

Sensitivity Analysis

How do scoring algorithm adjustments affect candidate rankings

Research Associates: Joanna Zuo, Jianing Yu

Research Advisor: Richard Huntsinger

Industry Collaboration Lead: Anna Burns

Date: 12/17/2021



“
How sensitive is the scoring algorithm to cluster weight changes?



Data



Data Summary

Data used is entirely synthetic and random, but replicates the identical structure of the new Brainwahve data

Simplifying assumptions:

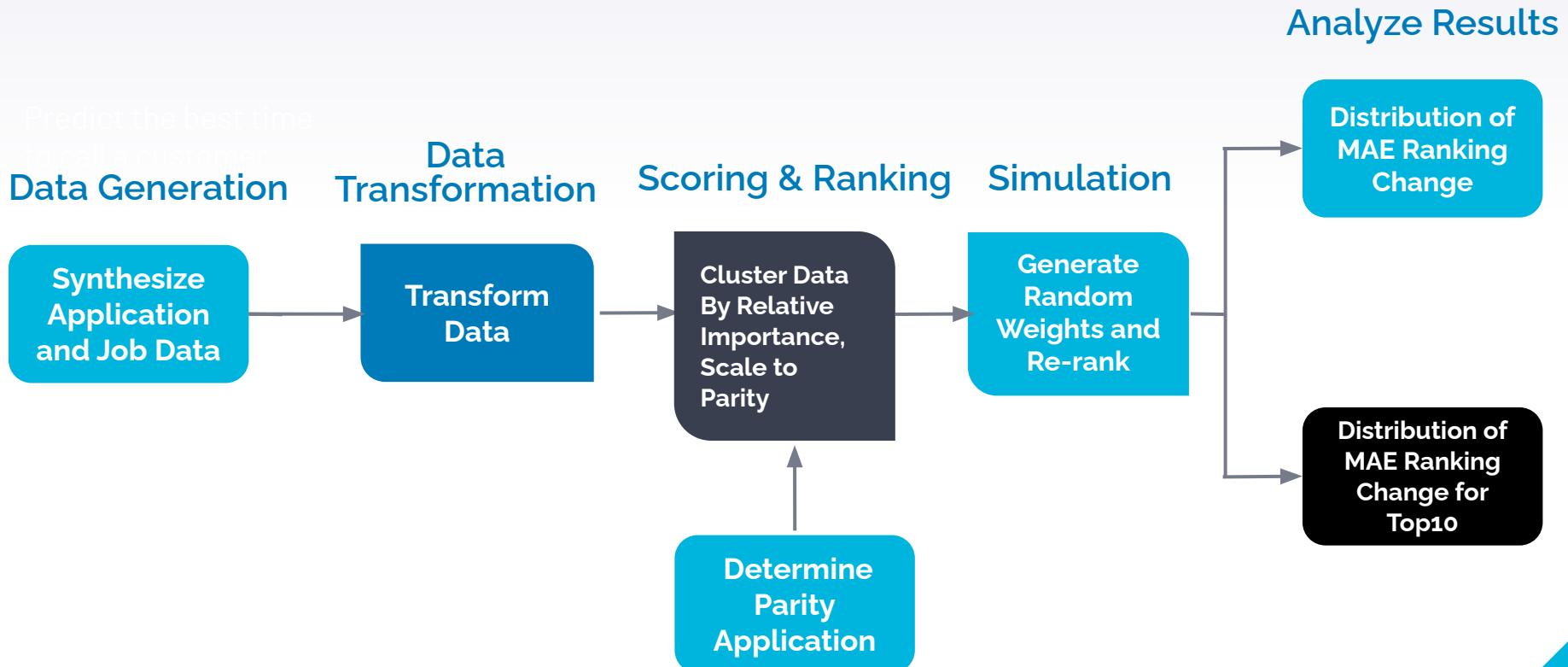
- Score all applications against one job
- Ignore disqualifying
- Treat job functions as categorical selections rather than text

Methodology



Methodology

Replicating Algorithm Scoring Process



Data Generation

- Generating Job and Application data tables
 - Data ranges accurate to software
- Creating parity application data
 - Using the job's requirement dataframes
 - Add configurable parity years' experience where necessary

Table Name	Application_Id	Title	Yrs_Exp	Proficiency	Last_Used	Ranking
application_sector.csv	X	X	X	X	X	
application_department.csv	X	X	X	X	X	
application_positions.csv	X	X	X	X	X	
application_areas_of_expertise.csv	X	X	X	X	X	
application_management_systems.csv	X	X	X	X	X	
application_carriers.csv	X	X	X	X	X	
application_computer_skills.csv	X	X	X	X	X	
application_languages.csv	X	X	X	X	X	
application_job_functions.csv	X	X	X	X	X	
application_degree_level.csv	X	X				
application_licenses.csv	X	X				
application_certs.csv	X	X				
application_candidate_traits.csv	X	X				X
application_company_traits.csv	X	X				X

Table Name	Job_Id	Title	Yrs_Exp	Proficiency	Importance	Ranking	Disqualifying
job_sector.csv	X	X	X		X		
job_department.csv	X	X	X		X		
job_positions.csv	X	X	X		X		
job_areas_of_expertise.csv	X	X		X	X		
job_management_systems.csv	X	X		X	X		
job_carriers.csv	X	X		X	X		
job_computer_skills.csv	X	X		X	X		
job_languages.csv	X	X		X	X		
job_job_functions.csv	X	X		X	X		
job_degree_level.csv	X	X			X		X
job_licenses.csv	X	X			X		X
job_certs.csv	X	X			X		X
job_company_traits.csv	X	X				X	
job_candidate_traits.csv	X	X				X	

Data Transformation

Replicating the algorithm scoring process on a large scale

- Filter tables based on title to find applications that match job specifications
- Apply feature transformation functions to the matched application / job data (sigmoid, MAE, proficiency comparison, etc.)
- Challenges
 - Ranked adjectives approach
 - Creation of worst-possible application ranking

Scoring and Ranking

- Cluster Data
 - For loop to merge tables within the same cluster
- For each application:
 - Scale raw scores using Relative Importance within cluster
 - Sum importance-weighted scores within each cluster
 - Scale importance-weighted cluster scores against parity application
 - Sum all cluster scores together to create final score
- Rank all applications by final score

Simulation

1. Set up baseline weights, e.g. [0.2, 0.2, 0.2, 0.2, 0.2]
2. Score and rank applications using baseline weights
3. 10,000 times repeat the following:
 - a. Randomly generate new weights (Details show in next slide)
 - b. Score and rank applications using randomly generated weights
 - c. Calculate MAE ranking change for each new weights → average of the absolute value of how much rankings for all applications changed in total due to the random weights compared to the baseline
4. Create visualizations for all 10,000 MAEs calculated above
5. Repeat the entire process, experimenting with different baseline weights and hyperparameters

► Simulation - Randomly Generating New Weights

- Start with baseline weights
- Use random normal distributions to generate new cluster weights that slightly deviate from the baseline
- Start with a different cluster each time to avoid bias and ensure the weights still sum to 100%

