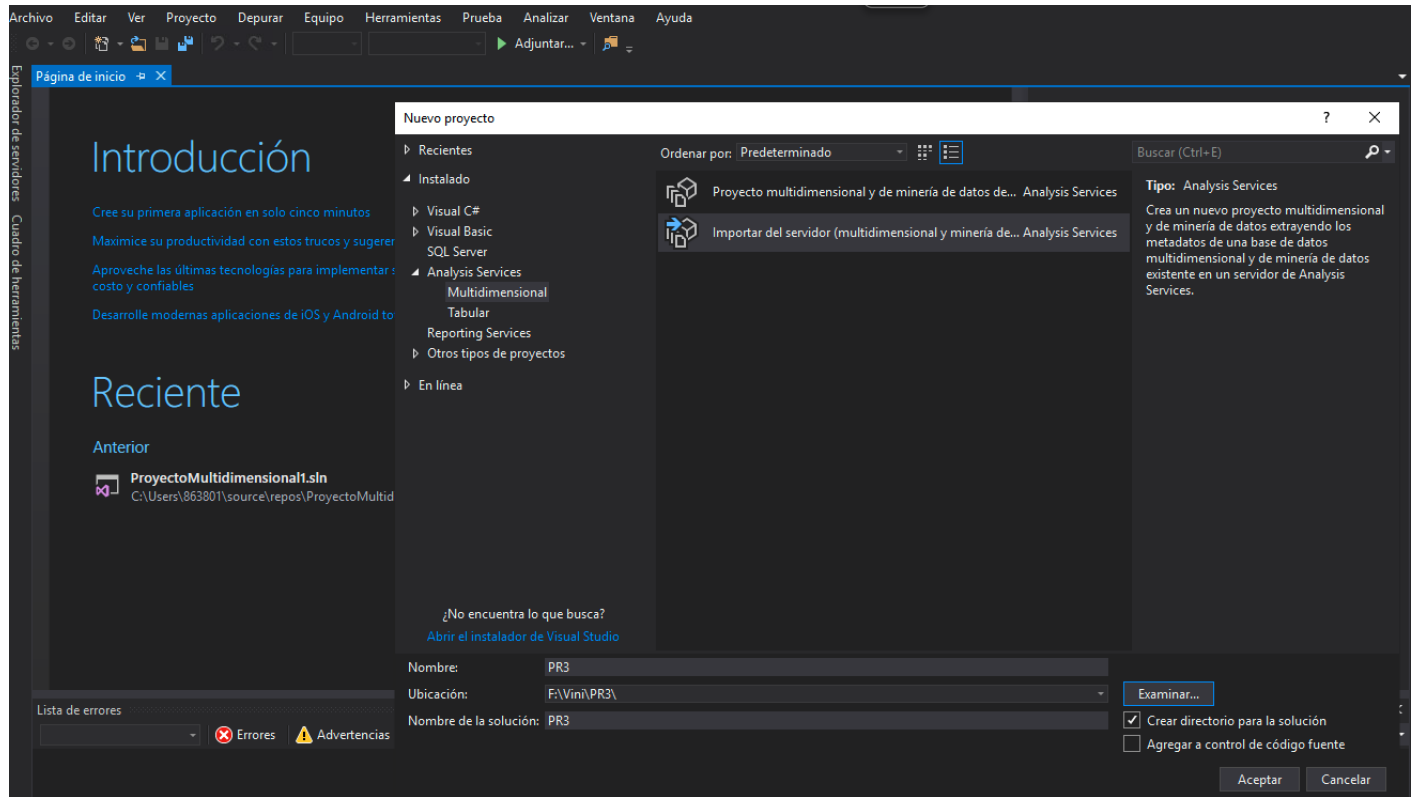


Creación del proyecto



Muy importante guardarlo en la unidad F.

Configuro la base de datos destino en mi servidor.

Base de datos de origen

Seleccione la base de datos de Analysis Services que se va a importar.



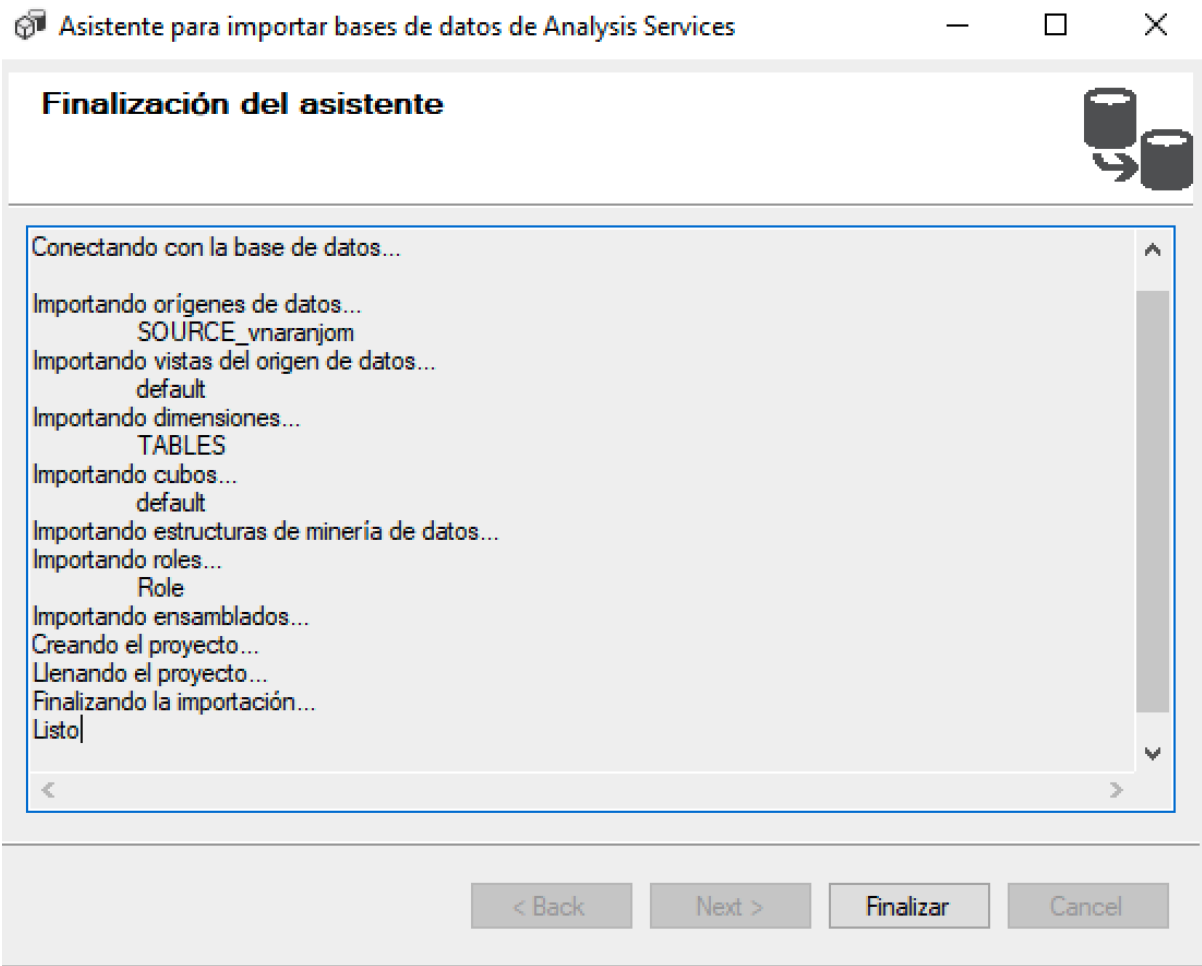
El asistente extrae el contenido de la base de datos elegida y crea un proyecto de Analysis Services basado en los datos extraídos.

Servidor:

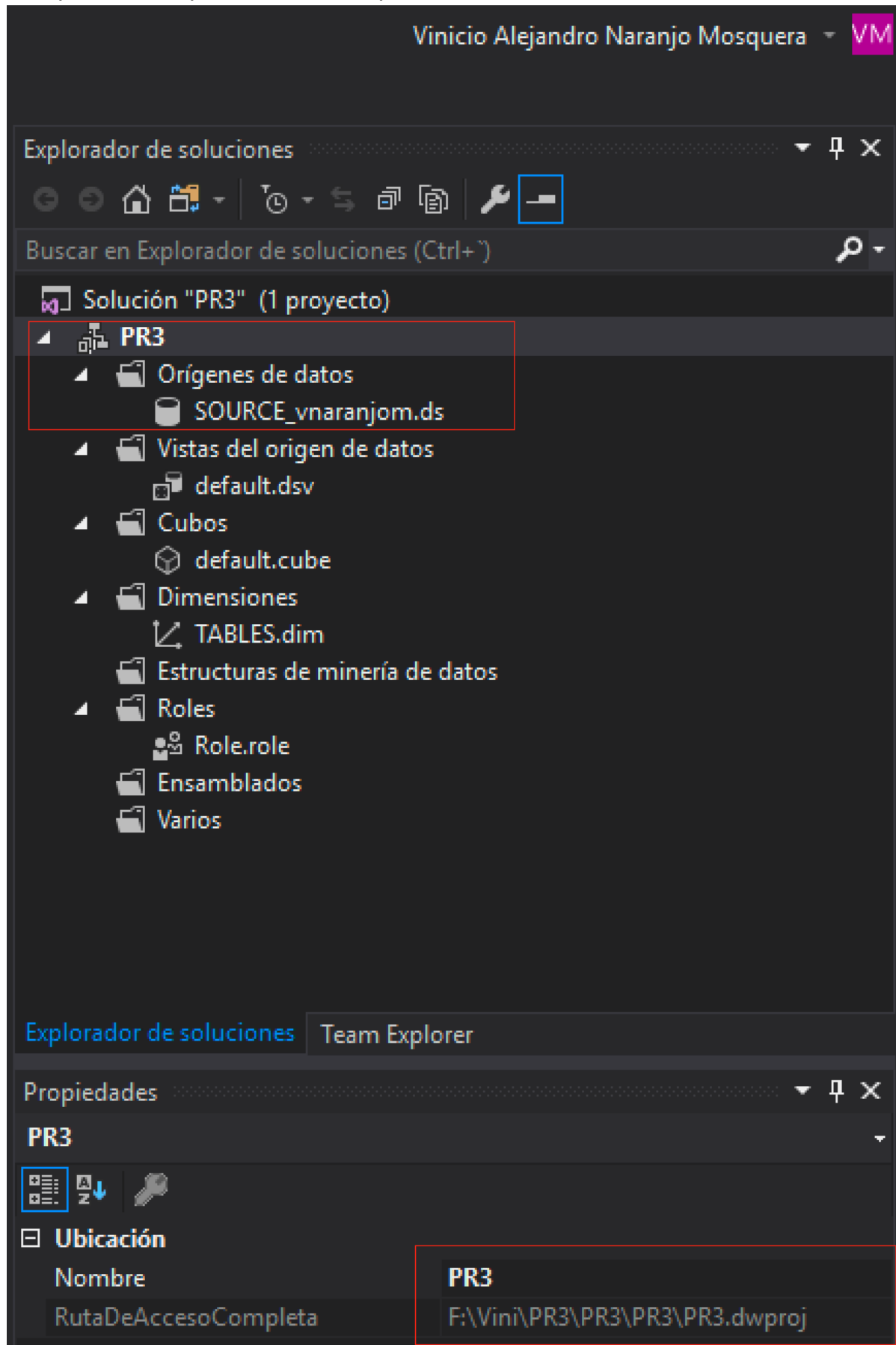
Base de datos:



Comprobamos que todo se ha creado correctamente.

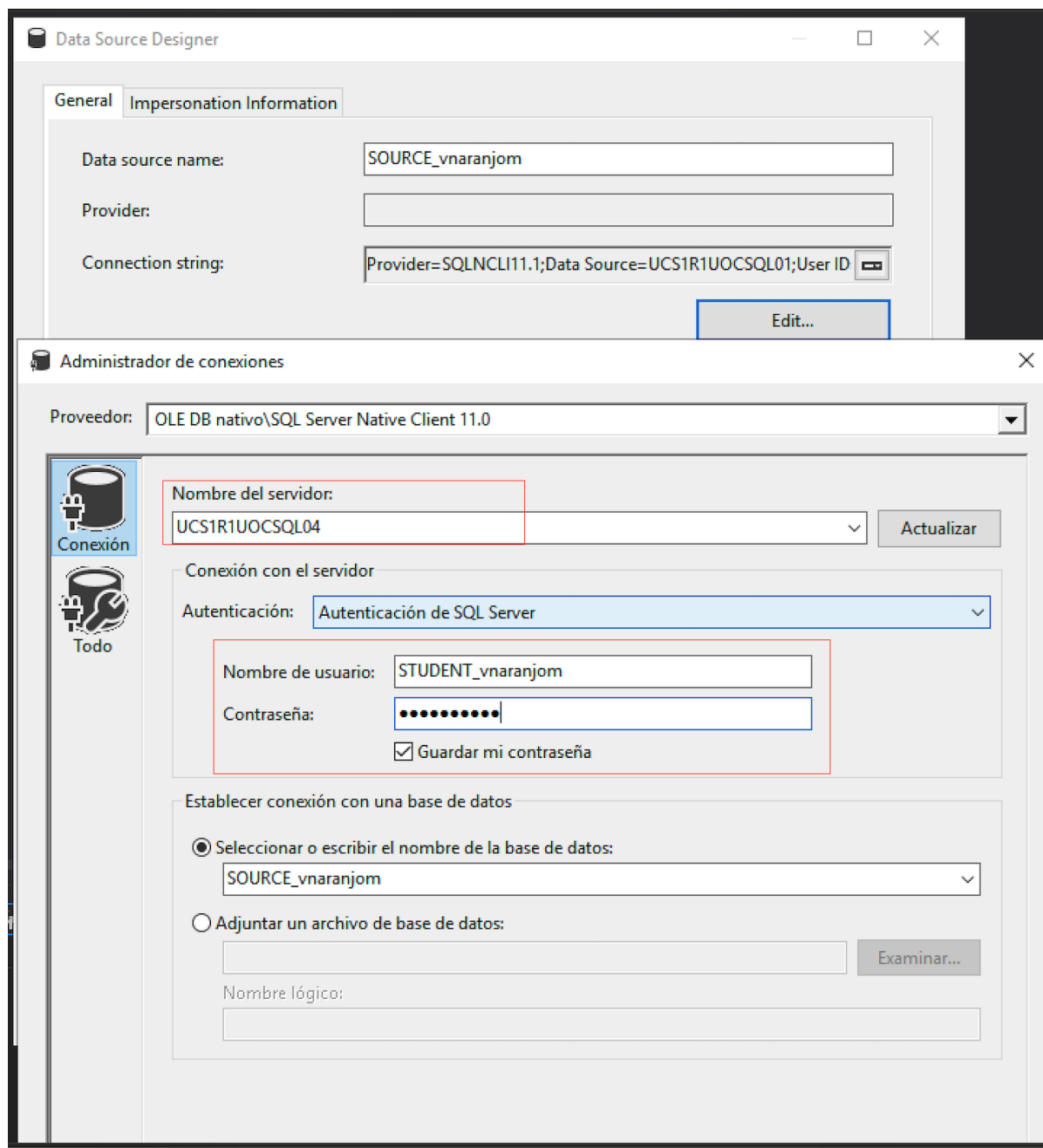


Compruebo la importación en el explorador de soluciones.

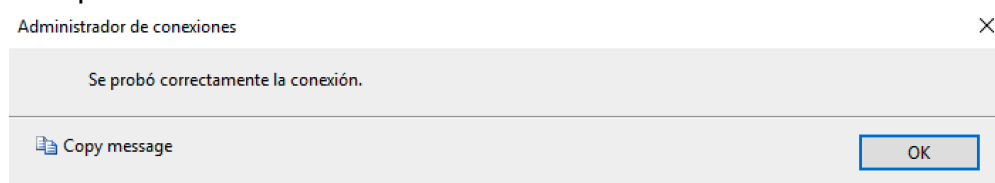


Configuración origen de datos

Configuro la nueva conexión al nuevo servidor UCS1R1UOCSQL04



Compruebo la conexión



Estoy configurando la suplantación para mi fuente de datos y he optado por 'Heredar' para simplificar la gestión de credenciales, aprovechando la configuración ya existente en el

Data Source Designer

General Impersonation Information

☐ Utilizar un nombre de usuario y una contraseña de Windows específicos

Nombre de usuario:

Contraseña:

☐ Utilizar la cuenta de servicio

☐ Utilizar las credenciales del usuario actual

☒ Heredar

OK Cancel Help

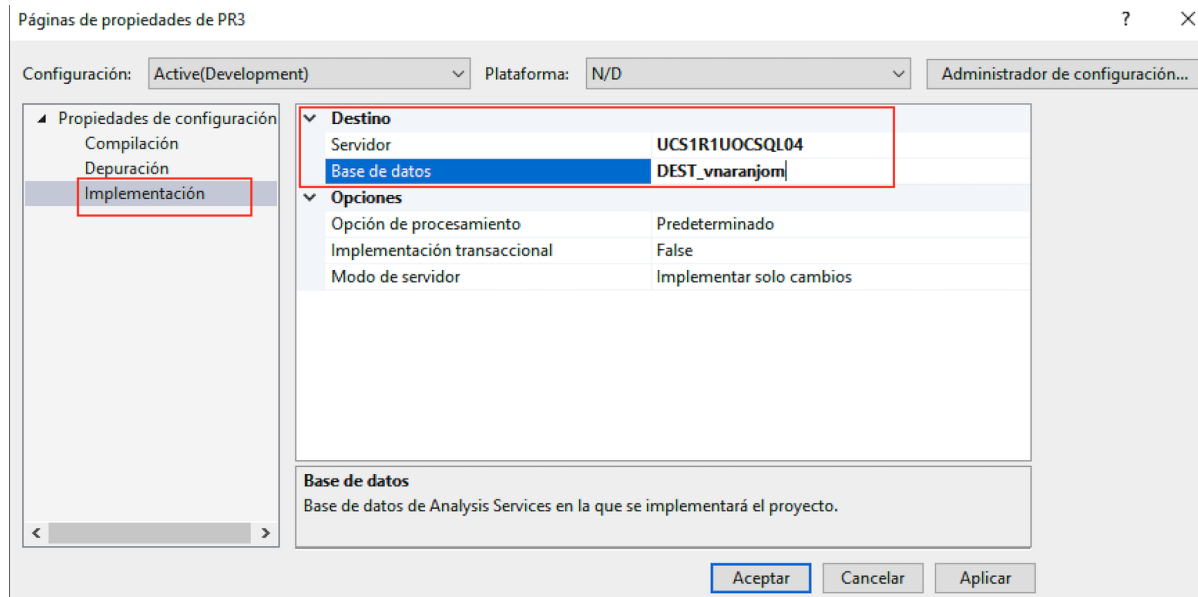
Configuración destino de datos

The screenshot shows the Visual Studio Solution Explorer with a context menu open for a project named 'PR3'. The menu includes options for database editing, schema generation, progress monitoring, processing, compilation, and deployment. It also includes options for limiting scope, changing views, adding, managing NuGet packages, setting as startup project, debugging, and source control. At the bottom, there are options for cutting, pasting, deleting, renaming, downloading the project, and properties.

Icon	Command	Shortcut
	Editar base de datos	
	Generar esquema relacional...	
	Mostrar progreso de implementación	
	Procesar...	
	Compilar	
	Recompilar	
	Implementar	
	Limitar el ámbito a esto	
	Nueva vista de Explorador de soluciones	
	Agregar	
	Administrar paquetes NuGet...	
	Establecer como proyecto de inicio	
	Depurar	
	Control de código fuente	
	Cortar	Ctrl+X
	Pegar	Ctrl+V
	Quitar	Supr
	Cambiar nombre	
	Descargar el proyecto	
	Propiedades	Alt+ Entrar

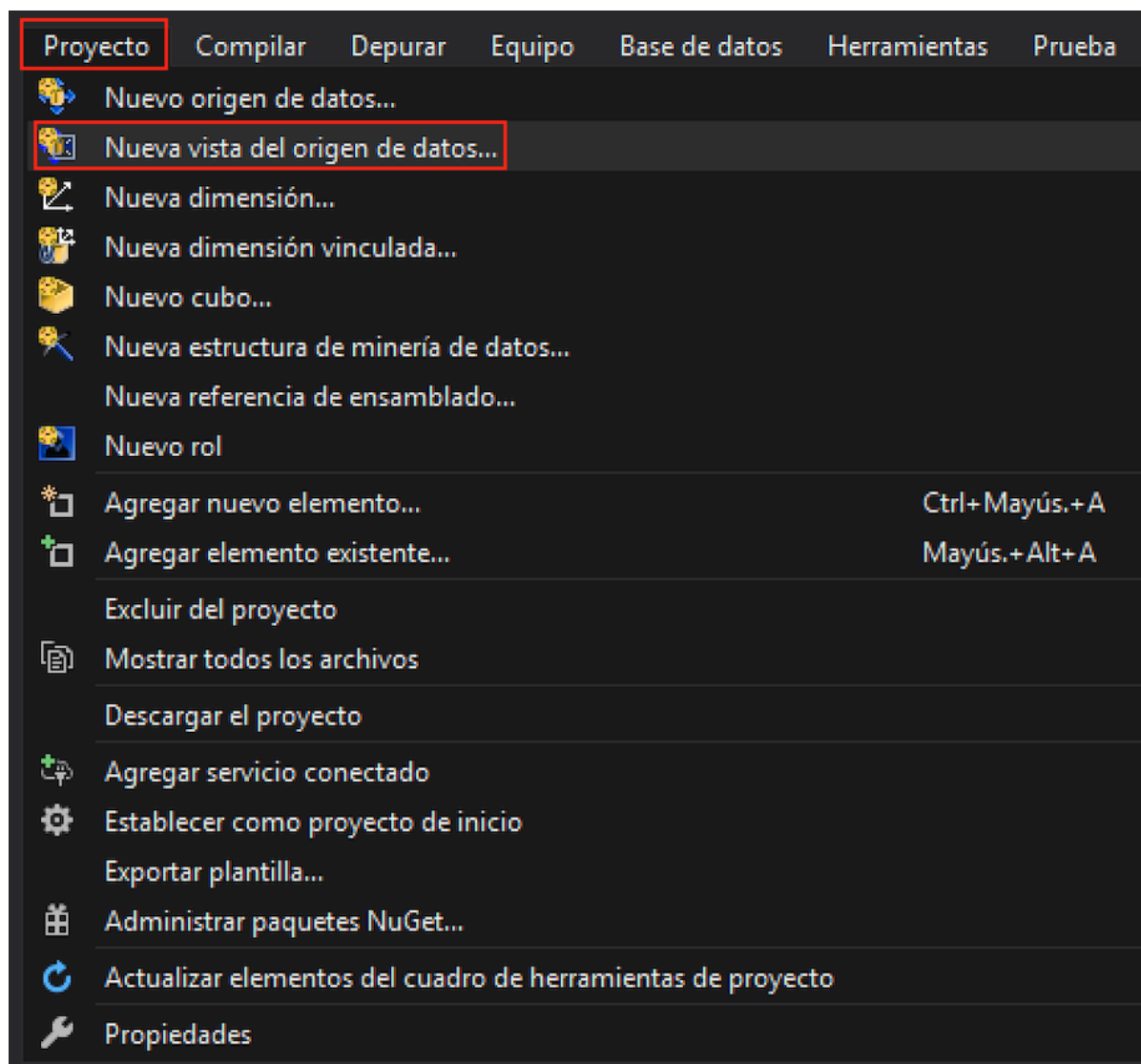
Hacemos click en Implementación y después configuramos el Servidor UCS1R1U0CSQL04

con la base de datos DEST_vnaranjom



Creando una vista del origen de datos

Elegimos Proyecto y después Nueva vista del origen de datos



Compruebo que el origen de datos es correcto.

Asistente para vistas del origen de datos

Seleccionar un origen de datos
Seleccione un origen de datos relacional o cree uno nuevo.

Origenes de datos relacionales:
SOURCE_vnaranjom

Propiedades del origen de datos:

Propiedad	Valor
Data Source	UCS1R1UOCSQL04
Initial Catalog	SOURCE_vnaranjom
Persist Securit...	True
Provider	SQLNCLI11.1
User ID	STUDENT_vnaranjom

Nuevo origen de datos... Avanzadas...

< Back Next > Finish >>| Cancel

Pulso en Next y selecciono las tablas necesarias para la vista de la tabla de hechos de FACT_NYTXI_TRIP

Seleccionar tablas y vistas

Seleccione los objetos de la base de datos relacional que deben incluirse en la vista del origen de datos.



Objetos disponibles:

Nombre	Tipo
DIM_LICENSE (dbo)	Tabla
FACT_FHV_TRIP (dbo)	Tabla

Filtrar:

☐ Mostrar objetos del sistema

Objetos incluidos:

Nombre	Tipo
FACT_NYTAXI_TRIP (d...	Tabla
DIM_PAYMENT (dbo)	Tabla
DIM_TIME (dbo)	Tabla
DIM_LOCATION (dbo)	Tabla
DIM_RATE (dbo)	Tabla
DIM_VENDOR (dbo)	Tabla

**Agregar tablas relacionadas**

< Back

Next >

Finish >>|

Cancel

Asistente para vistas del origen de datos

Finalización del asistente
Proporcione un nombre y haga clic a continuación en Finalizar para crear la nueva vista del origen de datos.

Nombre:
Vista_NYTAXI_TRIP

Vista previa:

Vista_NYTAXI_TRIP	
FACT_NYTAXI_TRIP (dbo)	
DIM_PAYMENT (dbo)	
DIM_TIME (dbo)	
DIM_LOCATION (dbo)	
DIM_RATE (dbo)	
DIM_VENDOR (dbo)	

< Back Next > **Finalizar** Cancel

La tabla **FACT_NYTAXI_TRIP** analiza los datos de los viajes en taxis amarillos en Nueva York, y está asociada con las siguientes dimensiones:

1. **FACT_NYTAXI_TRIP**: Tabla de hechos que contiene las métricas como duración, número de pasajeros, distancia, importe del viaje, entre otros.
2. **DIM_TIME**: Para analizar el tiempo (fecha y hora) de recogida y entrega de los viajes.
3. **DIM_LOCATION**: Para zonas de recogida y entrega en Nueva York.
4. **DIM_RATE**: Información sobre los tipos de tarifa aplicados.
5. **DIM_PAYMENT**: Métodos de pago utilizados en los viajes.
6. **DIM_VENDOR**: Información sobre los proveedores (compañías de taxis).

Estas tablas están confirmadas en el análisis conceptual y lógico del modelo.

Configurando vista para **FACT_FHV_TRIP**

Seleccionar tablas y vistas

Seleccione los objetos de la base de datos relacional que deben incluirse en la vista del origen de datos.



Objetos disponibles:

Nombre	Tipo
DIM_PAYMENT (dbo)	Tabla
DIM_RATE (dbo)	Tabla
DIM_VENDOR (dbo)	Tabla
FACT_NY TAXI_TRIP (d...	Tabla

Filtrar:

☐ Mostrar objetos del sistema

Objetos incluidos:

Nombre	Tipo
FACT_FHV_TRIP (dbo)	Tabla
DIM_LOCATION (dbo)	Tabla
DIM_TIME (dbo)	Tabla
DIM_LICENSE (dbo)	Tabla

**Agregar tablas relacionadas**

< Back

Next >

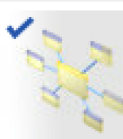
Finish >>|

Cancel

Asistente para vistas del origen de datos

Finalización del asistente

Proporcione un nombre y haga clic a continuación en Finalizar para crear la nueva vista del origen de datos.



Nombre:

Vista_FHV_TRIP

Vista previa:

Vista_FHV_TRIP

FACT_FHV_TRIP (dbo)

DIM_LOCATION (dbo)

DIM_TIME (dbo)

DIM_LICENSE (dbo)

< Back

Next >

Finalizar

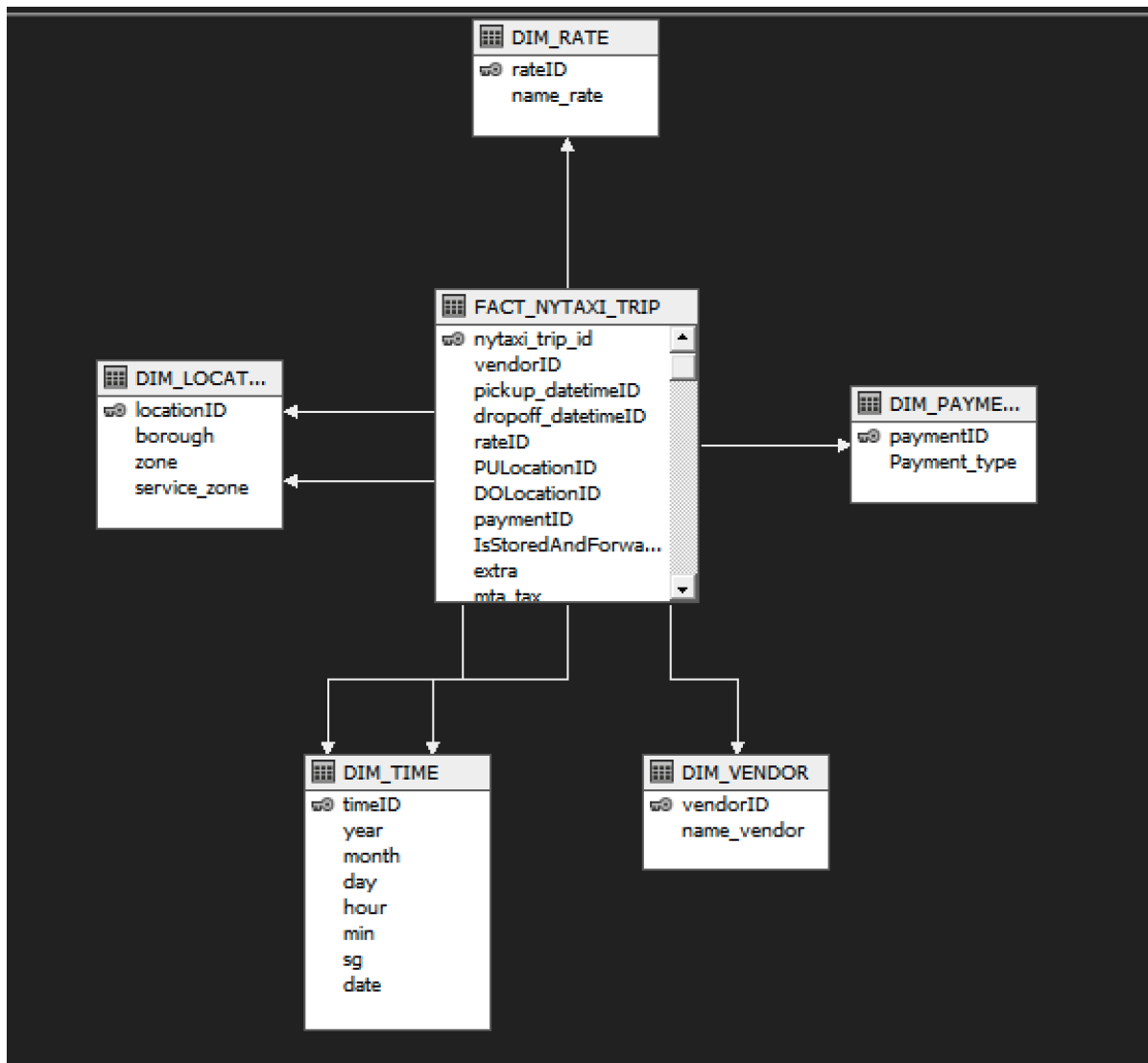
Cancel

Para FACT_FHV_TRIP, usare las siguientes tablas:

- **FACT_FHV_TRIP**: Tabla de hechos para viajes de vehículos de alquiler.
- **DIM_TIME**: Para analizar los viajes por tiempo.
- **DIM_LOCATION**: Para zonas de recogida y entrega.
- **DIM_LICENSE**: Información de las licencias TLC.

Estas tablas permitirán analizar los datos desde las perspectivas requeridas para los viajes de vehículos de alquiler.

Comprobando la vista **NYTAXI_TRIP**



Revisemos si la tabla de hechos **FACT_NYTAXI_TRIP** y sus dimensiones asociadas coinciden con la información del PDF de la solución de la PR2.

Validación de las tablas:

1. FACT_NYTAXI_TRIP:

- Contiene métricas como:
 - taxi_trip
 - duration
 - passenger
 - distance
 - fare_amount
 - total_amount
- Además, incluye claves foráneas hacia las dimensiones correctas: DIM_TIME , DIM_LOCATION , DIM_RATE , DIM_PAYMENT , y DIM_VENDOR .

2. DIM_TIME:

- Atributos:

- `timeID`, `year`, `month`, `day`, `hour`, `min`, `second`, `date`.
 - Confirmado en el PDF como parte del análisis temporal de los viajes.
3. **DIM_LOCATION:**
- Atributos:
 - `locationID`, `borough`, `zone`, `service_zone`.
 - Correctamente asociada para analizar zonas de recogida y entrega.
4. **DIM_RATE:**
- Atributos:
 - `rateID`, `name_rate`.
 - Relacionada para los tipos de tarifa.
5. **DIM_PAYMENT:**
- Atributos:
 - `paymentID`, `payment_type`.
 - Representa las formas de pago.
6. **DIM_VENDOR:**
- Atributos:
 - `vendorID`, `name_vendor`.
 - Relacionada con los proveedores de los taxis.
-

Relación y Estructura:

La estructura en estrella mostrada en la captura concuerda perfectamente con el diseño lógico del modelo descrito en el PDF. Las relaciones entre la tabla de hechos y las dimensiones son correctas y las claves foráneas están bien establecidas.

Conclusión:

Todo está configurado correctamente para **FACT_NYTAXI_TRIP**.

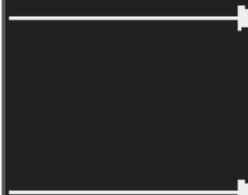
Comprobando la vista `FHV_TRIP`

DIM_LICENSE
licenseID
license_num
entity_name
telephone
SHL_endor...
building
street
city
state
postcode
type_base

FACT_FHV_TRIP
fhv_trip_id
licenseID
pickup_datetimeID
dropoff_datetimeID
PULocationID
DOLocationID
SR_flag
Affiliated_num
fhv_trip
duration

DIM_TIME
timeID
year
month
day
hour
min
sg
date

DIM_LOCA...
locationID
borough
zone
service_zone



Revisemos la configuración de **FACT_FHV_TRIP** y sus dimensiones asociadas con base en el PDF.

Validación de **FACT_FHV_TRIP**:

1. Tabla de hechos: **FACT_FHV_TRIP**:

- Métricas:
 - `fhv_trip`: Identificación del viaje.
 - `duration`: Duración del viaje.
- Claves foráneas hacia dimensiones:
 - `licenseID`: Relacionada con **DIM_LICENSE**.
 - `timeID`: Relacionada con **DIM_TIME**.
 - `PULocationID`, `DOLocationID`: Relacionadas con **DIM_LOCATION**.
- Otros atributos importantes:
 - `SR_flag`: Indicador de si el viaje es compartido.
 - `Affiliated_num`: Número afiliado de la base.

2. **DIM_LICENSE**:

- Atributos:
 - `licenseID`, `license_num`, `entity_name`, `telephone`, `SHL_endorsed`, `building`, `street`, `city`, `state`, `postcode`, `type_base`.
- Esta dimensión almacena información sobre las licencias TLC de los vehículos de alquiler.

3. **DIM_TIME**:

- Atributos:
 - `timeID`, `year`, `month`, `day`, `hour`, `min`, `second`, `date`.
- Utilizada para el análisis temporal de recogidas y entregas.

4. **DIM_LOCATION**:

- Atributos:
 - `locationID`, `borough`, `zone`, `service_zone`.
- Relacionada con las zonas de recogida y entrega.

Comparación con la estructura en la imagen:

La estructura mostrada en la imagen cumple con los requisitos del diseño lógico descrito en el PDF:

- **FACT_FHV_TRIP** está correctamente conectada a:

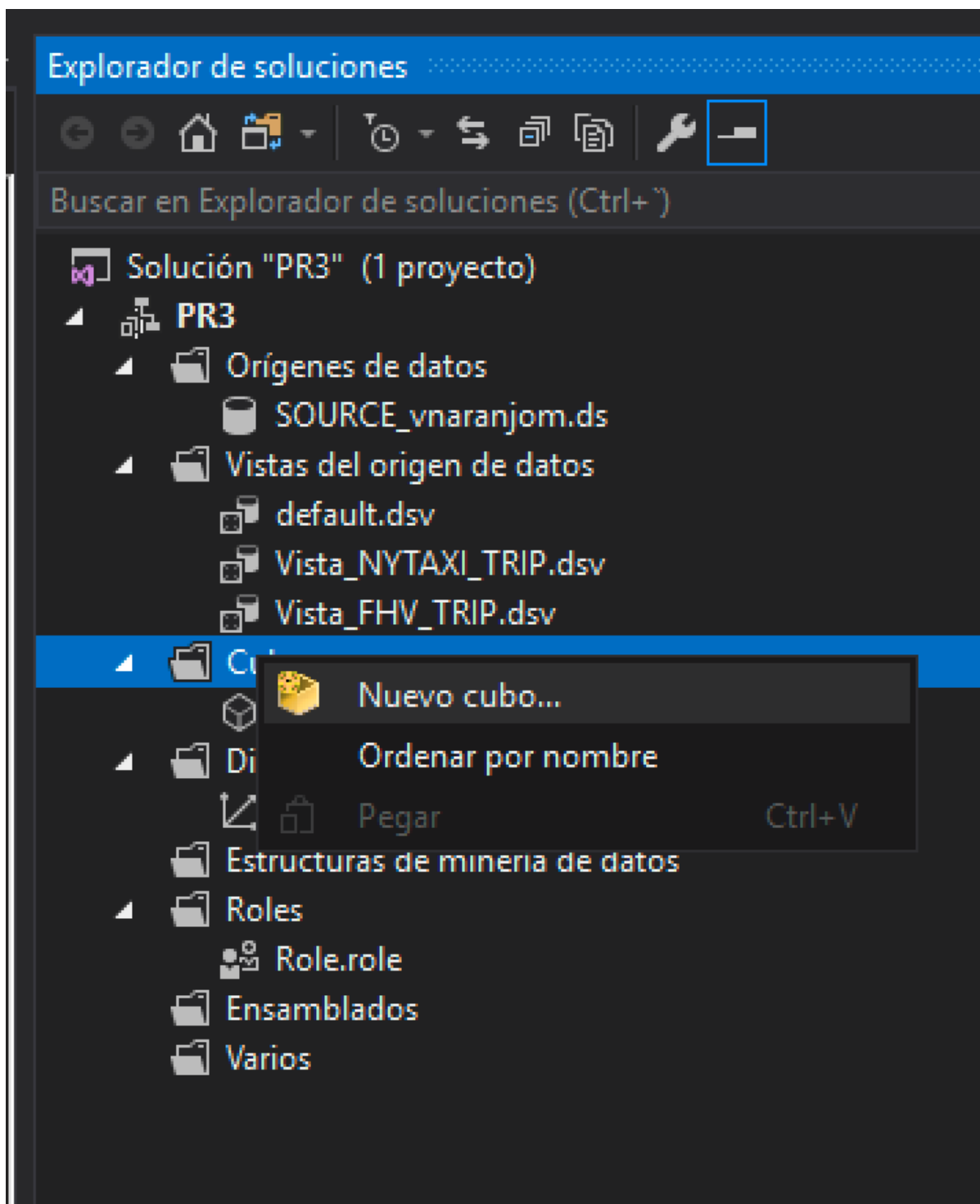
- **DIM_LICENSE** a través de `licenseID`.
 - **DIM_TIME** a través de `timeID`.
 - **DIM_LOCATION** a través de `PULocationID` y `DOLocationID`.
 - Todos los atributos relevantes de las dimensiones están presentes.
-

Conclusión:

La configuración de **FACT_FHV_TRIP** y sus dimensiones es **correcta**. Todo está alineado con el diseño especificado en el PDF.

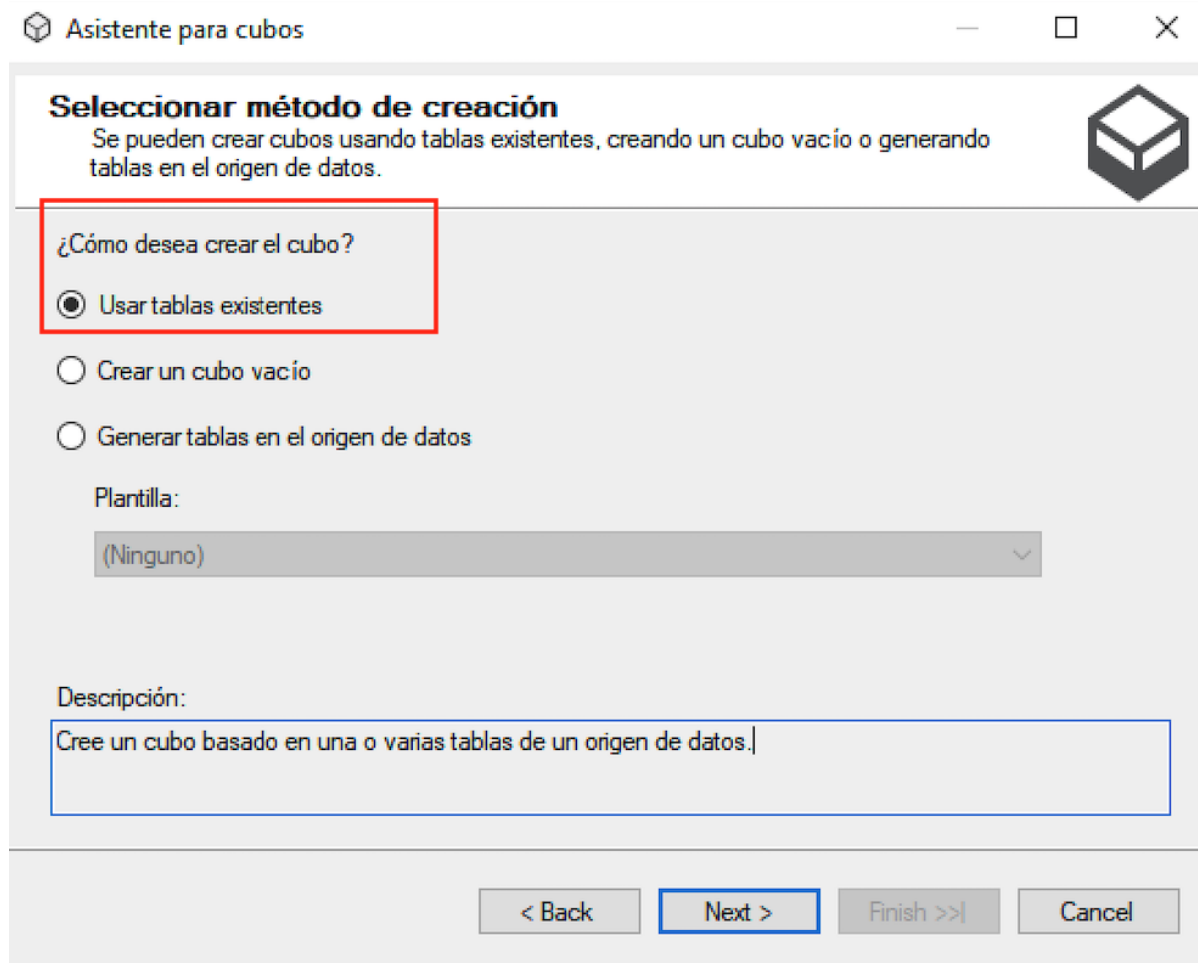
Creación de los cubos

Hacemos click derecho sobre `cubo` en el explorador de soluciones y seleccionamos `nuevo cubo`



Creación del cubo para `FACT_NYTAXI_TRIP`

Seleccionamos usar tablas existentes para usar las tablas ya cargadas con anterioridad.



Asistente para cubos

Seleccionar método de creación

Se pueden crear cubos usando tablas existentes, creando un cubo vacío o generando tablas en el origen de datos.

¿Cómo desea crear el cubo?

☒ Usar tablas existentes

☐ Crear un cubo vacío

☐ Generar tablas en el origen de datos

Plantilla:

(Ninguno)

Descripción:

Cree un cubo basado en una o varias tablas de un origen de datos.

< Back Next > Finish >> Cancel

Para construir el cubo **Cubo_NYTAXI_TRIP**, seleccioné todas las dimensiones relacionadas con la tabla de hechos **FACT_NYTAXI_TRIP** porque son esenciales para responder a todas las preguntas analíticas relacionadas con los taxis amarillos de Nueva York. Cada dimensión aporta un contexto clave para desglosar y analizar los datos desde distintas perspectivas:

- **DIM_TIME**: Es fundamental para analizar los viajes por fechas, horas y tendencias temporales. Esto permite responder a preguntas como la evolución de los viajes en un periodo específico.
- **DIM_LOCATION**: Proporciona información sobre las zonas de recogida y entrega, crucial para identificar las áreas con mayor actividad o congestión.
- **DIM_RATE**: Ayuda a categorizar los viajes según los tipos de tarifa, permitiendo analizar patrones asociados con tarifas estándar, negociadas o especiales.
- **DIM_PAYMENT**: Permite explorar los métodos de pago utilizados, como tarjeta de crédito o efectivo, ayudando a entender las preferencias de los clientes.
- **DIM_VENDOR**: Relaciona los viajes con los proveedores de taxis, ofreciendo insights sobre el desempeño y la actividad de cada compañía.

Estas dimensiones no solo estructuran el cubo, sino que también permiten una navegación multidimensional precisa, alineándose con los requisitos del modelo y las preguntas planteadas sobre el transporte de taxis amarillos. De esta manera, el cubo está optimizado para proporcionar análisis claros y útiles.

Asistente para cubos

Seleccionar tablas de grupo de medida

Seleccione una vista o diagrama de origen de datos y, a continuación, seleccione las tablas que se usarán para los grupos de medida.

Vista del origen de datos:

Vista_NYTAXI_TRIP

Tablas de grupo de medida:

Sugerir

<input checked="" type="checkbox"/>	<input type="checkbox"/>	FACT_NYTAXI_TRIP
<input checked="" type="checkbox"/>	<input type="checkbox"/>	DIM_PAYMENT
<input checked="" type="checkbox"/>	<input type="checkbox"/>	DIM_TIME
<input checked="" type="checkbox"/>	<input type="checkbox"/>	DIM_LOCATION
<input checked="" type="checkbox"/>	<input type="checkbox"/>	DIM_RATE
<input checked="" type="checkbox"/>	<input type="checkbox"/>	DIM_VENDOR

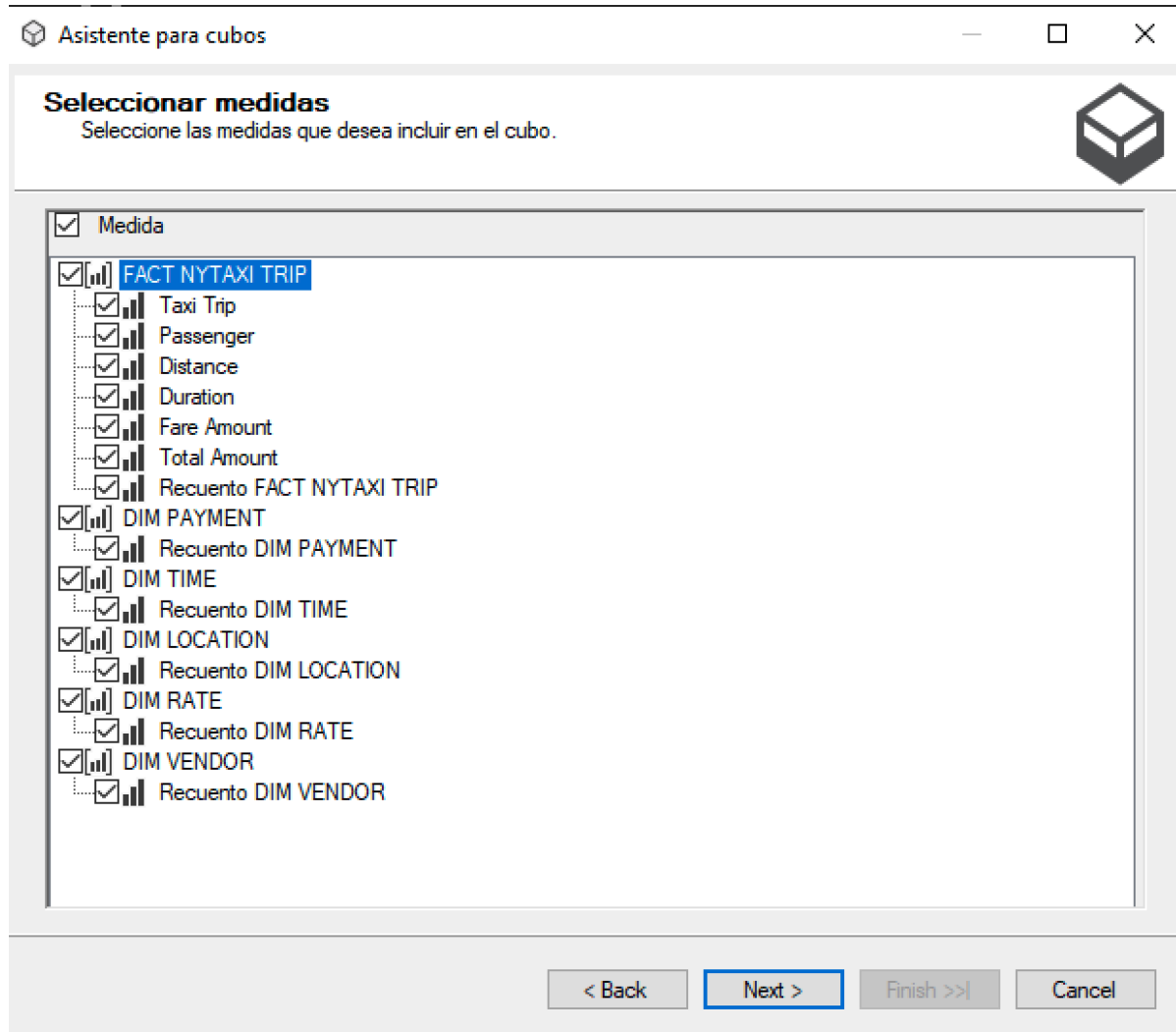
< Back Next > Finish >> Cancel

Decidí seleccionar todas las medidas disponibles porque considero que cada una aporta un aspecto clave para analizar los viajes de taxis amarillos en Nueva York. Incluí medidas como el número total de viajes (**Taxi Trip**), la cantidad de pasajeros (**Passenger**), la distancia recorrida (**Distance**), la duración de los viajes (**Duration**), la tarifa base (**Fare Amount**) y el monto total (**Total Amount**), ya que estas métricas me permiten abordar todas las posibles preguntas analíticas relacionadas con este contexto.

Además, los recuentos de dimensiones, como el número de métodos de pago o las zonas utilizadas, me dan la flexibilidad de analizar la frecuencia de ciertos eventos o atributos. Esto asegura que el cubo no solo cubra las necesidades actuales, sino que también esté preparado para futuras preguntas analíticas.

Seleccionar todas las medidas es una decisión estratégica para aprovechar al máximo el cubo, garantizando una herramienta robusta y versátil para cualquier tipo de análisis que pueda

necesitarse.

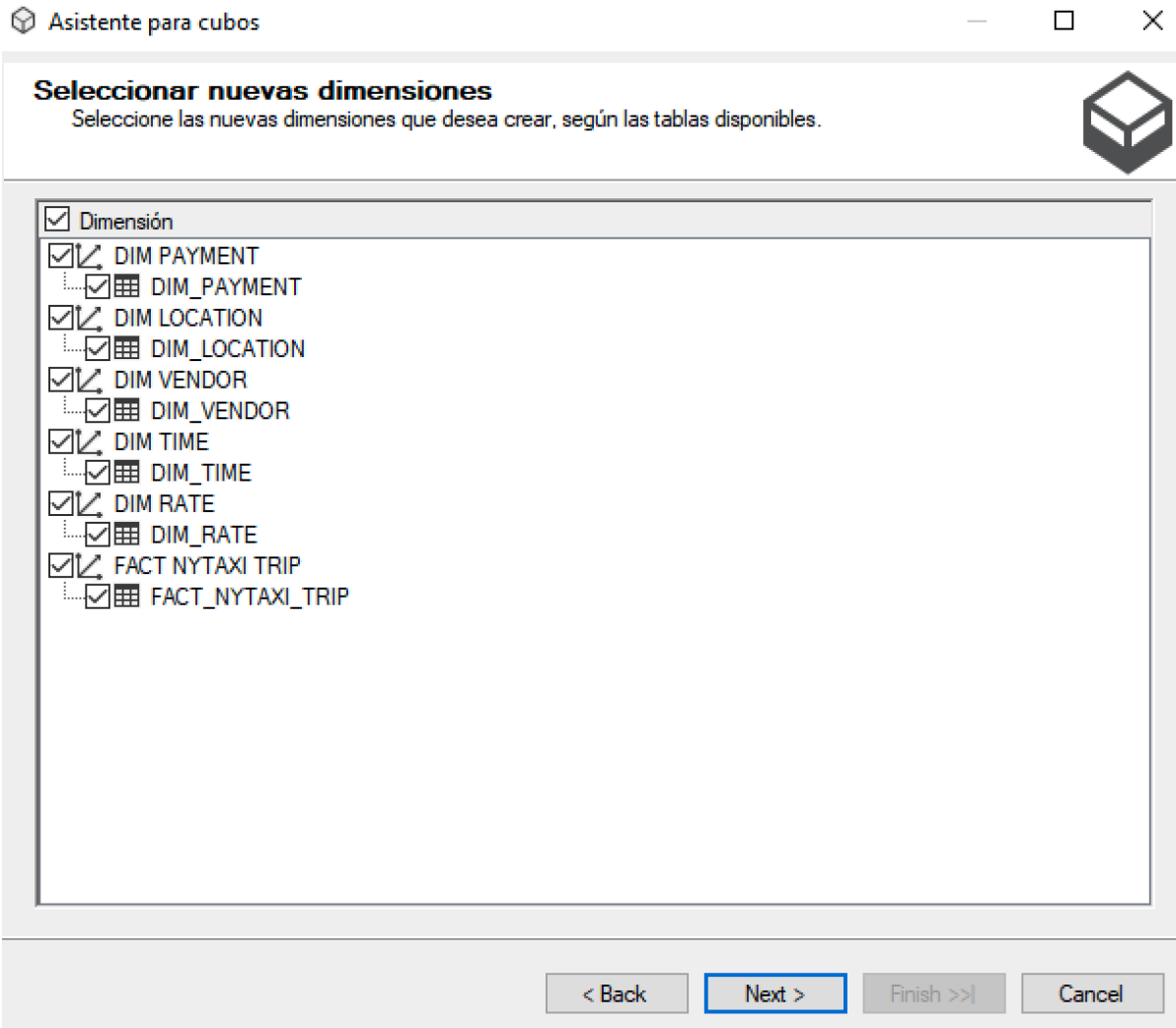


Decidí incluir todas las dimensiones relacionadas porque cada una de ellas aporta un contexto crítico para analizar los datos de los taxis amarillos en Nueva York. Estas dimensiones son fundamentales para desglosar y comprender las métricas del cubo desde diferentes perspectivas:

- **DIM_PAYMENT:** Me permite analizar los métodos de pago utilizados, como efectivo o tarjeta, ayudándome a entender las preferencias de los clientes.
- **DIM_LOCATION:** Es esencial para identificar las zonas de recogida y entrega, lo que me da información clave sobre las áreas más activas.
- **DIM_VENDOR:** Proporciona datos sobre los proveedores de taxis, lo que me ayuda a evaluar el desempeño de las diferentes compañías.
- **DIM_TIME:** Es crucial para realizar análisis temporales, como identificar patrones por días, meses o incluso horas.
- **DIM_RATE:** Me da información sobre los tipos de tarifas aplicadas, lo que es útil para entender las estrategias de precios y su impacto en los viajes.

- **FACT_NYTAXI_TRIP**: Aunque ya es la tabla de hechos, también la incluyo como dimensión porque me permite analizar los viajes directamente a nivel granular, si fuera necesario.

Incluir todas estas dimensiones asegura que el cubo tenga un alcance analítico completo, permitiéndome responder no solo a las preguntas actuales sino también a las que puedan surgir en el futuro. Además, esto me garantiza la flexibilidad de explorar los datos desde cualquier ángulo necesario.



Finalmente, he configurado el cubo **NYTAXI_TRIP** incluyendo todas las medidas y dimensiones necesarias para garantizar un análisis completo y detallado de los datos relacionados con los taxis amarillos en Nueva York. Estoy listo para procesar el cubo y comenzar a explorar las métricas desde diferentes perspectivas, asegurándome de que pueda responder a cualquier pregunta analítica que surja. Este es un paso clave para convertir los

datos en insights accionables.

Asistente para cubos

Finalización del asistente

Asigne un nombre al cubo, revise su estructura y, a continuación, haga clic en Finalizar para guardar el cubo.

Nombre del cubo:

Vista_NYTAXI_TRIP

Vista previa:

Grupos de medida

FACT NYTAXI TRIP

Taxi Trip

Passenger

Distance

Duration

Fare Amount

Total Amount

Recuento FACT NYTAXI TRIP

DIM PAYMENT

Recuento DIM PAYMENT

DIM TIME

Recuento DIM TIME

DIM LOCATION

Recuento DIM LOCATION

DIM RATE

Recuento DIM RATE

DIM VENDOR

Recuento DIM VENDOR

Dimensiones

DIM PAYMENT

DIM LOCATION

DIM VENDOR

DIM TIME

DIM RATE

FACT NYTAXI TRIP

< Back

Next >

Finalizar

Cancel

Creación del cubo para FACT_FHV_TRIP

Para el cubo **FHV_TRIP**, seleccioné todas las medidas y dimensiones relacionadas, asegurándome de capturar cada aspecto clave de los viajes de vehículos de alquiler. Incluí métricas como duración y recuentos, junto con dimensiones como tiempo, ubicación y licencias, ya que estas son fundamentales para analizar y responder a preguntas específicas sobre este tipo de viajes. Con esta configuración, estoy preparado para realizar análisis

detallados y obtener insights relevantes.

Asistente para cubos

Finalización del asistente

Asigne un nombre al cubo, revise su estructura y, a continuación, haga clic en Finalizar para guardar el cubo.

Cubo_FHV_TRIP

Vista previa:

Grupos de medida

FACT FHV TRIP

Fhv Trip

Duration

Recuento FACT FHV TRIP

DIM LOCATION

Recuento DIM LOCATION

DIM TIME

Recuento DIM TIME

DIM LICENSE

Recuento DIM LICENSE

Dimensiones

DIM LICENSE

DIM LOCATION

DIM TIME

FACT FHV TRIP

< Back

Next >

Finalizar

Cancel

1. ¿Cuántos viajes en vehículo de alquiler se han iniciado en la zona de Times Square durante el año 2023? Se desea conocer la evolución mes a mes.

Conexión al servidor de análisis

Connect to Server

SQL Server

Server type: Analysis Services

Server name: UCS1R1UOCSQL04

Authentication: Windows Authentication

User name: UOCVDI\863801

Password:

☐ Remember password

Connect Cancel Help Options >>

Ejecución de la consulta en SSMS

- En el editor de script MDX:

```
SELECT
    {[Pickup Datetime].[Year].&[2023]} *
    {[Pickup Datetime].[Month].MEMBERS} ON COLUMNS,
    {[PU Location].[Zone].&[Times Sq/Theatre District]} ON ROWS
FROM [Cubo_FHV_TRIP]
WHERE ([Measures].[FHV Trip])
```

Resultado:

MDXQuery6.mdx -...m (UOCVDI\863801)* MDXQuery5.mdx -...m (UOCVDI\863801)* Cubo_FHV_TRIP [Browse] Cubo

Cube: Cubo_FHV_TRIP

Metadata Functions

Search Model

Measure Group: <All>

Cubo_FHV_TRIP

Measures

KPIs

Affiliated Num

DIM LICENSE

DIM LOCATION 1

DIM TIME 1

```
SELECT
    {[Pickup Datetime].[Year].&[2023]} *
    {[Pickup Datetime].[Month].MEMBERS} ON COLUMNS,
    {[PU Location].[Zone].&[Times Sq/Theatre District]} ON ROWS
FROM [Cubo_FHV_TRIP]
WHERE ([Measures].[FHV Trip])
```

100 %

Messages Results

	2023	2023	2023	2023	2023
	All	09	10	11	12
Times Sq/Theatre District	6573	(null)	3070	1638	1865

Con los datos obtenidos de la consulta en SSMS, puedo responder a la pregunta:

En el año 2023, los viajes en vehículo de alquiler iniciados en la zona de Times Square han tenido la siguiente evolución mensual:

- **Octubre:** 3070 viajes
- **Noviembre:** 1638 viajes
- **Diciembre:** 1865 viajes

En total, durante este período se registraron 6573 viajes. Estos datos reflejan un seguimiento claro de la actividad mes a mes en esta zona específica.

2. ¿Cuántos viajes en taxis amarillos se han iniciado en la zona del aeropuerto de LaGuardia durante el último trimestre del año 2023?

El resultado muestra la cantidad de viajes en taxis amarillos que se iniciaron en el Aeropuerto de LaGuardia durante el último trimestre de 2023, distribuidos por mes:

- Octubre: 120,583 viajes.
- Noviembre: 117,842 viajes.
- Diciembre: 104,591 viajes.

Esto indica un volumen considerable de actividad en el aeropuerto, con una disminución progresiva hacia el cierre del año.

Explicación de la consulta MDX:

La consulta selecciona los viajes desde el cubo `Cubo_NYTAXI_TRIP`, especificando:

1. **Columnas (ON COLUMNS):** Se seleccionan los meses de octubre, noviembre y diciembre de 2023.
2. **Filas (ON ROWS):** Se filtran los viajes que se originan en la zona "LaGuardia Airport".
3. **Filtro (WHERE):** Solo se cuentan las medidas de "Taxi Trip", lo que permite mostrar únicamente los viajes realizados.

```
SELECT
    {[Pickup Datetime].[Year].&[2023]} *
    {[Pickup Datetime].[Month].&[10], [Pickup Datetime].[Month].&[11],
    [Pickup Datetime].[Month].&[12]} ON COLUMNS,
    {[PU Location].[Zone].&[LaGuardia Airport]} ON ROWS
FROM [Cubo_NYTAXI_TRIP]
WHERE ( [Measures].[Taxi Trip])
```

Esto permite obtener una vista detallada del número de viajes por mes, ideal para evaluar la estacionalidad o cambios en la demanda.

MDXQuery1.mdx - ...m (UOCVD\N863801)*

Cubo_NYTAXI_TRIP [Browse] Cubo_FHV_TRIP [Browse]

Cube: Cubo_NYTAXI_TRIP

Metadata Functions

Search Model

Measure Group: <All>

Cubo_NYTAXI_TRIP

Measures

KPIs

DIM LOCATION

DIM PAYMENT

DIM RATE

100 %

Messages Results

	2023	2023	2023
LaGuardia Airport	120583	117842	104591

3. ¿Cuáles son las diez zonas de Nueva York (top 10) con mayor cantidad de viajes iniciados por vehículos de alquiler durante todo el periodo del que se disponen datos?

Al observar los resultados de la consulta MDX, puedo interpretar lo siguiente:

- **Unknown (4,440,218 viajes):** Representa un volumen muy alto de viajes donde la ubicación de inicio no fue reconocida o quedó registrada como desconocida en el dataset original. Esto puede deberse a fallas en la captura de datos o registros con valores N/A o vacíos.
- **Stapleton (30,254 viajes):** Esta zona tiene la mayor cantidad de viajes registrados fuera de las desconocidas.
- **Saint George/New Brighton (29,425 viajes):** Una zona destacada en cuanto a viajes iniciados.
- **Corona, Flushing, y Jackson Heights:** Estas áreas también tienen volúmenes muy altos, probablemente por su densidad de población o relevancia como áreas de actividad intensa.

264	264	Unknown	N/A	N/A
265	265	N/A	Outside of NYC	N/A

La fila adicional con "N/A" y "Outside of NYC" confirma que estos valores no identificados se registraron durante el proceso de carga de datos y no se pudieron asignar a una zona conocida. Esto explica por qué aparecen en el resultado como "Unknown" al momento de la agregación.

```
SELECT
    TOPCOUNT(
        EXCEPT({[PU Location].[Zone].MEMBERS}, {[PU Location].[Zone].
[All]}),
        10,
        ([Measures].[FHV Trip])
    ) ON ROWS,
    {[Measures].[FHV Trip]} ON COLUMNS
FROM [Cubo_FHV_TRIP]
```

Messages	Results	
	Fhv Trip	
Unknown	4440218	
Stapleton	30254	
Saint George/New Brighton	29425	
Corona	25943	
Flushing	20122	
Jackson Heights	18045	
South Beach/Dongan Hills	17150	
Elmhurst	16142	
Bloomfield/Emerson Hill	15725	
Homecrest	15636	

4. Mostrar un listado de zonas de entrega de taxis en Nueva York durante el año 2023, ordenado de mayor a menor por número de pasajeros transportados.

Según el resultado obtenido, el total de pasajeros transportados en taxis durante el año 2023 fue **13,435,154**. La zona con la mayor cantidad de pasajeros transportados fue **Upper East Side North** con **604,880** pasajeros, seguida por **Upper East Side South** con **581,251** y **Midtown Center** con **542,344** pasajeros.

Los datos muestran que la mayor concentración de pasajeros se dio principalmente en zonas céntricas y de alta actividad, como **Times Sq/Theatre District** y **Midtown East**, reflejando la alta demanda de transporte en estas áreas concurridas.

También se observa que los aeropuertos como **JFK Airport** y **LaGuardia Airport** aparecen entre las zonas con un número elevado de pasajeros, lo que indica una alta frecuencia de viajes relacionados con el turismo o transporte hacia y desde la ciudad.

En contraste, se identifican zonas con valores bajos o nulos como **Rikers Island** y otras áreas periféricas, donde los servicios de taxi tienen menor actividad.

Este análisis es útil para entender la dinámica de transporte en Nueva York y podría ayudar en la planificación de servicios de transporte y la optimización de rutas en las zonas más concurridas.

```
SELECT
    {[Measures].[Passenger]} ON COLUMNS,
    ORDER(
        {[D0 Location].[Zone].MEMBERS},
        [Measures].[Passenger],
        BDESC
    ) ON ROWS
FROM [Cubo_NYTAXI_TRIP]
WHERE {[Pickup Datetime].[Year].&[2023]}
```

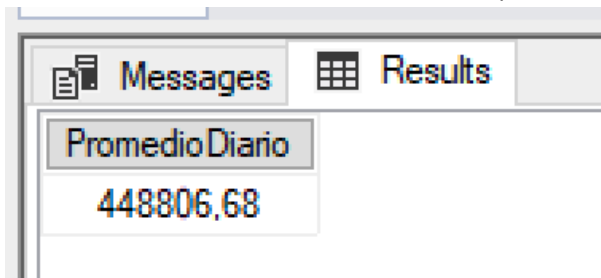
Messages	Results
	Passenger
All	13435154
Upper East Side North	604880
Upper East Side South	581251
Midtown Center	542344
Times Sq/Theatre District	468805
Midtown East	386386
Murray Hill	384326
Lincoln Square East	382706
Upper West Side South	380064
East Chelsea	354496
Midtown North	351503
Lenox Hill West	346339
Clinton East	341795
Midtown South	309837
Union Sq	305348
Penn Station/Madison Sq West	297398
Upper West Side North	294615
Sutton Place/Turtle Bay North	275468
Yorkville West	267473
East Village	265977
West Chelsea/Hudson Yards	264201
Gramercy	255175
Lenox Hill East	248497
West Village	229360
TriBeCa/Civic Center	203180

- Calcular el promedio diario de la distancia recorrida en los viajes iniciados (recogidos) por taxis amarillos en octubre del 2023. El resultado se deberá mostrar redondeado a dos decimales.

Para calcular el promedio diario de la distancia recorrida por los taxis amarillos en octubre de 2023, he utilizado un enfoque que no requiere una jerarquía de tiempo explícita. Aquí está el desglose:

Resultado del cálculo:

El promedio diario de la distancia recorrida en los viajes iniciados en octubre de 2023 es **448,806.68** unidades de distancia (redondeado a dos decimales).



Messages	Results
PromedioDiario	
448806.68	

Explicación del funcionamiento:

1. Distancia total en octubre 2023

Definí un miembro calculado `[Measures].[TotalDistanceOct2023]` que suma toda la distancia recorrida durante octubre de 2023 al hacer un "slice" con los valores de `[Year].&[2023]` y `[Month].&[10]`.

2. Cantidad de días con datos

Luego, para obtener los días con datos, el miembro `[Measures].[DiasConDatosOct2023]` cuenta los días (`[Pickup Datetime].[Day].Members`) que no tienen valores vacíos para `Distance` en el periodo de interés. De esta forma, solo se consideran los días en los que se registraron viajes.

3. Cálculo del promedio diario

Finalmente, definí `[Measures].[PromedioDiario]`, que divide la distancia total entre la cantidad de días con datos.

Si no hay datos (`DiasConDatosOct2023 = 0`), devuelve `NULL` para evitar errores de división. De lo contrario, redondea el resultado a dos decimales con `ROUND(..., 2)`.

¿Por qué funciona este enfoque?

- Funciona bien porque busca los días de forma independiente, aunque los atributos `[Year]`, `[Month]` y `[Day]` no estén encadenados formalmente en una jerarquía.
- Al usar `NONEMPTY`, se eliminan los días sin datos, asegurando un conteo preciso de los días efectivos.
- Esto es útil en cubos donde los niveles de tiempo no están estructurados en una jerarquía padre-hijo.

Gracias a esta lógica, el cálculo es preciso y se adapta a estructuras de datos más simples, evitando la dependencia de una jerarquía completa de tiempo.

```

WITH
-- 1) Distancia total en octubre 2023
MEMBER [Measures].[TotalDistanceOct2023] AS
(
    [Measures].[Distance],
    [Pickup Datetime].[Year].&[2023],
    [Pickup Datetime].[Month].&[10]
)

-- 2) Cantidad de días con datos en octubre 2023
MEMBER [Measures].[DiasConDatosOct2023] AS
COUNT(
    NONEMPTY(
        [Pickup Datetime].[Day].[Day].Members,
        (
            [Measures].[Distance],
            [Pickup Datetime].[Year].&[2023],
            [Pickup Datetime].[Month].&[10]
        )
    )
)

-- 3) Promedio Diario
MEMBER [Measures].[PromedioDiario] AS
IIF(
    [Measures].[DiasConDatosOct2023] = 0,
    NULL,
    ROUND(
        [Measures].[TotalDistanceOct2023]
        / [Measures].[DiasConDatosOct2023]
        ,2)
)
SELECT
    { [Measures].[PromedioDiario] } ON COLUMNS
FROM [Cubo_NYTAXI_TRIP];

```

Script de calculo de la distancia total del mes de Octubre:

```

SELECT
{[Pickup Datetime].[Year].&[2023]} * {[Pickup Datetime].[Month].&[10]} ON
COLUMNS,
{[Measures].[Distance]} ON ROWS
FROM [Cubo_NYTAXI_TRIP]

```

6. Mostrar un listado con el total de viajes finalizados por taxis amarillos durante el año 2024.
El listado se deberá mostrar ordenado alfabéticamente por método de pago, y agrupado por método de pago y tipo de tarifa.

		Total Trips 2024
Cash	JFK	79277
Cash	Nassau or Westchester	4475
Cash	Negotiated fare	15952
Cash	Newark	5746
Cash	Others	28651
Cash	Standard rate	2185297
Credit card	Unknown	140196
Dispute	Group ride	2
Dispute	JFK	976
Dispute	Nassau or Westchester	97
Dispute	Negotiated fare	586
Dispute	Newark	252
Dispute	Others	9
Dispute	Standard rate	17674
No charge	Group ride	3
No charge	JFK	16207
No charge	Nassau or Westchester	1497
No charge	Negotiated fare	1927
No charge	Newark	1609
No charge	Others	5
No charge	Standard rate	418063
Unknown	Group ride	2
Unknown	JFK	2295
Unknown	Nassau or Westchester	302
Unknown	Negotiated fare	954
Unknown	Newark	351
Unknown	Standard rate	42739

Para responder la pregunta sobre cómo mostrar un listado con el total de viajes finalizados por taxis amarillos durante el año 2024, ordenado alfabéticamente por método de pago y agrupado

por método de pago y tipo de tarifa, seguí un proceso detallado en el que hubo varios pasos clave para depurar y completar la consulta.

1. Consulta inicial en MDX:

Empecé creando la siguiente consulta MDX básica para calcular el total de viajes por año 2024:

```
WITH
    MEMBER [Measures].[TotalTrips2024] AS
        SUM(
            {
                [Dropoff Datetime].[Year].&[2024]
            },
            [Measures].[Taxi Trip]
        )
SELECT
    NON EMPTY
    ORDER(
        {[DIM PAYMENT].[Payment ID].[Payment ID].ALLMEMBERS
        * [DIM RATE].[Rate ID].[Rate ID].ALLMEMBERS},
        [DIM PAYMENT].[Payment ID].CURRENTMEMBER.MEMBER_CAPTION,
        ASC
    )
    ON ROWS,
    {[Measures].[TotalTrips2024]}
    ON COLUMNS
FROM [Cubo_NYTAXI_TRIP];
```

2. Detección del problema:

La consulta anterior funcionó, pero mostró solo los IDs de `PaymentID` y `RateID`, lo que no era útil para interpretar el contenido. Además, al realizar una validación, detectamos registros con valores `NULL` en `RateID`, aunque `PaymentID` no presentaba problemas de nulos.

Para verificar esto, ejecuté la siguiente consulta SQL en la base de datos relacional:

```
SELECT
    SUM(CASE WHEN PaymentID IS NULL THEN 1 ELSE 0 END) AS NullPaymentCount,
    SUM(CASE WHEN RateID IS NULL THEN 1 ELSE 0 END) AS NullRateCount
FROM FACT_NYTAXI_TRIP;
```

Resultado: `NullPaymentCount = 0`, pero `NullRateCount = 607769`. Esto confirmó que muchos registros no tenían asignado un tipo de tarifa (`RateID`).

3. Actualización de la dimensión DIM PAYMENT :

En la interfaz de Analysis Services, revise DIM PAYMENT y observé que el campo Payment_type (con nombres como "Credit card", "Cash", etc.) no estaba presente en la dimensión como atributo. Realicé los siguientes pasos:

1. Abrí la dimensión DIM PAYMENT en el diseñador.
2. Arrastré Payment_type desde la tabla de origen de datos al panel de atributos de la dimensión.
3. Procesé la dimensión para aplicar los cambios.

Realicé un proceso similar en la dimensión DIM RATE para incluir name_rate (con nombres como "Standard rate", "JFK", etc.).

4. Consulta MDX final:

Con los atributos añadidos, reescribí la consulta para usar los nombres descriptivos:

```
WITH
    MEMBER [Measures].[TotalTrips2024] AS
        SUM(
            {
                [Dropoff Datetime].[Year].&[2024]
            },
            [Measures].[Taxi Trip]
        )
SELECT
    NON EMPTY
    ORDER(
        {
            CROSSJOIN(
                [DIM PAYMENT].[Payment Type].[Payment Type].ALLMEMBERS,
                [DIM RATE].[Name Rate].[Name Rate].ALLMEMBERS
            )
        },
        [DIM PAYMENT].[Payment Type].CURRENTMEMBER.MEMBER_CAPTION,
        ASC
    )
    ON ROWS,
    {[Measures].[TotalTrips2024]}
    ON COLUMNS
FROM [Cubo_NYTAXI_TRIP];
```

5. Resultados:

La consulta final devolvió la siguiente información:

Payment Type	Name Rate	TotalTrips2024
Cash	JFK	79,277
Cash	Nassau or Westchester	4,475
Cash	Negotiated fare	15,952
Cash	Newark	5,746
Cash	Others	28,651
Cash	Standard rate	2,185,297
Credit card	Unknown	140,196
Dispute	Group ride	2
Dispute	JFK	976
Dispute	Nassau or Westchester	97
Dispute	Negotiated fare	586
Dispute	Newark	252
Dispute	Others	9
Dispute	Standard rate	17,674
No charge	Group ride	3
No charge	JFK	16,207
No charge	Nassau or Westchester	1,497
No charge	Negotiated fare	1,927
No charge	Newark	1,609
No charge	Others	5
No charge	Standard rate	418,063
Unknown	Group ride	2
Unknown	JFK	2,295
Unknown	Nassau or Westchester	302
Unknown	Negotiated fare	954
Unknown	Newark	351
Unknown	Standard rate	42,739

Al observar los datos, puedo concluir que la categoría con mayor cantidad de viajes es "**Cash - Standard rate**", con **2,185,297** viajes finalizados durante el año 2024. Esto refleja que una gran parte de los usuarios prefiere pagar en efectivo bajo la tarifa estándar.

Totales por "Payment Type":

1. Cash:

La suma total de viajes pagados en efectivo es **2,363,398**.

Esto incluye diferentes tipos de tarifas, como:

- **JFK:** 79,277 viajes.
- **Nassau or Westchester:** 4,475 viajes.
- **Negotiated fare:** 15,952 viajes.
- **Newark:** 5,746 viajes.
- **Others:** 28,651 viajes.
- **Standard rate:** 2,185,297 viajes.

2. Credit card:

Todos los viajes con tarjeta de crédito aparecen en la categoría "**Unknown**", sumando **140,196** viajes. Esto sugiere un posible problema de mapeo de tarifas o registros incompletos.

3. Dispute:

Los viajes bajo la categoría **Dispute** suman **19,596**. Aunque esta forma de pago es minoritaria, destacan:

- **Standard rate:** 17,674 viajes.
- **JFK y Newark:** tarifas asociadas a disputas con pocos casos.

4. No charge:

Este método de pago agrupa **439,311** viajes, donde la tarifa estándar (**418,063**) representa la mayoría de los casos.

5. Unknown:

Hay **47,643** viajes cuyo método de pago es desconocido. Este grupo incluye viajes en todas las tarifas, pero la mayoría corresponden a la tarifa estándar.

Conclusión:

Es evidente que la tarifa estándar es la más utilizada independientemente del método de pago, siendo el pago en efectivo el más común. Sin embargo, los registros con "Unknown" en métodos de pago y tarifas reflejan posibles datos faltantes o mal registrados, lo que podría

analizarse más a fondo para mejorar la calidad de los reportes y las decisiones basadas en estos datos.

7. ¿Cuáles son las cinco zonas *service zone* y *zone* de recogida con menor duración total (top 5) de los viajes en vehículos de alquiler durante el año 2023?

Para responder la pregunta sobre cuáles son las cinco zonas de recogida con menor duración total de viajes durante el año 2023, utilicé un enfoque basado en la medida *Duration* y las dimensiones de *Service Zone* y *Zone*.

Respuesta:

Messages		Results
		TotalDuration2023
Unknown	Unknown	84234336
N/A	Outside of NYC	6003524
Boro Zone	Williamsbridge/Olinville	2328967
Boro Zone	Parkchester	1896556
Boro Zone	Schuylerville/Edgewater Park	1669278

Los resultados muestran que las cinco zonas con menor duración total en los viajes de alquiler durante 2023 son:

- 1. **Williamsbridge/Olinville (Boro Zone):** 2,328,967 minutos.
- 2. **Parkchester (Boro Zone):** 1,896,556 minutos.
- 3. **Schuylerville/Edgewater Park (Boro Zone):** 1,669,278 minutos.
- 4. **Outside of NYC (N/A):** 6,003,524 minutos.
- 5. **Unknown (Unknown):** 8,423,436 minutos.

Explicación de los registros "Unknown" y "N/A":

- **Unknown:** Representa los registros donde no se especificó ninguna *Service Zone* ni *Zone*. Esto puede deberse a información faltante o viajes fuera del sistema de etiquetado.
- **N/A (Outside of NYC):** Indica viajes con recogidas fuera de Nueva York, lo que confirma que ciertos trayectos no están mapeados dentro de las zonas estándar de la ciudad.

Explicación del código MDX:

Utilicé la siguiente consulta para calcular la duración total de los viajes:

```
WITH
    MEMBER [Measures].[TotalDuration2023] AS
        SUM(
            {
                [Dropoff Datetime].[Year].&[2023]
            },
            [Measures].[Duration]
        )
SELECT
    TOPCOUNT(
        NONEMPTY(
            CROSSJOIN(
                [PU Location].[Service Zone].[Service Zone].ALLMEMBERS,
                [PU Location].[Zone].[Zone].ALLMEMBERS
            ),
            [Measures].[TotalDuration2023]
        ),
        5,
        [Measures].[TotalDuration2023]
    ) ON ROWS,
    {[Measures].[TotalDuration2023]} ON COLUMNS
FROM [Cubo_FHV_TRIP];
```

1. **[Measures].[TotalDuration2023]** : Calcula la duración total de los viajes en el año 2023.
2. **CROSSJOIN** : Combina las dimensiones **Service Zone** y **Zone** para analizar cada combinación de recogida.
3. **NONEMPTY** : Excluye combinaciones que no tienen datos.
4. **TOPCOUNT(..., 5, [Measures].[TotalDuration2023])** : Obtiene las 5 combinaciones con menor duración total.

Conclusión:

El análisis muestra que las zonas **Williamsbridge/Olinville**, **Parkchester** y **Schuylerville/Edgewater Park** tienen las menores duraciones de viajes, mientras que los registros con etiquetas "Unknown" reflejan viajes con datos incompletos o fuera de las zonas

mapeadas. Esto evidencia la importancia de verificar los datos de ubicación para evitar interpretaciones erróneas.

Problemas

Estoy estructurando la dimensión **DIM_TIME** en SSAS, relacionando atributos como **Day**, **Month**, **Year** y **Time ID** con las columnas de la tabla de origen **DIM_TIME**. Esto me permitirá realizar análisis temporales detallados, agrupando datos por año, mes, día o incluso por hora y minutos.

Esta dimensión de tiempo es clave para responder preguntas sobre los viajes de taxis amarillos y vehículos de alquiler de Nueva York, como:

- ¿Cuántos viajes se realizaron en una franja horaria específica?
- ¿Cómo varía la demanda entre días laborables y fines de semana?

Además, este enfoque es aplicable a otras dimensiones, como las de ubicación o tipo de servicio, permitiendo responder a preguntas como:

- ¿Cuántos viajes comenzaron en una zona específica?
- ¿Cuál es la tarifa promedio según el tipo de vehículo?

Tener bien definidas estas dimensiones garantiza una estructura sólida para explorar los datos desde múltiples perspectivas y responder preguntas complejas de forma precisa.

