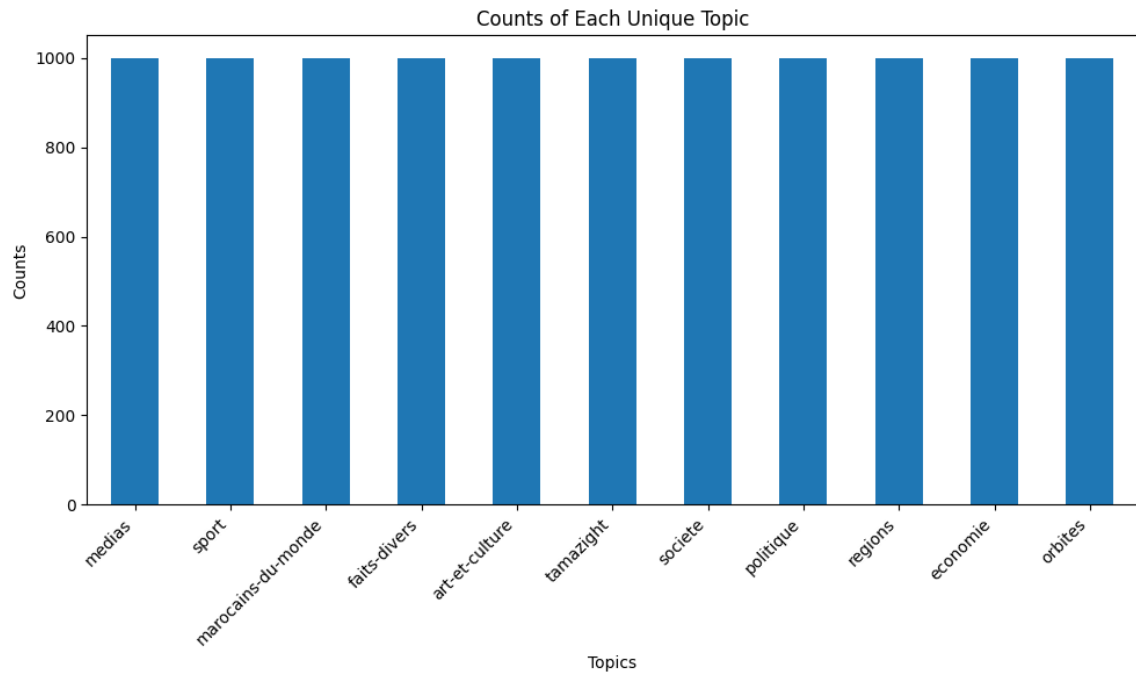


Insights about Task 1

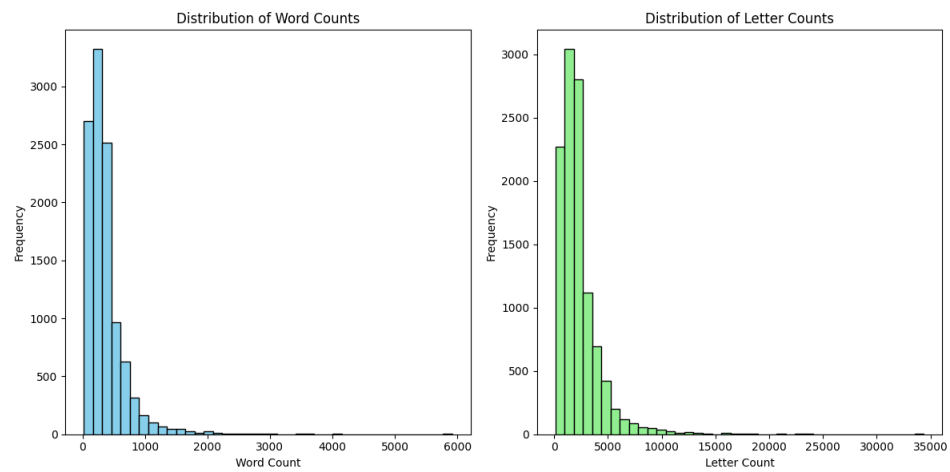
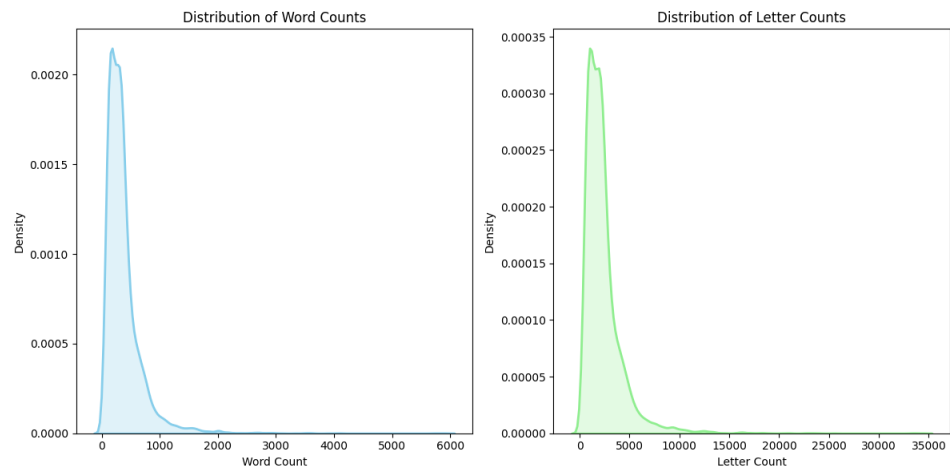
1. Number of Examples for each class says that the **data is balanced**



