

CAMADA
16175

Curso: Data Analytics

PROYECTO FINAL

EMPRESA DE TAXIS



INTEGRANTES

MATIAS GUTMAN
MATIAS BASSO

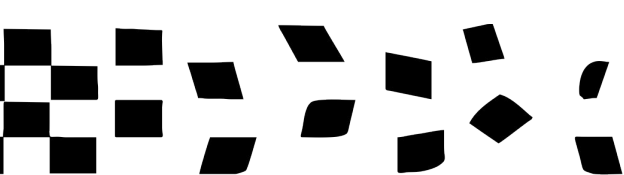
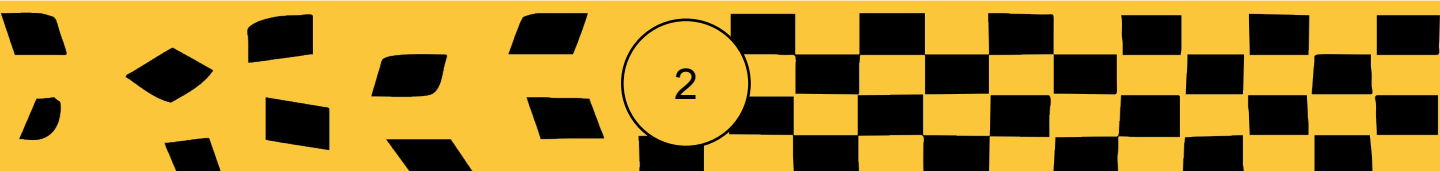
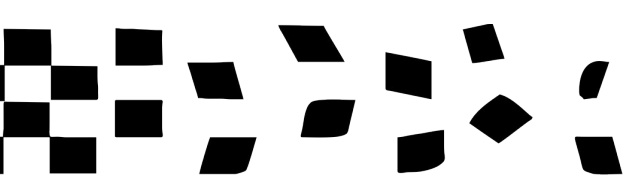


Tabla de contenido:

<u>Portada</u>	<u>PAG. 1</u>
<u>Tabla de contenido</u>	<u>PAG. 2</u>
<u>Acceso a base de datos y archivos Power BI</u>	<u>PAG. 3</u>
<u>Manual de marca</u>	<u>PAG. 4</u>
<u>Introducción</u>	
<u>Información de la empresa</u>	<u>PAG 5</u>
<u>Evolución del proyecto y base de datos utilizada</u>	
<u>Objetivo y alcance del proyecto</u>	
<u>Usuario final</u>	<u>PAG. 6</u>
<u>Diagrama entidad relación</u>	<u>PAG. 7</u>
<u>Descripción de tablas</u>	<u>PAG.8 a 11</u>
<u>Creación de tablas en Power BI</u>	<u>PAG. 11 a 12</u>
<u>Transformaciones realizadas</u>	<u>PAG. 13 a 14</u>
<u>Medidas calculadas</u>	<u>PAG. 15 a 16</u>
<u>Futuras líneas</u>	<u>PAG. 17</u>





Acceso base de datos y archivos Power Bi

Para obtener la base de datos se pueden descargar de la siguiente pagina:

<https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

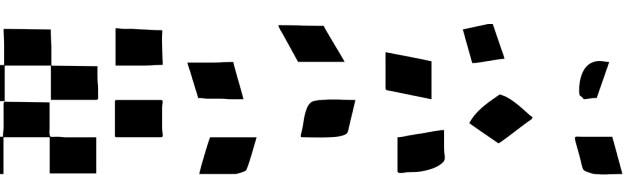
Para fines mas prácticos se cargo una carpeta con la base de datos en formato csv en mega, los cuales se podrán descargar de los siguientes link:

Link de base de datos: <https://mega.nz/folder/jphS2YCZ>

Contraseña de acceso: KH9AqMpCo9IX02x5_NraOQ

El archivo de power bi también se encuentra en mega ya que la plataforma no nos deja subirlo por el tamaño del archivo. El mismo se encuentra con el nombre de

Una vez descargado los archivos se procede a la carga de datos por carpeta o por archivo individual, en nuestro caso por una cuestión de capacidad de las maquinas se realizo tabla por tabla.



NYC TAXI

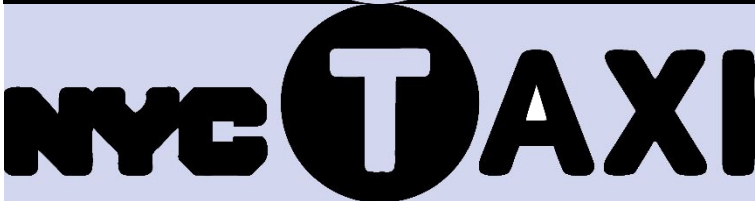
Manual de marca:

La tipografía a utilizar fue Arial, ya que se considero como una tipografía clara y elegante para un modelo empresarial. En el titulo principal del cada solapa esta utilizada en tamaño 18 y en el resto de títulos en tamaño 14.

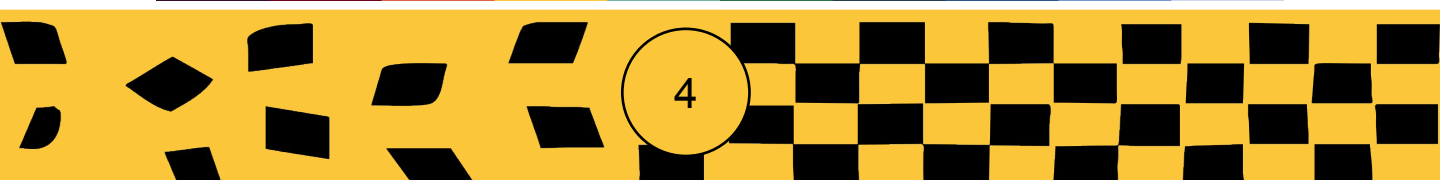
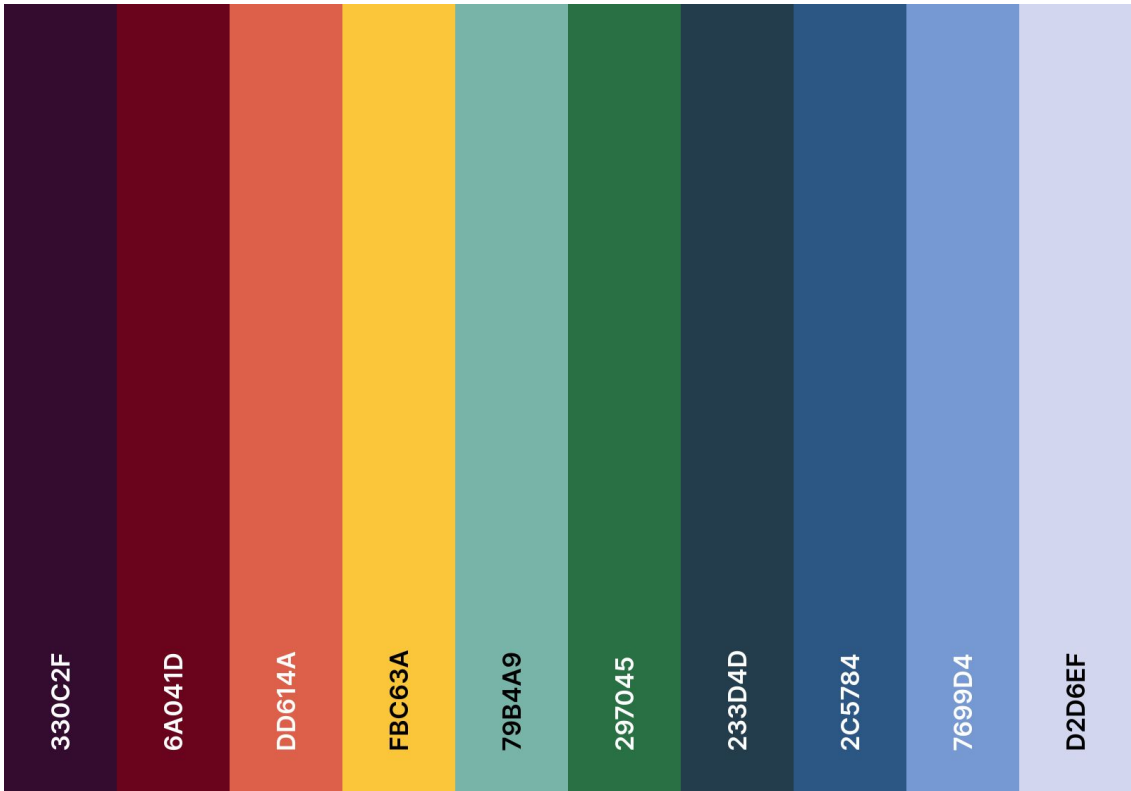
Logo Negativo

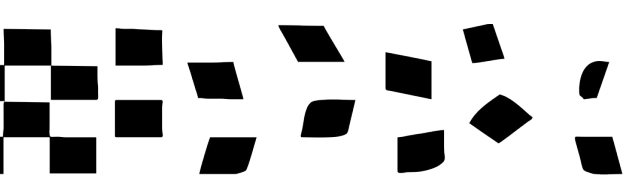


Logo Positivo



Se utilizo el siguiente esquema de colores, generado a partir del color amarillo que era el único con el que contábamos, y se busco colores apagados para no saturar la visión del usuario.





Introducción:

El objetivo del proyecto final es implementar y relacionar todo lo aprendido de forma teórica en la cursada de data analytics.

Para esto se desarrolla un dashboard en Power Bi aplicando todas las herramientas provistas por el profesor y tutor.

Información de la empresa:

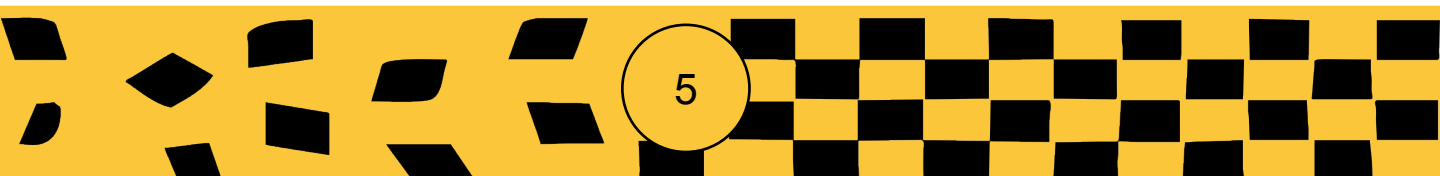
La empresa NYC TAXI es ficticia creada para los fines de este trabajo, lo que si es real es la base de datos que proviene de los movimiento de los taxis con licencia de la ciudad de Nueva York. Al mismo tiempo se uso el logo de la pagina donde se sacaron las bases de datos como logotipo de la empresa.

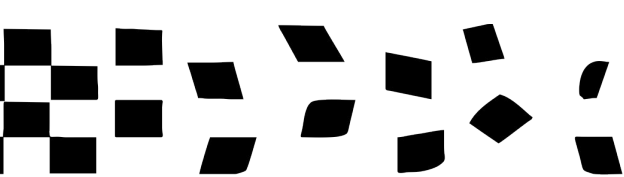
Evolución del proyecto y base de datos utilizada

Al principio se pensó en utilizar las cuatro base de datos que da la ciudad de Nueva York, las misma son Taxi Amarillos, Taxi Verdes, Vehículos de aplicaciones (Uber, etc.) y Vehículos de Alquiler.

Debido a que las bases de datos de los vehículos de alquiler y aplicaciones estaban vacías se opto por quedarnos con las de taxis amarillos y taxis verdes. Al empezar a utilizar las bases de datos se observo una gran disparidad entre los datos con lo cual investigando nos encontramos que ninguno de los dos tipos de taxis ejercía su actividad en la misma zona y uno tenia un publico cautivo en los aeropuertos. Con lo cual se decidió hacer la comparativa de los distintos meses de los taxis amarillos para generar un análisis mas fluido y con sentido.

Además de lo señalado anteriormente se recorto la base de datos a la mitad para poder trabajar con las maquinas disponibles.





Objetivo y alcance del proyecto

El alcance del proyecto es desarrollar un dashboard que permita tener de forma visual los principales indicadores y métricas que la empresa necesita para desarrollar su actividad de forma eficiente.

Los siguientes objetivos serán las variables desarrolladas en el dashboard:

TAXI Montos totales por periodos de mes.

TAXI Cantidad de pasajeros por periodos de mes

TAXI de barrios con mayor frecuencia de viajes

TAXI saturación mensual

TAXI comparativa de clientes diarios en los distintos meses.

TAXI al que atrae mas clientes.

TAXI al de pagos.

Usuario final

El proyecto esta generado en función de un usuario interno dentro de la empresa encargado de seguir las métricas para realizar los cambios pertinentes que le permitan a la organización cumplir con sus metas mensuales, las cuales repercutirán en el balance general de la empresa.

Debemos destacar que el tablero esta pensado de una forma estratégica y no operativa ya que lo que se podrá apreciar es la comparativa de los distintos meses y aplicar cambios para los meses siguientes.

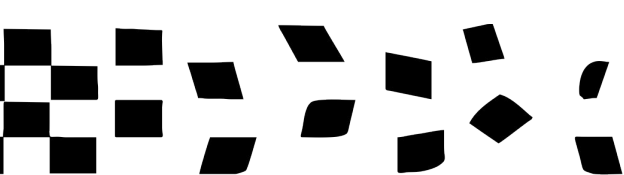
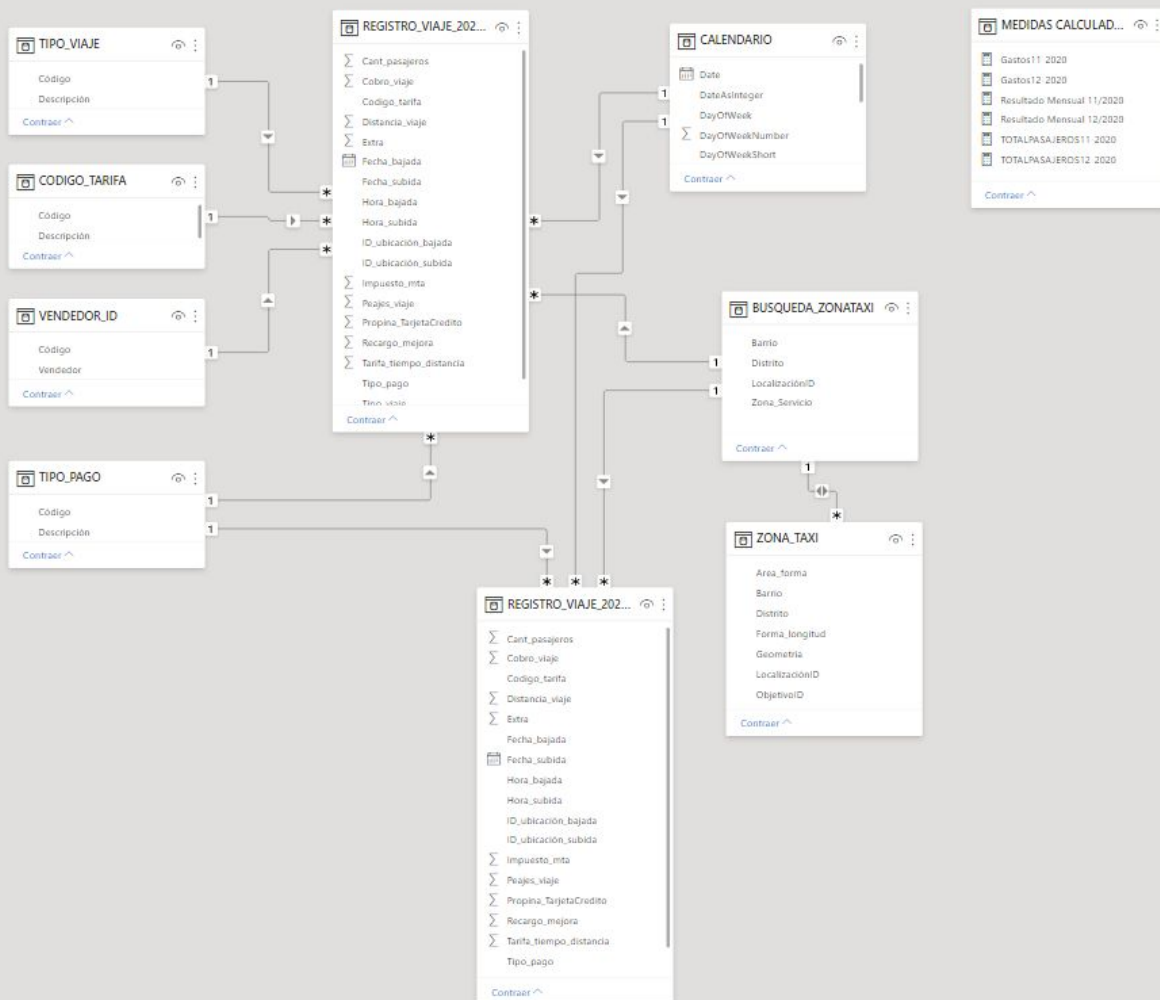
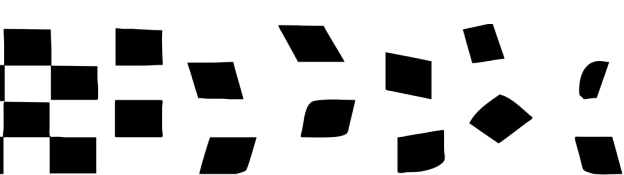


Diagrama entidad relación:





Descripción de tablas :





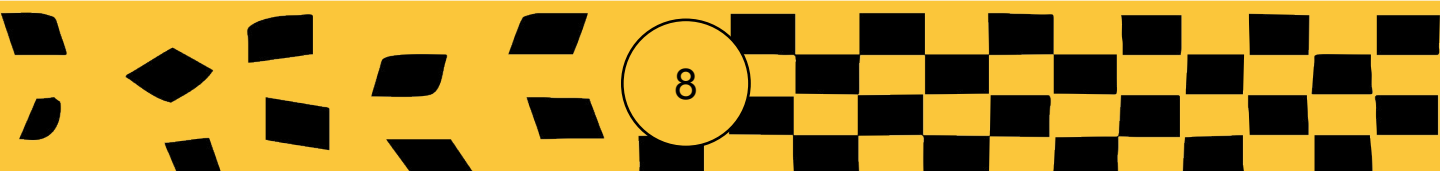
Clave Primaria

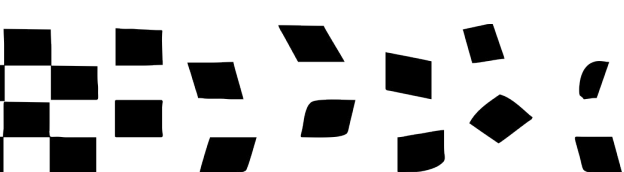


Clave Foránea



yellow_tripdata_2020-12 = REGISTRO_VIAJE_2020-12

Nombre del campo	Descripción	Tipo de dato	Traducción de columnas
VendorID	Un código que indica el proveedor de TPEP que proporcionó el registro. 1 = Creative Mobile Technologies, LLC; 2 = VeriFone Inc.	Int	VendedorID
tp Pickup_datetime	La fecha y hora en que se activó el medidor.	Timestamp	Fecha_hora_subida
tp Dropoff_datetime	La fecha y hora en que se desactivó el medidor.	Timestamp	Fecha_hora_bajada
Passenger_count	El número de pasajeros en el vehículo. Este es un valor ingresado por el conductor.	Int	Cant_pasajeros
Trip_distance	La distancia de viaje transcurrida en millas informada por el taxímetro.	Decimal	Distancia_viaje
PULocationID 	Zona de taxis de TLC en la que se utilizó el taxímetro.	Int	ID_ubicación_subida
DOLocationID 	Zona de taxis de TLC en la que se desactivó el taxímetro.	Int	ID_ubicación_bajada
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	Int	Codigo_tarifa
Store_and_fwd_flag	Esta bandera indica si el registro de viaje se mantuvo en el vehículo. Memoria antes de enviar al proveedor, también conocido como "almacenar y reenviar", porque el vehículo no tenía conexión al servidor. Y = viaje de almacenamiento y envío N = no es una tienda y un viaje hacia adelante	Text	Tipo_viaje
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	Int	Tipo_pago
Fare_amount	La tarifa de tiempo y distancia calculada por el medidor.	Decimal	Tarifa_tiempo_distancia
Extra	Extras y recargos varios. Actualmente, esto solo incluye las horas pico de \$ 0.50 y \$ 1 y los cargos nocturnos.	Decimal	Extra
MTA_tax	Impuesto de \$ 0.50 MTA que se activa automáticamente según el medidor tasa en uso.	Decimal	Impuesto_mta
Improvement_surcharge	Recargo de mejora de \$ 0.30 para viajes evaluados en la caída de la bandera.	Decimal	Recargo_mejora
Tip_amount	Importe de la propina: este campo se completa automáticamente para la tarjeta de crédito. Consejos: Las propinas en efectivo no están incluidas.	Decimal	Propina_TarjetaCredito
Tolls_amount	Importe total de todos los peajes pagados en el viaje.	Decimal	Peajes_viaje
Total_amount	El monto total cobrado a los pasajeros. No incluye propinas en efectivo.	Decimal	Cobro_viaje

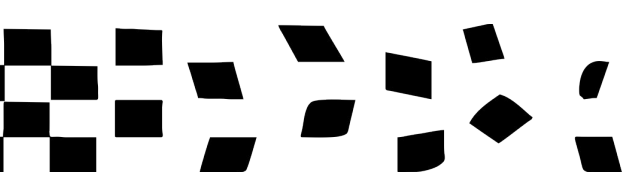






yellow_tripdata_2020-11 = REGISTRO_VIAJE_2020-11

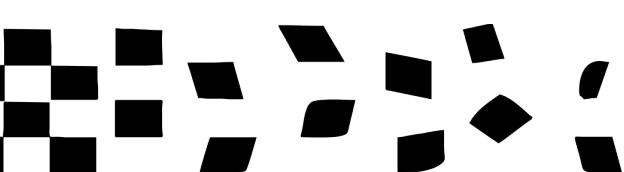
Nombre del campo	Descripción	Tipo de dato	Traducción de columnas
VendorID	Un código que indica el proveedor de TPEP que proporcionó el registro. 1 = Creative Mobile Technologies, LLC; 2 = VeriFone Inc.	Int	VendedorID
tpep_pickup_datetime	La fecha y hora en que se activó el medidor.	Timestamp	Fecha_hora_subida
tpep_dropoff_datetime	La fecha y hora en que se desactivó el medidor.	Timestamp	Fecha_hora_bajada
Passenger_count	El número de pasajeros en el vehículo. Este es un valor ingresado por el conductor.	Int	Cant_pasajeros
Trip_distance	La distancia de viaje transcurrida en millas informada por el taxímetro.	Decimal	Distancia_viaje
PULocationID 	Zona de taxis de TLC en la que se utilizó el taxímetro.	Int	ID_ubicación_subida
DOLocationID 	Zona de taxis de TLC en la que se desactivó el taxímetro.	Int	ID_ubicación_bajada
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	Int	Codigo_tarifa
Store_and_fwd_flag	Esta bandera indica si el registro de viaje se mantuvo en el vehículo. Memoria antes de enviar al proveedor, también conocido como "almacenar y reenviar", porque el vehículo no tenía conexión al servidor. Y = viaje de almacenamiento y envío N = no es una tienda y un viaje hacia adelante	Text	Tipo_viaje
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	Int	Tipo_pago
Fare_amount	La tarifa de tiempo y distancia calculada por el medidor.	Decimal	Tarifa_tiempo_distancia
Extra	Extras y recargos varios. Actualmente, esto solo incluye las horas pico de \$ 0.50 y \$ 1 y los cargos nocturnos.	Decimal	Extra
MTA_tax	Impuesto de \$ 0.50 MTA que se activa automáticamente según el medidor tasa en uso.	Decimal	Impuesto_mta
Improvement_surcharge	Recargo de mejora de \$ 0.30 para viajes evaluados en la caída de la bandera.	Decimal	Recargo_mejora
Tip_amount	Importe de la propina: este campo se completa automáticamente para la tarjeta de crédito. Consejos: Las propinas en efectivo no están incluidas.	Decimal	Propina_TarjetaCredito
Tolls_amount	Importe total de todos los peajes pagados en el viaje.	Decimal	Peajes_viaje
Total_amount	El monto total cobrado a los pasajeros. No incluye propinas en efectivo.	Decimal	Cobro_viaje








yellow_tripdata_2020-10 = REGISTRO_VIAJE_2020-10




Nombre del campo	Descripción	Tipo de dato	Traducción de columnas
VendorID	Un código que indica el proveedor de TPEP que proporcionó el registro. 1 = Creative Mobile Technologies, LLC; 2 = VeriFone Inc.	Int	VendedorID
tpep_pickup_datetime	La fecha y hora en que se activó el medidor.	Timestamp	Fecha_hora_subida
tpep_dropoff_datetime	La fecha y hora en que se desactivó el medidor.	Timestamp	Fecha_hora_bajada
Passenger_count	El número de pasajeros en el vehículo. Este es un valor ingresado por el conductor.	Int	Cant_pasajeros
Trip_distance	La distancia de viaje transcurrida en millas informada por el taxímetro.	Decimal	Distancia_viaje
PULocationID 	Zona de taxis de TLC en la que se utilizó el taxímetro.	Int	ID_ubicación_subida
DOLocationID 	Zona de taxis de TLC en la que se desactivó el taxímetro.	Int	ID_ubicación_bajada
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	Int	Codigo_tarifa
Store_and_fwd_flag	Esta bandera indica si el registro de viaje se mantuvo en el vehículo. Memoria antes de enviar al proveedor, también conocido como "almacenar y reenviar", porque el vehículo no tenía conexión al servidor. Y = viaje de almacenamiento y envío N = no es una tienda y un viaje hacia adelante	Text	Tipo_viaje
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	Int	Tipo_pago
Fare_amount	La tarifa de tiempo y distancia calculada por el medidor.	Decimal	Tarifa_tiempo_distancia
Extra	Extras y recargos varios. Actualmente, esto solo incluye las horas pico de \$ 0.50 y \$ 1 y los cargos nocturnos.	Decimal	Extra
MTA_tax	Impuesto de \$ 0.50 MTA que se activa automáticamente según el medidor tasa en uso.	Decimal	Impuesto_mta
Improvement_surcharge	Recargo de mejora de \$ 0.30 para viajes evaluados en la caída de la bandera.	Decimal	Recargo_mejora
Tip_amount	Importe de la propina: este campo se completa automáticamente para la tarjeta de crédito. Consejos: Las propinas en efectivo no están incluidas.	Decimal	Propina_TarjetaCredito
Tolls_amount	Importe total de todos los peajes pagados en el viaje.	Decimal	Peajes_viaje
Total_amount	El monto total cobrado a los pasajeros. No incluye propinas en efectivo.	Decimal	Cobro_viaje



taxi+_zone_lookup = BUSQUEDA_ZONATAXI

Nombre del campo	Descripción	Tipo de dato	Re nombre columnas
LocationID 	Número del 1 al 256 asignado a un barrio.	Int	LocalizaciónID
Borough	Distrito	Text	Distrito
Zone 	Barrio	Text	Barrio
service_zone 	Zona de taxi Yellow, Green (Boro), Airport	Text	Zona_Servicio

taxi_zone = ZONA_TAXI



Nombre del campo	Descripción	Tipo de dato	Traduccion de Columnas
ObjectID 	Duplica numero del LocationID para dejar la columna como clave primaria	Int	ObjetivoID
Sharpe_Leng	Forma de la longitud	Decimal	Forma_longitud
The_geom	Medidas geográficas.	Decimal	Geometría
Shape_area	Area de la forma	Decimal	Area_forma
Zone 	Barrio	Text	Barrio
Location ID 	Número del 1 al 256 asignado a un barrio.	Int	LocalizaciónID
Borough	Distrito	Text	Distrito

Creación de tablas en Power BI:

Las siguientes tablas se crearon en Power BI con el fin de dar nombre a valores que daban las tablas principales



Tipo_pago

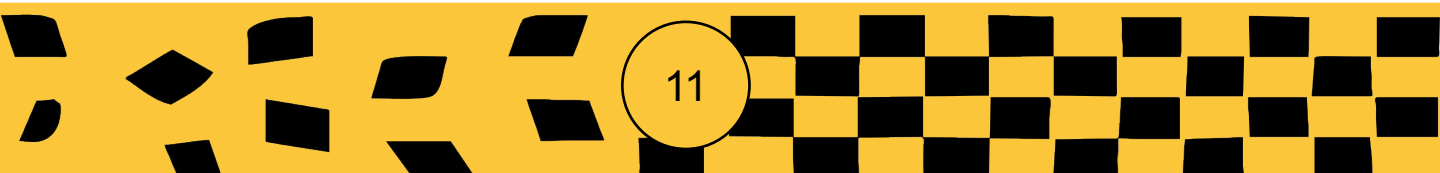
Código	Descripción
1	Tarjeta de Crédito
2	Efectivo
3	Sin Cargo
4	Disputa
5	Desconocido
6	Viaje Anulado

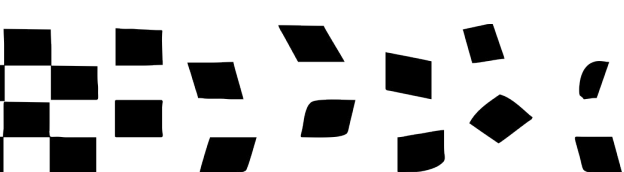
Nombre del campo	Descripción	Tipo de dato
Descripción 	Nombre del tipo de pago	Text
Código 	Código numerico del tipo de pago	Int

Vendedor_ID

Código	Descripción
1	Creative Mobile Technologies
2	VeriFone Inc.



Nombre del campo	Descripción	Tipo de dato
Descripción 	Nombre del Vendedor	Text
Código 	Numero asignado al vendedor	Int







Tipo_viaje

Código	Descripción
Y	viaje de almacenamiento y envío
N	no es una tienda y un viaje hacia adelante

Nombre del campo	Descripción	Tipo de dato
Descripción 	Descripción del tipo de viaje	Text
Código 	Letra que identifica el tipo de viaje	Text

Código_tarifa

Código	Descripción
1	Tarifa Estándar
2	JFK
3	Newark
4	Nassau o Westchester
5	Tarifa Negociada
6	Viaje en grupo

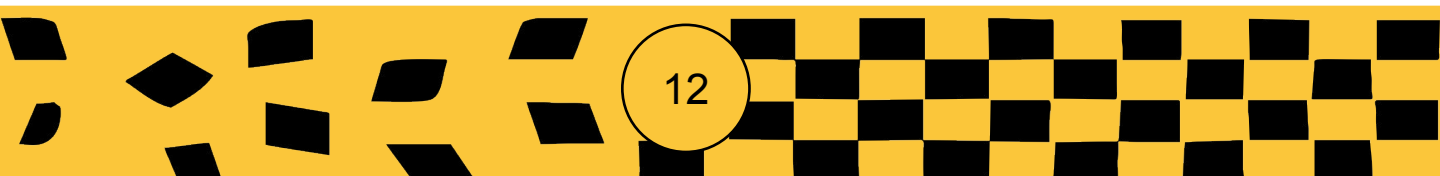
Nombre del campo	Descripción	Tipo de dato
Descripción 	Descripción de la tarifa	Text
Código 	Numero para identificar la tarifa	Int

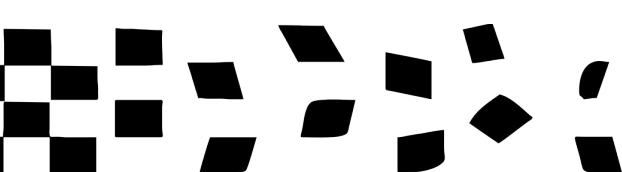
CALENDARIO

Se creo a partir de la siguiente formula:

```
CALENDARIO =  
ADDCOLUMNS (  
  CALENDAR(MIN('REGISTRO_VIAJE_2020-11'[Fecha_subida]),  
    MAX('REGISTRO_VIAJE_2020-12'[Fecha_bajada])),  
  "DateAsInteger", FORMAT ( [Date], "YYYYMMDD" ),  
  "Year", YEAR ( [Date] ),  
  "Monthnumber", FORMAT ( [Date], "MM" ),  
  "YearMonthnumber", FORMAT ( [Date], "YYYY/MM" ),  
  "YearMonthShort", FORMAT ( [Date], "YYYY/mmm" ),  
  "MonthNameShort", FORMAT ( [Date], "mmm" ),  
  "MonthNameLong", FORMAT ( [Date], "mmmm" ),  
  "DayOfWeekNumber", WEEKDAY ( [Date] ),  
  "DayOfWeek", FORMAT ( [Date], "dddd" ),  
  "DayOfWeekShort", FORMAT ( [Date], "ddd" ),  
  "Quarter", "Q" & FORMAT ( [Date], "Q" ),  
  "YearQuarter", FORMAT ( [Date], "YYYY" ) & "/" & "Q" & FORMAT ( [Date], "Q" ) )
```

Se utilizo esta formula ya que la misma fue dada en la clase y se apta a las fecha y horarios que poseen las tablas de la base de datos.





Transformaciones realizadas:

A continuación se verán impr de pantalla donde se detalla las transformaciones realizadas a la Base de datos.

Uno de los primeros cambios que se realizaron fue re nombrar todas las filas a español, al lado de la descripción de las tablas se le coloco como se llaman.

LocalizaciónID	Distrito	Barrio	Zona_Servicio
2	Queens	Jamaica Bay	Boro Zone
7	Queens	Astoria	Boro Zone
8	Queens	Astoria Park	Boro Zone
9	Queens	Auburndale	Boro Zone
10	Queens	Baisley Park	Boro Zone
15	Queens	Bay Terrace/Fort Totten	Boro Zone
16	Queens	Bayside	Boro Zone
19	Queens	Bellerose	Boro Zone
27	Queens	Breezy Point/Fort Tilden/Riis Beach	Boro Zone
28	Queens	Briarwood/Jamaica Hills	Boro Zone
30	Queens	Broad Channel	Boro Zone
38	Queens	Cambria Heights	Boro Zone
53	Queens	College Point	Boro Zone
56	Queens	Corona	Boro Zone
57	Queens	Corona	Boro Zone

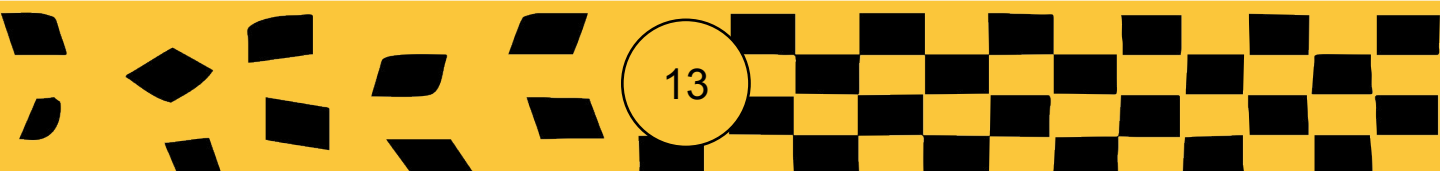
La columna de fecha y hora de subida y bajada del taxi se separo en dos columnas distintas.

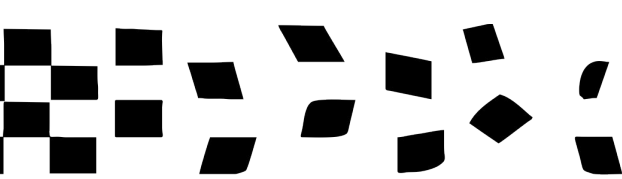
Fecha_subida	Hora_subida	Fecha_bajada	Hora_bajada
sábado, 21 de noviembre de 2020	15:56:00	sábado, 21 de noviembre de 2020	16:00:00
sábado, 21 de noviembre de 2020	11:23:00	sábado, 21 de noviembre de 2020	11:27:00
sábado, 21 de noviembre de 2020	11:17:00	sábado, 21 de noviembre de 2020	11:21:00
sábado, 21 de noviembre de 2020	10:27:00	sábado, 21 de noviembre de 2020	10:31:00
sábado, 21 de noviembre de 2020	10:10:00	sábado, 21 de noviembre de 2020	10:15:00
viernes, 20 de noviembre de 2020	16:06:00		
viernes, 20 de noviembre de 2020	15:06:00		
viernes, 20 de noviembre de 2020	15:03:00		
viernes, 20 de noviembre de 2020	14:58:00		

Nombre 'REGISTRO_VIAJE_2020-11'

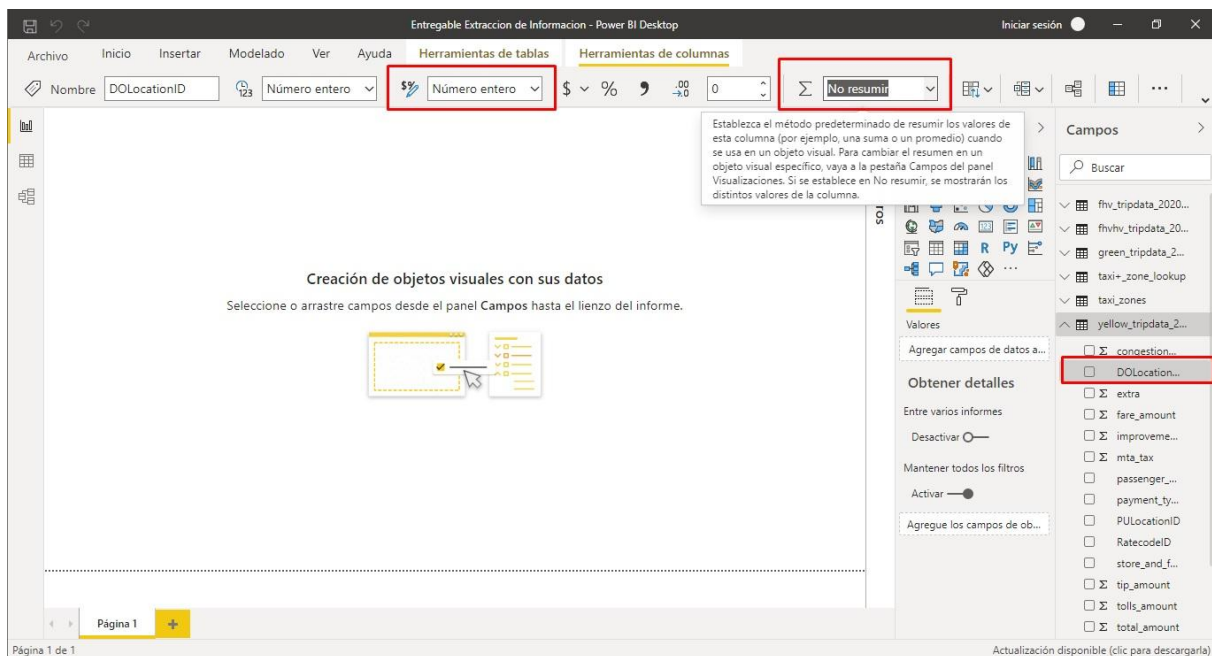
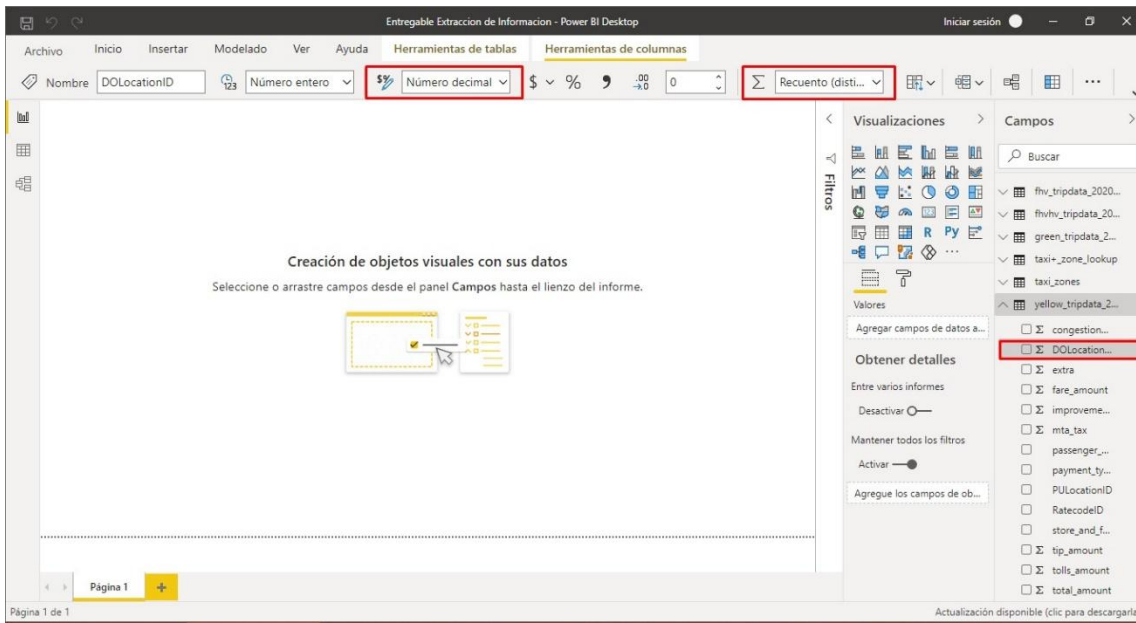
Modo de almacenamiento Importar

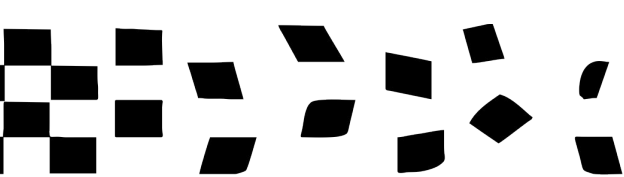
Datos actualizados 9/7/2021 17:09:53





Por otro lado se modificaron si los datos son de recuento o no para que no presente el símbolo de que es una medida que se puede sumar.





Medidas Calculadas

A continuación se detallara las formulas de las medidas calculadas utilizadas y se explicara porque se generaron.

TAXI

```
Resultado Mensual 11/2020 =  
VAR FacturacionMensual202011 =  
SUM('REGISTRO_VIAJE_2020-11'[Cobro_viaje])  
VAR CostoTotal = SUM('REGISTRO_VIAJE_2020-11'[Extra]) +  
SUM('REGISTRO_VIAJE_2020-11'[Impuesto_mta]) +  
SUM('REGISTRO_VIAJE_2020-11'[Peajes_viaje]) +  
SUM('REGISTRO_VIAJE_2020-11'[Recargo_mejora])  
Return FacturacionMensual202011 - CostoTotal
```

Se utiliza esta medida calculada para obtener el resultado de la facturación sin recargos para el mes de noviembre.

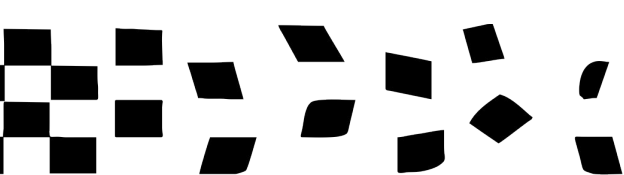
Se utilizan variables para no hacer mas pesada la base de datos.

TAXI

```
Resultado Mensual 12/2020 =  
VAR FacturacionMensual202012 =  
SUM('REGISTRO_VIAJE_2020-12'[Cobro_viaje])  
VAR CostoTotal = SUM('REGISTRO_VIAJE_2020-12'[Extra]) +  
SUM('REGISTRO_VIAJE_2020-12'[Propina_TarjetaCredito]) +  
SUM('REGISTRO_VIAJE_2020-12'[Peajes_viaje]) +  
SUM('REGISTRO_VIAJE_2020-12'[Recargo_mejora])  
Return FacturacionMensual202012 - CostoTotal
```

Se utiliza esta medida calculada para obtener el resultado de la facturación sin recargos para el mes de noviembre.

Se utilizan variables para no hacer mas pesada la base de datos.



TAXI

Gastos11-2020 = SUM
('REGISTRO_VIAJE_2020-11' [Impuesto_mta]) +
SUM('REGISTRO_VIAJE_2020-11' [Extra]) +
SUM('REGISTRO_VIAJE_2020-11' [Peajes_viaje])

Se suman los gastos que incrementan el cobro del viaje para luego poder sacar una ganancia neta de cada viaje.

TAXI

Gastos12-2020 = SUM
('REGISTRO_VIAJE_2020-12' [Propina_TarjetaCredito]) +
SUM('REGISTRO_VIAJE_2020-12' [Extra]) +
SUM('REGISTRO_VIAJE_2020-12' [Peajes_viaje])

Se suman los gastos que incrementan el cobro del viaje para luego poder sacar una ganancia neta de cada viaje.

TAXI

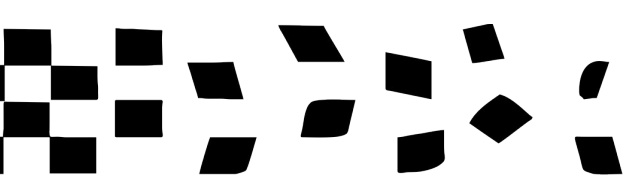
TOTALPASAJEROS11-2020 =
SUM('REGISTRO_VIAJE_2020-11' [Cant_pasajeros])

Se suman los de pasajeros para la creación de un KPI.

TAXI

TOTALPASAJEROS12-2020 =
SUM('REGISTRO_VIAJE_2020-12' [Cant_pasajeros])

Se suman los de pasajeros para la creación de un KPI.



Futuras líneas

- Poder mostrar la localización de los viajes con lo cual no solo podríamos hacer un mapa donde muestre donde se generan mas viajes sino que también se podría hacer un mapa de calor con los horarios y zonas con mayor demanda.
- Contar con la información de genero de los pasajes y grupo etario.
- Que los datos manuales que incorpora el conductor los realice un sensor o una aplicación de forma automática para generar menos errores.
- Poder identificar a los conductores, lo cual permitiría general un ranking de los mismo y poner la opción de que los usuarios finales lo califiquen.

Si bien son bases de datos muy extensas creemos que con la incorporación de mayor tecnología se podría automatizar varios procesos y brindar información mas detallada con lo cual generar mejores análisis, ayudando a la toma de decisiones de la organización.

