

PRESENTACIÓN DE PROYECTO

ANÁLISIS DEL SISTEMA DE TRANSPORTE ECOBICI

Pablo Noack

19 de abril del 2021

Universidad del Valle de Guatemala

BASES DE DATOS

1. Bases de datos empleadas en el análisis:

- **df19** : 11 variables y 275609 datos conformada por muestras aleatorias de 25,000 datos de cada mes del año 2019
- **dfdistancias**: Matriz de 480x480 con distancias entre estaciones
- **dffrec**: Matriz de 480x480 con las frecuencias con que de una estación se llegó a otra

CREACIÓN DEL MAPA DE ESTACIONES

MAPA DE ESTACIONES

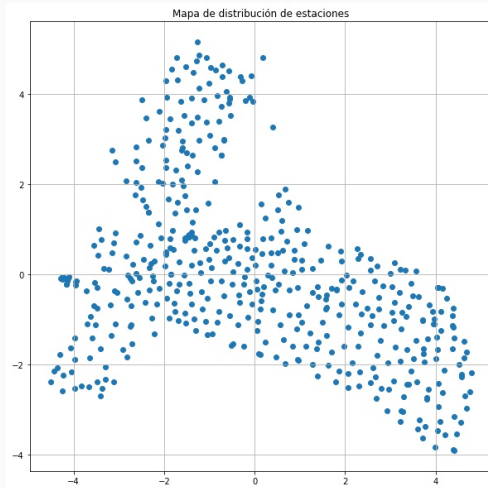


Figure 1: Recreación del mapa de estaciones empleando MDS

ANÁLISIS VARIABLES CONTINUAS



HISTOGRAMA PARA EL TIEMPO

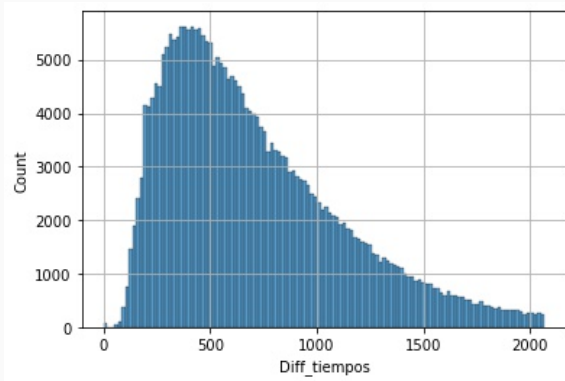


Figure 2: Histograma para la distribución de tiempos. Con media de 707.0 segundos, desviación estándar de 410.8 segundos y mediana de 614.0 segundos

QQ PLOT PARA EL TIEMPO ASUMIENDO UNA DISTRIBUCIÓN NORMAL

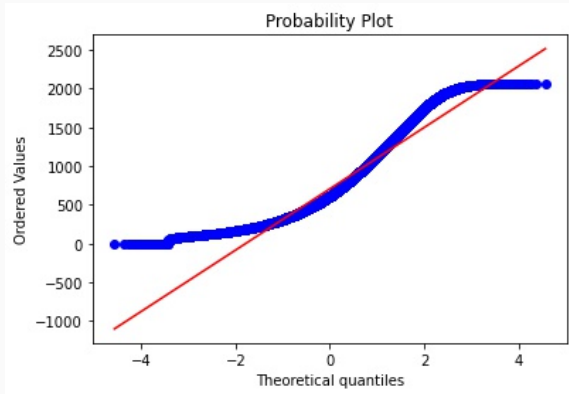


Figure 3: QQplot para una dsitribución normal

HISTOGRAMA PARA EL DESPLAZAMIENTO

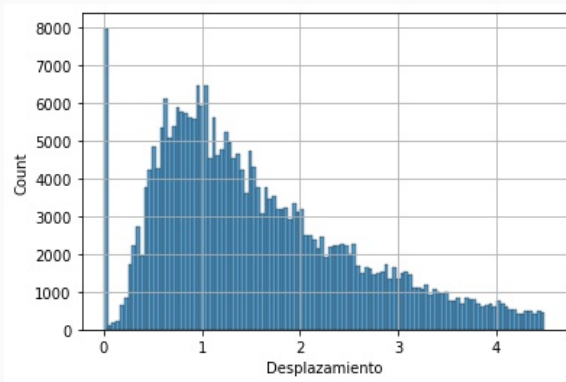


Figure 4: Histograma para la distribución de desplazamientos. Con media de 1.56 km, desviación estándar de 0.99 km y mediana de 1.33 km

