In [2]:
```python
# Clean end-to-end pipeline: load -> drop columns -> scale numeric -> TruncatedSVD
import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import TruncatedSVD
from sklearn.mixture import GaussianMixture

# ----------------------------
# 0) Settings
# ----------------------------
RANDOM_STATE = 42
N_SVD = 50
K_RANGE = range(2, 11)
GMM_N_INIT_SCAN = 5
GMM_N_INIT_FINAL = 10
MIN_GLOBAL_RATE = 0.02
TOP_N = 15

file_path = r"C:\Users\123wi\OneDrive\Desktop\duits uni\unsupervised\project\mental
out_dir = os.path.join(os.path.dirname(file_path), "clustering_outputs")
os.makedirs(out_dir, exist_ok=True)

# ----------------------------
# 1) Load data
# ----------------------------
df = pd.read_excel(file_path)

print("Original shape:", df.shape)

drop_prefixes = ("live_country_", "work_country_", "mhcond_")
cols_to_drop = [c for c in df.columns if str(c).startswith(drop_prefixes)]

print("Original shape:", df.shape)
print("Dropping:", len(cols_to_drop))
print("Examples:", cols_to_drop[:20])

df = df.drop(columns=cols_to_drop, errors="ignore")

print("New shape:", df.shape)

# Drop columns that contain "Unknown" in the *title*
unknown_cols = [c for c in df.columns if "Unknown" in str(c)]
df = df.drop(columns=unknown_cols, errors="ignore")
print("After dropping 'Unknown' columns:", df.shape)

# Ensure numeric (one-hot should already be 0/1). Any weird cells -> 0
X_df = df.apply(pd.to_numeric, errors="coerce").fillna(0)

# Keep age + pct_time_affected_ord (scale them)
num_cols = [c for c in ["What is your age?", "pct_time_affected_ord"] if c in X_df.
if num_cols:
```

```python
    X_df[num_cols] = StandardScaler().fit_transform(X_df[num_cols])

X = X_df.values

# Save the cleaned matrix
cleaned_path = os.path.join(out_dir, "cleaned_input_used_for_clustering.xlsx")
X_df.to_excel(cleaned_path, index=False)
print("Saved cleaned input to:", cleaned_path)


# ----------------------------
# 2) TruncatedSVD reduction
# ----------------------------
n_svd = min(N_SVD, X.shape[1] - 1) if X.shape[1] > 1 else 1
svd = TruncatedSVD(n_components=n_svd, random_state=RANDOM_STATE)
Z = svd.fit_transform(X)

print("SVD dims:", Z.shape[1])
print("Explained variance ratio (first 10):", np.round(svd.explained_variance_ratio
print("Cumulative explained variance:", float(svd.explained_variance_ratio_.sum()))

# Save SVD embeddings
Z_df = pd.DataFrame(Z, columns=[f"svd_{i+1}" for i in range(Z.shape[1])])
svd_embed_path = os.path.join(out_dir, "svd_embeddings.xlsx")
Z_df.to_excel(svd_embed_path, index=False)
print("Saved SVD embeddings to:", svd_embed_path)

# Save explained variance
svd_var_df = pd.DataFrame({
    "component": [f"svd_{i+1}" for i in range(Z.shape[1])],
    "explained_variance_ratio": svd.explained_variance_ratio_,
    "cumulative_explained_variance": np.cumsum(svd.explained_variance_ratio_)
})
svd_var_path = os.path.join(out_dir, "svd_explained_variance.xlsx")
svd_var_df.to_excel(svd_var_path, index=False)
print("Saved SVD explained variance to:", svd_var_path)

# ----------------------------
# 3) Pick k with BIC/AIC (GMM)
# ----------------------------
bics, aics = [], []
for k in K_RANGE:
    gmm = GaussianMixture(
        n_components=k,
        covariance_type="full",
        random_state=RANDOM_STATE,
        n_init=GMM_N_INIT_SCAN
    )
    gmm.fit(Z)
    bics.append(gmm.bic(Z))
    aics.append(gmm.aic(Z))

best_k_bic = int(list(K_RANGE)[int(np.argmin(bics))])
best_k_aic = int(list(K_RANGE)[int(np.argmin(aics))])

print("Best k by BIC:", best_k_bic)
print("Best k by AIC:", best_k_aic)
```

```python
# Save BIC/AIC table
bic_aic_df = pd.DataFrame({
    "k": list(K_RANGE),
    "BIC": bics,
    "AIC": aics
})
bic_aic_path = os.path.join(out_dir, "gmm_bic_aic_table.xlsx")
bic_aic_df.to_excel(bic_aic_path, index=False)
print("Saved BIC/AIC table to:", bic_aic_path)

# Plot and save BIC/AIC curves
plt.figure()
plt.plot(list(K_RANGE), bics)
plt.xlabel("k")
plt.ylabel("BIC (lower is better)")
plt.title("GMM model selection (BIC)")
bic_png = os.path.join(out_dir, "gmm_bic_curve.png")