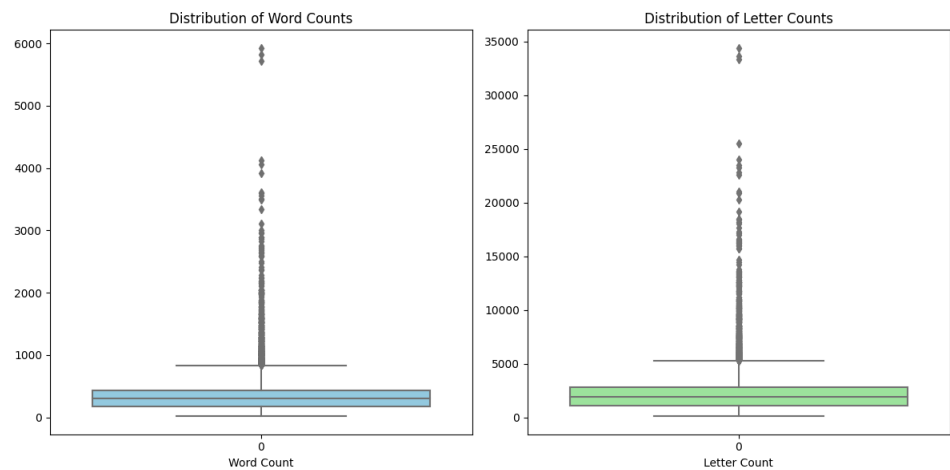
```
Number of values per class :  
medias 1000  
sport 1000  
marocains-du-monde 1000  
faits-divers 1000  
art-et-culture 1000  
tamazight 1000  
societe 1000  
politique 999  
regions 999  
economie 999  
orbites 999
```

2. The length of examples in words and letters

a.



b.



c.

d. As you can see the distribution is normally distribution and the data is a little skewed to the right. For the word counts it has the

same frequency distribution in the area from 150 near to 500 length of words. Also for the letters they have the same distribution.

- e. There are also some outliers that maybe affect on the model so that we need to normalize the data
- f. In the box plot there are some statistics:

```
Summary Statistics for Word Counts:
Median: 298.0
Q1: 173.0
Q3: 439.0
IQR: 266.0
Lower Whisker: -226.0
Upper Whisker: 838.0

Summary Statistics for Letter Counts:
Median: 1889.0
Q1: 1101.0
Q3: 2780.0
IQR: 1679.0
Lower Whisker: -1417.5
Upper Whisker: 5298.5
```

- i.
- ii. Now most of the data are in the area between Q1&Q3

3. Top frequent 5-grams generally and per class

a. Generally

```
Top 10 Frequent N-grams Overall:
0.001195      في تصريح لجريدة هسبريس الإلكترونية
0.000897      تحت إشراف النيابة العامة المختصة
0.000877      تصريح لجريدة هسبريس الإلكترونية أن
0.000811      الحراسة النظرية رهن إشارة البحث
0.000694      الشؤون الخارجية والتعاون الإفريقي والمغاربة
0.000694      الخارجية والتعاون الإفريقي والمغاربة المقيمين
0.000681      تحت تدبير الحراسة النظرية رهن
0.000675      تدبير الحراسة النظرية رهن إشارة
0.000667      والتعاون الإفريقي والمغاربة المقيمين بالخارج
0.000648      بفيروس كورونا المستجد خلال الـ24
dtype: float64
```

