# Title: Account Stats

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Account | Train Size | Test Size | ML Recall | ES | Block(50) | Block(15) | model (15) | Email recall | Model (without blocking) |
| 70541 | 206422 | 51606 | 0.982 |  |  |  |  |  |  |
| 53524 | 49484 | 12371 | 0.969 | 0.93 | 0.97 | 0.96 | 0.94 | 0.97 | 0.95 |
| 56025 | 36424 | 9106 | 0.997 | 0.86 | 0.99 | 0.98 | 0.86 | 0.96 | 0.90 |
| 70700 | 14192 | 3548 | 0.870 | 0.42 | 0.48 | 0.45 | 0.42 | 0.50 |  |
| 8302 | 12648 | 3162 | 0.964 | 0.57 | 0.68 | 0.63 | 0.58 | 0.65 | 0.70 |
| 69198 | 944 | 236 | 0.917 | 0.88 | 0.932 | 0.93 | 0.93 | 0.93 |  |

Account id: 53524  
Block cutoff 20  
Number of secondaries to read   
 DescribeResult(nobs=4512, minmax=(0, 68), mean=8.17508865248227, variance=57.649230761369765, skewness=1.5162532587347959, kurtosis=2.95958747751755)  
Top 3  
Recall of positive pairs 0.944813829787234  
Total token: 453452  
Token less than size 20 : 442742

Account id: 53524

Block size 20

Number of secondaries to read

DescribeResult(nobs=4512, minmax=(0, 68), mean=8.17508865248227, variance=57.649230761369765, skewness=1.5162532587347959, kurtosis=2.95958747751755)  
Top 5  
Recall of positive pairs 0.9621010638297872  
Total token: 453452  
Token less than size 20 : 442742

Account id: 53524  
Block size 20

Number of secondaries to read

DescribeResult(nobs=4512, minmax=(0, 68), mean=8.17508865248227, variance=57.649230761369765, skewness=1.5162532587347959, kurtosis=2.95958747751755)  
Top 10  
Recall of positive pairs 0.9660904255319149  
Total token: 453452  
Token less than size 20 : 442742

Account id: 53524  
Block size 10

Number of secondaries to read

DescribeResult(nobs=4512, minmax=(0, 19), mean=3.799645390070922, variance=6.510059727914899, skewness=1.838925065850597, kurtosis=3.934045683415846)  
Top 10  
Recall of positive pairs 0.9625443262411347  
Total token: 453452  
Token less than size 10 : 435574  
Tokens of size of size 1: 271690

Account id: 53524  
Block size 50

Number of secondaries to read

DescribeResult(nobs=4512, minmax=(0, 167), mean=30.97517730496454, variance=473.5773043356586, skewness=1.0125922741676523, kurtosis=1.8213173013751618)  
Top 10  
Recall of positive pairs 0.9680851063829787  
Total token: 453452  
Token less than size 50 : 450145  
Tokens of size of size 1: 271690

---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Account id: 56025  
Block size 10

Number of secondaries to read

DescribeResult(nobs=1852, minmax=(0, 21), mean=5.558315334773218, variance=9.883685545026738, skewness=0.8893058500054926, kurtosis=0.7935155189100271)  
Top 10  
Recall of positive pairs 0.9730021598272138  
Total token: 398582  
Token less than size 10 : 381968  
Tokens of size of size 1: 220961

Account id: 56025  
Block size 10

Number of secondaries to read

DescribeResult(nobs=1852, minmax=(0, 21), mean=5.558315334773218, variance=9.883685545026738, skewness=0.8893058500054926, kurtosis=0.7935155189100271)  
Top 3  
Recall of positive pairs 0.8639308855291576  
Total token: 398582  
Token less than size 10 : 381968  
Tokens of size of size 1: 220961

Account id: 56025  
Block size 5

Number of secondaries to read

DescribeResult(nobs=1852, minmax=(0, 10), mean=2.8461123110151187, variance=2.031951382301085, skewness=-0.18132148502785847, kurtosis=0.4336441181716677)  
Top 3  
Recall of positive pairs 0.8250539956803455  
Total token: 398582  
Token less than size 5 : 355095  
Tokens of size of size 1: 220961

---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Account id: 8302  
Block size 20

Number of secondaries to read

DescribeResult(nobs=636, minmax=(0, 42), mean=6.45440251572327, variance=38.733350666072404, skewness=1.7061275864000933, kurtosis=3.551921015353395)  
Top 3  
Recall of positive pairs 0.5754716981132075  
Total token: 969025  
Token less than size 20 : 962753  
Tokens of size of size 1: 820986

Account id: 8302  
Block size 20

Number of secondaries to read

DescribeResult(nobs=636, minmax=(0, 42), mean=6.45440251572327, variance=38.733350666072404, skewness=1.7061275864000933, kurtosis=3.551921015353395)  
Top 10  
Recall of positive pairs 0.6367924528301887  
Total token: 969025  
Token less than size 20 : 962753  
Tokens of size of size 1: 820986

Account id: 8302  
Block size 50

Number of secondaries to read

DescribeResult(nobs=636, minmax=(0, 94), mean=14.995283018867925, variance=294.6818674788293, skewness=1.7310061028860941, kurtosis=3.125585927726325)  
Top 3  
Recall of positive pairs 0.5660377358490566  
Total token: 969025  
Token less than size 50 : 966378  
Tokens of size of size 1: 820986

