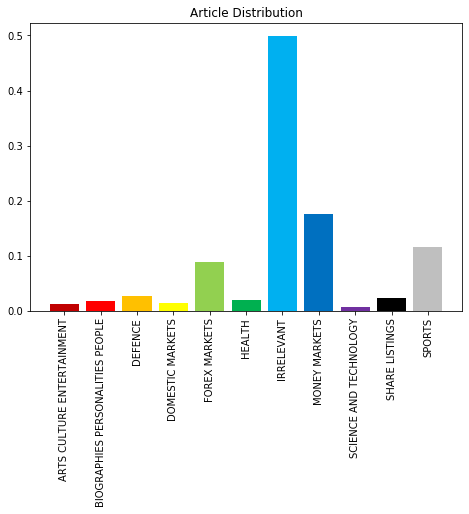
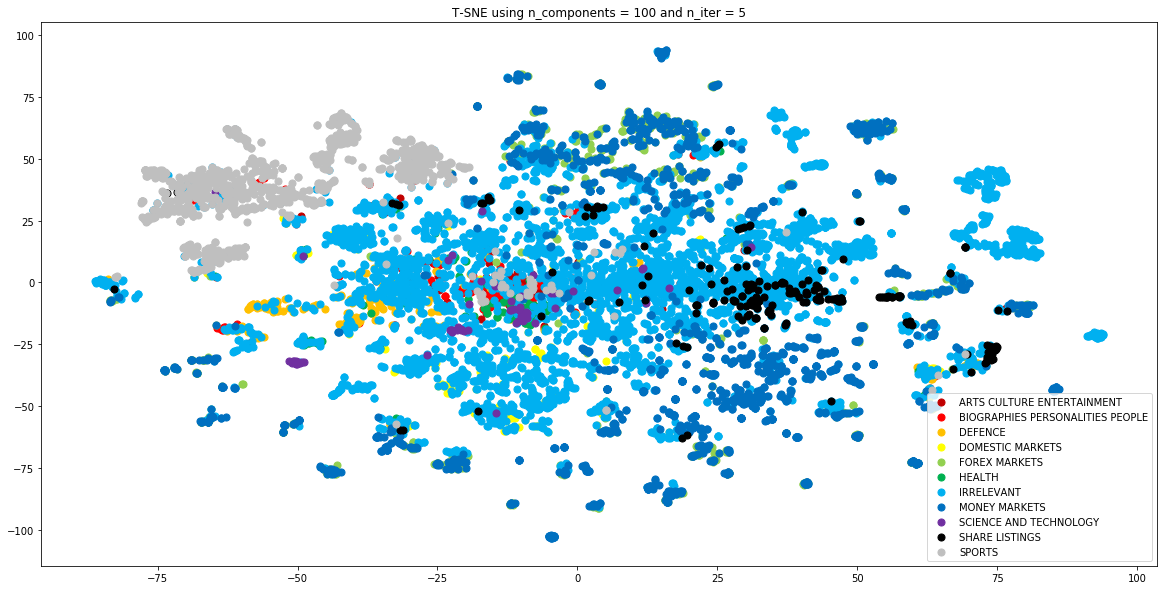
**Explanatory Data Analysis**



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **FOREX MARKETS** | | **IRRELEVANT** | | **MONEY MARKETS** | |
| dollar | 3,358 (2.92%) | percent | 5,945 (1.00%) | bank | 4,447 (2.40%) |
| bank | 2,178 (1.89%) | year | 5,237 (0.88%) | rate | 3,772 (2.03%) |
| rate | 1,755 (1.53%) | million | 4,740 (0.80%) | percent | 3,515 (1.89%) |
| trad | 1,678 (1.46%) | trad | 3,678 (0.62%) | dollar | 3,387 (1.83%) |
| market | 1,668 (1.45%) | market | 3,556 (0.60%) | market | 2,901 (1.56%) |
| percent | 1,336 (1.16%) | shar | 3,294 (0.56%) | trad | 2,490 (1.34%) |
| currenc | 1,292 (1.12%) | compan | 3,190 (0.54%) | day | 1,775 (0.96%) |
| deal | 1,208 (1.05%) | stat | 3,160 (0.53%) | deal | 1,707 (0.92%) |
| mark | 1,191 (1.04%) | bank | 3,095 (0.52%) | cent | 1,533 (0.83%) |
| yen | 962 (0.84%) | pric | 2,611 (0.44%) | currenc | 1,472 (0.79%) |



