

In [1]:

```
import pandas as pd
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.cluster import KMeans
from nltk.tokenize import word_tokenize
import nltk
from nltk.corpus import stopwords
nltk.download('stopwords')
nltk.download('punkt')
import string
import matplotlib.pyplot as plt
import requests
from bs4 import BeautifulSoup
import re
```

```
[nltk_data] Downloading package stopwords to C:\Users\Mannahil
[nltk_data]   Miftah\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to C:\Users\Mannahil
[nltk_data]   Miftah\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

In [2]:

```
links = ['/bukhari', '/muslim', '/nasai', '/abudawud', '/tirmidhi', '/ibnmajah']
total = [97, 56, 51, 43, 49, 37]
base_url = 'https://sunnah.com/'

urls = [f'{base_url}{link}/{i}' for link, count in zip(links, total) for i in range(1, count + 1)]

book_names, collected_by, arabic_hadith, english_hadith, references = [], [], [], [], []

for url in urls:
    req = requests.get(url)
    soup = BeautifulSoup(req.text, 'html.parser')

    # Collect Arabic Hadith
    arabic_hadith.extend([arabic.text for arabic in soup.find_all("div", class_="arabic_hadith_full arabic")])

    # Collect Collection Information
    collection_info = soup.find("div", class_="crumbs").text
    collection_parts = re.sub(r'Home » ', '', collection_info).split(" » ")
    collected_by.extend([collection_parts[0]] * len(collection_parts))
    book_names.extend([collection_parts[1]] * len(collection_parts))

    # Collect English Hadith
    english_hadith.extend([hadith.text for hadith in soup.find_all("div", class_="text_details")])

    # Collect References
    references.extend([reference.find("a").text for reference in soup.find_all("tr") if reference.find("a") is not None])

data = list(zip(book_names, collected_by, arabic_hadith, english_hadith, references))

df = pd.DataFrame(data, columns=['Book Name', 'Collected By', 'Arabic', 'English', 'References'])

# Clean the data by removing unwanted characters
df = df.replace(to_replace=r'\\r|\\r|\\n|\\t|Narrated:|\\u200f\\.\\u200f|\\u200f"\\u200f|\\n', value='', regex=True)
```

In [3]:

df

Out [3]:

	Book Name	Collected By	Arabic	English	References
0	Revelation	Sahih al-Bukhari	... حَدَّثَنَا الْحُمَيْدِيُّ عَبْدُ اللَّهِ بْنُ	I heard Allah's Messenger (ﷺ) saying, "Th...	Sahih al-Bukhari 1
1	Revelation	Sahih al-Bukhari	... حَدَّثَنَا عَبْدُ اللَّهِ بْنُ يُوسُفَ، قَالَ	\r\r (the mother of the faithful believers...	Sahih al-Bukhari 2
2	Belief	Sahih al-Bukhari	... حَدَّثَنَا يَحْيَى بْنُ بُكَيْرٍ، قَالَ	The commencement of the Divine Inspiration to ...	Sahih al-Bukhari 3
3	Belief	Sahih al-Bukhari	... قَالَ ابْنُ شِهَابٍ وَأَخْبَرَنِي أَبُو سَلَمَ	"While I was walking, all of a sudden I heard ...	Sahih al-Bukhari 4
4	Knowledge	Sahih al-Bukhari	... حَدَّثَنَا مُوسَى بْنُ إِسْمَاعِيلَ، قَالَ حَد	\r\r Ibn 'Abbas in the explanation of the ...	Sahih al-Bukhari 5
...
661	Interpretation of Dreams	Sunan Ibn Majah	... حَدَّثَنَا أَبُو الْيَمَانِ، قَالَ أَخْبَرَنَا	Anas bin Malik Al-Ansari, told me, "Abu Bakr u...	Sahih al-Bukhari 680
662	Tribulations	Sunan Ibn Majah	... حَدَّثَنَا أَبُو مَعْمَرٍ، قَالَ حَدَّثَنَا عَ	The Prophet (ﷺ) did not come out for three day...	Sahih al-Bukhari 681
663	Tribulations	Sunan Ibn Majah	... حَدَّثَنَا يَحْيَى بْنُ سُلَيْمَانَ، قَالَ حَد	My father said, "When Allah's Messenger (ﷺ) be...	Sahih al-Bukhari 682
664	Zuhd	Sunan Ibn Majah	... حَدَّثَنَا زَكْرِيَّا بْنُ يَحْيَى، قَالَ حَد	`Aisha said, "Allah's Messenger (ﷺ) ordered Ab...	Sahih al-Bukhari 683
665	Zuhd	Sunan Ibn Majah	... حَدَّثَنَا عَبْدُ اللَّهِ بْنُ يُوسُفَ، قَالَ	Allah's Messenger (ﷺ) went to establish peace ...	Sahih al-Bukhari 684

666 rows x 5 columns

In [8]:

```
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(df['English'])

num_clusters = 5
kmeans = KMeans(n_clusters=num_clusters, random_state=42)
df['Cluster'] = kmeans.fit_predict(X)

cluster_mapping = {
    0: 'The Profession of Faith (the Shahada)',
    1: 'Daily Prayers (Salat)',
    2: 'Alms-giving (Zakat)',
    3: 'Fasting during Ramadan (Saum)',
    4: 'Pilgrimage to Mecca (Hajj)'
}
df['Pillar'] = df['Cluster'].map(cluster_mapping)

print(df[['Book Name', 'Collected By', 'Arabic', 'English', 'References', 'Cluster', 'Pillar']])
```

	Book Name	Collected By	\
0	Revelation	Sahih al-Bukhari	
1	Revelation	Sahih al-Bukhari	
2	Belief	Sahih al-Bukhari	
3	Belief	Sahih al-Bukhari	
4	Knowledge	Sahih al-Bukhari	
..	
661	Interpretation of Dreams	Sunan Ibn Majah	
662	Tribulations	Sunan Ibn Majah	
663	Tribulations	Sunan Ibn Majah	
664	Zuhd	Sunan Ibn Majah	
665	Zuhd	Sunan Ibn Majah	

Arabic \

0	... حَدَّثَنَا الْحُمَيْدِيُّ عَبْدُ اللَّهِ بْنُ
1	... حَدَّثَنَا عَبْدُ اللَّهِ بْنُ يُوسُفَ، قَالَ

```

2... حَدَّثَنَا بِحَدِيثِ بْنِ أَبِي كَيْسٍ، قَالَ حَدَّثَ
3... قَالَ ابْنُ شِهَابٍ وَأَخْبَرَنِي أَبُو سَلَمَةَ
4... حَدَّثَنَا مُوسَى بْنُ إِسْمَاعِيلَ، قَالَ حَدَّثَ
...
661... حَدَّثَنَا أَبُو الْيَمَانِ، قَالَ أَخْبَرَنَا
662... حَدَّثَنَا أَبُو مَعْمَرٍ، قَالَ حَدَّثَنَا ع
663... حَدَّثَنَا بِحَدِيثِ بْنِ سُلَيْمَانَ، قَالَ حَدَّثَ
664... حَدَّثَنَا زَكَرِيَاءُ بْنُ يَحْيَى، قَالَ حَدَّثَ
665... حَدَّثَنَا عَبْدُ اللَّهِ بْنُ يُوسُفَ، قَالَ

```

```

English References \
0 I heard Allah's Messenger (ﷺ) saying, "Th... Sahih al-Bukhari 1
1 \r\r (the mother of the faithful believers... Sahih al-Bukhari 2
2 The commencement of the Divine Inspiration to ... Sahih al-Bukhari 3
3 "While I was walking, all of a sudden I heard ... Sahih al-Bukhari 4
4 \r\r Ibn 'Abbas in the explanation of the ... Sahih al-Bukhari 5
...
661 Anas bin Malik Al-Ansari, told me, "Abu Bakr u... Sahih al-Bukhari 680
662 The Prophet (ﷺ) did not come out for three day... Sahih al-Bukhari 681
663 My father said, "When Allah's Messenger (ﷺ) be... Sahih al-Bukhari 682
664 `Aisha said, "Allah's Messenger (ﷺ) ordered Ab... Sahih al-Bukhari 683
665 Allah's Messenger (ﷺ) went to establish peace ... Sahih al-Bukhari 684

```

