

Analizando Mil Millones de Filas: del bosque a la hoja



José Luis Cano
Next-Step Consultores

josemanuelcano@next-step.es
[@jlcanoc](https://twitter.com/jlcanoc)



Power BI
Spain Users Group

2 Octubre 2017



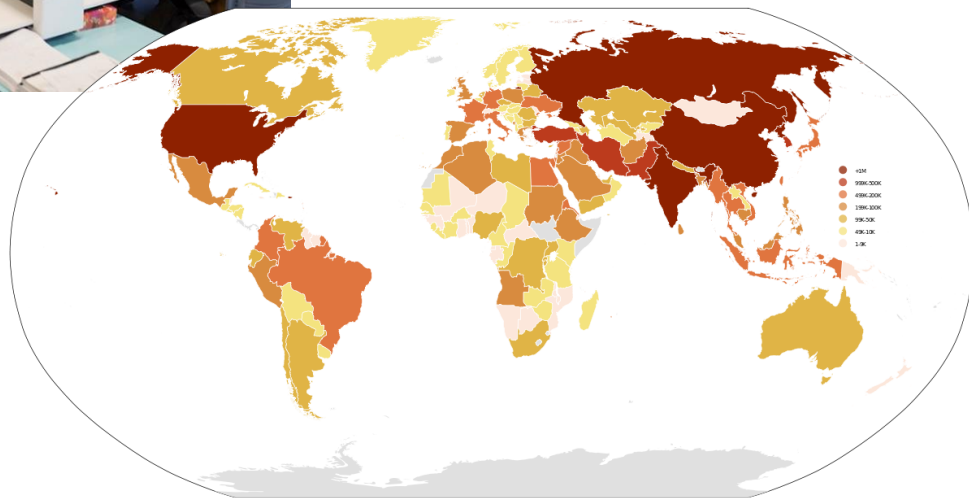
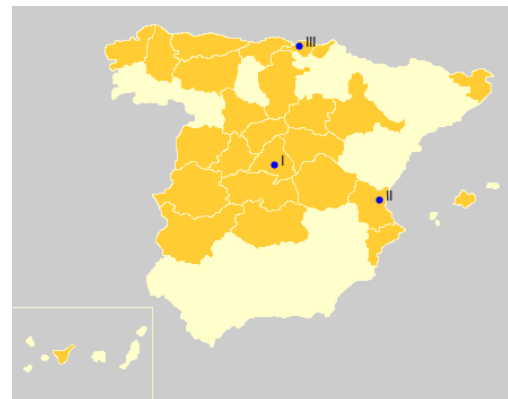
Objetivos



- Escenarios de Power BI
- Azure Analysis Services
- Del bosque a la hoja: demo Taxi NYC

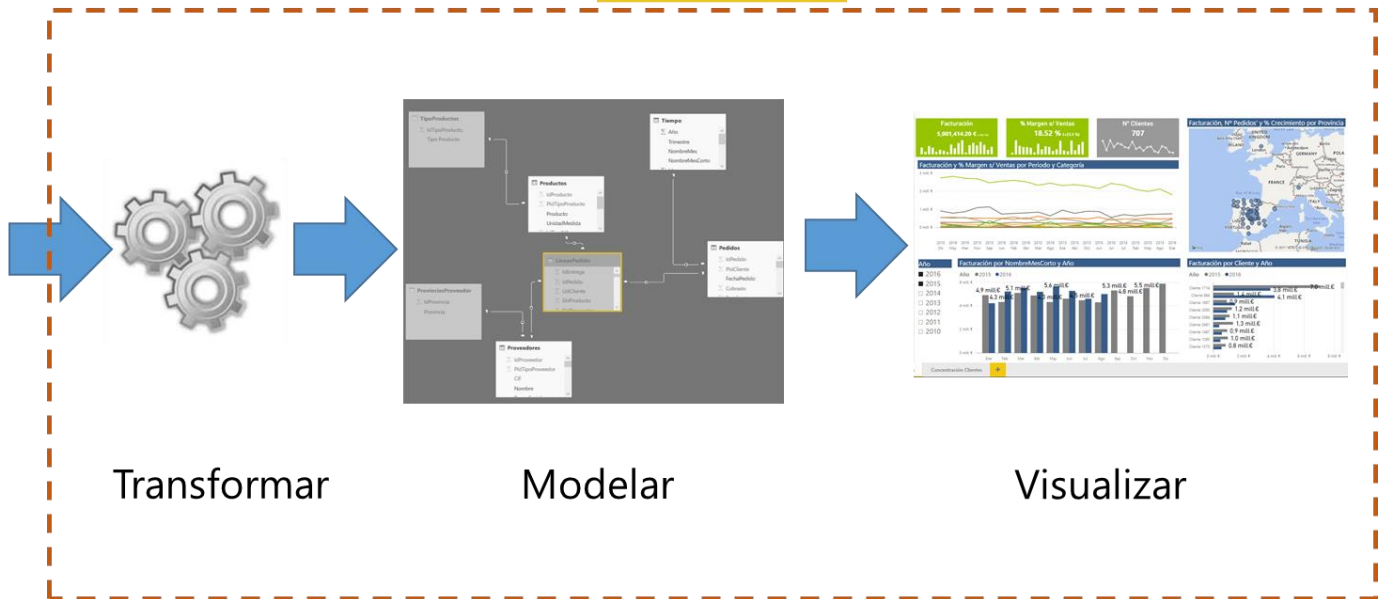


Escenarios de Power BI



Escenarios de Power BI

Personal - Departamental



¿Cuántas filas caben en mi Power BI?



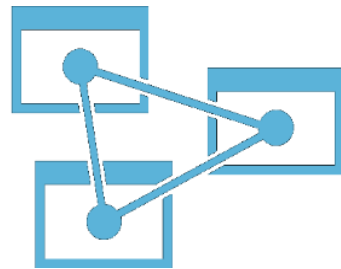
Vertipaq



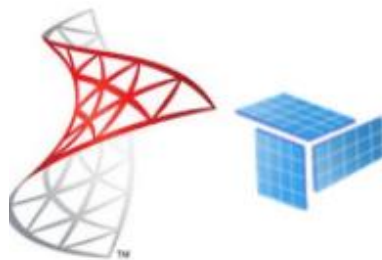
Power BI



Power Pivot



Azure Analysis
Services



SSAS Tabular



Power BI
Spain Users Group

Vertipaq



Mes	Producto	País	Importe
Enero	Bicicletas	España	51.352
Enero	Calzado	España	47.847
Enero	Calzado	Francia	41.748
Enero	Calzado	Italia	84.471
Febrero	Bicicletas	España	64.475
Febrero	Calzado	España	51.214
Febrero	Calzado	Francia	63.894
Febrero	Calzado	Italia	49.623



Vertipaq



Mes	Producto	País	Importe
Enero	Bicicletas	España	51.352
Enero	Calzado	España	47.847
Enero	Calzado	Francia	41.748
Enero	Calzado	Italia	84.471
Febrero	Bicicletas	España	64.475
Febrero	Calzado	España	51.214
Febrero	Calzado	Francia	63.894
Febrero	Calzado	Italia	49.623



Vertipaq



Mes
Enero
Enero
Enero
Enero
Febrero
Febrero
Febrero
Febrero
...



Mes	Inicio	Fin
Enero	1	4
Febrero	5	8
...



¿Cuántas ~~filas~~ caben en mi Power BI?

¿Cuánto espacio cabe en mi Power BI?

Quiero analizar miles de millones de filas

Quiero a e filas

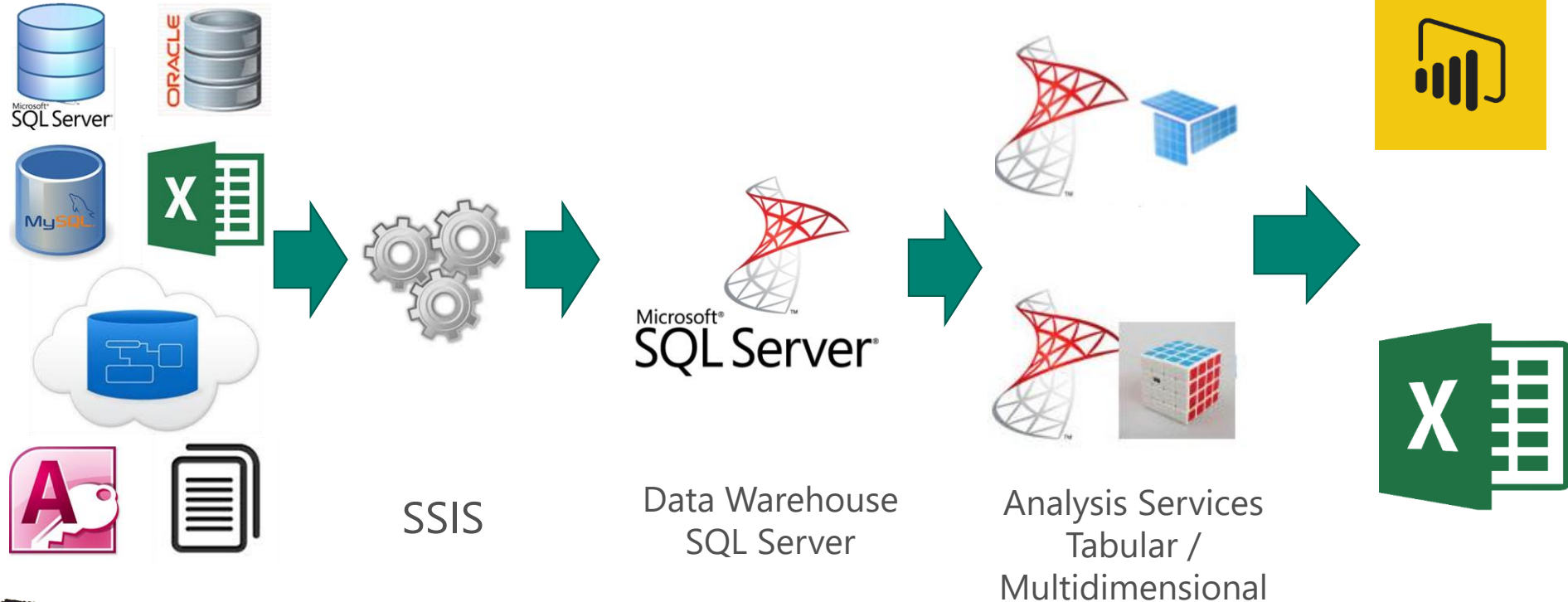


Find More @
NerdTests.com

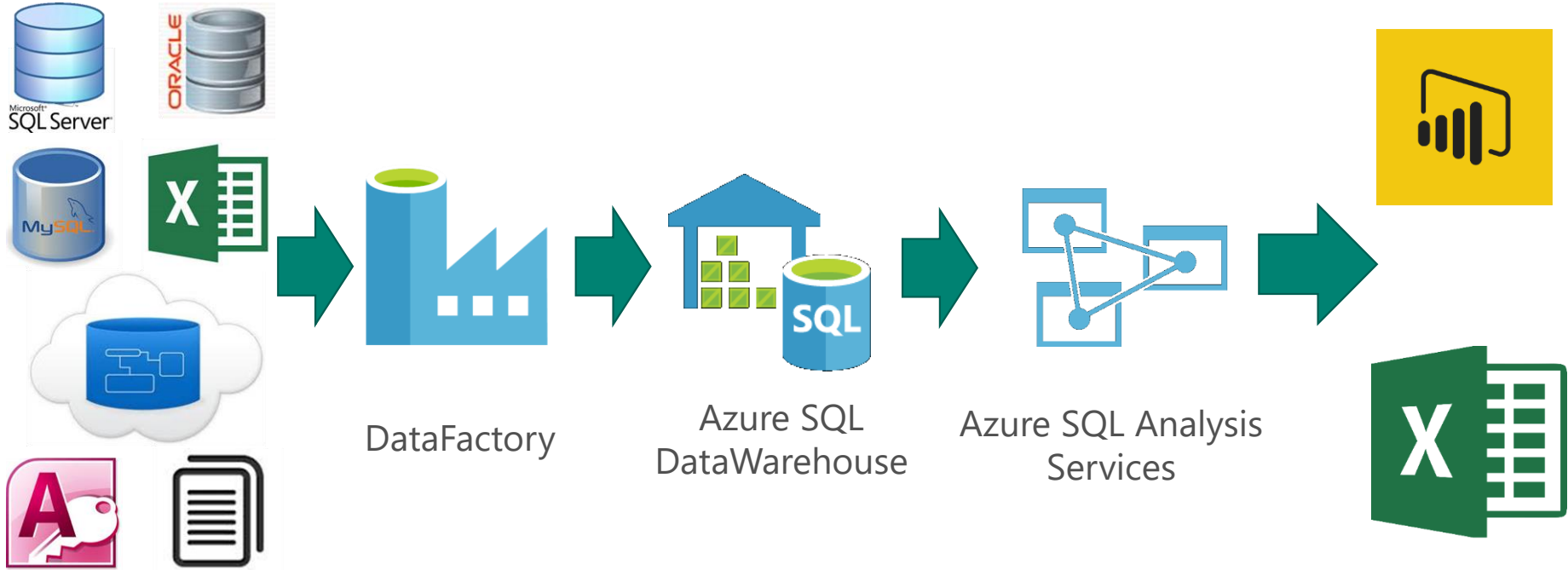


Escenarios de Power BI

Corporativo On-Premise

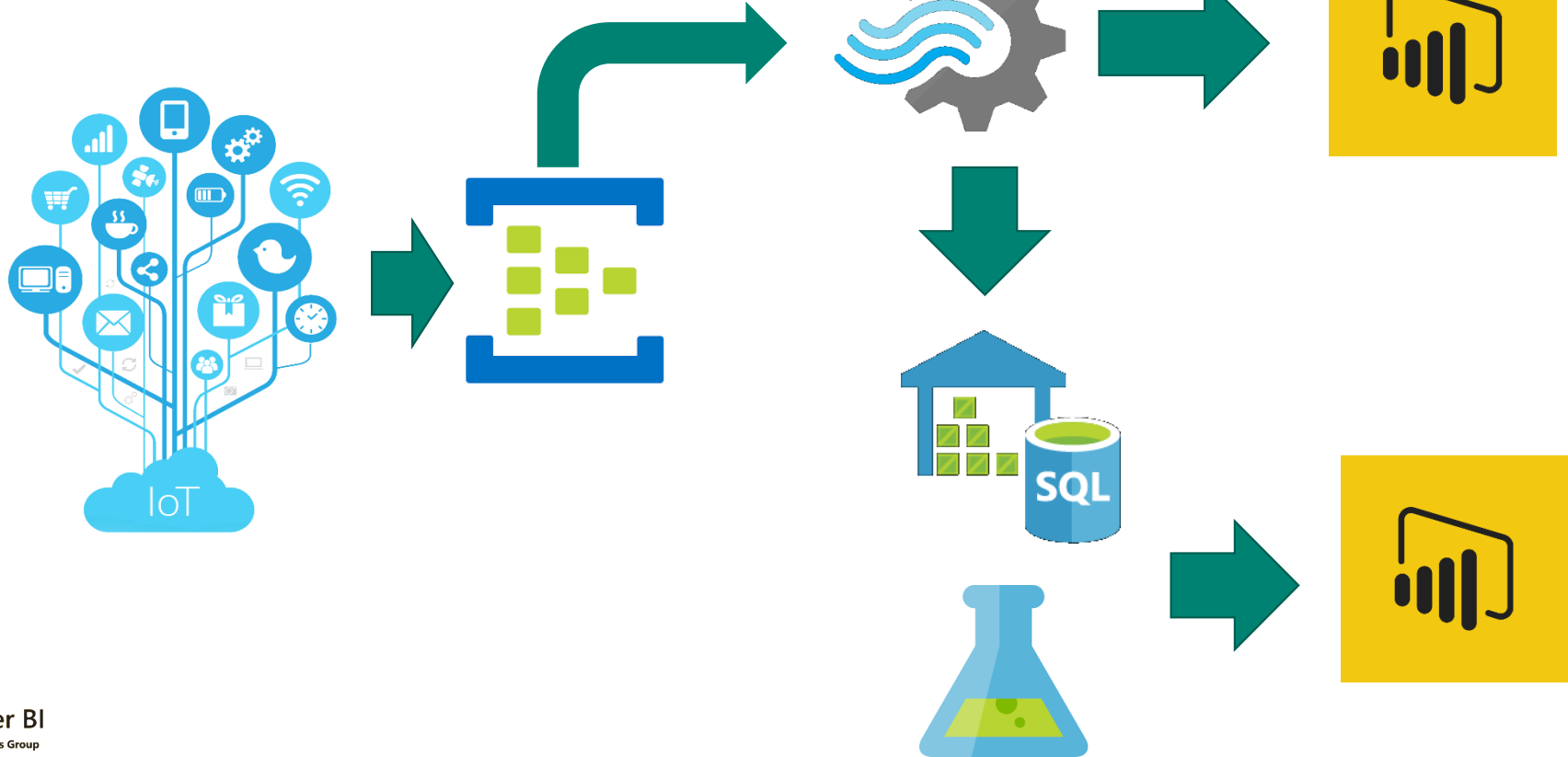


Escenarios de Power BI Corporativo PaaS

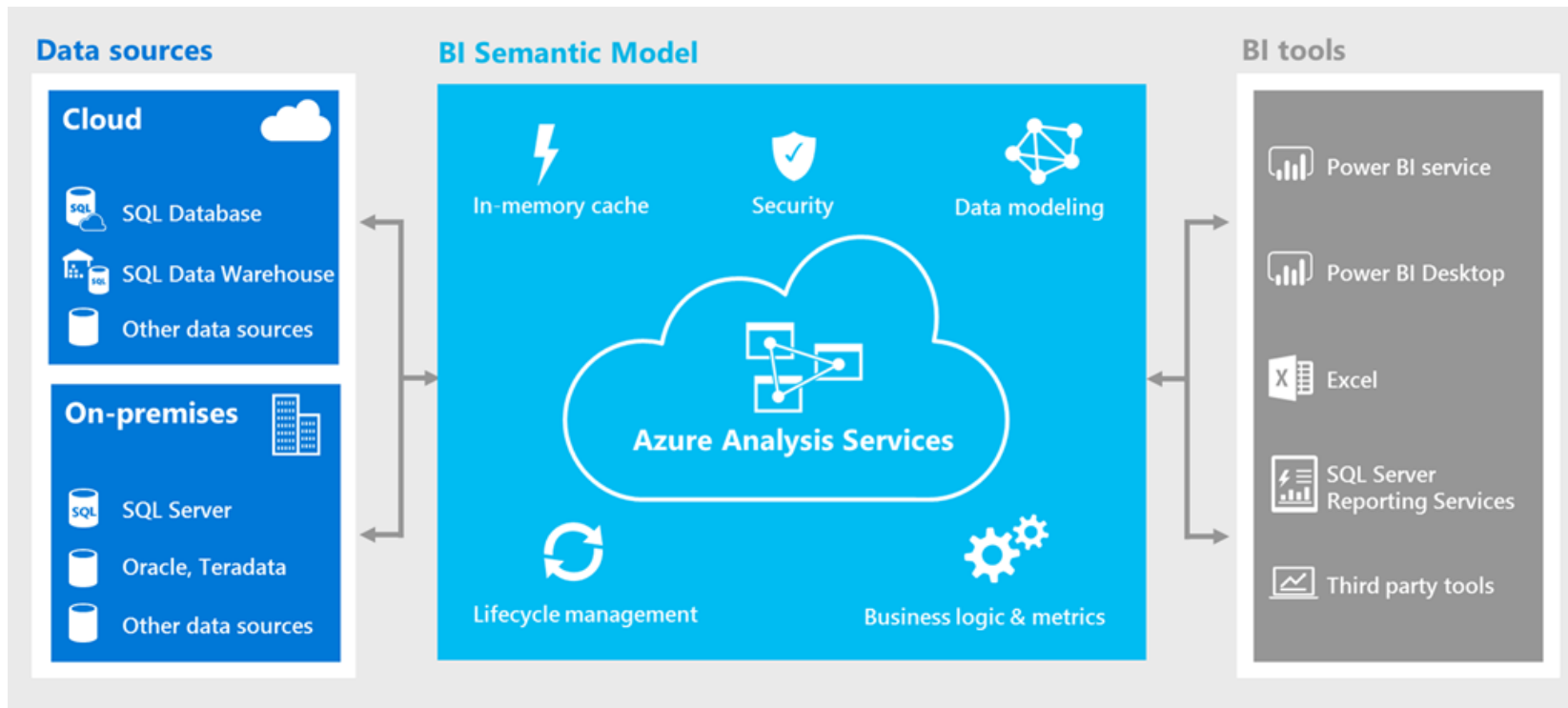
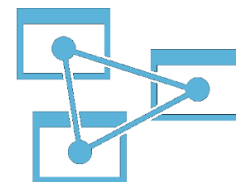


Escenarios de Power BI

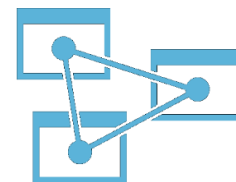
Otros escenarios



Azure Analysis Services



Demo Azure Analysis Services



 Visual Studio
2017



Power BI
Spain Users Group

Demo Taxi Nueva York



http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

← → ↺ 🏠

Frequently Asked Questions

TLC News

TLC Site Map

Contact/Visit TLC

Facebook Twitter YouTube Instagram LinkedIn

Online Transactions (LARS)

Apply for a License

Pay Renewal Fee

Pay Summons

Pay Other Fees

Update License Information

Additional Information

I am a...
Choose One

I am here to...
Please select an option above

nyc.gov/html/tlc/html/about/trip_record_data.shtml

payment types, and driver-reported passenger counts. The data used in the attached datasets were collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology providers authorized under the TaxiCab & Livery Passenger Enhancement Programs (TPEP/LPEP). The trip data was not created by the TLC, and TLC makes no representations as to the accuracy of these data.

The For-Hire Vehicle ("FHV") trip records include fields capturing the dispatching base license number and the pick-up date, time, and taxi zone location ID (shape file below). These records are generated from the FHV Trip Record submissions made by bases. Note: The TLC publishes base trip record data as submitted by the bases, and we cannot guarantee or confirm their accuracy or completeness. Therefore, this may not represent the total amount of trips dispatched by all TLC-licensed bases. The TLC performs routine reviews of the records and takes enforcement actions when necessary to ensure, to the extent possible, complete and accurate information.

For trip record data including TLC taxi zone location IDs, location names and corresponding boroughs for each ID can be found [here](#). A shapefile containing the boundaries for the taxi zones can be found [here](#).

Trip Sheet Data (CSV Format)

2017

2016

2015

2014

2013

2012

2011

2010

2009

Data Dictionary – Yellow Taxi Trip Records

September 28, 2015

Page 1 of 1

This data dictionary describes yellow taxi trip data. For dictionaries describing green taxi and FHV data, please visit http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml.

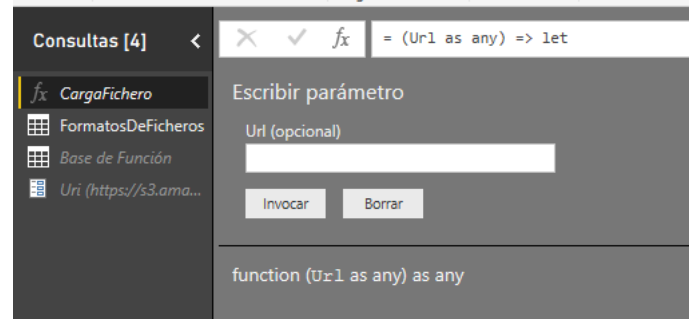
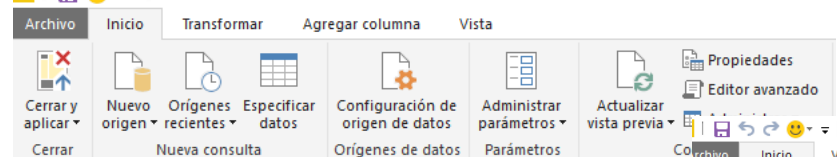
Field Name	Description
VendorID	A code indicating the TPEP provider that provided the record. 1= Creative Mobile Technologies, LLC; 2= VeriFone Inc.
tpcp_pickup_datetime	The date and time when the meter was engaged.
tpcp_dropoff_datetime	The date and time when the meter was disengaged.
Passenger_count	The number of passengers in the vehicle. This is a driver-entered value.
Trip_distance	The elapsed trip distance in miles reported by the taximeter.
Pickup_longitude	Longitude where the meter was engaged.
Pickup_latitude	Latitude where the meter was engaged.
RateCodeID	The final rate code in effect at the end of the trip. 1= Standard rate 2=JFK



Formatos



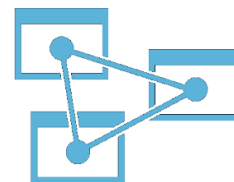
Formatos De Ficheros - Editor de consultas



Año	Mes	Column1	Column2	Column3	Column13	Column16	Column19
2009	1	vendor_name	Trip_Pickup_DateTime	Trip_Dropoff_DateTime	Fare_Amt	Total_Amt	
2009	2	vendor_name	Trip_Pickup_DateTime	Trip_Dropoff_DateTime	Fare_Amt	Total_Amt	
2009	3	vendor_name	Trip_Pickup_DateTime	Trip_Dropoff_DateTime	Fare_Amt	Total_Amt	
2009	4	vendor_name	Trip_Pickup_DateTime	Trip_Dropoff_DateTime	Fare_Amt	Total_Amt	
2009	5	vendor_name	Trip_Pickup_DateTime	Trip_Dropoff_DateTime	Fare_Amt	Total_Amt	
2009	6	vendor_name	Trip_Pickup_DateTime	Trip_Dropoff_DateTime	Fare_Amt	Total_Amt	
2009	7	vendor_name	Trip_Pickup_DateTime	Trip_Dropoff_DateTime	Fare_Amt	Total_Amt	
2009	8	vendor_name	Trip_Pickup_DateTime	Trip_Dropoff_DateTime	Fare_Amt	Total_Amt	
2009	9	vendor_name	Trip_Pickup_DateTime	Trip_Dropoff_DateTime	Fare_Amt	Total_Amt	
2009	10	vendor_name	Trip_Pickup_DateTime	Trip_Dropoff_DateTime	Fare_Amt	Total_Amt	
2009	11	vendor_name	Trip_Pickup_DateTime	Trip_Dropoff_DateTime	Fare_Amt	Total_Amt	
2009	12	vendor_name	Trip_Pickup_DateTime	Trip_Dropoff_DateTime	Fare_Amt	Total_Amt	
2010	1	vendor_id	pickup_datetime	dropoff_datetime	fare_amount	total_amount	
2010	2	vendor_id	pickup_datetime	dropoff_datetime	fare_amount	total_amount	
2010	3	vendor_id	pickup_datetime	dropoff_datetime	fare_amount	total_amount	
2010	4	vendor_id	pickup_datetime	dropoff_datetime	fare_amount	total_amount	
2010	5	vendor_id	pickup_datetime	dropoff_datetime	fare_amount	total_amount	
2010	6	vendor_id	pickup_datetime	dropoff_datetime	fare_amount	total_amount	
2010	7	vendor_id	pickup_datetime	dropoff_datetime	fare_amount	total_amount	
2010	8	vendor_id	pickup_datetime	dropoff_datetime	fare_amount	total_amount	
2010	9	vendor_id	pickup_datetime	dropoff_datetime	fare_amount	total_amount	
2010	10	vendor_id	pickup_datetime	dropoff_datetime	fare_amount	total_amount	
2010	11	vendor_id	pickup_datetime	dropoff_datetime	fare_amount	total_amount	
2010	12	vendor_id	pickup_datetime	dropoff_datetime	fare_amount	total_amount	

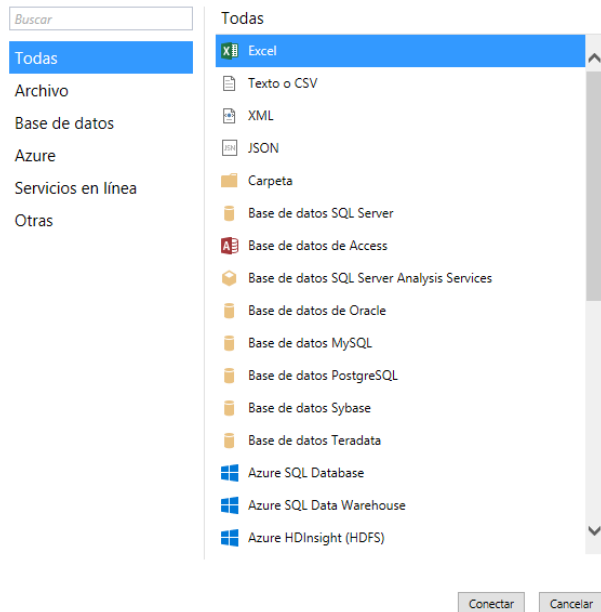


Azure Analysis Services

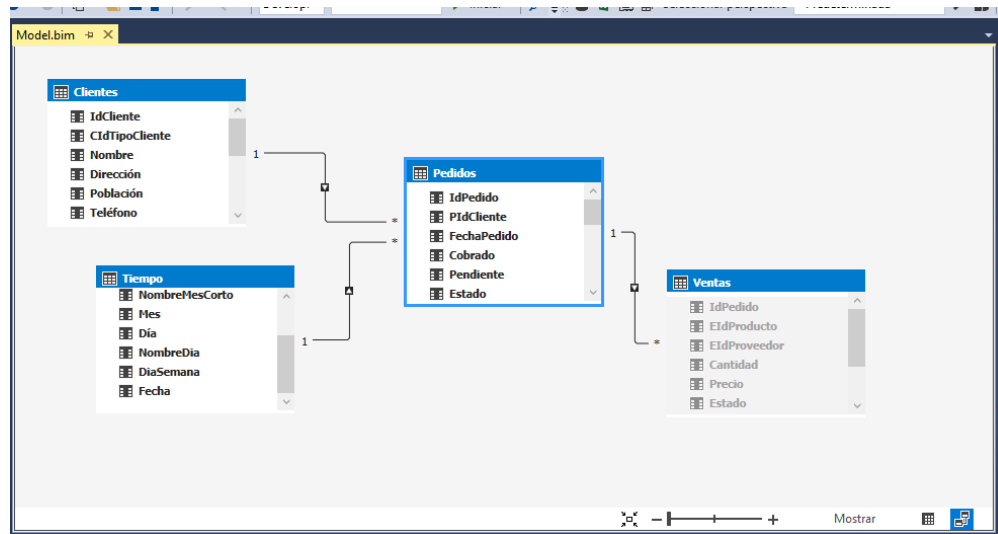


Experiencia Power BI

Obtener datos



Transformación



Modelo de Datos



Particiones en AAS

- Optimiza consumo de memoria
- Permite anexar distintos orígenes

Administrador de partición

Use particiones para dividir una tabla en partes lógicas que se puedan procesar de forma independiente.

Tabla: YellowTaxi

Buscar nombres de partición

Nombre de la partición	Último procesado
2009	Nunca
2010	Nunca
2011	Nunca

Nuevo Copiar Eliminar

Detalles - 2011

Nombre de la partición: 2011

Expresión de consulta:

```
let
Source = #SQL/nextstepdw database windows net;NextStep",
dbo_V_YellowTaxi = Source[{Schema="dbo",Item="V_YellowTaxi"}][Data],
#"Otras columnas quitadas" = Table.SelectColumns(dbo_V_YellowTaxi,{"Pickup_date","Passenger_count","Trip_distance","Pickup_longitude",
"Pickup_latitude","Dropoff_longitude","Dropoff_latitude","Payment_type","Total_amount","DriverId","PULocationID","DOLocationID",
"DurationInSeconds","MinuteKey"}),
#"Filas filtradas" = Table.SelectRows("#Otras columnas quitadas",each [Pickup_date] >= #date(2011, 1, 1) and [Pickup_date] <= #date(2011, 12,
31)),
#"Índice agregado" = Table.AddIndexColumn("#Filas filtradas","Índice", 2000000000, 1),
#"Columnas con nombre cambiado" = Table.RenameColumns("#Índice agregado",{{"Índice","IdViaje"}})
in
#"Columnas con nombre cambiado"
```

Validar Diseño...

Último procesado: Nunca

Aceptar Cancelar

Del Bosque a la hoja

```

DEFINE
    MEASURE YellowTaxi[Nº Filas] =
        COUNTROWS ( 'YellowTaxi' )
EVALUATE
    (
        ROW (
            "Nº de Filas", [Nº Filas],
            "Nº de Filas en Millones", FORMAT ( [Nº Filas] / 1E6, "###,###,##0.00" )
        )
    )
    
```

100 %

Messages Results	
[Nº de Filas]	[Nº de Filas en Millones]
1179737655	1.179,74

Del Bosque a la hoja

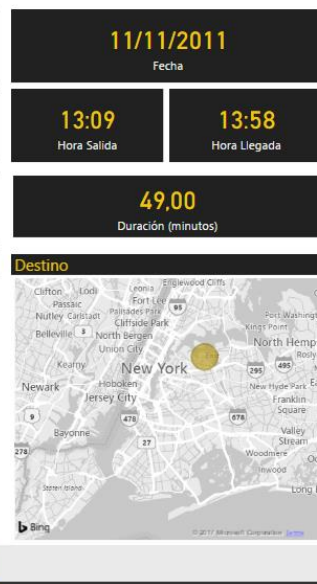
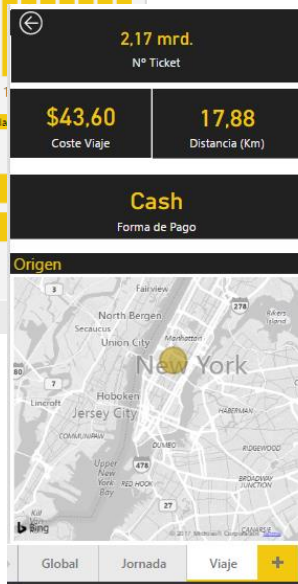
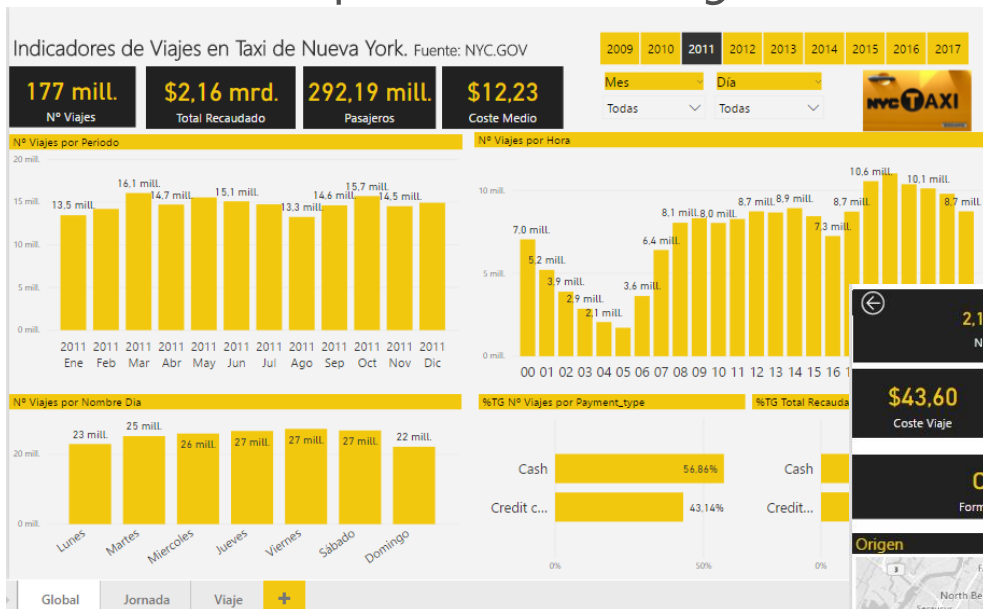
```
1 EVALUATE
2 ( CALCULATETABLE ( 'YellowTaxi', YellowTaxi[IdViaje] = 1112719186 ) )
```

100 %

YellowTaxi[Pickup_date]	YellowTaxi[Pass...	YellowTaxi[Trip_...	YellowTaxi[Pick...	YellowTaxi[Pick...	YellowTaxi[Drop...	YellowTaxi[Drop...
11/06/2010 0:00:00	1	36,04	-73,85166	40,661323	-74,335137	40,56191

asazure://westeurope.asazur... | joseluis.cano@next-ste... | NYC_Taxi | 00:00:01

Del Bosque a la hoja



Visualizar Detalle – Novedad Sep 2017

Indicadores de Viajes en Taxi de Nueva York. Fuente: NYC.GOV

169 mill.

Nº Viajes

\$1,99 mrd.

Total Recaudado

282,93 mill.

Pasajeros

\$11,77

Coste Medio

2009

2010

2011

2012

2013

2014

2015

Mes

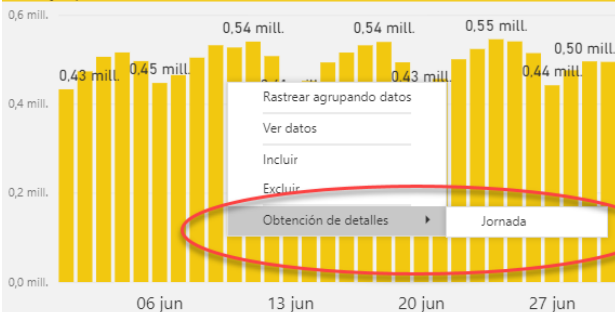
Día

Todas

Todas



Nº Viajes por Fecha

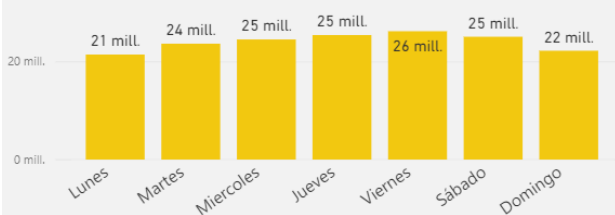


Rastrear agrupando datos
Ver datos
Incluir
Excluir

Obtención de detalles

Jornada

Nº Viajes por Nombre Día



Global

Jornada

Viaje

Nº Viajes por Hora



Datos del día 11/06/2010

541 mil

Nº Viajes

\$6,49 mill.

Total Recaudado

903.89 mil

Pasajeros

12,38

Duración Media (minutos)

\$11,99

Coste Medio

1,67

Promedio Pasajeros

%TG Nº V

IdViaje	Coste Medio	Duración Media (minutos)	Pasajeros	Time	Nombre
1.014.301.960,00	\$18,30	27,00	4	00:00	Richar
1.014.705.940,00	\$27,10	28,00	2	00:00	John I
1.014.722.920,00	\$9,10	9,00	2	00:00	John I
1.014.837.580,00	\$7,10	7,00	2	00:00	John I
1.015.541.440,00	\$14,30	17,00	5	00:00	Richar
1.015.788.460,00	\$5,90	4,00	5	00:00	David
1.016.066.740,00	\$15,50	12,00	2	00:00	Peter
1.016.067.880,00	\$13,50	16,00	5	00:00	Peter
1.017.162.760,00	\$5,10	3,08	2	00:00	John I
1.017.305.620,00	\$5,50	3,72	2	00:00	Peter
1.017.343.360,00	\$7,50	5,83	2	00:00	Peter
1.017.360.940,00	\$8,70	9,78	2	00:00	Greg I
Total	\$11,99	12,38	903895		

Global

Jornada

Viaje

Hora:

00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15

Min:

00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16

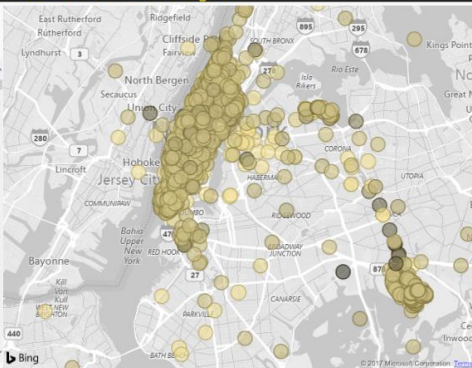
2,44 mill.

Distancia Total (Km)

4,51

Distancia Media (Km)

Muestra de Puntos de Recogida



FILTROS

Filtros de obtención de

detalles

Fecha

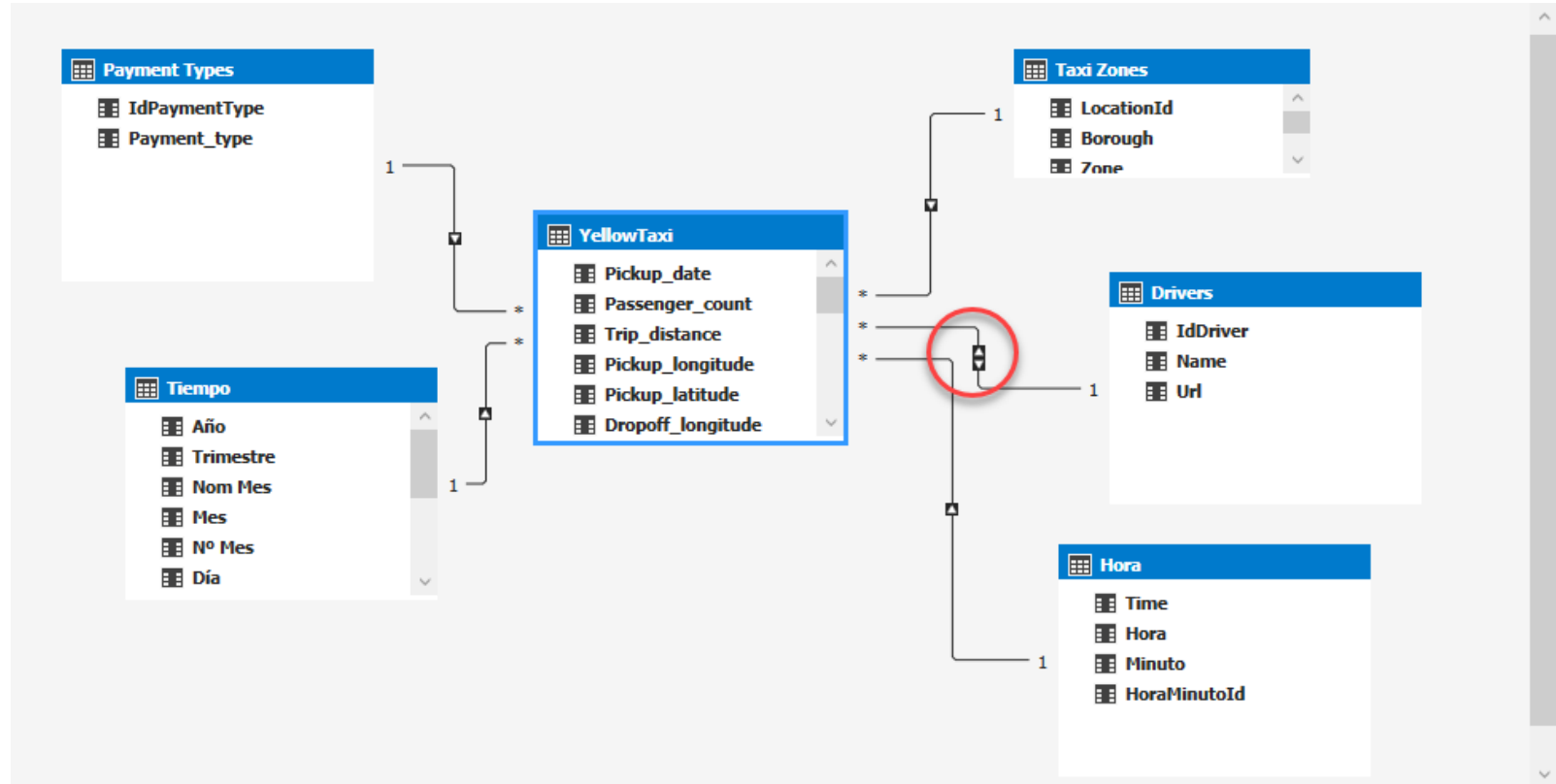
es 11/06/2010 00:00

Filtros de nivel de informe

Año

no es 2017 o 2016

Relación de doble filtro



¡Muchas gracias!

José Luis Cano
Next-Step Consultores