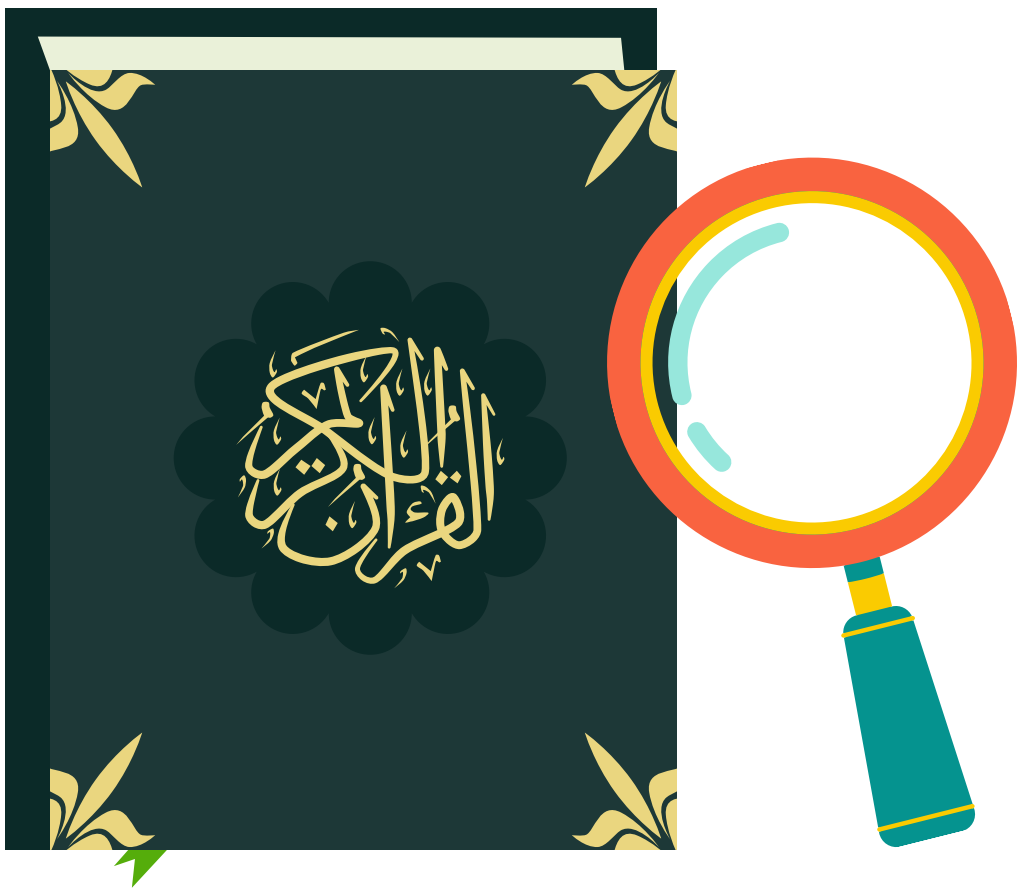


Exploratory Data Analysis

Quran Dataset (English version)

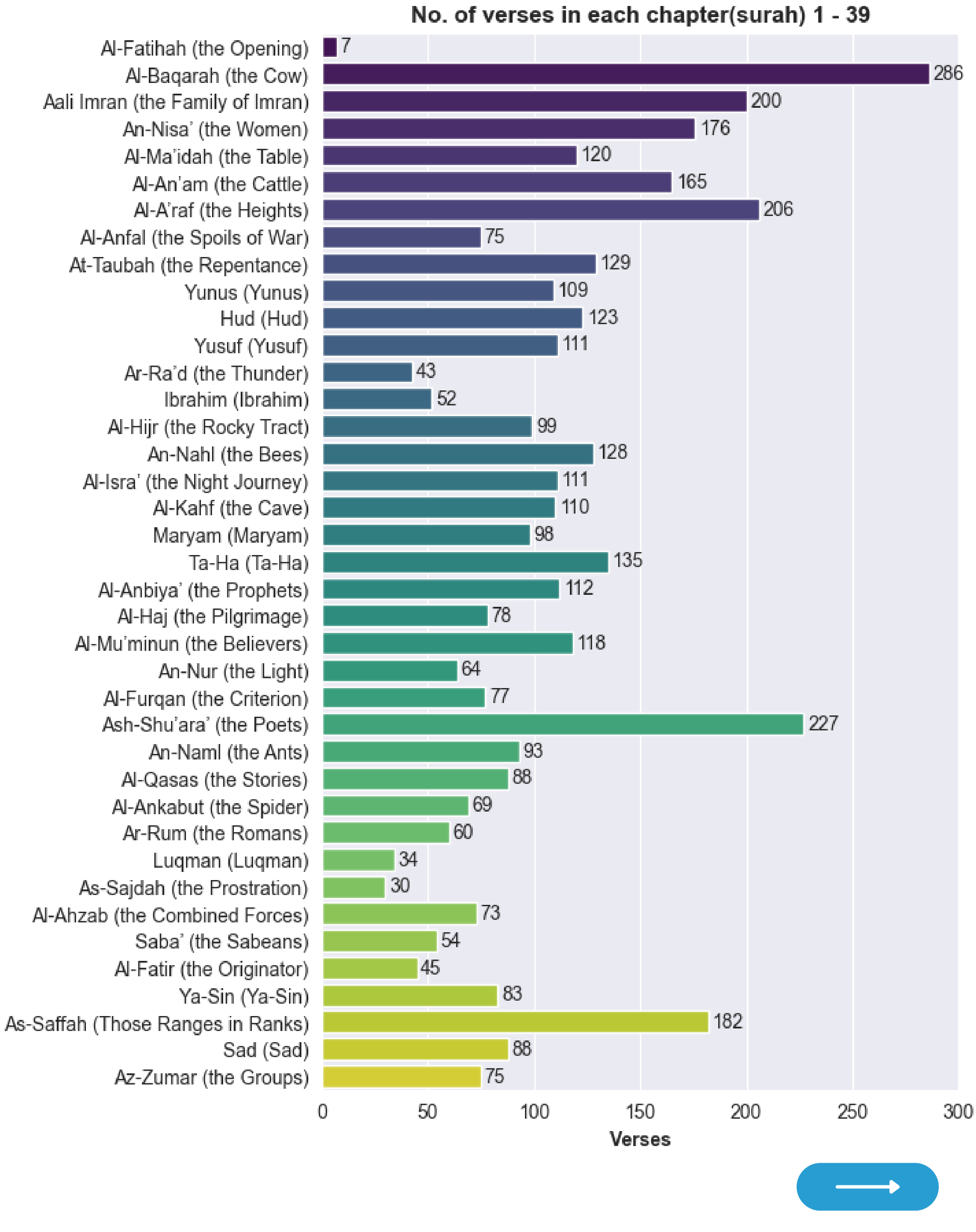
using pandas, matplotlib and seaborn

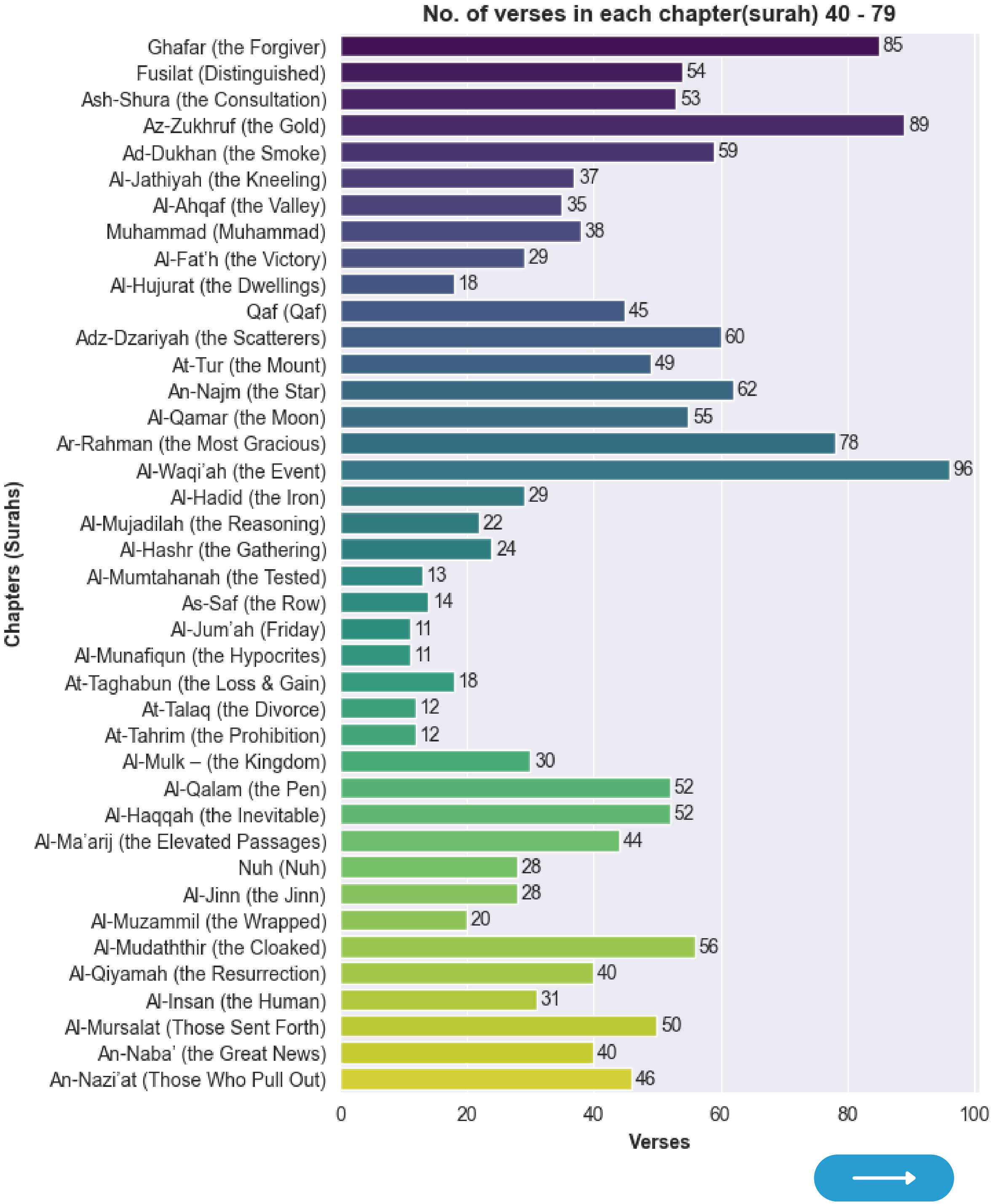
Dataset is available on **kaggle**

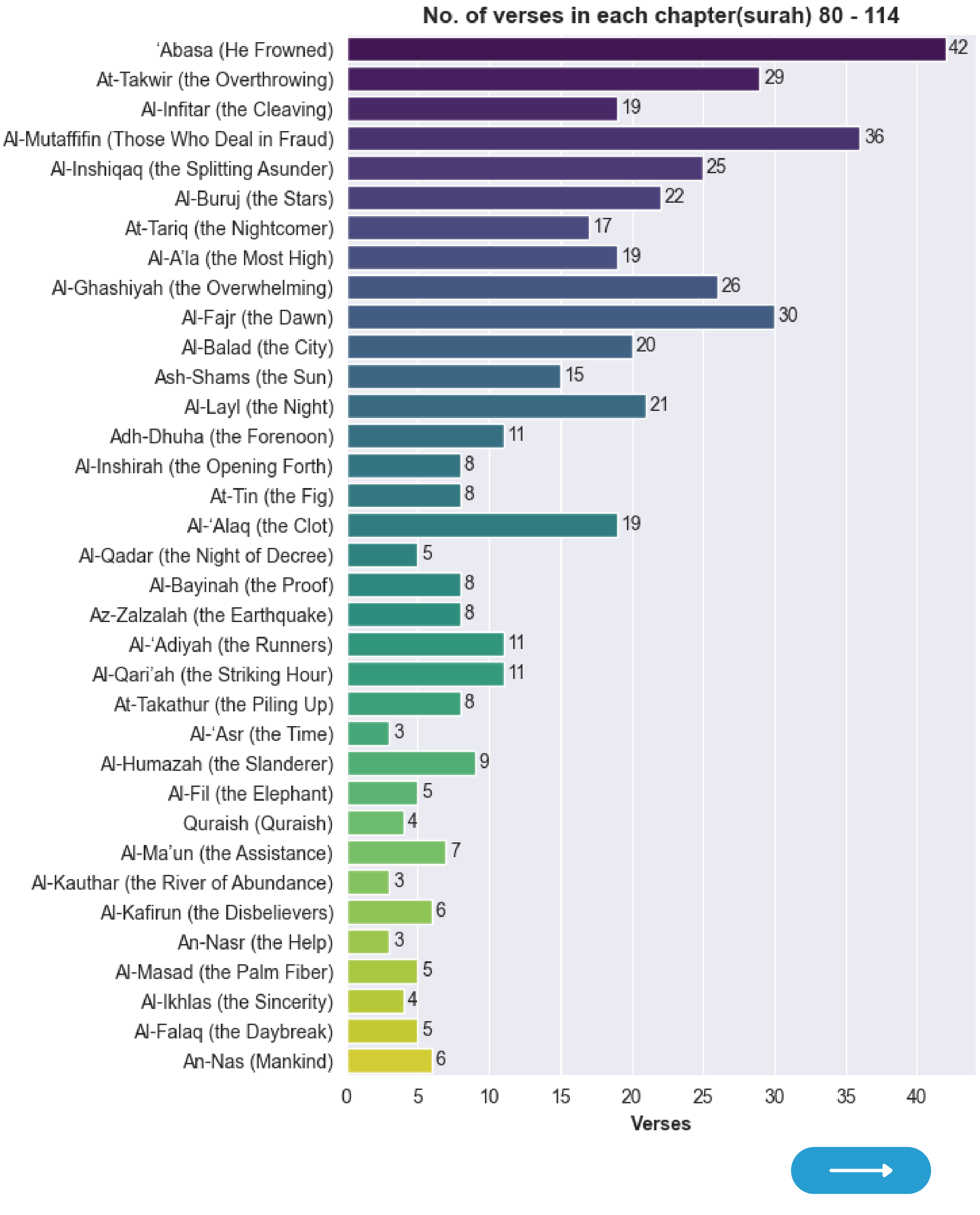
ali-bin-kashif

Number of verses in each chapter (Surah)?

```
# Function for making and displaying a count plot
def quran_chapters_verses_countplot(dataframe, y_axis, title):
    # Setting figures size and theme
    plt.figure(figsize=(6,10))
    sns.set_style('darkgrid')

    # Countplot to count the number of verses of each chapter and plot
    ax = sns.countplot(data=dataframe, y=y_axis, hue=y_axis, palette='viridis')

    # Setting labels for each bar
    for container in ax.containers:
        ax.bar_label(container, size=10, padding=2)

    # Customizing plot
    ax.set_title(title, fontweight='bold', fontsize=12)
    ax.set_ylabel('Chapters (Surahs)', fontweight='bold')
    ax.set_xlabel('Verses', fontweight='bold')

    plt.show()

quran_chapters_verses_countplot(surah_1_39, 'Surah Name', 'No. of verses in each chapter(surah) 1 - 39')
```


