Significance of tips to classifying a user as “elite”:

Questions to explore:

1. Using NLP techniques, can we predict what type of tips elite users give (i.e. the language or advice or topic contained within the tip)?
2. What percent of tips per year are given by elite users vs. non-elite users?
3. Are there particular elite or non-elite users that are giving a large proportion of tips?

For our dataset, we have 946,600 total tips. Tips are only given between years 2009-2017.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **Total** |
| tips | 912 | 40,371 | 141,422 | 179,214 | 161,283 | 158,001 | 125,512 | 133,568 | 6,317 | 946,600 |
| elite tips | 287 | 9,928 | 30,462 | 46,004 | 36,981 | 28,927 | 26,247 | 30,965 | 1,442 | 211,243 |
| % elite | 31.5 | 24.6 | 21.5 | 25.7 | 22.9 | 18.3 | 20.9 | 23.2 | 22.8 | 22.3 |



What is the distribution of tips coming from unique elite users?

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** |
| # unique elite tip-givers | 58 | 1080 | 2505 | 3887 | 3621 | 3514 | 3947 | 4740 | 660 |
| # 1-time tip-givers | 26 | 324 | 703 | 1153 | 1263 | 1269 | 1546 | 1860 | 415 |
| # 2-time tip-givers | 6 | 181 | 371 | 526 | 553 | 540 | 633 | 773 | 99 |
| # 3-time tip-givers | 7 | 100 | 230 | 362 | 302 | 318 | 361 | 420 | 48 |
| # 4-time tip-givers | 3 | 53 | 155 | 264 | 233 | 203 | 228 | 294 | 33 |
| # 5-time tip-givers | 4 | 49 | 133 | 179 | 157 | 154 | 174 | 180 | 10 |





|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** |
| # unique elite tip-givers | 58 | 1080 | 2505 | 3887 | 3621 | 3514 | 3947 | 4740 | 660 |
| # 10≤x<20-time tip-givers | 6 | 135 | 273 | 439 | 344 | 341 | 344 | 414 | 17 |
| # 20≤x<30-time tip-givers | 0 | 43 | 111 | 181 | 145 | 121 | 104 | 150 | 2 |
| # 30≤x<40-time tip-givers | 1 | 24 | 56 | 95 | 75 | 68 | 52 | 57 | 0 |



Compare this to the distribution of tips given by non-elite users per year:

**New Project Aim: Patterns of Elitism**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** |
| Elites | 6510 | 10,396 | 13,072 | 17,572 | 19,550 | 20,025 | 25,009 | 28,846 | 22,628 |
| Non-Elites | 1,022,922 | 1,019,036 | 1,016,360 | 1,011,860 | 1,009,882 | 1,009,407 | 1,004,423 | 1,000,586 | 1,006,804 |

Total User Counts

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** |
| Elite | Non-Elite | 1118 | 2100 | 2619 | 2683 | 4508 | 3732 | 4182 | 6512 | X |
| Elite | Elite | 5392 | 8296 | 10,523 | 14889 | 15042 | 16293 | 20827 | 22334 | X |
| Non-Elite | Elite | 5004 | 4776 | 7119 | 4661 | 4983 | 8716 | 8019 | 294 | X |
| *At Least 1 yr Elite*  Non-Elite | Non-Elite | 38190 | 34532 | 29513 | 27471 | 25171 | 20963 | 16676 | 20564 | X |
| *Never Elite*  Non-Elite | Non-Elite | 979728 | 979728 | 979278 | 979278 | 979278 | 979278 | 979278 | 979278 | X |
| *Total*  Non-Elite | Non-Elite | 1017918 | 1014260 | 1009241 | 1007199 | 1004899 | 1000691 | 996404 | 1000292 | X |

User Percentages

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** |
| Elite | Non-Elite | .0011 | .0020 | .0025 | .0026 | .0044 | .0036 | .0041 | .0063 | X |
| Elite | Elite | .0052 | .0081 | .0102 | .0145 | .0146 | .0158 | .0202 | .0217 | X |
| Non-Elite | Elite | .0049 | .0046 | .0069 | .0045 | .0048 | .0085 | .0078 | .0003 | X |
| *At Least 1 yr Elite*  Non-Elite | Non-Elite | .0371 | .0335 | .0287 | .0267 | .0245 | .0204 | .0162 | .0200 | X |
| *Never Elite*  Non-Elite | Non-Elite | .9517 | .9517 | .9517 | .9517 | .9517 | .9517 | .9517 | .9517 | X |
| *Total*  Non-Elite | Non-Elite | .9888 | .9853 | .9804 | .9784 | .9762 | .9721 | .9679 | .9717 | X |

|  |  |  |
| --- | --- | --- |
|  | **Overall: 2009-2016** | **Percentages: 2009-2016** |
| Elite | Non-Elite | 27454 | 0.0033 |
| Elite | Elite | 113596 | 0.0137 |
| Non-Elite | Elite | 43572 | 0.0052 |
| Non-Elite | Non-Elite | 805094 | 0.9775 |
|  | 8235456 | 1.0 |

\*The overall total is essentially the total number of users (1,029,432) \* the number of years (8).