---------------------------------------------------------------------------------------------------------------------------------------------------------------------

Account id: 70700  
Block size 100

Number of secondaries to read

DescribeResult(nobs=724, minmax=(0, 204), mean=25.591160220994475, variance=1049.5753574348746, skewness=1.7734236770363594, kurtosis=3.5935132040430773)  
Top 3  
Recall of positive pairs 0.4350828729281768  
Total token: 2216394  
Token less than size 100 : 2213523  
Tokens of size of size 1: 1942573

-------------------------------------------------------

Account id: 69198  
Block size 15

Number of secondaries to read

DescribeResult(nobs=44, minmax=(2, 44), mean=16.022727272727273, variance=81.51109936575054, skewness=0.8169431058824503, kurtosis=1.003928673288339)  
Top 3  
Recall of positive pairs 0.9318181818181818  
Total token: 29397  
Token less than size 15 : 28859  
Tokens of size of size 1: 22759

---------------------------------------------------------------

Number of accounts in leads table = 103584

Number of accounts with > 1 leads in leads table = 103538

Number of accounts with > 10 leads in leads table = 17774

Average account size = ~3760

Max account size = 3438530 (account\_id:47 Freshdesk)

------------------------------------------------------------------

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **account\_id** | **Account size** | **Total no. of blocks** | **No. Of blocks of size 1** | **No. Of blocks of size > 1** | **Number of leads stored in data store** |
| 53524 | 293301 | 453452 | 271690 | 181762 | 2468749 |
| 56025 | 216777 | 398582 | 220961 | 177621 | 2110788 |
| 70700 | 985863 | 2216394 | 1942573 | 273821 | 5489042 |
| 8302 | 396508 | 969025 | 820986 | 148039 | 2879162 |
| 69198 | 187636 | 29397 | 22759 | 6638 | 796065 |

Link to account stats:

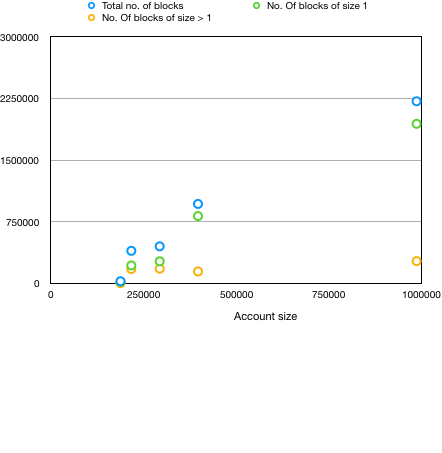
[Account 8302](https://docs.google.com/spreadsheets/d/1bY0lvrDaLUkJzMjcJ6en3Ht3noIPls4FD30SqnRPw6M/edit#gid=0)

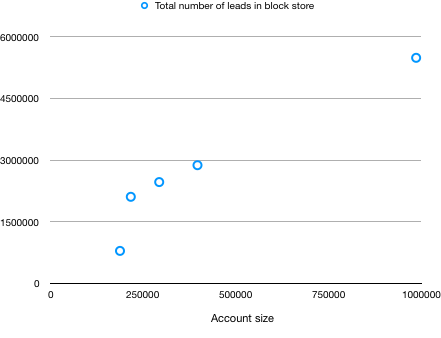
[Account 53524](https://docs.google.com/spreadsheets/d/1FEiDdweS-T-xpWf-iucQlaIiq3dmIyOmSVLD2WXi-cY/edit#gid=0)

[Account 69198](https://docs.google.com/spreadsheets/d/1aMvk3e5CgxZcG-86o_Cy7qfV1IKM5ramGR8Ox4W0ebc/edit#gid=0)

[Account 70700](https://docs.google.com/spreadsheets/d/1mmpnmC9IRzcKumRFD-R2Q1N4SGHoBUryPKOswdOA5e4/edit#gid=0)

[Account 56025](https://docs.google.com/spreadsheets/d/1idwDGac2W42tj0dVVYwv7IGj3M6o7fXio8m1kjYb2vE/edit#gid=0)





Storage requirement calculation for redis(block data storage):

Based on graph1, we projected the no. of blocks per account to be 1.5 times the account size

Based on graph2, we projected the no. of leads in block storage per account to be 7 times the account size.

[Link to block storage data calculation](https://docs.google.com/spreadsheets/d/1H0f2GE_cNWVTn9_gJOt_bIJ1uFb83jrPKrfy1Bo7TPA/edit#gid=0)

(Projected) Total no. of leads stored in all blocks of all accounts = 467887210

Each lead\_id will be of size ~10 bytes.

(Projected) Total 4678872100 bytes(4.67GB) for storing the lead\_ids

(Projected) Total no. of blocks for all accounts = 100266041

Each key will have lead/contact\_accountid\_blockvalue which will be of size ~20(6 + 5 + 7) bytes.

(Projected) Total 2005320820 bytes(2GB) for storing the keys

Overall 6.67GB for just the leads block store.

Contacts would be similar size or less.

So, 14GB for storing all the existing leads and contacts in block storage.