Analysis Platform



Analysis Platform

Combine

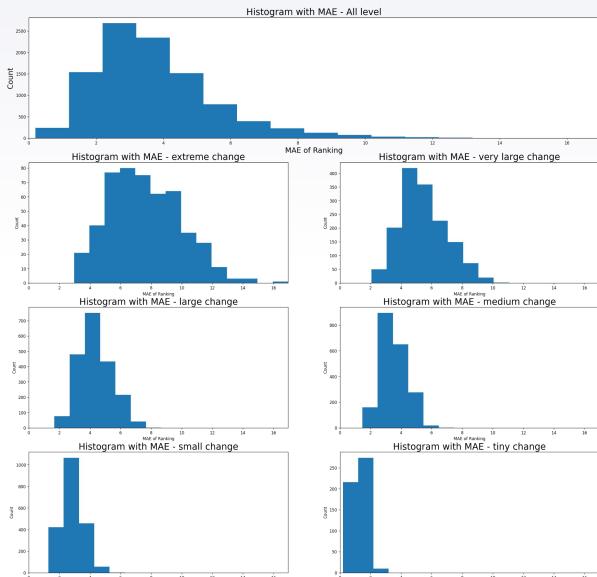
Approximately 1000 lines
of python code

Results

- ❖ Baseline Weights 1: [0.2, 0.2, 0.2, 0.2, 0.2]
- ❖ Baseline Weights 2: [0.2018, 0.3028, 0.1014, 0.1821, 0.2119]
- ❖ Baseline Weights 3: [0.3, 0.2, 0.3, 0.1, 0.1]
- ❖ Top 10 analysis

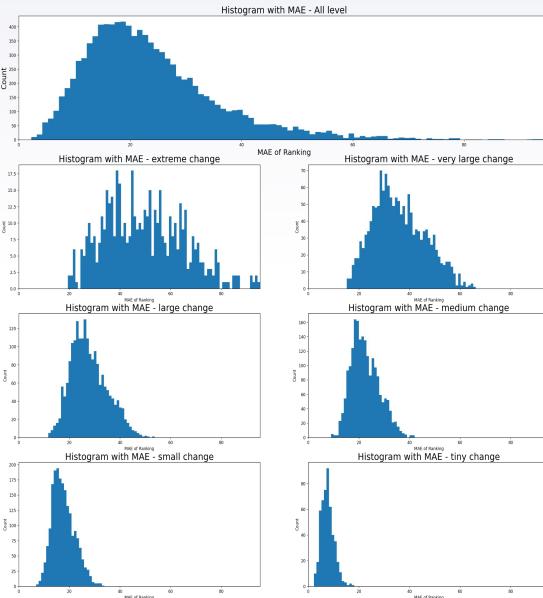


Baseline 1: Vary zoom level



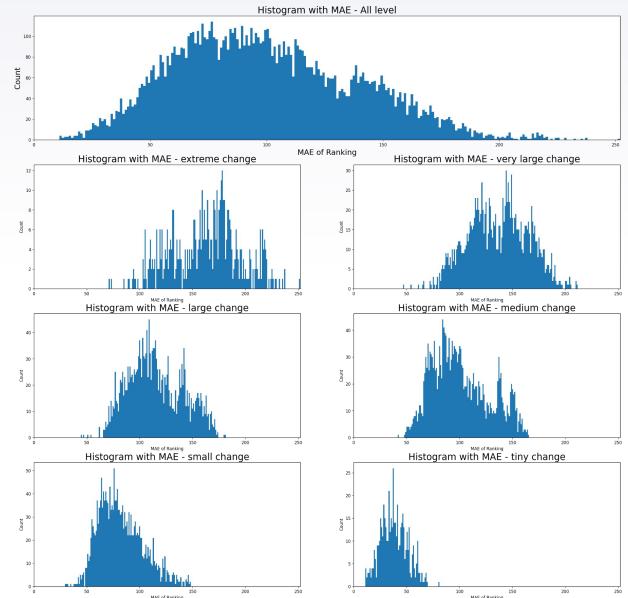
Zoom in - 0.003

Weight change from 0.00015 to 0.0099
Mean Ranking MAE (All) : 3.727



Normal - 0.02

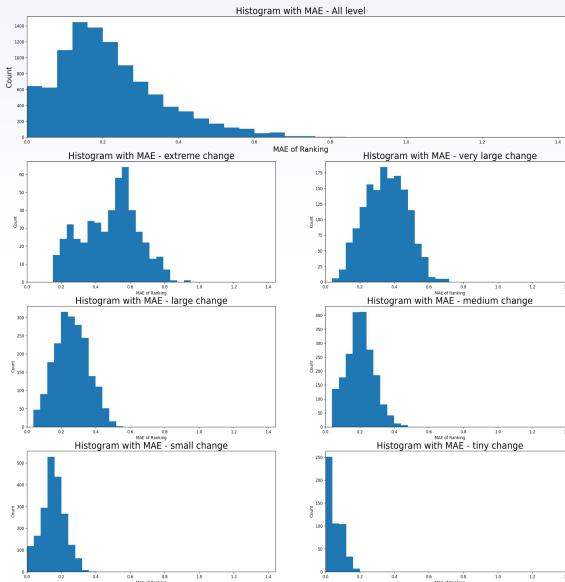
Weight change from 0.0015 to 0.0702
Mean Ranking MAE (All) : 23.743



Zoom out - 0.1

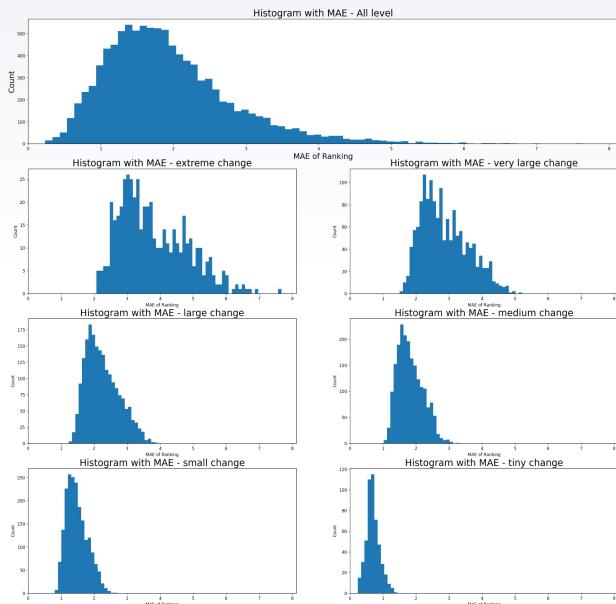
Weight change from 0.009 to 0.285
Mean Ranking MAE (All) : 99.663

