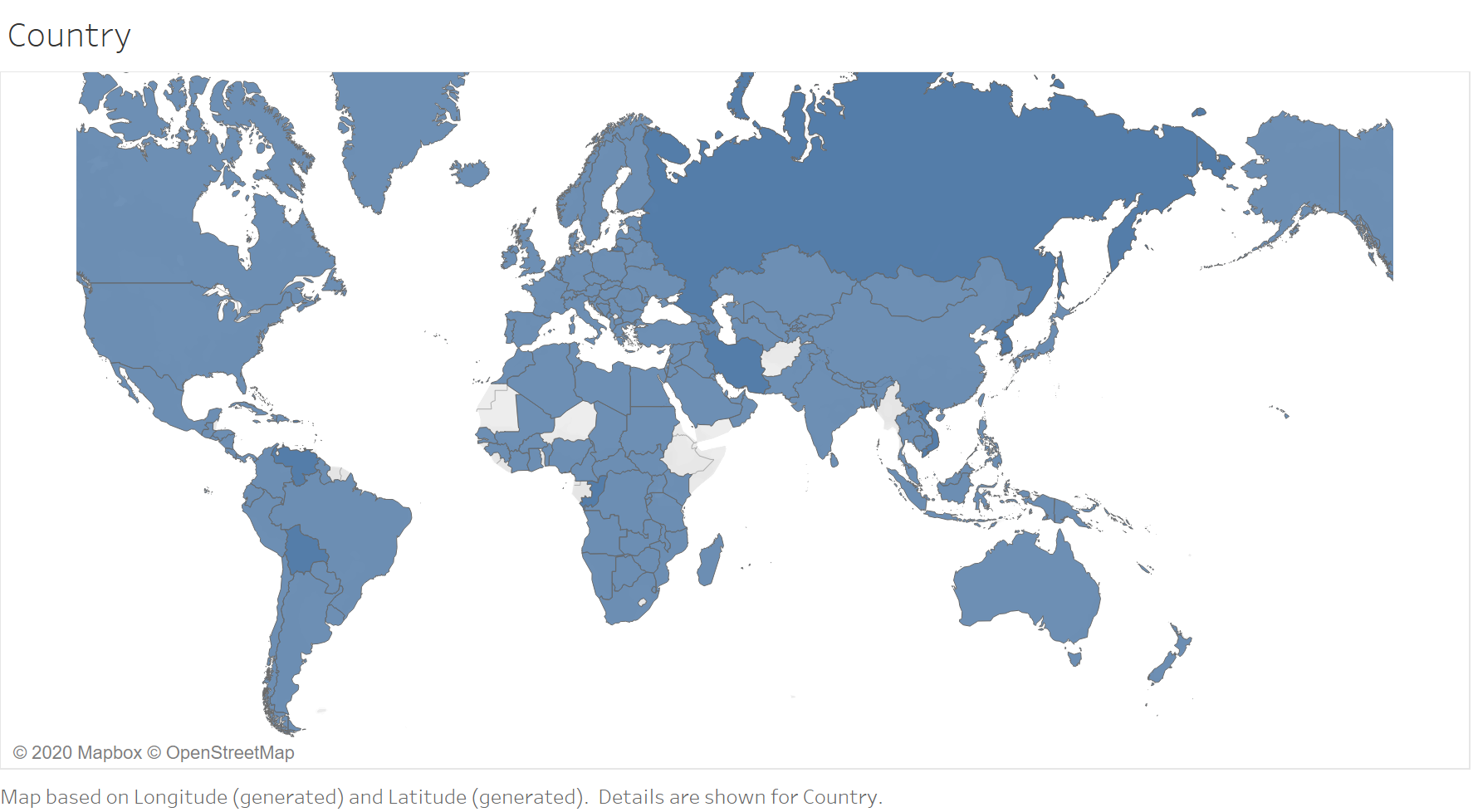
Data Description Report

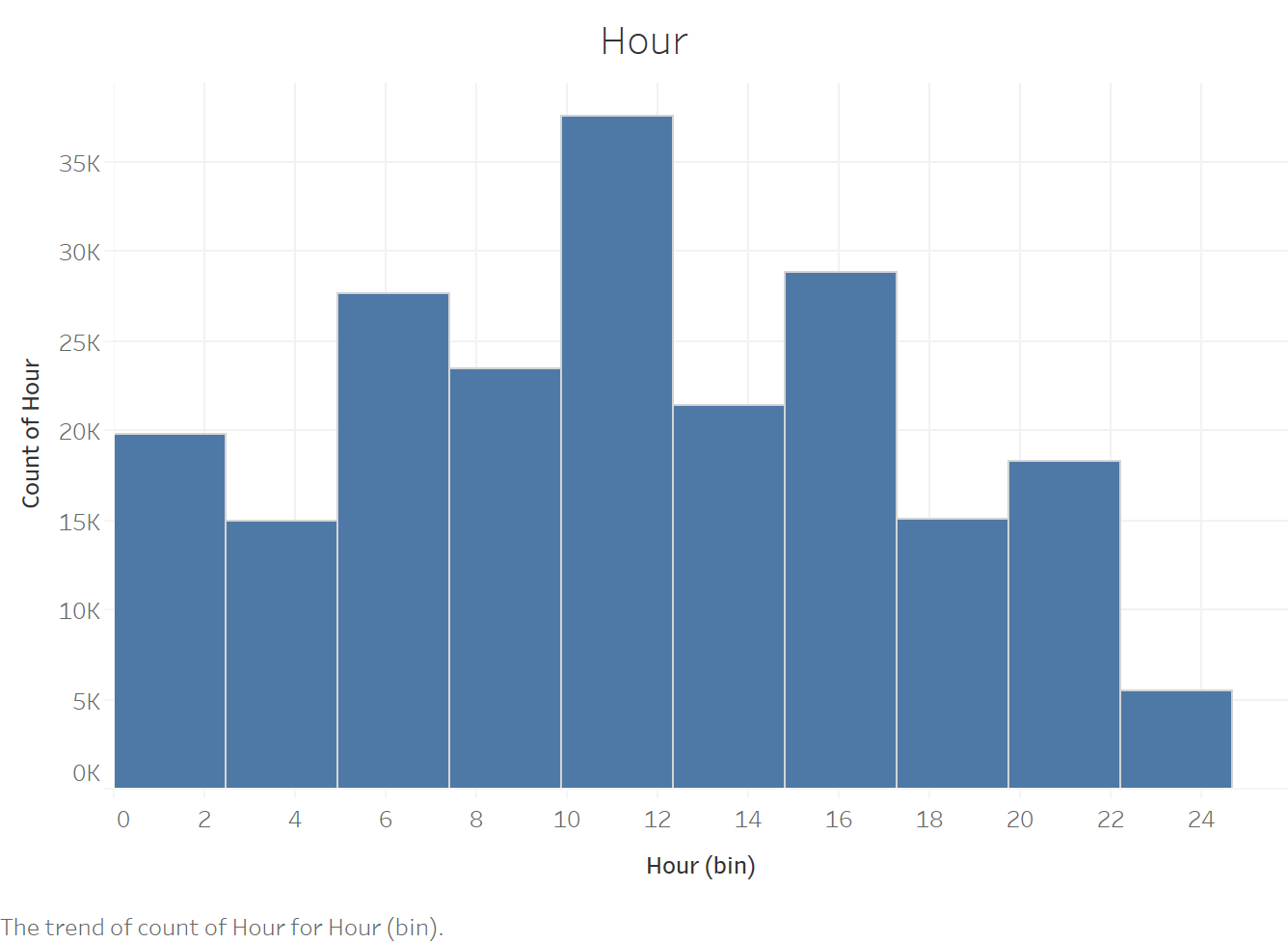
* Country
  + Country of the tweet’s origin
* State
  + State within the Country where the tweet has been retweeted
* State Code
  + State code of the state where the tweet has been retweeted
* City
  + City from which the tweet has been retweeted
* Weekday
  + Specific day of the week when the has been retweeted
* Hour
  + Specific hour of the day has been retweeted
* Day
  + Date on which the tweet has been retweeted
* Lang
  + Language of the tweet
* Is Reshare
  + If the tweet is a retweet or original
* Reach
  + Followers of an account handle
* Retweet Count
  + Number of times the tweet has been retweeted
* Likes
  + Number of likes for a retweet
* Klout
  + Klout score (popularity score). This is an artificial score.
