1-Values in the “score” column follows a normal distribution, and there is no skewness present in the distribution.

2-The code "df.isna().sum()" appears to be used to check for missing values (NaN) in the DataFrame "df." The output indicates no missing values , This means that there are no missing values in any of the columns in the DataFrame, as the count of missing values for each column is 0. All columns contain complete data.

3- The output shows the number of occurrences for each unique topic:

- "sport": 1521 occurrences

- "medias": 1509 occurrences

- "art-et-culture": 1378 occurrences

This information indicates the distribution of topics in the DataFrame. The "sport" topic is the most common with 1521 occurrences, followed by "medias" with 1509 occurrences, and "art-et-culture" with 1378 occurrences.

4-The provided data shows the top 3-grams (sequences of three words) in the entire corpus along with their respective frequencies. Here are the top 10 3-grams:

1. ('ان', 'شاء', 'الله'): 52 occurrences

2. ('لله', 'وانا', 'اليه'): 26 occurrences

3. ('لله', 'وإنا', 'إليه'): 24 occurrences

4. ('وانا', 'اليه', 'راجعون'): 23 occurrences

5. ('قوة', 'الا', 'بالله'): 22 occurrences

6. ('وإنا', 'إليه', 'راجعون'): 22 occurrences

7. ('الله', 'ونعم', 'الوكيل'): 22 occurrences

8. ('حول', 'قوة', 'الا'): 17 occurrences

9. ('n', 'a', 'pas'): 17 occurrences

10. ('انا', 'لله', 'وانا'): 16 occurrences

From this insight, we can observe the following:

1. The most frequent 3-gram in the corpus is "ان شاء الله with a frequency of 52 occurrences.

2. The second and third most common 3-grams are variations of "لله وانا اليه" , with frequencies of 26 and 24 occurrences, respectively.

3. The phrase "وانا اليه راجعون" appears frequently with 23 occurrences.

Overall, the insight suggests that the corpus contains content with a strong religious context, It's important to note that without additional context or knowledge of the corpus content, the significance of these 3-grams may vary, and further analysis may be needed to understand the underlying themes or topics within the text.

4- Most Common Words per Topic:

Topic: art-et-culture

ان: 382

الله: 354

المغرب: 181

او: 174

الى: 155

de: 125

et: 119

كانت: 115

المغاربة: 113

العربية: 101

الفن: 101

المغربي: 95

البخاري: 87

يجب: 87

المغربية: 85

الا: 85

فقط: 82

الكتاب: 77

le: 76

la: 75

Topic: medias

ان: 457

الله: 349

المغرب: 263

الى: 204

او: 180

de: 171

يجب: 139

الشعب: 123

الجنسية: 119

et: 115

الدولة: 105

التربية: 104

la: 104

le: 103

المغاربة: 97

المغربية: 94

les: 89

فقط: 87

الا: 82

المغربي: 77

Topic: sport

الله: 580

ان: 406

الرجاء: 231

الوداد: 190

فريق: 137

الى: 128

او: 119

الفريق: 116

مبروك: 114

المغرب: 113

de: 96

يجب: 95

كورونا: 95

لاعب: 94

المدرب: 89

حمد: 89

et: 84

المغاربة: 84

le: 84

الكرة: 83

From the most common words in each topic, we can derive some insights:

1. \*\*Religious Context:The word "الله" appears prominently in all three topics, indicating a strong religious context within the corpus.

2. \*\*Cultural Focus:The topics "art-et-culture" and "medias" include words related to Moroccan culture, such as "المغرب", "المغاربة", and "العربية".

3. \*\*Sports Terminology:The "sport" topic contains words related to football teams like "الرجاء" and "الوداد", as well as terms like "فريق", "لاعب" , and "المدرب".

4. \*\*Common Words Across Topics: Certain words like”ان” and "الى" appear frequently in all three topics, suggesting they might be common connectors or conjunctions used in various contexts.