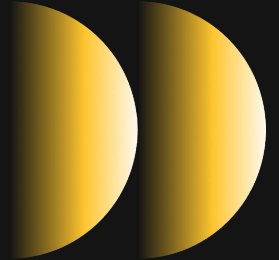


# Enunciado del Proyecto final

DSMarket



# DSMarket: Your next generation store





# DSMarket: Your next generation store

Bienvenidos al Proyecto final del Máster en Data Science: el caso de uso de DSMarket.

El proyecto se os plantea como un ejercicio de role play, dividido en varias tareas que tendréis que completar durante los meses siguientes, y que se os pedirá que presentéis al final del máster.

Este ejercicio práctico tiene como objetivo recrear un escenario de trabajo realista para un científico de datos, en el que los requerimientos están poco (o nada) definidos (¡bienvenido a la incertidumbre de la ciencia de datos!). El éxito de los diferentes proyectos dependerá a menudo de la combinación de los tres tipos principales de habilidades de las que ya hemos hablado tanto (programación + analítica + negocio).



# DSMarket: Your next generation store

Este proyecto también os brindará la oportunidad de trabajar en grupos, trabajar con los códigos de los demás y tener vuestra primera exposición a las herramientas de colaboración que se utilizan con frecuencia en casi todos los proyectos de DS.

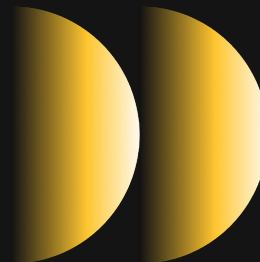
Estáis a punto de convertirlos en Nicole, una científica de datos sénior que se une al departamento financiero de una pequeña cadena de supermercados: DSMarket.

¡Que os divirtáis!





# Contexto





# Contexto

**DSMarket** (anteriormente conocida como TradiStores) es una pequeña cadena de centros comerciales en los Estados Unidos que llega algo más tarde a la transformación digital que ha revolucionado el sector de retail durante los últimos años. Su cambio de nombre es solo el primer paso de un plan disruptivo de cinco años para remodelar completamente cada proceso dentro de la empresa.

DSMarket ha incorporado a Michelle Huggins como su nueva Chief Digital Officer. Después de más de 15 años de experiencia liderando las áreas de Marketing Digital de empresas clave dentro del sector retail, Michelle está planeando explotar el activo clave que DSMarket ha estado acumulando pero ignorando durante demasiado tiempo: ¡sus datos!

Junto con los muchos especialistas en marketing digital que Michelle está contratando durante el primer año, solo ha incluido a un científico de datos. Las iniciativas de datos durante el primer año se centrarán en la estandarización y transformación de las fuentes de datos de la compañía, y en la migración de todas las fuentes y procesos de datos a la nube.



# Contexto

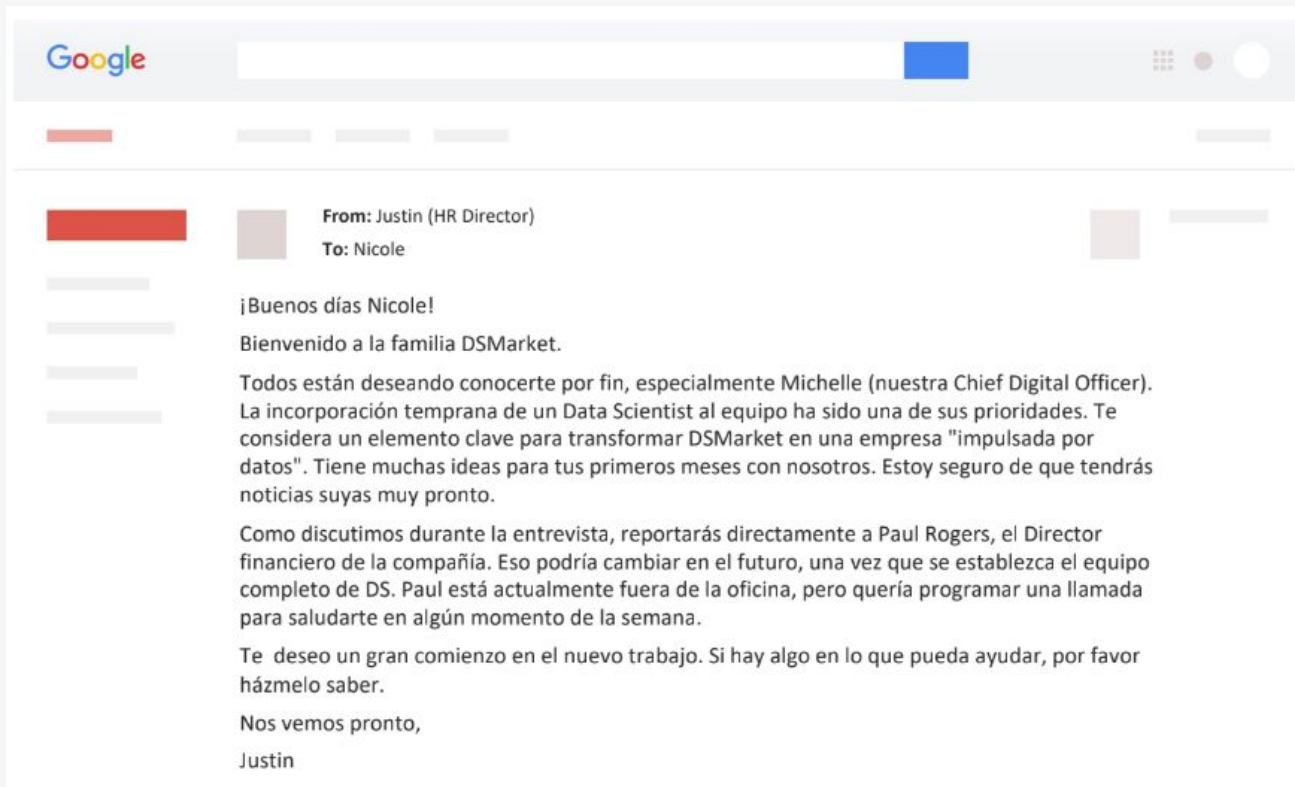
Los ingenieros de datos y los arquitectos de datos serán los principales perfiles tecnológicos requeridos. Sin embargo, DSMarket está interesado en incorporar un científico de datos senior para impulsar las iniciativas de DS con mayor prioridad. Se planea que el equipo de DS aumente rápidamente a partir del segundo año.