No. of verses in each chapter(surah) 1 - 39
Chapters (Surahs)
Al-Fatihah (the Opening) 7
Al-Baqarah (the Cow) 286
Aali Imran (the Family of Imran) 200
An-Nisa' (the Women) 176
Al-Ma'idah (the Table) 120
Al-An'am (the Cattle) 165
Al-A'raf (the Heights) 206
Al-Anfal (the Spoils of War) 75
At-Taubah (the Repentance) 129
Yunus (Yunus) 109
Hud (Hud) 123
Yusuf (Yusuf) 111
Ar-Ra'd (the Thunder) 43
Ibrahim (Ibrahim) 52
Al-Hijr (the Rocky Tract) 99
An-Nahl (the Bees) 128
Al-Isra' (the Night Journey) 111
Al-Kahf (the Cave) 110
Maryam (Maryam) 98
Ta-Ha (Ta-Ha) 135
Al-Anbiya' (the Prophets) 112
Al-Haj (the Pilgrimage) 78
Al-Mu'minun (the Believers) 118
An-Nur (the Light) 64
Al-Furqan (the Criterion) 77
Ash-Shu'ara' (the Poets) 227
An-Naml (the Ants) 93
Al-Qasas (the Stories) 88
Al-Ankabut (the Spider) 69
Ar-Rum (the Romans) 60
Luqman (Luqman) 34
As-Sajdah (the Prostration) 30
Al-Ahzab (the Combined Forces) 73
Saba' (the Sabeans) 54
Al-Fatir (the Originator) 45
Ya-Sin (Ya-Sin) 83
As-Saffah (Those Ranges in Ranks) 182
Sad (Sad) 88
Az-Zumar (the Groups) 75
0
50
100
150
200
250
300
Verses


