

## Datos de FHV (vehículos de alquiler).

FHS puede tener varios significados dependiendo del contexto. En este caso particular, dentro de los registros de FHV de 2018, **FHS** se refiere a **High-Volume Fleet Service (Servicio de Flota de Alto Volumen)**.

Las empresas de FHS son aquellas que ofrecen servicios de transporte compartido a gran escala, como **Uber Pool** o **Lyft Line**. En los registros de FHV, el campo **SR\_Flag** se utiliza para identificar si un viaje fue parte de un servicio de FHS. Un valor de **1** en este campo indica que el viaje fue compartido, mientras que un valor nulo indica que no lo fue.

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 1290116 entries, 0 to 1290115
```

```
Data columns (total 7 columns):
```

```
#   Column          Non-Null Count  Dtype
```

```
---  -----  -
```

```
0   dispatching_base_num    1290116 non-null object
```

```
1   pickup_datetime         1290116 non-null datetime64[us]
```

```
2   dropOff_datetime        1290116 non-null datetime64[us]
```

```
3   PUlocationID            274816 non-null float64
```

```
4   DOlocationID            1107055 non-null float64
```

```
5   SR_Flag                  0 non-null    float64
```

```
6   Affiliated_base_number  1290116 non-null object
```

```
dtypes: datetime64[us](2), float64(3), object(2)
```

```
memory usage: 68.9+ MB
```

**Dispatching\_base\_num:** El número de licencia base de TLC (Comisión de Taxis y Limusinas de la Ciudad de Nueva York) de la base que envió el viaje.

**Pickup\_datetime:** La fecha y hora de recogida del viaje.

**DropOff\_datetime:** Fecha y hora de llegada del viaje

**PULocationID:** Zona ubicación de taxis TLC en la que comenzó el viaje.

**DOLocationID:** Zona de ubicación de taxis TLC en la que finalizó el viaje.

**SR\_Flag:** Indica si el viaje fue parte de una cadena de viajes compartidos ofrecida por una empresa de FHV de alto volumen (p. ej., Uber Pool, Lyft Line). Para los viajes compartidos, el valor es 1. Para los viajes no compartidos, este campo es nulo.

**Affiliated\_base\_number:** Podría ser un identificador único para una base a la que la base despachadora (dispatching\_base\_num) está afiliada. Esto podría ser relevante para empresas que operan con múltiples bases o franquicias.

## **FHVHV: datos de viajes de FHV de alto volumen (High Volume FHV).**

**Hvfhs\_license\_num:** El número de licencia de TLC de la base o negocio de HVFHS. A partir de septiembre de 2019, los licenciarios de HVFHS son los siguientes: HV0002: Juno; HV0003: Uber; HV0004: Via; HV0005: Lyft.

**Dispatching\_base\_num:** El número de licencia de base de TLC de la base que envió el viaje.

**originating\_base\_num:** número de base de la base que recibió la solicitud de viaje original.

Es igual a **Dispatching\_base\_num:** El número de licencia de base de TLC de la base que envió el viaje.

**request\_datetime:** fecha/hora en la que el pasajero solicitó ser recogido.

**on\_scene\_datetime:** fecha/hora en la que el conductor llegó al lugar de recogida (Accesible Solo vehículos).

**Pickup\_datetime:** La fecha y hora de la recogida del viaje.

**DropOff\_datetime:** La fecha y hora de la llegada del viaje.

**PULocationID:** Zona de taxis de TLC en la que comenzó el viaje.

**DOLocationID:** Zona de taxis de TLC en la que terminó el viaje.

**trip\_miles:** millas totales para el viaje del pasajero.

**trip\_time:** tiempo total en segundos para el viaje del pasajero.

**base\_passenger\_fare:** tarifa base del pasajero antes de peajes, propinas, impuestos y tarifas.

**tolls:** monto total de todos los peajes pagados en el viaje.

**Bcf:** monto total recaudado en el viaje para el Black Car Fund.

**sales\_tax:** monto total recaudado en el viaje para el impuesto a las ventas de NYS.

**congestion\_surcharge:** monto total recaudado en el viaje para el recargo por congestión de NYS.

**airport\_fee:** tarifa \$2.50 tanto para dejar como para recoger en los aeropuertos LaGuardia, Newark y John F. Kennedy.

**Tips:** monto total de propinas recibidas del pasajero.

**driver\_pay:** pago total del conductor (sin incluir peajes ni propinas y neto de comisiones, recargos o impuestos).

**shared\_request\_flag:** ¿El pasajero aceptó un viaje compartido, independientemente de si fue emparejado? (S/N).

**shared\_match\_flag:** ¿El pasajero compartió el vehículo con otro pasajero que reservó por separado en algún momento durante el viaje? (S/N).

**access\_a\_ride\_flag:** ¿El viaje se administró en nombre de la Autoridad de Transporte Metropolitano (MTA)? (S/N).

**wav\_request\_flag:** ¿El pasajero solicitó un vehículo accesible para sillas de ruedas (WAV)? (S/N).

**wav\_match\_flag:** ¿El viaje se realizó en un vehículo accesible para sillas de ruedas (WAV)? (S/N).

## Taxis amarillos:

```
<class 'pandas.core.frame.DataFrame'>
```

RangeIndex: 2964624 entries, 0 to 2964623

Data columns (total 19 columns):

#	Column	Dtype
0	VendorID	int32
1	tpep_pickup_datetime	datetime64[us]
2	tpep_dropoff_datetime	datetime64[us]
3	passenger_count	float64
4	trip_distance	float64
5	RatecodeID	float64
6	store_and_fwd_flag	object
7	PULocationID	int32
8	DOLocationID	int32
9	payment_type	int64

```
10 fare_amount      float64
11 extra            float64
12 mta_tax          float64
13 tip_amount       float64
14 tolls_amount     float64
15 improvement_surcharge float64
16 total_amount     float64
17 congestion_surcharge float64
18 Airport_fee      float64
dtypes: datetime64[us](2), float64(12), int32(3), int64(1), object(1)
memory usage: 395.8+ MB
```

**Vendor ID:** Código del proveedor de tecnología para taxi que provee el registro. 1: Creative Mobile Technologies, LLC; 2: VeriFone Inc.

**tpep\_pickup\_datetime:** la fecha y la hora cuando se activa el taxímetro.

**tpep\_dropoff\_datetime:** la fecha y la hora cuando se desactiva el taxímetro.

**Passenger\_count:** el número de pasajeros en el vehículo. Este es un valor ingresado por el controlador.

**Trip\_distance:** La distancia del viaje transcurrida en millas informada por el taxímetro.

**PULocationID:** Zona de Taxis TLC en la que se activó el taxímetro.

**DOLocationID:** Zona de Taxis TLC en la que se desactivó el taxímetro.

**RateCodeID:** Código de tarifa final vigente al final del viaje: 1= Tarifa estándar; 2=JFK; 3=Newark; 4=Nassau o Westchester; 5=Tarifa negociada; 6=Viaje en grupo.

**Store\_and\_fwd\_flag:** Esta bandera indica si el registro del viaje se mantuvo en la memoria del vehículo antes de enviarlo al proveedor, es decir, si se “almacenó y reenvió”, porque el vehículo no tenía una conexión con el servidor. Y = almacenar y reenviar el viaje; N = no es un viaje almacenado y reenviado.

**Payment\_type:** Un código numérico que indica cómo pagó el pasajero el viaje. 1= Tarjeta de crédito; 2= Efectivo; 3= Sin cargo; 4= Disputa; 5= Desconocido; 6= Viaje anulado.

**Fare\_amount:** La tarifa de tiempo y distancia calculada por el taxímetro.

**Extra:** Extras y recargos varios. Actualmente, solo se incluyen los cargos de \$0,50 y \$1 por hora punta y por noche.

**MTA\_tax:** Impuesto MTA de \$0,50 que se activa automáticamente en función de la tarifa medida en uso.

**Improvement\_surcharge:** Recargo por mejoras de \$0,30 para viajes en el momento de la bajada de bandera. El recargo por mejoras comenzó a aplicarse en 2015.

**Tip\_amount:** Monto de la propina: este campo se completa automáticamente para las propinas con tarjeta de crédito. Las propinas en efectivo no están incluidas.

**Tolls\_amount:** Importe total de todos los peajes pagados en el viaje.

**Total\_amount:** Importe total de todos los peajes pagados en el viaje.

**Congestión\_Surcharge:** Monto total recaudado en el viaje por el recargo por congestión en NYS.

**Airport\_fee:** Tarifa de \$1.25 por retiro únicamente en los aeropuertos LaGuardia y John F. Kennedy.

## Taxis verdes, datos: utilizaremos los de los Años 2020-2021.

\*Enero del 2024 tiene 56 000 datos (muestra abajo); Enero del 2020 tiene 447 769 datos y Diciembre del 2021 tiene 100 000 aprox. Estos son datos extraídos desde la página web de datos oficiales de la Ciudad de Nueva York. [TLC Trip Record Data - TLC \(nyc.gov\)](https://www.tlc.nyc.gov/trip-record)

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 56551 entries, 0 to 56550
```

```
Data columns (total 20 columns):
```

```
#   Column              Non-Null Count  Dtype
```

```
---  -----  -
```

```
0   VendorID          56551 non-null  int32
```

```
1   lpep_pickup_datetime  56551 non-null  datetime64[us]
```

```
2   lpep_dropoff_datetime  56551 non-null  datetime64[us]
```

```
3   store_and_fwd_flag    53136 non-null  object
```

```
4   RatecodeID           53136 non-null  float64
```

```
5   PULocationID         56551 non-null  int32
```

```
6   DOLocationID         56551 non-null  int32
```

```
7   passenger_count       53136 non-null  float64
```

```
8   trip_distance         56551 non-null  float64
```

```
9   fare_amount           56551 non-null  float64
```

10 extra	56551 non-null float64
11 mta_tax	56551 non-null float64
12 tip_amount	56551 non-null float64
13 tolls_amount	56551 non-null float64
14 ehail_fee	0 non-null float64
15 improvement_surcharge	56551 non-null float64
16 total_amount	56551 non-null float64
17 payment_type	53136 non-null float64
18 trip_type	53133 non-null float64
19 congestion_surcharge	53136 non-null float64

dtypes: datetime64[us](2), float64(14), int32(3), object(1), memory usage: 8.0+ MB

### Diccionario de estos datos:

**VendorID:** Un código que indica el proveedor de LPEP que proporcionó el registro. 1= Creative Mobile Technologies, LLC; 2= VeriFone Inc.

**lpep\_pickup\_datetime:** La fecha y la hora en que se activó el taxímetro.

**lpep\_dropoff\_datetime:** La fecha y la hora en que se desactivó el taxímetro.

**Store\_and\_fwd\_flag:** Esta bandera indica si el registro del viaje se mantuvo en la memoria del vehículo antes de enviarlo al proveedor, también conocido como "almacenamiento y reenvío", porque el vehículo no tenía una conexión con el servidor. Y = viaje de almacenamiento y reenvío; N = no es un viaje de almacenamiento y reenvío.

**RateCodeID:** El código de tarifa final vigente al final del viaje. 1 = Tarifa estándar; 2 = JFK

3 = Newark; 4 = Nassau o Westchester; 5 = Tarifa negociada; 6 = Viaje en grupo.

**PULocationID:** Zona de taxis de TLC en la que se activó el taxímetro.

**DOLocationID:** Zona de taxis de TLC en la que se desactivó el taxímetro.

**Passenger\_count:** La cantidad de pasajeros en el vehículo. Este es un valor ingresado por el conductor.

**Trip\_distance:** La distancia del viaje transcurrida en millas informada por el taxímetro.

**Fare\_amount:** La tarifa por tiempo y distancia calculada por el taxímetro.

**Extra:** Extras y recargos varios. Actualmente, esto solo incluye los cargos de \$0,50 y \$1 por hora pico y por noche.

**MTA\_tax:** Impuesto de MTA de \$0,50 que se activa automáticamente según la tarifa del taxímetro en uso.

**Tip\_amount:** Monto de la propina: este campo se completa automáticamente para las propinas con tarjeta de crédito. Las propinas en efectivo no están incluidas.

**Tolls\_amount:** Monto total de todos los peajes pagados en el viaje.

**ehail\_fee:** Esta columna probablemente indica una tarifa de llamada electrónica, que posiblemente se cobre por reservar el viaje en taxi a través de una aplicación móvil. No todos los viajes en taxi pueden tener esta tarifa.

**Improvement\_surcharge:** Se aplica un recargo por mejora de \$0,30 a los viajes con pasajeros en el momento de la bajada de bandera. El recargo por mejora comenzó a aplicarse en 2015.

**Total\_amount:** El monto total cobrado a los pasajeros. No incluye las propinas en efectivo.

**Payment\_type:** Un código numérico que indica cómo pagó el pasajero el viaje. 1 = Tarjeta de crédito; 2 = Efectivo; 3 = Sin cargo; 4 = Disputa; 5 = Desconocido; 6 = Viaje anulado.

**Trip\_type:** Un código que indica si el viaje fue un viaje con conductor en la calle o un viaje con conductor que se asigna automáticamente en función de la tarifa medida en uso, pero que el conductor puede modificar. 1 = Viaje con conductor en la calle; 2 = Viaje con conductor.

**congestion\_surcharge:** Esta columna probablemente almacena cualquier recargo aplicado por congestión de tráfico durante el viaje.