plt.savefig(bic_png, dpi=200, bbox_inches="tight")
plt.show()

plt.figure()
plt.plot(list(K_RANGE), aics)
plt.xlabel("k")
plt.ylabel("AIC (lower is better)")
plt.title("GMM model selection (AIC)")
aic_png = os.path.join(out_dir, "gmm_aic_curve.png")
plt.savefig(aic_png, dpi=200, bbox_inches="tight")
plt.show()

print("Saved BIC curve to:", bic_png)
print("Saved AIC curve to:", aic_png)

# ----------------------------
# 4) Fit GMM twice: BIC-k and AIC-k
# ----------------------------
def fit_and_plot(k, title_suffix):
    gmm = GaussianMixture(
        n_components=k,
        covariance_type="full",
        random_state=RANDOM_STATE,
        n_init=GMM_N_INIT_FINAL
    )
    labels = gmm.fit_predict(Z)

    plt.figure(figsize=(10, 6))
    plt.scatter(Z[:, 0], Z[:, 1], c=labels, s=18, alpha=0.85)
    plt.title(f"GMM clusters ({title_suffix}, k={k}) in 2D SVD space")
    plt.xlabel("SVD1")
    plt.ylabel("SVD2")

    fig_path = os.path.join(out_dir, f"gmm_clusters_{title_suffix}_k{k}.png")
    plt.savefig(fig_path, dpi=200, bbox_inches="tight")
    plt.show()

    counts = pd.Series(labels).value_counts().sort_index()
```

```python
        print(f"Cluster counts ({title_suffix}):", counts.to_dict())
        print("Saved cluster plot to:", fig_path)

        return labels, gmm, fig_path

labels_bic, gmm_bic, bic_cluster_plot = fit_and_plot(best_k_bic, "BIC_selected")
labels_aic, gmm_aic, aic_cluster_plot = fit_and_plot(best_k_aic, "AIC_selected")

# Save cluster labels (both) + first 2 SVD coords for easy plotting in Excel
clusters_out = pd.DataFrame({
    "svd_1": Z[:, 0],
    "svd_2": Z[:, 1],
    "cluster_bic": labels_bic,
    "cluster_aic": labels_aic
})
clusters_path = os.path.join(out_dir, "clusters_labels_and_2d_coords.xlsx")
clusters_out.to_excel(clusters_path, index=False)
print("Saved cluster labels + 2D coords to:", clusters_path)

# ----------------------------
# 5) Global "most important" features driving SVD1/SVD2
# ----------------------------
feature_names = X_df.columns.to_list()

def top_loadings(component_idx, top_n=20):
    load = svd.components_[component_idx]
    return pd.Series(np.abs(load), index=feature_names).sort_values(ascending=False

svd1_top = top_loadings(0, 20).to_frame("SVD1_loading_abs")
print("\nTop variables driving SVD1 (absolute loadings):")
display(svd1_top)

svd2_top = None
if Z.shape[1] >= 2:
    svd2_top = top_loadings(1, 20).to_frame("SVD2_loading_abs")
    print("\nTop variables driving SVD2 (absolute loadings):")
    display(svd2_top)

# Save loadings tables
svd_loadings_path = os.path.join(out_dir, "svd_top_loadings.xlsx")
with pd.ExcelWriter(svd_loadings_path, engine="openpyxl") as writer:
    svd1_top.to_excel(writer, sheet_name="SVD1_top20")
    if svd2_top is not None:
        svd2_top.to_excel(writer, sheet_name="SVD2_top20")
print("Saved top loadings to:", svd_loadings_path)

# ----------------------------
# 6) Cluster "reasons" = feature enrichment per cluster
# ----------------------------
def cluster_reasons_table(X_df, labels, top_n=15, min_global_rate=0.02, eps=1e-9):
    """

    Returns a dict of {cluster_id: DataFrame} with top enriched features by lift:
      lift = P(feature>0 | cluster) / P(feature>0 | all)
    """

    Xbin = (X_df > 0).astype(int)
    global_rate = Xbin.mean(axis=0)
```

```python
        keep = global_rate[global_rate >= min_global_rate].index.tolist()

        Xbin_k = Xbin[keep]
        global_rate_k = global_rate[keep]

        out_dict = {}
        for cl in sorted(pd.unique(labels)):
            idx = (labels == cl)
            n = int(idx.sum())
            cl_rate = Xbin_k.loc[idx].mean(axis=0)
            lift = (cl_rate + eps) / (global_rate_k + eps)

            out = pd.DataFrame({
                "lift": lift,
                "cluster_rate": cl_rate,
                "global_rate": global_rate_k
            }).sort_values("lift", ascending=False).head(top_n)

            out_dict[int(cl)] = out
        return out_dict

# AIC reasons
print("\n### Reasons for AIC-selected clustering ###")