No. of verses in each chapter(surah) 40 - 79
Chapters (Surahs)
Ghafar (the Forgiver) 85
Fusilat (Distinguished) 54
Ash-Shura (the Consultation) 53
Az-Zukhruf (the Gold) 89
Ad-Dukhan (the Smoke) 59
Al-Jathiyah (the Kneeling) 37
Al-Ahqaf (the Valley) 35
Muhammad (Muhammad) 38
Al-Fat'h (the Victory) 29
Al-Hujurat (the Dwellings) 18
Qaf (Qaf) 45
Adz-Dzariyah (the Scatterers) 60
At-Tur (the Mount) 49
An-Najm (the Star) 62
Al-Qamar (the Moon) 55
Ar-Rahman (the Most Gracious) 78
Al-Waqi'ah (the Event) 96
Al-Hadid (the Iron) 29
Al-Mujadilah (the Reasoning) 22
Al-Hashr (the Gathering) 24
Al-Mumtahanah (the Tested) 13
As-Saf (the Row) 14
Al-Jum'ah (Friday) 11
Al-Munafiqun (the Hypocrites) 11
At-Taghabun (the Loss & Gain) 18
At-Talaq (the Divorce) 12
At-Tahrim (the Prohibition) 12
Al-Mulk – (the Kingdom) 30
Al-Qalam (the Pen) 52
Al-Haqqah (the Inevitable) 52
Al-Ma'arij (the Elevated Passages) 44
Nuh (Nuh) 28
Al-Jinn (the Jinn) 28
Al-Muzammil (the Wrapped) 20
Al-Mudaththir (the Cloaked) 56
Al-Qiyamah (the Resurrection) 40
Al-Insan (the Human) 31
Al-Mursalat (Those Sent Forth) 50
An-Naba' (the Great News) 40
An-Nazi'at (Those Who Pull Out) 46
0
20
40
60
80
100
Verses


No. of verses in each chapter(surah) 80 - 114
Chapters (Surahs)
'Abasa (He Frowned) 42
At-Takwir (the Overthrowing) 29
Al-Infitar (the Cleaving) 19
Al-Mutaffifin (Those Who Deal in Fraud) 36
Al-Inshiqaq (the Splitting Asunder) 25
Al-Buruj (the Stars) 22
At-Tariq (the Nightcomer) 17
Al-A'la (the Most High) 19
Al-Ghashiyah (the Overwhelming) 26
Al-Fajr (the Dawn) 30
Al-Balad (the City) 20
Ash-Shams (the Sun) 15
Al-Layl (the Night) 21
Adh-Dhuha (the Forenoon) 11
Al-Inshirah (the Opening Forth) 8
At-Tin (the Fig) 8
Al-'Alaq (the Clot) 19
Al-Qadar (the Night of Decree) 5
Al-Bayinah (the Proof) 8
Az-Zalzalah (the Earthquake) 8
Al-'Adiyah (the Runners) 11
Al-Qari'ah (the Striking Hour) 11
At-Takathur (the Piling Up) 8
Al-'Asr (the Time) 3
Al-Humazah (the Slanderer) 9
Al-Fil (the Elephant) 5
Quraish (Quraish) 4
Al-Ma'un (the Assistance) 7
Al-Kauthar (the River of Abundance) 3
Al-Kafirun (the Disbelievers) 6
An-Nasr (the Help) 3
Al-Masad (the Palm Fiber) 5
Al-Ikhlas (the Sincerity) 4
Al-Falaq (the Daybreak) 5
An-Nas (Mankind) 6
0 5 10 15 20 25 30 35 40
Verses

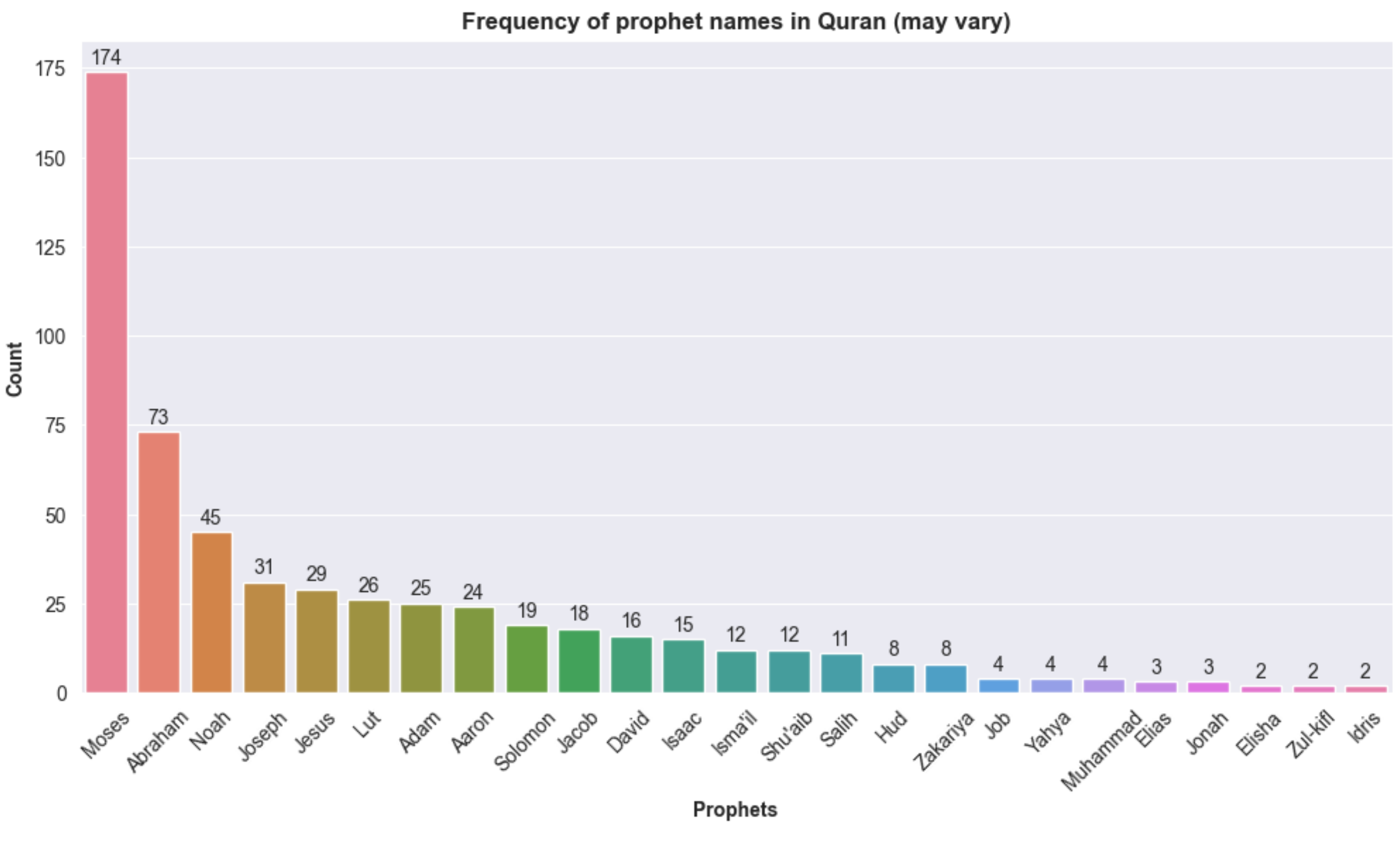
- **Quran** has **114 chapters** and **6,236 verses**, not including Bismillah.
- **Al-Baqarah** (The Cow) is the lengthiest chapter.
- **Al-Asr** (the Time) , **Al-Kauthar** (the River of Abundance) and **Al-Nasr**(the Help) are the shortest chapters with 3 verses only.