Vosotros seréis Nicole. Nicole ha sido contratada como científica de datos senior por la nueva Chief Digital Officer. Sin embargo, reportará directamente a Paul Rogers, el Director Financiero de la compañía. Las iniciativas de DS que se han priorizado son las de mayor importancia para el departamento financiero. Las predicciones de ventas en DSMarket siempre se han hecho utilizando enfoques muy rudimentarios, y los márgenes de error obtenidos están afectando a muchas áreas de la empresa. La magnitud de esos errores dejó de ser aceptable hace mucho tiempo.

Además, muchos de los procesos internos dentro de la empresa (estimaciones de stock, optimización de precios, entregas, predicciones de stockout,...) son ahora mismo procesos muy manuales con una fuerte dependencia de la experiencia empresarial, y su optimización mediante métodos de IA se ha incluido como parte del plan de 5 años que la empresa ha elaborado.

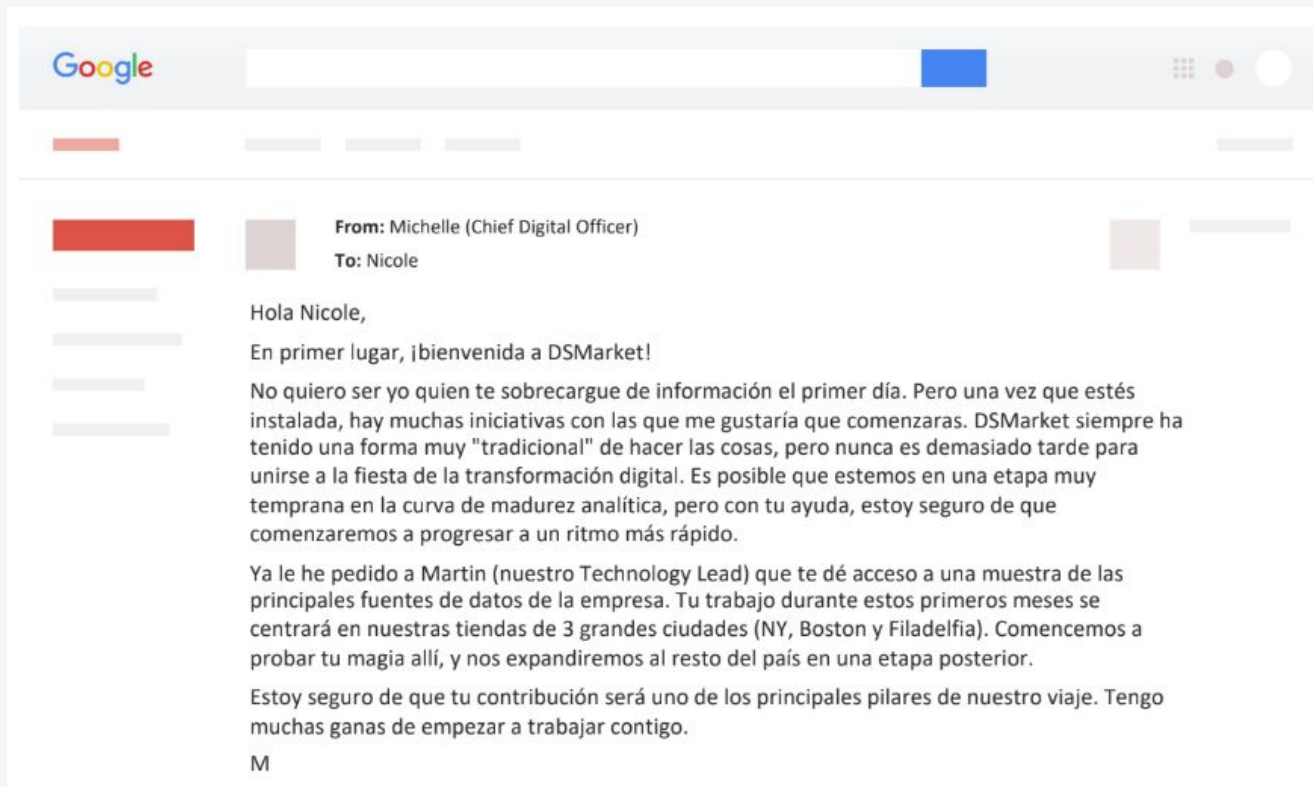


# Contexto



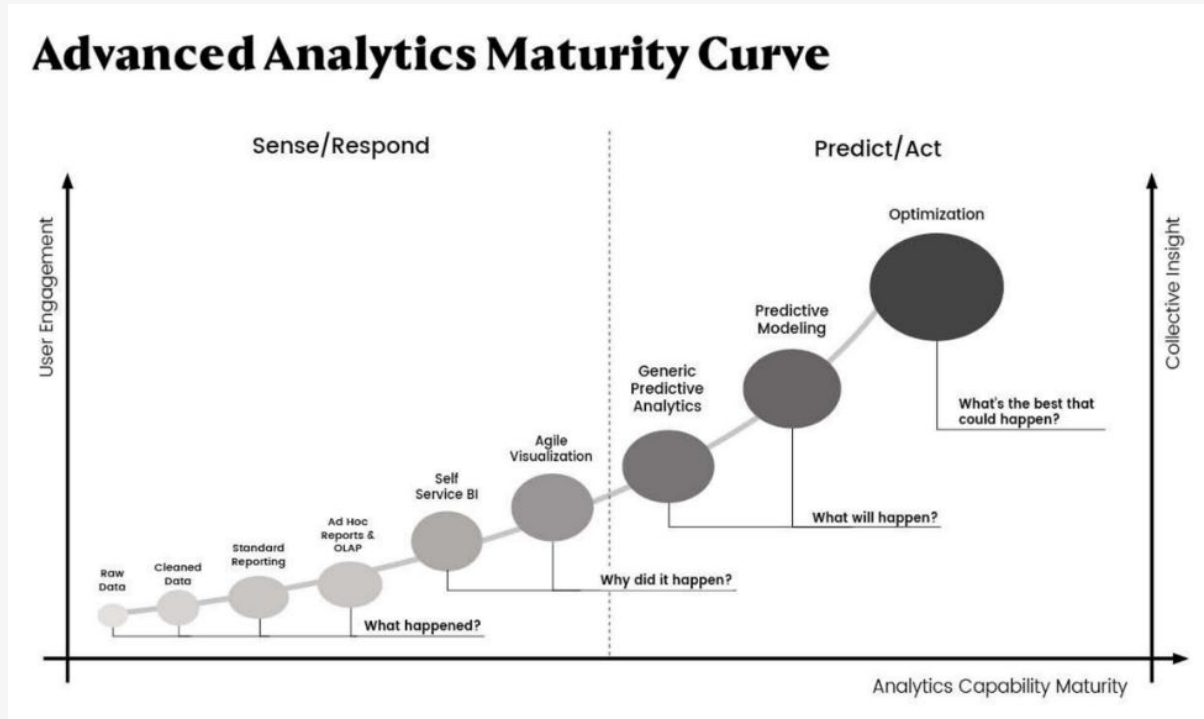


# Contexto



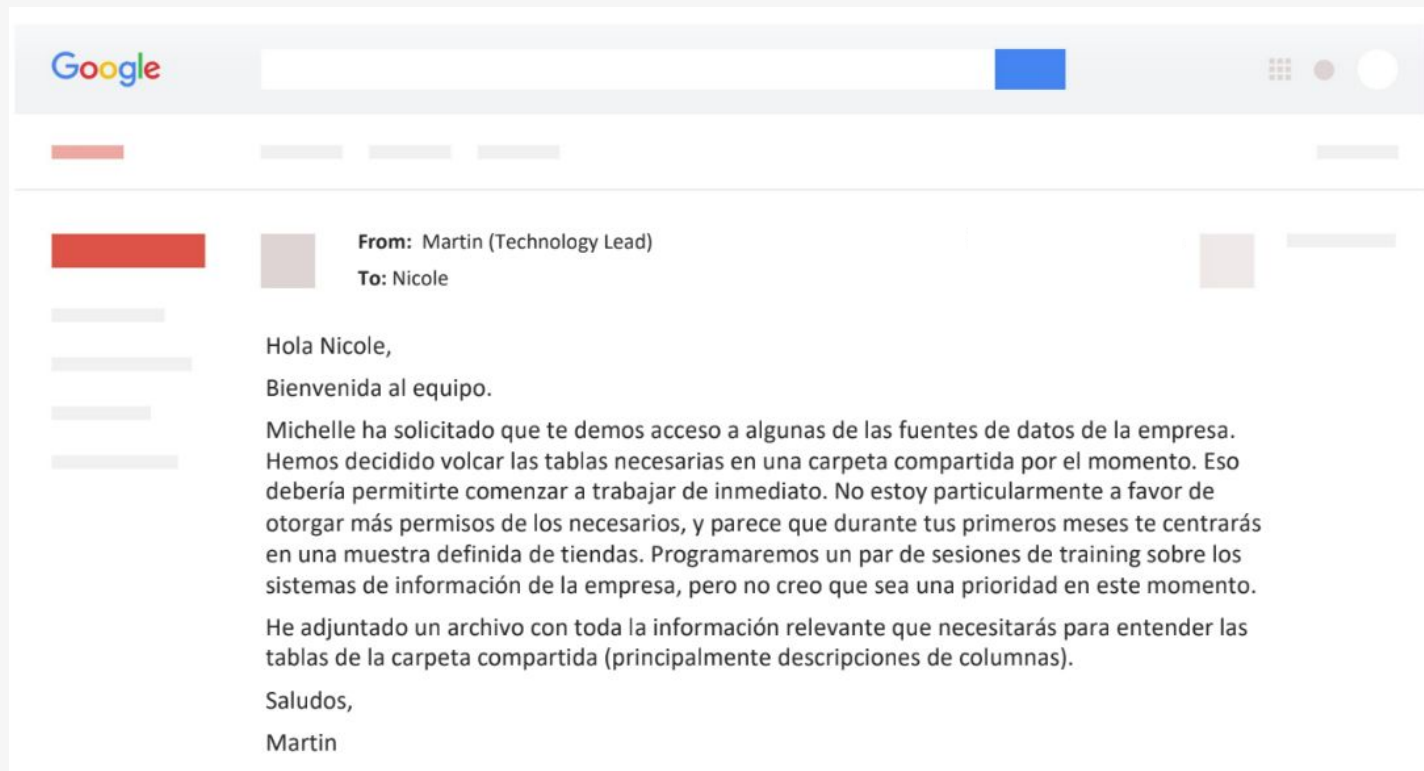


# Contexto





# Contexto





# Contexto

## FILE 1. daily\_calendar\_with\_events.csv

<u>Name</u>	<u>Table</u>	<u>Description</u>
date	calendar	date in y-m-d format
weekday	calendar	day of the week
weekday_int	calendar	numeric day of the week (Saturday day 1, Friday day 7)
d	calendar	day identifier
event	calendar	if the date includes an event, the name of this event (only a few are included)

## FILE 2. item\_prices.csv

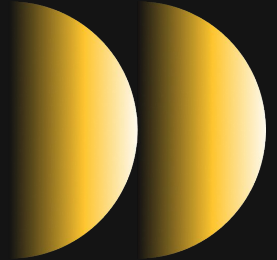
<u>Name</u>	<u>Table</u>	<u>Description</u>
item	prices	product id
category	prices	product category
store_code	prices	alphanumeric code of the store
yearweek	prices	date period for the price (year-week format)
sell_price	prices	price for the product "item" for the period in "yearweek". Prices are provided per week (average across 7 days). If not available, there were no sales for the product during that week

## FILE 3. item\_sales.csv

<u>Name</u>	<u>Table</u>	<u>Description</u>
id	sales	sales series id (combination of item + store_code)
item	sales	product id
category	sales	product category
department	sales	department id (different identifier for different stores)
store	sales	store name
store_code	sales	store id
region	sales	region
d_1,d_2,d_...	sales	number of units sold per day