Baseline 1: Vary number of applications



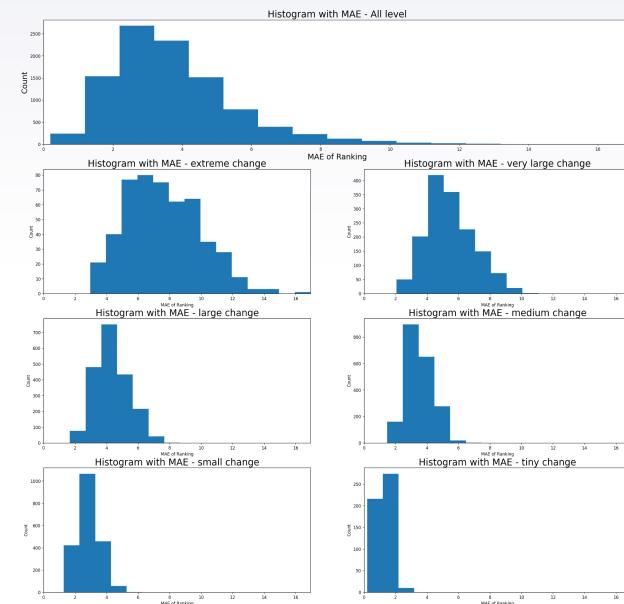
50 random applications

Weight change from 0.0002 to 0.0098 - std : 0.003
Mean Ranking MAE (All) : 0.225



500 random applications

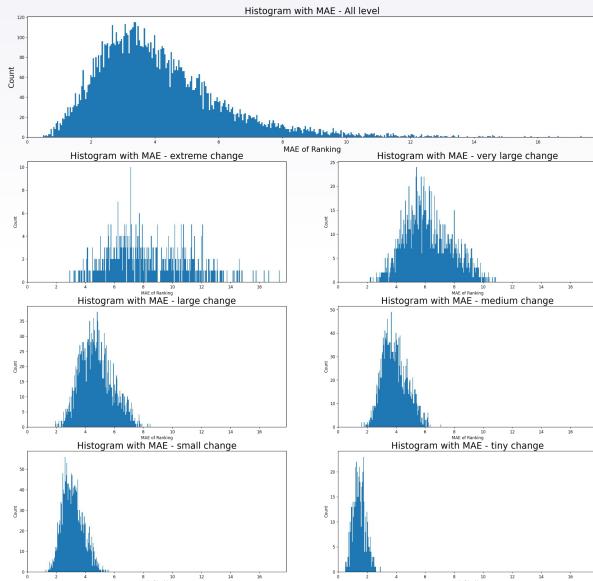
Weight change from 0.0002 to 0.0095 - std : 0.003
Mean Ranking MAE (All) : 2.155



1000 random applications

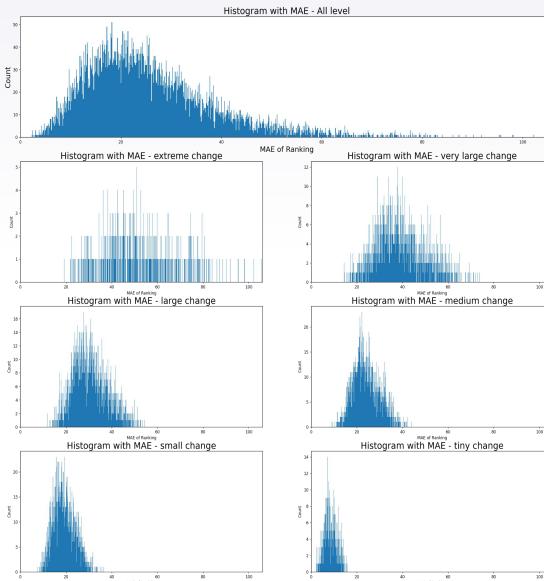
Weight change from 0.00015 to 0.0099 - std : 0.003
Mean Ranking MAE (All) : 4.137

Baseline 2: Vary zoom level



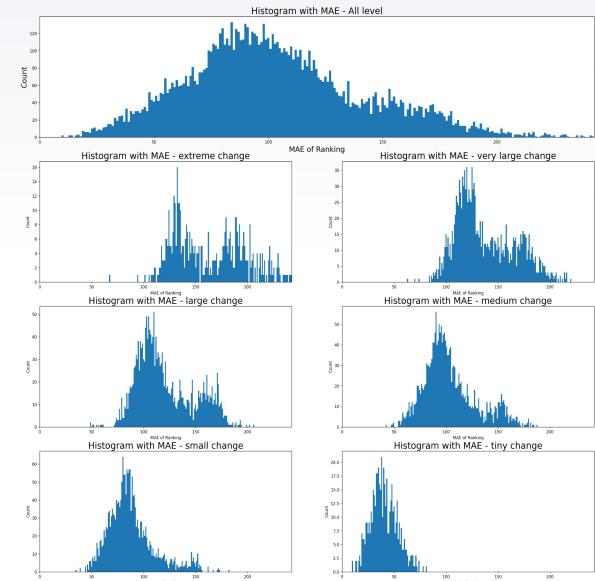
Zoom in - 0.003

Weight change from 0.00029 to 0.0091
Mean Ranking MAE (All) : 4.138



Normal - 0.02

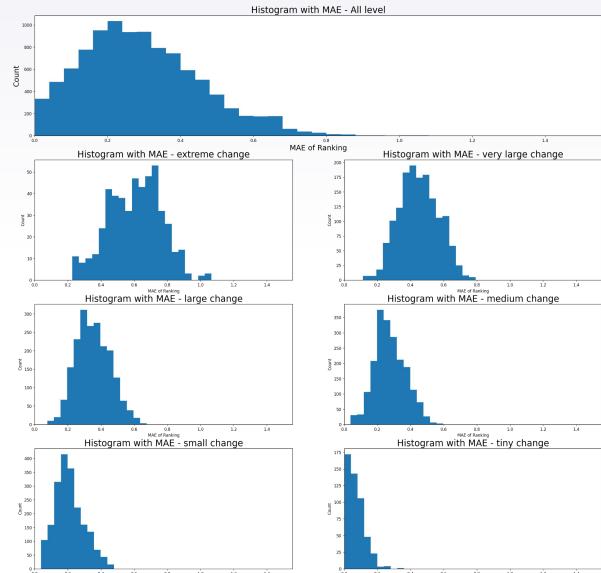
Weight change from 0.0012 to 0.0684
Mean Ranking MAE (All) : 25.814



Zoom out - 0.1

Weight change from 0.0054 to 0.2554
Mean Ranking MAE (All) : 103.038

Baseline 2: Vary number of applications



50 random applications

Weight change from 0.00018 to 0.0098 - std : 0.003
Mean Ranking MAE (All) : 0.296