We want to focus on what tips and reviews that non-elite users leave on Yelp that allow them to become elite users in the following year, as well as what tips and reviews that elite users are leaving that allow them to continue their elite status.

Using NLP techniques, how are the N.E. and E.E. users’ comments and reviews distinguishable from E.N. users? Are these statistically significant differences?

Looking at tips (counts and percentages) across the elite/non-elite patterns:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tip Counts** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** |
| Elite | Non-Elite Tips | 5 | 920 | 4659 | 4336 | 4496 | 2729 | 2229 | 2879 | X |
| Elite | Elite Tips | 282 | 9008 | 25803 | 41668 | 32485 | 26198 | 24018 | 28086 | X |
| Non-Elite | Elite Tips | 29 | 1942 | 6375 | 4938 | 4053 | 5695 | 3804 | 358 | X |
| Non-Elite | Non-Elite Tips | 596 | 28501 | 104585 | 128272 | 120249 | 123379 | 95461 | 102245 | X |
| Total Tips | 912 | 40371 | 141422 | 179214 | 161283 | 158001 | 125512 | 133568 | 6317 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **User Counts** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** |
| Elite | Non-Elite | 1118 | 2100 | 2619 | 2683 | 4508 | 3732 | 4182 | 6512 | X |
| Elite | Elite | 5392 | 8296 | 10,523 | 14889 | 15042 | 16293 | 20827 | 22334 | X |
| Non-Elite | Elite | 5004 | 4776 | 7119 | 4661 | 4983 | 8716 | 8019 | 294 | X |
| Non-Elite | Non-Elite | 1017918 | 1014260 | 1009241 | 1007199 | 1004899 | 1000691 | 996404 | 1000292 | X |



What is the tips-per-user (TPU) average for each group?

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Counts** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** |
| Elite | Non-Elite TPU | .0044 | .4381 | 1.7789 | 1.6161 | 0.9973 | 0.7312 | 0.5329 | 0.4421 | X |
| Elite | Elite TPU | .0522 | 1.086 | 2.4520 | 2.7986 | 2.1596 | 1.6079 | 1.1532 | 1.2575 | X |
| Non-Elite | Elite TPU | .0057 | 0.4066 | 0.8955 | 1.0594 | 0.8133 | 0.6533 | 0.4744 | 1.2176 | X |
| Non-Elite | Non-Elite TPU | .0006 | 0.0281 | 0.1036 | 0.1273 | 0.1196 | 0.1233 | 0.1259 | 0.1022 | X |



Next questions:

Q1: What is the distribution of individual users giving these tips (i.e. is a large portion of the tips given by a small subset of users or are the tips pretty evenly distributed across the user groups)?

Q2: Which particular users are leaving the tips?

Q3: How many outliers do we have in each of the distributions (show boxplots for each year)?

Elite/Elite Group

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** |
| # unique elite tip-givers |  |  |  |  |  |  |  |  |  |
| # 1-time tip-givers |  |  |  |  |  |  |  |  |  |
| # 2-time tip-givers |  |  |  |  |  |  |  |  |  |
| # 3-time tip-givers |  |  |  |  |  |  |  |  |  |
| # 4-time tip-givers |  |  |  |  |  |  |  |  |  |
| # 5-time tip-givers |  |  |  |  |  |  |  |  |  |

Non-Elite/Elite Group

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** |
| # unique elite tip-givers |  |  |  |  |  |  |  |  |  |
| # 1-time tip-givers |  |  |  |  |  |  |  |  |  |
| # 2-time tip-givers |  |  |  |  |  |  |  |  |  |
| # 3-time tip-givers |  |  |  |  |  |  |  |  |  |
| # 4-time tip-givers |  |  |  |  |  |  |  |  |  |
| # 5-time tip-givers |  |  |  |  |  |  |  |  |  |

Elite/Non-Elite Group

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** |
| # unique elite tip-givers |  |  |  |  |  |  |  |  |  |
| # 1-time tip-givers |  |  |  |  |  |  |  |  |  |
| # 2-time tip-givers |  |  |  |  |  |  |  |  |  |
| # 3-time tip-givers |  |  |  |  |  |  |  |  |  |
| # 4-time tip-givers |  |  |  |  |  |  |  |  |  |
| # 5-time tip-givers |  |  |  |  |  |  |  |  |  |

**Meeting Wednesday 4/11 Discussion:**

What makes a person more likely to become elite?

* Higher number of tips?
* Higher number of reviews?
* More thoughtful or helpful reviews/tip content (NLP)?
* Seasonal, i.e. year-round, “elite behavior”?
* Lengths of review?

Next steps:

1. Look at distributions and numbers of reviews for users (as we did for tips).
2. Answer the question: leaving what number of tips maximizes the probability of becoming elite for already elite members? How about for non-elite members wanting to become elite?



  
  


Q3: What are the monthly differences (i.e. posting pattern) for users writing reviews or tips? Measure the monthly deviation and plot.