

Loading the packages

```
In [1]: import os
import pandas as pd

#https://www.sbert.net/docs/pretrained_models.html "to see the new models"

from sentence_transformers import SentenceTransformer
embedder = SentenceTransformer('all-mpnet-base-v2')
```

```
In [2]: from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
from wordcloud import WordCloud
```

```
In [3]: # Loading the data
df = pd.read_excel('cluster dos cluster.xlsx')
df.head()
```

Out[3]:

	cluster	cluster_name	Description
0	0	Social and Geoligical	Collaborative work friendly. Collaborative wor...
1	1	Agile and training	Team training and monitoring. Types of trainin...
2	2	Team organization and attitude	Participation and support to solve issues. Thi...
3	3	Stakeholders	Understanding over the client's business proce...
4	4	Green software development	Interfacing with different layers of developme...

Clustering

```
In [4]: corpus = list(df['Description'])
```

```
In [5]: corpus
```

```
Out[5]: ['Collaborative work friendly. Collaborative work friendly is the ability to function on multidisciplinary teams. In today’s multicultural world, this outcome also implies an ability to collaborate with people from different cultures, abilities, and background s. Frequency of social events. Social events include but are not limited to telling people what to do, spending time with sharp and witty people, giving speeches, attending parties, laughing without reservation, voicing strong personal values and opinions in a group, telling jokes, criticizing someone, and asking for help or advice. Social interaction. Social interaction is how individuals act and react concerning one another. Climatic condition. Weather generally refers to day-to-day temperature and precipitation activity, whereas climate is the term for the average atmospheric conditions over more extended periods. Climate is the average weather conditions for a particular geographical location over notable years. Social facilities. Inequality manifests in the unequal provision of social amenities within the districts of the local government. The social amenities are educational, health, and market facilities, like drinking water, sanitation, electricity, housing qualities, and drainage arrangement. It is connected to geogr aphic dimension. Geological condition. Ecological-geological conditions are considered a geographical environment created by a set of contemporary morphologically expressed geological factors that influence specific features of the functioning of the biota, in cluding human beings, within the framework of the ecological-geological system. Communality. Communality refers to the personal characteristics that the trustor has in common with the trustee, like a similar goal they wish to achieve, shared language use, comm on identity characteristics, or shared values. It is a personal technical dimension. Internalised norms. Internalized norms are integrity, discretion, honesty, fairness, and loyalty. This criterion refers to the intrinsic moral norms a trustee uses to guard hi s/her actions. The language Analysis regarding how a trustee uses the chat and walls could infer some people’s values. Communication protocols and customs. A communication Protocol is a system of rules that allows two or more entities in a communications syste m to transmit information via any variation of a physical quantity. Communication protocols are formal descriptions of formats and rules for producing digital messages for electronic data exchange.',
'Team training and monitoring. Types of training: Induction Program, Training on Application Functionality, On the job Training, Trainee ramp-up. The education and support to distributed team members are essential in GSD. Advance and Uniform Development Envir onment and Training. For the Advance and Uniform Development Environment, all the development teams in GSD must use the same development environment. Even to use the latest technology and tools, the developers need to be trained appropriately to acquire the re quired skill and knowledge to ease the integration process in the long run. Agile training. Adopting Agile-driven team training methodologies is an efficient way of excelling in agile software project management with significant advantages in production costs, time-to-market, complexity, and quality improvement over traditional human resource management methodologies. It is a human resource dimension. Scrum hours. Scrum is the most widely applied Agile methodology and is a process framework for delivering products a nd services of the highest possible quality and handling complex problems or situations. Iterative and incremental approaches are used to develop products using cross-functional teams. Number os sprints. The number of sprints is short work cycles for increment al development. Scaling tools and standards. Beyond regular global projects, agile scaling involves many challenges, including coordination among multiple agile teams and the need for an initial architecture and requirement analysis. Several frameworks for sca ling agile software development have been suggested, such as the Scaled Agile Framework (SAFe), Disciplined Agile Delivery, Large-Scale Scrum, Nexus, and Scrum@Scale. Training of DevOps activities. Training in DevOps activities has a positive impact while impl ementing DevOps. Formal training sessions are required to understand the concept and DevOps environment properly. The organization must support its teams with training sessions to help their organization successfully work on DevOps activities.',
"Participation and support to solve issues. This category consists of ideas to increase the acceptance of, and commitment to, both the organization and the transformation process. This can be achieved by the employee's participation in organizational processe
```

```
In [6]: corpus_embeddings = embedder.encode(corpus)
```

```
In [7]: corpus_embeddings
```

```
Out[7]: array([[ 0.02656563, -0.00952512, -0.01910171, ...,  0.03667088,
                0.01796212,  0.01580494],
               [ 0.01627997, -0.09900156, -0.0416006 , ..., -0.00957133,
                0.00788075, -0.00171264],
               [ 0.04984569, -0.04072112, -0.01857589, ...,  0.03353123,
                0.0144909 , -0.01424925],
               ...,
               [ 0.03298348, -0.01486641, -0.02557615, ...,  0.00109911,
                -0.0069869 , -0.00196207],
               [ 0.06567325, -0.04980197, -0.02835721, ..., -0.00270857,
                -0.01901965, -0.01220288],
               [ 0.01808701, -0.03959769, -0.03459328, ..., -0.00744437,
                0.00292018, -0.00765984]], dtype=float32)
```

```
In [17]: clustering_model = KMeans(n_clusters=6, random_state=0, n_init=300)
clustering_model.fit(corpus_embeddings)
cluster_assignment = clustering_model.labels_
```

