



REPORT ON SOCIAL MEDIA TWEETS



About Dataset

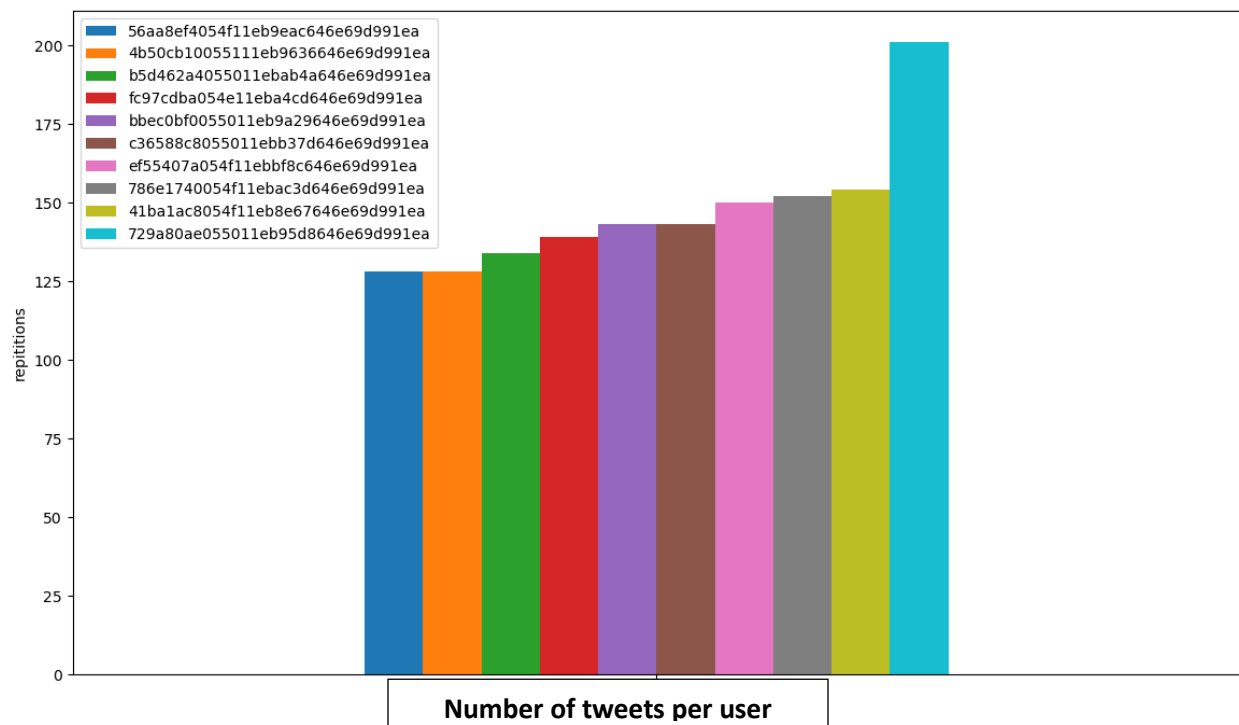
The dataset is social media posts that are generated by different users, each has a specific id related to the platform.