- i.
- b. Per Class

Top 10 Frequent N-grams Per Class:
Topic: art-et-culture
0.001356 محمد السادس للفن الحديث والمعاصر
0.001049 في تصريح لجريدة هسبريس الإلكترونية
0.000930 أكاديمية المملكة المغربية للتعاون الثقافي
0.000904 الثقافة والشباب والرياضة قطاع الثقافة
0.000862 النقابة المغربية لمهنيي الفنون الدرامية
0.000752 مؤسسة أكاديمية المملكة المغربية للتعاون
0.000700 متحف محمد السادس للفن الحديث
0.000698 معرض كنوز الإسلام في إفريقيا
0.000694 الثقافة والاتصال والرياضة والمؤسسة الوطنية
0.000694 العالم العربي بباريس ووزارة الثقافة
Name: art-et-culture, dtype: float64

Topic: economie
0.002579 وزير الاقتصاد والمالية وإصلاح الإدارة
0.002218 درهم على شكل تسبيقات لمدة
0.002143 وزارة الاقتصاد والمالية وإصلاح الإدارة
0.001960 مليار درهم على شكل تسبيقات
0.001838 درهم في الفترة نفسها من
0.001723 في المائة من الناتج الداخلي
0.001694 الصناعة والتجارة والاقتصاد الأخضر والرقمي
0.001498 على شكل تسبيقات لمدة أيام
0.001408 الفصل الأول من السنة الجارية
0.001381 البحري والتنمية القروية والمياه والغابات
Name: economie, dtype: float64

i.

Topic: faits-divers
0.009345 تحت إشراف النيابة العامة المختصة
0.008810 الحراسة النظرية رهن إشارة البحث
0.007422 تحت تدبير الحراسة النظرية رهن
0.007356 تدبير الحراسة النظرية رهن إشارة
0.005759 الذي تشرف عليه النيابة العامة
0.005564 تشرف عليه النيابة العامة المختصة
0.005155 النظرية رهن إشارة البحث التمهيدي
0.004799 رهن إشارة البحث التمهيدي الذي
0.004763 للكشف عن جميع ظروف وملابسات
0.004541 بلاغ للمديرية العامة للأمن الوطني
Name: faits-divers, dtype: float64

Topic: marocains-du-monde
0.003095 بالمغاربة المقيمين بالخارج وشؤون الهجرة
0.002804 المنتدبة المكلفة بالمغاربة المقيمين بالخارج
0.002367 المكلفة بالمغاربة المقيمين بالخارج وشؤون
0.002278 الوزارة المنتدبة المكلفة بالمغاربة المقيمين
0.002035 الخارجية والتعاون الإفريقي والمغاربة المقيمين
0.001974 والتعاون الإفريقي والمغاربة المقيمين بالخارج
0.001972 الشؤون الخارجية والتعاون الإفريقي والمغاربة
0.001575 لدى وزير الشؤون الخارجية والتعاون
0.001346 في تصريح لجريدة هسبريس الإلكترونية
0.001325 وزير الشؤون الخارجية والتعاون الإفريقي
Name: marocains-du-monde, dtype: float64

ii.

```

Topic: medias
0.001951    المساء التي ورد بها أن
0.001658    من المساء التي ورد بها
0.001497    من المساء التي نشرت أن
0.001397    ونسبة إلى مصادر الجريدة فإن
0.001382    قراءة مواد بعض الجرائد الورقية
0.001345    مواد بعض الجرائد الورقية الخاصة
0.001331    بعض الجرائد الورقية الخاصة بيوم
0.001217    ونقرأ في المساء كذلك أن
0.001123    رئيس الحكومة سعد الدين العثماني
0.001040    الأحداث المغربية التي نشرت أن
Name: medias, dtype: float64