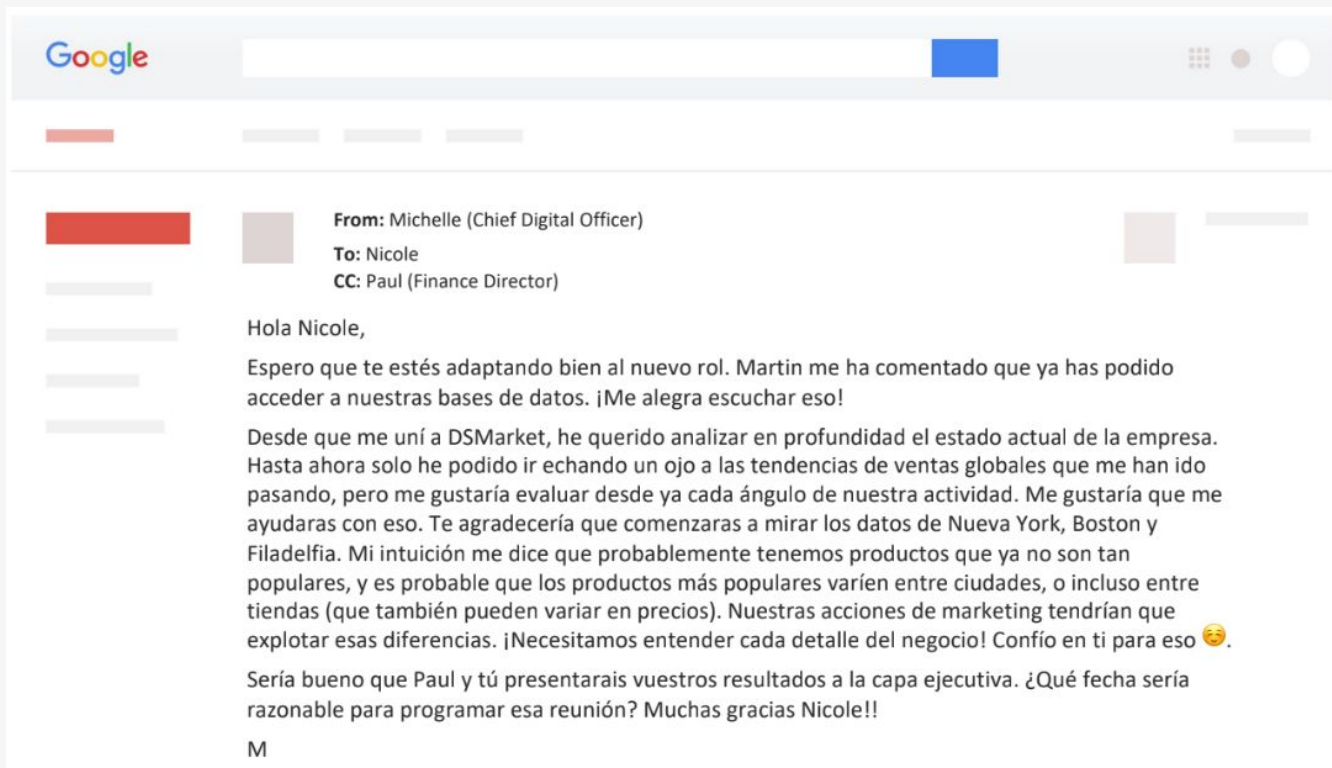


# Tarea 1: Análisis



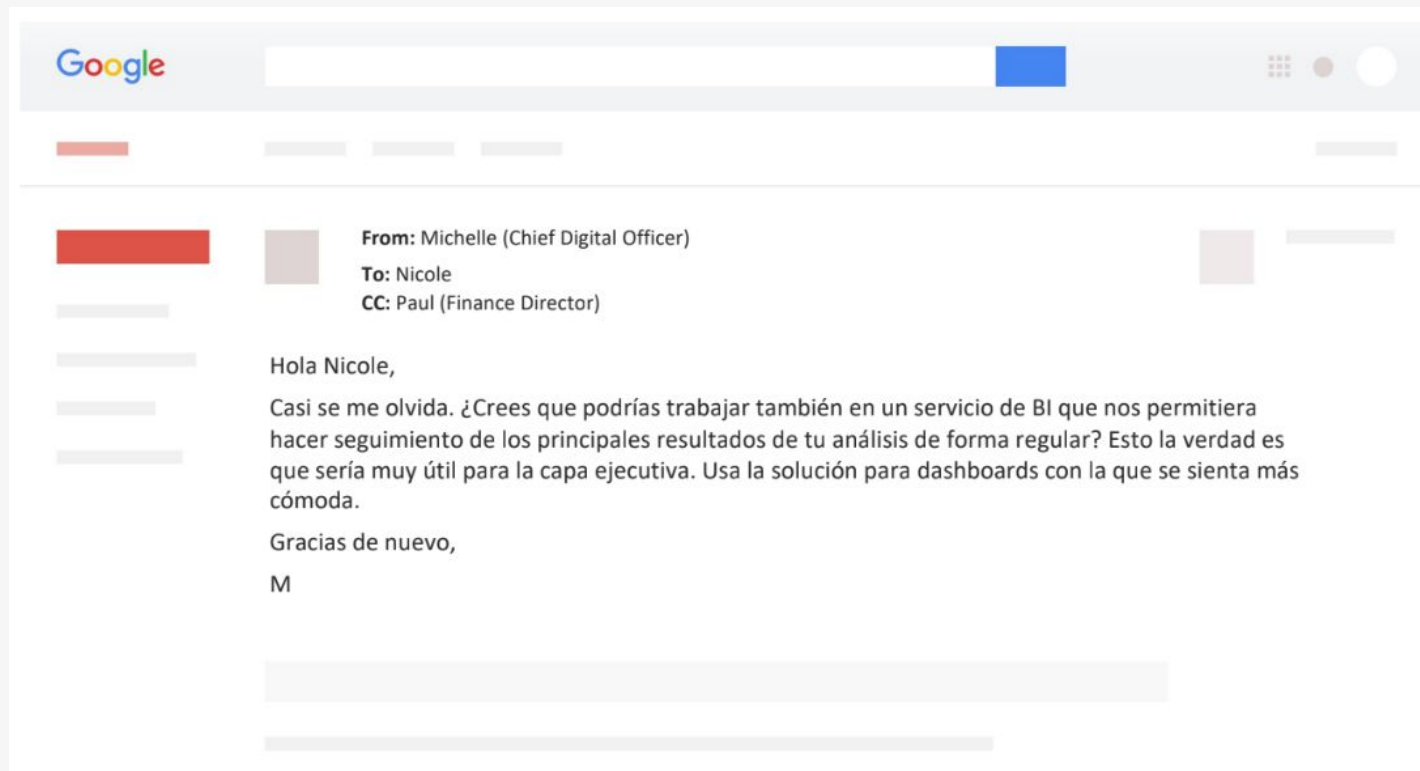


# Tarea 1



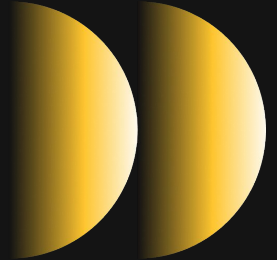


# Tarea 1



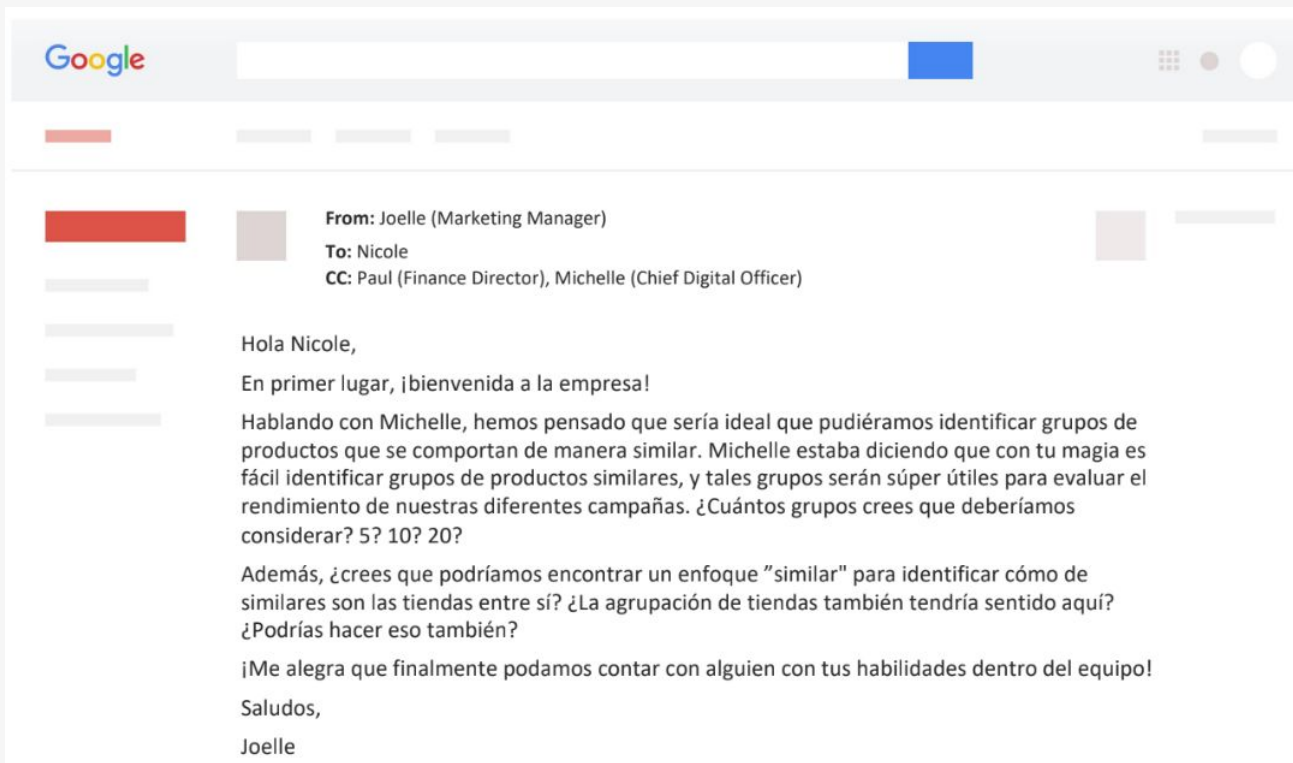


# Tarea 2: Clustering



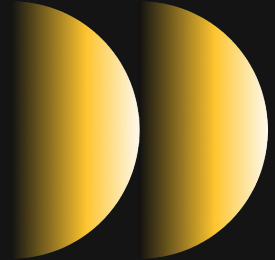


# Tarea 2



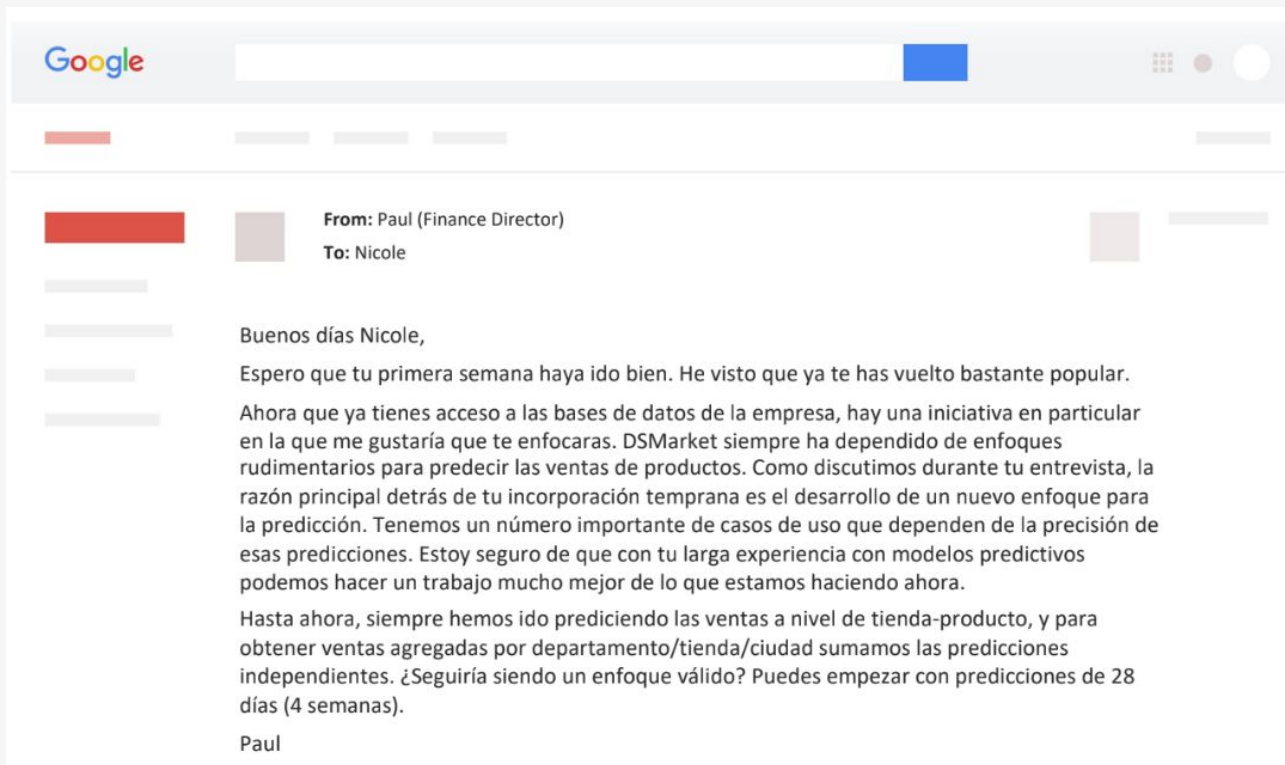


# Tarea 3: Predicción de ventas



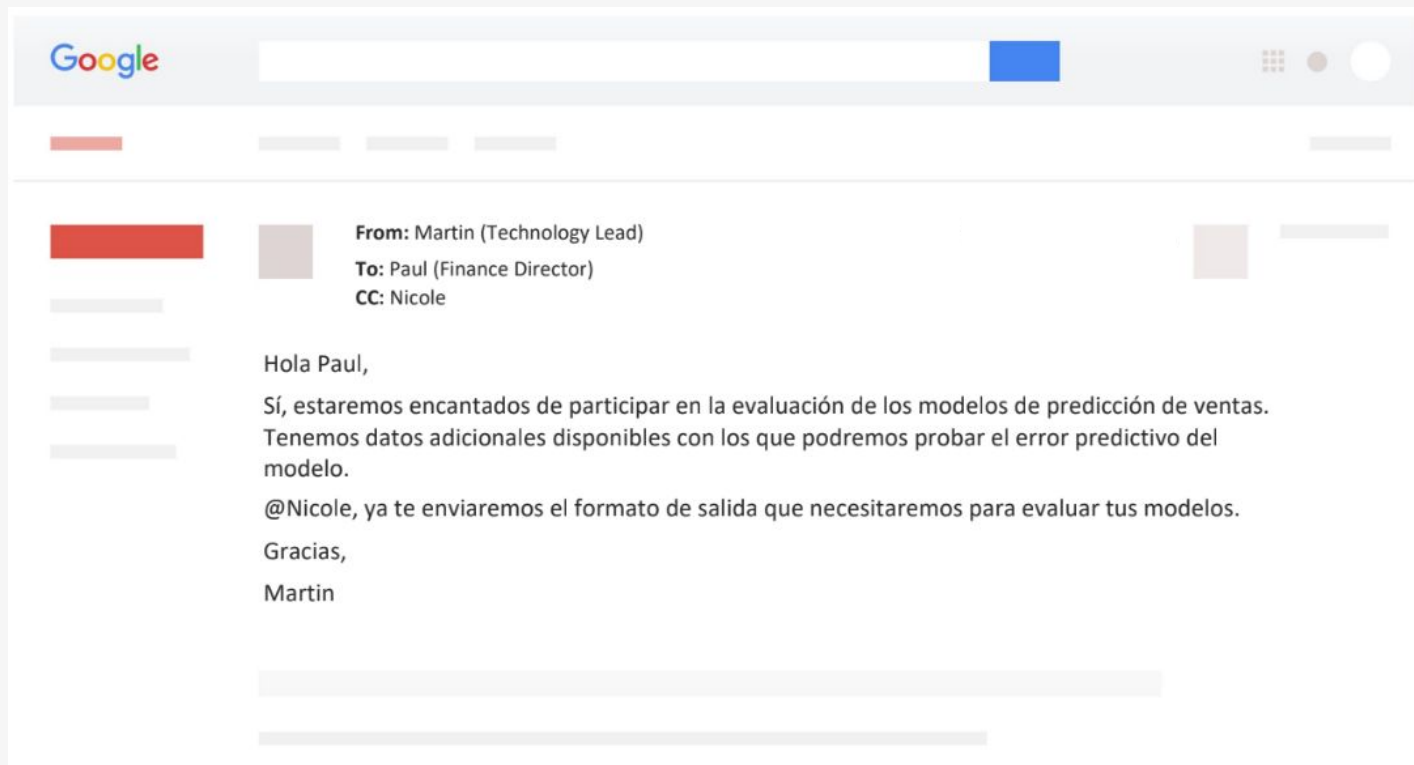


# Tarea 3



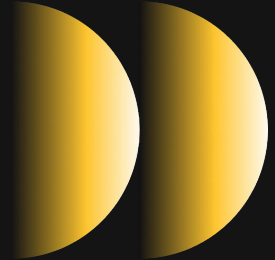


# Tarea 3



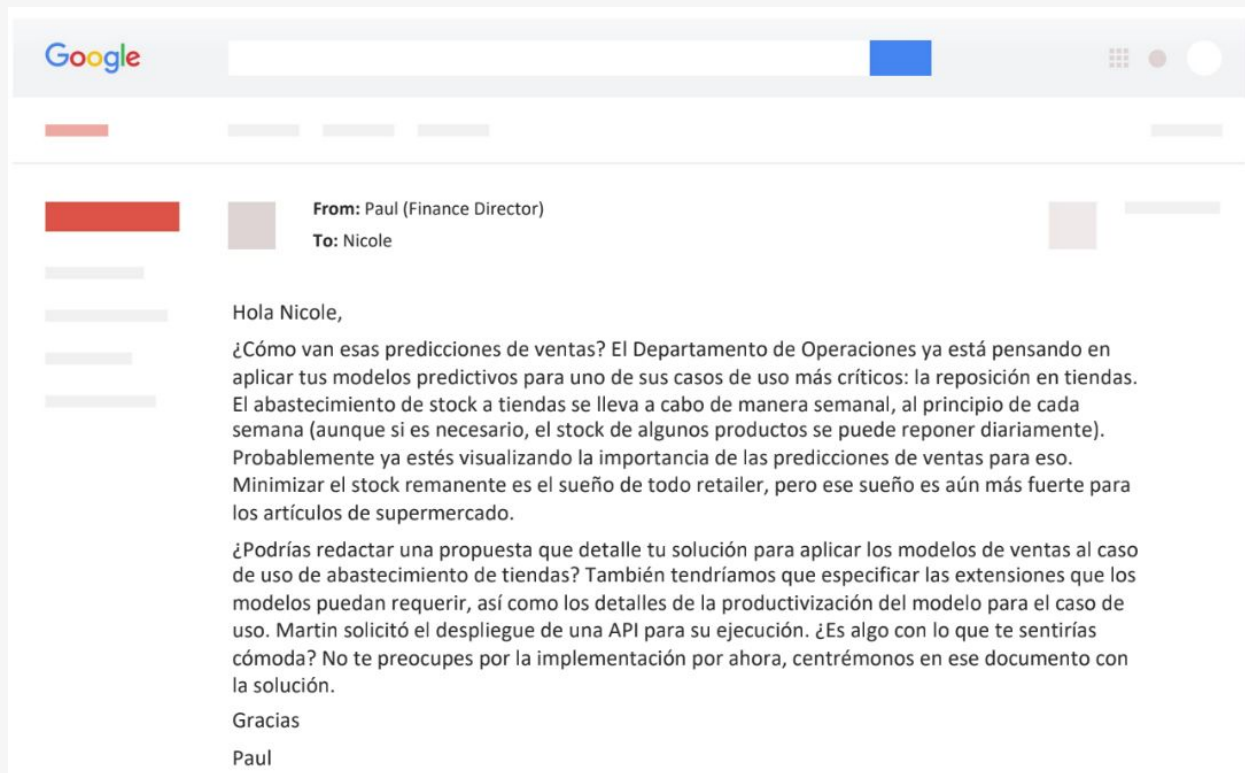


# Tarea 4: Caso de uso de abastecimiento de tiendas



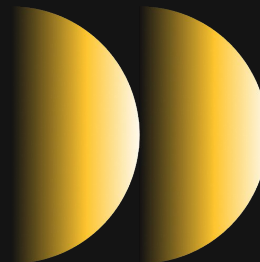


# Tarea 4





# Algunas reglas





# Algunas reglas

- El proyecto se realizará en grupos de 4 personas
- Se valorará el desarrollo técnico de las tareas, pero también la creatividad en la propuesta, la orientación a negocio y la comunicación de resultados
- Cualquier consulta sobre el evaluable se realizará en slack en los grupos al efecto

