



# PROBLEMA FINAL

Visualización de datos

# Cuadro de mando de BiciMAD

## Proyecto final de visualización de datos

### Objetivos

El objetivo de este proyecto final es realizar un cuadro de mando interactivo de análisis sobre los datos de BiciMAD.

### Dataset

Los datos provienen del [portal de datos abiertos de la EMT](#). Son los correspondientes al mes de junio de 2019. Existen estos ficheros de datos:

- stations.csv: contiene las propiedades de cada estación. Sus columnas son:
  - station\_id: el ID de la estación
  - name: su nombre
  - total\_bases: el número de bases (anclajes) que tiene para bicis
  - address: la dirección
  - number: el número de estación
  - latitude: latitud de su localización
  - longitude: longitud de su localización
- stations\_status.csv: correspondiente al estado guardado por hora de cada una de las estaciones. Sus columnas son:
  - date: la fecha en la que se guardó el estado
  - hour: la hora en la que se guardó el estado
  - station\_id: el ID de la estación
  - activate: indica si la estación está activa (1) o inactiva (0)
  - no\_available: indica si la estación está disponible (0) o no disponible (1)
  - reservations\_count: indica el número de reservas activas
  - free\_bases: indica el número de bases (anclajes) libres, es decir, aquellas en las que se puede dejar la bici para finalizar un viaje
  - dock\_bikes: indica el número de bicicletas ancladas, es decir, aquellas disponibles para iniciar un viaje
  - light: indica el grado de ocupación (desde 0 = baja hasta 3 = alta)
- trips.csv: con el detalle de viajes realizados. Las columnas son:
  - date\_unplug: la fecha en la que se inicia el viaje
  - hour\_unplug: la hora en la que se inicia el viaje
  - station\_id\_unplug: el ID de la estación desde la que se inicia el viaje (donde se coge la bici)
  - station\_id\_plug: el ID de la estación donde se finaliza el viaje (donde se deja la bici)
  - travel\_time: el tiempo de viaje en segundos
  - user\_type: el tipo de usuario (0: Desconocido, 1: Usuario anual (poseedor de un pase anual), 2: Usuario ocasional, 3: Trabajador de la empresa)

- user\_age\_range: el rango de edad del usuario (0: Desconocido, 1: entre 0 y 16 años, 2: entre 17 y 18 años, 3: entre 19 y 26 años, 4: entre 27 y 40 años, 5: entre 41 y 65 años, 6: 66 años o más)
- user\_zip\_code: el código postal del usuario

## Cuadro de mando

Como aspectos generales:

- El cuadro de mando será un documento de flexdashboard generado con R Markdown.
- Puedes generar los gráficos con ggplot2 o plotly, pero se valorará más positivamente plotly.
- Ídem con los mapas, puedes generarlos con ggplot2 o leaflet, pero se valorará más positivamente leaflet.
- Los indicadores superiores deberán ser value boxes. Puedes acompañarlos con un icono de fondo representativo o un color determinado para que den más información (p.e. si todos los cuadros son azules y pones uno naranja, este último indicará "alerta")
- En algunos casos, no detallo el tipo de gráfico a pintar. Elige el tipo (líneas, puntos, barras, histogramas, ...) que represente mejor la información que se pide representar.

El objetivo de este cuadro de mando es hacer un análisis de un mes de BiciMAD, en este caso, junio de 2019. El cuadro de mando deberá tener las siguientes páginas:

- Operativa. Esta página contendrá:
  - Indicadores superiores:
    - Número de estaciones
    - Número de bases: corresponde a la suma de las bases de todas las estaciones
    - % tiempo sin disponibilidad: corresponde al porcentaje de tiempo en el que las estaciones se encontraban no disponibles (es decir, al número de veces que veas en los datos de estado de estaciones que no está disponible, con respecto al total de observaciones de estado de estaciones).
  - Gráficos:
    - Mapa con todas las estaciones. El color del marcador con el que se representan deberá ser verde si esa estación ha estado sin disponibilidad < 24 horas, naranja si lo ha estado entre 24 y 72 horas, y roja si lo ha estado > 72 horas. El tooltip deberá mostrar el ID de estación, nombre, número total de bases y horas no disponibles.
    - Gráfico de barras que muestre las 15 estaciones con mayor número de horas no disponibles, y dicha cantidad de horas.
- Actividad. Esta página contendrá:
  - Indicadores superiores:
    - Número de viajes
    - % viajes con pase anual