Prophets mentioned in the Quran and how many times?

```
# List of all the prophets mentioned in the Quran
prophet_names = [ "Isma'il", 'Elisha', 'Zul-Kifl', 'Jesus', 'Moses', "Shu'aib", 'Jacob', 'Lut',
                 'Joseph', 'Isaac', 'Job', 'Aaron', 'Abraham', 'Noah', 'Adam', 'Hud', 'Solomon',
                 'David', 'Zakariya', 'Yahya', 'Elias', 'Jonah', 'Idris', 'Salih', 'Muhammad']

# Making a empty dicitonary
prophet_names_freq = {}

# Initializing with all prophet names as key and 0 value(i.e count)
for prophet in prophet_names:
    prophet_names_freq[prophet.lower()] = 0

# Code for counting prophet names, start by looping in Quran dataframe
for lab,row in quran_english.iterrows():

    # Extract each verse and split the words and store in a list
    ayah_word_list = row['Text'].split(' ')

    # Loop through verse list to access each word
    for word in ayah_word_list:
        # Lower casing the word and remove any special characters from it through RE
        word = word.lower()
        pattern = re.compile('[^a-zA-Z-\']')
        word = pattern.sub('', word)

        # If the word is present in prophet names dict then increase count value
        if word in prophet_names_freq:
            prophet_names_freq[word] += 1

# Making a dataframe with the prophets dictionary
prophet_names_df = pd.DataFrame({ 'Name' : list(prophet_names_freq.keys()),
                                  'Frequency' : list(prophet_names_freq.values())})

prophet_names_df['Name'] = prophet_names_df['Name'].apply(str.capitalize)

prophet_names_df.sort_values('Frequency', ascending=False, inplace=True)
```

Prophets mentioned in the Quran and how many times?

```
# Plotting the graph
plt.figure(figsize=(12,6))
sns.set_style('darkgrid')
ax = sns.barplot(data=prophet_names_df, x='Name', y='Frequency', palette='husl')

for container in ax.containers:
    ax.bar_label(container, size=10, padding=2)

ax.set_title("Frequency of prophet names in Quran (may vary)", fontweight='bold', fontsize=12)
ax.set_ylabel("Count", fontweight='bold')
ax.set_xlabel("Prophets", fontweight='bold')
ax.tick_params('x', rotation=45)

plt.show()
```



- **Moses (Musa A.S)** is mentioned the most in Quran followed by **Abraham**, **Noah** and **Joseph(Yusuf) A.S**.
- These frequencies may vary from translation to translation of differents languages and Mufasirs.

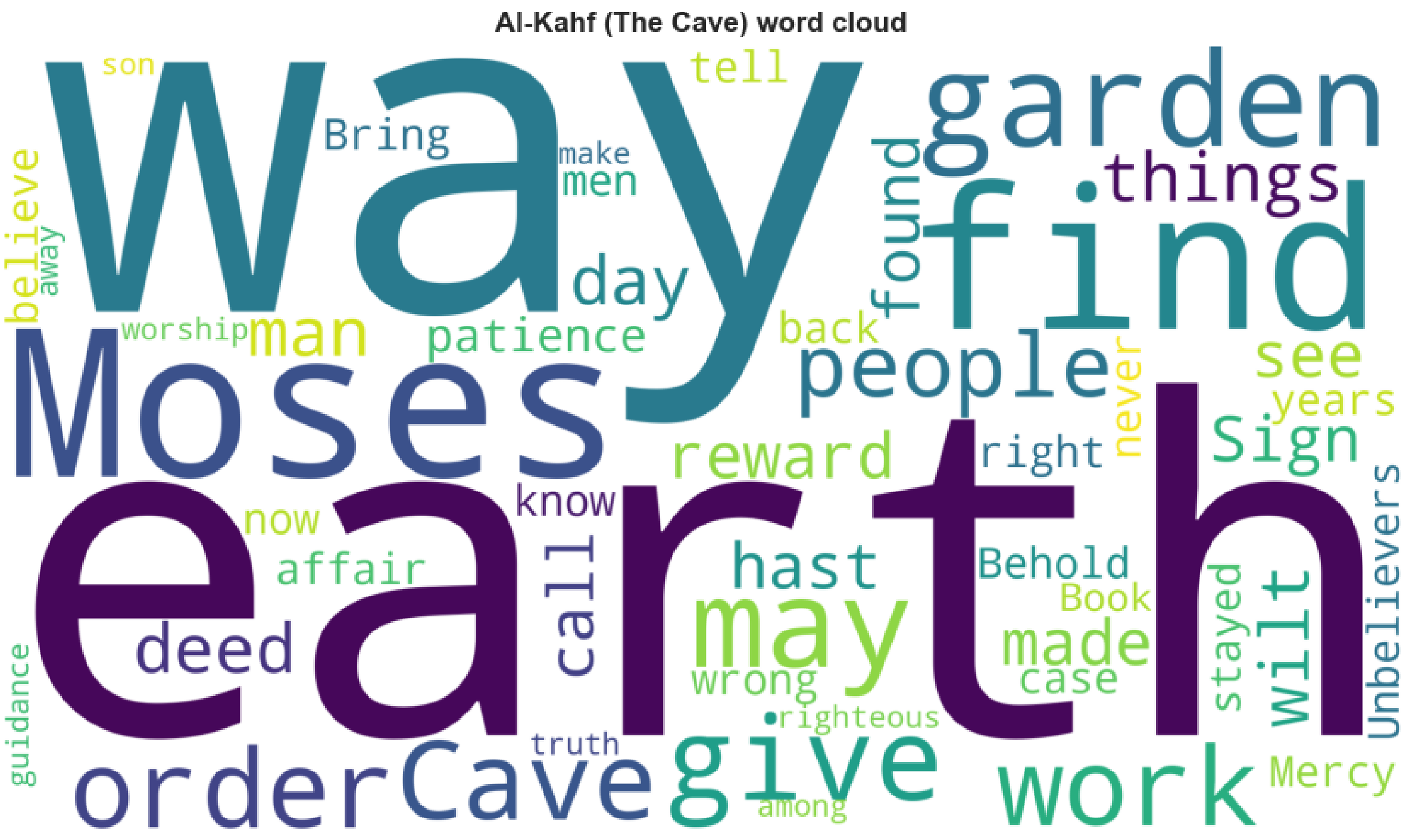
Word cloud of Quran and different chapters.