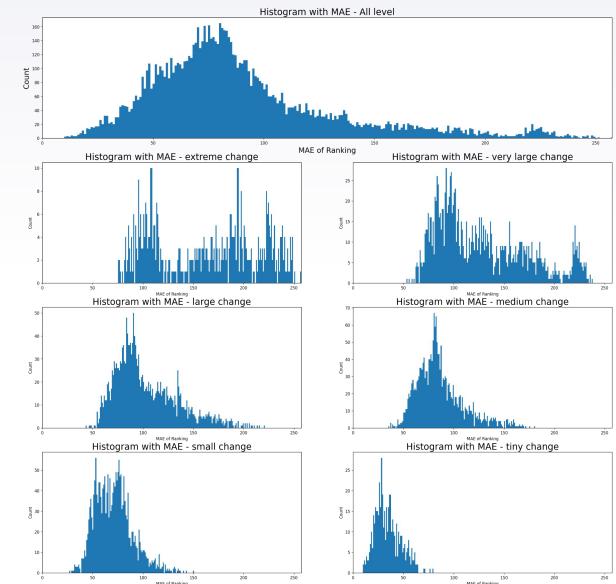
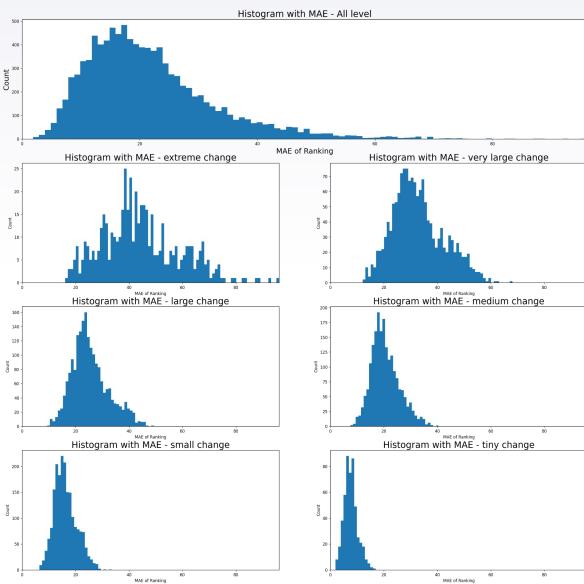
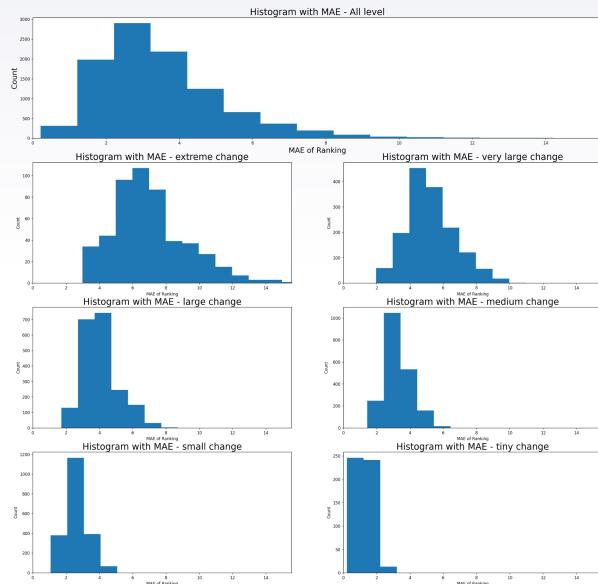
500 random applications

Weight change from 0.0003 to 0.0102 - std : 0.003
Mean Ranking MAE (All) : 1.932

1000 random applications

Weight change from 0.00029 to 0.0091 - std : 0.003
Mean Ranking MAE (All) : 3.727

Baseline 3: Vary zoom level



Zoom in - 0.003

Weight change from 0.00015 to 0.0096
Mean Ranking MAE (All) : 3.49

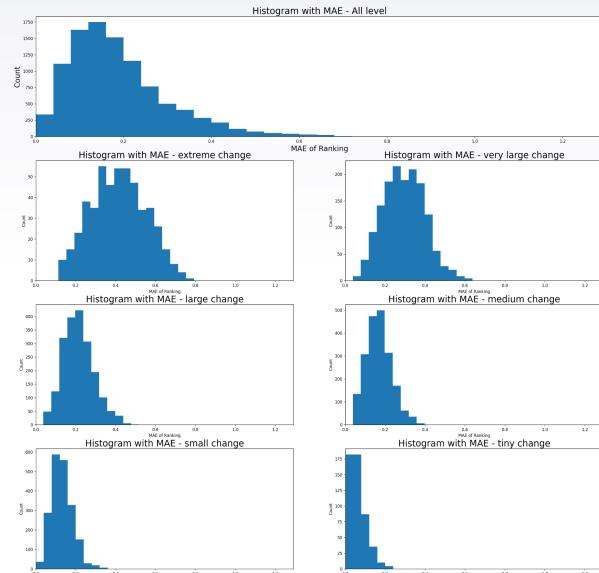
Normal - 0.02

Weight change from 0.0013 to 0.0655
Mean Ranking MAE (All) : 21.75

Zoom out - 0.1

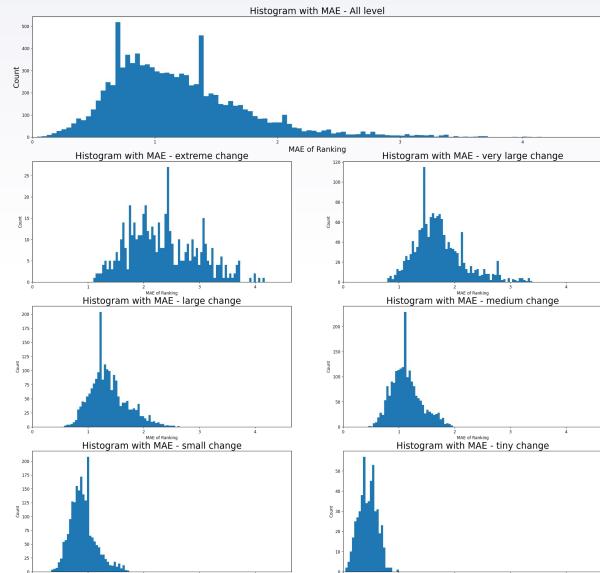
Weight change from 0.0061 to 0.24
Mean Ranking MAE (All) : 89.934

Baseline 3: Vary number of applications



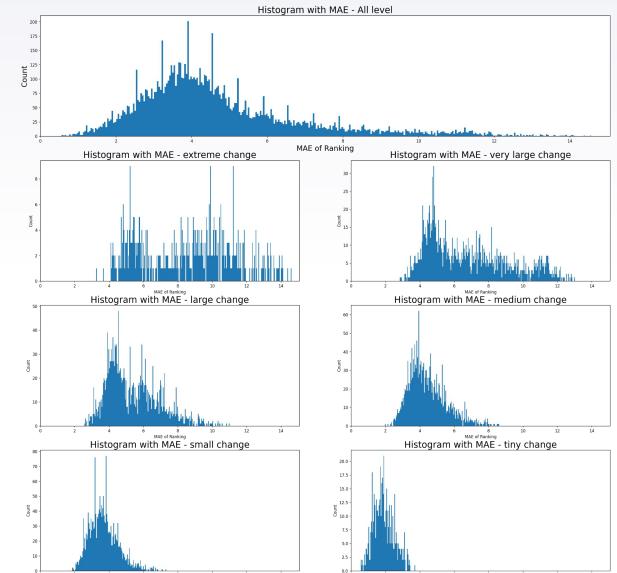
50 random applications

Weight change from 0.0002 to 0.0099 - std : 0.003
Mean Ranking MAE (All) : 0.1951