* Sentiment
  + Classification of the tweet (in terms of the tweet meaning and influence). Positive/negative
* Text
  + The content of the tweet
* Gender
  + Biological sex of the person associated with the tweet

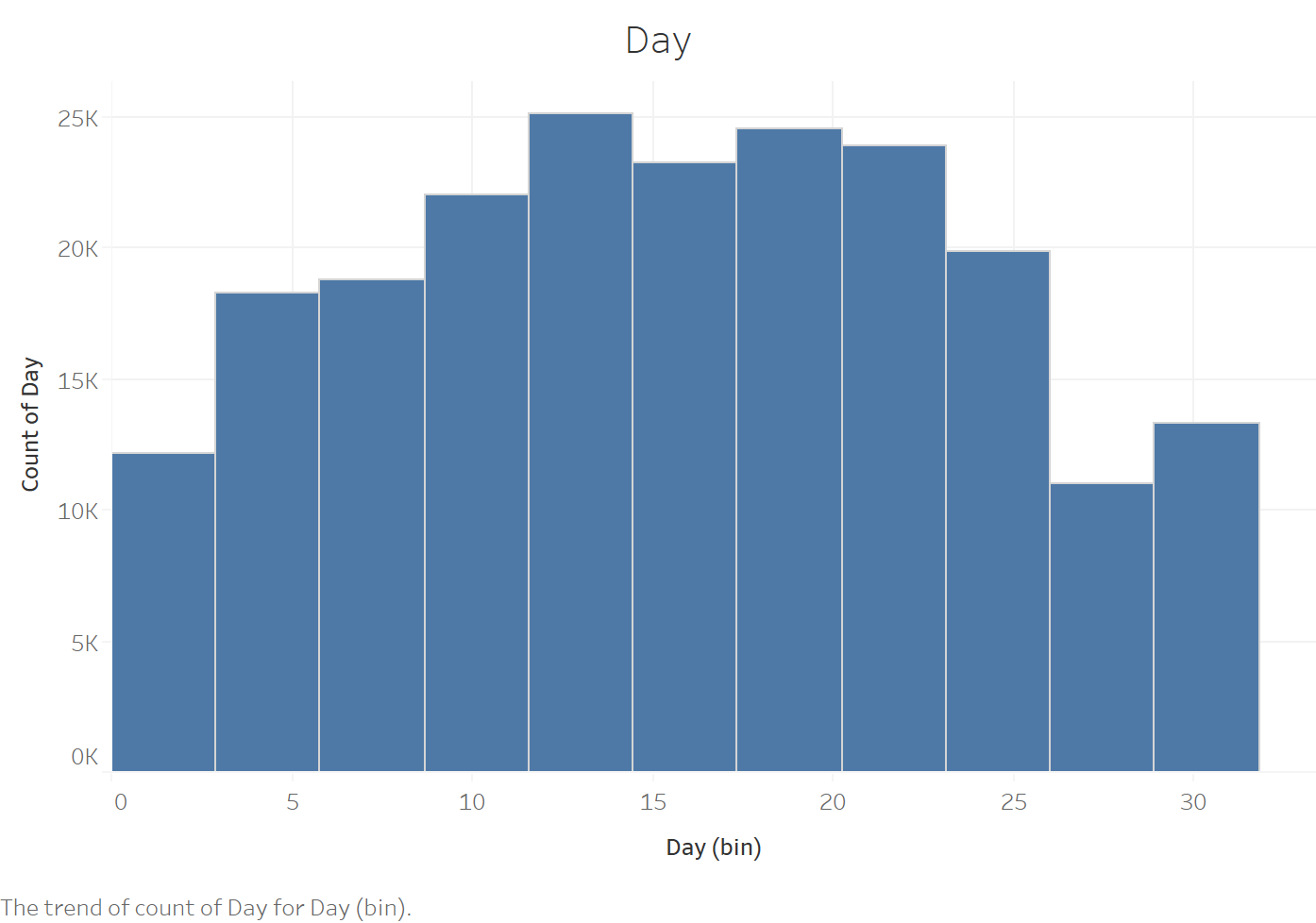
# Univariate Properties

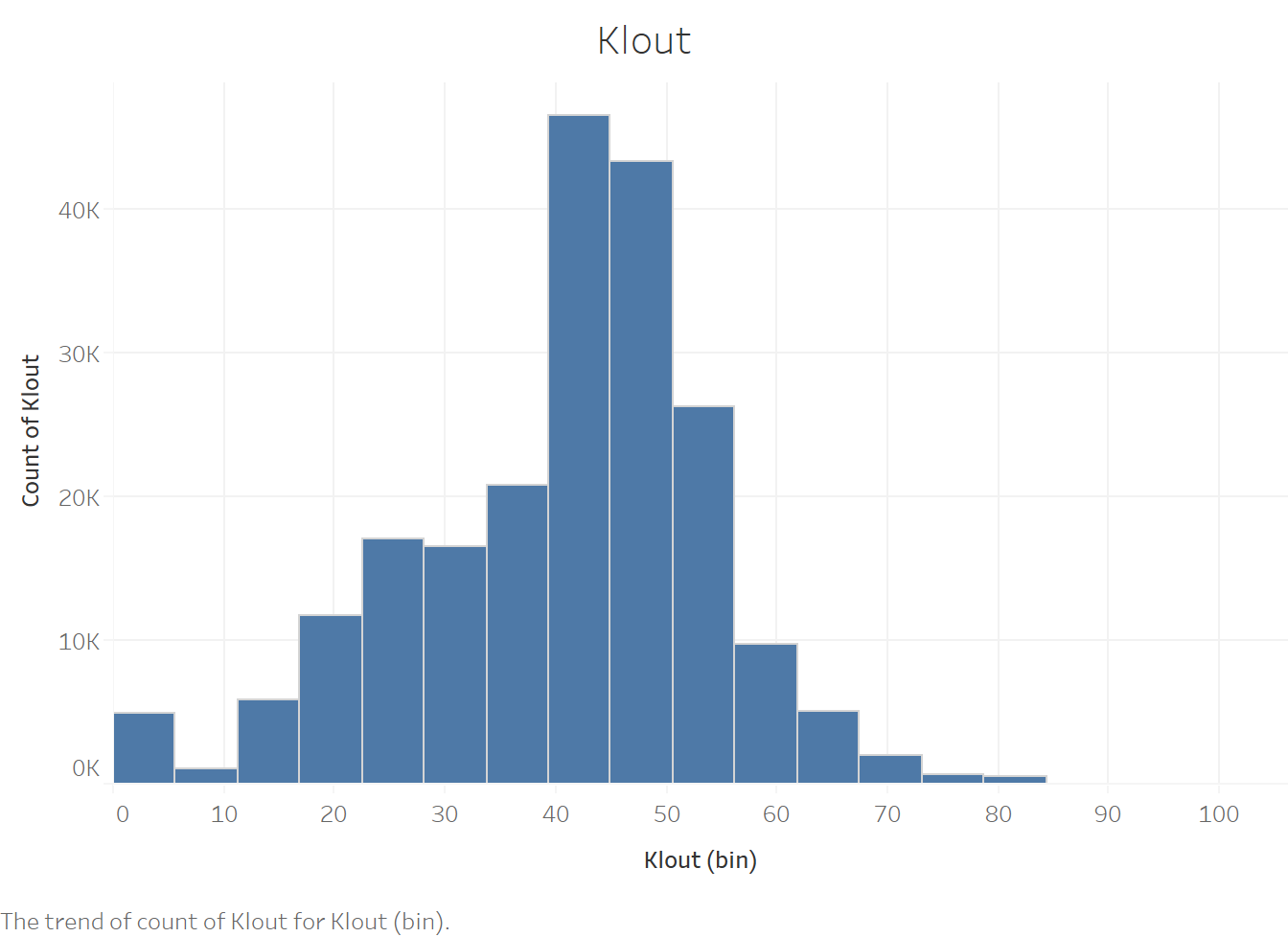
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Feature** | **Type** | **Data Type** | **Count** | **Missing** | **%** | **Unique** | **Min** | **Q1** | **Med** | **Q3** | **Max** | **Mean** | **SD** | **Skew** | **Kurt** |