- Word clouds or tag clouds are graphical representations of word frequency that give greater prominence to words that appear more frequently in a source text. The larger the word in the visual the more common the word was in the document.

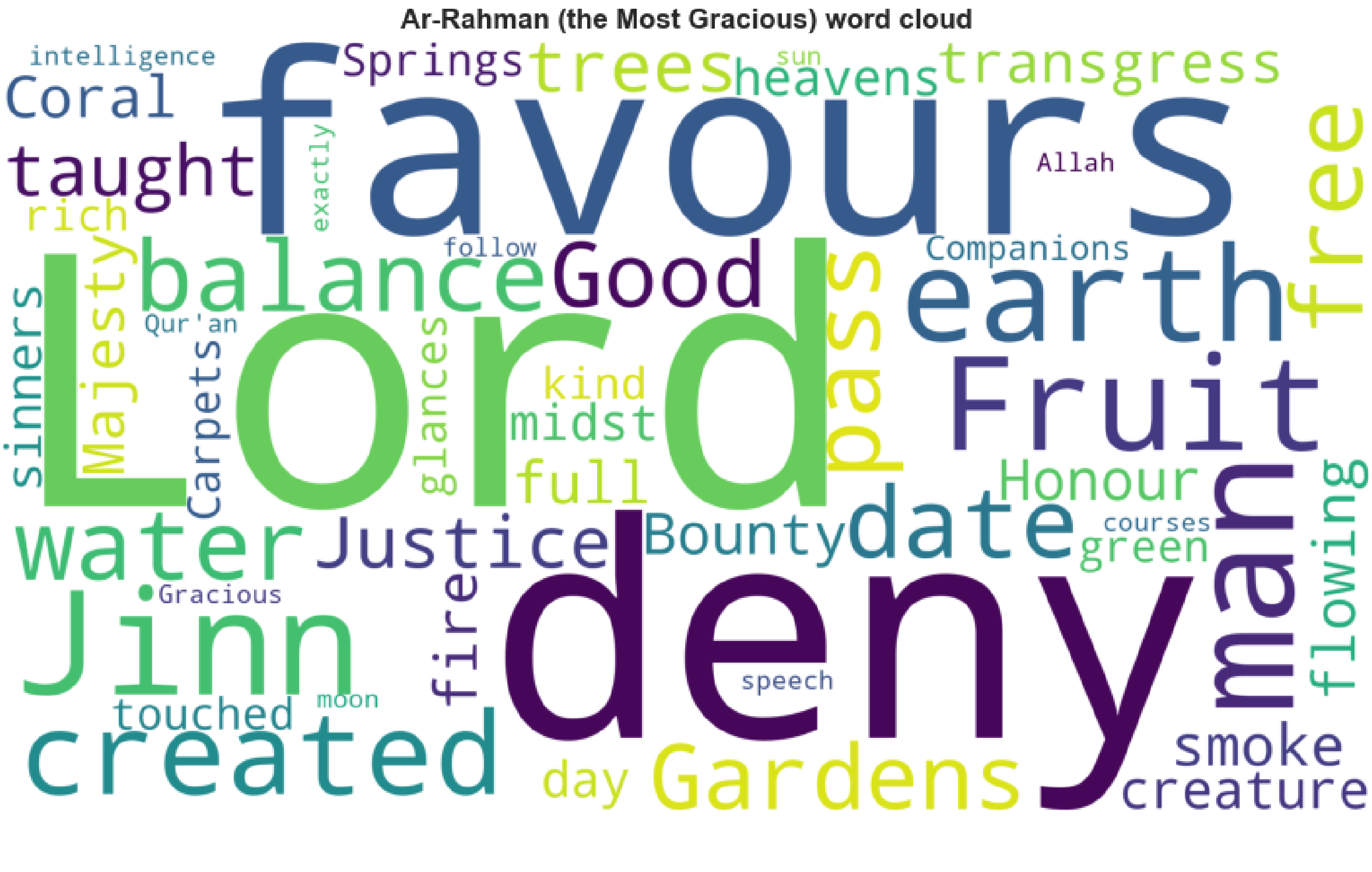
```
# Function for generating word cloud, it takes, a string, stopwords list, title for the image, maxwords to show
def word_cloud_generator(text_corpus, stopwords, title, maxwords):
    word_cloud = WordCloud( background_color='white', stopwords=stopwords, height=1080, width=1920, max_words=maxwords)
    word_cloud.generate(text_corpus)

    plt.figure(figsize=( 12,6))
    plt.imshow(word_cloud)
    plt.title(title, fontweight='bold', fontsize=12)
    plt.axis('off')
    plt.show()

# Collected some stopwords from the English Quran and updating it in STOPWORDS set
quranic_stop_words = ['ye', 'verily', 'will', 'said', 'say', 'us', 'thy', 'thee', 'O', 'except', 'Nay',
                      'thou' ,'hath', 'Thus', 'none', 'therein', 'come', 'came', 'even', 'two', 'word',
                      'every', 'let', 'thing', 'with', 'whose', 'forth', 'wouldst', 'set', 'unto']
stopwords = STOPWORDS.update(quranic_stop_words)


# Extracting each verse and storing them in a string
quran_text = ""
for lab,row in quran_english.iterrows():
    quran_text += row['Text']

word_cloud_generator(quran_text, stopwords, 'Quran Word Cloud', 100)
```