500 random applications

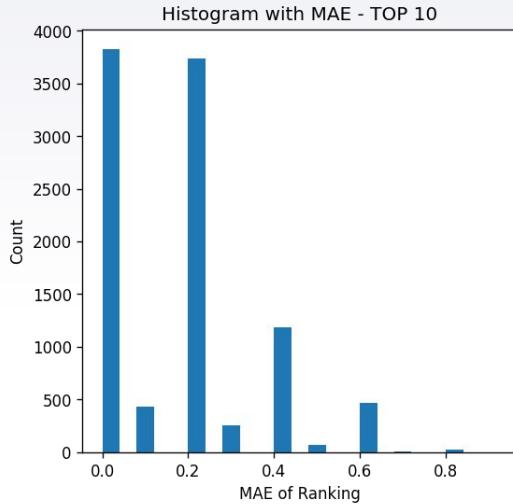
Weight change from 0.00016 to 0.0114 - std : 0.003
Mean Ranking MAE (All) : 1.852



1000 random applications

Weight change from 0.00015 to 0.0096 - std : 0.003
Mean Ranking MAE (All) : 3.485

Top 10 analysis

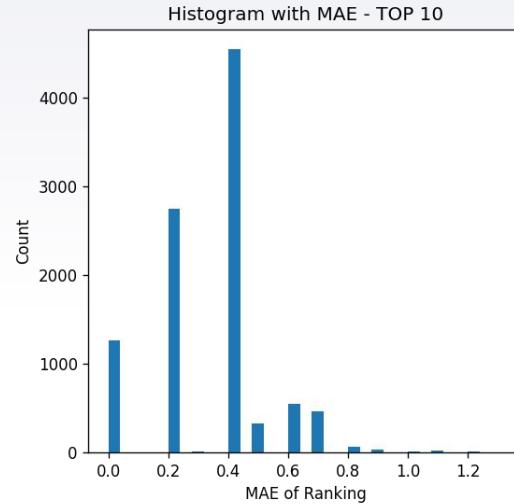


Baseline 1

mean of MAE for Ranking (TOP 10 change):
0.168

mode of MAE for Ranking (TOP 10 change):
[(0.0, 3824/10000)]

Weight change from 0.00015 to 0.0099 - std: 0.003

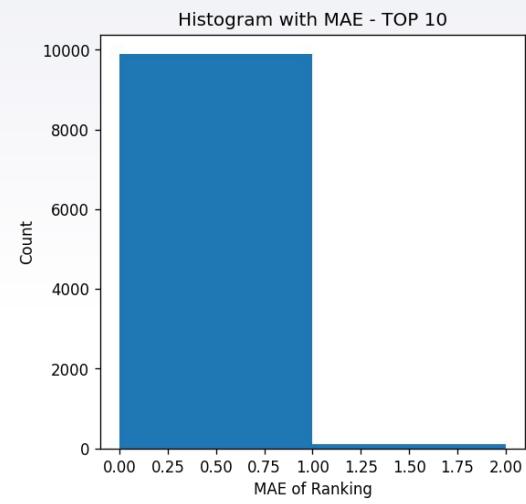


Baseline 2

mean of MAE for Ranking (TOP 10 change): mean of MAE for Ranking (TOP 10 change):
0.328

mode of MAE for Ranking (TOP 10 change): mode of MAE for Ranking (TOP 10 change):
[(0.4, 4544/10000)]

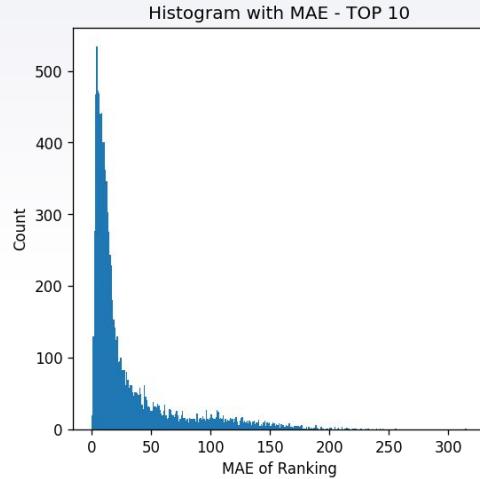
Weight change from 0.00029 to 0.0091 - std: 0.003



Baseline 3

Weight change from 0.00015 to 0.0096 - std: 0.003

Top 10 analysis

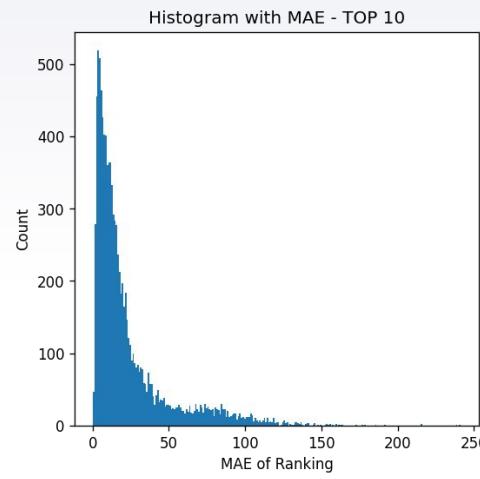


Baseline 1

mean of MAE for Ranking (TOP 10 change):
30.28

mode of MAE for Ranking (TOP 10 change):
[(4.5, 67/10000)]

Weight change from 0.0093 to 0.2852 - std: 0.01

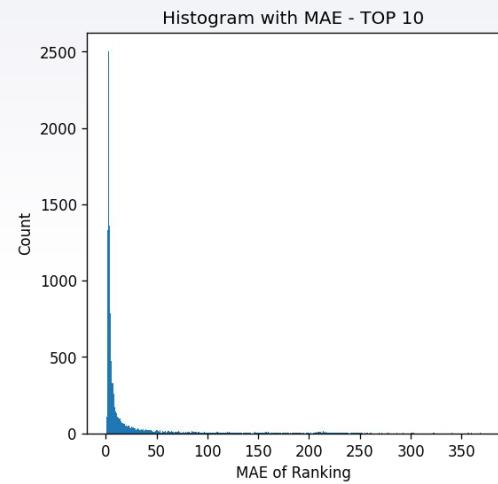


Baseline 2

mean of MAE for Ranking (TOP 10 change): mean of MAE for Ranking (TOP 10 change):
23.034

mode of MAE for Ranking (TOP 10 change): mode of MAE for Ranking (TOP 10 change):
[(4, 63/10000)]

Weight change from 0.0054 to 0.2554 - std: 0.01



Baseline 3

Weight change from 0.0061 to 0.24 - std: 0.001

Conclusion



Conclusion

For the top 10 applications:

- Small adjustments to the scoring algorithm (average weight change of less than 1%) cause only a small change to the overall application rankings (ranks change less than about 0.3 positions on average), independent on the baseline weights.
- Medium adjustments to the scoring algorithm (average weight change of about 10%) cause a moderate change to the overall application rankings (ranks change 8-10 positions on average), depending on the baseline weights.
- Large adjustments to the scoring algorithm (average weight change of more than about 25%) cause a large change to the overall application rankings (ranks change 18-30 positions on average), depending on the baseline weights.

