



Privacidad en aplicaciones móviles



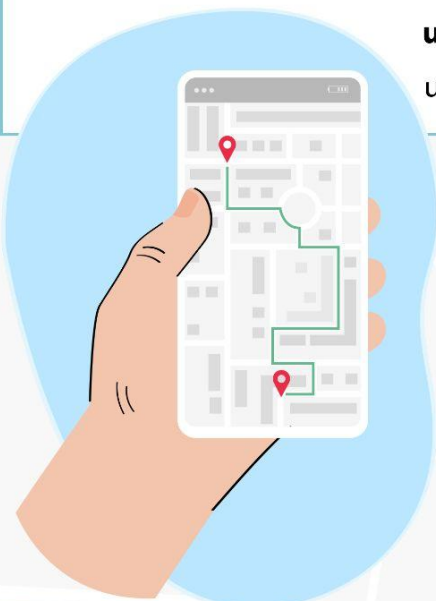
Reflexionemos

1. ¿Qué aplicaciones de celulares que usan tu ubicación conoces?
2. ¿Cuán seguido utilizan este tipo de aplicaciones?
3. ¿Creen que existen algunos riesgos al utilizar algunas de estas?

Infografía “Aplicaciones móviles que requieren tu ubicación”

APLICACIONES MÓVILES QUE REQUIEREN TU UBICACIÓN

Existe muchas aplicaciones de dispositivos móviles que entregan un **servicio basado en la ubicación** que un usuario transmite.



SERVICIOS QUE UTILIZAN LA POSICIÓN:



Previsión
meteorológica



Sistema de
recomendaciones



Aplicaciones
de navegación



Seguimiento
deportivo

A partir del **historial de ubicaciones** compartidas a las aplicaciones, se pueden **identificar aquellos lugares** en los que pasas más tiempo durante el día.



Infografía “Aplicaciones móviles que requieren tu ubicación”

A partir de la infografía respondamos:

1. ¿Qué es un punto de interés de una persona y cómo las aplicaciones pueden identificarlo?
2. Digan ejemplos de sus puntos de interés del día.
3. ¿De qué manera amenaza tu privacidad si terceras personas reconocen tus puntos de interés?

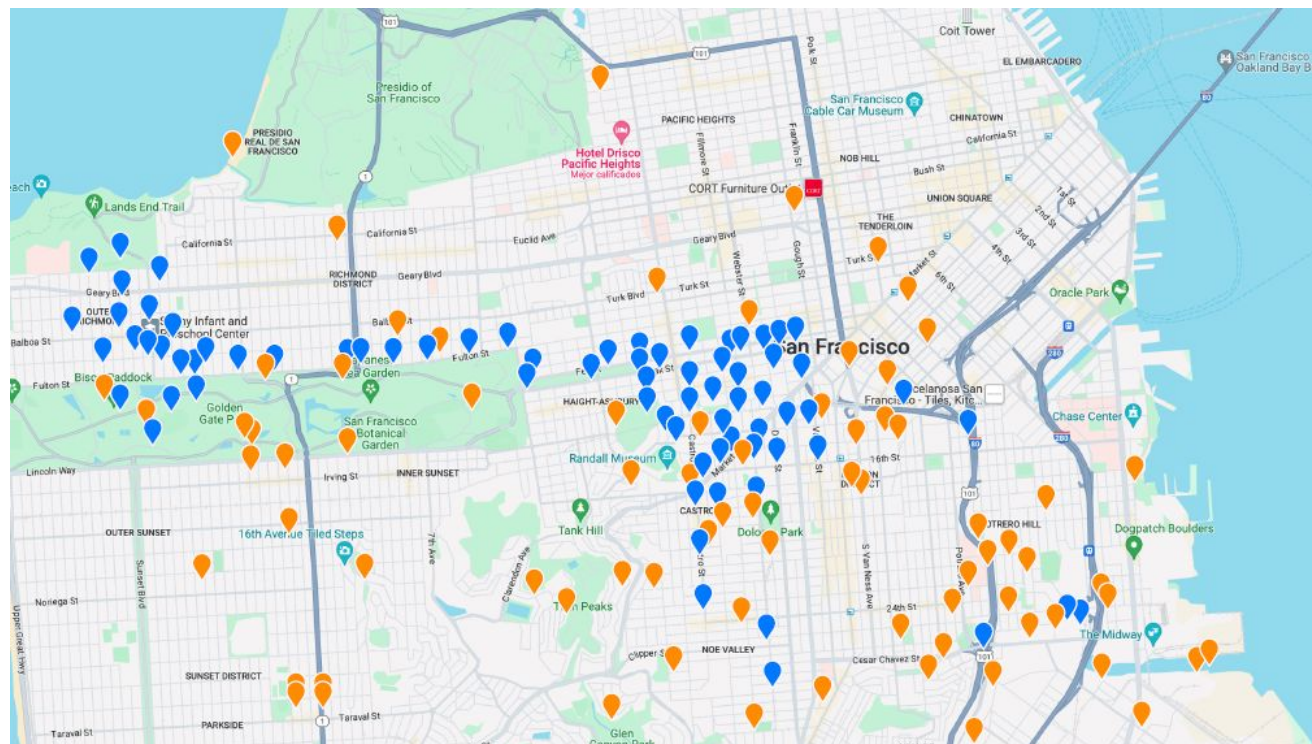
Infografía “Aplicaciones móviles que requieren tu ubicación”

Entonces:

1. Un punto de interés es un lugar en que una persona pasa un tiempo considerable del día. Las aplicaciones los pueden identificar accediendo al historial de ubicaciones.
2. Puntos de interés de sus estudiantes pueden ser el hogar, el colegio, una plaza en la que se encuentran con sus amigos-as, etc.
3. Amenaza la privacidad en el sentido que terceras personas no solo saben las ubicaciones, sino que también pueden inferir las preferencias “sensibles” como religiosas, políticas, sexuales, etc.

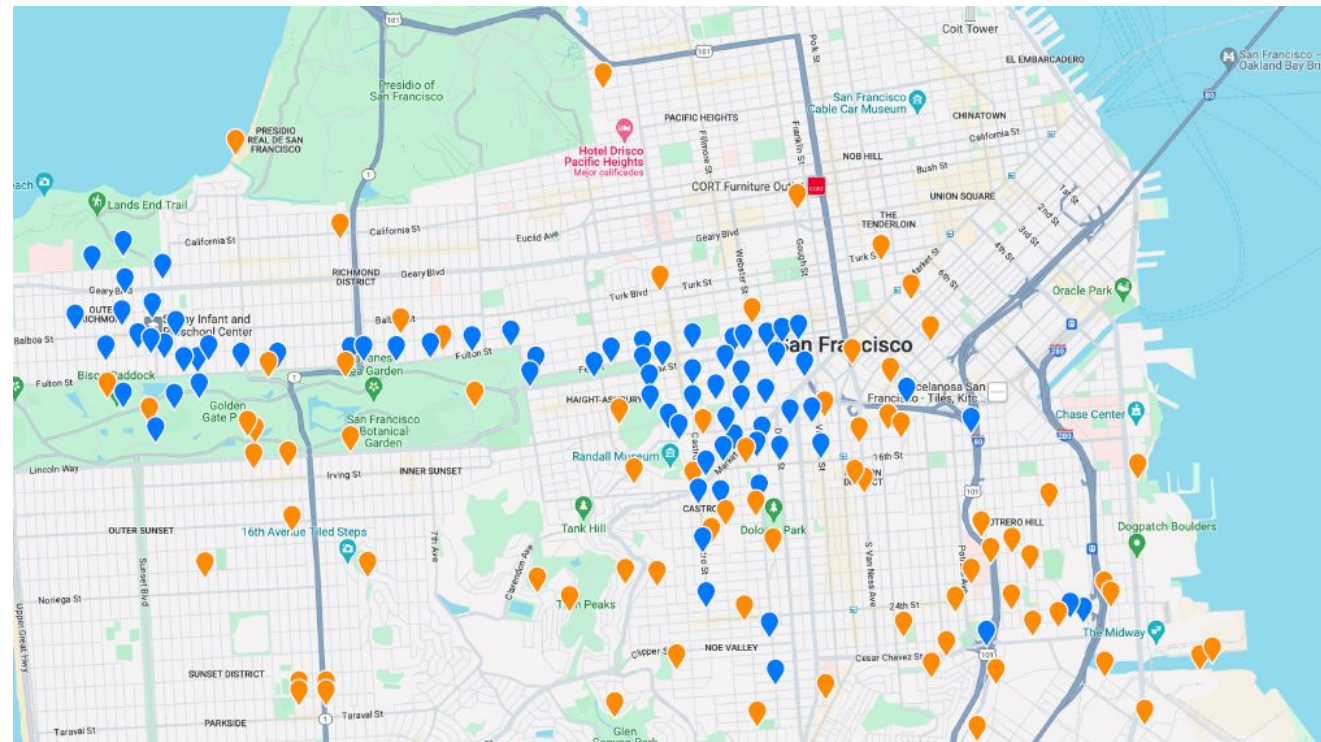
Problema

En la situación vista en la infografía contamos con una muestra de las posiciones de dos usuarios: Usuario Azul y Usuario Naranja. Esta muestra recopiló sus posiciones cada 30 segundos, durante 15 minutos.