| Location ID (tw Location) | Feature | Integer | 212191 | 0 | 0 | 6289 | 1 | 1494 | 3767 | 3776 | 6289 | 2,817.86 | 1355 | -0.55 | 2.3 |
| Country | Feature | String | 212191 | 0 | 0 | 175 | x | x | x | x | x | x | x | x | x |
| State | Feature | String | 212191 | 0 | 0 | 1225 | x | x | x | x | x | x | x | x | x |
| State Code | Feature | String | 212191 | 0 | 0 | 388 | x | x | x | x | x | x | x | x | x |
| City | Feature | String | 212191 | 0 | 0 | 5516 | x | x | x | x | x | x | x | x | x |
| TweetID | Feature | String | 212191 | 0 | 0 | 207850 | x | x | x | x | x | x | x | x | x |
| Weekday | Feature | String | 212191 | 0 | 0 | 7 | x | x | x | x | x | x | x | x | x |
| Hour | Feature | Integer | 212191 | 0 | 0 | 24 | 0 | 7 | 11 | 16 | 23 | 11.25 | 6.13 | 0.05 | 2.12 |
| Day | Feature | Integer | 212191 | 0 | 0 | 31 | 1 | 9 | 16 | 22 | 31 | 15.7 | 8.28 | -0.01 | 1.96 |
| Lang | Feature | String | 212191 | 0 | 0 | 41 | x | x | x | x | x | x | x | x | x |
| Is Reshare | Feature | Boolean | 212191 | 0 | 0 | 2 | x | x | x | x | x | x | x | x | x |