Recommendations

This project has quantified the stability of various baseline weights (how resilient rankings are to weight changes). It has not determined the quality of the baseline weights (are the applications ranked at the top the best for the job?).

Next steps for Brainwahve:

- Validate the baseline weights to create an informed plan for future weight adjustments.
 - If the baseline weights are good, it is safe to make changes (<1%) to refine the algorithm without causing volatility in the rankings.
 - If the baseline is not good, larger changes are needed in order to improve the rankings.
- Extend methodology beyond weights to consider sensitivity to data feature transformation functions.
- Extend methodology to make synthetic data more realistic.
- Extend methodology to multiple jobs.

▶ Personal Takeaways

- Functionality - converting code to one function with configurable parameters to change feature transformations
- Shuffle before or after

THANK YOU!



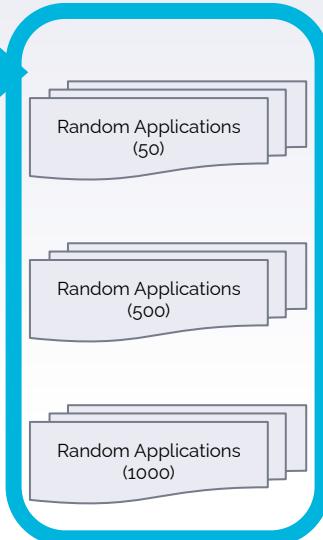
Questions?



Appendix



Data set



Baseline

1: Evenly - all 0.2

2: provided by Anna

3: Our Assumption

Score and rank applications



Std for generate random weight

1: small std. - 0.003

2: median std. - 0.02

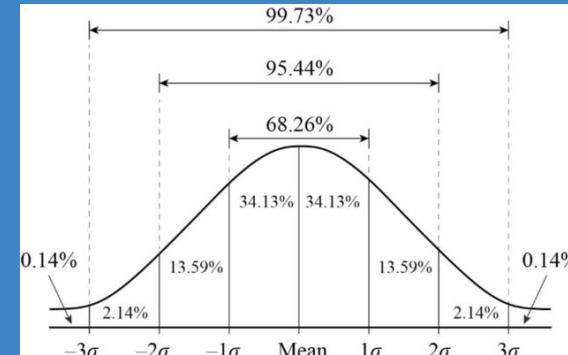
3: large std. - 0.1

10,000 times repeat

Output

Test set !

MAE
for
different level of changes



Baseline 1: 1000 random applications with different standard deviation all 0.2

std:0.003

```
mean of MAE for Ranking (All level): 3.727  
mean of MAE for Ranking (Extreme change): 7.595  
mean of MAE for Ranking (very large change): 5.478  
mean of MAE for Ranking (large change): 4.349  
mean of MAE for Ranking (medium change): 3.519  
mean of MAE for Ranking (small change): 2.851  
mean of MAE for Ranking (tiny change): 1.283
```

std:0.02

```
mean of MAE for Ranking (All level): 23.743  
mean of MAE for Ranking (Extreme change): 49.509  
mean of MAE for Ranking (very large change): 35.892  
mean of MAE for Ranking (large change): 27.736  
mean of MAE for Ranking (medium change): 22.499  
mean of MAE for Ranking (small change): 17.655  
mean of MAE for Ranking (tiny change): 7.635
```

std:0.1

```
mean of MAE for Ranking (All level): 99.663  
mean of MAE for Ranking (Extreme change): 162.770  
mean of MAE for Ranking (very large change): 136.553  
mean of MAE for Ranking (large change): 115.232  
mean of MAE for Ranking (medium change): 101.021  
mean of MAE for Ranking (small change): 82.194  
mean of MAE for Ranking (tiny change): 38.246
```

```
mode of MAE for Ranking (All level): [(2.804, 8)]  
mode of MAE for Ranking (Extreme change): [(9.021, 2)]  
mode of MAE for Ranking (very large change): [(5.266, 4)]  
mode of MAE for Ranking (large change): [(3.63, 5)]  
mode of MAE for Ranking (medium change): [(2.89, 5)]  
mode of MAE for Ranking (small change): [(2.877, 6)]  
mode of MAE for Ranking (tiny change): [(1.185, 5)]
```

```
mode of MAE for Ranking (All level): [(28.537, 4)]  
mode of MAE for Ranking (Extreme change): [(35.135, 2)]  
mode of MAE for Ranking (very large change): [(37.694, 2)]  
mode of MAE for Ranking (large change): [(23.566, 3)]  
mode of MAE for Ranking (medium change): [(27.017, 3)]  
mode of MAE for Ranking (small change): [(11.942, 3)]  
mode of MAE for Ranking (tiny change): [(5.994, 2)]
```

```
mode of MAE for Ranking (All level): [(114.833, 4)]  
mode of MAE for Ranking (Extreme change): [(177.935, 2)]  
mode of MAE for Ranking (very large change): [(137.647, 2)]  
mode of MAE for Ranking (large change): [(114.833, 3)]  
mode of MAE for Ranking (medium change): [(87.353, 2)]  
mode of MAE for Ranking (small change): [(77.139, 2)]  
mode of MAE for Ranking (tiny change): [(47.192, 1)]
```

Baseline 1: standard deviation is 0.003 with different size of random applications all 0.2

Applications:
50

```
mean of MAE for Ranking (All level): 0.225
mean of MAE for Ranking (Extreme change): 0.482
mean of MAE for Ranking (very large change): 0.351
mean of MAE for Ranking (large change): 0.265
mean of MAE for Ranking (medium change): 0.209
mean of MAE for Ranking (small change): 0.162
mean of MAE for Ranking (tiny change): 0.064
```

Applications:
500

```
mean of MAE for Ranking (All level): 2.155
mean of MAE for Ranking (Extreme change): 4.29
mean of MAE for Ranking (very large change): 3.190
mean of MAE for Ranking (large change): 2.511
mean of MAE for Ranking (medium change): 2.038
mean of MAE for Ranking (small change): 1.647
mean of MAE for Ranking (tiny change): 0.776
```

Applications:
1000

```
mean of MAE for Ranking (All level): 4.137
mean of MAE for Ranking (Extreme change): 8.365
mean of MAE for Ranking (very large change): 6.212
mean of MAE for Ranking (large change): 4.757
mean of MAE for Ranking (medium change): 3.912
mean of MAE for Ranking (small change): 3.135
mean of MAE for Ranking (tiny change): 1.474
```

```
mode of MAE for Ranking (All level): [(0.151, 1446)]
mode of MAE for Ranking (Extreme change): [(0.567, 64)]
mode of MAE for Ranking (very large change): [(0.339, 184)]
mode of MAE for Ranking (large change): [(0.226, 315)]
mode of MAE for Ranking (medium change): [(0.226, 412)]
mode of MAE for Ranking (small change): [(0.151, 528)]
mode of MAE for Ranking (tiny change): [(0.038, 164)]
```

```
mode of MAE for Ranking (All level): [(1.788, 29)]
mode of MAE for Ranking (Extreme change): [(3.496, 4)]
mode of MAE for Ranking (very large change): [(3.136, 9)]
mode of MAE for Ranking (large change): [(2.412, 14)]
mode of MAE for Ranking (medium change): [(1.728, 15)]
mode of MAE for Ranking (small change): [(1.636, 15)]
mode of MAE for Ranking (tiny change): [(0.832, 9)]
```

```
mode of MAE for Ranking (All level): [(3.152, 12)]
mode of MAE for Ranking (Extreme change): [(10.582, 2)]
mode of MAE for Ranking (very large change): [(3.396, 4)]
mode of MAE for Ranking (large change): [(4.514, 6)]
mode of MAE for Ranking (medium change): [(4.176, 7)]
mode of MAE for Ranking (small change): [(2.664, 8)]
mode of MAE for Ranking (tiny change): [(1.276, 4)]
```

Baseline 2: 1000 random applications with different standard deviation [0.2018,0.3028,0.1014,0.1821,0.2119]

std:0.003

```
mean of MAE for Ranking (All level): 4.138  
mean of MAE for Ranking (Extreme change): 8.365  
mean of MAE for Ranking (very large change): 6.213  
mean of MAE for Ranking (large change): 4.757  
mean of MAE for Ranking (medium change): 3.913  
mean of MAE for Ranking (small change): 3.136  
mean of MAE for Ranking (tiny change): 1.474
```

std:0.02

```
mean of MAE for Ranking (All level): 25.814  
mean of MAE for Ranking (Extreme change): 51.804  
mean of MAE for Ranking (very large change): 39.277  
mean of MAE for Ranking (large change): 30.582  
mean of MAE for Ranking (medium change): 24.265  
mean of MAE for Ranking (small change): 19.209  
mean of MAE for Ranking (tiny change): 8.573
```

std:0.1

```
mean of MAE for Ranking (All level): 103.038  
mean of MAE for Ranking (Extreme change): 162.684  
mean of MAE for Ranking (very large change): 137.367  
mean of MAE for Ranking (large change): 120.537  
mean of MAE for Ranking (medium change): 103.515  
mean of MAE for Ranking (small change): 87.615  
mean of MAE for Ranking (tiny change): 40.197
```

```
mode of MAE for Ranking (All level): [(3.152, 12)]  
mode of MAE for Ranking (Extreme change): [(10.582, 2)]  
mode of MAE for Ranking (very large change): [(3.396, 4)]  
mode of MAE for Ranking (large change): [(4.514, 6)]  
mode of MAE for Ranking (medium change): [(4.176, 7)]  
mode of MAE for Ranking (small change): [(2.664, 8)]  
mode of MAE for Ranking (tiny change): [(1.276, 4)]
```

```
mode of MAE for Ranking (All level): [(26.336, 6)]  
mode of MAE for Ranking (Extreme change): [(70.242, 2)]  
mode of MAE for Ranking (very large change): [(33.224, 4)]  
mode of MAE for Ranking (large change): [(35.452, 4)]  
mode of MAE for Ranking (medium change): [(18.112, 4)]  
mode of MAE for Ranking (small change): [(17.88, 4)]  
mode of MAE for Ranking (tiny change): [(8.086, 2)]
```

```
mode of MAE for Ranking (All level): [(74.432, 4)]  
mode of MAE for Ranking (Extreme change): [(140.172, 2)]  
mode of MAE for Ranking (very large change): [(119.033, 2)]  
mode of MAE for Ranking (large change): [(102.758, 3)]  
mode of MAE for Ranking (medium change): [(80.126, 3)]  
mode of MAE for Ranking (small change): [(90.162, 2)]  
mode of MAE for Ranking (tiny change): [(37.418, 2)]
```

Baseline 2: standard deviation is 0.003 with different size of random applications [0.2018,0.3028,0.1014,0.1821,0.2119]

Applications: 5
0

```
mean of MAE for Ranking (All level): 0.296
mean of MAE for Ranking (Extreme change): 0.604
mean of MAE for Ranking (very large change): 0.450
mean of MAE for Ranking (large change): 0.356
mean of MAE for Ranking (medium change): 0.281
mean of MAE for Ranking (small change): 0.218
mean of MAE for Ranking (tiny change): 0.083
```

Applications: 500

```
mean of MAE for Ranking (All level): 1.932
mean of MAE for Ranking (Extreme change): 3.796
mean of MAE for Ranking (very large change): 2.874
mean of MAE for Ranking (large change): 2.255
mean of MAE for Ranking (medium change): 1.826
mean of MAE for Ranking (small change): 1.473
mean of MAE for Ranking (tiny change): 0.693
```

Applications: 1000

```
mean of MAE for Ranking (All level): 3.727
mean of MAE for Ranking (Extreme change): 7.595
mean of MAE for Ranking (very large change): 5.478
mean of MAE for Ranking (large change): 4.349
mean of MAE for Ranking (medium change): 3.519
mean of MAE for Ranking (small change): 2.850
mean of MAE for Ranking (tiny change): 1.283
```

```
mode of MAE for Ranking (All level): [(0.226, 1034)]
mode of MAE for Ranking (Extreme change): [(0.717, 53)]
mode of MAE for Ranking (very large change): [(0.415, 195)]
mode of MAE for Ranking (large change): [(0.302, 311)]
mode of MAE for Ranking (medium change): [(0.226, 375)]
mode of MAE for Ranking (small change): [(0.189, 415)]
mode of MAE for Ranking (tiny change): [(0.075, 143)]
```

```
mode of MAE for Ranking (All level): [(1.588, 39)]
mode of MAE for Ranking (Extreme change): [(3.48, 5)]
mode of MAE for Ranking (very large change): [(2.256, 9)]
mode of MAE for Ranking (large change): [(1.92, 13)]
mode of MAE for Ranking (medium change): [(1.568, 21)]
mode of MAE for Ranking (small change): [(1.404, 17)]
mode of MAE for Ranking (tiny change): [(0.604, 10)]
```

```
mode of MAE for Ranking (All level): [(2.804, 8)]
mode of MAE for Ranking (Extreme change): [(9.021, 2)]
mode of MAE for Ranking (very large change): [(5.266, 4)]
mode of MAE for Ranking (large change): [(3.63, 5)]
mode of MAE for Ranking (medium change): [(2.89, 5)]
mode of MAE for Ranking (small change): [(2.877, 6)]
mode of MAE for Ranking (tiny change): [(1.185, 5)]
```

