

# **UNIVERSIDAD NACIONAL DE SAN AGUSTÍN**

**Facultad de Ingeniería, Producción y Servicios**

**Escuela Profesional de Ciencia de la Computación**



**Curso:**

**Estructuras de Datos Avanzados**

**Trabajo:**

**La maldición de la dimensionalidad**

**Profesor:**

**ROSA YULIANA GABRIELA PACCOTACYA**

**De:**

**- Jharold Alonso Mayorga Villena**

**AREQUIPA - PERÚ 2023**

# *Informe de análisis de histogramas de distancias en diferentes dimensiones*

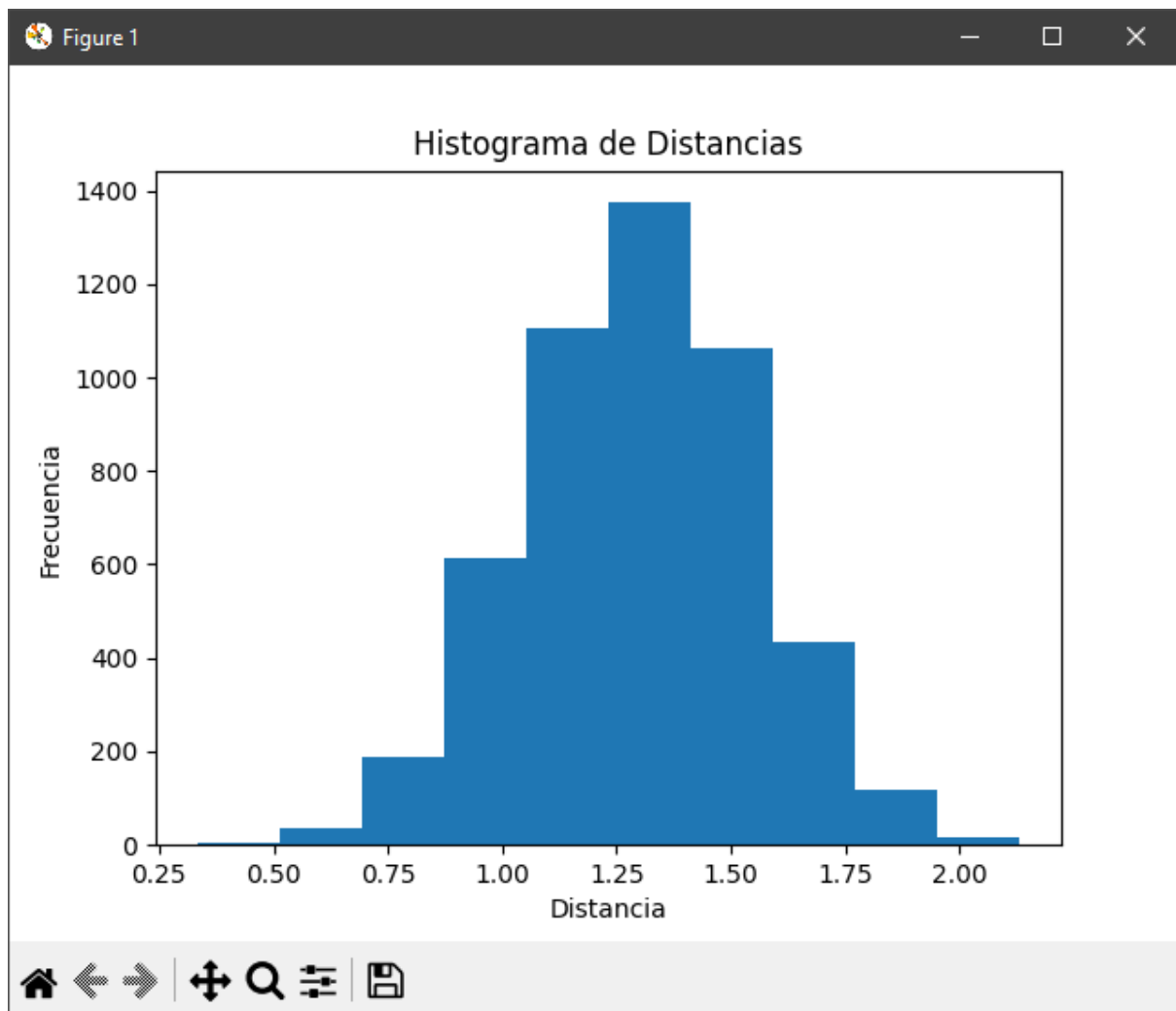
## **Introducción**

En este informe, se analizarán los histogramas de las distancias obtenidas a partir de un conjunto de puntos en diferentes dimensiones. El objetivo es examinar cómo cambia la distribución de las distancias a medida que se varía la dimensión del espacio. Se generaron histogramas para las dimensiones 10, 50, 100, 500, 1000, 2000 y 5000.

