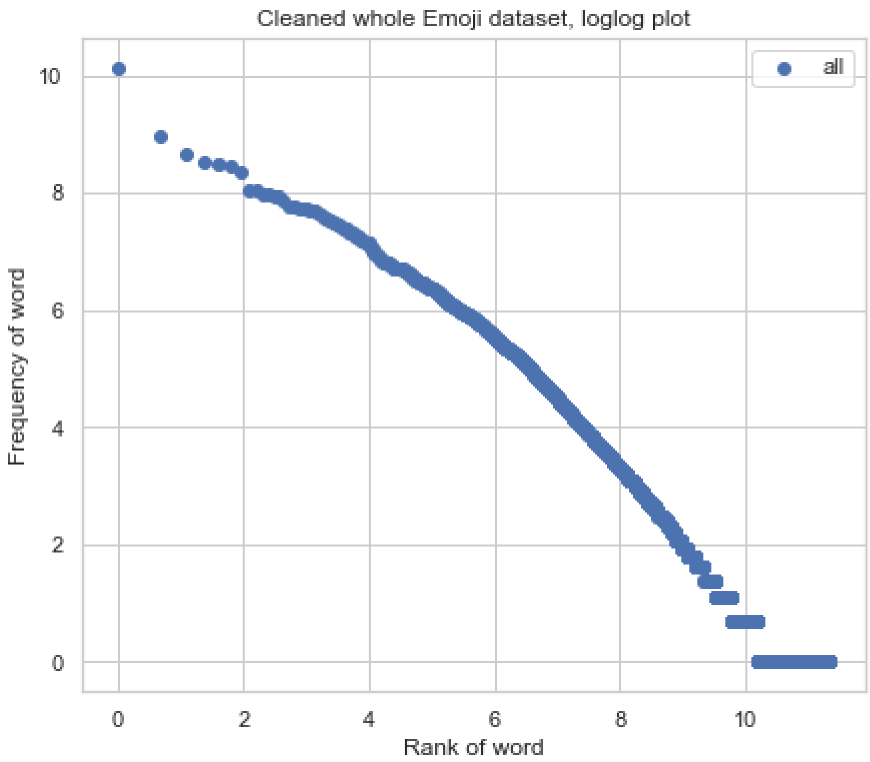
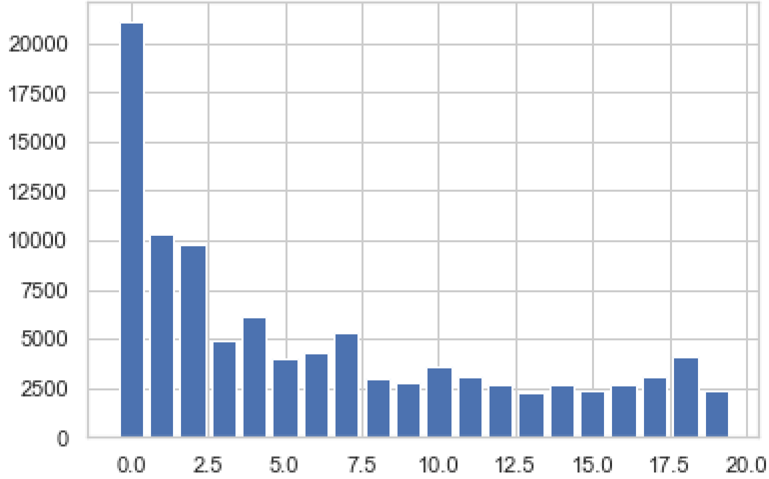
```
Statistics for whole Emoji dataset:

Top 20 in freq_words
 {'user': 25106, 'love': 7771, 'california': 5797, 'new': 5100, 'amp': 4940, 'h
appy': 4646, 'day': 4312, 'beach': 3113, 'today': 3072, 'night': 2938, 'time':
2919, 'one': 2852, 'christmas': 2846, 'good': 2663, 'los': 2392, 'like': 2383,
'san': 2350, 'park': 2312, 'ca': 2293, 'angeles': 2288}

Top 20 least common words in freq_words
 [('unday', 1), ('belch', 1), ('happybirthdaybubb', 1), ('flipnout', 1), ('obse
ssedwithmydog', 1), ('tinyepic', 1), ('tinyepicwestern', 1), ('calebrancourt',
1), ('parrisproject', 1), ('sundaymarket', 1), ('roadraceengineering', 1), ('cr
stylestravel', 1), ('notcaturday', 1), ('sausageparty', 1), ('happygayunclesda
y', 1), ('happygunclesday', 1), ('mauricethewhale', 1), ('colorsworldwide', 1),
('rnbonly', 1), ('powerrangerszeo', 1)]

The number of characters: 7237437
The number of tokens is: 1493251
The number of sentences is: 52054
The average number of tokens per sentence is: 20
The number of unique tokens are: 83462
The tokens ratio is: 0.056
The number of total tokens after removing stopwords are: 694344
Zipf's law, LogLog Plot
```



```
           Distribution of true labels:

           Plot of labels distribution:
```



```
    Label no. Emoji                          Emoji as txt  Count  Percentage
0           0   ❤                              _red_heart_  21057       21.06
1           1   😍            _smiling_face_with_hearteyes_  10252       10.25
2           2   😂                  _face_with_tears_of_joy_   9750        9.75
3           3   💕                             _two_hearts_   4956        4.96
4           4   🔥                                   _fire_   6105        6.11
5           5   😊          _smiling_face_with_smiling_eyes_   3983        3.98
6           6   😎              _smiling_face_with_sunglasses_   4278        4.28
7           7   ✨                               _sparkles_   5293        5.29
8           8   💙                             _blue_heart_   3013        3.01
9           9   😘                      _face_blowing_a_kiss_   2737        2.74
10         10   📷                                 _camera_   3573        3.57
11         11  🇺🇸                         _United_States_   3038        3.04
12         12   ☀                                    _sun_   2639        2.64
13         13   💜                           _purple_heart_   2247        2.25
14         14   😉                           _winking_face_   2659        2.66
15         15   💯                         _hundred_points_   2326        2.33
16         16   😁          _beaming_face_with_smiling_eyes_   2640        2.64
17         17   🎄                         _Christmas_tree_   3063        3.06
18         18   📸                       _camera_with_flash_   4056        4.06
19         19   😜                  _winking_face_with_tongue_   2332        2.33
```