Analysis Of the Data

1 – Number of examples per class

```
max_sorted_num_of_data_per_calss
```

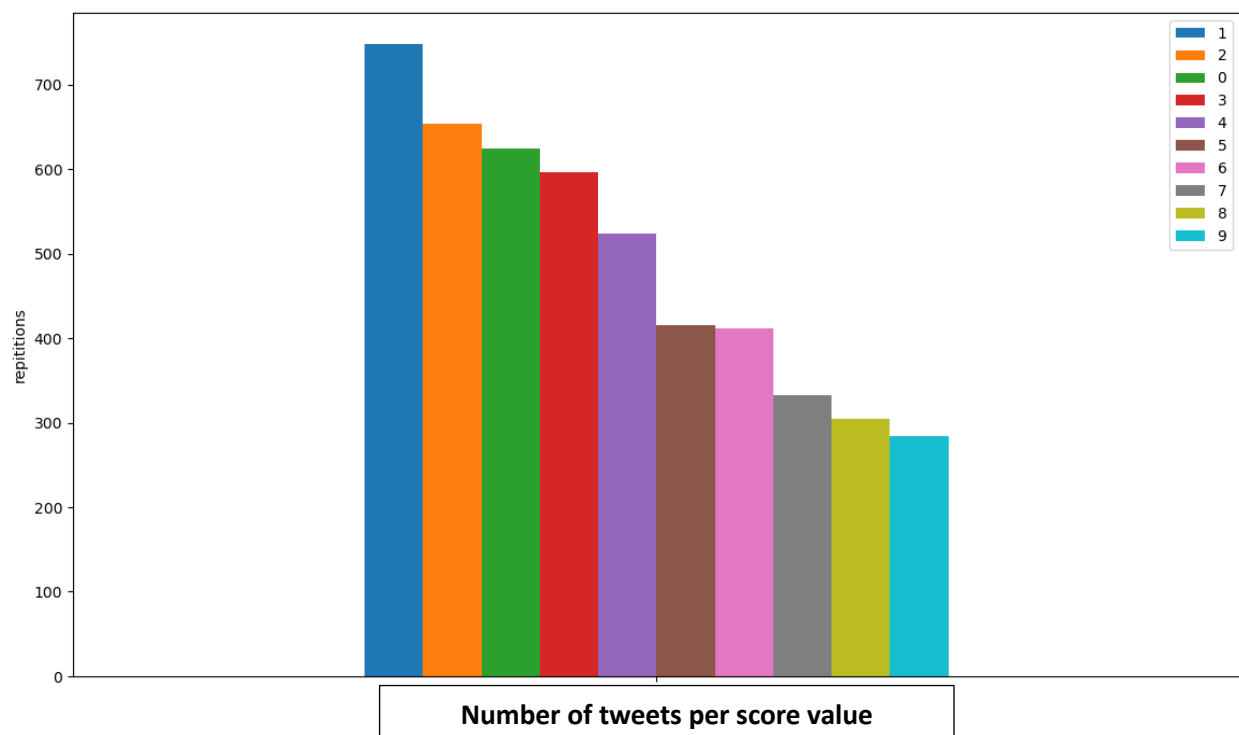
```
{'56aa8ef4054f11eb9eac646e69d991ea': 128,  
'4b50cb10055111eb9636646e69d991ea': 128,  
'b5d462a4055011ebab4a646e69d991ea': 134,  
'fc97cdba054e11eba4cd646e69d991ea': 139,  
'bbec0bf0055011eb9a29646e69d991ea': 143,  
'c36588c8055011ebb37d646e69d991ea': 143,  
'ef55407a054f11ebbf8c646e69d991ea': 150,  
'786e1740054f11ebac3d646e69d991ea': 152,  
'41ba1ac8054f11eb8e67646e69d991ea': 154,  
'729a80ae055011eb95d8646e69d991ea': 201}
```



The data and the bar plot above show the most active postId(s) (persons tweets on social media), as each Id corresponds to one of the application users.

```
dict(itertools.islice(df.score.value_counts().items(), 10))
```

```
{1: 748,  
 2: 654,  
 0: 625,  
 3: 597,  
 4: 524,  
 5: 416,  
 6: 411,  
 7: 333,  
 8: 305,  
 9: 284}
```