```

```

Topic: orbites
0.002083    النبي صلى الله عليه وسلم
0.001336    بعث الملك محمد السادس برقية
0.001205    لقضاء المقاومين وأعضاء جيش التحرير
0.001096    الشؤون الخارجية والتعاون الإفريقي والمغاربة
0.001079    والتعاون الإفريقي والمغاربة المقيمين بالخارج
0.001070    الخارجية والتعاون الإفريقي والمغاربة المقيمين
0.001045    في تصريح لجريدة هسبريس الإلكترونية
0.001021    وزير الشؤون الخارجية والتعاون الإفريقي
0.000933    الملك محمد السادس برقية تهنئة
0.000842    السامية لقضاء المقاومين وأعضاء جيش
Name: orbites, dtype: float64

```

iii.

```

Topic: politique
0.004171    الشؤون الخارجية والتعاون الإفريقي والمغاربة
0.004090    الخارجية والتعاون الإفريقي والمغاربة المقيمين
0.003908    والتعاون الإفريقي والمغاربة المقيمين بالخارج
0.003716    وزير الشؤون الخارجية والتعاون الإفريقي
0.002166    بوريطة وزير الشؤون الخارجية والتعاون
0.002164    رئيس الحكومة سعد الدين العثماني
0.002160    سعد الدين العثماني رئيس الحكومة
0.002048    ناصر بوريطة وزير الشؤون الخارجية
0.001668    الحبيب المالكي رئيس مجلس النواب
0.001665    الأمين العام لحزب الأصالة والمعاصرة
Name: politique, dtype: float64

```

```

Topic: regions
0.004210    إصابة جديدة بفيروس كورونا المستجد
0.003137    المستبعدة بعد تحليل مخبري سلبي
0.003103    الحالات المستبعدة بعد تحليل مخبري
0.002409    عند الساعة السادسة من مساء
0.002313    الحالات النشطة التي تتلقى العلاج
0.002090    حالة إصابة جديدة بفيروس كورونا
0.002085    حالة بإقليم الفقيه بن صالح
0.002003    إجمالي الحالات المستبعدة بعد تحليل
0.001926    مجموع الحالات النشطة التي تتلقى
0.001879    النشطة التي تتلقى العلاج بالجهة
Name: regions, dtype: float64

```

iv.

Topic: societe
0.006674 الصحية والانتخراط في التدابير الاحترازية
0.006674 النظافة والسلامة الصحية والانتخراط في
0.006674 بقواعد النظافة والسلامة الصحية والانتخراط
0.006674 والسلامة الصحية والانتخراط في التدابير
0.006662 في التدابير الاحترازية التي اتخذتها
0.006662 والانتخراط في التدابير الاحترازية التي
0.006651 اتخذتها السلطات المغربية بكل وطنية
0.006651 الاحترازية التي اتخذتها السلطات المغربية
0.006651 الالتزام بقواعد النظافة والسلامة الصحية
0.006651 التي اتخذتها السلطات المغربية بكل
Name: societe, dtype: float64

Topic: sport
0.004631 الجامعة الملكية المغربية لكرة القدم
0.002213 للجامعة الملكية المغربية لكرة القدم
0.001502 العصبة الوطنية لكرة القدم الاحترافية
0.001239 البالغ من العمر 21 سنة
0.001222 البالغ من العمر 22 سنة
0.001197 رئيس الجامعة الملكية المغربية لكرة
0.001140 التابعة للجامعة الملكية المغربية لكرة
0.001081 الانتقالات الصيفية المقبلة وحسب مصادر
0.001013 بسبب تفشي فيروس كورونا المستجد
0.000980 فوزي لقجع رئيس الجامعة الملكية
Name: sport, dtype: float64

Topic: tamazight
0.003055 الشبكة الأمازيغية من أجل المواطننة
0.002781 القانون التنظيمي لتفعيل الطابع الرسمي
0.002772 التنظيمي لتفعيل الطابع الرسمي للأمازيغية
0.002488 عميد المعهد الملكي للثقافة الأمازيغية
0.002368 المجلس الوطني للغات والثقافة المغربية
0.001971 مراحل تفعيل الطابع الرسمي للأمازيغية
0.001893 في تصريح لجريدة هسبريس الإلكترونية
0.001836 التنظيمي المتعلق بتفعيل الطابع الرسمي
0.001803 القانون التنظيمي المتعلق بتفعيل الطابع
0.001704 المتعلق بتفعيل الطابع الرسمي للأمازيغية
Name: tamazight, dtype: float64

v.