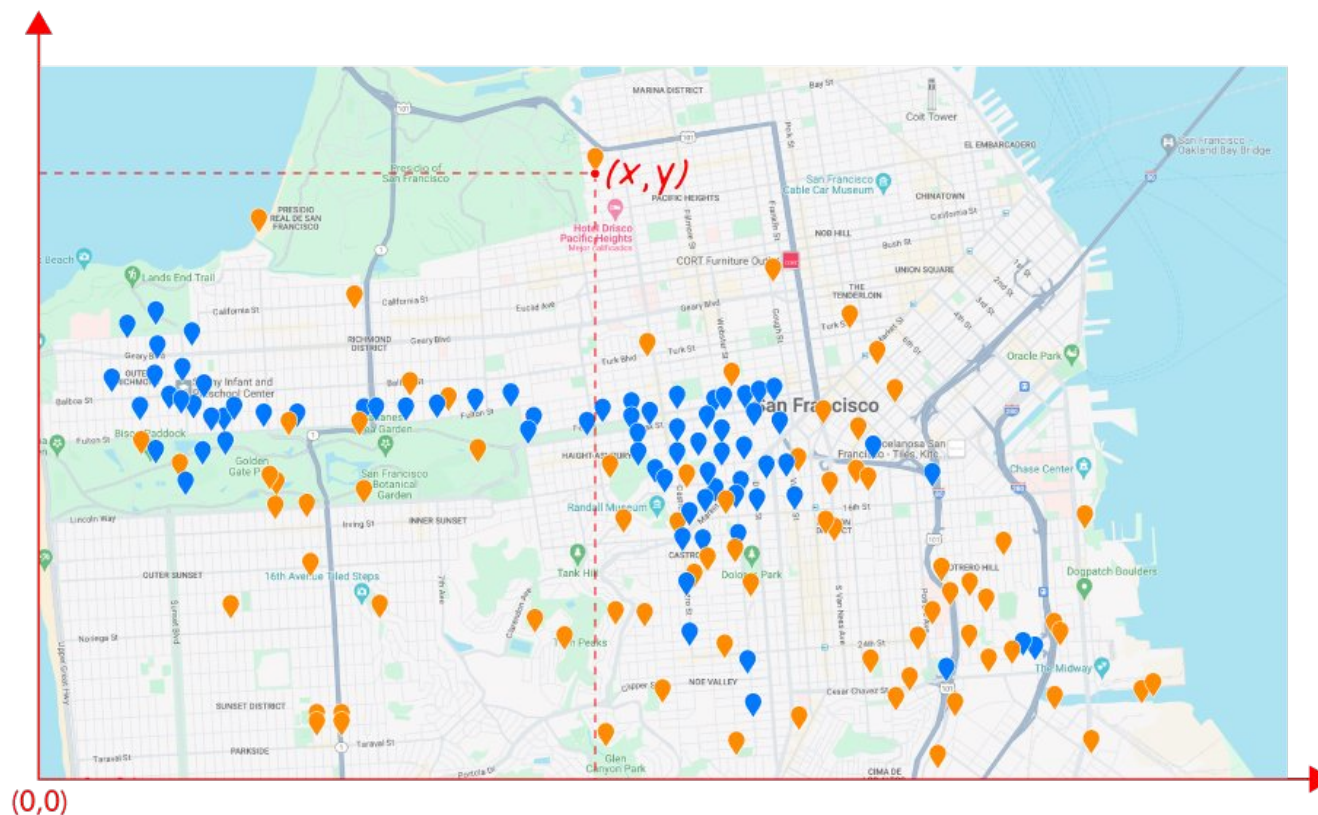
Problema

¿Qué usuario tiene menor nivel de privacidad respecto a sus puntos de interés, el azul o el naranja?, ¿Por qué?



Problema

Podemos visualizar las ubicaciones anteriores con el siguiente sistema de coordenadas con origen en (0, 0).

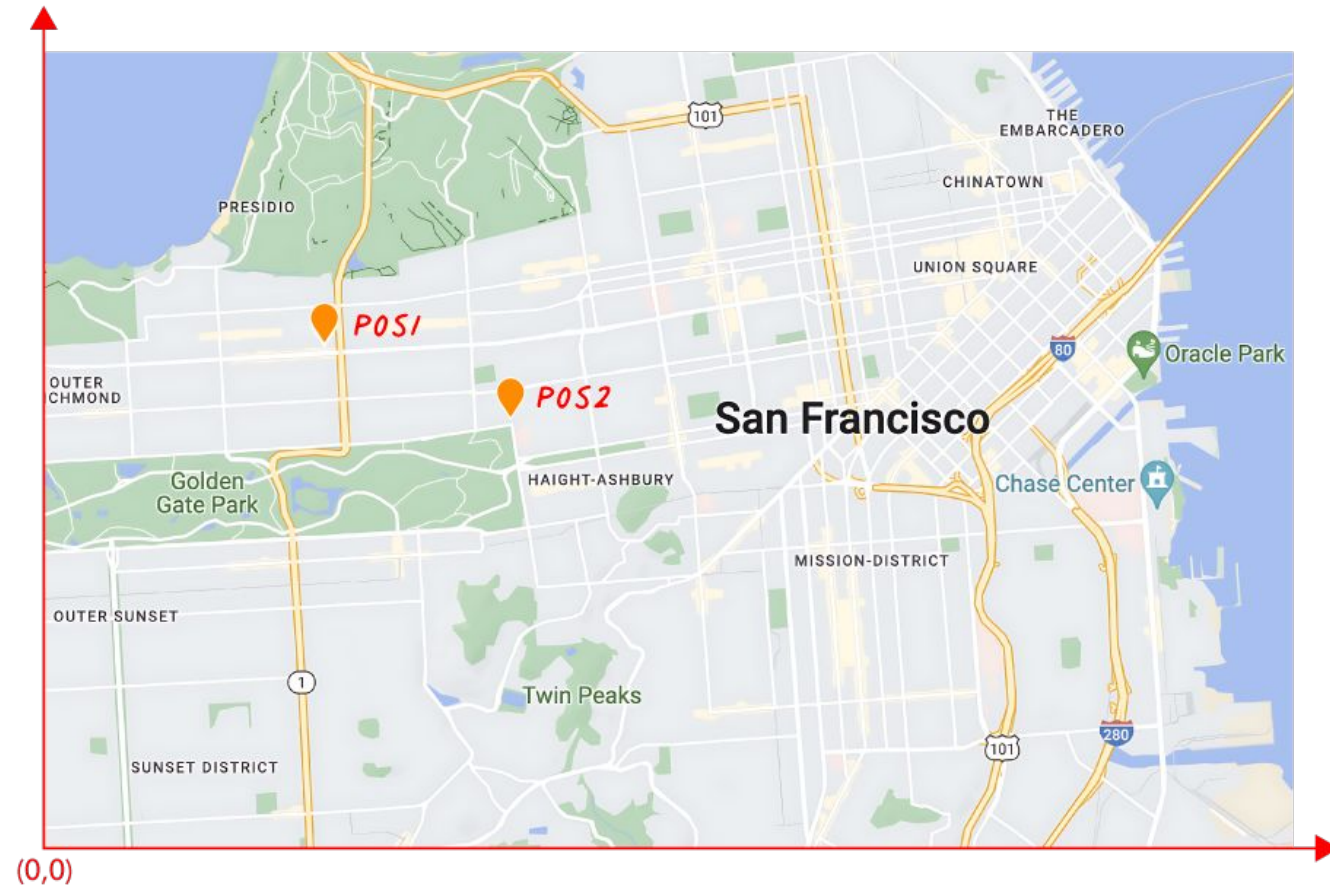


Problema

¿ Cómo podemos comparar matemáticamente los niveles de privacidad entre dos usuarios, utilizando sus posiciones en los últimos 15 minutos?

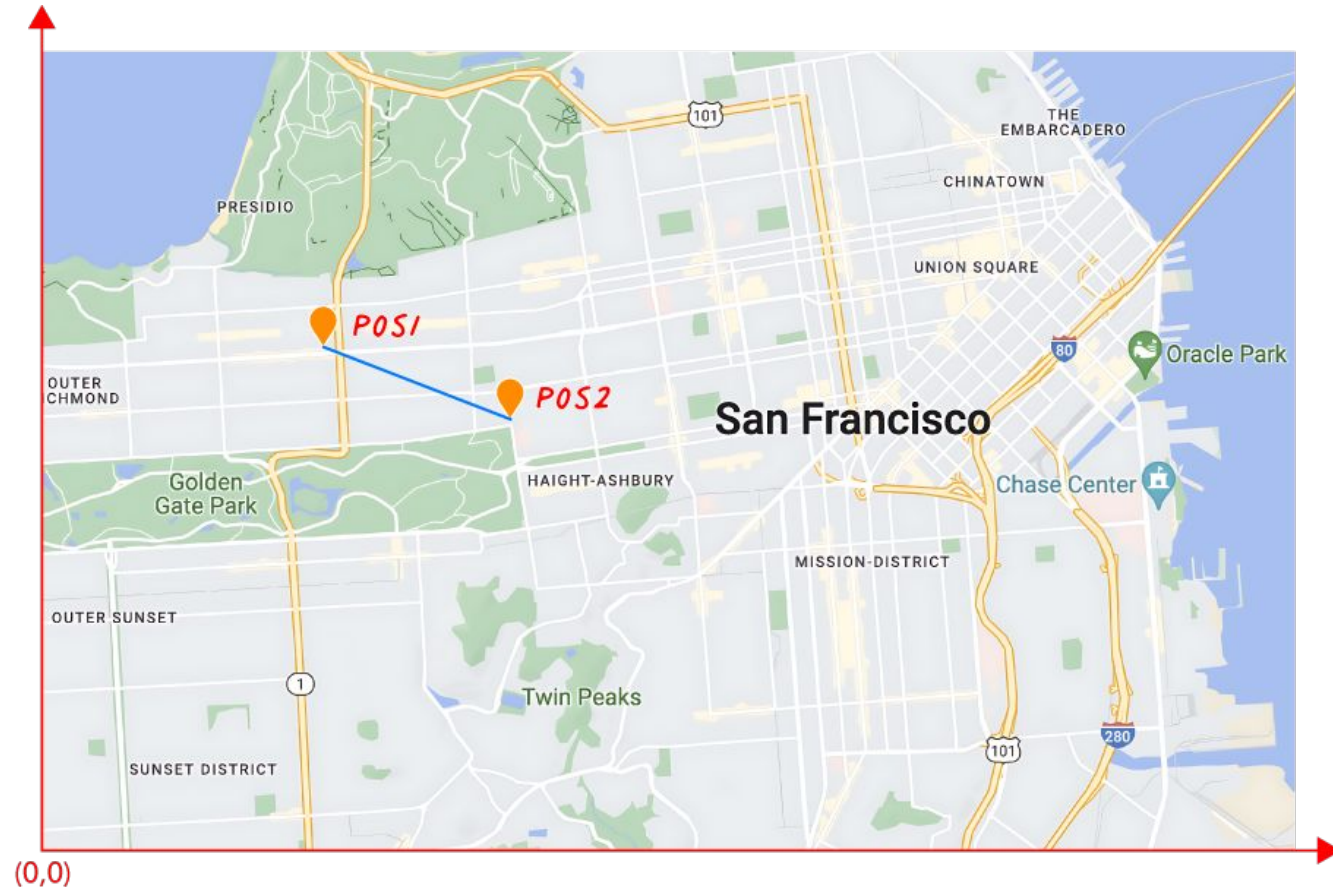
Medida de la cercanía del conjunto de posiciones de un usuario.

¿Cómo medirías la cercanía entre ambas posiciones?



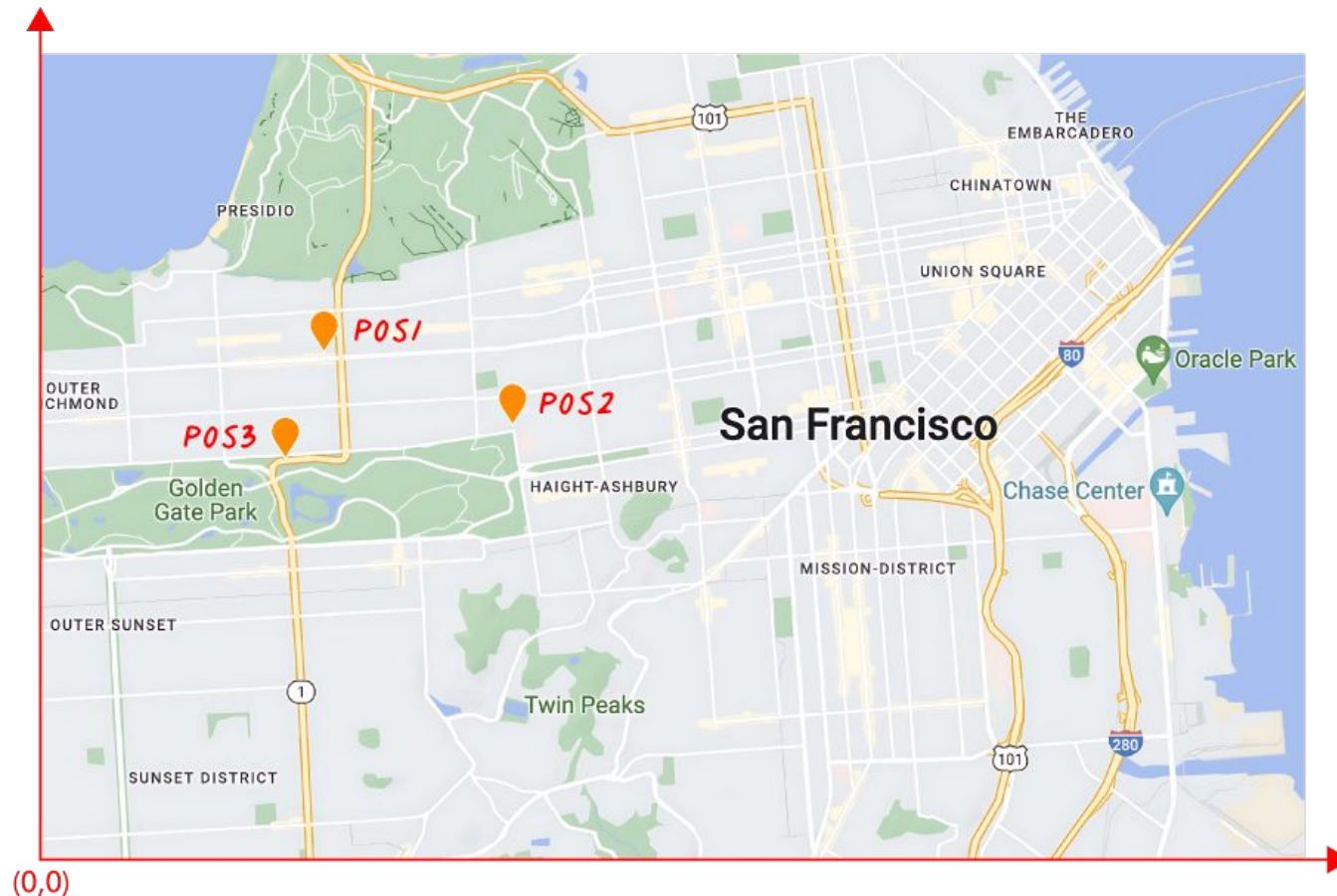
Medida de la cercanía del conjunto de posiciones de un usuario.