Baseline 3: 1000 random applications with different standard deviation [0.3,0.2,0.3,0.1,0.1]

std:0.003

```
mean of MAE for Ranking (All level): 3.49
mean of MAE for Ranking (Extreme change): 7.118
mean of MAE for Ranking (very large change): 5.218
mean of MAE for Ranking (large change): 4.092
mean of MAE for Ranking (medium change): 3.296
mean of MAE for Ranking (small change): 2.626
mean of MAE for Ranking (tiny change): 1.25
```

std:0.02

```
mean of MAE for Ranking (All level): 21.75
mean of MAE for Ranking (Extreme change): 44.55
mean of MAE for Ranking (very large change): 33.13
mean of MAE for Ranking (large change): 25.34
mean of MAE for Ranking (medium change): 20.42
mean of MAE for Ranking (small change): 16.30
mean of MAE for Ranking (tiny change): 7.331
```

std:0.1

```
mean of MAE for Ranking (All level): 89.934
mean of MAE for Ranking (Extreme change): 163.305
mean of MAE for Ranking (very large change): 128.663
mean of MAE for Ranking (large change): 104.025
mean of MAE for Ranking (medium change): 85.988
mean of MAE for Ranking (small change): 71.2875485
mean of MAE for Ranking (tiny change): 35.158
```

```
mode of MAE for Ranking (All level): [(2.998, 14)]
mode of MAE for Ranking (Extreme change): [(5.212, 2)]
mode of MAE for Ranking (very large change): [(4.224, 5)]
mode of MAE for Ranking (large change): [(3.344, 6)]
mode of MAE for Ranking (medium change): [(2.998, 9)]
mode of MAE for Ranking (small change): [(1.958, 7)]
mode of MAE for Ranking (tiny change): [(1.314, 5)]
```

```
mode of MAE for Ranking (All level): [(13.644, 5)]
mode of MAE for Ranking (Extreme change): [(60.806, 1)]
mode of MAE for Ranking (very large change): [(24.496, 3)]
mode of MAE for Ranking (large change): [(22.11, 3)]
mode of MAE for Ranking (medium change): [(19.458, 4)]
mode of MAE for Ranking (small change): [(15.504, 4)]
mode of MAE for Ranking (tiny change): [(9.096, 2)]
```

```
mode of MAE for Ranking (All level): [(68.108, 4)]
mode of MAE for Ranking (Extreme change): [(168.645, 1)]
mode of MAE for Ranking (very large change): [(124.648, 2)]
mode of MAE for Ranking (large change): [(104.85, 2)]
mode of MAE for Ranking (medium change): [(68.108, 3)]
mode of MAE for Ranking (small change): [(59.51, 3)]
mode of MAE for Ranking (tiny change): [(26.652, 2)]
```

Baseline 3: standard deviation is 0.003 with different size of random applications [0.3,0.2,0.3,0.1,0.1]

Applications:
50

```
mean of MAE for Ranking (All level): 0.1951
mean of MAE for Ranking (Extreme change): 0.416
mean of MAE for Ranking (very large change): 0.299
mean of MAE for Ranking (large change): 0.223
mean of MAE for Ranking (medium change): 0.179
mean of MAE for Ranking (small change): 0.142
mean of MAE for Ranking (tiny change): 0.073
```

Applications:
500

```
mean of MAE for Ranking (All level): 1.852
mean of MAE for Ranking (Extreme change): 3.616
mean of MAE for Ranking (very large change): 2.76
mean of MAE for Ranking (large change): 2.162
mean of MAE for Ranking (medium change): 1.755
mean of MAE for Ranking (small change): 1.417
mean of MAE for Ranking (tiny change): 0.656
```

Applications:
1000

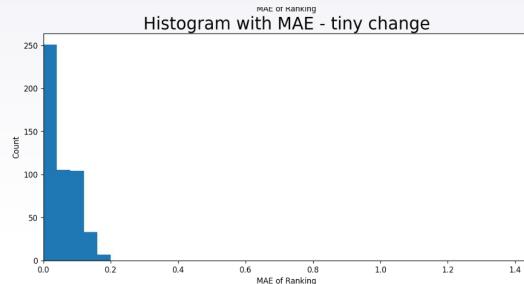
```
mean of MAE for Ranking (All level): 3.485
mean of MAE for Ranking (Extreme change): 6.995
mean of MAE for Ranking (very large change): 5.236
mean of MAE for Ranking (large change): 4.051
mean of MAE for Ranking (medium change): 3.257
mean of MAE for Ranking (small change): 2.637
mean of MAE for Ranking (tiny change): 1.273
```

```
mode of MAE for Ranking (All level): [(0.151, 1747)]
mode of MAE for Ranking (Extreme change): [(0.340, 55)]
mode of MAE for Ranking (very large change): [(0.264, 215)]
mode of MAE for Ranking (large change): [(0.226, 422)]
mode of MAE for Ranking (medium change): [(0.189, 499)]
mode of MAE for Ranking (small change): [(0.113, 587)]
mode of MAE for Ranking (tiny change): [(0.075, 182)]
```

```
mode of MAE for Ranking (All level): [(1.568, 33)]
mode of MAE for Ranking (Extreme change): [(3.548, 4)]
mode of MAE for Ranking (very large change): [(2.192, 11)]
mode of MAE for Ranking (large change): [(2.08, 15)]
mode of MAE for Ranking (medium change): [(1.776, 16)]
mode of MAE for Ranking (small change): [(1.216, 18)]
mode of MAE for Ranking (tiny change): [(0.696, 9)]
```

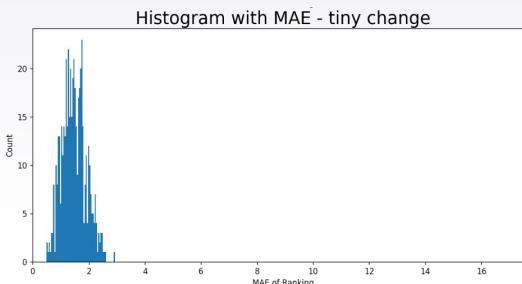
```
mode of MAE for Ranking (All level): [(2.736, 17)]
mode of MAE for Ranking (Extreme change): [(6.308, 5)]
mode of MAE for Ranking (very large change): [(4.768, 4)]
mode of MAE for Ranking (large change): [(4.35, 6)]
mode of MAE for Ranking (medium change): [(2.736, 9)]
mode of MAE for Ranking (small change): [(2.634, 8)]
mode of MAE for Ranking (tiny change): [(1.094, 5)]
```

Histogram for tiny change level MAEs with different baseline (std = 0.003, 1000 random applications)



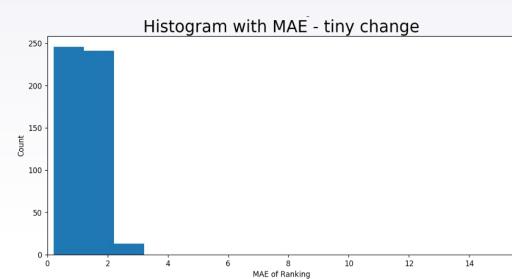
Baseline 1: Evenly

mean of MAE for Ranking (tiny change) :
0.064



Baseline 2: provided by Anna

mean of MAE for Ranking (tiny change) :
1.474



Baseline 3: our assumption

mean of MAE for Ranking (tiny change) :
1.273

mode of MAE for Ranking (tiny change) :
[(0.037, 164)]

mode of MAE for Ranking (TOP 10 change) :
[(0.4, 4544)]

mode of MAE for Ranking (tiny change) :
[(1.094, 5)]