**Appendix**

**A1 – Article Distribution %**

|  |  |
| --- | --- |
| **Article Distribution (%)** | |
| ARTS CULTURE ENTERTAINMENT | 1.23% |
| BIOGRAPHIES PERSONALITIES PEOPLE | 1.76% |
| DEFENCE | 2.72% |
| DOMESTIC MARKETS | 1.40% |
| FOREX MARKETS | 8.89% |
| HEALTH | 1.93% |
| IRRELEVANT | 49.83% |
| MONEY MARKETS | 17.61% |
| SCIENCE AND TECHNOLOGY | 0.74% |
| SHARE LISTINGS | 2.29% |
| SPORTS | 11.60% |

**A2 – Top 10 most frequent words under bag of words methodology**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Top 10 most frequent words** | | | | | | | |
| **ARTS CULTURE ENTERTAINMENT** | | **BIOGRAPHIES PERSONALITIES PEOPLE** | | **DEFENCE** | | **DOMESTIC MARKETS** | |
| year | 156 (0.75%) | year | 249 (0.80%) | nato | 511 (1.06%) | import | 541 (3.14%) |
| film | 153 (0.73%) | presid | 198 (0.64%) | stat | 410 (0.85%) | ton | 324 (1.88%) |
| million | 102 (0.49%) | yeltsin | 140 (0.45%) | milit | 406 (0.84%) | percent | 265 (1.54%) |
| stat | 83 (0.40%) | stat | 133 (0.43%) | forc | 381 (0.79%) | year | 262 (1.54%) |
| work | 81 (0.39%) | minist | 126 (0.41%) | minist | 364 (0.76%) | oil | 191 (1.11%) |
| offic | 80 (0.38%) | hospit | 124 (0.40%) | defend | 357 (0.74%) | trad | 190 (1.10%) |
| show | 72 (0.34%) | offic | 122 (0.39%) | offic | 317 (0.66%) | chin | 186 (1.08%) |
| releas | 72 (0.34%) | lead | 121 (0.39%) | russia | 299 (0.62%) | export | 161 (0.94%) |
| peopl | 70 (0.33%) | told | 117 (0.38%) | presid | 238 (0.50%) | million | 152 (0.88%) |
| music | 69 (0.33%) | report | 113 (0.37%) | plan | 235 (0.49%) | produc | 140 (0.81%) |
| **FOREX MARKETS** | | **HEALTH** | | **IRRELEVANT** | | **MONEY MARKETS** | |
| dollar | 3,358 (2.92%) | year | 253 (0.85%) | percent | 5,945 (1.00%) | bank | 4,447 (2.40%) |
| bank | 2,178 (1.89%) | health | 248 (0.83%) | year | 5,237 (0.88%) | rate | 3,772 (2.03%) |
| rate | 1,755 (1.53%) | stat | 241 (0.81%) | million | 4,740 (0.80%) | percent | 3,515 (1.89%) |
| trad | 1,678 (1.46%) | diseas | 184 (0.62%) | trad | 3,678 (0.62%) | dollar | 3,387 (1.83%) |
| market | 1,668 (1.45%) | tobacc | 169 (0.57%) | market | 3,556 (0.60%) | market | 2,901 (1.56%) |
| percent | 1,336 (1.16%) | drug | 165 (0.55%) | shar | 3,294 (0.56%) | trad | 2,490 (1.34%) |
| currenc | 1,292 (1.12%) | peopl | 151 (0.51%) | compan | 3,190 (0.54%) | day | 1,775 (0.96%) |
| deal | 1,208 (1.05%) | canc | 148 (0.50%) | stat | 3,160 (0.53%) | deal | 1,707 (0.92%) |
| mark | 1,191 (1.04%) | study | 143 (0.48%) | bank | 3,095 (0.52%) | cent | 1,533 (0.83%) |
| yen | 962 (0.84%) | percent | 139 (0.47%) | pric | 2,611 (0.44%) | currenc | 1,472 (0.79%) |
| **SCIENCE AND TECHNOLOGY** | | **SHARE LISTINGS** | | **SPORTS** | |  | |
| spac | 141 (1.23%) | shar | 740 (3.56%) | play | 1,491 (1.06%) |  |  |
| mir | 110 (0.96%) | compan | 392 (1.89%) | win | 1,054 (0.75%) |  |  |
| shuttl | 83 (0.72%) | list | 347 (1.67%) | game | 947 (0.67%) |  |  |
| scient | 77 (0.67%) | million | 296 (1.42%) | year | 938 (0.67%) |  |  |
| year | 73 (0.63%) | stock | 282 (1.36%) | world | 928 (0.66%) |  |  |
| station | 70 (0.61%) | trad | 252 (1.21%) | match | 925 (0.66%) |  |  |
| crew | 68 (0.59%) | exchang | 238 (1.14%) | champ | 880 (0.63%) |  |  |
| mission | 67 (0.58%) | pric | 209 (1.01%) | cup | 786 (0.56%) |  |  |
| nasa | 62 (0.54%) | percent | 200 (0.96%) | team | 734 (0.52%) |  |  |
| research | 61 (0.53%) | market | 195 (0.94%) | scor | 707 (0.50%) |  |  |

