

## PEC 1 VISUALIZACIÓN DATOS

CARLOS ALLO LATORRE

### 0.1 .- El enlace al conjunto de datos que usaréis posteriormente en la Práctica II.

Para la realización de esta práctica se usarán varios dataSets y varios endpoints de una API expuesta por el ayuntamiento de Zaragoza:

DataSets sobre los datos del uso del bus y el tranvía por meses y años: <https://www.zaragoza.es/sede/portal/datos-abiertos/servicio/catalogo/1960>

- Datos de uso del tranvía desde el año 2013 hasta 2018
- Datos de uso de bus por líneas y meses para el año 2013 desde Agosto
- Datos de uso de bus por líneas y meses para el año 2014
- Datos de uso del bus por líneas y meses para el año 2015
- Datos de uso por líneas de bus y meses para el año 2016
- Datos de uso por líneas de bus y meses para el año 2017

Datasets sobre los datos de usos del servicio de transporte Bizi (bicicletas compartidas en la ciudad de Zaragoza) en intervalos de 6 meses: <https://www.zaragoza.es/sede/portal/datos-abiertos/servicio/catalogo/70>

- Contiene todos los viajes realizados durante 6 Meses (Octubre 16 - Marzo 17)
- Contiene todos los viajes realizados durante 6 Meses (Marzo 17 - Agosto 17)
- Contiene todos los usuarios que realizaron viajes incluyendo el género y el rango de edad del usuario durante el periodo de Marzo 17 hasta Agosto 17
- Contiene todos los usuarios que realizaron viajes incluyendo el género y el rango de edad del usuario durante el periodo de Octubre 16 hasta Marzo 17

Api que nos proporciona información en tiempo real (con actualización de los datos cada minuto/dos minutos) sobre los servicios bus, tranvía y bizi:

- Información buses y tranvía: [https://www.zaragoza.es/docs-api\\_sede/#/Equipamientos%20y%20movilidad%3A%20Transporte%20urbano](https://www.zaragoza.es/docs-api_sede/#/Equipamientos%20y%20movilidad%3A%20Transporte%20urbano)
- Información bizi: [https://www.zaragoza.es/docs-api\\_sede/#/Equipamientos%20y%20movilidad%3A%20Estaciones%20Bizi](https://www.zaragoza.es/docs-api_sede/#/Equipamientos%20y%20movilidad%3A%20Estaciones%20Bizi)

### 0.2.- Idea en líneas generales de la visualización:

En este punto, se presenta la idea en líneas general que se desea realizar en esta práctica, para que sea más fácil responder a las preguntas posteriores sin caer en la repetición de información. Se espera crear un panel con 3 pestañas, en donde cada una de ellas contendrá gráficas con diferentes propósitos:

- La primera, mostrará información en tiempo real gracias a llamadas recurrentes a las APIS mencionadas. Podremos ver el tiempo que le falta a un bus a llegar a una parada en concreto, o las bicicletas que quedan en una estación Bizi. Para hacerlo visual, podríamos dar otra visualización al usuario, en donde en las bicicletas, se pinte un punto en el mapa más o menos grande en función de la densidad de bicis en cada parada. En buses, se podría hacer de forma similar, teniendo en cuenta el tiempo de espera a un bus, pero habría que darle una vuelta para hacerlo visual y entendible rápido (en una parada podríamos tener diferentes líneas, cada una con tiempos concretos)

- La segunda, se realizarán gráficas interactivas que muestren datos de los dataSets estáticos, en donde se pueda filtrar por años, numero de viajeros y poder sacar diferentes conclusiones una vez se tenga los gráficos realizados.
- La tercera, se quieren realizar análisis algo más complejos que permitan sacar información o conclusiones no tan triviales como las de la pestaña anterior. Esta podría ser, teniendo en cuenta que contamos con las ubicaciones de las paradas del transporte público, como de las estaciones Bizis, fraccionarlas por barrios, y visualizar en cual de ellos hay más paradas de bus/tranvía o bicis, más usos de los mismos y ver puede tener alguna relación con el nivel económico de los mismos.