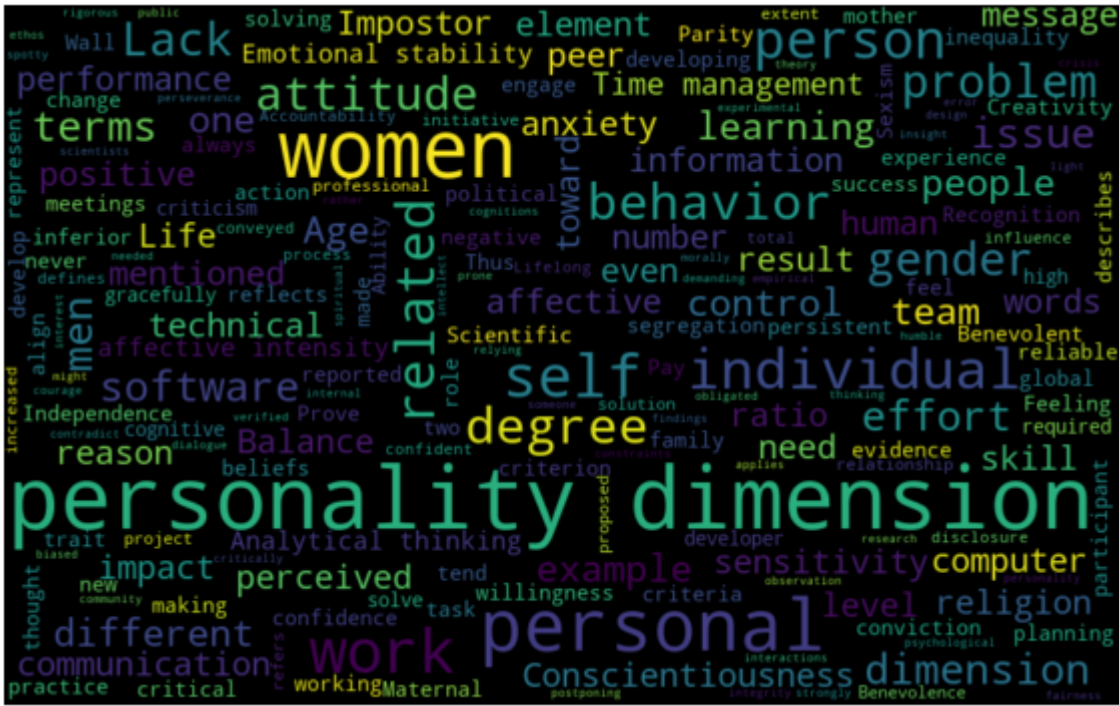
```
In [18]: cluster_assignment
```

```
Out[18]: array([4, 5, 1, 1, 3, 1, 3, 0, 5, 1, 1, 5, 2, 0, 1, 4, 5, 1, 5, 0, 4, 1,
                3, 5, 3])
```

Out[19]:		corpus	cluster	cluster_name
0	Collaborative work friendly. Collaborative wor...		4	Social and Geological
1	Team training and monitoring. Types of trainin...		5	Agile and training
2	Participation and support to solve issues. Thi...		1	Team organization and attitude
3	Understanding over the client's business proce...		1	Stakeholders
4	Interfacing with different layers of developme...		3	Green software development

```
Out[21]: <function print>
```

```
In [23]: word_cloud(cluster_df, 0)
```



A word cloud visualization of terms related to organizational knowledge management. The words are arranged in a dense, overlapping manner, with colors ranging from purple to yellow. The most prominent words include 'knowledge', 'experience', 'information', 'project', 'work', 'skill', 'management', 'lack', 'communication', 'task', 'team', 'member', 'site', 'issue', 'different', 'need', 'problem', 'interaction', 'planning', 'responsibilities', 'level', 'mutual', 'knowledge transfer', 'development', 'learning', 'dimension', 'working', 'risk', 'client', 'lead', 'production', 'use', 'coordination', 'stakeholder', 'distributed', 'role', 'integration', 'selection', 'technology', 'difference', 'employee', 'culture', 'software development', 'distance practice', 'global understanding', 'environment', 'effectiveness', 'organization', 'ability conflict', 'context regular meetings', 'failure', 'attitude', 'change personal', 'location', 'may', 'group', 'impact', 'size', 'language', 'increase', 'one effort', 'among due', 'staff', 'addition', 'defined action', 'manager', 'time', 'individual', 'business survey', 'domain will power actual', 'frequency', 'relationship', 'organizational tool', 'effectively improve', 'understand', 'regional culture', 'value \$', 'human resource', 'related', 'goal', 'technical', 'quality', 'effective', 'number', 'process', 'support', 'result may', 'productivity', 'satisfaction', 'degree', 'thus', 'informal', 'cultural', 'software development'.



[illegible][illegible]