The plot and the data above show the number of tweets for the top 10 scores

2 – Top frequent n-grams generally and per class

```
bigram = most_frequent_general_ngrams(df, "comment", 2)
```

```
for val in bigram[:5]:  
    print(val)
```

```
[(1, ('علينا', 'الاهتمام'))]  
[(2, ('ما', 'نعرف'))]  
[(1, ('مبادرات', 'مهمة'))]  
[(1, ('شكرا', 'لمكتب'))]  
[(1, ('كل', 'من'))]
```

```
trigram = most_frequent_general_ngrams(df, "comment", 3)
```

```
for val in trigram[:5]:  
    print(val)
```

```
[(1, ('علينا', 'الاهتمام', 'ببيت'))]  
[(2, ('ما', 'نعرف', 'إذا'))]  
[(1, ('مبادرات', 'مهمة', 'يقوم'))]  
[(1, ('شكرا', 'لمكتب', 'بيت'))]  
[(1, ('كل', 'من', 'عليها'))]
```

```
quadgram = most_frequent_general_ngrams(df, "comment", 4)
```

```
for val in quadgram[:5]:  
    print(val)
```

```
[(1, ('علينا', 'الاهتمام', 'ببيت', 'الابحاث'))]  
[(1, ('بيت', 'الشعر', 'في', 'المغرب'))]  
[(1, ('مبادرات', 'مهمة', 'يقوم', 'بها'))]  
[(1, ('شكرا', 'لمكتب', 'بيت', 'الشعر'))]  
[(1, ('كل', 'من', 'عليها', 'فان'))]
```

```
tengram = most_frequent_general_ngrams(df, "comment", 10)
```

```
for val in tengram[:5]:  
    print(val)
```

```
[(1, ('علينا', 'الاهتمام', 'ببيت', 'الابحاث', 'العلمية', 'و', 'المختبرات', 'بماذا', 'نفعلنا', 'الشعر'))]  
[(1, ('بيت', 'الشعر', 'في', 'المغرب', 'يحظى', 'بالتقدير', 'داخل', 'وخارج', 'أرض', 'الوطن'))]  
[(1, ('مبادرات', 'مهمة', 'يقوم', 'بها', 'بيت', 'الشعر', 'الصون', 'الذاكرة', 'الثقافية', 'ببلادنا'))]  
[(1, ('شكرا', 'لمكتب', 'بيت', 'الشعر', 'في', 'المغرب', 'وعلى', 'رأسه', 'الشاعر', 'مراد'))]  
[(1, ('كل', 'من', 'عليها', 'فان', 'ويبقى', 'وجه', 'ربك', 'ذو', 'الجلال', 'والإكرام.....'))]
```

Top frequent n-grams generally

```
bigrams = most_frequent_ngrams_per_class(df, "postId", "comment", "f06aa998054e11eba66e646e69d991ea", 2)
```

```
for val in bigrams[:5]:
    print(val)
```

```
[(1, ('علينا', 'الاهتمام'))]
[(2, ('ما', 'نعرف'))]
[(1, ('مبادرات', 'مهمة'))]
[(1, ('شكرا', 'لمكتب'))]
```

```
trigrams = most_frequent_ngrams_per_class(df, "postId", "comment", "f50f0476054e11eba31b646e69d991ea", 3)
```

```
for val in trigrams[:5]:
    print(val)
```

```
[(1, ('فضيحة', 'من', 'العبار'))]
[(1, ('حد', 'لو', 'تم'))]
[(2, ('لولا', 'ضغط', 'الجمهور'))]
[(1, ('بدوري', 'أحيي', 'الفنان'))]
[(1, ('نتمنى', 'أن', 'يتبع'))]
```

Top frequent n-grams per specific class

3 – Lengths of examples in words and in letters

```
example, length = examples_length(df, "postId", "comment")
```

```
for i in range(5):
    print(example[i], length[i])
```