### **1.- La justificación de vuestra selección del conjunto de datos.**

La justificación de la elección de estos datos ha sido por dos motivos:

- El primero de ellos sería el personal, ya que en mi trabajo de programador web uso diariamente APIs en el código para obtener datos de diferentes servicios, y me parece interesante aprender a integrar estas en un panel de visualización, no basándonos únicamente en un dataSet estático. Otro motivo que se sumaría a este punto sería el descubrimiento en la práctica anterior del repositorio de datos abiertos que presta el ayuntamiento de mi ciudad, y el uso en esta práctica me va a permitir familiarizarme con el (tanto con el de datos estáticos de dataSet como el dinámico mediante la API) para posiblemente en un futuro poder extraer datos fiables de estas webs que antes no sabía que existían. También, dentro de este punto, se podría englobar que el realizar una visualización que personalmente pueda usar en mi día a día al ser usuario diario del transporte público (sobre todo la primera pestaña) me motiva a realizar esta práctica.
- El segundo sería ya que considero que son datos interesantes. Cada una de las diferentes pestañas que quiero crear con los datos, presentan diferentes particularidades. La primera de ellas, puede ser de ayuda a cualquier usuario que vaya a usar el transporte en Zaragoza, tanto a gente local (ver cuanto le queda al siguiente bus, ver en que parada puede encontrar más bicis...) como a gente visitante (consultar las líneas y tiempos de espera para cada una de ellas). La segunda pestaña, puede servir para ver la evolución del uso del transporte público, y ver si se puede considerar si la ciudad sigue un patrón de sostenibilidad y uso del transporte) o no. Por último, la tercera, puede ser de utilidad o bien al comprar una casa (para ver que barrios están mejor comunicados), como para perseguir una nueva ruta por parte del ayuntamiento si se encuentran discrepancias grandes entre barrios.

### **2.- La relevancia del conjunto de datos en su contexto.**

Los datos, como se puede deducir de los párrafos anteriores, son totalmente actuales si hablamos de los obtenidos por medio a las llamadas a las APIs (tienen refresco cada uno o dos minutos en función del endpoint al que se llame) los cuales nos proporcionarán información sobre la ubicación de las diferentes parada o estaciones Bizis, tiempos de esperas hasta el siguiente bus o tranvía u ocupación de las estaciones Bizi. Por otro lado, los datos resumen del uso del transporte público no son lo más recientes posibles ya que el ayuntamiento ha facilitado en este repositorio datos que van del 2013 al 2018 por lo que nos faltarían datos resumen de los años 2019, 2020 y 2021. Sin embargo, considero que no es un intervalo de tiempo muy grande, y que pueden seguir siendo de utilidad los resúmenes y gráficas obtenidas de estas ya

que el año 2020 y 2021 ha sido afectado por el covid y nos podrían haber distorsionado los análisis y conclusiones (ya que en el periodo de confinamiento y días posteriores por miedo de la población, se redujeron las frecuencias y la utilización del mismo).

En complemento a la pregunta anterior, como he mencionado en esta, creo que son interesantes. Son datos que al ser usados en 3 pestañas diferentes, hacen que cada una de ellas afecten a un colectivo específico; la primera de ellas, a cualquier persona que esté en Zaragoza y quiera hacer uso del servicio público. La segunda, al propio ayuntamiento para poder evaluar de forma rápida el uso del transporte público en zaragoza y poder evaluar si el uso del mismo es el esperado, si hay líneas con mayor uso que otras (para ver si tienen los buses bien repartidos por líneas)... Por último, la tercera dependerá de las gráficas que se hagan, pero si se realiza el fraccionamiento por barrios, puede ser útil tanto a un ciudadano que quiera comprarse una vivienda y quiera tener un barrio conectado (o poco conectado si solo quiere usar su propio coche), como al propio ayuntamiento para conocer cómo de repartidas tiene las marquesinas y si tiene algún barrio con poco o muchas paradas de transporte.

Finalmente, respecto a la perspectiva de género, se aprecia como no son datos que este hecho afecte. En los datos de usos del transporte público, en el propio número se encuentran los datos masculinos y femeninos, y en los tiempos de espera a los buses o tranvías no influye este hecho. Mencionar que en uno de los dataSets se cuenta con la distinción sobre el sexo y rango de edad sobre el uso del servicio Bizi, lo que nos permitirá poder realizar una pequeña clasificación del tipo de usuarios que hicieron uso de este servicio.

### **3.- La complejidad (medida, variables disponibles, tipos de datos, etc.).**

Para responder a esta pregunta, nos tenemos que fijar en cada uno de los dataSet que vamos a usar en la práctica, ya que partimos de 3 tipos de dataSets diferentes.

