

# Tarea Grande 1

Profesores Vicente Domínguez – Luis Ramírez

Anunciada: 27 de agosto de 2018

---

## Indicaciones

- Fecha de Entrega: 10 de septiembre de 2018.
  - Debes entregar tu tarea en tu repositorio GitHub privado asignado para esta evaluación.
  - Cada hora o fracción de atraso descuenta 0,5 puntos de la nota que obtengas.
  - La tarea es *individual*. La copia será evaluada con nota 1 en el la tarea, además de las sanciones disciplinarias correspondientes.
- 

## Objetivo

El objetivo de esta tarea es que aprendas a:

- Utilizar Python para manipular datos.
- Utilizar los datos para poder hacer gráficos mediante las librería `matplotlib` o `Altair`.
- Utilizar HTML+CSS+JS para construir una página web sencilla en la que se muestren las visualizaciones construidas.

## Descripción de los datos

### Archivos

En tu repositorio de la tarea, encontrarás un archivo denominado `puntos_bip_fixed.csv`. Este fichero contiene información sobre 2092 puntos bip! (lugares donde puedes comprar o recargar tu tarjeta del sistema Transantiago) ubicados en el Gran Santiago. Sus columnas son las siguientes:

1. **Código**<sup>1</sup>: número de identificación único para cada punto bip!.
2. **Entidad**: nombre de la empresa que administra el punto.
3. **Nombre Fantasia**: nombre con el que se conoce *popularmente* al punto bip!.
4. **Comuna**: comuna donde se encuentra el punto.
5. **Longitud**: coordenada de longitud de la ubicación del lugar de recarga.
6. **Latitud**: coordenada de latitud de la ubicación del punto bip!.

## Instrucciones

### Parte 1: generación, manipulación y visualización de datos (5 puntos)

#### Generación de datos aleatorios (2 puntos)

Para hacer esta tarea más interesante, tendrás que generar datos relativos a recargas en los puntos bip! durante el año 2017, a partir de la información contenida en `puntos_bip_fixed.csv`. En particular, tendrás que generar un archivo `.csv` que contenga 10000 (diez mil) recargas bip!, **generadas aleatoriamente**. Sus columnas deben ser:

1. Código: código del punto bip! donde se realiza la transacción.
2. Fecha: fecha (en formato YYYY-MM-DD<sup>2</sup>). Recuerda que los datos deben pertenecer al año 2017, por lo que las fechas deben moverse entre 2017-01-01 y 2017-12-31.
3. Monto Carga: monto que se recargó en la transacción. La recarga mínima es \$750 y la máxima \$25500. Solo se permiten recargas con  $\delta = 10$ , por lo que el monto de recarga podrá tomar su valor según el siguiente patrón: 750, 760, ..., 25500.

Tanto las fechas y montos de carga deben generarse de forma aleatoria según lo descrito anteriormente. En el caso de los puntos bip!, como hay menos puntos bip! que transacciones solicitadas, se debe tener precaución de admitir la **selección** de elementos *con reemplazo*<sup>3</sup>.

---

<sup>1</sup>Los tildes de todas las palabras del archivo fueron omitidos intencionalmente.

<sup>2</sup>*Hint*: la librería `datetime` te permitirá realizar esto con relativa facilidad.

<sup>3</sup>El conjunto desde el que se hace esta selección corresponde al conjunto de códigos entregados en el primer archivo.

## Manipulación y visualización de los datos (3 puntos)

En esta parte de tu tarea, trabajarás con los datos entregados y los que acabas de generar, para obtener lo necesario para la confección del *dashboard*. **El entregable de esta sección es un *jupyter notebook* (extensión `.ipynb`)**, donde se aprecie el código, gráficos realizados y explicaciones que te parezcan pertinentes. De no cumplir con el tipo de archivo, recibirás una penalización en tu nota. Además, se recomienda fuertemente que utilices *pandas*. Los gráficos deben realizarse **obligatoriamente** con la ayuda de `matplotlib` o de `Altair`. En particular, debes realizar lo siguiente:<

1. Obtener y disponer en forma de **tabla** las 5 comunas con mayor cantidad de puntos de recarga bip!. Posteriormente, guardar la información en un archivo CSV de nombre `top_5_puntos.csv` y realizar un **gráfico de torta** (ver figura 1 en anexo) con esta información (0.5 puntos).
2. Generar un **gráfico de línea** (ver figura 1 en el anexo) con la cantidad de cargas realizadas según día de la semana (0.75 puntos).
3. Confeccionar un **gráfico de barra** (ver figura 2 en el anexo) donde en el eje  $x$  se encuentren los meses y en el  $y$  los montos totales recargados en cada mes (0.75 puntos).
4. Obtener una **tabla** con los montos totales recargados por comuna<sup>4</sup>. Luego, guardar las 5 comunas con mayor cantidad de dinero recargado en un archivo CSV de nombre `top_5_montos.csv` (1 punto).

**Para poder utilizar los gráficos en la Parte 2, estos deben ser exportados a un archivo de imagen.**

## Parte 2: *dashboard* (1 punto)

Con los datos y gráficos ya obtenidos, debes construir una página web que contenga cada uno de los gráficos y tablas mencionados en el enunciado. Como la confección de los gráficos será evaluada en la parte 1, la distribución de puntaje de esta parte es:

1. La página es **responsive** y los elementos que la componen están correctamente alineados y no se superponen unos con otros (**0.5 puntos**).
2. Uso de JavaScript para cargar los datos `.csv` generado en la parte 1 y así construir las tablas (**0.5 puntos**).

---

<sup>4</sup>*Hint*: es recomendable que hagas *merge* entre los *DataFrames* de recargas y de puntos bip! según algún atributo único que sea común en ambos.

## ¡Bonus! (0.5 puntos adicionales<sup>5</sup>)

¿Notaste que en el archivo `puntos_bip_fixed.csv` contiene información sobre la latitud y longitud de la ubicación de los puntos bip!? Pues bien, si logras disponer esta información en un bonito mapa incluido en tu página web, de forma que esta siga cumpliendo con los requisitos de la parte 2. Para la realización del mapa, te recomendamos explorar la librería de JavaScript Leaflet o la API de Google Maps.

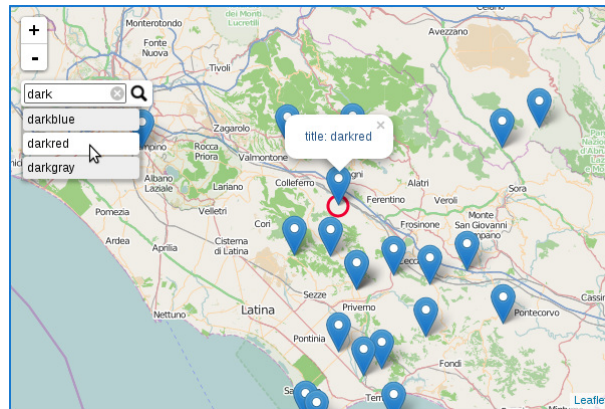


Figura 1: Mapa de referencia con pines según coordenadas en Leaflet (el menú mostrado en la imagen es totalmente opcional)

## Formato de entrega

Para evaluar con facilidad la parte 1 de esta tarea, deberás realizarla en un *jupyter notebook* y subir este archivo de extensión `.ipynb` a tu repositorio asignado. En cuanto a la parte 2, debes incluir **todos** los archivos que permitan visualizar tu página correctamente. En el repositorio base de la tarea, encontrarás una estructura propuesta que te ayudará a cumplir con el formato esperado.

## 1. Anexo

A continuación podrás encontrar imágenes que podrán guiar la confección de tus gráficos:

---

<sup>5</sup>Sí, ¡puedes optar a una nota final igual a 7.5!

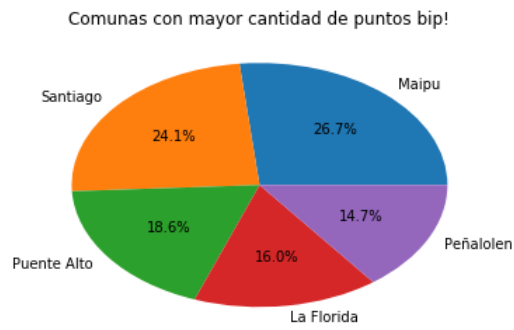


Figura 2: Top 5 comunas con mayor cantidad de puntos bip!

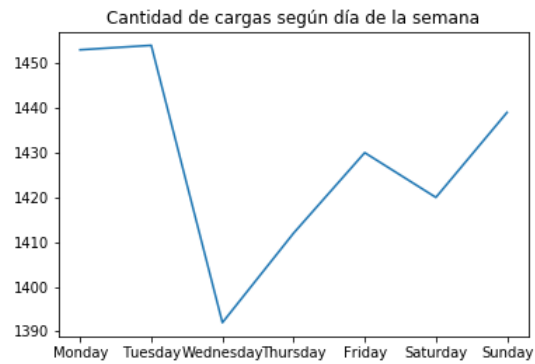


Figura 3: Cantidad de cargas según día de la semana

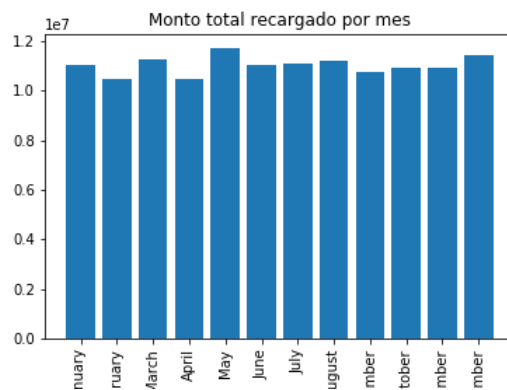


Figura 4: Monto total recargado en cada mes