Cercanía entre ambas posiciones, como la medida del segmento azul



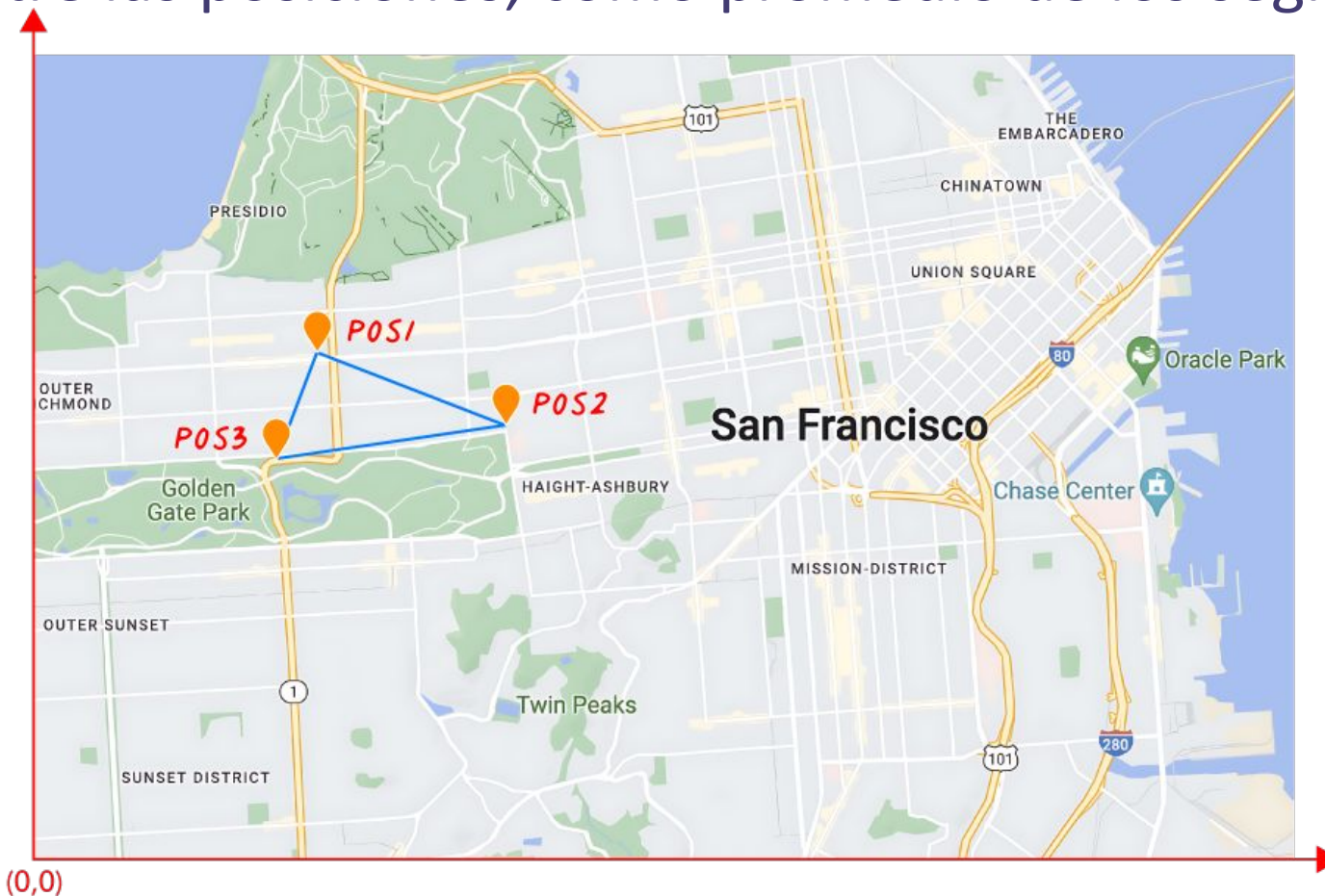
Medida de la cercanía del conjunto de posiciones de un usuario.

¿Cómo medirías ahora la cercanía entre las 3 posiciones?



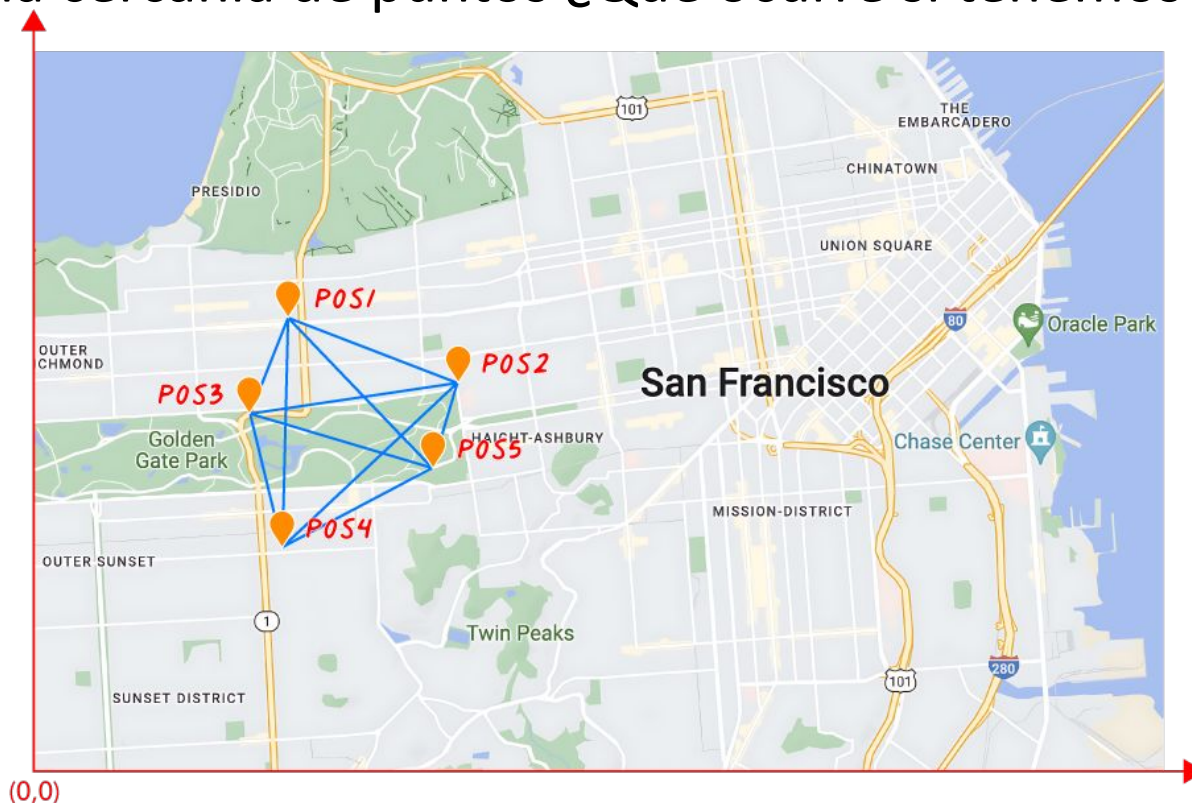
Medida de la cercanía del conjunto de posiciones de un usuario.

Cercanía entre las posiciones, como promedio de los segmento azules



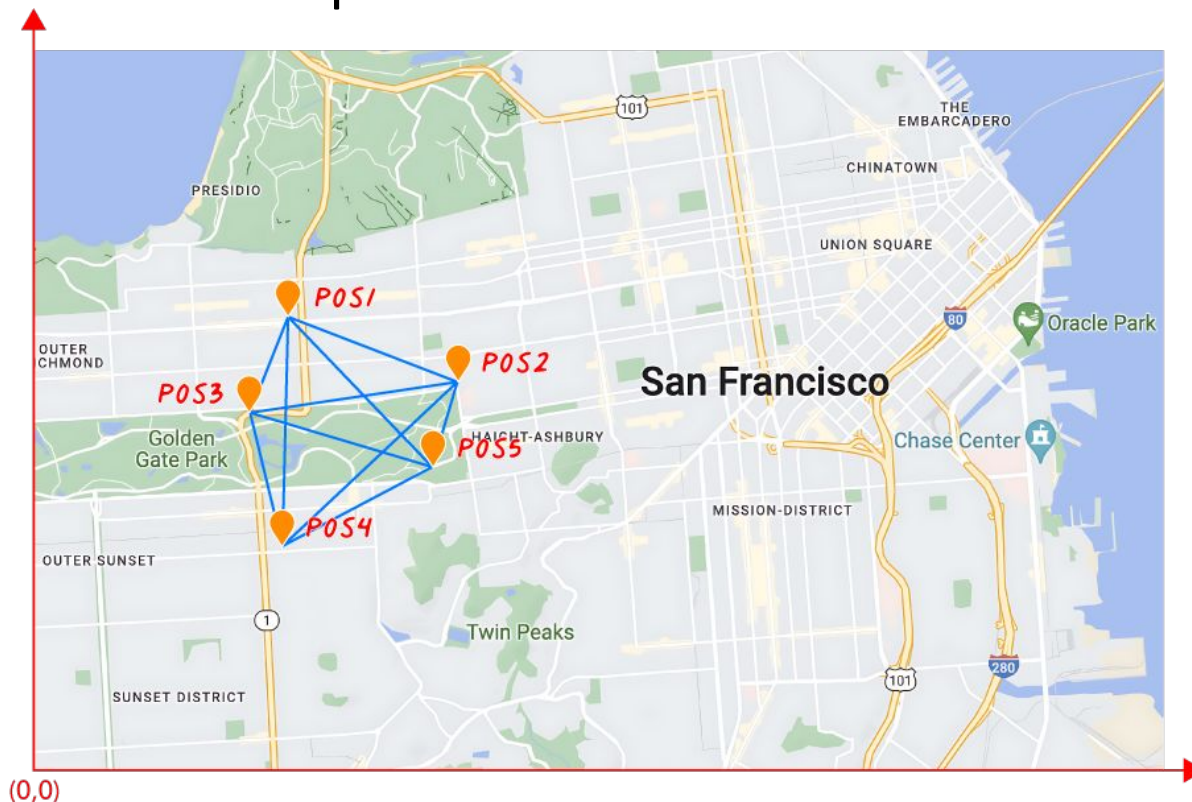
Medida de la cercanía del conjunto de posiciones de un usuario.

Si consideramos ahora 5 ubicaciones, tenemos que calcular las distancias de cada segmento azul y luego calcular los promedios de estas distancias, y así obtenemos una medida de la cercanía de puntos ¿Que ocurre si tenemos más puntos?



Centroide de un conjunto de puntos.

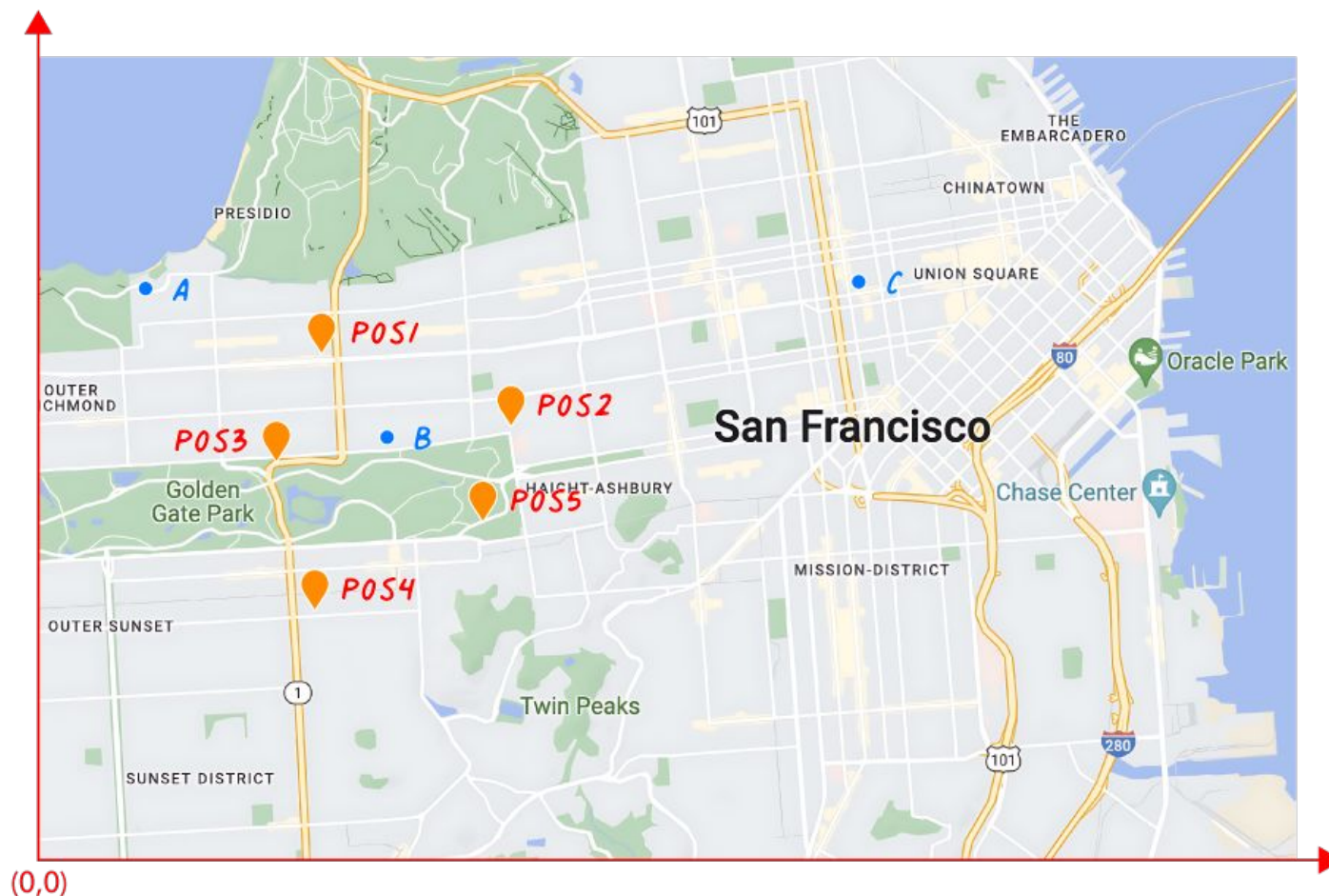
Con 30 ubicaciones, tendríamos $\binom{30}{2} = 435$ combinaciones de segmentos azules, un cálculo engorroso para medir la cercanía. Para esto es mejor que consideremos un punto de referencia a todos puntos.



Actividad 1

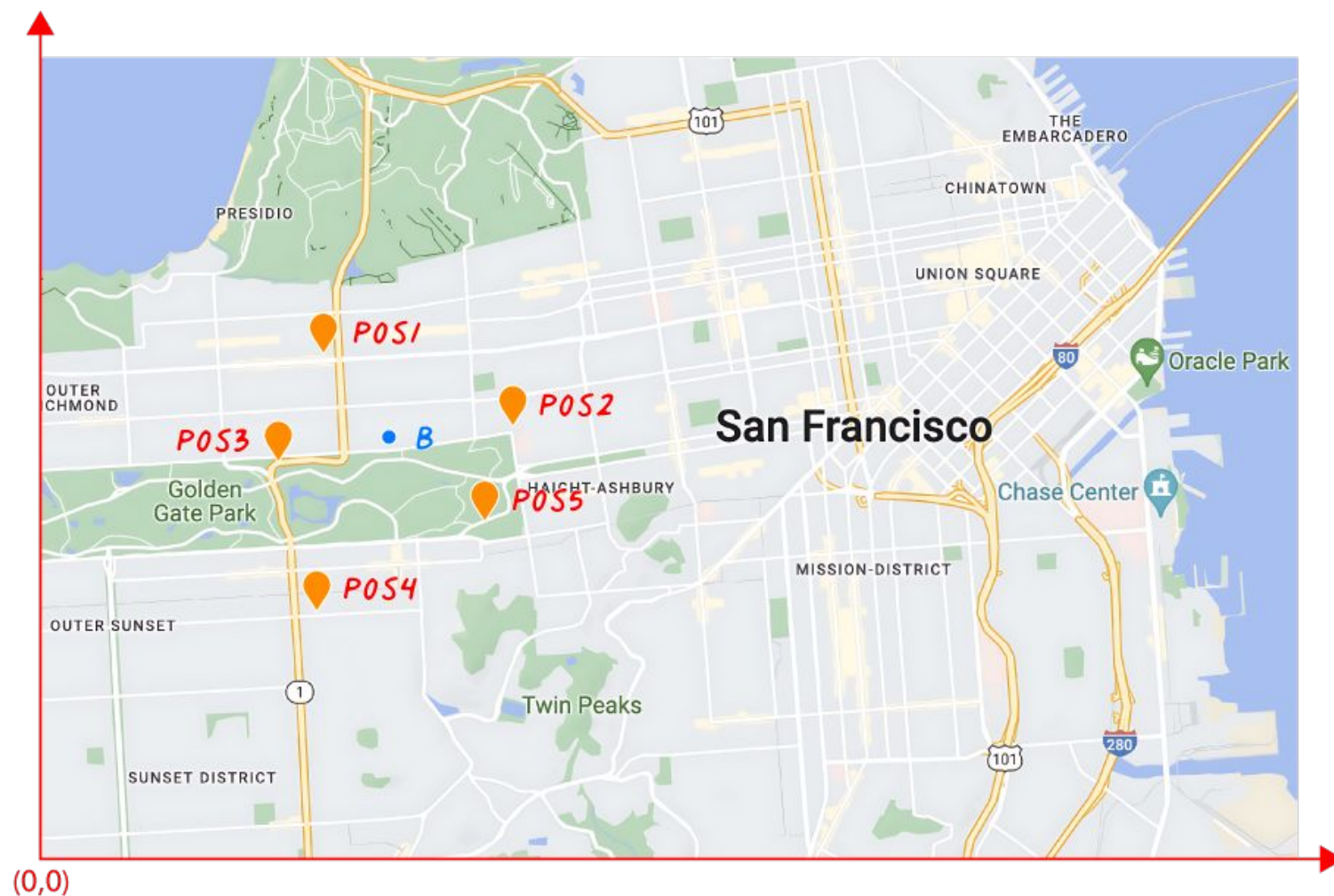
Para comprender el concepto de centroide respondamos la siguiente pregunta

1. Si tuvieses que elegir un punto aproximado que represente un lugar “promedio” dónde se encontraba el usuario entre estas 5 posiciones, ¿cuál elegirías: A, B o C? Justifica.



Actividad 1

1. Tiene más sentido seleccionar el punto B, pues se encuentra “en medio” de las posiciones, mientras que los puntos A y C pueden descartarse dada la lejanía al conjunto de puntos.



Centroide de un conjunto de puntos.