علينا الاهتمام ببيت الأبحاث العلمية و المختبرات، بماذا نفعلنا الشعر و الغناء سوى التفاهة و تضيق الوقت
بيت الشعر في المغرب يحظى بالتقدير داخل وخارج أرض الوطن لما يسديه للشعر نصوصاً ودراسات وأنشطة في منتهى الجدية... لا نعرف ما إذا كانت منشوراته قد حظيت بالدعم هذه السنة، ولا
نعرف ما إذا كان الوزير الجديد يعرف عن هذا البيت وعن قيمته الاعتبارية التي جعلت اليونسكو تُحدث يوماً عالمياً للشعر باقتراح منه. أما صباح الزيداني، فهي الآن وحدها سرب، وكان أولى بالو
زارة أن تمكنها من أسباب التحليق...
شكراً لمراد والذين معه. 409
مبادرات مهمة يقوم بها بيت الشعر لصون الذاكرة الثقافية ببلادنا 61
شكرا لمكتب بيت الشعر في المغرب وعلى رأسه الشاعر مراد في الدفاع عن الرموز الثقافية المغربية وجعل الثقافة في صلب القضايا والاهتمامات الإنسانية 140
كل من عليها فان ويبقى وجه ربك ذو الجلال والإكرام... ميزانية هذه السينما يجب ان توزع على البحث العلمي و ليس على الفنانين، نشكر فيروس كرونا على تنوير الرأي العام 160

Lengths of general examples in letters

```
example, length = examples_length(df, "postId", "comment", length = "words")
for i in range(5):
    print(example[i], length[i])
```

علينا الاهتمام ببيت الأبحاث العلمية و المختبرات، بماذا نفعلنا الشعر و الغناء سوى التفاهة و تضيق الوقت
بيت الشعر في المغرب يحظى بالتقدير داخل وخارج أرض الوطن لما يسديه للشعر نصوصاً ودراسات وأنشطة في منتهى الجدية... لا نعرف ما إذا كانت منشوراته قد حظيت بالدعم هذه السنة، ولا
نعرف ما إذا كان الوزير الجديد يعرف عن هذا البيت وعن قيمته الاعتبارية التي جعلت اليونسكو تُحدث يوماً عالمياً للشعر باقتراح منه. أما صباح الزيداني، فهي الآن وحدها سرب، وكان أولى بالو
زارة أن تمكنها من أسباب التحليق...
شكراً لمراد والذين معه. 72
مبادرات مهمة يقوم بها بيت الشعر لصون الذاكرة الثقافية ببلادنا 10
شكرا لمكتب بيت الشعر في المغرب وعلى رأسه الشاعر مراد في الدفاع عن الرموز الثقافية المغربية وجعل الثقافة في صلب القضايا والاهتمامات الإنسانية 23
كل من عليها فان ويبقى وجه ربك ذو الجلال والإكرام... ميزانية هذه السينما يجب ان توزع على البحث العلمي و ليس على الفنانين، نشكر فيروس كرونا على تنوير الرأي العام 30

Lengths of general examples in words

```
example, length = examples_length(df, "postId", "comment", "class")
for i in range(len(example)):
    print(example[i], length[i])
```

class_name: 015031ba055111eb92e2646e69d991ea

كورونا يبعد التفاهة عن شاشات الإفساد الاخلاقي ، ربما هكذا كان يجب ان يكون عنوان المقال. 87

عيننا بالافلام الثقافية ولي خارجة عن الواقع كل من هب ودب افتحم السينما من أجل مصالح الشخصية مشا التمثيل مع ماله بالي ماستافدو والو في الحياة ولا المماة. 153

كورونا" يبعد أفلاماً سينمائية عن الشاشة السوداء 47

Lengths of randomly selected class examples in letters

```
example, length = examples_length(df, "postId", "comment", "class", "words")
for i in range(len(example)):
    print(example[i], length[i])
```

class_name: c7e6d31c054f11eba77a646e69d991ea

فيلم جميل وممتع.

(أتمنى مقالات أكثر من هذا النوع مستقبلا)

شكرا. 11

فيلم رائع و اتيق. انصح به ليهواة السينما 8

جميل.. حبذا لو تم التعريف في مقالات قابلة بأفلام سينمائية ذات مواضيع أدبية وثقافية مميزة 15

Lengths of randomly selected class examples in words