# Análisis de los histogramas

A continuación, se presentarán los análisis correspondientes a cada dimensión junto con las gráficas de los histogramas generados.

## 1. Dimensión 10:



```
= RESTART: C:\Users\pc\OneDrive\UNSA\VI SEMESTRE\Estructura de Datos Avanzados\L  
a maldicion de la dimensionalidad\histogram.py  
Media: 1.2955301717415293  
Mediana: 1.297823912407519  
Desviación estándar: 0.24899281086727293  
Los datos siguen una distribución normal.
```

- Distribución de las distancias:

En la dimensión 10, se observa una distribución de distancias similar al informe anterior, con una concentración de distancias cortas y una disminución gradual a medida que las distancias aumentan. También se aprecia un sesgo hacia la derecha, indicando la presencia de agrupaciones de puntos.

- Estadísticas descriptivas:

Media: 1.2955301717415293

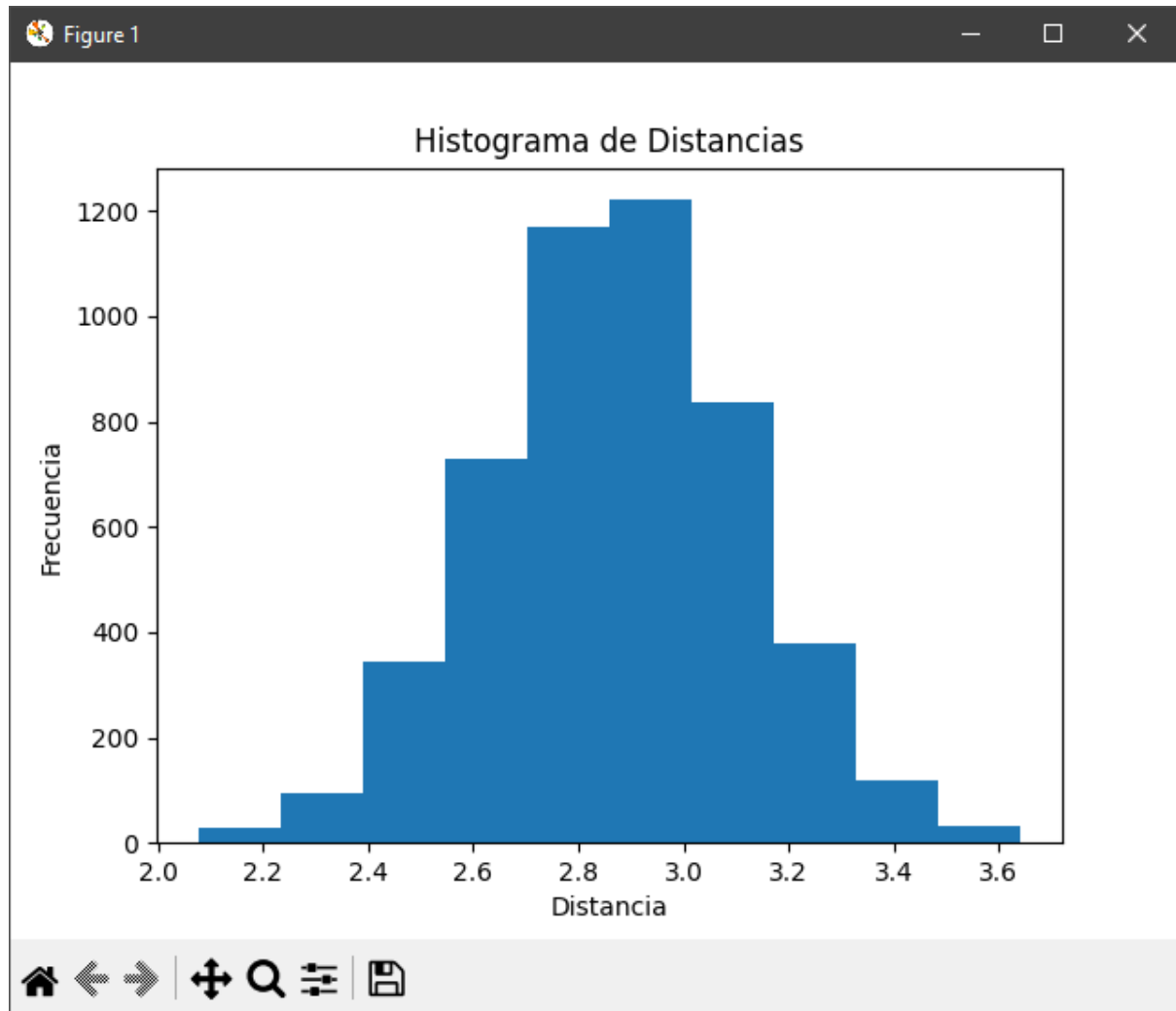
Mediana: 1.297823912407519

Desviación estándar: 0.24899281086727293

- Prueba de normalidad:

Los datos siguen una distribución normal.