QQ PLOT PARA EL DESPLAZAMIENTO ASUMIENDO UNA DISTRIBUCIÓN NORMAL

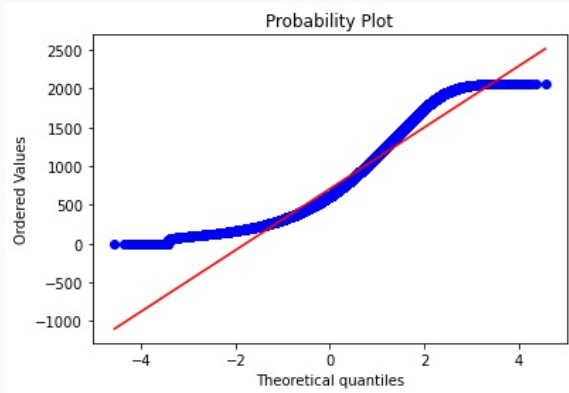


Figure 5: QQplot para una dsitribución normal

HISTOGRAMA PARA LA EDAD

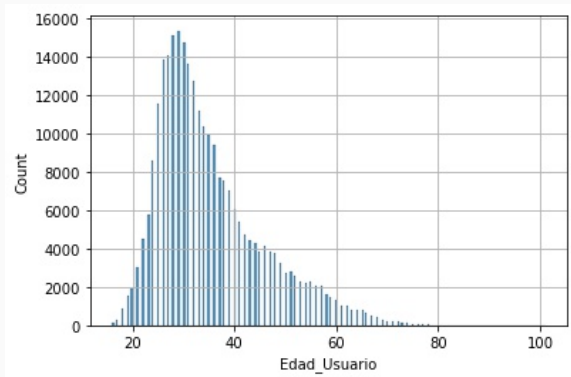


Figure 6: Histograma para la distribución de edades. Con media: de 35 años, desviación estándar de 10.41 años y mediana de 32 años

QQ PLOT PARA LAS EDADES ASUMIENDO UNA DISTRIBUCIÓN NORMAL

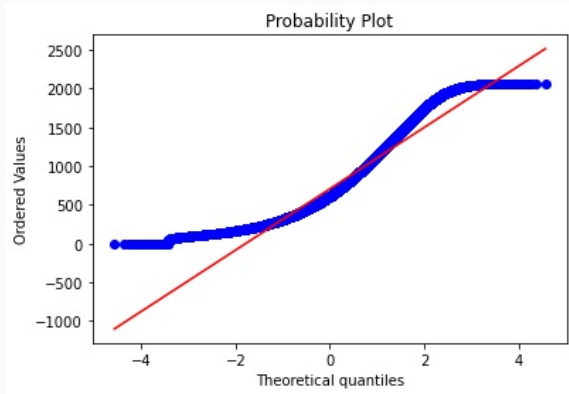


Figure 7: QQplot para una dsitribución normal

PRIMERA PREGUNTA : ¿ CÓMO SE
COMPORTA EL TIEMPO PARA LOS
CASOS EN QUE EL DESPLAZAMIENTO
ES 0?



