

# Trabajo práctico 1 - Análisis con Python

## Viajes de taxis NYC 2016

Este trabajo práctico consiste en identificar buenas preguntas sobre un dataset particular, contestarlas aplicando lo visto en clase usando Python y volcarlas en un informe.

Se evaluará: la calidad del informe, de las preguntas propuestas y del hilo conductor que lleve entre preguntas. No se evaluará la calidad de los códigos implementados para responder estas preguntas

## Descripción del conjunto de datos

### Datos de Viajes de Taxis NYC 2016

NYC Open data<sup>1</sup> ofrece diferentes datasets. En este trabajo práctico usaremos los registros de los viajes de los **taxis amarillos** en el año **2016** ([links](#)). Los registros de viaje de taxis amarillos incluye campos de fechas, horas de recogida, hora de devolución, ubicaciones de recogida y devolución, distancias de viaje, tarifas detalladas, tipos de tarifas, tipos de pago y recuentos de pasajeros informados por el conductor entre otras.

En el siguiente link pueden encontrar una buena descripción del dataset: <https://data.cityofnewyork.us/Transportation/2018-Yellow-Taxi-Trip-Data/t29m-gskq> donde pueden entender a que se refiere cada campo, que tipo de dato es, entre otros.

### Datos de clima

Es un conjunto de datos que almacena diferente información del estado del tiempo para el año 2016 tomado diferentes mediciones por hora. Pueden encontrar una buena descripción de los datos en

<https://www.kaggle.com/meinertsen/new-york-city-taxi-trip-hourly-weather-data>, lean con atención la descripción de los datos faltantes.

## Consigna e Informe

Deberán entregar un informe de entre 5-10 páginas en el cual deben incluirse figuras, citas y todo lo que corresponda. Deberán entregarlo por email (fpousa@udesa.edu.ar) hasta el **miércoles 30 de Marzo 23:59**, el asunto del email deberá ser: "Entrega TP Python - [Lista de apellidos]", deberá enviarlo un solo integrante del grupo con copia a los demás.

---

<sup>1</sup> <https://opendata.cityofnewyork.us/>

El informe deberá incluir una descripción de los datos, las preguntas que se formularon, una breve explicación de cómo fueron abordadas las implementaciones para resolver estas preguntas. Deberá haber un hilo conductor en todo el informe, es decir, a partir de la descripción de los datos es importante entender de donde surgen las preguntas que plantean y luego cómo los métodos que usaron para responderlas impactan en los resultados, a su vez, cómo estos resultados dispararon otras preguntas y conclusiones.

A continuación se presentan algunas preguntas que pueden ser exploradas. Se sugiere que tomen esto como una guía, no hace falta responder todas, pero **sí hace falta incluir preguntas nuevas**, pueden chequear estas preguntas con el docente para corroborar el alcance de las mismas, esto lo pueden hacer durante la clase o por email.

También se verá con mucho aprecio y expectativa **la inclusión de nuevos datasets que complementen el actual**, esto permitirá hacer preguntas diferentes. Por ejemplo: podrían buscar dataset que tengan precio de diferentes activos financieros y entender si algunos cambios abruptos impactaron de alguna manera en cómo se movieron las personas en taxis, también incluir datos sobre eventos particulares a lo largo del año (festivales, catástrofes, manifestaciones, elecciones, etc) y ver cómo esto impacta en la dinámica

**Aclaración:** el dataset es muy grande, si justifican bien, podrán hacer un análisis muestral de los datos, no haciendo falta tomar todos los datos para responder una pregunta.

## Preguntas disparadoras:

¿Cómo es la distribución del tiempo de los viajes? ¿Es igual en todo el año? ¿Cambia a lo largo de las horas del día?

¿Cómo es la distribución de pasajeros en los viajes? ¿Es igual en todo el año? ¿Cambia a lo largo de las horas del día? ¿Cambia según el día de la semana?

¿Cómo es la distribución de viajes? ¿Es igual en todo el año? ¿Cambia a lo largo de las horas del día?

¿Cómo es la distribución de distancia en los viajes? ¿Es igual en todo el año? ¿Cambia a lo largo de las horas del día? ¿Cambian las zonas donde se suben los pasajeros y de donde bajan en función del día, fecha, hora?

¿Cómo covarían todas estas preguntas respecto al clima?

## Notebook con ejemplos:

<https://drive.google.com/file/d/1PpVjr-OWXW2PHZqbcgboftSsJ8pwluce/view?usp=sharing>