- % viajes ocasionales
- % viajes de trabajadores
- Gráficos:
  - Número de viajes realizados por cada rango de edad. Nota: pon variables que identifiquen bien a estos grupos de edad (p.e. "19-26", "27-40", y no 1, 2, 3, ...)
  - Evolución del número de viajes por fecha, con granularidad diaria (es decir, agregado por día).
  - [OPCIONAL] Comparación del número de viajes medios por hora según el día de la semana (lunes, martes, ... domingo). La idea de este gráfico es ver si cambia el uso según el día de la semana, a qué horas hay más y menos viajes dependiendo de ello.
  - [OPCIONAL] Comparación del número de viajes medios por hora según el rango de edad. La idea de este gráfico es ver si cambia el uso según la edad, a qué horas usan más y menos el servicio.
  - [OPCIONAL] Distribución del número de minutos por viaje. Nota: piensa cómo tratar los outliers (la cola larga, es decir, unos pocos viajes muy largos), puedes eliminarlos o utilizar otro tipo de escala.
  - Gráfico de barras con las 15 estaciones desde la que más se inician los viajes (junto con el número de viajes iniciados desde ellas)
  - Gráfico de barras con las 15 estaciones en la que se finalizan más los viajes (junto con el número de viajes finalizados desde ellas)
- [OPCIONAL] Estado estaciones. Nota: las páginas anteriores no requerían utilizar shiny, ya que no dependen de entradas de usuario. Esta página es la única que sí lo necesita. Puedes investigar un poco más sobre esto en las referencias de apoyo. Esta página contendrá:
  - Menú lateral con inputs de usuario:
    - La estación.
  - Indicadores superiores:
    - Número de viajes iniciados en ella
    - Número de viajes finalizados en ella
  - Gráficos:
    - Evolución del % de bicis con respecto al total de bases (anclajes) por fecha y hora. Nos mostrará la capacidad a lo largo del tiempo para poder iniciar / finalizar viajes desde / en ella (100% es que todas las bases tienen bici, y 0% es que todas las bases están libres)
    - Evolución del número de viajes por fecha y hora.

[OPCIONAL] Puntos extra:

- Puedes añadir nuevas páginas, indicadores superiores o gráficas con análisis que resulten de tu interés. Se valorarán positivamente, especialmente aquellos que muestren algún patrón o sean un tipo de gráfico nuevo.

## Referencias de apoyo

Puedes consultar [esta referencia](#) de cómo integrar shiny en un cuadro de mando de flexdashboard.

### Formato de entrega

El proyecto final entregado debe ser un .zip que se llame nombre\_apellido.zip con la siguiente estructura:

- \* dashboard.Rmd: El informe en R Markdown.
- \* dat/: una carpeta con los datos (los .csv).
- \* resources/ (opcional): si necesitas otros recursos para tu informe (imágenes, un css, ...) los puedes situar en una carpeta con este nombre.

### Criterios de calificación

#### 1. Reproducibilidad

Para poder calificar tu trabajo, lo primero es que yo pueda ejecutarlo en mi ordenador sin problemas. Esto es, en una sesión nueva de R, se debe generar el html resultante sin errores. Ten cuidado con las rutas a ficheros. Deben ser relativas (p.e. dat/xxxx.csv), y apuntar a ficheros contenidos en tu entregable. Si tienes una ruta absoluta del estilo c:/users/pepa/blahla.csv, fallará en otro ordenador y será un suspenso.

#### 2. Gráficos

Los gráficos incluidos deberán representar correctamente su objetivo (mediante el uso correcto de su tipología, el mapeo de propiedades, ...) y además tendrán una presentación correcta, con nombrado de ejes adecuado al usuario (en lugar de utilizar nombres de columna), un uso adecuado de escalas de colores, etc.

Los números deben estar redondeados a cierto número de cifras significativas. No muestres, por ejemplo, un porcentaje como 5.287368%. En su lugar, utiliza 5.3% o 5%.

#### 3. Limpieza y calidad del código

Se evaluará también que el código se pueda leer fácilmente y esté correctamente modularizado. Asegúrate de cuidar los siguientes aspectos:

- Nomenclatura del código: buenos nombres a columnas, variables, funciones, ...
- DRY (Don't repeat yourself): si utilizas el mismo código varias veces, debería estar encapsulado en una función y no copiado y pegado varias veces.
- Código limpio: el código debe ser fácilmente interpretable, no intentes hacer muchas cosas en una sola línea muy compleja; sepáralo y ve nombrando adecuadamente a las variables intermedias.

- Comentarios adecuados: comenta tu código adecuadamente indicando lo que estás haciendo en cada bloque.

#### 4. Aspecto visual

Se valorará positivamente un aspecto visual cuidado y personalizado.

#### 5. Parte básica y parte opcional

De cara a la evaluación, está dividido la parte básica (que hay que realizar para conseguir el APTO) y la parte opcional (para quien quiera retarse e ir a por el 10). Las secciones sin comentario son de la parte básica, y aquellas marcadas con la etiqueta [OPCIONAL] son la opcional.