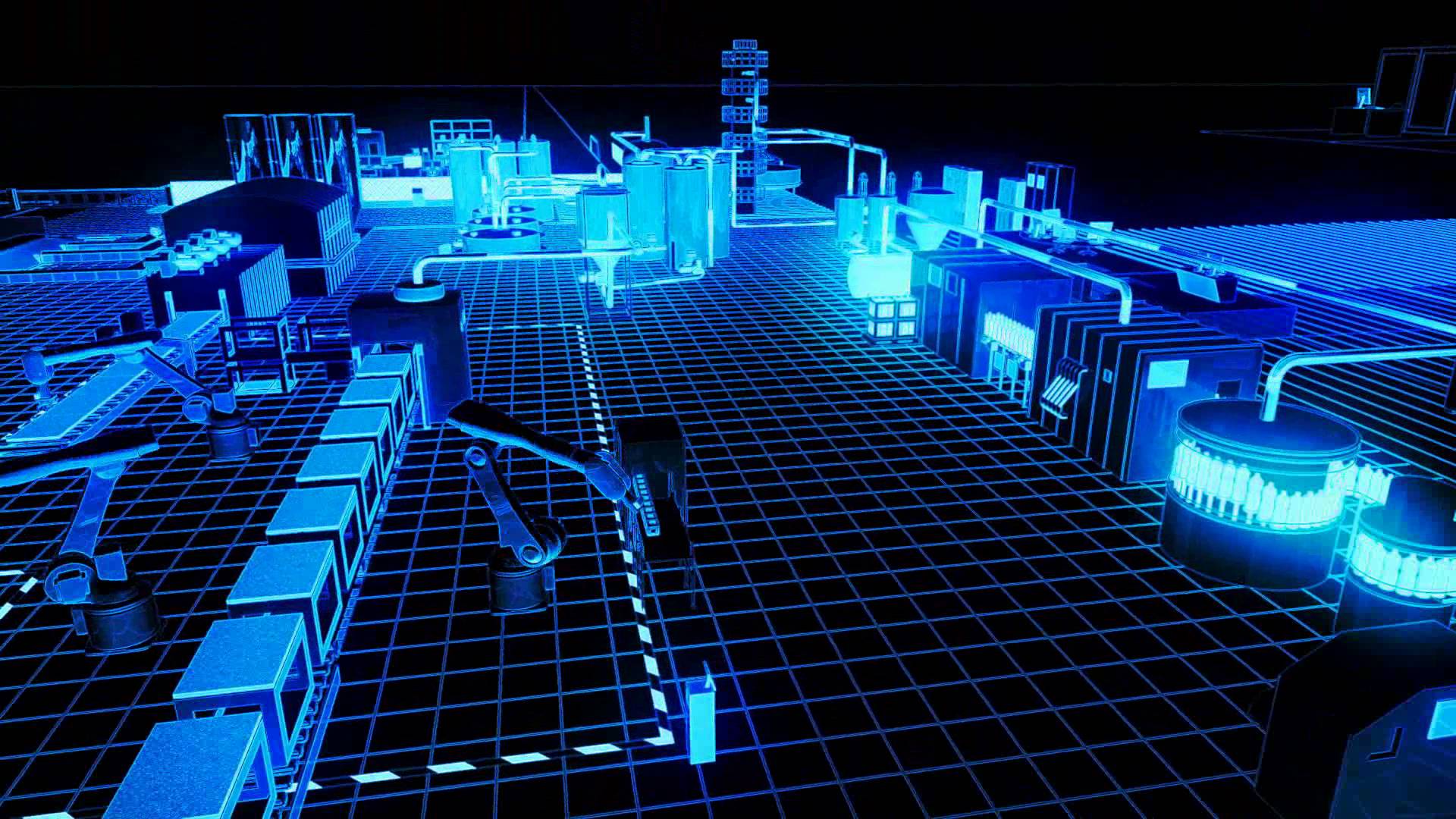
**UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS**

**ESCUELA PROFECIONAL DE INGENIERIA DE SISTEMAS**

**TOPICOS**

**DOCUMENCIÓN**

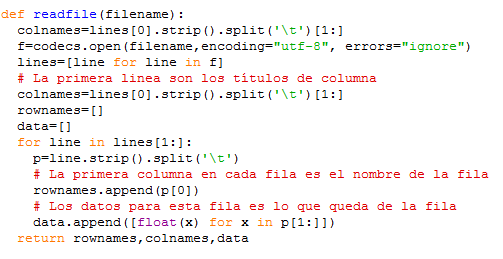
INTEGRANTE:

* **JOEL RUIZ TAIPE**

**Andahuaylas**

1. **Modifique el código de parsear blogs para agrupar entradas individuales en lugar de blogs enteros. ¿Las entradas del mismo blog se agrupan unas con otras?¿Que hay de las entradas de la misma fecha?**

El código para modificar

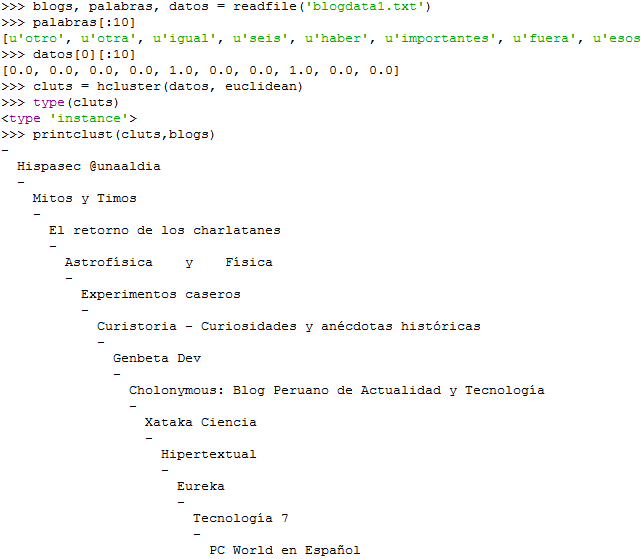


1. **Pruebe usando la distancia euclidiana para clustering de los blogs. ¿Cómo cambia esto los resultados?**

Esto es la distancia euclidiana que contiene la variable1 y la variable2 para que saque la dicha distancia

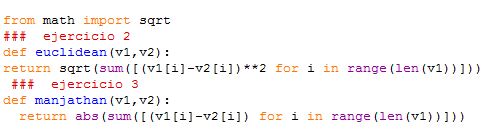


Aquí la distancia euclidiana solo te muestra las distancias de cada blog. En esta caso la distancias de sus títulos.

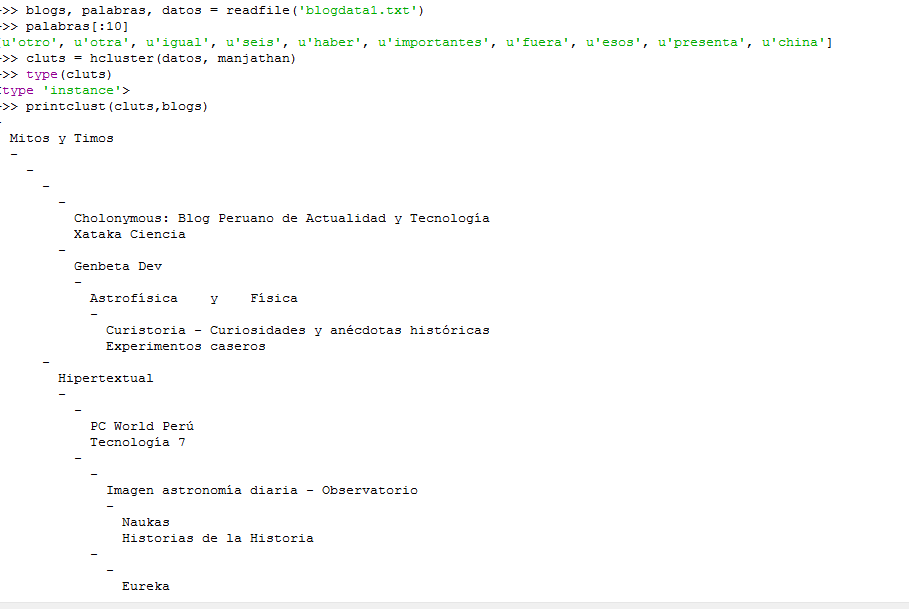


1. **Investigue acerca de la distancia de Manhattan. Cree una función para esta y vea cómo cambia los resultados.**

La fórmula de manjathan es el valor absoluto de 2 variables o de dos puntos

****

El resultado del algoritmo de manjathan te devuelve los títulos de cada blog y su grupo que pertenece a ese título de cada uno por separado



1. **Modifique la función del clustering K-means para que devuelva la distancia total ente todos los ítems de sus respectivos centroides.**
2. **Cree una función que corra el clustering K-means para diferentes valores de k. ¿Cómo cambia la distancia cuando el número de clusters se incrementa?¿Hasta que punto mejora tener más clusters?**