

## Política para el tratamiento de datos

- **VendorId** sólo acepta 1 o 2, si se da algo fuera de esto se imputa por la moda.
- **tpep\_pickup\_datetime** acepta registros entre 01-01-2018 y 31-01-2018 (23:59:59 hs). Fuera del rango se desestima
- **tpep\_dropoff\_datetime** Acepta valores entre 01-01-2018 y 01-02-2018 (hasta el mediodía 12pm). Fuera del rango se desestima.
- **tpep\_dropoff\_datetime** Se desestima, se reemplaza por el campo 'Duración del viaje'.
- **Passanger\_count** acepta registros entre 1-4, fuera de esto se imputa por moda. Imputar los ceros (0) por unos (1).
- **Trip\_distance** 3 sigmas para detectar outliers, se desestiman los outliers superiores. Los viajes con trip\_distance = 0 se desestiman.
- **PU\_Location** y **DO\_location** aceptan valores dentro del rango (1, 265), fuera de este rango se imputa por 265 (Unknown). Si ambos son 265 en simultáneo, se desestiman.
- **Rate\_CodeID** acepta valores dentro del rango (1-6), fuera de este rango se imputa por moda.
- **Store\_n\_fwd** Eliminar columna. No es relevante para el análisis.
- **Payment\_type** acepta valores (1-6), fuera de este rango se imputa por 5 (Unknown).
- **Fare\_Amount** Valores negativos se desestiman.
- **Extra** acepta valores en (0.0, 0.5, 1.0) valores fuera del rango se imputan por moda.
- **MTA\_tax** Sólo acepta valores de 0.5. Si el viaje no es outlier, imputar, caso contrario, se desestima.
- **Improvement\_surcharge** Sólo acepta valores de (0.3). Si el viaje no es outlier, imputar, caso contrario, se desestima.
- **Tip\_amount**: Eliminar columna. No es relevante para el análisis.
- **Tolls\_amount** no acepta valores negativos, marcar como outliers lo que esté arriba de 3 sigmas. Se marca como outliers los que estén por arriba de 45 dólares
- **Total\_amount** payment\_type = 3 o 4 son problematicos, marcado como outliers + desestimado