```

Cluster Pillar
0 1 Daily Prayers (Salat)
1 1 Daily Prayers (Salat)
2 1 Daily Prayers (Salat)
3 1 Daily Prayers (Salat)
4 1 Daily Prayers (Salat)
...
661 0 The Profession of Faith (the Shahada)
662 0 The Profession of Faith (the Shahada)
663 0 The Profession of Faith (the Shahada)
664 0 The Profession of Faith (the Shahada)
665 0 The Profession of Faith (the Shahada)

```

[666 rows x 7 columns]

In [9]:

```
df
```

Out[9]:

	Book Name	Collected By	Arabic	English	References	Cluster	Pillar
0	Revelation	Sahih al-Bukhari	حَدَّثَنَا الْحُمَيْدِيُّ عَبْدُ اللَّهِ بْنُ ...	I heard Allah's Messenger (ﷺ) saying, "Th...	Sahih al-Bukhari 1	1	Daily Prayers (Salat)
1	Revelation	Sahih al-Bukhari	حَدَّثَنَا عَبْدُ اللَّهِ بْنُ ... يُوسُفَ، قَالَ	\r\r (the mother of the faithful believers...	Sahih al-Bukhari 2	1	Daily Prayers (Salat)
2	Belief	Sahih al-Bukhari	حَدَّثَنَا يَحْيَى بْنُ بُكَيْرٍ، قَالَ حَدَّثَ ...	The commencement of the Divine Inspiration to ...	Sahih al-Bukhari 3	1	Daily Prayers (Salat)
3	Belief	Sahih al-Bukhari	قَالَ ابْنُ شِهَابٍ ... وَأَخْبَرَنِي أَبُو سَلَمَةَ	"While I was walking, all of a sudden I heard ...	Sahih al-Bukhari 4	1	Daily Prayers (Salat)
4	Knowledge	Sahih al-Bukhari	حَدَّثَنَا مُوسَى بْنُ ... إِسْمَاعِيلَ، قَالَ حَدَّثَ	\r\r Ibn 'Abbas in the explanation of the ...	Sahih al-Bukhari 5	1	Daily Prayers (Salat)
...
661	Interpretation of Dreams	Sunan Ibn Majah	حَدَّثَنَا أَبُو الْيَمَانِ، قَالَ ... أَخْبَرَنَا	Anas bin Malik Al-Ansari, told me, "Abu Bakr u...	Sahih al-Bukhari 680	0	The Profession of Faith (the Shahada)
662	Tribulations	Sunan Ibn Majah	حَدَّثَنَا أَبُو مَعْمَرٍ، قَالَ ... حَدَّثَنَا ع	The Prophet (ﷺ) did not come out for three day...	Sahih al-Bukhari 681	0	The Profession of Faith (the Shahada)
663	Tribulations	Sunan Ibn Majah	حَدَّثَنَا يَحْيَى بْنُ ... سُلَيْمَانَ، قَالَ حَدَّثَ	My father said, "When Allah's Messenger (ﷺ) be...	Sahih al-Bukhari 682	0	The Profession of Faith (the Shahada)
664	Zuhd	Sunan Ibn Majah	حَدَّثَنَا زَكَرِيَاءُ بْنُ يَحْيَى، ... قَالَ ع	`Aisha said, "Allah's Messenger (ﷺ) ordered Ab...	Sahih al-Bukhari 683	0	The Profession of Faith (the Shahada)
		Sunan	حَدَّثَنَا عَبْدُ اللَّهِ بْنُ يُوسُفَ، قَالَ	Allah's Messenger (ﷺ) went to establish peace ...	Sahih al-Bukhari 684		The Profession of Faith (the Shahada)

665	Zuhd	Sunan	عن أبي عبد الله بن	Allah's messenger (ﷺ), went	Sunan al	0	The Profession of
	Book Name	Collected	عنه، قال	to establish peace	Bukhari 684	Cluster	Faith (the Shahada)
		By	Arabic	English	References		Pillar

666 rows x 7 columns

In [10]:

```
columns_to_remove = ["English", "Cluster"]

data_df = df.drop(columns=columns_to_remove)
```

In [11]:

```
data_df.to_csv("Hadith(Mashood Bokhari).csv", index=False)
```

In []: