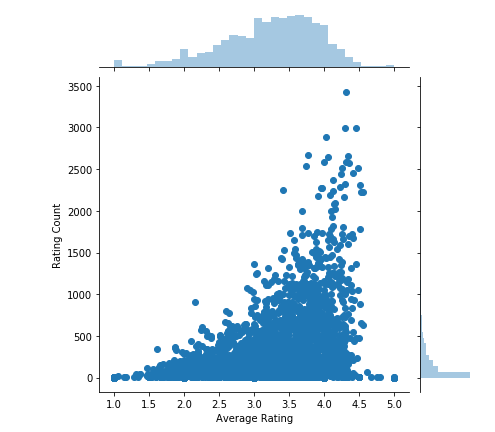
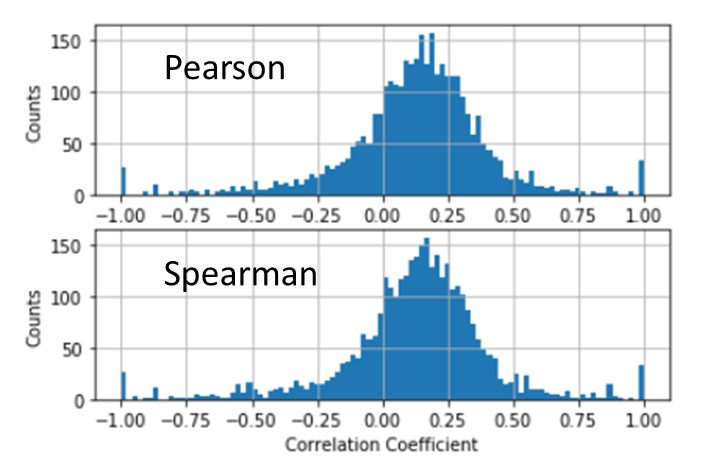
**Initial Approach: (Correlation)**





Observations:

There is a positive correlation between rating counts and average rating, which means a movie that is rated by many people is more likely to have a higher rating

Most movies are highly uncorrelated () due to large sparsity

Even if sparsity was not a problem, the results are a combined overview of all users, thus allowing no clear way to incorporate specific user preference

🡪 everyone gets the same recommendations

**Synthetic Data:**

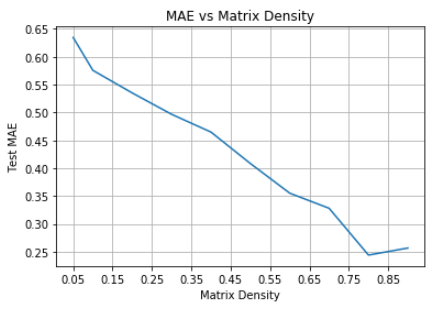
Impact of matrix density:

Dataset Shape: (100, 100)

Dataset rank: 11

Maximum MAE: 0.63 (Density = 5%)

Minimum MAE: 0.20 (Density = 80%)



Validity of Iterative SVD:

Density: 0.2

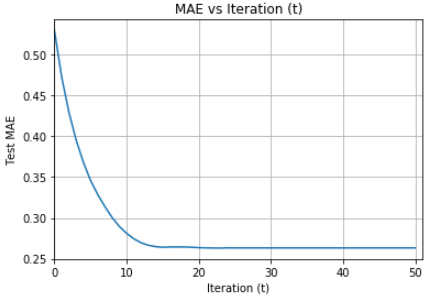
Train Num Count: 2026

Test Num Count: 7974  
Dataset Shape: (100, 100)

Dataset rank: 11

Maximum MAE: 0.53 (t = 1)

Minimum MAE: 0.26 (t = 50)



**Method 1: SVD**

Part1: Matrix completion

MAE results on different approaches

Filled using zeros: 3.58054281753226

First Iteration: 2.3107

Last Iteration: 1.0338

Filled using user means: 0.8293739398197625

First Iteration: 0.7486

Last Iteration: 0.7136

Filled using movie means: 0.7818449943655356

First Iteration: 0.7169

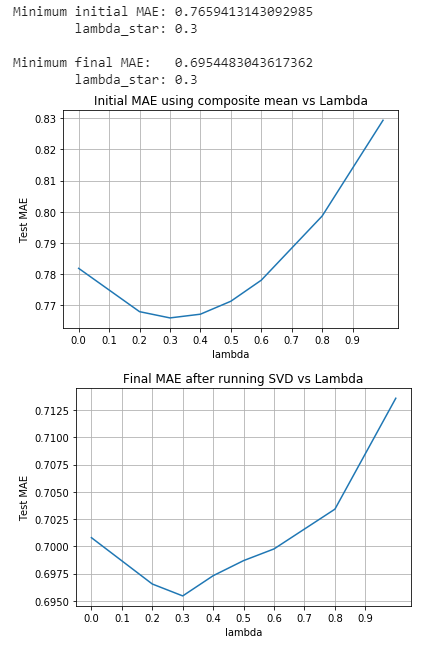
Last Iteration: 0.7008

Filled using average of both: 0.7712933153690582

First Iteration: 0.7118

Last Iteration: 0.6987

Part 2: Iterative SVD to tune weights between means



**Method 2: Regression**