HISTOGRAMA DEL TIEMPO

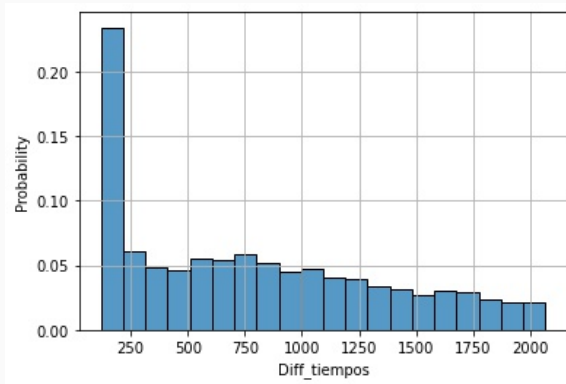


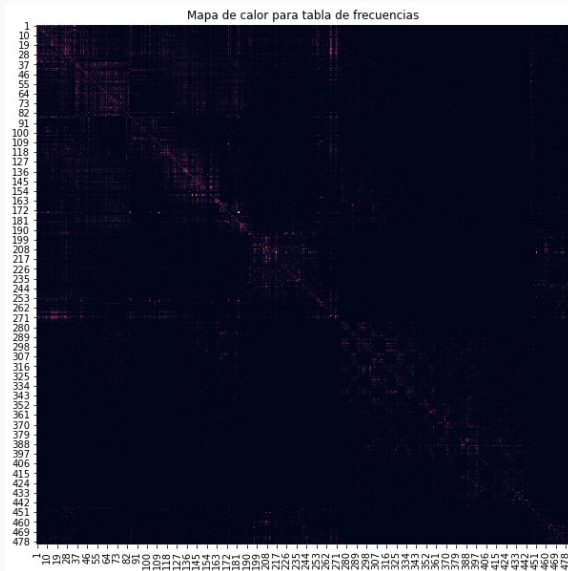
Figure 8: Histograma del tiempo para el caso en que el desplazamiento fué de 0 km

Primera conclusión

Existen por lo menos 2 casos en los cuales se emplea el transporte en bicicleta: uno en el que existe un desplazamiento entre estaciones y otro en el que el desplazamiento es 0 , caracterizandose este último por tiempos cortos de uso, siendo la categoría del histograma más común menor a los 250 segundos (4.4 minutos).

SEGUNDA PREGUNTA: ¿QUÉ TAN
RELACIONADAS SE ENCUENTRAN LAS
ESTACIONES DE SALIDA Y LLEGADA?
¿EXISTE ALGUNA RELACIÓN
SUBYACENTE ENTRE ELLAS?

MAPA DE CALOR TABLA DE FRECUENCIAS



Segunda conclusión

La tabla de frecuencias muestra que las estaciones más frecuentadas aparecen al rededor de la diagonal principal. Se puede interpretar como que las estaciones más cercanas a la diagonal principal son las más cercanas en el mapa real

TERCERA PREGUNTA : DADO QUE
EXISTE UNA RELACIÓN SUBYACENTE
ENTRE LAS ESTACIONES Y ESTA SE
PUEDE CARACTERIZAR MEDIANTE LAS
FRECUENCIAS ; ESTAS
DISTRIBUCIONES DE FRECUENCIAS
EXHIBEN CARACTERÍSTICAS TAL QUE
SE PUEDAN CLÚSTERIZAR?