aic_reasons = cluster_reasons_table(X_df, labels_aic, top_n=TOP_N, min_global_rate=

for cl, tab in aic_reasons.items():
    print(f"\n--- Cluster {cl} (n={(labels_aic==cl).sum()}) top {TOP_N} enriched fe
    display(tab)

# Save AIC "reasons" into an Excel workbook (one sheet per cluster)
aic_reasons_path = os.path.join(out_dir, "AIC_cluster_reasons.xlsx")
with pd.ExcelWriter(aic_reasons_path, engine="openpyxl") as writer:
    for cl, tab in aic_reasons.items():
        tab.to_excel(writer, sheet_name=f"cluster_{cl}")
print("Saved AIC cluster reasons to:", aic_reasons_path)

# Also save a single combined "Long" table for easier sorting/filtering in Excel
rows = []
for cl, tab in aic_reasons.items():
    t = tab.copy()
    t.insert(0, "cluster", cl)
    rows.append(t.reset_index().rename(columns={"index": "feature"}))
aic_reasons_long = pd.concat(rows, ignore_index=True)
aic_reasons_long_path = os.path.join(out_dir, "AIC_cluster_reasons_long.xlsx")
aic_reasons_long.to_excel(aic_reasons_long_path, index=False)
print("Saved AIC cluster reasons (long) to:", aic_reasons_long_path)

# Save cluster counts (AIC) nicely
aic_counts = pd.Series(labels_aic).value_counts().sort_index()
aic_counts_df = aic_counts.rename("count").to_frame()
aic_counts_path = os.path.join(out_dir, "AIC_cluster_counts.xlsx")
aic_counts_df.to_excel(aic_counts_path)
print("Saved AIC cluster counts to:", aic_counts_path)

# -----------------------------
# 7) Final "report" workbook (everything in ONE Excel file)
```

```python
# ----------------------------
report_path = os.path.join(out_dir, "FINAL_GMM_SVD_REPORT.xlsx")
with pd.ExcelWriter(report_path, engine="openpyxl") as writer:
    X_df.to_excel(writer, sheet_name="cleaned_input", index=False)
    Z_df.to_excel(writer, sheet_name="svd_embeddings", index=False)
    svd_var_df.to_excel(writer, sheet_name="svd_variance", index=False)
    bic_aic_df.to_excel(writer, sheet_name="bic_aic", index=False)
    clusters_out.to_excel(writer, sheet_name="labels_2d", index=False)
    svd1_top.to_excel(writer, sheet_name="svd1_top20")
    if svd2_top is not None:
        svd2_top.to_excel(writer, sheet_name="svd2_top20")
    aic_counts_df.to_excel(writer, sheet_name="aic_cluster_counts")
    aic_reasons_long.to_excel(writer, sheet_name="aic_cluster_reasons_long", index=

print("\n✅ DONE")
print("All outputs saved in:", out_dir)
print("Main report workbook:", report_path)
```

```
Original shape: (1433, 204)
Original shape: (1433, 204)
Dropping: 49
Examples: ['live_country_Australia', 'live_country_Brazil', 'live_country_Canada',
'live_country_France', 'live_country_Germany', 'live_country_Ireland', 'live_country
_Midwest', 'live_country_Netherlands', 'live_country_Northeast', 'live_country_Othe
r', 'live_country_Russia', 'live_country_South America', 'live_country_Sweden', 'liv
e_country_Switzerland', 'live_country_United Kingdom', 'live_country_West America',
'work_country_Australia', 'work_country_Brazil', 'work_country_Canada', 'work_countr
y_France']
New shape: (1433, 155)
After dropping 'Unknown' columns: (1433, 121)
Saved cleaned input to: C:\Users\123wi\OneDrive\Desktop\duits uni\unsupervised\proje
ct\clustering_outputs\cleaned_input_used_for_clustering.xlsx
SVD dims: 50
Explained variance ratio (first 10): [0.0213 0.0827 0.0654 0.0557 0.0473 0.034  0.02
88 0.025  0.0227 0.0217]
Cumulative explained variance: 0.8328190831130015
Saved SVD embeddings to: C:\Users\123wi\OneDrive\Desktop\duits uni\unsupervised\proj
ect\clustering_outputs\svd_embeddings.xlsx
Saved SVD explained variance to: C:\Users\123wi\OneDrive\Desktop\duits uni\unsupervi
sed\project\clustering_outputs\svd_explained_variance.xlsx
Best k by BIC: 2
Best k by AIC: 5
Saved BIC/AIC table to: C:\Users\123wi\OneDrive\Desktop\duits uni\unsupervised\proje
ct\clustering_outputs\gmm_bic_aic_table.xlsx
```
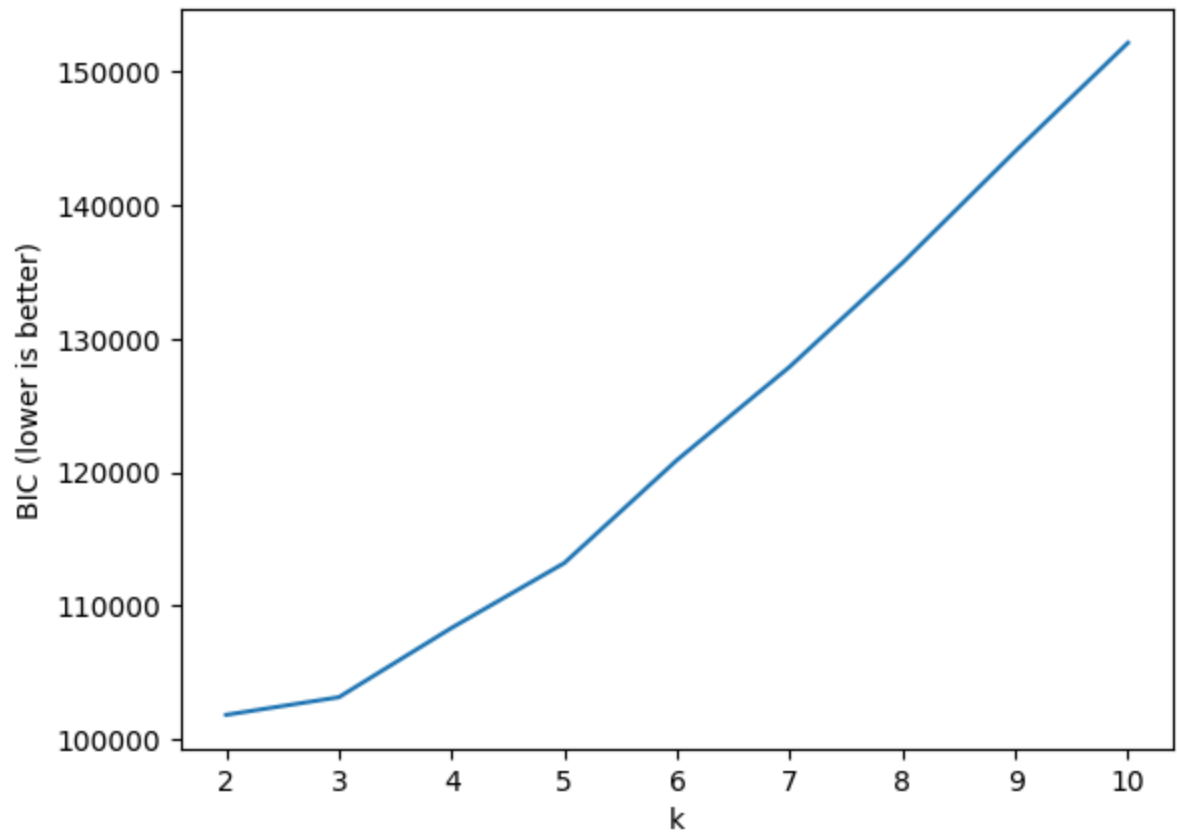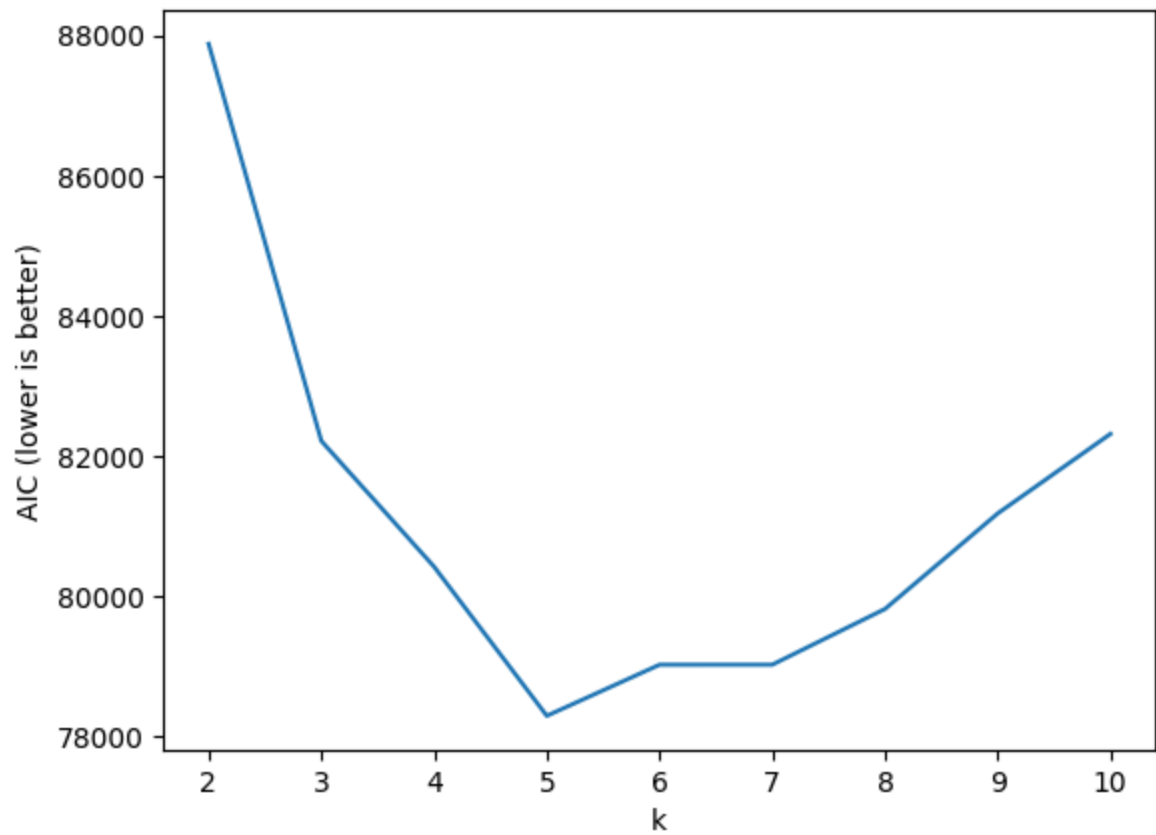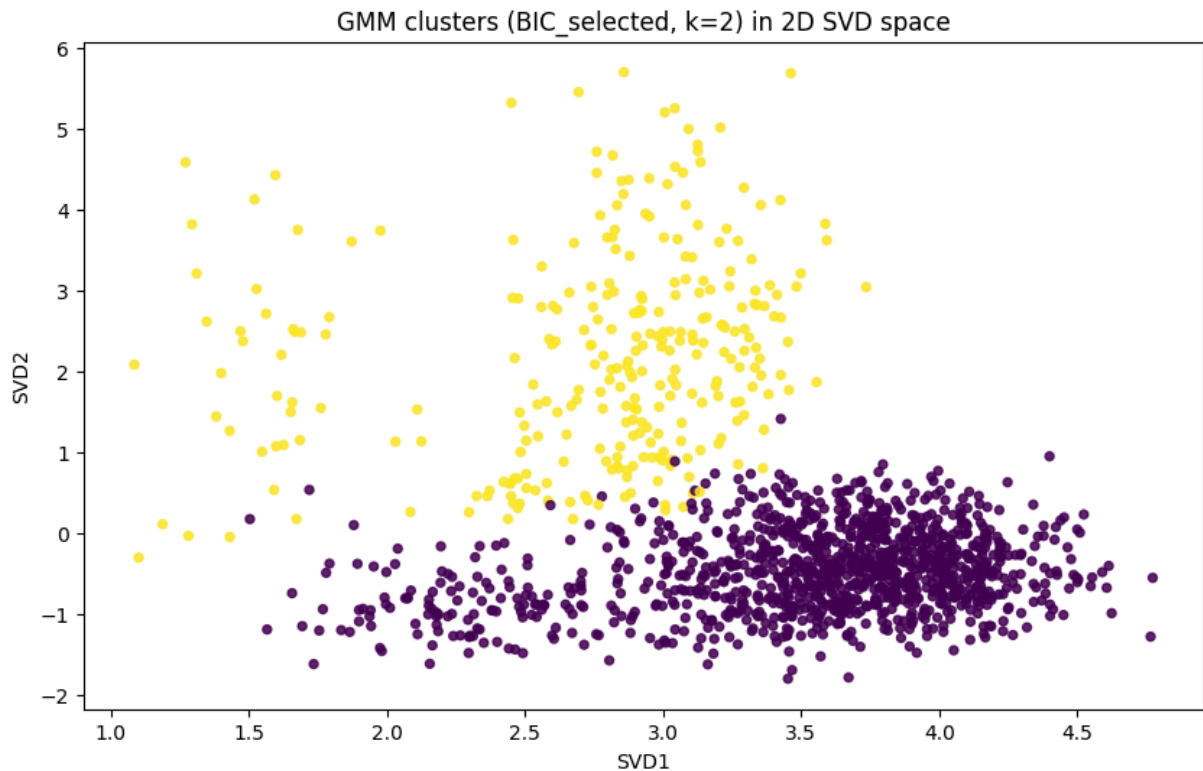
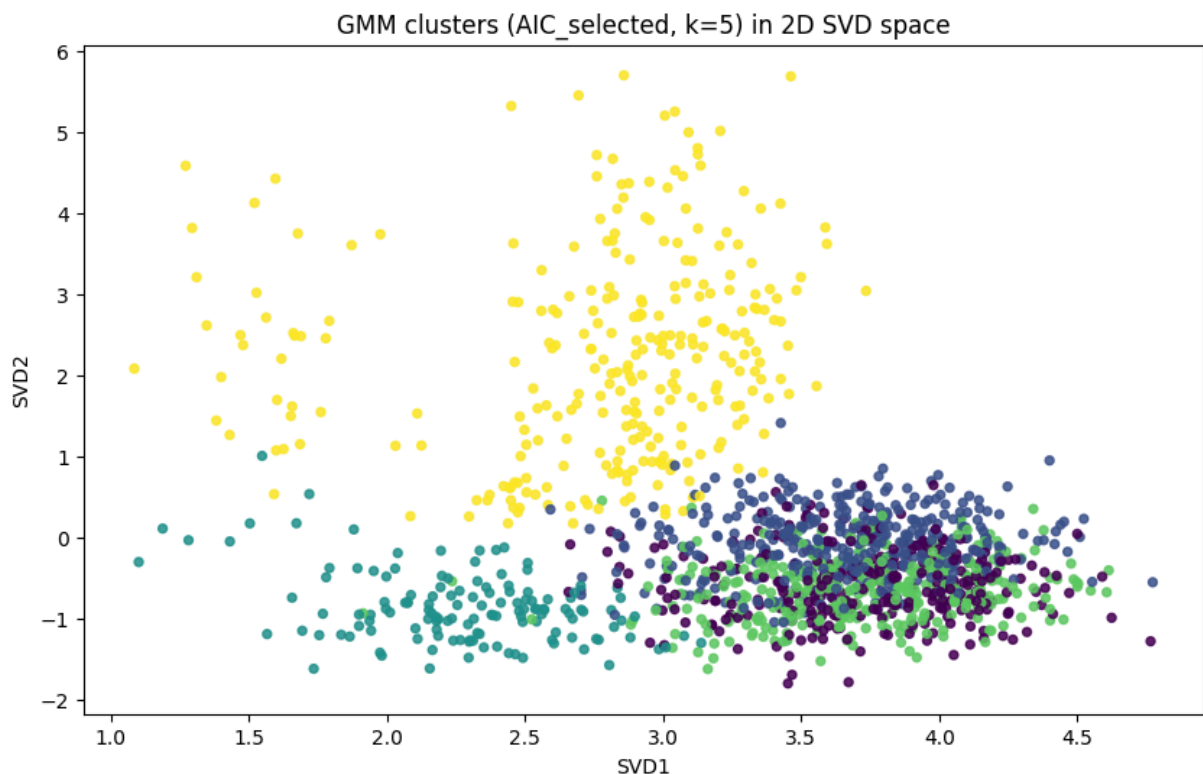## GMM model selection (BIC)