## 2. Dimensión 50:



```
= RESTART: C:\Users\pc\OneDrive\UNSA\VI SEMESTRE\Estructura de Datos Avanzados\La maldición de la dimensionalidad\histogram.py
Media: 2.871252224857471
Mediana: 2.8715050118840644
Desviación estándar: 0.2437498559125815
Los datos siguen una distribución normal.
```

- Distribución de las distancias:

En la dimensión 50, se observa una distribución de distancias similar a la dimensión 10, con una concentración de distancias cortas y una disminución gradual a medida que las distancias aumentan. El sesgo hacia la derecha también es evidente, lo que sugiere la presencia de agrupaciones de puntos.

- Estadísticas descriptivas:

Media: 2.871252224857471

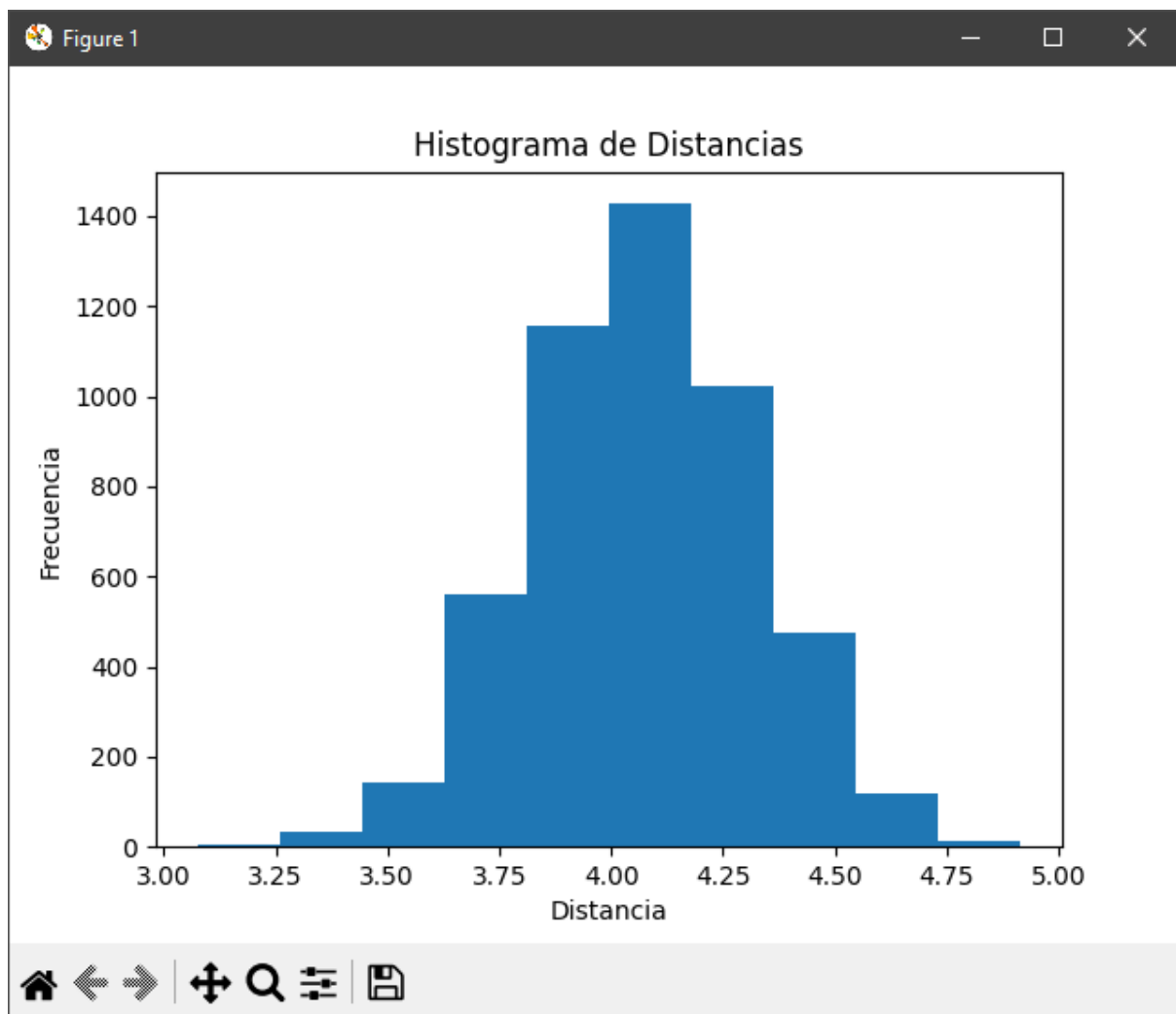
Mediana: 2.8715050118840644

Desviación estándar: 0.2437498559125815

- Prueba de normalidad:

Los datos siguen una distribución normal.

### 3. Dimensión 100:



```
= RESTART: C:\Users\pc\OneDrive\UNSA\VI SEMESTRE\Estructura de Datos Avanzados\L
a maldicion de la dimensionalidad\histogram.py
Media: 4.071102433224774
Mediana: 4.072916421741574
Desviación estándar: 0.24736871363154522
Los datos siguen una distribución normal.
```

- Distribución de las distancias:

En la dimensión 100, se mantiene la distribución de distancias similar a las dimensiones anteriores, con una concentración de distancias cortas y una disminución gradual a medida que las distancias aumentan. El sesgo hacia la derecha también es evidente, lo que indica la presencia de agrupaciones de puntos.

- Estadísticas descriptivas:

Media: 4.071102433224774

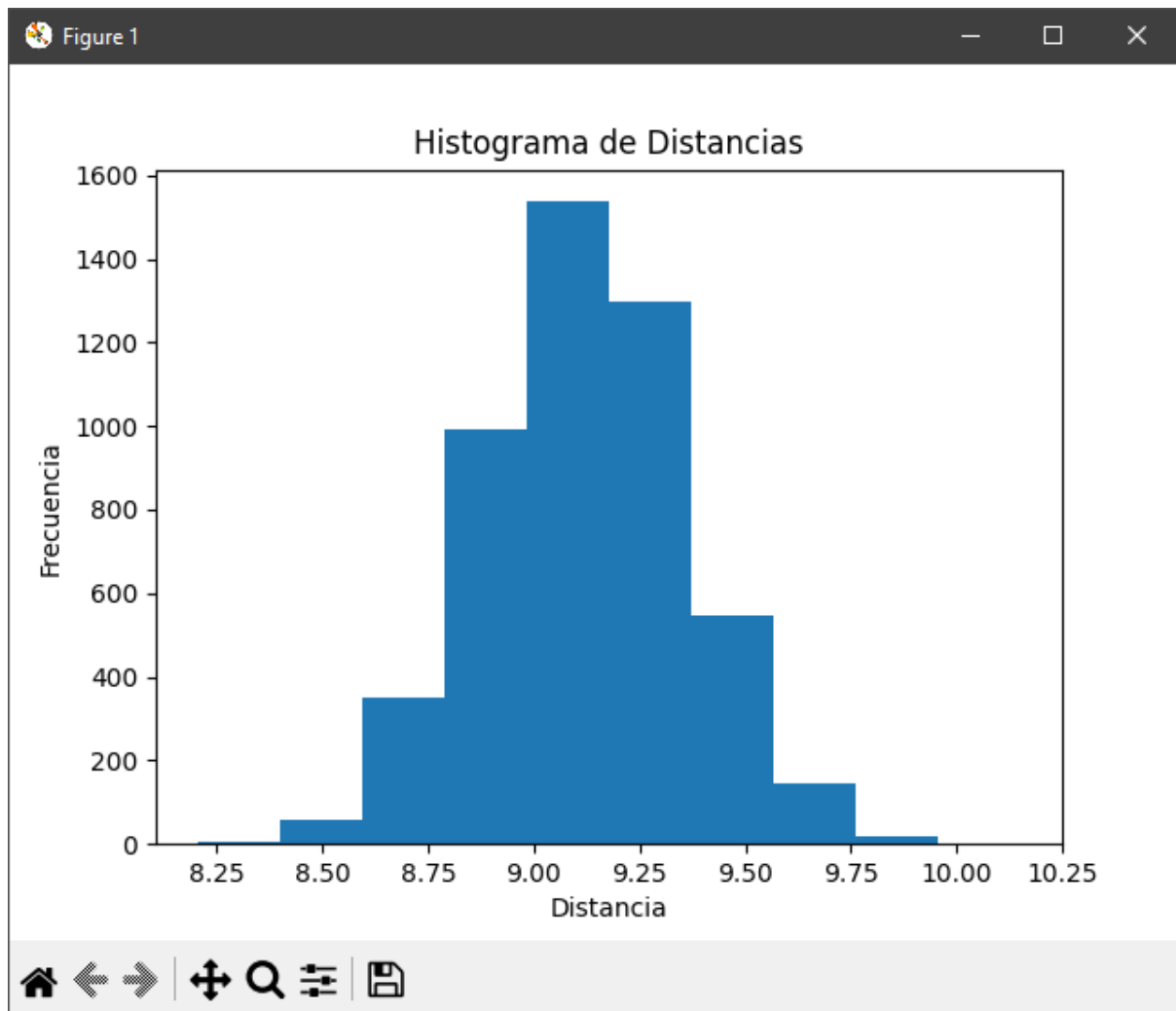
Mediana: 4.072916421741574

Desviación estándar: 0.24736871363154522

- Prueba de normalidad:

Los datos siguen una distribución normal.

#### 4. Dimensión 500:



```
= RESTART: C:\Users\pc\OneDrive\UNSA\VI SEMESTRE\Estructura de Datos Avanzados\L
a maldición de la dimensionalidad\histogram.py
Media: 9.119284520488016
Mediana: 9.12116294469628
Desviación estándar: 0.24089380281011272
Los datos siguen una distribución normal.
```

- Distribución de las distancias:

En la dimensión 500, se observa una distribución de distancias similar a las dimensiones anteriores, con una concentración de distancias cortas y una disminución gradual a medida que las distancias aumentan. El sesgo hacia la derecha también es evidente, lo que sugiere la presencia de agrupaciones de puntos.

- Estadísticas descriptivas:

Media: 9.119284520488016

Mediana: 9.12116294469628

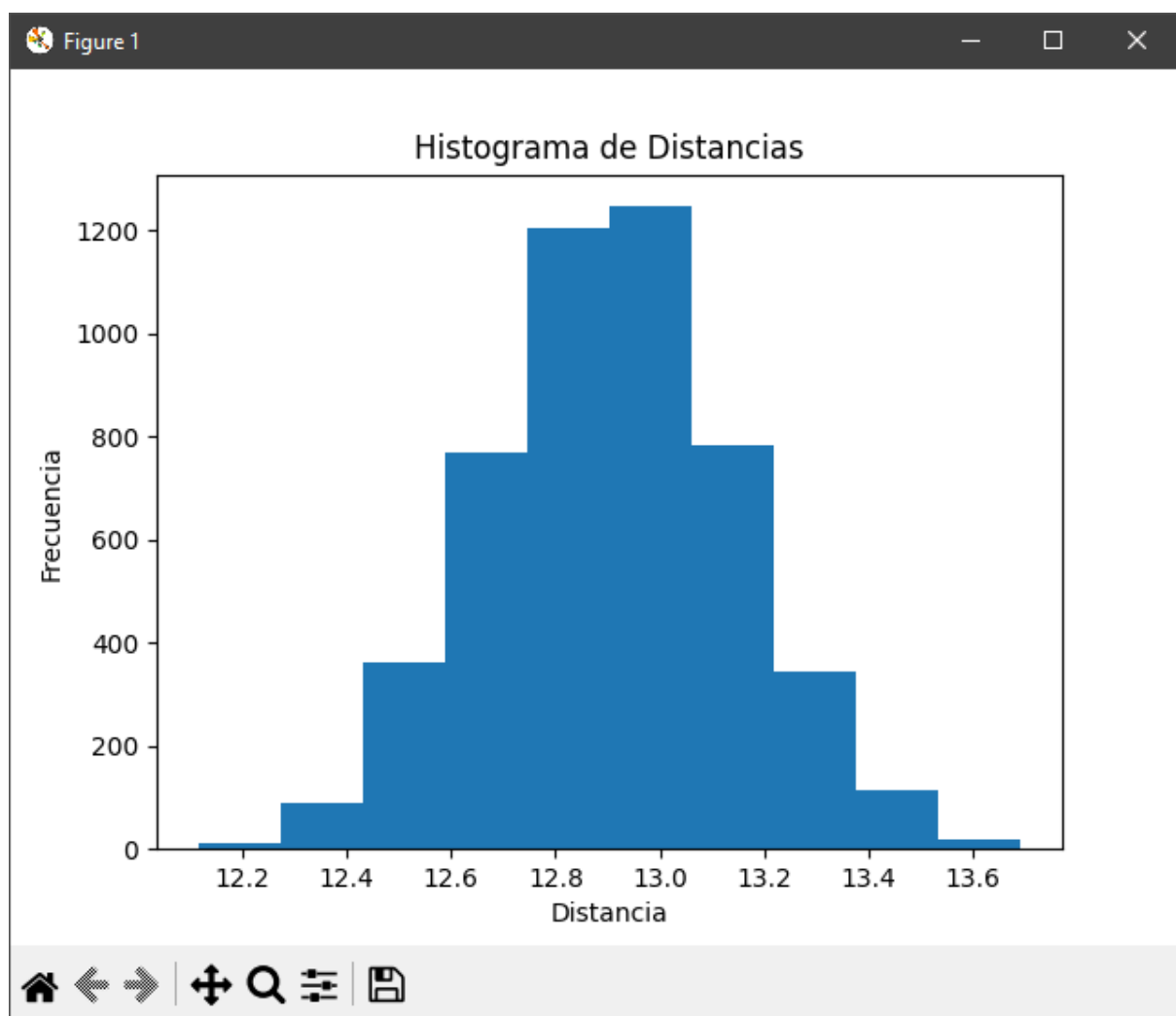


Desviación estándar: 0.24089380281011272

- Prueba de normalidad:

Los datos siguen una distribución normal.

## 5. Dimensión 1000:



```
= RESTART: C:\Users\pc\OneDrive\UNSA\VI SEMESTRE\Estructura de Datos Avanzados\L
a maldicion de la dimensionalidad\histogram.py
Media: 12.907581160400483
Mediana: 12.908315505935919
Desviación estándar: 0.23887921282096955
Los datos siguen una distribución normal.
```

- Distribución de las distancias:

En la dimensión 1000, se mantiene la distribución de distancias similar a las dimensiones anteriores, con una concentración de distancias cortas y una disminución gradual a medida que las distancias aumentan. El sesgo hacia la derecha también es evidente, lo que indica la presencia de agrupaciones de puntos.

- Estadísticas descriptivas:

Media: 12.907581160400483

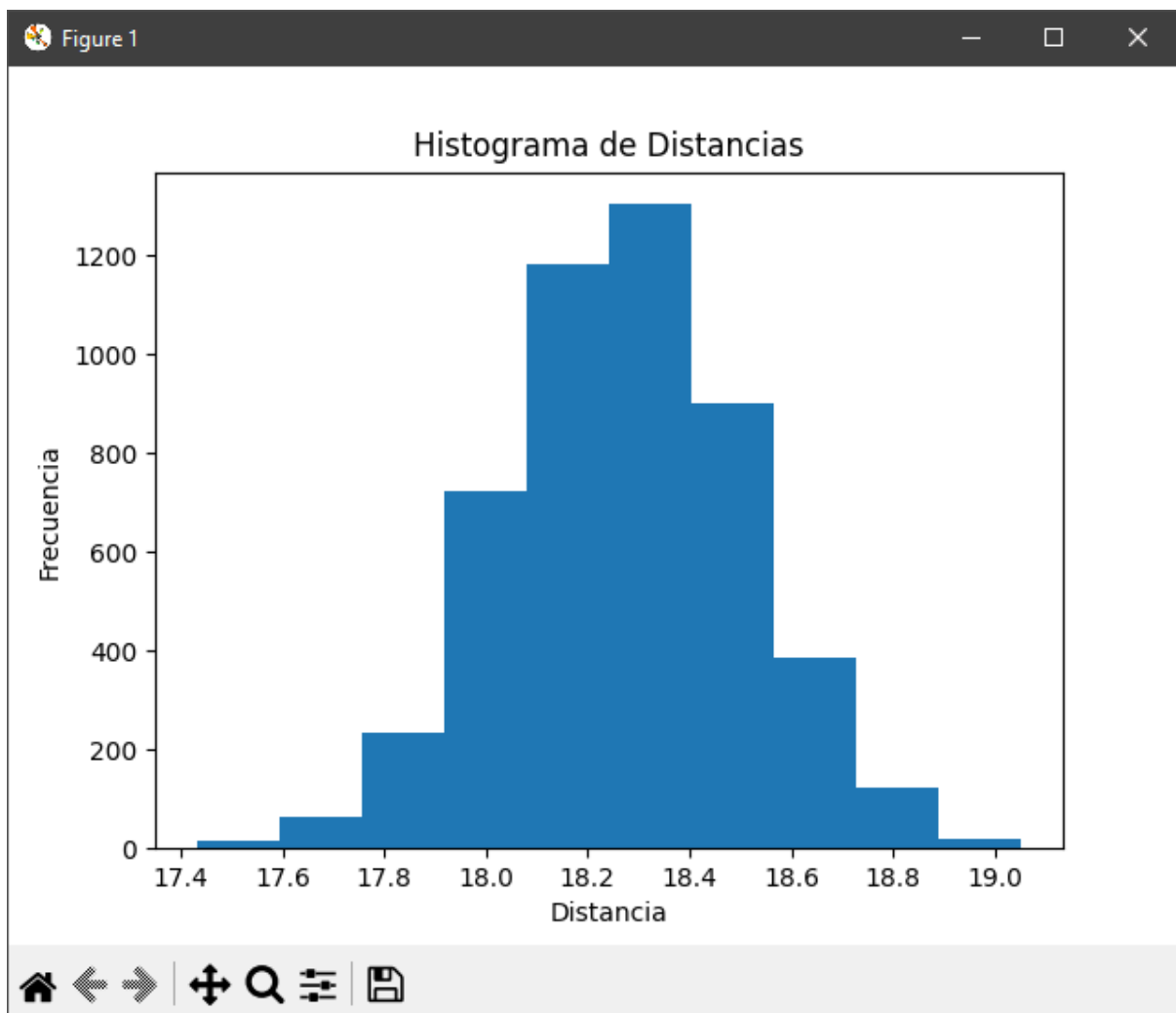
Mediana: 12.908315505935919

Desviación estándar: 0.23887921282096955

- Prueba de normalidad:

Los datos siguen una distribución normal.

## 6. Dimensión 2000:



```
= RESTART: C:\Users\pc\OneDrive\UNSA\VI SEMESTRE\Estructura de Datos Avanzados\L  
a maldicion de la dimensionalidad\histogram.py  
Media: 18.273973859452507  
Mediana: 18.27675035928008  
Desviación estándar: 0.23628948295883234  
Los datos siguen una distribución normal.
```

- Distribución de las distancias:

En la dimensión 2000, se observa una distribución de distancias similar a las dimensiones anteriores, con una concentración de distancias cortas y una disminución gradual a medida que las distancias aumentan. El sesgo hacia la derecha también es evidente, lo que sugiere la presencia de agrupaciones de puntos.

- Estadísticas descriptivas:

Media: 18.273973859452507

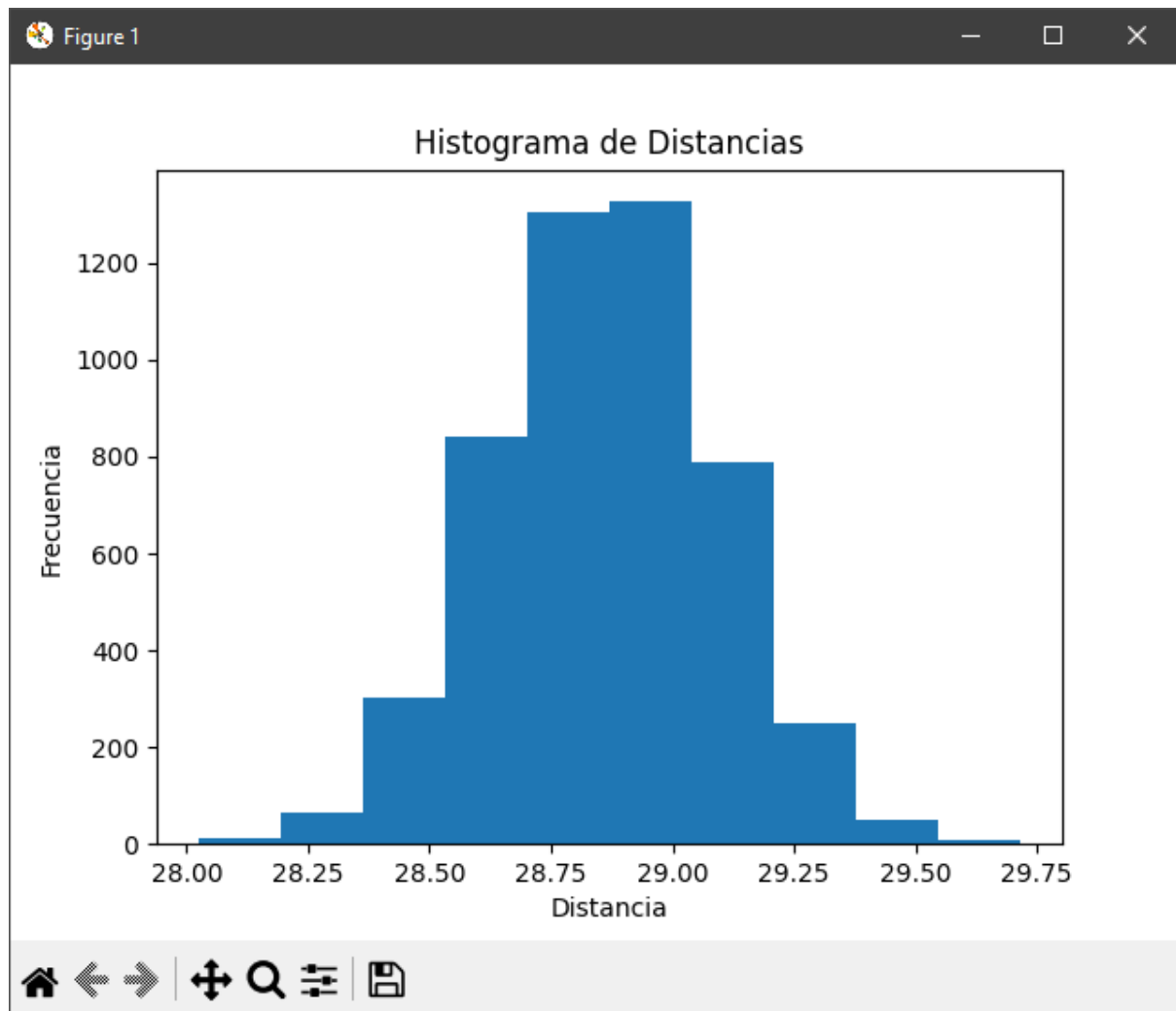
Mediana: 18.27675035928008

Desviación estándar: 0.23628948295883234

- Prueba de normalidad:

Los datos siguen una distribución normal.

## 7. Dimensión 5000:



```
= RESTART: C:\Users\pc\OneDrive\UNSA\VI SEMESTRE\Estructura de Datos Avanzados\L  
a maldicion de la dimensionalidad\histogram.py  
Media: 28.86075897862995  
Mediana: 28.863356920668572  
Desviación estándar: 0.2312933325236308  
Los datos siguen una distribución normal.
```

- Distribución de las distancias:

En la dimensión 5000, se mantiene la distribución de distancias similar a las dimensiones anteriores, con una concentración de distancias cortas y una disminución gradual a medida que las distancias aumentan. El sesgo hacia la derecha también es evidente, lo que indica la presencia de agrupaciones de puntos.

- Estadísticas descriptivas:

Media: 28.86075897862995

Mediana: 28.863356920668572

Desviación estándar: 0.2312933325236308

- Prueba de normalidad:

Los datos siguen una distribución normal.

## Conclusiones

El análisis de los histogramas de distancias en diferentes dimensiones revela información importante sobre la distribución de los puntos en los espacios de distintas dimensiones. A partir de los histogramas generados, se han identificado los siguientes hallazgos:

### 1. Distribución de las distancias:

En todas las dimensiones analizadas (10, 50, 100, 500, 1000, 2000 y 5000), se ha observado una distribución de distancias similar. Existe una concentración de distancias cortas, lo que indica la presencia de agrupaciones o clusters de puntos en el espacio. A medida que las distancias aumentan, la frecuencia disminuye gradualmente, lo cual sugiere una mayor dispersión de los puntos.