REDUCCIÓN DE DIMENSIONES : TSNE

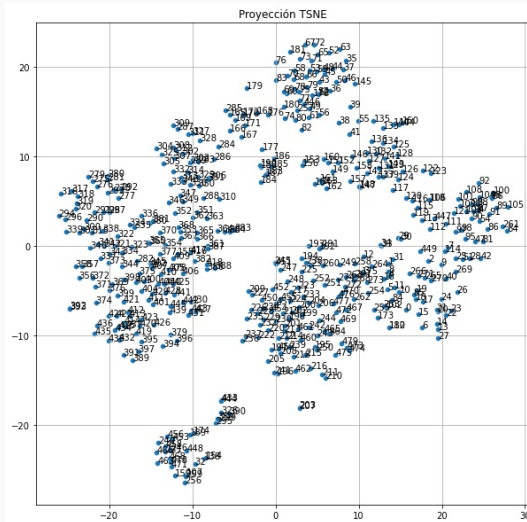


Figure 9: Tsne con parámetros: perplexity = 30 e init= pca

MAPA GENERADO POR TSNE APLICANDO FUZZY CMEANS

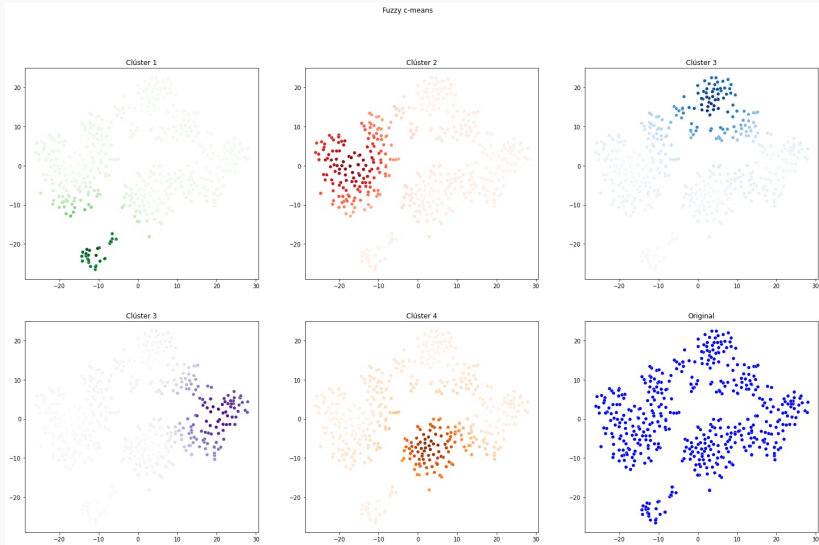


Figure 10: Clústerización de mapa generado por tsne

Tercera conclusión

La tabla de frecuencias exhibe características que permite a los algoritmos de reducción de dimensiones proyectar clústers de datos sobre 2 dimensiones tal que un algoritmo de clústering pueda captarlos y clasificarlos.

CUARTA PREGUNTA: SI ES POSIBLE
QUE TSNE HAYA CAPTADO
CARARCTÍSTICAS COMÚNES ENTRE
LAS TABLAS DE FRECUENCIAS ¿CÚAL
SON ESAS CARACTERÍSTICAS?