GMM model selection (AIC)

```
Saved BIC curve to: C:\Users\123wi\OneDrive\Desktop\duits uni\unsupervised\project\c
lustering_outputs\gmm_bic_curve.png
Saved AIC curve to: C:\Users\123wi\OneDrive\Desktop\duits uni\unsupervised\project\c
lustering_outputs\gmm_aic_curve.png
```

GMM clusters (BIC_selected, k=2) in 2D SVD space


```
Cluster counts (BIC_selected): {0: 1145, 1: 288}
Saved cluster plot to: C:\Users\123wi\OneDrive\Desktop\duits uni\unsupervised\projec
t\clustering_outputs\gmm_clusters_BIC_selected_k2.png
```

GMM clusters (AIC_selected, k=5) in 2D SVD space

```
Cluster counts (AIC_selected): {0: 304, 1: 371, 2: 137, 3: 339, 4: 282}
Saved cluster plot to: C:\Users\123wi\OneDrive\Desktop\duits uni\unsupervised\projec
t\clustering_outputs\gmm_clusters_AIC_selected_k5.png
Saved cluster labels + 2D coords to: C:\Users\123wi\OneDrive\Desktop\duits uni\unsup
ervised\project\clustering_outputs\clusters_labels_and_2d_coords.xlsx

Top variables driving SVD1 (absolute loadings):
```

	SVD1_loading_abs
Do you have previous employers?	0.263501
Is your employer primarily a tech company/organization?	0.234073
yn__Have you heard of or observed negative consequences for co-workers who have been open about mental health issues in your workplace?_No	0.219207
gender_Male	0.210631
prev__Did your previous employers ever formally discuss mental health (as part of a wellness campaign or other official communication)? _NoneDid	0.187390
yn__Would you bring up a mental health issue with a potential employer in an interview?_No	0.179876
yn__Do you think that discussing a physical health issue with your employer would have negative consequences?_No	0.178006
prev__Did your previous employers provide resources to learn more about mental health issues and how to seek help?_NoneDid	0.176481
Have you ever sought treatment for a mental health issue from a mental health professional?	0.173527
yn__Has your employer ever formally discussed mental health (for example, as part of a wellness campaign or other official communication)?_No	0.170841
prevneg__Did you hear of or observe negative consequences for co-workers with mental health issues in your previous workplaces?_None	0.159622
prev6__Would you have been willing to discuss a mental health issue with your previous co-workers?_Some	0.156612
remote_Sometimes	0.152691
role_Back-end Developer	0.150415
yn__Do you have a family history of mental illness?_Yes	0.139090
prev6__Would you have been willing to discuss a mental health issue with your direct supervisor(s)?_Some	0.139065
prevneg__Do you think that discussing a mental health disorder with previous employers would have negative consequences?_Some	0.132118
mhatt__Do you feel that being identified as a person with a mental health issue would hurt your career?_Yes	0.131878
prevneg__Do you think that discussing a physical health issue with previous employers would have negative consequences?_Some	0.131773
mhatt__How willing would you be to share with friends and family that you have a mental illness?_SomewhatOpen	0.130559

```
Top variables driving SVD2 (absolute loadings):
```

	SVD2_loading_abs
pct_time_affected_ord	0.674921
Are you self-employed?	0.262566
What is your age?	0.227886
yn__Have you heard of or observed negative consequences for co-workers who have been open about mental health issues in your workplace?_No	0.220542
yn__Do you think that discussing a physical health issue with your employer would have negative consequences?_No	0.206256
yn__Has your employer ever formally discussed mental health (for example, as part of a wellness campaign or other official communication)?_No	0.157037
role_One-person shop	0.135443
yn__Would you feel comfortable discussing a mental health disorder with your direct supervisor(s)?_Yes	0.122274
Have you ever sought treatment for a mental health issue from a mental health professional?	0.117016
mhatt__Do you feel that being identified as a person with a mental health issue would hurt your career?_Yes	0.109074
yn__Does your employer offer resources to learn more about mental health concerns and options for seeking help?_No	0.106927
prev__Did your previous employers ever formally discuss mental health (as part of a wellness campaign or other official communication)?_NoneDid	0.106217
prevneg__Did you hear of or observe negative consequences for co-workers with mental health issues in your previous workplaces?_Some	0.101312
benefits_Yes	0.101286
mhatt__Do you think that team members/co-workers would view you more negatively if they knew you suffered from a mental health issue?_Yes	0.099660
interf__If you have a mental health issue, do you feel that it interferes with your work when NOT being treated effectively?_Often	0.098883
yn__Do you feel that your employer takes mental health as seriously as physical health?_Yes	0.098401
prev__Did your previous employers provide resources to learn more about mental health issues and how to seek help?_NoneDid	0.096829
Is your employer primarily a tech company/organization?	0.095554
mhatt__Have you observed or experienced an unsupportive or badly handled response to a mental health issue in your current or previous workplace?_No	0.093162

```
Saved top loadings to: C:\Users\123wi\OneDrive\Desktop\duits uni\unsupervised\projec
t\clustering_outputs\svd_top_loadings.xlsx

### Reasons for AIC-selected clustering ###

--- Cluster 0 (n=304) top 15 enriched features ---
```

	lift	cluster_rate	global_rate
prev__Did your previous employers provide resources to learn more about mental health issues and how to seek help?_AllDid	3.142544	0.111842	0.035590
prev__Did your previous employers ever formally discuss mental health (as part of a wellness campaign or other official communication)?_AllDid	3.142544	0.072368	0.023029
prev__Did your previous employers ever formally discuss mental health (as part of a wellness campaign or other official communication)?_SomeDid	2.514035	0.447368	0.177948
prev6__Was your anonymity protected if you chose to take advantage of mental health or substance abuse treatment resources with previous employers?_Sometimes	2.416326	0.200658	0.083043
prev6__Did you feel that your previous employers took mental health as seriously as physical health?_AllDid	2.411720	0.072368	0.030007
prev__Did your previous employers provide resources to learn more about mental health issues and how to seek help?_SomeDid	2.401378	0.621711	0.258897
prev6__Was your anonymity protected if you chose to take advantage of mental health or substance abuse treatment resources with previous employers?_YesAlways	2.356908	0.269737	0.114445
prev6__Have your previous employers provided mental health benefits?_AllDid	2.331834	0.305921	0.131193
prev6__Were you aware of the options for mental health care provided by your previous employers?_AwareSome	2.307806	0.618421	0.267969
prev6__Have your previous employers provided mental health benefits?_SomeDid	2.302657	0.628289	0.272854
prev6__Were you aware of the options for mental health care provided by your previous employers?_AwareAll	2.031368	0.256579	0.126308
prevneg__Do you think that discussing a mental health disorder with previous employers would have negative consequences?_None	2.002329	0.157895	0.078856
mhcare_options_Yes	1.950015	0.417763	0.214236
yn__Would you feel comfortable discussing a mental health disorder with your coworkers?_Yes	1.936950	0.371711	0.191905
yn__Would you feel comfortable discussing a mental health disorder with your direct supervisor(s)?_Yes	1.872310	0.559211	0.298674

```
--- Cluster 1 (n=371) top 15 enriched features ---
```

	lift	cluster_rate	global_rate
yn__Do you think that discussing a physical health issue with your employer would have negative consequences?_Yes	3.203077	0.091644	0.028611
mhatt__If a mental health issue prompted you to request a medical leave from work, asking for that leave would be:_VeryDifficult	2.585934	0.212938	0.082345
yn__Do you think that discussing a mental health disorder with your employer would have negative consequences?_Yes	2.459348	0.439353	0.178646
yn__Would you feel comfortable discussing a mental health disorder with your direct supervisor(s)?_No	2.414084	0.566038	0.234473
yn__Would you feel comfortable discussing a mental health disorder with your coworkers?_No	2.335256	0.638814	0.273552
yn__Have you heard of or observed negative consequences for co-workers who have been open about mental health issues in your workplace?_Yes	2.285989	0.156334	0.068388
prevneg__Do you think that discussing a mental health disorder with previous employers would have negative consequences?_All	2.204721	0.347709	0.157711
mhatt__Do you think that team members/co-workers would view you more negatively if they knew you suffered from a mental health issue?_Yes	2.187322	0.679245	0.310537
prevneg__Do you think that discussing a physical health issue with previous employers would have negative consequences?_All	2.140052	0.110512	0.051640
yn__Do you feel that your employer takes mental health as seriously as physical health?_No	2.116108	0.447439	0.211445
mhatt__If a mental health issue prompted you to request a medical leave from work, asking for that leave would be:_SomewhatDifficult	2.057430	0.285714	0.138870
prev6__Was your anonymity protected if you chose to take advantage of mental health or substance abuse treatment resources with previous employers?_No	1.947228	0.164420	0.084438
prev6__Would you have been willing to discuss a mental health issue with your direct supervisor(s)?_None	1.912697	0.555256	0.290300
mhatt__How willing would you be to share with friends and family that you have a mental illness?_NotOpen	1.905517	0.099730	0.052338
mhatt__Do you feel that being identified as a person with a mental health issue would hurt your career?_Yes	1.867662	0.870620	0.466155

--- Cluster 2 (n=137) top 15 enriched features ---

	lift	cluster_rate	global_rate
yn__Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources provided by your employer?_No	2.241397	0.131387	0.058618
remote_Never	1.790425	0.416058	0.232380
yn__Would you feel comfortable discussing a mental health disorder with your direct supervisor(s)?_No	1.712178	0.401460	0.234473
mhatt__If a mental health issue prompted you to request a medical leave from work, asking for that leave would be:_SomewhatEasy	1.600618	0.313869	0.196092
company_size_1-5	1.568978	0.065693	0.041870
benefits_No	1.554843	0.321168	0.206560
company_size_>1000	1.552635	0.277372	0.178646
mhatt__How willing would you be to share with friends and family that you have a mental illness?_NotOpen	1.534112	0.080292	0.052338
mhatt__If a mental health issue prompted you to request a medical leave from work, asking for that leave would be:_VeryDifficult	1.506928	0.124088	0.082345
yn__Do you feel that your employer takes mental health as seriously as physical health?_No	1.484402	0.313869	0.211445
mhatt__Have you observed or experienced an unsupportive or badly handled response to a mental health issue in your current or previous workplace?_No	1.457369	0.576642	0.395673
yn__Would you feel comfortable discussing a mental health disorder with your coworkers?_No	1.414215	0.386861	0.273552
yn__Does your employer offer resources to learn more about mental health concerns and options for seeking help?_No	1.378888	0.510949	0.370551
yn__Do you have a family history of mental illness?_No	1.328916	0.452555	0.340544
mhatt__How willing would you be to share with friends and family that you have a mental illness?_NotApplicable	1.307482	0.102190	0.078158

```
--- Cluster 3 (n=339) top 15 enriched features ---
```

