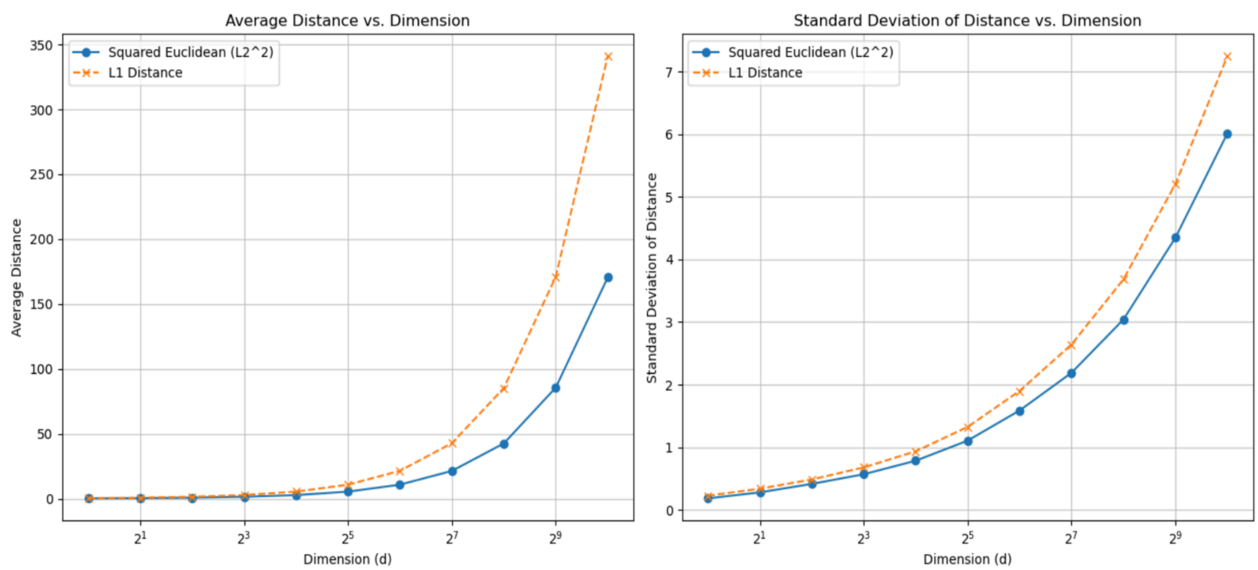

Observations on Distance and Standard Deviation in High Dimensions

After running the simulation and plotting the results, you'll typically observe the following trends:

1. Average Distance Increases with Dimension:

- **Observation:** Both the average Squared Euclidean (l22) distance and the average l1 distance will generally increase significantly as the dimension d increases.
- **Explanation:** As you add more dimensions, there are simply more "ways" for two random points to be different. Each additional dimension adds another component to the difference vector (xj−yj), contributing to the overall distance. Points that might be relatively close in low dimensions have much more "room to spread out" in higher dimensions.

2. Standard Deviation as a Proportion of the Mean Decreases (Distances Become More Concentrated):

- **Observation:** While the *absolute value* of the standard deviation might initially increase and then level off or even slightly decrease, the most striking observation is how the **standard deviation becomes a much smaller proportion of the average distance** as the dimension grows. In other words, the distribution of distances becomes very "tight" or concentrated around the mean.
- **Explanation (The "Curse of Dimensionality" in action):** This is a key manifestation of the "curse of dimensionality." In very high dimensions, almost all pairs of randomly

sampled points from a unit cube tend to be roughly the same "average" distance from each other.
    - **Intuition:** Imagine a point in 1D. It can be 0 or 1. Most points are close to the ends. In 2D, points cluster more towards corners. In very high dimensions, *all* points tend to be near the "corners" of the hypercube (i.e., their coordinates are mostly close to 0 or 1, rarely near 0.5 for all dimensions simultaneously). When two random points are chosen, they are both likely to be "corner-like" points, and the vector connecting them will have most of its components close to ±1 or 0. This makes their distances surprisingly similar.
    - **Practical Implication:** This phenomenon implies that in high-dimensional spaces, the concept of "proximity" or "neighborhoods" becomes less meaningful. All points become "far away" from each other, making tasks like clustering, nearest neighbor search, and density estimation very challenging because the *relative* differences in distance diminish.

3. Differences Between l22 and l1 Distances:

- **Observation:**
    - The **Squared Euclidean (l22) distances** will generally be larger than the l1 distances for the same points, and their growth rate might differ.
    - The **l1 distance** often shows a more linear growth with dimension compared to the l22 distance, which squares the differences.
- **Explanation:**
    - **l22 (Squared Euclidean):** This metric heavily penalizes large differences in any single dimension due to the squaring operation. Its value scales roughly with d times the average squared difference per dimension.
    - **l1 (Manhattan):** This metric sums the absolute differences. Its value scales roughly with d times the average absolute difference per dimension.

In summary: The experiment visually demonstrates that in high-dimensional spaces, random points become increasingly sparse and "isolated." The distances between them not only grow, but their distribution becomes remarkably uniform, with most points ending up approximately the same distance from each other. This counter-intuitive behavior is the core of the "curse of dimensionality."