**A3 – Top 10 highest TF-IDF words**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Top 10 highest TF-IDF words** | | | | | | | |
| **ARTS CULTURE ENTERTAINMENT** | | **BIOGRAPHIES PERSONALITIES PEOPLE** | | **DEFENCE** | | **DOMESTIC MARKETS** | |
| findley | 0.7795 | jagan | 0.8203 | okinaw | 0.7728 | mln | 0.8674 |
| hemingway | 0.7748 | farrakh | 0.8018 | gore | 0.7380 | cocoa | 0.7981 |
| ballet | 0.7606 | elvi | 0.7950 | ralston | 0.7134 | nil | 0.7839 |
| dor | 0.7272 | sobhraj | 0.7722 | taiw | 0.6701 | motor | 0.7533 |
| iram | 0.7178 | anh | 0.7429 | sudan | 0.6649 | maiz | 0.7472 |
| elvi | 0.6927 | hoff | 0.7385 | rokhlin | 0.6601 | sugar | 0.7329 |
| vasar | 0.6914 | patnaik | 0.7360 | symet | 0.6551 | tea | 0.7165 |
| cronkit | 0.6879 | gandh | 0.7133 | oau | 0.6551 | coconut | 0.7089 |
| mgm | 0.6863 | uday | 0.7107 | lebe | 0.6363 | lumb | 0.7065 |
| pirat | 0.6838 | manley | 0.7090 | bahrain | 0.6306 | rakver | 0.6956 |
| **FOREX MARKETS** | | **HEALTH** | | **IRRELEVANT** | | **MONEY MARKETS** | |
| shekel | 0.8604 | kevork | 0.7863 | unq | 0.9542 | goi | 0.9948 |
| forward | 0.8305 | scan | 0.7657 | lbs | 0.9389 | day | 0.9419 |
| unch | 0.8255 | canc | 0.7520 | plc | 0.9281 | uak | 0.8904 |
| shill | 0.8200 | vaccin | 0.7353 | mt | 0.9245 | oct | 0.8819 |
| bir | 0.8192 | mushroom | 0.7309 | na | 0.9168 | sept | 0.8351 |
| yuan | 0.8019 | measl | 0.7273 | hallwood | 0.9084 | forward | 0.8276 |
| lek | 0.7959 | cesarean | 0.7184 | footstar | 0.9060 | shekel | 0.8153 |
| snb | 0.7935 | smok | 0.7034 | gallon | 0.9048 | reai | 0.8085 |
| lev | 0.7811 | marijuan | 0.7001 | gradc | 0.8793 | june | 0.7998 |
| teng | 0.7776 | fat | 0.6838 | ostrich | 0.8707 | mln | 0.7972 |
| **SCIENCE AND TECHNOLOGY** | | **SHARE LISTINGS** | | **SPORTS** | |  | |
| pand | 0.7262 | easdaq | 0.8428 | bartov | 0.8660 |  |  |
| clon | 0.7068 | worldxchang | 0.8037 | daly | 0.8494 |  |  |
| arian | 0.6909 | sone | 0.7838 | kanu | 0.8118 |  |  |
| diabet | 0.6717 | title | 0.7704 | edberg | 0.7860 |  |  |
| crocodil | 0.6553 | nk | 0.7629 | besikt | 0.7789 |  |  |
| femal | 0.6354 | olivet | 0.7488 | inc | 0.7779 |  |  |
| cjd | 0.6320 | sgb | 0.7466 | riedel | 0.7621 |  |  |
| shuttle | 0.6267 | miki | 0.7438 | beat | 0.7538 |  |  |
| comput | 0.6153 | peapod | 0.7408 | hapoel | 0.7497 |  |  |
| robot | 0.6105 | antah | 0.7398 | skat | 0.7458 |  |  |