El punto anterior se puede determinar exactamente como el promedio de las coordenadas de las posiciones, como se indica a continuación con 30 puntos:

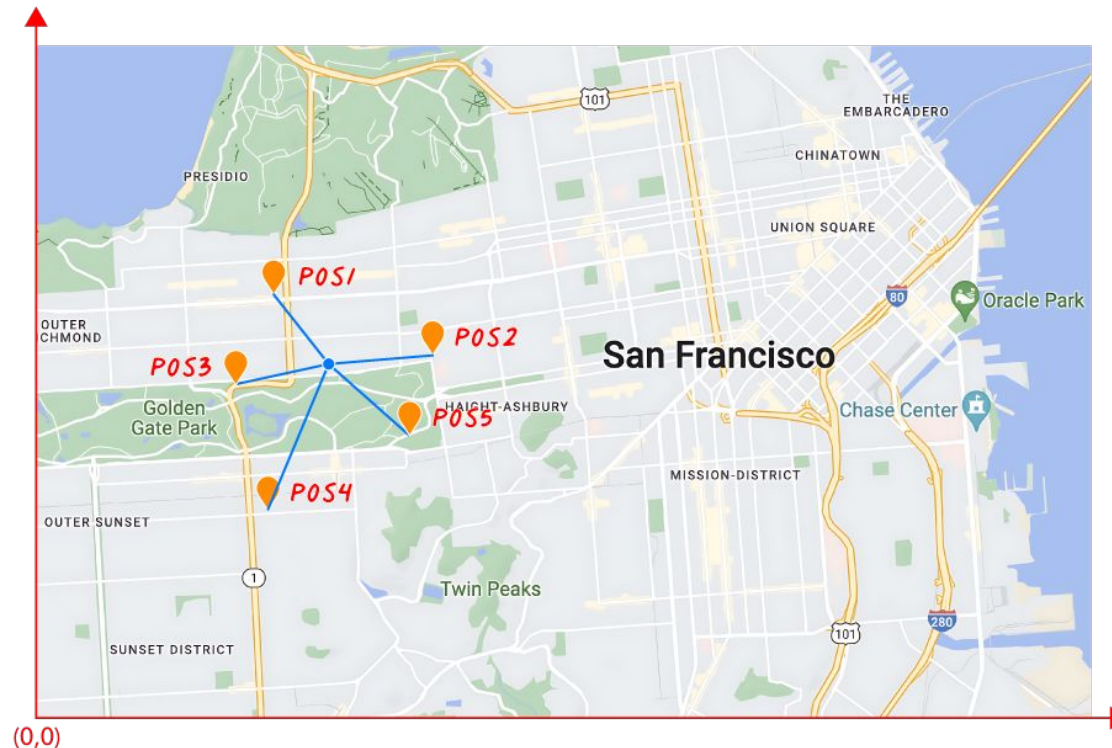
$$(x_c, y_c) = \left(\frac{x_1 + \cdots + x_{30}}{30}, \frac{y_1 + \cdots + y_{30}}{30} \right)$$

Este punto se llama “**baricentro**” o “**centroide**” o “**centro de gravedad**” de un conjunto de puntos.

Esta definición coincide con la del centro de gravedad de un triángulo, punto donde se intersectan las transversales de gravedad.

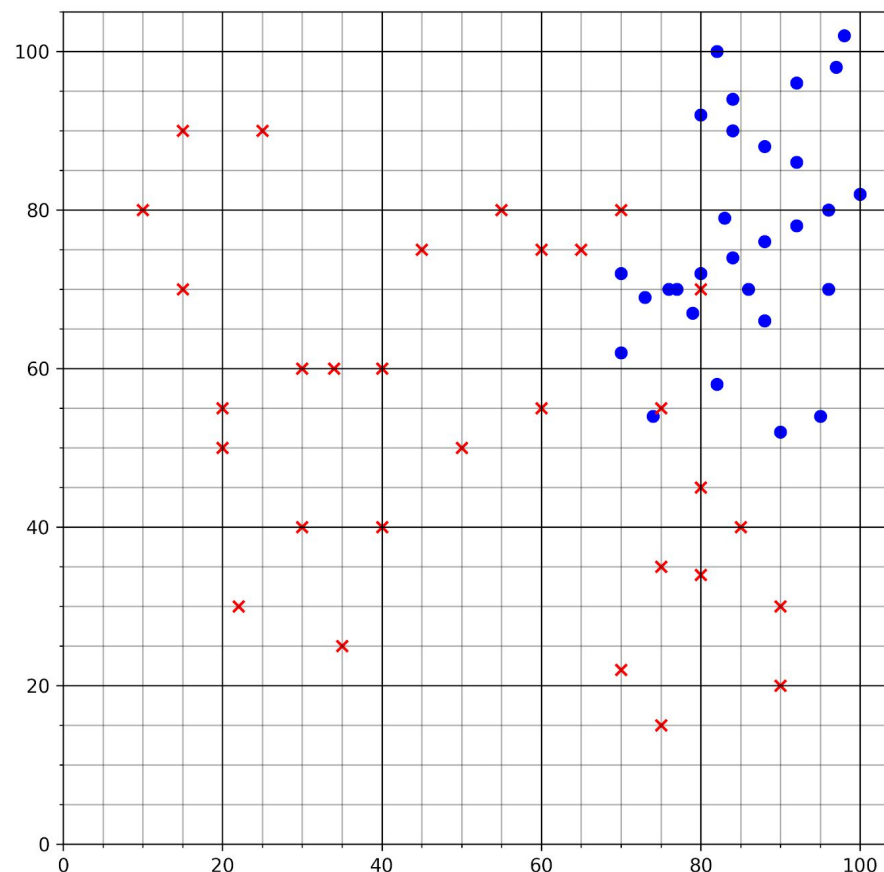
Centroide de un conjunto de puntos.

Ahora que sabemos cómo calcular el centroide de un conjunto de puntos, la medida de la cercanía de puntos se simplifica, pues dado un conjunto de 5 puntos solo resta calcular sus distancias al centroide y promediarlas, por ejemplo para el caso de 30 ubicaciones, no calculamos 435 distancias, sino solo 30, y las promediamos.



Cálculo del centroide del Usuaría Azul y Rojo.

A continuación tenemos las ubicaciones de los usuarios azul y rojo. Podemos notar que el cálculo del centroide de cada usuario se puede calcular usando herramientas de excel.



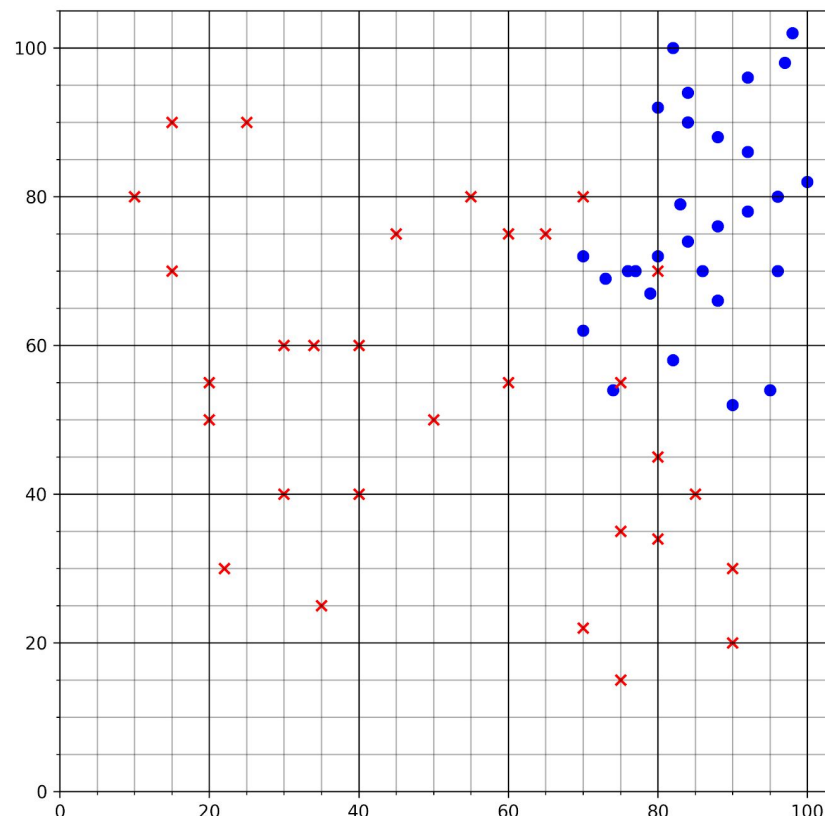
Cálculo del centroide del Usuario Azul y Rojo.