	lift	cluster_rate	global_rate
mhatt__If a mental health issue prompted you to request a medical leave from work, asking for that leave would be:_VeryEasy	1.998284	0.306785	0.153524
prev6__Were you aware of the options for mental health care provided by your previous employers?_NotAware	1.990096	0.808260	0.406141
mhatt__Do you think that team members/co-workers would view you more negatively if they knew you suffered from a mental health issue?_No	1.895291	0.525074	0.277041
yn__Would you feel comfortable discussing a mental health disorder with your direct supervisor(s)?_Yes	1.846904	0.551622	0.298674
yn__Would you feel comfortable discussing a mental health disorder with your coworkers?_Yes	1.844570	0.353982	0.191905
prev6__Would you have been willing to discuss a mental health issue with your direct supervisor(s)?_All	1.818124	0.117994	0.064899
yn__Do you feel that your employer takes mental health as seriously as physical health?_Yes	1.811631	0.442478	0.244243
mhatt__Do you feel that being identified as a person with a mental health issue would hurt your career?_No	1.791160	0.221239	0.123517
yn__Would you bring up a mental health issue with a potential employer in an interview?_Yes	1.698404	0.132743	0.078158
prev6__Have your previous employers provided mental health benefits?_NoneDid	1.636312	0.424779	0.259595
prev__Did your previous employers provide resources to learn more about mental health issues and how to seek help?_NoneDid	1.581412	0.929204	0.587579
prevneg__Did you hear of or observe negative consequences for co-workers with mental health issues in your previous workplaces?_None	1.550323	0.820059	0.528960
yn__Do you think that discussing a physical health issue with your employer would have negative consequences?_No	1.545406	0.902655	0.584089
mhatt__Do you feel that being identified as a person with a mental health issue would hurt your career?_Maybe	1.524070	0.625369	0.410328
mhatt__If a mental health issue prompted you to request a medical leave from work, asking for that leave would be:_SomewhatEasy	1.519363	0.297935	0.196092

--- Cluster 4 (n=282) top 15 enriched features ---

	lift	cluster_rate	global_rate
pct_time_affected_ord	5.081560	0.723404	0.142359
Are you self-employed?	4.975326	0.996454	0.200279
role_One-person shop	4.324061	0.485816	0.112352
role_Sales	3.278426	0.070922	0.021633
role_Executive Leadership	2.062812	0.145390	0.070482
remote_Always	1.866696	0.446809	0.239358
role_Designer	1.731495	0.163121	0.094208
interf__If you have a mental health issue, do you feel that it interferes with your work when being treated effectively?_Often	1.563557	0.070922	0.045359
yn__Do you think that discussing a mental health disorder with your employer would have negative consequences?_No	1.496956	0.656028	0.438241
mhatt__Have you observed or experienced an unsupportive or badly handled response to a mental health issue in your current or previous workplace?_Observed	1.366632	0.251773	0.184229
prevneg__Did you hear of or observe negative consequences for co-workers with mental health issues in your previous workplaces?_All	1.311370	0.056738	0.043266
What is your age?	1.310821	0.546099	0.416609
mhatt__How willing would you be to share with friends and family that you have a mental illness?_NotApplicable	1.270390	0.099291	0.078158
prev6__Have your previous employers provided mental health benefits?_NoneDid	1.256730	0.326241	0.259595
prevneg__Did you hear of or observe negative consequences for co-workers with mental health issues in your previous workplaces?_Some	1.236055	0.382979	0.309839

```
Saved AIC cluster reasons to: C:\Users\123wi\OneDrive\Desktop\duits uni\unsupervised
\project\clustering_outputs\AIC_cluster_reasons.xlsx
Saved AIC cluster reasons (long) to: C:\Users\123wi\OneDrive\Desktop\duits uni\unsup
ervised\project\clustering_outputs\AIC_cluster_reasons_long.xlsx
Saved AIC cluster counts to: C:\Users\123wi\OneDrive\Desktop\duits uni\unsupervised
\project\clustering_outputs\AIC_cluster_counts.xlsx

✅  DONE
All outputs saved in: C:\Users\123wi\OneDrive\Desktop\duits uni\unsupervised\project
\clustering_outputs
Main report workbook: C:\Users\123wi\OneDrive\Desktop\duits uni\unsupervised\project
\clustering_outputs\FINAL_GMM_SVD_REPORT.xlsx
```

In [ ]:

file:///C:/Users/123wi/OneDrive/Desktop/duits%20uni/unsupervised/project/Dimension%20reduction%20and%20clustering%20Final.html

18/18