Social Media Usage

A study of how people indulge in this impactful and large industry

Barok Tsegaye, Christopher Booth, Logan Kerns

What was our goal?

Our goal was simply to analyze and understand the behavioral patterns in terms of social media usage and see what conclusions we could derive from them based on such info

	age	income	gender	time_spent	preferred_platform	interests	demographics	posts_per_week	exercise_hours
0	51	94413	female	4.5	Facebook	Sports	Suburban	7	5.5
1	64	87851	female	1.0	Facebook	Technology	Urban	14	8.0
2	41	68087	female	4.5	Instagram	Food	Rural	1	10.0
3	27	30988	male	5.0	Instagram	Music	Urban	3	7.0
4	55	79214	non-binary	1.0	Instagram	Sports	Rural	11	4.5
...	...	...	...	...	...	...	...	...	...
1995	32	45309	male	5.0	TikTok	Food	Suburban	3	9.5
1996	34	33802	male	1.0	Facebook	Technology	Urban	6	5.0
1997	23	27369	male	3.5	Snapchat	Music	Suburban	6	2.0
1998	39	61083	female	2.5	Snapchat	Technology	Suburban	11	0.0
1999	15	16970	male	0.5	Instagram	Food	Suburban	8	5.0

How did it go?

- Our Algorithms kinda suck

- Hierarchical Clustering was our best

- We ended up with around 420 total clusters

What did we come up with?

```
Cluster Summary:
              age          income    gender  time_spent  preferred_platform
Cluster
0        40.083333  50122.333333  0.416667    3.250000            1.416667
1        59.333333  81862.666667  0.666667    3.333333            1.333333
2        16.777778  18507.777778  0.666667    3.111111            1.000000
3        48.625000  61159.000000  0.625000    2.875000            1.375000
4        19.818182  22031.909091  0.545455    2.636364            1.727273
```

```
gender_mapping = {'male': 0, 'female': 1, 'non-binary': 2}
df['gender'] = df['gender'].replace(gender_mapping)

platform_mapping = {'Facebook': 0, 'Instagram': 1, 'Twitter': 2, 'Snapchat': 3, 'TikTok': 4}
df['preferred_platform'] = df['preferred_platform'].replace(platform_mapping)

interests_mapping = {'Sports': 0, 'Technology': 1, 'Food': 2, 'Music': 3, 'Art': 4, 'Fashion': 5, 'Travel': 6}
df['interests'] = df['interests'].replace(interests_mapping)

demographics_mapping = {'Suburban': 0, 'Urban': 1, 'Rural': 2}
df['demographics'] = df['demographics'].replace(demographics_mapping)
```

```
Averages for cluster 31:
user_id               1996.166667
age                     19.500000
gender                   0.333333
time_spent               2.750000
preferred_platform       2.166667
interests                1.833333
Cluster                 31.000000
dtype: float64
```

silhouette score: 0.723098507387255

THANKS!