- You can see the words which are more frequent and have more emphasis.

```python
# Subsetting 18 Surah from the Quran dataframe
surah_kahf = quran_english_with_surah[quran_english_with_surah['Surah'] == 18]

# Concating each verse in a string
surah_kahf_text = ""
for lab,row in surah_kahf.iterrows():
    surah_kahf_text += row['Text']

stopwords2 = list(STOPWORDS) + ['Allah', 'Lord', 'one', 'best']
word_cloud_generator(surah_kahf_text, stopwords2, 'Al-Kahf (The Cave) word cloud', 50)
```



```python
surah_takasur = quran_english_with_surah[quran_english_with_surah['Surah'] == 102]

surah_takasur_text = ""
for lab,row in surah_takasur.iterrows():
    surah_takasur_text += row['Text']

word_cloud_generator(surah_takasur_text, stopwords, 'Al-Takasur (The piling up) word cloud', 50)
```



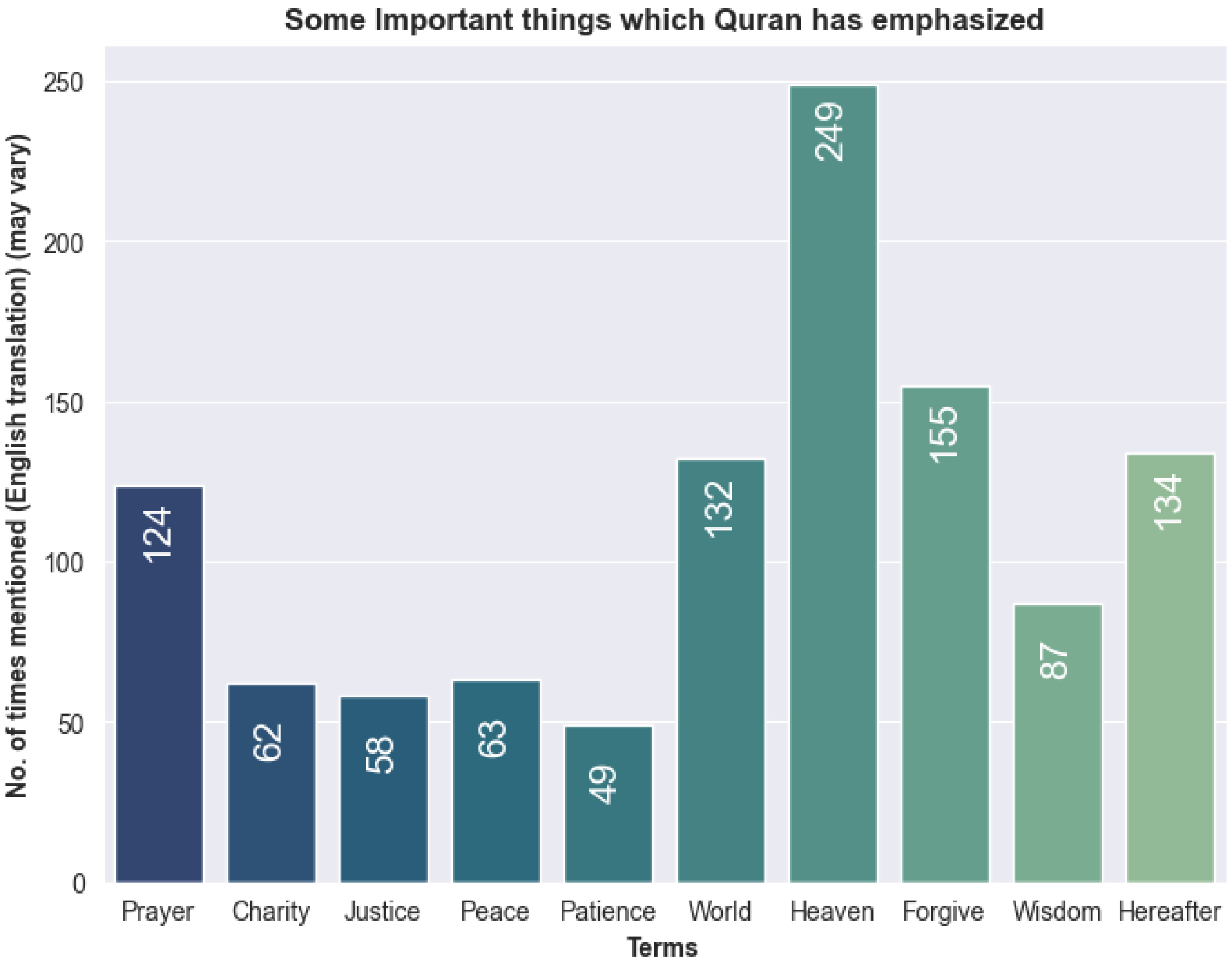
- The theme of Surah Ar-Rahman is established in the recurring statement, **“So which of the favours of your Lord would you deny?”**

Important things which Quran has emphasized on.