Comenzando con el dataSet sobre el uso del bus, nos encontramos que estos dataSet, nos presentan los mismos datos tanto mensualmente como acumulados a lo largo que pasa el propio año que consultamos. Se usarán los datos mensuales, dejando al programa de visualización que se use estos para realizar las agregaciones que el propio usuario demande. Las variables de las que partiremos serán los kilómetros realizados por cada línea de bus, el número de viajeros y el ratio de viajeros por kilómetro. Para este caso, encontramos que estas variables serán cuantitativas, mientras que la línea de bus a la que hacen referencia cada dato serán cuantitativas (ya que hay líneas que no corresponden con un número como serían las líneas Nocturnas que empiezan con N). Tendremos una línea de información por cada línea de bus, siendo esta de 43 líneas totales. De forma similar encontramos para los datos del tranvía, que no están agrupados por líneas del mismo (al tan solo existir una, la línea 1), proporcionándonos en el mismo documento los resúmenes mensuales (viajeros y km recorridos por los tranvías, en una línea por mes).

Si ahora pasamos a los datos de usos de la bizi, al no funcionar esta por líneas como en el caso anterior, encontramos dataSet menos agrupados. Así pues, encontramos dos tipos de dataSet con miles de registros. El primero de ellos, nos indicará datos sobre cada uno de los viajes, y nos dará cada uno de ellos el id de usuario que ha realizado el viaje (numérico), la fecha y hora de cada uno de ellos (dato fecha) y la estación de inicio y fin del mismo (numérico, al darnos el id de la estación para que sea más fácil su identificación y no su nombre como tal). Por otro lado, como complemento de cada usuario anterior, encontramos una relación con su id, que nos da el sexo de cada usuario (categórico) y su rango de edad (que aunque podría ser cuantitativo, es

categorico ya que no da la edad exacta del usuario, sino que nos da el rango de edad en la que se encuentra).

Por último, se encuentran los endpoints de la API. Comenzando por el endpoint del servicio bizi, encontramos dos endpoints que nos traen la misma información, con la diferencia que el primero de ellos trae la información de todas las estaciones, mientras que el segundo trae sólo la información de la estación con el id especificado. En ambos casos recuperaremos el id (dato numérico), about (link a una web del ayuntamiento de Zaragoza que la pinta en el mapa), title, estado, estadoEstacion, address, descripcion (datos cualitativos que dan info sobre la estación), tipoEquipamiento (link a una web RDF con info sobre el equipamiento), bicisDisponibles, anclajes disponibles (dato numérico cuantitativo sobre la cantidad de bicis que hay y los huecos que hay para dejarlas), lastUpdated (fecha en UTC sobre la ultima actualización) y un icon (link a un png con imagen genérica del servicio de bizi). Adicionalmente, también tenemos las coordenadas de cada estación.

Respecto al servicio del tranvía, de los endpoints que usaremos, encontramos los dos endpoints que usaremos tienen una lógica idéntica al anterior pero para este transporte. Contamos con dos endpoints, uno para la información general de todas las paradas, y otro para cada una de las paradas. Los datos que se nos proporcionan aquí serán las coordenadas para poder representarla en el mapa, y otros datos cualitativos como el id, title, mensajes (es el mensaje que se muestra en la parada en el panel digital), o descripción. Presenta también la fecha de actualización (lastUpdated) junto con una uri que muestra la ubicación de la parada en una web externa. Finalmente, para saber cuánto tiempo le queda al tranvía, ofrece un atributo "destinos" donde da los tiempos que le quedan al tranvía para cada una de las direcciones (en Zaragoza tan solo hay una línea de tranvía). Para el caso de los buses, encontramos un hecho similar con la diferencia que en la ubicación de las marquesinas, no nos da tiempo de espera al bus. Si que nos dará en la variable title el número de marquesina, la calle y las líneas que pasan por la misma (ej: "(688) P. Reyes De Aragón N.º 24 Líneas: 58"). Para obtener los tiempos de espera, se ha de realizar la petición al endpoint propio especificándole la propia marquesina de la que se quiere obtener información.

Por último, mencionar que toda esta información obtenida por API se obtendrá en un JSON, por lo que requerirá un pequeño procesamiento inicial para que los datos puedan ser representados en los diferentes paneles.

#### **4.- La originalidad.**

Considero que los datos no son los datos clásicos al no ser un conjunto de datos que están presentes en librerías o programas por defecto, al ser encontrados en la web oficial del ayuntamiento de Zaragoza.

Respecto a la existencia de otras visualizaciones, creo que tendríamos que centrarnos en cada una de las pestañas para abordar este punto:

- La primera de ellas, al obtener los datos constantemente sobre una API pública, entiendo que tiene que ser utilizada por diferentes aplicaciones que proporcionan información al usuario sobre lo que falta de llegar a un transporte público, o los recorridos que realiza una línea de bus. Sin embargo, estas visualizaciones tienden a estar separadas por tipo de transporte, por lo que aprovechando que estas APIs nos dan datos de todas ellas, creo que esta visualización se podría considerar una evolución para tener concentradas todas ellas en un mismo lugar en donde el usuario mediante su

interacción pueda filtrar por la o las que el mismo desee. Adicionalmente, la mayor parte son aplicaciones que requieren descargar una aplicación en el móvil (ViveZaragoza, ZUM) y este hecho puede ser una desventaja para algunos usuarios que sólo van a hacer uso del transporte un día, o que tienen poco espacio en su dispositivo. Como nuestra visualización será accesible por una URL, solventaría este problema también.