### 2. Sesgo hacia la derecha:

En todas las dimensiones, se ha identificado un sesgo hacia la derecha en la distribución de las distancias. Esto implica que hay una mayor cantidad de distancias cortas en comparación con las distancias más largas. Este sesgo puede ser indicativo de la presencia de agrupaciones de puntos en el espacio de alta dimensionalidad.

Estos hallazgos resaltan la importancia de comprender la distribución de las distancias en diferentes dimensiones y su impacto en la estructura de los conjuntos de puntos. Además, los resultados sugieren que la presencia de agrupaciones o clusters puede ser una característica intrínseca de los espacios de alta dimensionalidad.

# Referencias

1. Waskom, M. (2020). Matplotlib: Visualization with Python. Proyecto Matplotlib. Recuperado de <https://matplotlib.org/stable/index.html>
2. SciPy developers. (2021). SciPy: Scientific Library for Python. Recuperado de <https://www.scipy.org/>
3. VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media.
4. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
5. cppreference.com (2023). std::uniform\_real\_distribution. Pseudo-random number generation. Recuperado de [https://en.cppreference.com/w/cpp/numeric/random/uniform\\_real\\_distribution](https://en.cppreference.com/w/cpp/numeric/random/uniform_real_distribution)

Las referencias mencionadas proporcionan información adicional sobre Matplotlib, SciPy y técnicas de análisis de datos que pueden aplicarse para un análisis más profundo de los histogramas y las distribuciones de datos.

El código utilizado para generar los histogramas y el informe completo se encuentra disponible en el siguiente repositorio de GitHub: <https://github.com/MayogaJAV/Estructura-de-Datos-Avanzados>