| Reach | Feature | Integer | 212191 | 0 | 0 | 18793 | 0 | 151 | 469 | 1556 | 10,342,452 | 8485.7 | 85,507 | 51.31 | 4137.2 |
| Likes | Feature | Integer | 212191 | 0 | 0 | 106 | 0 | 0 | 0 | 0 | 291 | 0.15 | 2.7 | 28 | 1307 |
| Klout | Feature | Integer | 212191 | 0 | 0 | 87 | 0 | 32 | 43 | 49 | 99 | 40.44 | 13.81 | -1 | 0.69 |
| Sentiment | Feature | Float | 212191 | 0 | 0 | 496 | -8 | 0 | 0 | 1 | 8 | 0 | 1.03 | 0.77 | 6 |
| Text | Feature | String | 212191 | 0 | 0 | 142862 | x | x | x | x | x | x | x | x | x |
| LocationID | Feature | Integer | 212191 | 0 | 0 | 6289 | 1 | 1494 | 3767 | 3776 | 6289 | 2818 | 1355 | -1 | 2.3 |
| UserID | Feature | String | 212191 | 0 | 0 | 61990 | x | x | x | x | x | x | x | x | x |
| UserID (tw User) | Feature | String | 212191 | 0 | 0 | 61990 | x | x | x | x | x | x | x | x | x |
| Gender | Feature | String | 212191 | 0 | 0 | 4 | x | x | x | x | x | x | x | x | x |
| Retweet Count | Label | Integer | 212191 | 0 | 0 | 2069 | 0 | 0 | 0 | 3 | 26127 | 24.6 | 183 | 24.38 | 2164.9 |