Por ejemplo si quisiéramos calcular el centroide de 3 ubicaciones, podemos usar la función “PROMEDIO” o “AVERAGE” aplicada a las coordenadas de los ejes de las abscisa y ordenada, como se muestra a continuación.

	abscisa	ordenada
	75,4	10,7
	83,2	17,8
	52,5	21,4
centroide	=AVERAGE (C3:C5)	16,6

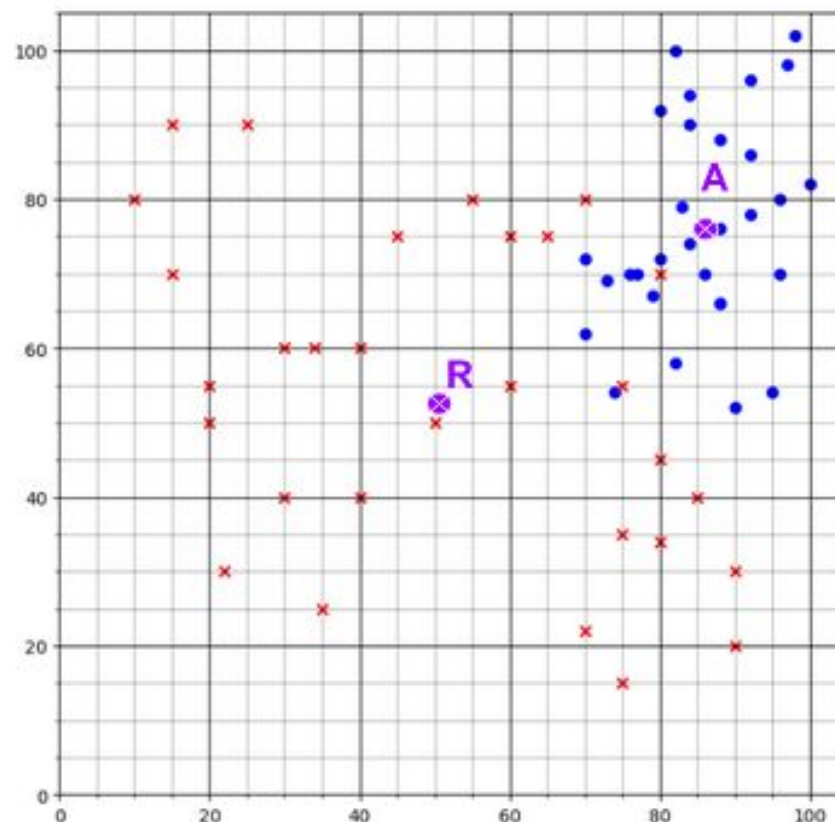
Actividad 1

2. En la siguiente imagen ubica aproximadamente los centroides de los dos usuarios. El centroide del usuario rojo es $R=(51.4, 53.5)$ y del usuario azul es $A=(85.5, 76.2)$.



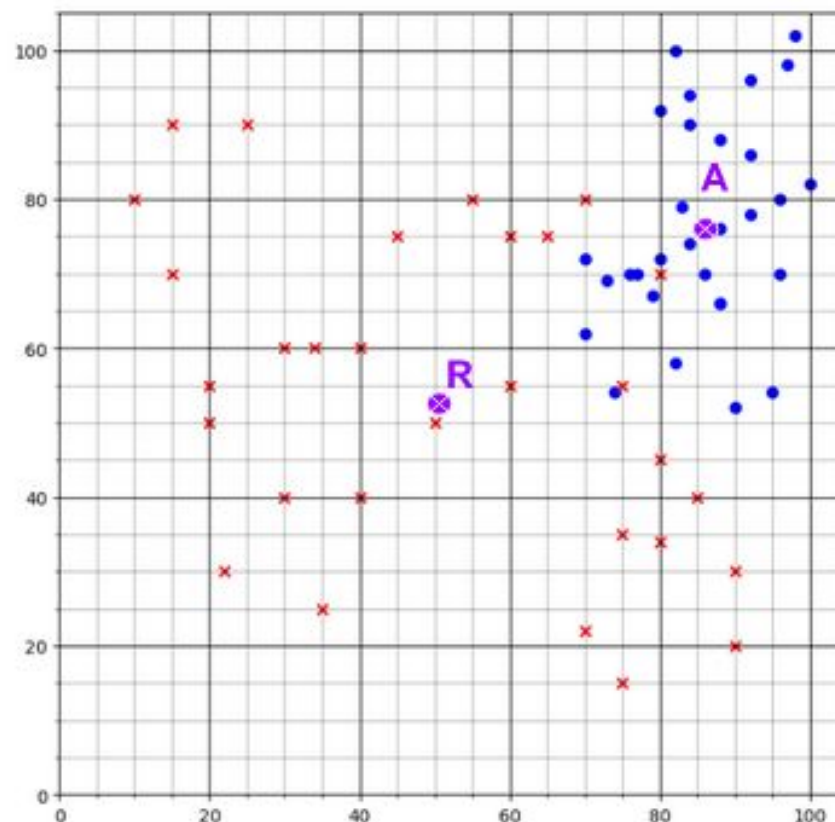
Actividad 1

3. ¿Para cuál de los dos usuarios, te parece que las distintas posiciones se encuentran más cercanas al centroide?



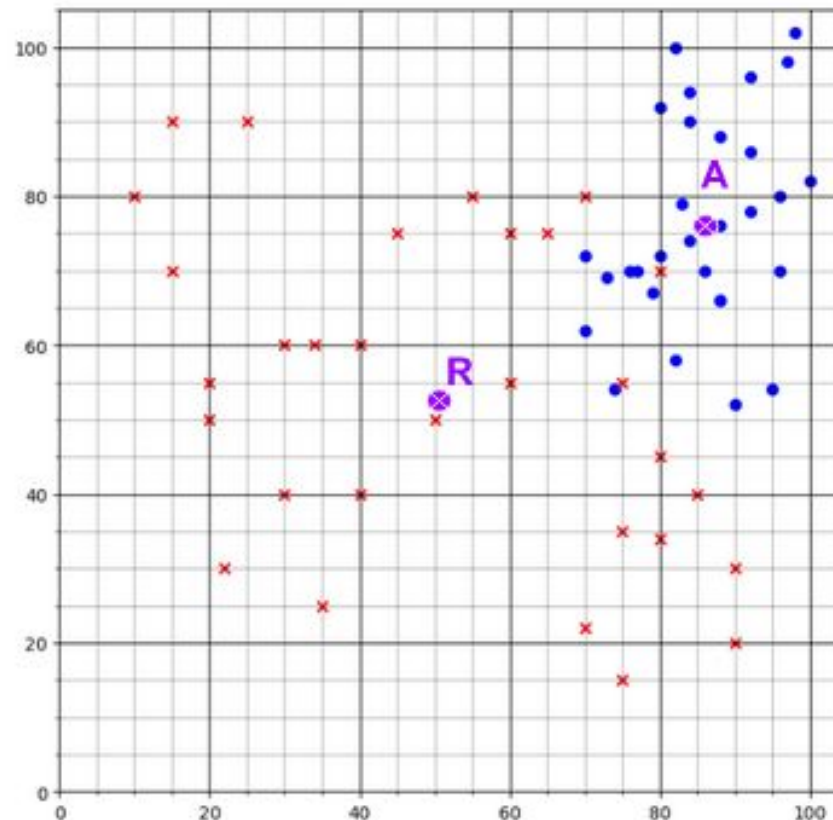
Actividad 1

3. Es evidente que para el usuario azul sus posiciones se encuentran más cercanas al centroide.



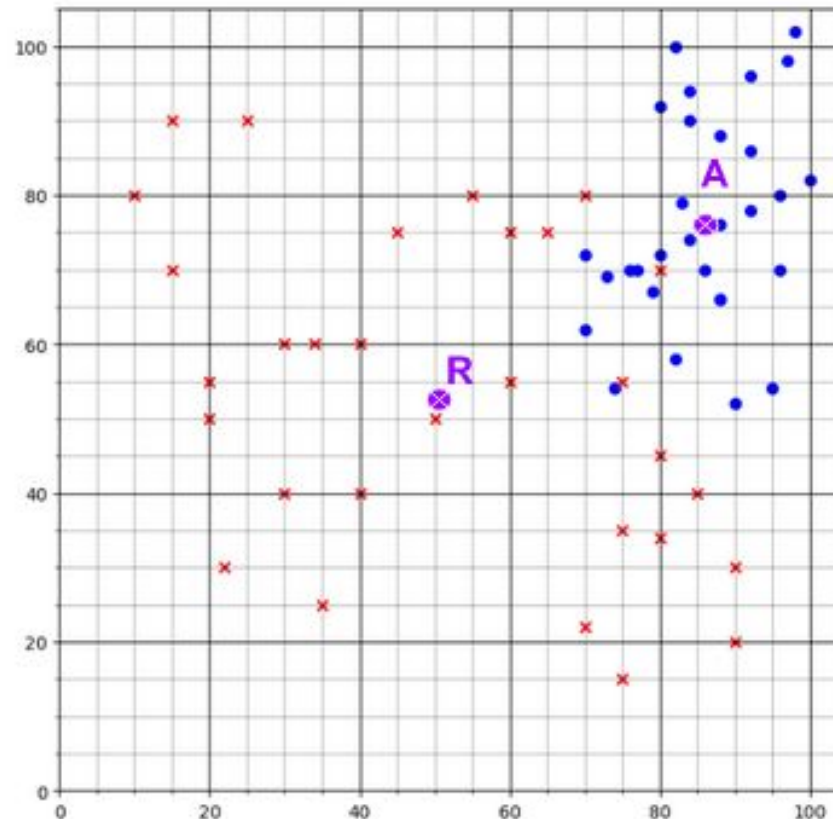
Varianza como medida de dispersión del conjunto de puntos.

