



DATA NEXUS

C O N S U L T I N G

DICCIONARIO DE DATOS

1. Índice

1. Índice.....	2
2. Datanexus.....	3
3. Diagrama E/R.....	4
4. Tablas primarias.....	5
4.1. Business_yelp.....	6
4.2. Reviews_google.....	7
4.3. Reviews_yelp.....	8
4.4. Metadata_google.....	9
5. Tablas derivadas.....	10
5.1. Google_yelp_business.....	11
5.2. Google_yelp_reviews.....	12
6. Tabla de vinculación.....	13
6.1. Tabla Calendar.....	13

2. Datanexus

Datanexus es un *datawarehouse* creado en *BigQuery* del servicio *Google Cloud Platform* (GCP).

Para poblar el mismo se utilizan datos proporcionados por las plataformas *Google Maps* y *Yelp* que nos proporciona información valiosa para los inversores, ayudándoles en la toma de decisiones estratégicas.

Datanexus se encuentra compuesto por 4 tablas primarias llamadas:

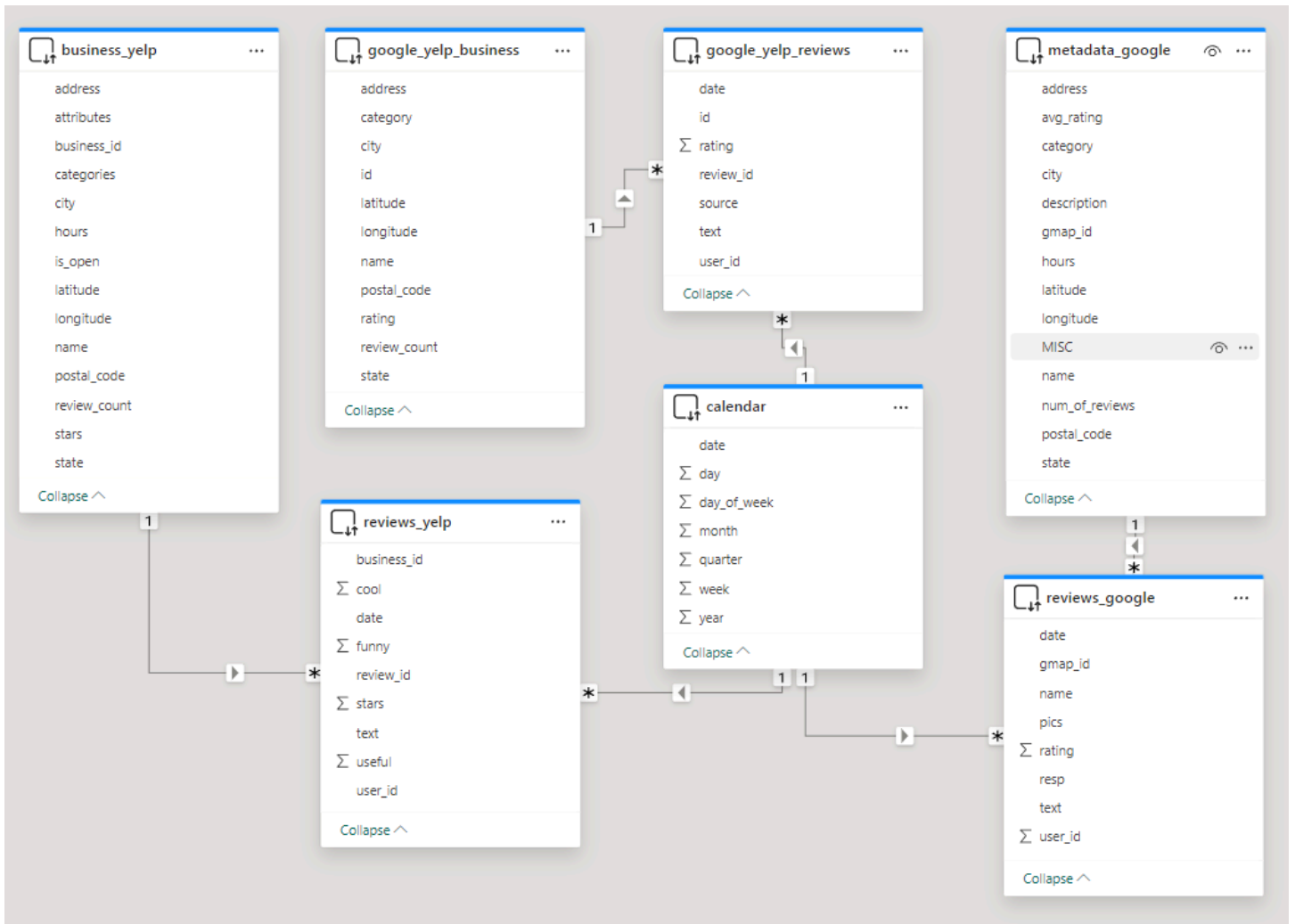
- *Business_yelp*
- *Reviews_google*
- *Reviews_yelp*
- *Metadata_google*

Por otro lado, a partir de la información proporcionada en las tablas primarias, se generan 2 tablas derivadas que integran los locales y las reseñas de *Google Maps* y de *Yelp*:

- *Google_yelp_business*
- *Google_yelp_reviews*

Finalmente, la tabla *calendar* se utiliza para poder establecer las relaciones entre las tablas primarias y las derivadas.

3. Diagrama E/R



4. Tablas primarias

Las tablas primarias poseen por separado la información en cuanto a locales gastronómicos y reseñas de los datos proporcionados por las empresas Google Maps y Yelp.

Business_yelp, es una tabla generada a partir de uno de los datasets proporcionados por la empresa *Yelp!*, asociado principalmente con los distintos negocios.

Contiene información del comercio, incluyendo localización, atributos y categorías.

Reviews_yelp, es una tabla generada a partir de uno de los datasets proporcionados por la empresa *Yelp!*

Contiene las reseñas de los usuarios pertenecientes a los 5 estados comprendidos en el alcance del proyecto: Pensilvania, California, Texas, Florida y New York. Incluye el *user_id* que escribió el *review* y el *business_id* el cual conecta con el negocio correspondiente detallado en la tabla *business_yelp*.

Reviews_google, es una tabla generada a partir de varios datasets proporcionados por la empresa *Google* que contienen las reseñas de los usuarios pertenecientes a los 5 estados comprendidos en el alcance del proyecto: Pensilvania, California, Texas, Florida y New York.

Metadata_google, es una tabla generada a partir de varios datasets proporcionados por la empresa *Google* que contiene información del comercio, incluyendo localización, atributos y categorías.

4.1. Business_yelp

Contiene información del comercio, incluyendo localización, atributos y categorías.

	Nombre de la columna	Tipo de datos	Descripción
PK	business_id	STRING	22 caracteres id del negocio, refiere al negocio en reviews_yelp
	name	STRING	Nombre del negocio
	address	STRING	Dirección completa del negocio
	city	STRING	Ciudad
	state	STRING	Código de 2 letras del Estado donde se ubica el negocio
	postal_code	STRING	Código postal
	latitude	FLOAT	Latitud
	longitude	FLOAT	Longitud
	stars	FLOAT	Rating en estrellas, redondeado a 0 o 0.5
	review_count	INTEGER	Número de reseñas
	is_open	INTEGER	0 si está cerrado, 1 si está abierto
	attributes	STRING	Atributos o características del negocio
	categories	STRING	Categorías de los negocios
	hours	STRING	Horarios en que el local permanece abierto. De día a hora, las horas son en 24 horas

4.2. Reviews_google

Contiene las reviews de los usuarios de *Google* de los estados California, Texas, New York, Pensilvania y Florida.

	Nombre de la columna	Tipo de datos	Descripción
PK	user_id	FLOAT	Id del usuario
	name	STRING	Nombre del usuario
	rating	FLOAT	Rating en estrellas
	text	STRING	Reseña en inglés
	pics	STRING	URL que contiene las imágenes que incluyen los usuarios en la reseña
	resp	STRING	Respuesta en inglés del negocio a la reseña
	gmap_id	STRING	Id del negocio que conecta con metadata_google
FK	time	STRING	Fecha y hora de la reseña formato YYYY-MM-DD

4.3. Reviews_yelp

Contiene las reseñas de los usuarios de *Yelp* en los estados de California, Texas, New York, Pensilvania y Florida. Incluye el *user_id* que escribió el *review* y el *business_id* por el cual se escribe la reseña.

	Nombre de la columnas	Tipo de datos	Descripción
PK	review_id	STRING	22 caracteres id de reseña
	user_id	STRING	22 caracteres id único de usuario
FK	business_id	STRING	22 caracteres id del negocio, refiere al negocio en business_yelp
	stars	FLOAT	Puntaje en estrellas de 1 al 5
	useful	INTEGER	Número de votos como reseña útil
	funny	INTEGER	Número de votos como reseña graciosa
	cool	INTEGER	Número de votos como reseña cool
	text	STRING	Reseña en inglés
	date	TIMESTAMP	Fecha y hora de la reseña en formato YYYY-MM-DD

4.4. Metadata_google

Contiene información de los comercios de *Google*, incluyendo localización, atributos y categorías.

	Nombre de la columna	Tipo de datos	Descripción
	name	STRING	Nombre del negocio
	address	STRING	Dirección completa del negocio
	gmap_id	STRING	Id del negocio que conecta con reviews_google
	description	STRING	Descripción del negocio en inglés
	latitude	FLOAT	Latitud
	longitude	FLOAT	Longitud
	category	STRING	Categoría de los negocios
	avg_rating	FLOAT	Puntaje en estrellas de 1 al 5
	num_of_reviews	INTEGER	Número de reseñas
	hours	STRING	Horarios de los negocios
	MISC	STRING	Atributos o características del negocio
	state	STRING	Código de 2 letras del Estado donde se ubica el negocio
	postal_code	STRING	Código postal
	city	STRING	Ciudad

5. Tablas derivadas

Las tablas derivadas surgen de la combinación de los datos de locales gastronómicos y de reseñas, proporcionados por ambas empresas.

Google_yelp_business, es una tabla que unifica todos los locales gastronómicos proporcionados por ambas empresas. Esta tabla combina los negocios pertenecientes a los 5 estados comprendidos en el alcance del proyecto: Pensilvania, California, Texas, Florida y New York.

Google_yelp_reviews, es una tabla que unifica todas las reseñas proporcionadas por ambas empresas. Esta tabla combina las *reviews* pertenecientes a los 5 estados comprendidos en el alcance del proyecto: Pensilvania, California, Texas, Florida y New York.

5.1. Google_yelp_business

Contiene información del comercio, incluyendo localización, atributos, categorías, entre otros.

	Nombre de la columnas	Tipo de datos	Descripción
	name	STRING	Nombre del negocio
	address	STRING	Dirección completa del negocio
PK	gmap_id	STRING	Id del negocio que conecta solamente con la columna id de la tabla derivada google_yelp_reviews
	latitude	FLOAT	Latitud
	longitude	FLOAT	Longitud
	category	STRING	Categoría de los negocios
	rating	FLOAT	Puntaje en estrellas de 1 al 5
	reviews_count	INTEGER	Número de reseñas
	state	STRING	Código de 2 letras del Estado donde se ubica el negocio
	postal_code	STRING	Código postal
	city	STRING	Ciudad donde se ubica el negocio

5.2. Google_yelp_reviews

Contiene las reviews de los usuarios de ambas empresas de los estados California, Texas, New York, Pensilvania y Florida.

	Nombre de la columnas	Tipo de datos	Descripción
FK	id	STRING	Id que conecta solamente con la columna gmap_id de la tabla derivada google_yelp_business
	review_id	STRING	22 caracteres id de reseña
FK	user_id	STRING	22 caracteres id único de usuario
	text	STRING	Reseña en inglés
	rating	FLOAT	Puntaje en estrellas de 1 al 5
	date	TIMESTAMP	Fecha y hora de la reseña en formato YYYY-MM-DD
	source	STRING	Fuente de la información: <i>Google Maps</i> o <i>Yelp!</i>

6. Tabla de vinculación

6.1. Tabla Calendar

Esta tabla sirve principalmente para poder establecer relación entre las tablas primarias y las tablas derivadas.

	Nombre de la columnas	Tipo de datos	Descripción
	date	DATE	Fecha expresada en formato YYYY/MM/DD
	year	INTEGER	Año expresado en entero YYYY
	month	INTEGER	Mes expresado en entero MM
	day	INTEGER	Día expresado en entero DD
	day_of_week	INTEGER	1: Domingo, 2: Lunes, 3: Martes, 4: Miércoles, 5: Jueves, 6: Viernes y 7: Domingo
	week	INTEGER	Semana del año. Va de 0 a 52.
	quarter	INTEGER	Representa el trimestre del año