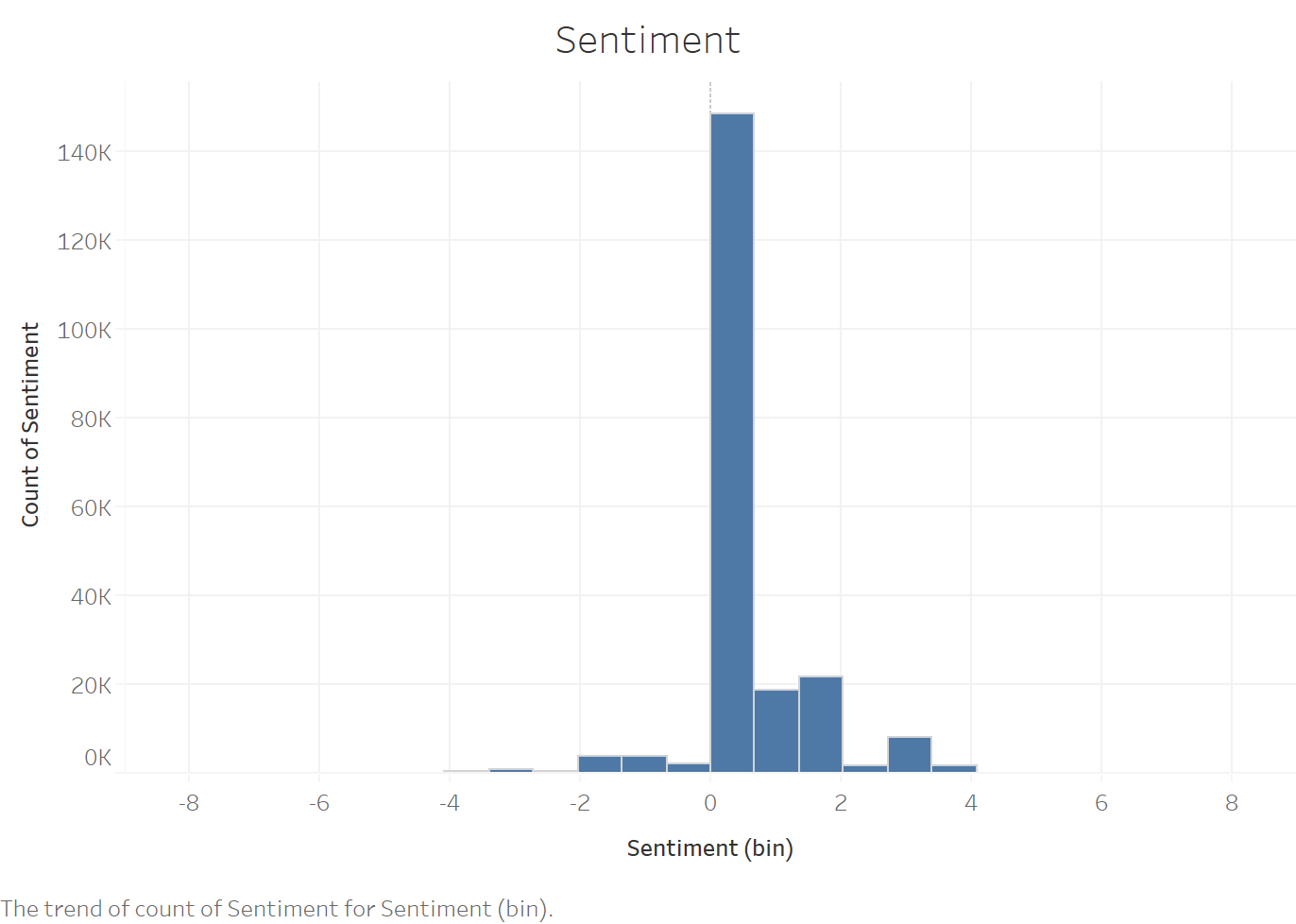
# Univariate Visualizations

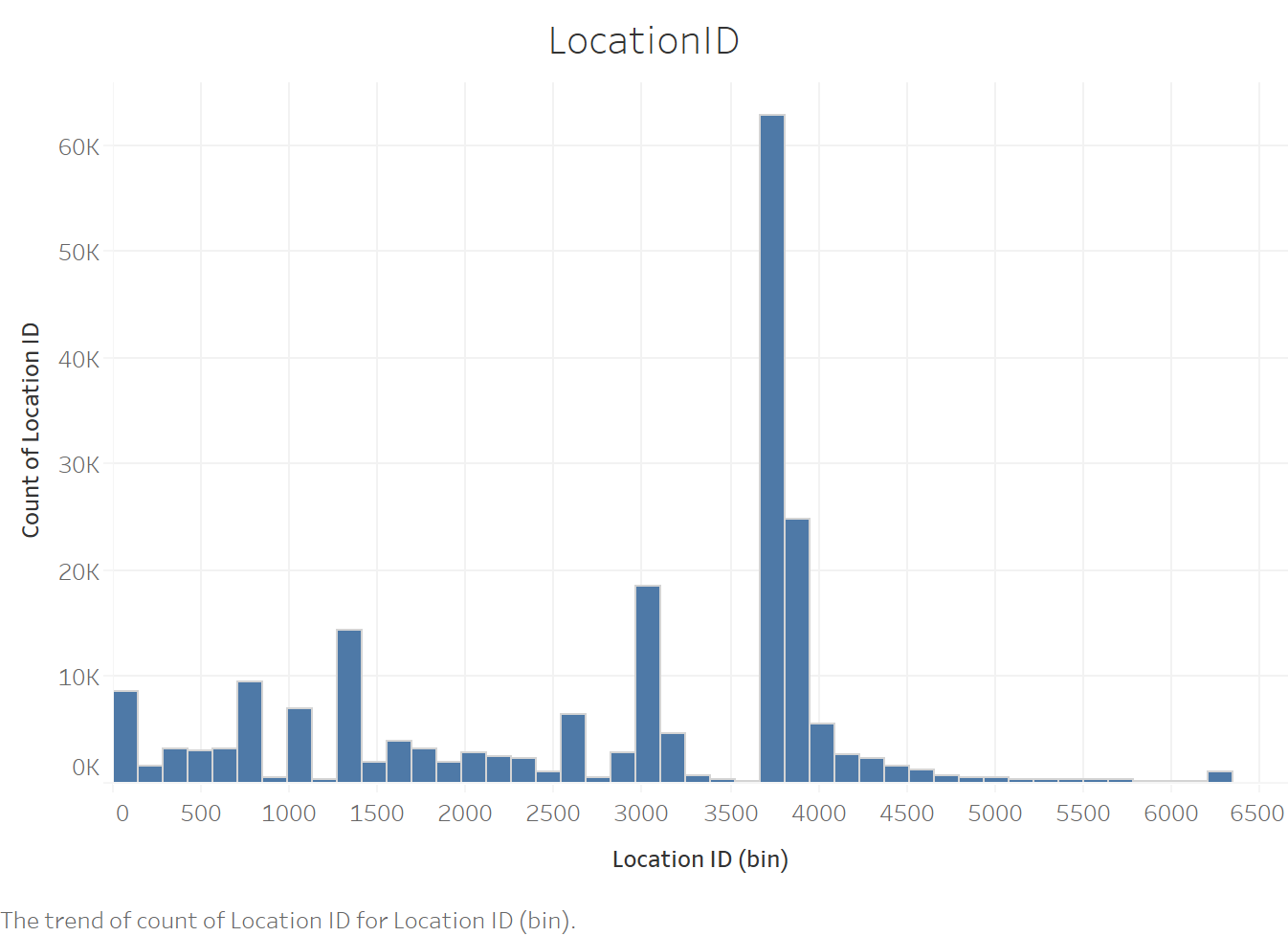


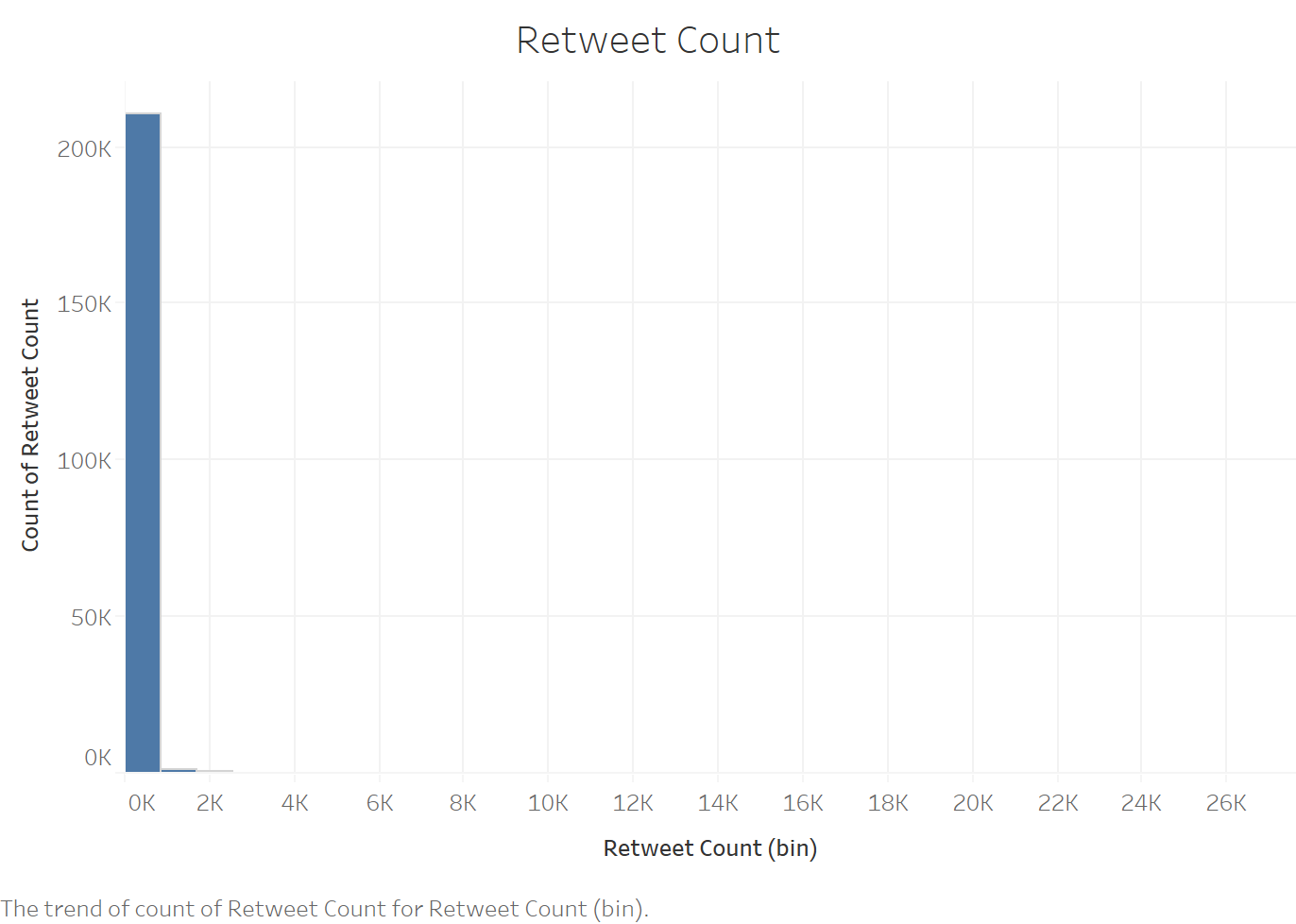












Reference:

Rao, Adithya, et al. "Klout score: Measuring influence across multiple social networks." *2015 IEEE International Conference on Big Data (Big Data)*. IEEE, 2015.

Chen, Jundong, et al. "Sentiment analysis of the correlation between regular tweets and retweets." *2017 IEEE 16th international symposium on network computing and applications (NCA)*. IEEE, 2017.