MAPA GENERADO REAL APLICANDO FUZZY CMEANS

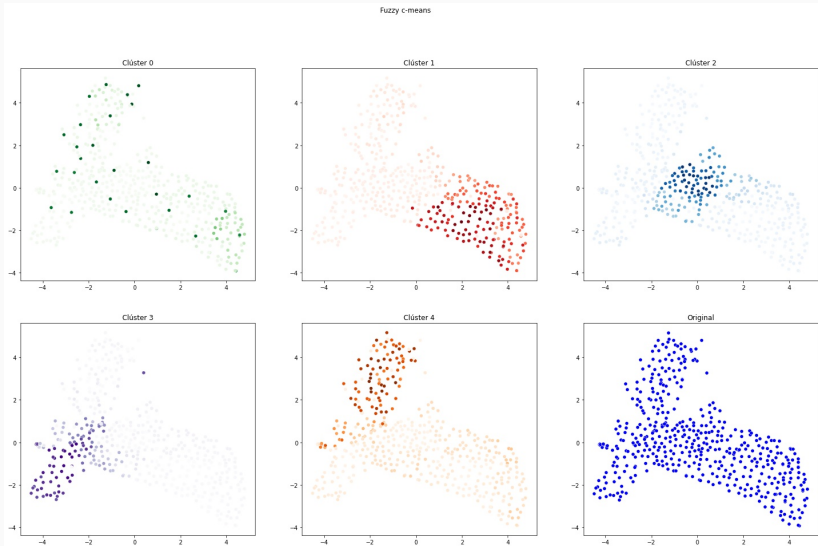


Figure 11: Clústerización de mapa real

MAPA GENERADO REAL CLUSTERIZADO CON FUZZY_cmeans

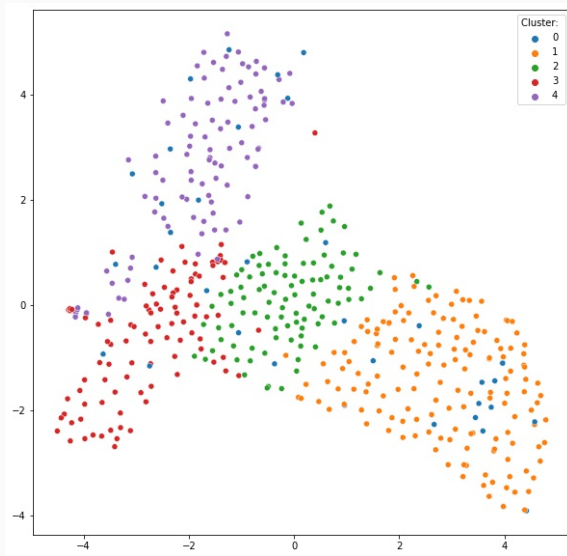


Figure 12: Clústerización de mapa real

HISTOGRAMA DE LOS TIEMPOS COMPARANDO CLÚSTERS

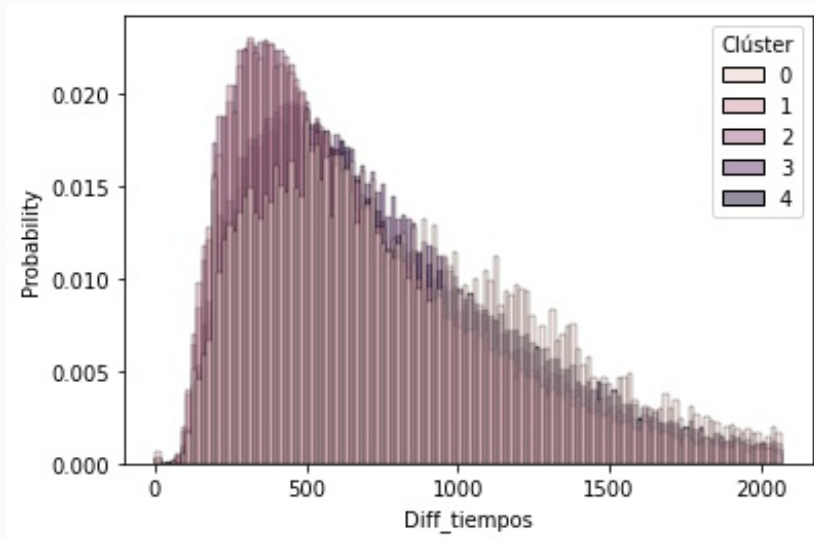


Figure 13: Histograma clústers para el tiempo

HISTOGRAMA DE LAS DISTANCIAS COMPARANDO CLÚSTERS

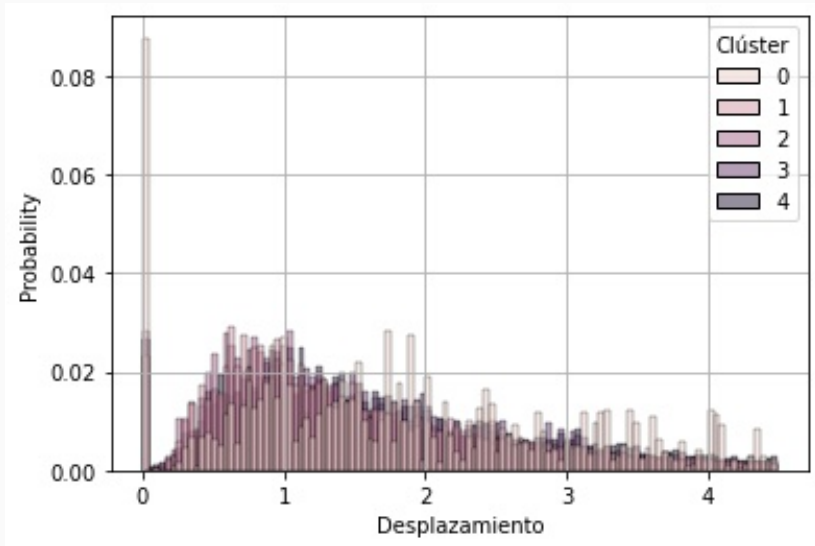


Figure 14: Histograma clústers para el desplazamiento

BOXPLOT DE LAS DISTANCIAS COMPARANDO CLÚSTERS