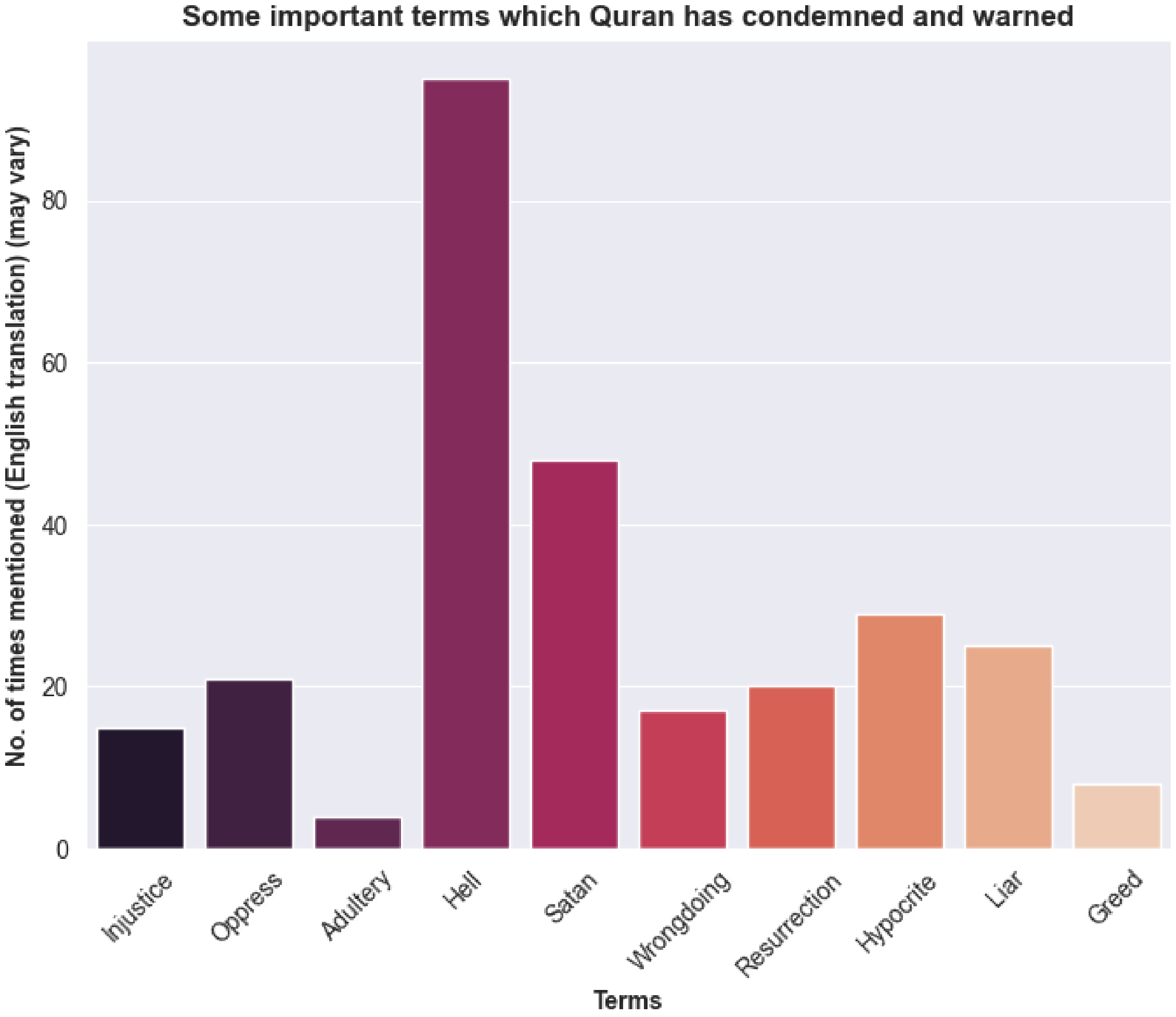
```
# List of terms
terms_list = [ 'prayer', 'charity', 'justice', 'peace', 'patience', 'world', 'heaven', 'forgive', 'wisdom', 'hereafter']

# Making a dictionary of terms with value 0 initially
terms_list_freq = {}
for term in terms_list:
    terms_list_freq[term] = 0

# Code for counting the matched words and incrementing the count
for lab,row in quran_english.iterrows():
    ayat_word_list = row['Text'].split(' ')
    for word in ayat_word_list:
        word = word.lower()
        pattern = re.compile('[^a-zA-Z]')
        word = pattern.sub('', word)
        for term in terms_list_freq:
            if word == term:
                terms_list_freq[word] += 1
            elif word.__contains__(term):
                terms_list_freq[term] += 1

# Make a dataframe from the dictionary
terms_df = pd.DataFrame({'Term' : terms_list_freq.keys(),
                         'Freq' : terms_list_freq.values()})

# Capitialize all term names
terms_df['Term'] = terms_df['Term'].apply(str.capitalize)

```

Important things which Quran has emphasized on.

```python
# Plotting the graph
plt.figure(figsize=(8,6))
sns.set_style('darkgrid')
ax = sns.barplot(data=terms_df, x='Term', y='Freq',palette='crest_r')

for container in ax.containers:
    ax.bar_label(container, size=15, padding=-30, color='white', rotation=90)

ax.set_title("Some Important things which Quran has emphasized", fontweight='bold', fontsize=12)
ax.set_ylabel("No. of times mentioned (English translation) (may vary)", fontweight='bold')
ax.set_xlabel("Terms", fontweight='bold')

plt.show()
```



- Quran has emphasized on the topics which are necessary for the betterment of the society and one's life.



- There are many acts and harmful deeds which Quran has condemned and warned about their consequences.

Thank you, thats all uptill now!