- Respecto a la segunda, se desconoce tras una búsqueda visualizaciones que muestren este tipo de información resumen más a allá que los datos csv de los que se van a partir. Además, al poder combinar datos de uso sobre bus, tranvía y bizi, se pueden combinar por superposición las cantidades de todas ellas enriqueciendo la información que proporciona cada una de las excels de forma separada creando gráficos agregados de toda esta información (si volvemos al punto 0.1, podemos visualizar que en los dataSets proporcionan la información dividida tanto por transporte, como por años (y dentro de cada documento, por líneas y meses).
- Finalmente, la tercera, persiguiendo la idea de estudios de análisis por barrios y paradas (que se podrá extrapolar a otras áreas) tampoco se ha visto ninguna visualización similar. Como mucho, se podría hacer un ejercicio manual de conteo en maps de cuantas paradas hay por barrio. Por lo que, en nuestro caso, vamos a hacer una pequeña actualización de los mapas tradicionales que tan solo muestran las paradas o estaciones en el, con el fin de dividirlos por barrios y crear gráficas a partir de esta información, enriqueciendo los conjuntos de datos iniciales al hacer agrupaciones por barrios y combinando toda esta información.

## **5.- Las cuestiones que responderéis con la visualización de este conjunto de datos**

Para responder a estas preguntas, y para seguir con la dinámica que se ha seguido hasta el momento, creo que sería interesante volver a fraccionar este apartado en 3 apartados, uno por cada pestaña.

- La primera de ellas, las que se muestran los datos más actuales al ser refrescados continuamente, podríamos responder a cuestiones del tipo ¿Cuánto le falta a mi próximo bus o tranvía? ¿Dónde están las paradas del transporte público (tanto bus, tranvía o estaciones bizi)? ¿Hay disponibilidad de bizi en una parada en concreto? ¿Y hay disponibilidad de hueco para dejar una bici en el caso de que esté usándola en este momento? ¿Qué líneas de buses dispongo en zaragoza? ¿Cuál es la media de espera de todas las líneas de buses en todas las paradas de Zaragoza? ¿Y las de tranvía? ¿Cuál es la media de bizi disponibles para ser usadas (estacionadas en una parada) en todo Zaragoza? ¿Y los huecos totales disponibles?
- En la segunda pestaña, la que obteníamos a partir de datos estáticos, podemos resolver preguntas del tipo ¿En qué años se produjeron más viajes en transporte público? ¿Se han incrementado los viajeros en el paso del tiempo? ¿Hay más buses con el paso de los años para aumentar las frecuencias (mayor número de km recorridos por líneas)? ¿Se deberían de poner más buses/tranvías para líneas concretas (ratio viajeros/km por cada línea)? ¿Los trayectos realizados con las bicicletas públicas son de largo o corto recorrido? ¿Cuál es el perfil (género y rango de edad) que más usa el servicio Bizi? ¿Y el que menos?
- Por último, la tercera pregunta, permitirá solucionar preguntas más avanzadas. Por el momento, centrándonos en la distribución por barrios que se quiere conseguir, nos podemos plantear preguntas como ¿Qué barrios cuentan con más transporte público? ¿Cuáles con menos? ¿Por cuantos barrios pasa de media una línea de bus? ¿Y de

transporte? ¿Corresponde que el barrio más poblado se el que más transporte público posee? ¿Cuántas paradas por metro cuadrado hay en cada barrio?

Considero que estas cuestiones tienen en cuenta todos los puntos anteriores, y que responden a las preguntas que se podría plantear un usuario en base a los objetivos que se persiguen con estas visualizaciones, comentados ya brevemente en la introducción y a lo largo de la elaboración de este informe. Además, si prestamos atención a las preguntas que se pretenden responder, vemos que los atributos comentados en la pregunta 3 son los suficientes para poder obtener las respuestas. Como nota, simplemente quedaría estudiar como se realizará la clasificación de parada – barrio, para examinar si esta se hará manual en la carga de datos en base a los diferentes barrios de Zaragoza, o si el programa/tecnología que se use lo permite, introducir las coordenadas de los barrios para que por si mismo realice esta distribución.