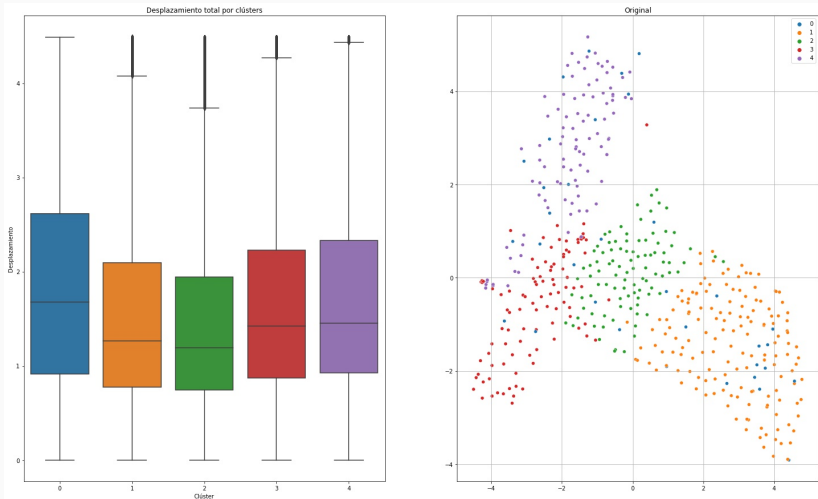


Figure 15: Boxplots clústers para el desplazamiento

BOXPLOT DE LOS TIEMPOS COMPARANDO CLÚSTERS

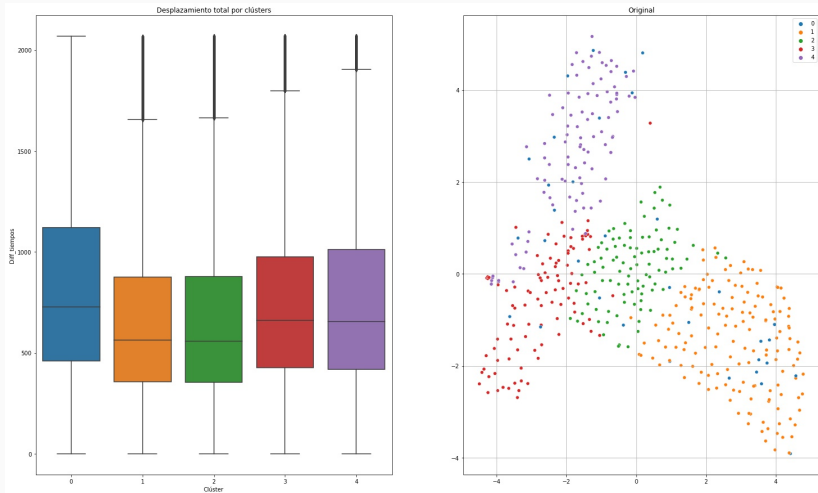


Figure 16: Boxplots clústers para el tiempo

última conclusión

A partir de la tabla de frecuencias t-sne fué capaz de captar r la distribución geográfica de las estaciones indirectamente. Con cierto error debido a la naturaleza del movimiento entre estaciones. Por otro lado, se comprueba que cada clúster parece exhibir comportamientos estadísticos distintos con respecto a la distancia y al tiempo.

FIN