¿ Cómo podemos medir la dispersión de las posiciones matemáticamente?



Varianza como medida de dispersión del conjunto de puntos.

Podemos calcular el promedio de las distancias de cada posición al centroide



Varianza como medida de dispersión del conjunto de puntos.

La distancia de la primera ubicación al centroide viene dada por:

$$d_1 = \sqrt{(x_1 - x_c)^2 + (y_1 - y_c)^2}$$

En cambio para evitar raíces, se considera:

$$d_1^2 = (x_1 - x_c)^2 + (y_1 - y_c)^2$$

El promedio de las 30 distancias al cuadrado se escribiría de forma compacta de la siguiente manera:

$$\frac{(x_1 - x_c)^2 + (y_1 - y_c)^2 + \cdots + (x_{30} - x_c)^2 + (y_{30} - y_c)^2}{30}$$

Varianza como medida de dispersión del conjunto de puntos.

La expresión anterior se puede reordenar algebraicamente como sigue:

$$\frac{(x_1 - x_c)^2 + \dots + (x_{30} - x_c)^2 + (y_1 - y_c)^2 + \dots + (y_{30} - y_c)^2}{30}$$

Finalmente al distribuir la fracción se obtiene que:

$$\frac{(x_1 - x_c)^2 + \dots + (x_{30} - x_c)^2}{30} + \frac{(y_1 - y_c)^2 + \dots + (y_{30} - y_c)^2}{30}$$

La expresión anterior es igual a la suma de la varianza de las abscisas más la varianza de las ordenadas de todas las posiciones:

$$Var_x + Var_y = \sigma_x^2 + \sigma_y^2$$

Denominaremos Varianza del conjunto de puntos a esta suma de varianzas:

$$Var = Var_x + Var_y$$

Varianza como medida de dispersión del conjunto de puntos.

¿En qué caso creen que la varianza de los datos es más pequeña, cuando las posiciones del usuario están más juntas o más separadas?

Cálculo de la varianza del Usuario Azul y Rojo.

Por ejemplo si quisiéramos calcular la varianza de 3 ubicaciones, podemos usar la función “VAR” aplicada a las coordenadas de los ejes de las abscisa y ordenada, como se muestra a continuación.

	abscisa	ordenada
	75,4	10,7
	83,2	17,8
	52,5	21,4
centroide	70,4	16,6
varianza	254,6	=VAR(D3:D5)

Actividad 1

4. A continuación se entregan los valores de las varianzas en el eje X y el Y para los usuarios Azul y Rojo.

	Usuario Azul	Usuario Rojo
Varianza X	70,8	653,8
Varianza Y	199,8	472,9

¿Se confirma con estos valores que un usuario tiene mucha más privacidad que el otro? ¿Cuál? Justifica.

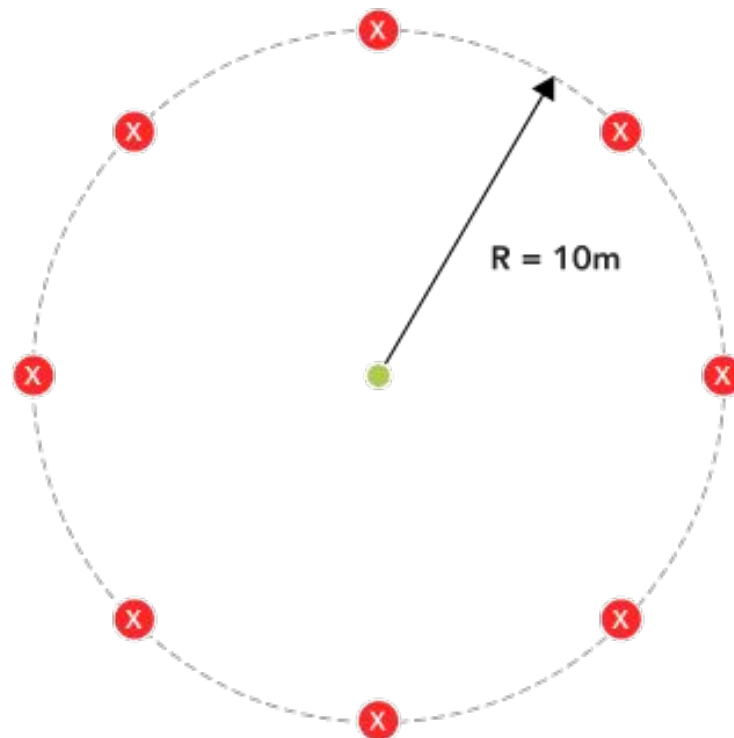
Actividad 1

4. Al sumar las varianzas en los dos ejes para cada usuario, se obtiene que la varianza del usuario Azul es 270,6 , mientras que la del usuario Rojo es 1126,7. Se observa entonces que para el usuario Azul, la suma de las varianzas es menor que la del usuario Rojo, por lo tanto el nivel de privacidad del usuario Azul es menor que la del Rojo. Lo que confirma lo que se visualizó anteriormente en el gráfico con las posiciones de ambos usuarios.

Identificación de puntos de interés utilizando la varianza

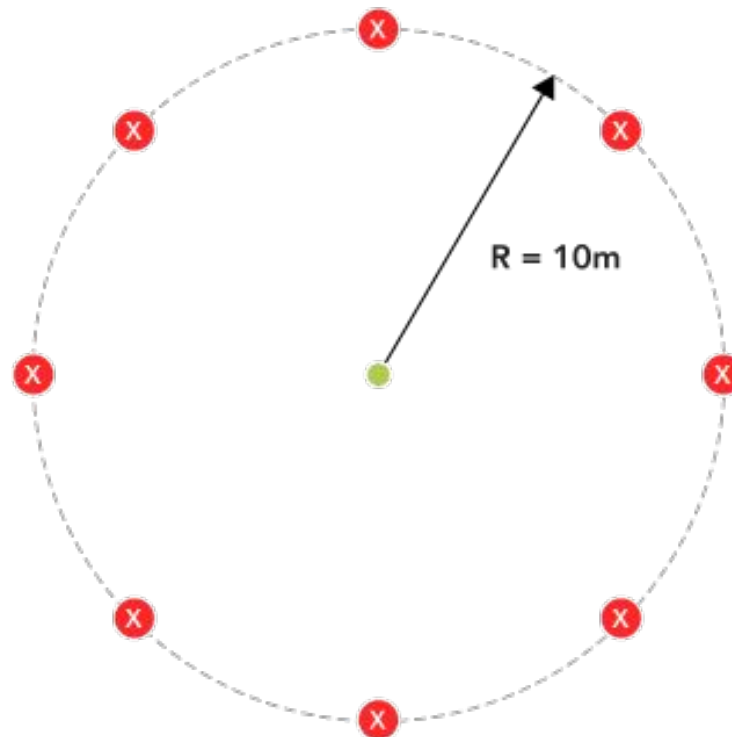
Si un usuario ha permanecido los últimos 15 minutos cerca de un punto de interés ¿La varianza de este es grande o pequeña?

Identificación de puntos de interés utilizando la varianza



En este caso, ¿cuál es el valor de la varianza?

Identificación de puntos de interés utilizando la varianza



Este resultado es $10^2 = 100$ lo cual se sigue cumpliendo independiente de la cantidad de puntos sobre la circunferencia, en general para puntos “regulares” sobre la circunferencia de radio R , se tiene una varianza de R^2

Actividad 1

5. **¿Cómo utilizarías la varianza para detectar un punto de interés?**
6. **¿A partir de qué valores de la varianza te parecen razonables para definir un punto de interés? Pueden inspirarse del caso de la circunferencia de radio (10 m)**

Actividad 1

5. Una varianza pequeña con un valor cercano a 0 permite identificar un punto de interés, por lo que bastaría determinar la varianza y verificar si es cercana a 0.

6. Se puede definir un criterio que si los datos están a menos de (10 m) del centroide entonces es un punto de interés. Luego, si hacemos el cálculo de la dispersión tomando este valor, tendríamos que un límite para la varianza puede ser $10^2 = 100$.

Sistematización

Podemos sistematizar las siguientes ideas:

- El punto promedio de una nube de puntos en el plano se llama “centroide”, cuyas coordenadas corresponden al promedio de sus abscisas y de sus ordenadas.
- La varianza de un conjunto de datos es una medida de dispersión de los mismos. Así por ejemplo, datos más concentrados tendrán menor varianza que datos más dispersos.
- Para datos bidimensionales, consideramos como varianza a la suma de las varianzas de cada coordenada.

Sistematización

Podemos sistematizar las siguientes ideas:

- Cuando los datos corresponden a las ubicaciones durante el día de una persona, su medida de dispersión o varianza, sirve para definir el nivel de privacidad de dicha persona.



Privacidad en aplicaciones móviles

