

In []:

```
In [3]: import pandas as pd
from bs4 import BeautifulSoup
import time
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.chrome.options import Options
import csv
import requests
```

```
In [4]: base_url = "https://www.clearquran.com/"
with open('quran_trans.csv', 'w', newline='', encoding='utf-8') as csvfile:
    writer = csv.writer(csvfile)
    writer.writerow(["Surah Number", "Verse"])
# Iterate over surahs from 1 to 114
    for surah_number in range(1, 115):
        surah_url = f"{base_url}{surah_number:03d}.html"
        print("Extracting from: ", surah_url)
        response = requests.get(surah_url)
        if response.status_code == 200:
            soup = BeautifulSoup(response.content, 'html.parser')
            quran_text_div = soup.find('div', {'id': 'quran-text'})
            verse_elements = quran_text_div.find_all('p')
            for verse_number, verse in enumerate(verse_elements, start=1):
                verse_text = verse.get_text(strip=True)
                writer.writerow([surah_number, verse_number, verse_text])
                print(f"Surah {surah_number}, Verse {verse_number} saved successfully")
            else:
                print(f"Error fetching website content from {surah_url}")
```

Extracting from: <https://www.clearquran.com/001.html> (<https://www.clearquran.com/001.html>)

Surah 1, Verse 1 saved successfully.
 Surah 1, Verse 2 saved successfully.
 Surah 1, Verse 3 saved successfully.
 Surah 1, Verse 4 saved successfully.
 Surah 1, Verse 5 saved successfully.
 Surah 1, Verse 6 saved successfully.
 Surah 1, Verse 7 saved successfully.

Extracting from: <https://www.clearquran.com/002.html> (<https://www.clearquran.com/002.html>)

Surah 2, Verse 1 saved successfully.
 Surah 2, Verse 2 saved successfully.
 Surah 2, Verse 3 saved successfully.
 Surah 2, Verse 4 saved successfully.
 Surah 2, Verse 5 saved successfully.
 Surah 2, Verse 6 saved successfully.
 Surah 2, Verse 7 saved successfully.
 Surah 2, Verse 8 saved successfully.

```
In [5]: df = pd.read_csv('quran_trans.csv')

pillar_keywords = {
    "Shahada": ["Praise", "testimony", "witness", "faith", "guidance"],
    "Salat": ["prayer", "Tahajudd", "worship"],
    "Zakat": ["charity", "give", "poor"],
    "Hajj": ["pilgrimage", "Mekkah", "Kaaba", "Hajj"]
}

def categorize_ayah(ayah_text):
    for pillar, keywords in pillar_keywords.items():
        if any(keyword in ayah_text.lower() for keyword in keywords):
            return pillar
    return "Other"

df['pillar'] = df['Verse'].apply(categorize_ayah)
df.to_csv('Quran_Verse.csv', index=False)
print(df.head(20))
```

	Surah	Number	Verse	pillar
1	1	1	In the name of God, the Gracious, the Merciful.	Other
1	2	2	Praise be to God, Lord of the Worlds.	Other
1	3	3	The Most Gracious, the Most Merciful.	Other
1	4	4	Master of the Day of Judgment.	Other
1	5	5	It is You we worship, and upon You we call f...	Salat
1	6	6	Guide us to the straight path.	Other
1	7	7	The path of those You have blessed, not of t...	Other
2	1		In the name of God, the Gracious, the Merciful.	Other
2	2		1.Alif, Lam, Meem.	Other
2	3	2	This is the Book in which there is no doubt,...	Other
2	4	3	Those who believe in the unseen, and perform...	Salat
2	5	4	And those who believe in what was revealed t...	Other
2	6	5	These are upon guidance from their Lord. The...	Shahada
2	7	6	As for those who disbelieve—it is the same f...	Other
2	8	7	God has set a seal on their hearts and on th...	Other
2	9	8	Among the people are those who say, “We beli...	Other
2	10	9	They seek to deceive God and those who belie...	Other
2	11	10	In their hearts is sickness, and God has in...	Other
2	12	11	And when it is said to them, “Do not make t...	Other
2	13	12	In fact, they are the troublemakers, but th...	Other

In []: