

BIG DATA for BUSINESS

2.10 Modelos analíticos



Conecta Empleo

Contenido desarrollado por
Synergic Partners





Regresión

Caso de uso I: Restaurantes Italianos en NY

MODELOS ANALÍTICOS

Caso de uso I: Restaurantes Italianos en NY



Descripción

A partir de datos sobre el rating de los clientes sobre la comida, la decoración y el servicio de cada restaurante, predecir el precio de la cena.

URL del dataset

https://s3-eu-west-1.amazonaws.com/academy-formation/datasets/italian_restaurants/nyc.csv

MODELOS ANALÍTICOS

Caso de uso I: Restaurantes Italianos en NY



Variables

Y	Price: El precio (en \$US) de la cena (incluyendo una bebida y la propina)
x_1	Food : Rating de los clientes sobre la comida(max 30)
x_2	Décor: Rating de clientes sobre la decoración (max 30)
x_3	Service: Rating de clientes sobre el servicio (max 30)
x_4	East: Variable dummy= 1 (0) si el restaurante está al este (oeste) de la 5ª Avenida

MODELOS ANALÍTICOS

Caso de uso I: Restaurantes Italianos en NY



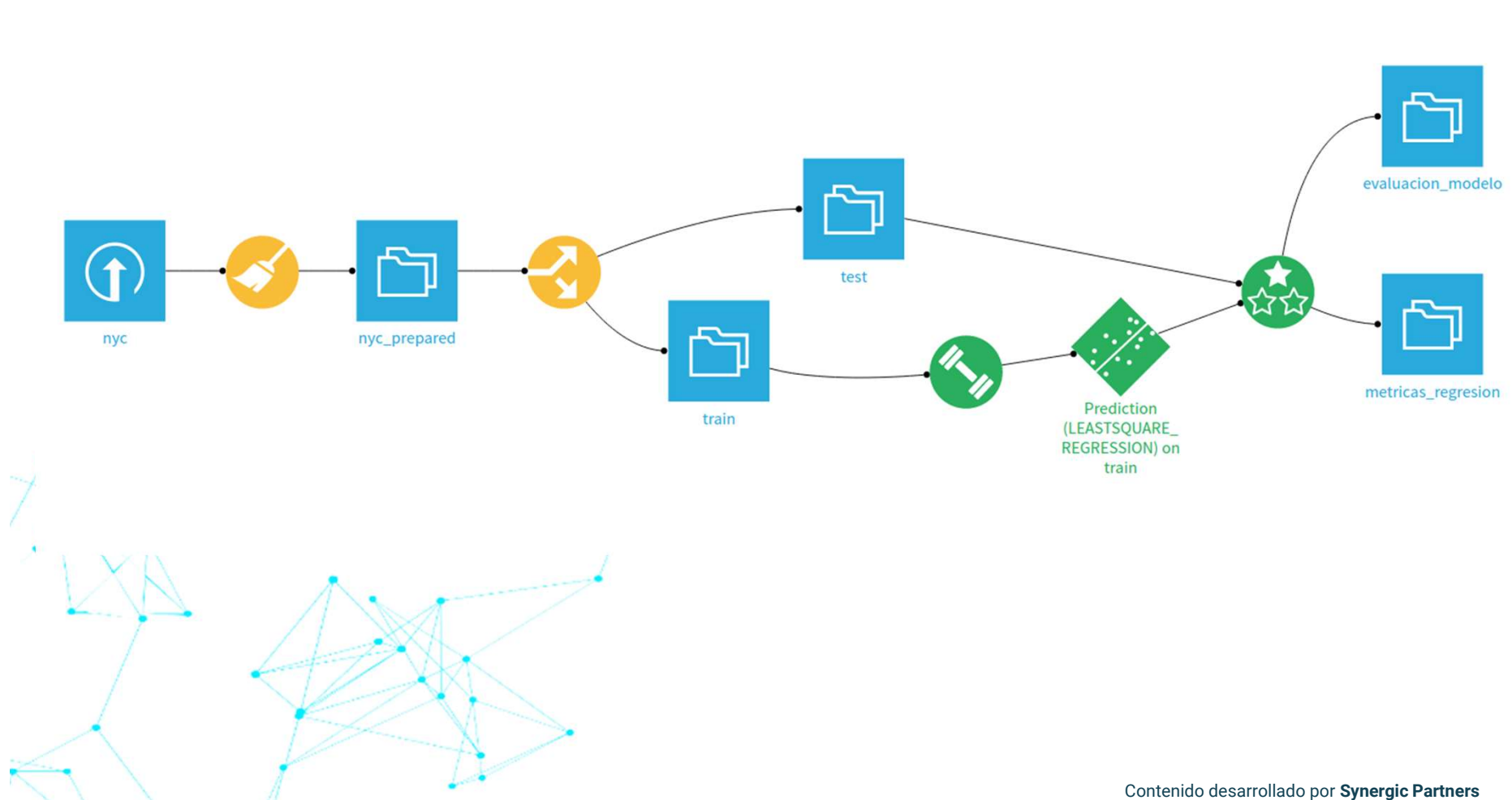
Tareas a realizar

1. Carga de los datos
2. Calidad y limpieza de los datos
3. Análisis exploratorio gráfico y descriptivo
4. Modelo predictivo de regresión
5. Evaluación del modelo
6. Predicción/Scoring

MODELOS ANALÍTICOS

Caso de uso I: Restaurantes Italianos en NY

Workflow final



MODELOS ANALÍTICOS

Caso de uso I: Restaurantes Italianos en NY

DEMO EN VIVO





Clasificación

Caso de uso II: Mantenimiento de Coches (segunda parte)

MODELOS ANALÍTICOS

Caso de uso II: Mantenimiento de Coches (segunda parte)



Descripción

El cliente es el director de un agencia de alquiler de coches. Entre sus tareas se encuentra planificar el mantenimiento de la flota de coches, pero también resolver averías inesperadas.

Para evitar sobrecostes y retrasos por falta de repuestos se quiere ajustar un modelo que permita predecir averías futuras. Para ello tenemos datos históricos sobre averías pasadas, datos de uso y mantenimiento de la flota de coches.

MODELOS ANALÍTICOS

Caso de uso II: Mantenimiento de Coches (segunda parte)



Datasets

1. **usage_data.csv**: información sobre los coches. El histórico de uso de los coches, medido en millas recorridas en ciertos instantes del tiempo.

URL

https://s3-eu-west-1.amazonaws.com/academy-formation/datasets/predictive_maintenance/usage_data.csv.gz

MODELOS ANALÍTICOS

Caso de uso II: Mantenimiento de Coches (segunda parte)



Datasets

2. **maintenance_failure.csv**: indica si una parte ha sufrido una avería (no siempre disponible)

URL

https://s3-eu-west-1.amazonaws.com/academy-formation/datasets/predictive_maintenance/maintenance_failure.csv

MODELOS ANALÍTICOS

Caso de uso II: Mantenimiento de Coches (segunda parte)



Datasets

3. **maintenance_part_consumption.csv**: contiene las partes que han requerido servicio, por coche, y motivo.

URL

https://s3-eu-west-1.amazonaws.com/academy-formation/datasets/predictive_maintenance/maintenance_part_consumption.csv.gz

MODELOS ANALÍTICOS

Caso de uso II: Mantenimiento de Coches (segunda parte)



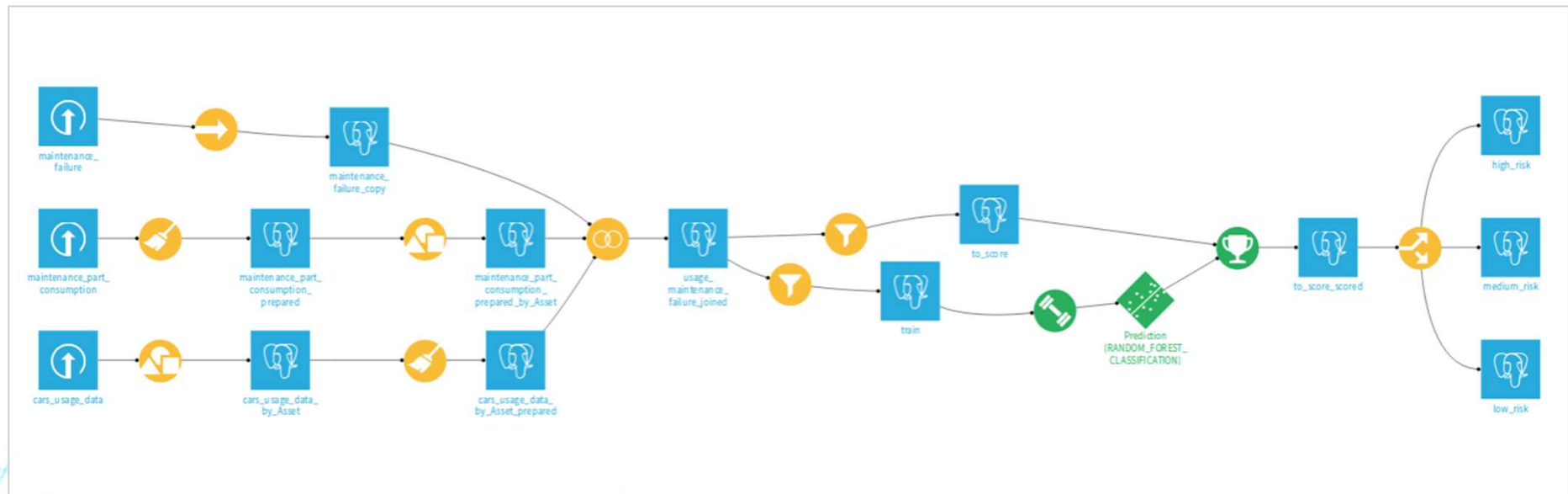
Tareas a realizar

1. Dividir entre casos conocidos y desconocidos.
2. Entrenar y analizar el modelo usando los casos conocidos (datos de entrenamiento).
3. Aplicar el modelo a los casos desconocidos (datos de prueba).
4. Desplegar el modelo.

MODELOS ANALÍTICOS

Caso de uso II: Mantenimiento de Coches (segunda parte)

Workflow final



MODELOS ANALÍTICOS

Caso de uso II: Mantenimiento de Coches (segunda parte)

DEMO EN VIVO





Clustering

Caso de uso III: Weblogs

MODELOS ANALÍTICOS

Caso de uso III: Weblogs



Descripción del problema

El equipo de marketing de Dataiku está interesado en analizar los visitantes de la página web corporativa con el fin de entender su comportamiento y clasificarlos de acuerdo a:

- **Engagement:** número de sesiones, páginas vistas, y acciones realizadas.
- **Intereses:** categorías y tipo de contenido de interés.

MODELOS ANALÍTICOS

Caso de uso III: Weblogs



Objetivos

- Segmentar a los potenciales clientes a partir de su comportamiento de navegación, y diferenciar si su interés es marketing o ventas.
- Diferenciar a los usuarios de acuerdo al tipo dispositivo usado durante la navegación.

Construir un sistema personalizado de envío de emails a visitantes conocidos

MODELOS ANALÍTICOS

Caso de uso III: Weblogs



Datasets

1. **LogsDataiku.csv.gz**: Contiene dos mese de información sobre cada visita a la página web de Dataiku.

URL

<https://s3-eu-west-1.amazonaws.com/academy-formation/datasets/weblogs/LogsDataiku.csv.gz>

MODELOS ANALÍTICOS

Caso de uso III: Weblogs



Datasets

2. **CRM_full.csv.gz**: Base de datos CRM con información transaccional y sociodemográfica de los clientes de la empresa.

URL

https://s3-eu-west-1.amazonaws.com/academy-formation/datasets/weblogs/CRM_full.csv.gz

MODELOS ANALÍTICOS

Caso de uso III: Weblogs



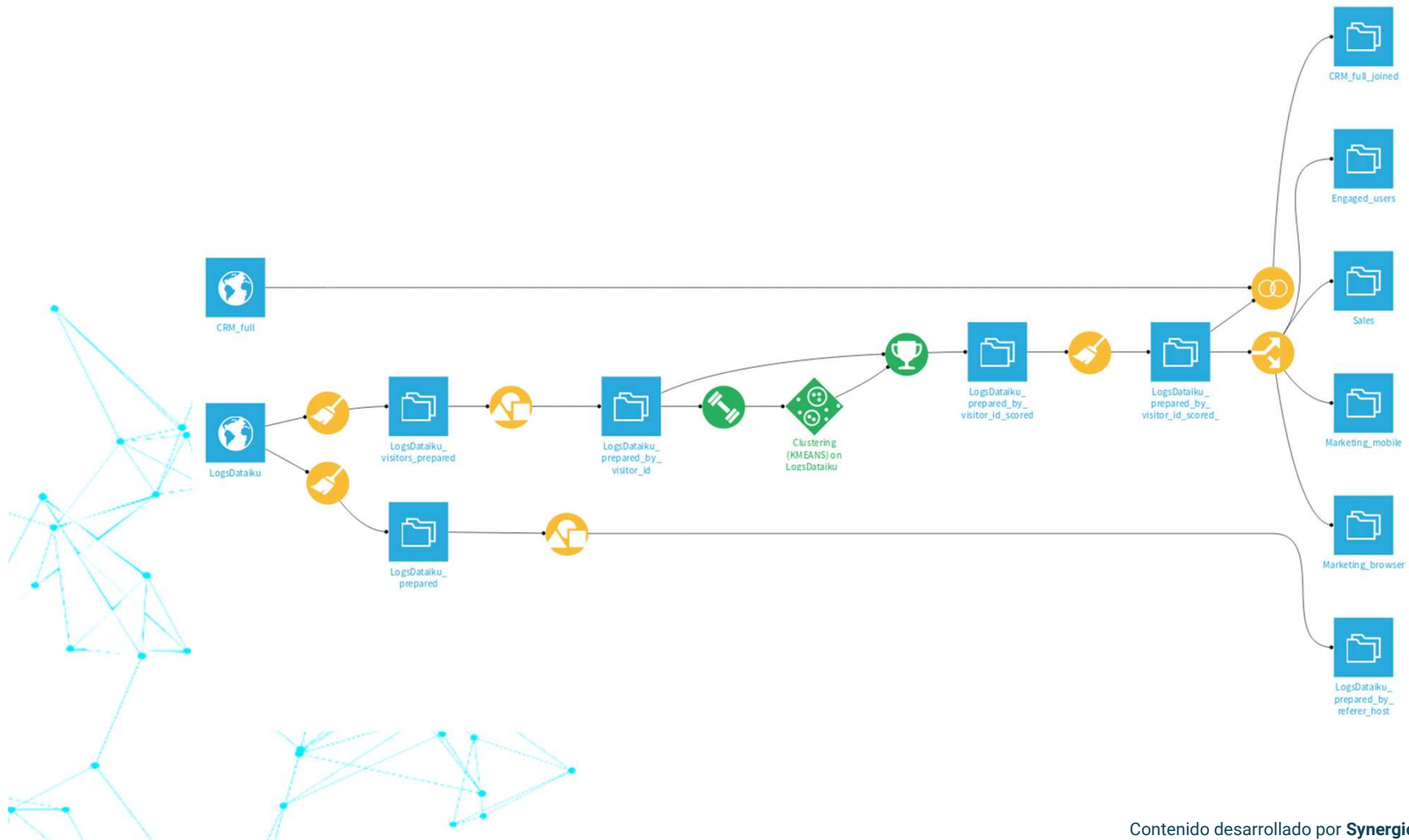
Tareas a realizar

1. Cargar datasets
2. Limpieza y enriquecimiento de los datos de navegación
3. Agrupación y agregación de los datos
4. Construcción del modelo de clustering
5. Perfilado y división de clientes
6. Enriquecimiento de datos CRM con los segmentos obtenidos de visitantes conocidos

MODELOS ANALÍTICOS

Caso de uso III: Weblogs

Workflow final



MODELOS ANALÍTICOS

Caso de uso III: Weblogs

DEMO EN VIVO



Telefónica
FUNDACIÓN

Conecta Empleo

