210 209 208 207 206 205 204 203 202 201 200 199 198 197
196 195 194 193 192 191 190 189 188 187 186 185 184 183 182 181 180 179
178 177 176 175 174 173 172 171 170 169 168 167 166 165 164 163 162 161
160 159 158 157 156 155 154 153 152 151 150 149 148 147 146 145 144 143
142 141 140 139 138 137 136 135 134 133 132 131 130 129 128 127 126 125
124 123 122 121 120 119 118 117 116 115 114 113 112 111 110 109 108 107
106 105 104 103 102 101]

Cluster 3
[2500 2499 2498 ... 336 335 334]

Cluster 4
[51441 48806 47434 ... 3344 3339 3337]

Cluster 5
[4995 4994 4993 ... 1670 1668 1667]

Cluster 6
[274047 190771 190689 162163 158363 127952 122031 120753 116196 115188
113997 111372 109532 104752 104112 101482 97357 95234 94458 93247
90095 87381 85090 84327 84218 84179 78261 75640 75109 74357
68233 67692 67564 67261 66632 65839 65335 64046 62685 61803
61638 61040 60602 60324 59993 59712 59588 58819 58538 57720
57255 55945 54117 51721 50715 50303 46523 44799 42043 38917]

```

⇒ Các cụm phân chia theo quy mô công ty. Trong đó đặc biệt với cụm 6 gồm các công ty với quy mô lớn và số lượng nhân viên đông đảo nhất tập trung ở các quốc gia thuộc châu Âu, châu Á và châu Mỹ. Trong khi đó các cụm còn lại với quy mô thấp hơn phân bổ đều cho các châu lục

## 7. Kết luận

Sau khi thực hiện khai phá dữ liệu, nhóm chúng tôi đã đạt được những kết quả quan trọng, đồng thời nhận thức được những hạn chế cũng như đề xuất phát triển cho quá trình tiếp theo. Dưới đây là những điểm chính:

- **Kết Quả:**
  - **Khai Phá Góc Nhìn:** Nhóm đã áp dụng các kỹ thuật khai phá dữ liệu để tìm hiểu các góc nhìn khác nhau từ tập dữ liệu. Điều này đã là bước quan trọng để xác định các xu hướng, mối quan hệ, và tri thức tiềm ẩn trong dữ liệu.
- **Hạn Chế:**
  - **Kinh Nghiệm Đọc Dữ Liệu:** Nhóm gặp hạn chế trong việc đọc hiểu kết quả sau phân tích. Kinh nghiệm này thường được thể hiện một cách rõ ràng thông qua việc trực quan hóa kết quả, và nhóm hiện chưa có đủ kinh nghiệm để hiệu quả trong việc khai thác sâu dữ liệu và tri thức.

- **Hạn Chế Kiến Thức:** Do hạn chế về kiến thức, nhóm chỉ thực hiện một mô hình. Điều này gây ra hạn chế trong việc đảm bảo tối ưu và độ chính xác của kết quả khi áp dụng nhiều mô hình.
- **Xử Lý Dữ Liệu:** Xử lý dữ liệu chưa đạt tối ưu, dẫn đến việc các kết quả sau khi gom cụm vẫn còn nhiều giá trị nhiễu, ảnh hưởng đến kết quả chung.
- **Đề Xuất Phát Triển:**
  - **Tham Khảo Các Phương Pháp Khác:** Nhóm có thể dành thêm thời gian nghiên cứu và tham khảo các phương pháp khai phá dữ liệu khác nhau, mở rộng kiến thức và kỹ năng trong lĩnh vực này. **Cải Tiến Xử Lý Dữ Liệu:** Tập trung vào việc tối ưu hóa xử lý dữ liệu để loại bỏ nhiễu một cách hiệu quả, đảm bảo rằng mô hình chỉ sử dụng dữ liệu chất lượng, đáp ứng yêu cầu của mô hình.
  - **Áp Dụng Nhiều Mô Hình và Tối Ưu:** Mở rộng áp dụng nhiều loại mô hình và thực hiện các quá trình tối ưu hóa mô hình để thu thập nhiều tri thức và góc nhìn sâu hơn khi đối mặt với các tập dữ liệu phức tạp.

Những điều kiện này sẽ giúp nhóm không chỉ nắm bắt được sự đa dạng của dữ liệu mà còn đảm bảo rằng mỗi quá trình khai phá mang lại giá trị lâu dài và kiến thức sâu sắc về doanh nghiệp và thị trường.

### • Ứng Dụng

Bộ dữ liệu với quy mô lớn về doanh nghiệp và thông tin liên quan có thể được ứng dụng vào nhiều mục đích khác nhau, tùy thuộc vào yêu cầu và mục tiêu cụ thể của người sử dụng. Dưới đây là một số mục đích chính mà bạn có thể sử dụng bộ dữ liệu này.

#### 1. Nghiên Cứu Thị Trường và Ngành Công Nghiệp:

Phân tích cấu trúc thị trường và ngành công nghiệp thông qua thông tin về quy mô, địa điểm, ngày thành lập, và các thông số kinh doanh khác.

#### 2. Dự Báo Xu Hướng và Phát Triển:

Xây dựng mô hình dự đoán và phân tích xu hướng phát triển trong ngành công nghiệp, giúp dự báo các thay đổi trong môi trường kinh doanh.

#### 3. Phân Tích Cạnh Tranh:

So sánh và đánh giá vị thế cạnh tranh của các doanh nghiệp dựa trên các thông số như doanh thu, số lượng nhân viên, và quốc gia hoạt động.

#### 4. Tư Duy Chiến Lược:

Hỗ trợ quyết định chiến lược doanh nghiệp bằng cách hiểu rõ hơn về môi trường kinh doanh, khách hàng, và đối thủ cạnh tranh.

#### 5. Xây Dựng Mô Hình Dự Đoán:

Sử dụng dữ liệu để xây dựng mô hình dự đoán, ví dụ như mô hình dự đoán doanh thu tương lai dựa trên lịch sử hoạt động của doanh nghiệp.

#### 6. Phân Loại và Gom Cụm:

Áp dụng các thuật toán phân loại để gom cụm các doanh nghiệp có đặc điểm tương tự, giúp hiểu rõ hơn về các nhóm ngành công nghiệp.

#### 7. Nghiên Cứu Đối Tác và Liên Kết:

Phân tích các mối quan hệ giữa các doanh nghiệp và đối tác có thể mang lại cơ hội hợp tác và liên kết chiến lược.

#### 8. Quản lý Rủi Ro và An Toàn:

Đánh giá rủi ro và an toàn trong kinh doanh dựa trên các thông tin về ngành công nghiệp, vị trí hoạt động, và lịch sử hoạt động.

#### 9. Nghiên Cứu Xu Hướng Toàn Cầu:

Sử dụng dữ liệu liên quan đến Thế vận hội Tokyo 2021 để nghiên cứu về ảnh hưởng của sự kiện toàn cầu

---

đối với doanh nghiệp và ngành công nghiệp.

### **10. Đào Sâu vào Thông Tin Chi Tiết:**

Tìm hiểu chi tiết về từng doanh nghiệp, từ đó đào sâu vào các biểu đồ, thống kê, và thông tin cụ thể để có cái nhìn toàn diện về thị trường.

Bằng cách ứng dụng vào những mục đích trên, người sử dụng có thể đạt được sự hiểu biết sâu sắc và có cơ hội ra quyết định chiến lược thông minh trong lĩnh vực doanh nghiệp và kinh doanh.

## **Phục lục:**

Github: [link](#)

Colab: [link](#)