

PRA 2 – Creación de la visualización

Alumno: Luis Manuel Molina Coca (luimoco)

Horarios del servicio de transporte ferroviario.

Servicio Euskotren en Bizkaia

Contenido

1.	Título de la visualización	2
2.	Descripción del documento	2
2.1.	¿A qué pregunta pretende dar respuesta la visualización?	2
3.	Elaboración del conjunto de datos.....	2
3.1.	Adquisición de los datos.....	2
3.2.	Diseño.....	3
3.2.1.	Descubrimiento de variables.....	3
3.2.2.	Arquitectura	4
3.3.	Generación del dataset	5
4.	Creación de la visualización.....	6
4.1.	Título de la visualización	6
4.2.	El conjunto de datos.....	6
4.2.1.	Exploración	6
4.3.	Boceto	9
4.4.	Implementación	10
4.4.1.	Importar los datos	10
4.4.2.	Hoja Mapa Estaciones	10
4.4.3.	Hoja Selección de estación.....	11
4.4.4.	Hoja de visualización de rutas y horarios.....	12
4.4.5.	Dashboard Horarios Euskotren Bizkaia	13
5.	Implantación y uso de la visualización	15
5.1.	Respuestas de la visualización a la pregunta planteada	15
6.	Visualización derivada: Carteles de estación	17
6.1.	Implementación	18
6.2.	Uso de la visualización.....	19
7.	Conclusiones y carencias.....	20

1. Título de la visualización

Horarios Euskotren Bizkaia

Publicación en [Tableau Public](#):

https://public.tableau.com/views/PRA2/HorariosEuskotrenBizkaia?:language=es-ES&:display_count=n&:origin=viz_share_link

2. Descripción del documento

El siguiente documento, narra el proceso de creación de una visualización que trata de responder cuestiones relativas a los horarios establecidos por el servicio de transporte ferroviario, brindado por la sociedad *Eusko Trenbideak – Ferrocarriles Vascos, S.A.* (Euskotren)¹ dependiente del Gobierno Vasco, en el Territorio Histórico de Bizkaia.

Tradicionalmente, los horarios de paso del tren se han presentado en un formato tabular difícil de interpretar por la multitud de estaciones, líneas, horas de paso y condicionantes en los que ciertos trenes no llegan a hacer parada en la estación. Poder generar una visualización a partir de la adquisición y tratamiento previo de estos datos, puede significar obtener una herramienta capaz de ofrecer una experiencia mucho más enriquecedora tanto en la interpretación de la información como generación de conocimiento que también queda oculta entre la vorágine de horas de paso y paradas.

2.1. ¿A qué pregunta pretende dar respuesta la visualización?

Con la visualización, pretendemos responder la siguiente pregunta principal:

¿A qué hora debería coger el tren en mi estación de origen para llegar a mi destino a una hora adecuada?

En las distintas secciones del documento, abordaremos el proceso de elaboración del conjunto de datos y el de creación de la visualización que permita dar respuesta a esta pregunta e incluso ofrecer información en formato visual que ayude a generar el conocimiento que es mucho más difícil de encontrar en el formato tabular clásico.

3. Elaboración del conjunto de datos

3.1. Adquisición de los datos

Los datos se obtienen desde la página web <https://www.euskotren.eus/tren/horarios>.

Se descargan los [PDF](#) con las tablas de horarios correspondientes a las líneas [E1](#), [E3](#), [E4](#) y [L3](#) (en ambos sentidos) puesto que sus rutas hacen parada en estaciones del Territorio Histórico de Bizkaia.

¹ Euskotren: <https://www.euskotren.eus/>

3.2. Diseño

3.2.1. Descubrimiento de variables

Se realiza un análisis de la información contenida en las tablas de horarios para la extracción de los distintos datos existentes.



Figura 1. Descubrimiento de variables en los carteles de horarios.

La Figura 1 muestra el diseño del cartel de horarios para una de las líneas. Con esta información podemos extraer los siguientes datos:

- **Código de línea:** Codifica una ruta para denominar un conjunto de estaciones por las que discurren distintos convoyes en distintos sentidos, horarios y horas.
- **Tipo Horario:** Identifica el conjunto de días de la semana en los que pueden variar las horas de paso de convoyes en la ruta.
Dominio:
 - **Laboral** (L): Comprende los días de la semana de lunes a viernes.
 - **Festivo** (F): Comprende sábado, domingo y cualquier día Laboral si coincide con una festividad.
- **Sentido:** Identifica la dirección en la que el convoy realiza la ruta de la línea.
Dominio: Ida (I) o Vuelta (V)
- **Estación:** Ubicación donde se encuentra una posible parada del convoy dentro de la ruta de una línea. Una estación se define con los siguientes atributos:
 - **Nombre:** Denominación por el que es conocida la parada. Generalmente coincide con el municipio donde se sitúa.
 - **Latitud:** Coordenada para la ubicación de la estación en un mapa.
 - **Longitud:** Coordenada para la ubicación de la estación en un mapa.
- **Convoy:** Se trata del recorrido efectuado por un tren en una ruta concreta, sobre una o varias líneas, la cual transcurre entre un conjunto específico de estaciones donde realiza parada en una hora de paso concreta.
 - No está especificado en la información, pero supone poder tener una identificación interna de todos los trayectos que se realizan.

- Hay convoyes que están agregados en una banda sombreada, donde no se desglosan los horarios de las paradas y sí la frecuencia desde la estación de salida. Esta banda de tiempo es la más difícil de interpretar para descubrir el horario puesto que el usuario ha de calcularlo.
- Un mismo convoy puede dar servicio a más de una línea. Es decir, un convoy puede tener paradas comunes en la misma hora de paso con otras líneas que están en la misma ruta. Las rutas consideradas se muestran en la Figura 2.

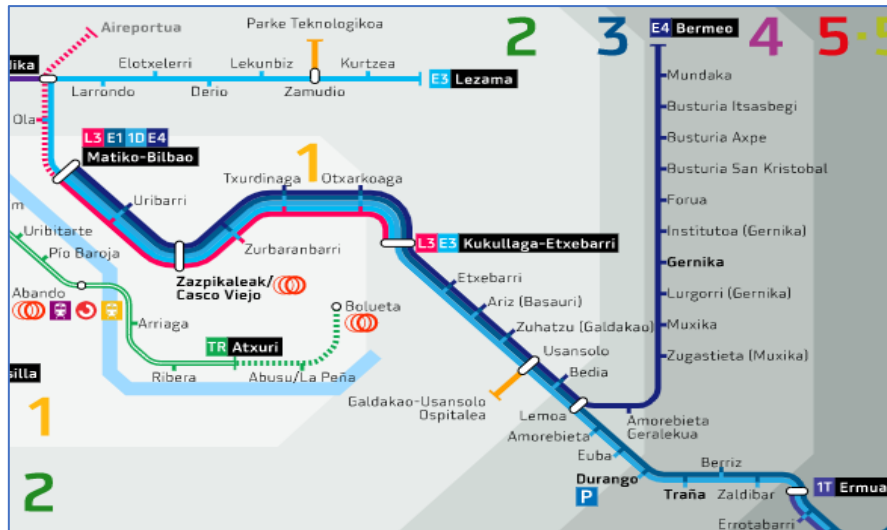


Figura 2. En el plano de rutas, se pueden apreciar los tramos donde las líneas se solapan. Un convoy que parte de Gernika a las 6:58 en la línea E4, comparte en Kukullaga Etxebarri parada de las líneas E3 y L3 a las 7:36.

- **Hora Paso:** Marca la hora en la que el convoy para en una estación determinada con algunas excepciones:
 - '>': El convoy solamente para en horarios concretos especificados en un pie de tabla.
 - '<': El convoy no realiza parada en la estación.

3.2.2. Arquitectura

Crearemos una visión desnormalizada del siguiente modelo lógico obtenido a partir de los conceptos extraídos del descubrimiento de variables.

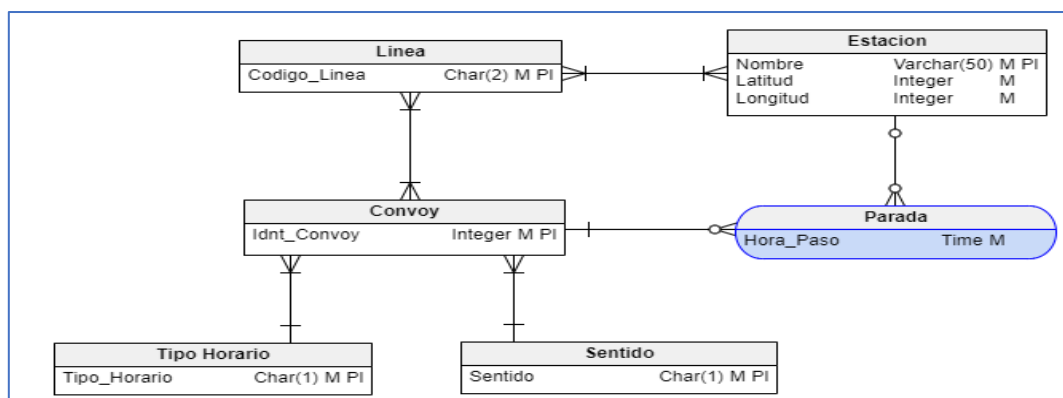


Figura 3. Modelo lógico de datos descubiertos.

Es decir, atendiendo al modelo planteado en la Figura 2, desnormalizaremos la asociación de hechos **Parada** atendiendo a sus dimensiones **Estacion**, **Convoy**, **Línea**, **Tipo Horario** y **Sentido**.

Así pues, los registros de información contienen los siguientes datos desnormalizados:

- **Convoy** (Convoy) (PK)
- **Nombre** (Estacion) (PK)
- **Codigo_Linea** (Linea)
- **Latitud** (Estacion)
- **Longitud** (Estacion)
- **Tipo_Horario** (Tipo Horario)
- **Hora_Paso** (Parada)

Finalmente, atendiendo a las relaciones existentes entre las entidades (dado que un convoy participa en varias líneas) se agrega la información para considerar los registros finales del conjunto de datos:

- **Lineas**: Códigos de las posibles líneas que sirve el convoy con su parada en la estación.
- **Sentido**: Sentido de la marcha del convoy dentro de la ruta de las líneas que sirve.
- **Horario**: Señala el tipo de horario al que se ajusta la hora de paso del convoy.
- **Estacion**: Nombre de la estación donde hace parada el convoy.
 - **Latitud**: Coordenada de situación de la estación en un mapa.
 - **Longitud**: Coordenada de situación de la estación en un mapa.
- **Hora**: Hora en que efectúa parada el convoy en la estación.

	Lineas	Sentido	Horario	Estacion	Latitud	Longitud	Hora
1	L3	I	L	Kukullaga Etxebarri	43.2511609	-2.8899548	05:36:00
2	L3	I	L	Otxarkoaga Bilbao	43.2589089	-2.8980102	05:38:00
3	L3	I	L	Txurdinaga Bilbao	43.2579646	-2.9070018	05:39:00
4	L3	I	L	Zurbaranbarri Bilbao	43.2642909	-2.9129307	05:41:00
5	L3	I	L	Zazpikaleak Bilbao	43.2599549	-2.9224017	05:43:00
6	L3	I	L	Uribari Bilbao	43.2668168	-2.9197002	05:45:00
7	L3	I	L	Matiko Bilbao	43.2690035	-2.9254383	05:46:00

Tabla 1. Ejemplo de los registros del conjunto de datos correspondiente a un convoy sirviendo la línea L3 en sentido ida, un día Laboral en las distintas estaciones en las que efectúa parada a la hora determinada.

3.3. Generación del dataset

Para digitalizar los datos se va a utilizar la herramienta **Microsoft Excel**. Aprovecharé las ventajas que ofrecen las hojas de cálculo tanto para copiar celdas y obtener los datos que se duplican debido a la desnormalización, como para establecer fórmulas que sumen las frecuencias en los distintos rangos horarios para no transcribir todo el **PDF** y conseguir desplegar automáticamente las bandas únicamente con la información de frecuencias.

Para extraer la información de **latitud** y **longitud** de las estaciones, se utiliza **Google Maps**.

Como en el **Excel** se realiza la transcripción de todos los archivos **PDF**, no se tiene en cuenta el hecho de que un convoy puede servir distintas líneas a su paso por estaciones que comparten ruta. Para realizar esta agregación, se exportan los datos del **Excel** a una base de datos creada en **Microsoft SQL Server 2017**. Haciendo servir el lenguaje **SQL**, generamos una consulta que realice la agregación de convoyes y devuelva los registros finales que conforman el conjunto de datos definitivo exportando el resultado a **CSV**.

4. Creación de la visualización

4.1. Título de la visualización

Horarios Euskotren Bizkaia: Publicación en [Tableau Public](#)

https://public.tableau.com/views/PRA2/HorariosEuskotrenBizkaia?:language=es-ES&:display_count=n&:origin=viz_share_link

4.2. El conjunto de datos

Para la creación de la visualización, se emplea el conjunto de datos *Horarios Euskotren Bizkaia* contenido en formato [CSV](#).

4.2.1. Exploración

8.286 registros.

7 atributos:

- **Convoy** (continuo): Número secuencial y único que describe cada unidad que realiza un determinado recorrido por estaciones determinadas en un horario estipulado.

Rango	
Mínimo	1
Máximo	521

- **Lineas** (categórico): Códigos de las posibles líneas que sirve el convoy con su parada en la estación.

Dominio	Ocurrencias
E4	1.767
E3	1.704
E1	1.490
E3 L3	1.106
L3	776
E1 L3	698
E4 L3	550
E4 E3 L3	103
E1 E3 L3	77
E4 E3	15

- **Sentido** (categórico): Sentido de la marcha del convoy dentro de la ruta de las líneas que sirve.

Dominio	Descripción	Ocurrencias
I	Ida	4.150
V	Vuelta	4.136

- **Horario** (categórico): Señala el tipo de horario al que se ajusta la hora de paso del convoy.

Dominio	Descripción	Ocurrencias
L	Laboral	5.259
F	Festivo	3.027

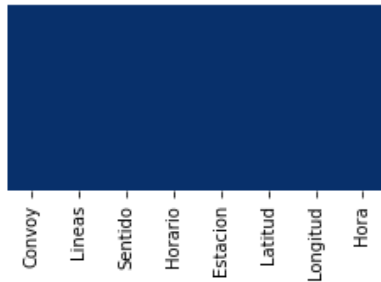
- **Estacion** (categórico): Nombre de la estación donde hace parada el convoy.
 - **Latitud** (Float): Coordenada de situación de la estación en un mapa.
 - **Longitud** (Float): Coordenada de situación de la estación en un mapa.

Dominio	Latitud, Longitud	Ocurrencias
Matiko Bilbao	43.269003, -2.925438	483
Uribarri Bilbao	43.266817, -2.919700	483
Zazpikaleak Bilbao	43.259955, -2.922402	483
Txurdinaga Bilbao	43.257965, -2.907002	482
Zurbaranbarri Bilbao	43.264291, -2.912931	482
Otxarkoaga Bilbao	43.258909, -2.898010	482
Kukullaga Etxebarri	43.251161, -2.889955	482
Amorebieta	43.216584, -2.737983	205
Bedia	43.206434, -2.800638	204
Etxebarri	43.244727, -2.892260	204
Ariz Basauri	43.241532, -2.881722	204
Usansolo Galdakao	43.218087, -2.824452	204
Lemoa	43.206607, -2.778291	204
Zuhatsu Galdakao	43.228743, -2.855595	204
Sondika	43.298951, -2.926364	189
Ola Sondika	43.292220, -2.926419	189
Lezama	43.273424, -2.835660	187
Elotxelerrri Loiu	43.294339, -2.905634	187
Lekunbiz Zamudio	43.287134, -2.874575	187
Zamudio	43.283989, -2.864961	187
Larrondo Loiu	43.294301, -2.916263	187
Kurtzea Lezama	43.276479, -2.846537	187
Derio	43.290625, -2.886510	187
Durango	43.169968, -2.638537	116
Traña Abadiño	43.165084, -2.614469	112
Euba Amorebieta	43.196710, -2.699737	107
Gernika	43.314466, -2.675672	105
Zaldibar	43.171635, -2.552394	99
Berriz	43.168511, -2.576055	99
Lurgorri Gernika	43.306266, -2.680399	98
Zugastietia Muxika	43.252026, -2.692617	97
Muxika	43.287247, -2.689866	97
Axpe Busturia	43.380391, -2.692939	96
Errotabarri Ermua	43.183351, -2.503351	96
S. Kristobal Busturia	43.367670, -2.690204	96
Ermua	43.185551, -2.513451	96
Mundaka	43.408912, -2.700770	96
Forua	43.331945, -2.673586	96
Bermeo	43.417798, -2.723196	96
Institutoa Gernika	43.319788, -2.676718	96
Itsasbegi Busturia	43.391328, -2.693875	96

- **Hora** (Time): Hora en que efectúa parada el convoy en la estación.

Rango	Ocurrencias
Primera	05:36:00
Última	23:49:00

El dataset no presenta valores nulos:



No se observan valores extremos:

- La línea **E4** (**1.767**) es la que más paradas tiene en la provincia de Bizkaia en comparación con las otras líneas. La línea **E3** tiene menos paradas pero mayor servicio que la anterior. Por otro lado, los trenes que cubren servicio de las paradas en las líneas **E4** y **E3** (**15**) son pocas puesto que son líneas que se sitúan una a continuación de otras con intención de abarcar el máximo recorrido sobre paradas de la provincia. Estos convoyes corresponden a unos servicios especiales que hacen un recorrido “express” a primera hora para unir los puntos más alejados de Bizkaia.
- El número de viajes en ambos sentidos es similar.
- Evidentemente el número de viajes en jornadas **laborales** es mayor que en jornadas **festivas**.
- Las estaciones más transitadas son las que corresponden a la línea de metro **L3** que transcurre desde **Kukullaga Etxebarrri** a **Matiko Bilbao**. Por otro lado, las estaciones menos transitadas son las que se encuentran cerca de los límites de la provincia.
- El rango de horas es lógico, prestando servicio desde las **5:36** hasta las **23:49**.

4.3. Boceto

La motivación del proyecto era conseguir una visualización directa y potente capaz de sustituir las típicas tablas de horarios de paradas complicadas de interpretar. El siguiente es el boceto con la visualización original para poder interpretar las salidas desde una estación concreta y sus llegadas a las estaciones donde puedas llegar

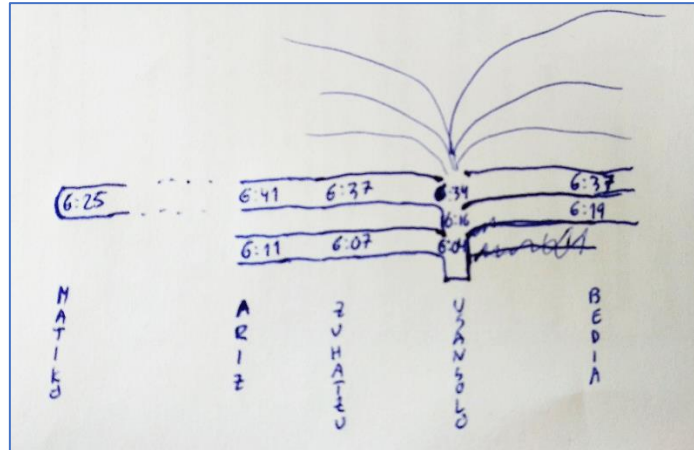


Figura 4. Boceto de una visualización para representar las salidas y llegadas desde una estación concreta.

Se trata de una visualización estática y dependiente de la estación donde se representa en forma de palmera las diferentes salidas de trenes desde la estación en el sentido de circulación del tren. La idea es seguir el tronco para ver las distintas salidas desde la estación y la rama hasta la estación de destino.

Para responder a la pregunta planteada, puedes buscar en la columna de la estación de destino, la rama con la hora deseada de llegada y seguir hacia el tronco para conocer la hora de salida del convoy en tu estación de origen donde estás consultando la información.

El primer obstáculo durante la preparación del concepto, fue que la visualización debía servir para distintas estaciones por lo que ésta tendría que ser interactiva con capacidad de filtrar la información para una visualización concreta según necesidad. También poder dar información de la red a un usuario no habitual de la provincia. De ahí surge el boceto que dará forma a la visualización implementada.

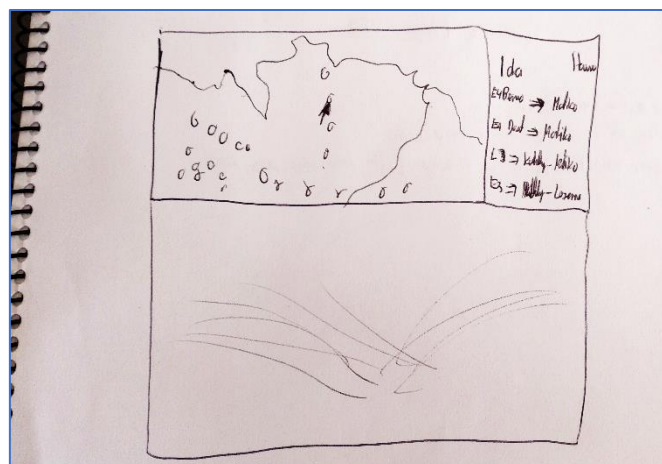


Figura 5. Boceto de la visualización interactiva

4.4. Implementación

Para implementar la visualización, vamos a hacer uso de la herramienta [Tableau](#). Uno de los motivos de elección de esta herramienta, es académico puesto que fruto de un despiste, no reparé en los recursos disponibles en el aula sobre esta herramienta. Consecuencia de ello fue no trabajar con ella la PEC2 y entiendo que es un recurso muy valorado en el mundo de la visualización y merece la pena realizar un esfuerzo introductorio en la misma.

Como aspectos técnicos, destacar que sus capacidades de creación de diversos tipos de visualizaciones, la integración de estos gráficos en cuadros de mando interactivos y la recopilación final en historias, van a ser muy ventajosas para el desarrollo y uso de la visualización.

4.4.1. Importar los datos

Creamos un nuevo libro importando los datos del archivo [Horarios Euskotren Bizkaia.csv](#) generado.

Modificamos la configuración regional a [inglés \(Estados Unidos\)](#) para que [Tableau](#) interprete el carácter `'.'` de los campos `Latitud` y `Longitud` como separador decimal. De esta manera, ya me ofrece la posibilidad de darle [función geográfica](#) y poder usarlo así posteriormente.

La interpretación del campo `Hora` como fecha parece ser adecuada en un primer momento, pero quizá pueda ser contraproducente el hecho de que le anteponga un día concreto, [30/12/1989](#) por defecto.

4.4.2. Hoja Mapa Estaciones

Para elaborar el mapa de estaciones, se establecen las medidas geográficas `Latitud` y `Longitud` como columnas y filas y se establece la visualización [Mapa de Símbolos](#). Como detalles a visualizar en el mapa se incluye la dimensión `Estacion`.

Se configuran las propiedades [Capas de mapas](#) para establecer un diseño más limpio que haga destacar los puntos de las estaciones.

Se establece como [etiqueta](#) la dimensión `Estacion` para que aparezcan los nombres sin permitir que se solapen. La desventaja es que, en una visión general, no aparecen los nombres de las estaciones más próximas, las cuales son bastante importantes por ser la línea del metro de la capital. Para tratar de solucionarlo:

- Se permite hacer *zoom* al usuario para que pueda observar un terreno concreto y si existe separación suficiente aparecen.
- Se generan [descripciones emergentes](#) donde se muestra el nombre de la estación e información adicional.

En las [descripciones emergentes](#), se ha pretendido mostrar al usuario, además de la estación concreta, información general sobre las horas de paso de trenes por dicha estación agrupadas por las dimensiones `Horario` y `Sentido` para una mejor agregación de datos.

4.4.2.1. Hoja HorasEstacionHorarioSentido

Para conseguir mostrar el contenido en las descripciones emergentes, se ha creado una visualización de tipo **Tabla** utilizando la dimensión **Estacion** y tres nuevas dimensiones calculadas a partir de las dimensiones **Horario**, **Sentido** y **Hora**. Se han salvado los siguientes inconvenientes para poder elaborar la visualización final:

- Como la dimensión **Hora** es interpretada por **Tableau** como una fecha, nos mostraba la información del día delante de la hora de paso. Se ha solucionado creando una nueva dimensión **SoloHora** trabajada sobre la información tipo fecha de dicha dimensión, transformándola en formato cadena y añadiendo ceros a la izquierda en minutos y segundos para conseguir un formato 24h.
- Se han creado a su vez dos dimensiones calculadas que pretenden ofrecer un mensaje más descriptivo de las codificaciones para **Sentido** y **Horario** contenidas en el conjunto de datos. Las dimensiones **SentidoDesc** y **HorarioDesc**, únicamente traducen la codificación de la variable a una descripción entendible para el usuario.
- La gran cantidad de horas de paso de trenes por cada estación hacía que cada hora se representara en una fila de la tabla y ésta no cupiera en el desplegable de la visualización del mapa. Pude encontrar una solución al problema gracias al usuario Sam L. a la pregunta *“Concatenate values in one row with comma as delimiter in TABLEAU worksheet”*² de Dhruvil P. Creando un ranking ordenado de todas las combinaciones de horas, separadas por comas, ordenado y filtrado por la primera de ellas que contiene la agrupación de todas las entradas combinadas. Esta información se incorpora como etiqueta a la fila correspondiente.

4.4.3. Hoja Selección de estación

Se trata de una visualización sencilla de tipo **Tabla** que servirá de filtro de selección de los parámetros que condicionan la decisión de un usuario. Para conseguir obtener la respuesta a la pregunta ¿A qué hora pasa el tren que me transporte a mi destino a una hora adecuada? Es necesario conocer:

- La estación de partida.
- El horario del día en el que necesito viajar.
- El sentido del viaje.

Se implementa mediante una **tabla** cuyas filas corresponden al listado de estaciones del conjunto de datos y agrupado en columnas por las distintas salidas que puedan ocurrir según el horario. Para conseguir el símbolo de enlace, se utiliza como detalle de la tabla, una operación de cuenta de distintos de las líneas que pasan por la estación por **Horario** y **Sentido**.

Se aplican ciertos formatos a la tabla con el color corporativo de la compañía y se omite la representación del cálculo del detalle sustituyéndolo por un icono en forma de flecha existente en **Tableau** que pueda representar como metáfora la acción de *ir, marchar, dirigirse a*.

² Sam L. respuesta a Dhruvil P. *Concatenate values in one row with comma as delimiter in TABLEAU worksheet*. Disponible en línea en: <https://stackoverflow.com/questions/58189803/concatenate-values-in-one-row-with-comma-as-delimiter-in-tableau-worksheet>. Stack Overflow 01 oct. 2019. Visitado el 19 may. 2021.

4.4.4. Hoja de visualización de rutas y horarios

Es la visualización principal donde se muestra la información que pretende resolver la consulta principal: *Qué tren debo coger en mi estación para llegar al destino a una hora concreta.*

En la visualización se observan las rutas seguidas por todos los convoyes presentes en el conjunto de datos. En el eje de abscisas se sitúan las estaciones por las que transcurre la ruta y en el eje de ordenadas el tramo de horas en el día donde el servicio está activo. Así en cada intersección, se sitúa el dato de la hora de paso del tren en cada estación.

Se trata de una visualización de tipo **Gantt** que refleja la evolución de cada convoy en el tiempo a lo largo de las estaciones como hitos. Se han salvado los siguientes problemas:

- Ordenar adecuadamente las estaciones atendiendo a la propia infraestructura ferroviaria. Inicialmente las estaciones aparecen ordenadas alfabéticamente por lo que la ruta entre los puntos presenta saltos.
 - Las líneas **E1** y **E4** se bifurcan en la estación de **Amorebieta**. Es por esto que, lamentablemente, va a existir un salto en los trenes de la línea **E1**. Desde **Amorebieta** a **Euba** **Amorebieta**.
- La medida **Hora** en el eje de ordenadas ha de ser configurada como **continua** y **mostrar la fecha exacta**. Existen detalles a destacar en la configuración del eje:
 - Se estipula el intervalo como **Automático** para que la visualización pueda hacer **zoom** con el filtro por hora.
 - Se establecen las marcas de graduación principal en modo **Automáticamente** pero se induce a que las marcas secundarias tengan un intervalo de **15 minutos**.
- La dimensión a representar es el **Convoy** lo que dará continuidad al diagrama.
 - Se establece un color corporativo a las líneas del diagrama.
 - Se establece como marca para su representación el tipo **Línea**.
- Se crean tres **filtros** para poder interactuar con la visualización:
 - **Hora**: Para que el usuario pueda seleccionar y se realice un **zoom** automático en el intervalo de horas de su interés. Se activa el panel para que el usuario pueda interactuar con la visualización.
 - **Sentido**: Sirve para poder filtrar la visualización desde la hoja **SeleccionEstacion**, en la integración de la visualización dentro del **cuadro de mando**, para mostrar únicamente las rutas que discurren en un sentido concreto.
 - **Horario**: Utilizado para filtrar la visualización desde la hoja **SeleccionEstacion**, en la integración de la visualización dentro del **cuadro de mando**, para mostrar únicamente las rutas que discurren en un tipo de día concreto.
- Es necesario que, en las opciones de **Análisis**, no agregue las medidas.
- Se reduce la información de la **descripción emergente** para mostrar:
 - El nombre de la **Estacion**, debido a que lejos del eje de abscisas es fácil perder la ubicación.
 - La **Hora** de paso, debido a que lejos del eje de ordenadas es fácil perder la ubicación y en este eje no se pueden representar todas las marcas horarias.

- Se configura la utilidad de **Etiquetas** para que muestre las horas de paso de los convoyes por la **Estación** resaltada. Como hay escenarios donde aparecen muchas rutas superpuestas, se opta por no mostrar la etiqueta en caso de solapamiento.
- Establecer la **Vista completa** de la visualización para que su presentación escale según el tamaño asignado a la hoja y las rutas tengan mayor sensación de extensión.
- Se ha formateado la rejilla del gráfico para ofrecer líneas guía de los distintos valores de ambos ejes. Me he decidido a incluirlas puesto que es bastante la información que genera el gráfico y era sencillo perderse tratando de localizar la estación y la hora sin una ayuda visual. Para no entorpecer la visión del detalle principal que es la ruta del convoy, se asigna un color tenue y con contraste sobre el corporativo.

4.4.5. Dashboard Horarios Euskotren Bizkaia

Este cuadro de mando, compone la visualización final en la cual se aglutinan las hojas descritas en los apartados anteriores y se añade la interactividad entre ellas para ser funcional al usuario.

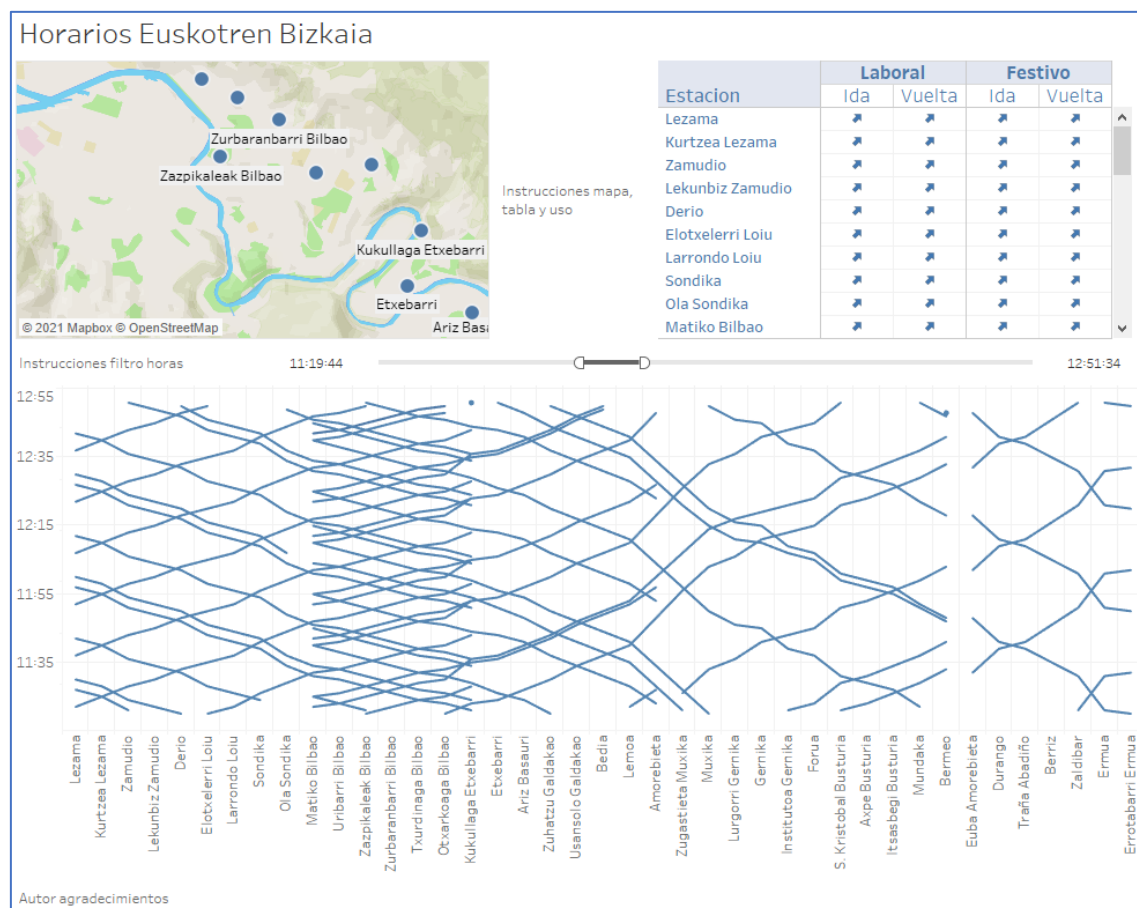


Figura 6. Estructura del cuadro de mandos de la visualización.

La visualización la componen (de arriba abajo y de izquierda a derecha):

- Título de la visualización.
- Hoja **Mapa Estaciones**: En la *overview* será posible situar geográficamente las estaciones y la configuración de la línea ferroviaria a través de los puntos que representan la ubicación de las estaciones. Será posible:
 - Realizar **zoom** para localizar las estaciones. Si el espacio es suficiente, se mostrarán las etiquetas con el nombre de la estación.

- Realizar una acción para *seleccionar* una o varias estaciones de interés para el usuario.
 - Realizar una acción para ofrecer un *detalle* de las horas en las que pasan los trenes en la estación sobre la que se posiciona el ratón
- Etiqueta de instrucciones simples para el uso de la visualización.
- Hoja *Selección Estación*: Permite elegir una o varias estaciones, como en la hoja *Mapa de Estaciones*, y además establecer un *filtro* adicional por el *Horario* (laboral o festivo) y el *Sentido* (ida o vuelta).
- Etiqueta instrucciones filtro de rango de horas.
 - Permite realizar *zoom* sobre la visualización de la hoja de rutas al establecer un rango determinado en el eje x que hace más visibles los recorridos de los convoyes entre estaciones.
- Filtro rango de horas: Permite establecer un rango de horas de interés para el usuario que además permita realizar la acción de *zoom* en la visualización de rutas situada debajo.
- Hoja *Horas Ruta*: Realiza la visualización condicionada a las acciones realizadas sobre las hojas anteriores y el filtro de rango de horas.
 - Su *overview*, consiste en todas las rutas que discurren en todos los horarios y sentidos de la marcha de los convoyes durante las horas que dura el servicio.
 - Permite la *selección* y *detalle* de un convoy determinado pinchando sobre éste lo que lo hace destacar sobre el resto.
 - Permite la *selección* de estaciones en el eje de abscisas lo que permite ofrecer detalles informativos del paso del tren por la estación.
- Autoría y agradecimientos.

Técnicamente se establecen las siguientes interacciones entre las hojas que componen el cuadro de mando:

- *FiltroSelecciónHorarioSentido*: Permite, al hacer clic sobre la configuración (*Horario*, *Sentido*) de la hoja *Selección Estación*, filtrar los convoyes que coinciden con esta disposición en la hoja *Horas Ruta*. Se permite realizar más de una elección y la acción al descartar las selecciones se establece a volver a mostrar todos los valores.
- *ResaltarEstación*: Permite, al hacer clic sobre la configuración (*Estación*, *Horario*, *Sentido*) de la hoja *Selección Estación*, resaltar la columna correspondiente a las estaciones seleccionadas en la hoja *Horas Ruta*.

Para la realización del cuadro de mandos, se han corregido o adaptado los siguientes aspectos:

- Eliminar los filtros *Horario* y *Sentido* establecidos en la hoja *Horas Ruta*, sustituyéndolos por la acción de *filtro* incluida en el cuadro de mandos *FiltroSelecciónHorarioSentido*.
- Ordenación de las estaciones en la hoja *Selección Estación*. Puesto que en su planteamiento inicial las estaciones quedaban ordenadas alfabéticamente y cobra mayor sentido que estén ordenadas igual que la disposición de la hoja *Horas Ruta*. De esta manera, también, es mucho más sencillo para el usuario poder seleccionar dos estaciones puesto que, al establecer el orden cierta dependencia conceptual de su ubicación, es más intuitiva la elección de las estaciones de inicio y fin del trayecto.

También se ha ordenado la disposición de las columnas de **Horario**, las cuales aparecían en orden alfabético dando prioridad al horario **Festivo** que es menos relevante que el **Laboral**. Como perceptualmente los usuarios leen de izquierda a derecha, de esta manera se da preferencia al horario más habitual.

5. Implantación y uso de la visualización

La implantación de la visualización en conjunto, se guarda localmente en el fichero `PRA2.twb` adjunto en la práctica, utilizando el conjunto de datos `Horarios Euskotren Bizkaia.csv` también adjunto.

A su vez se publica guardando en los servidores de [Tableau Public](#) de manera accesible en la siguiente dirección web:

<https://public.tableau.com/views/PRA2/HorariosEuskotrenBizkaia?:language=es-ES&:display count=n&:origin=viz> share link

Previamente se ha solicitado exportar los datos, los cuales se adjuntan en la práctica con el fichero [Horarios Euskotren Bizkaia.hyper](#).

5.1. Respuestas de la visualización a la pregunta planteada

Supongamos ser un usuario del transporte público que necesita viajar de la estación **ZazpiKaleak** (Casco Viejo) de Bilbao a **Usansolo Galdakao**. Debemos estar en nuestro destino antes de las **14:15** y tenemos disponibilidad para comenzar el viaje a partir de las **13:00 horas** de un día **laborable**.

Una primera información nos la proporciona situarnos encima de la estación de origen para observar los horarios de paso del tren por esta estación

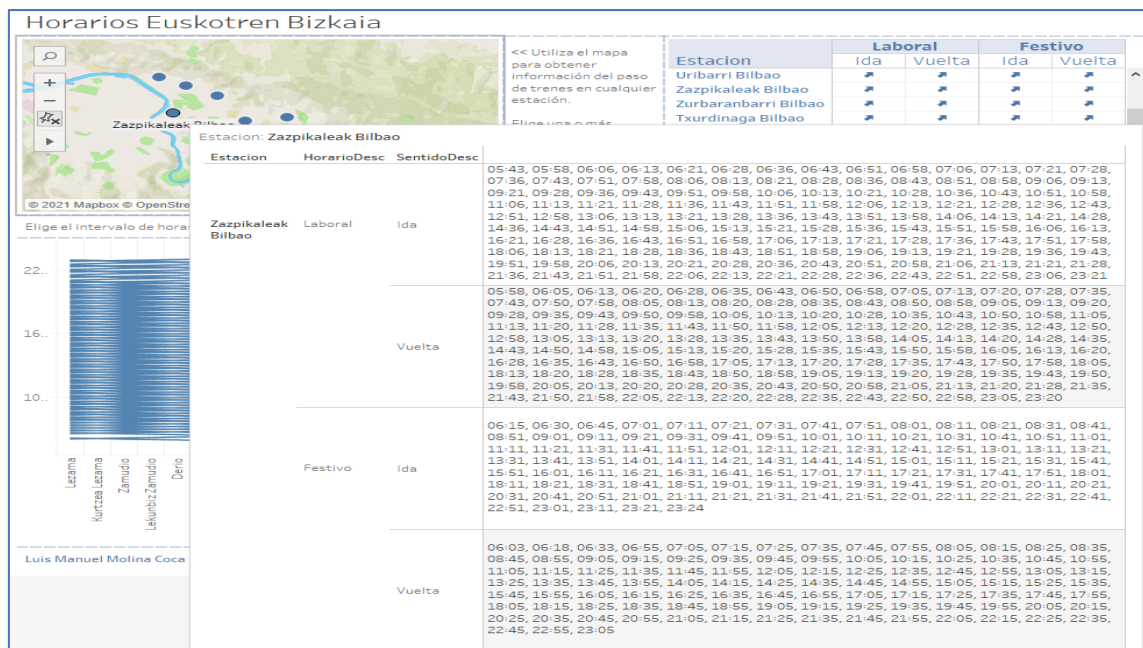


Figura7. Utilización de la herramienta de visualización de salidas en una estación determinada ubicada en el mapa.

Situándonos en el mapa sobre la parada de origen, mediante el listado de partidas de un día laborable en sentido de vuelta, nos informamos de las salidas en horario laboral del sentido

vuelta a partir de las **13 horas** son: **13:05**, **13:13**, **13:20**, **13:28**, **13:35**, **13:43**, **13:50**, **13:58**, **14:05** y **14:13**.

Esta primera aproximación, no obstante, no nos indica la duración del trayecto hasta el destino y al incluir distintas líneas, no nos garantiza que todas estas salidas tengan destino en la estación de **Usansolo**.

Para encontrar el convoy que nos transporta a la hora adecuada, seleccionamos la estación de origen **Zazpikaleak Bilbao** y manteniendo la tecla **Ctrl**, la estación destino **Usansolo Galdakao** para el **Horario** y **Sentido** adecuados. Éstas se resaltan en el **overview** de la visualización de rutas.

Establecemos el rango horario con el control deslizante desde las **13 horas** hasta las **14:15**, hora en la que tenemos nuestra cita concertada en el destino.

Se realiza un **zoom** de manera automática en la visualización de rutas y podemos observar los convoyes que la realizan. El tren que nos garantiza estar antes de la hora llegaría a las **14:04** y siguiendo su línea de ruta, podemos ver que debemos coger el tren en la estación de origen a las **13:43**.

También distinguimos dos posibles opciones más dentro del intervalo en el que podemos viajar: **13:28 – 13:46** y **13:13 – 13:34**. Siguiendo la línea de ruta, nos encontramos con información del punto de partida del convoy y su destino para contrastar en los rótulos del andén y de la locomotora a su llegada a la estación.

Por último, también se visualizan salidas desde el origen que o bien no pasan por la estación de destino (**13:06**, **13:20**, **13:35**, **13:50**, **14:05**) o bien no nos conviene porque llega más tarde de la hora (**13:58**)

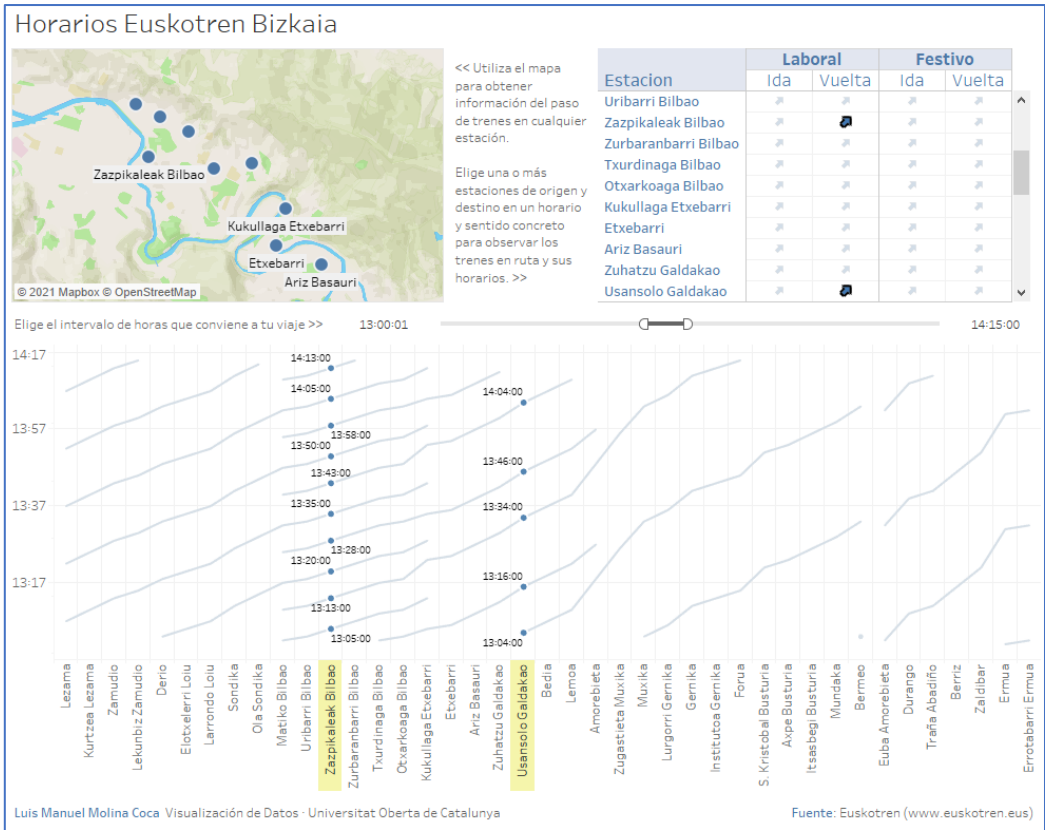


Figura 8. Utilización de la visualización interactiva para concretar el horario de salida y llegada conveniente.

6. Visualización derivada: Carteles de estación

Utilizando el conocimiento generado por la visualización descrita, es posible crear una visualización innovadora para la cartelera de las estaciones que informa de los horarios de paso.



Figura 9. Cartelería de la estación de Usansolo Galdakao

Como se comentaba en el apartado 2 de descripción del documento, la tarea para el usuario de interpretar el horario del tren que más le conviene para llegar a su destino puede ser difícil presentada con la visión tabular que se observa en la Figura 9.

La siguiente visualización sirve como alternativa a la cartelera presentada en cada estación para dar de nuevo respuesta a la pregunta *¿A qué hora debería coger el tren en mi estación de origen para llegar a mi destino a una hora adecuada?*

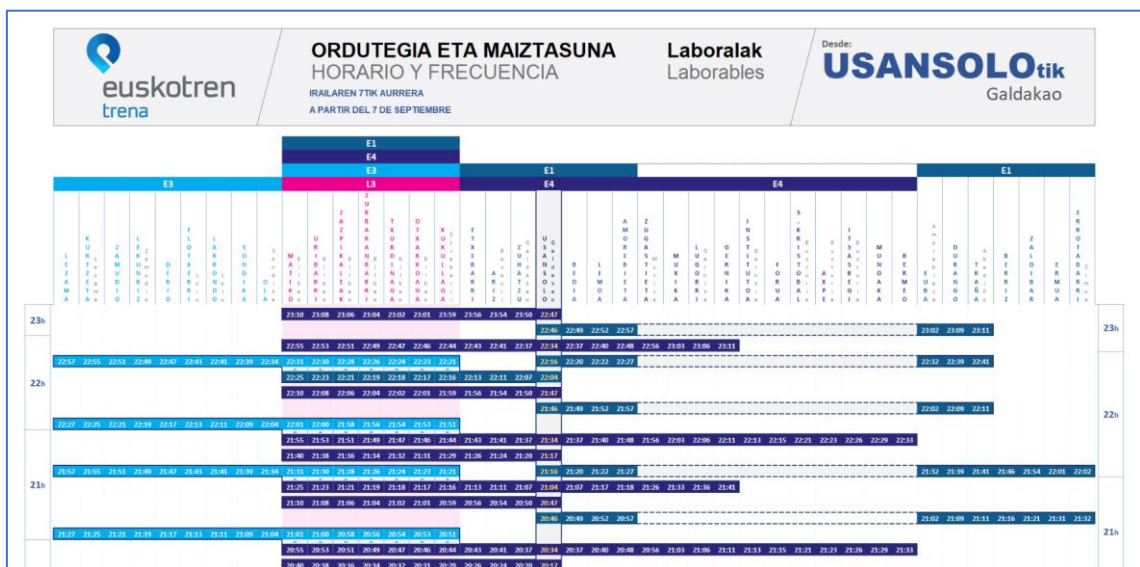


Figura 10. Extracto de la visualización del cartel de horarios en días laborables de la estación de Usansolo Galdakao.

La visualización completa del cartel representado en la Figura 10, se encuentra adjunta en el fichero *Parada Usansolo* en formato PNG.

6.1. Implementación

Se usa la herramienta *Microsoft Excel* para organizar y presentar estéticamente la información obtenida utilizando la visualización *Horarios Euskotren Bizkaia* seleccionando la estación *Usansolo Galdakao* y recorriendo el intervalo de horas sobre un día laborable en ambos sentidos.

Aspectos a destacar de la visualización:

- Aglutina la información de los carteles de cada línea, que se ven en la Figura 9, en un único cartel y en ambos sentidos de circulación.
- Da respuesta directa a la hora de paso puesto que muestra la información completa sin tener que calcular frecuencias.
- Cada cartel es dependiente de la estación. Los carteles de la Figura 9 son comunes para cualquier estación.
- Estética corporativa respetando la paleta de la organización y de representación de cada línea.
- Sería posible incluir la información de los horarios laboral y festivo en el mismo cartel, pero la condensación de información y la mezcla haría difícil la interpretación. Por lo tanto, se opta por realizar un cartel distinto por Horario.
- La cabecera presenta el logo, título del cartel, información adicional de validez, el horario que representa y la parada desde la que tiene sentido su interpretación.
- Se describen las paradas, resaltando la representativa del cartel, en orden lógico de ubicación dentro de la red. Cada parada se agrupa por un bloque en la parte superior con el código de la línea que pasa por ellas.
 - La bifurcación de las líneas E1 y E4 se representa con un lapso punteado sobre la línea E4 que represente un salto de paradas.
 - Cada línea tiene su color representativo y las paradas donde tiene exclusividad también se representan de este color.
 - En el nombre de la estación, se resalta el nombre y como secundario el municipio.
- Para localizar en el árbol el tramo horario de llegada, se dibujan unas bandas guía a izquierda y derecha que agrupen, aproximadamente, la mayor cantidad de horas de parada en las estaciones. Esto pretende acelerar la búsqueda por una hora concreta.
- Se dibujan los recorridos como una línea continua clásica y habitual en las representaciones de transporte público.
 - Cada recorrido se pinta del color oficial de la línea que lo sirve. De esta manera, el usuario sabe la línea del convoy a coger cuando varias líneas circulan por la estación.
 - Se representan los posibles transbordos como un recorrido adicional, servido por una línea distinta a tenor del color, con un icono de flecha.
- Al final de la visualización de recorridos, se replica de manera invertida la información de líneas y paradas.
 - Tras realizar la visualización, me di cuenta que quizá fuera mejor seguir un orden horario mañana-tarde-noche de arriba abajo que no al contrario que puede resultar menos intuitivo.

6.2. Uso de la visualización

De modo similar a la visualización principal, partiendo desde la estación de Usansolo si quisiera estar en Zamudio antes de las 22h de un día laboral, utilizaría el cartel de la siguiente manera:

Localizar la estación de destino **Zamudio** → bajar con la mirada hasta encontrar una hora conveniente **21:51** (puedo guiarme por la banda lateral de rangos horarios) → por el color observo que es un tren de la línea **E3** → sigo su recorrido hasta una de las estaciones donde haría transbordo. Por ejemplo, en **Matiko Bilbao** saldría a las **21:31** llegando desde un tren de la línea **E4** a las **21:25** → sigo el recorrido hasta la estación de origen **Usansolo Galdakao** y me informo que tengo que coger el tren que sale a las **21:04**.

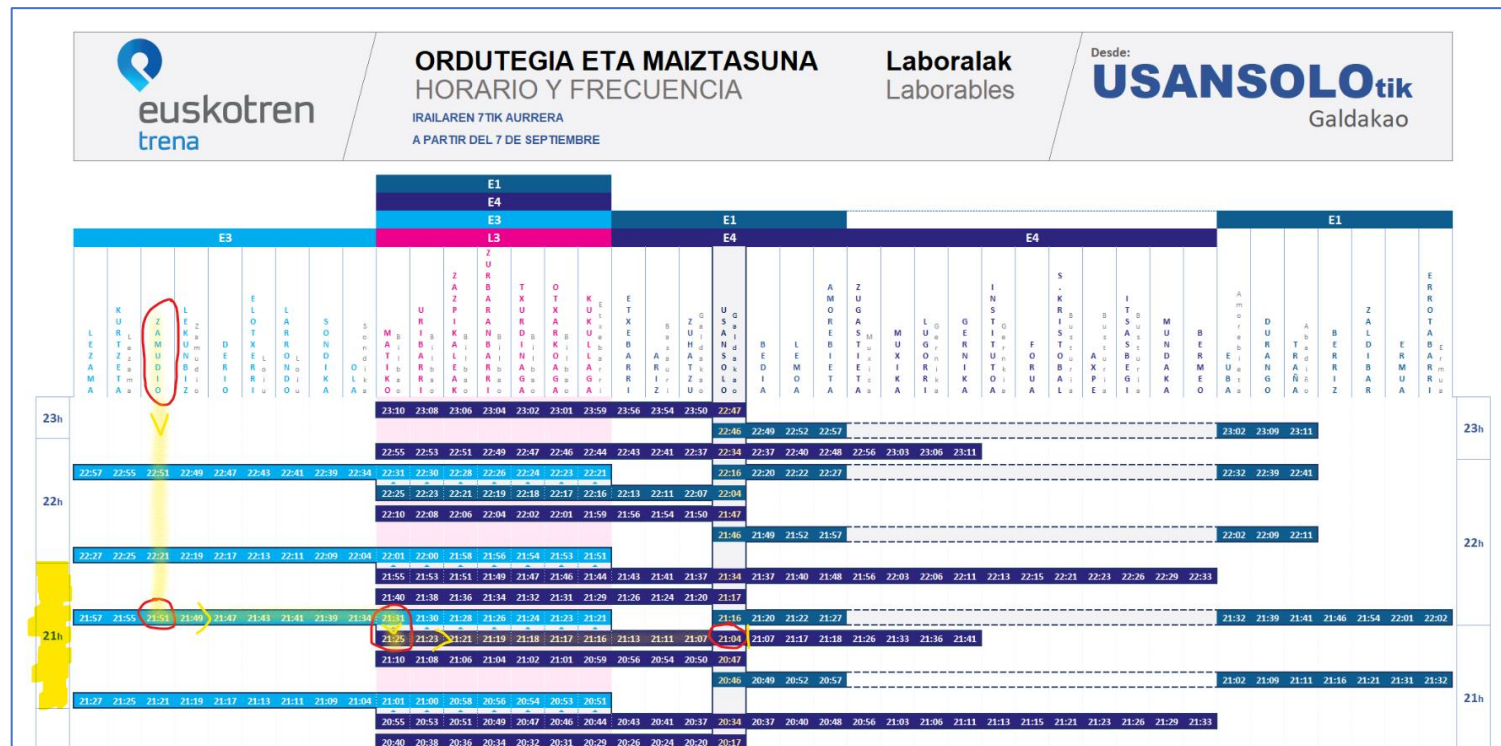


Figura 11. Utilización de la visualización Cartel de Estación de Usansolo

7. Conclusiones y carencias

Me gustaría expresar dos conclusiones a las que he llegado mediante la realización de la práctica:

- He quedado impresionado con el potencial que ofrece la visualización de datos para poder resolver problemas. Es decir, en un desarrollo tradicional, al que estoy acostumbrado, habría que recoger requisitos, obtener los datos, programar una interfaz de presentación, programar un negocio para resolver las funcionalidades requeridas y también una capa de acceso a los datos. Gracias a la evolución de la visualización de datos y a las herramientas actuales de generación, sin una línea de código y sustentado únicamente en un fichero de texto plano con datos, he podido ser capaz de realizar prácticamente una aplicación interactiva y funcional.

- También me ha sorprendido el potencial de la herramienta [Tableau](#). He cumplido el objetivo que me había marcado y que no pude cumplir en la PEC2 de tener un primer contacto con esta herramienta tan utilizada.

En más de una ocasión, me ha impactado la sencillez con la que, simplemente arrastrando los campos adecuados, conseguía reprogramarse para ofrecer una visualización nueva o una funcionalidad concreta. También el alto grado de personalización: como a la hora de permitir establecer un orden propio de las estaciones sin un criterio específico como el alfabético.

Otro punto a favor es la comunidad de usuarios que tiene la herramienta con la que he podido solventar algunas trabas en la programación de objetivos concretos sin tener experiencia previa con la suite de [Tableau](#).

- Con todo, también me queda algo de insatisfacción por no poder alcanzar aquello que tenía en la mente. Entiendo que puede ser algo habitual en el trabajo de un profesional en la visualización de datos. No obstante, con la limitación del tiempo que puedo dedicar y mi limitación por el desconocimiento inicial de la herramienta, creo que ha salido una visualización decente, pero con carencias que me gustaría también reconocer en este apartado.
 - No he podido aprovechar la información de las líneas del conjunto de datos. Sí es un dato relevante para el usuario puesto que se informa en el andén de la línea a la que pertenece el tren que llega, pero debido a que en la realidad las líneas se superponen según el tramo que sirven, este dato habría saturado de información la visualización lastrando el objetivo concreto de la visualización y añadiendo complejidad a su realización.
 - La interpretación de los valores `Ida` y `Vuelta` se dificulta al no conocer las paradas de origen y destino del convoy por ser dependiente de la línea que sirve. Debido al espacio para las etiquetas, no poder ofrecer una mayor descripción de la línea por no disponer del dato en el dataset y que se han agregado en estas líneas que transcurren por la misma estación: como consecuencia, se hace mucho más difícil seleccionar los filtros de estaciones por no tener herramientas para conocer si el sentido de ida o de vuelta es el que me dirige a mi destino concreto.
 - Pretendía que la estación de partida tuviera más importancia, es decir, fuera el centro de la visualización y sus ramas izquierda y derecha marcarán el sentido y las horas de llegada, pero no he sido capaz de encontrar la visualización adecuada en [Tableau](#).