

END-TO-END MACHINE LEARNING

Proyecto final: DSMarket

DSMarket - your next generation store

Bienvenidos al Proyecto final del Master 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).

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

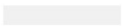
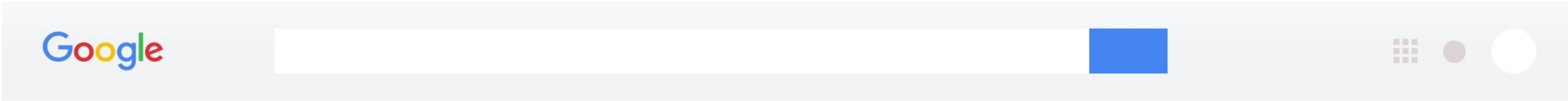


DSMarket (anteriormente conocida como TradiStores) es una pequeña cadena de centros comerciales en los Estados Unidos que llega algo más tarde a la transformación digital que ha revolucionado el sector de retail durante los últimos años. Su cambio de nombre es solo el primer paso de un plan disruptivo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

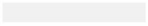
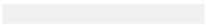
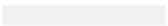
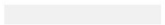
Vosotros seréis Nicole. Nicole ha sido contratada como **científica de datos senior** por la nueva Chief Digital Officer. Sin embargo, **reportará directamente a Paul Rogers, el Director Financiero de la compañía**. Las iniciativas de DS que se han priorizado son las de mayor importancia para el departamento 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.





From: Michelle (Chief Digital Officer)
To: Nicole

September 20th, 2021



Hola Nicole,

En primer lugar, ¡bienvenida a DSMarket!

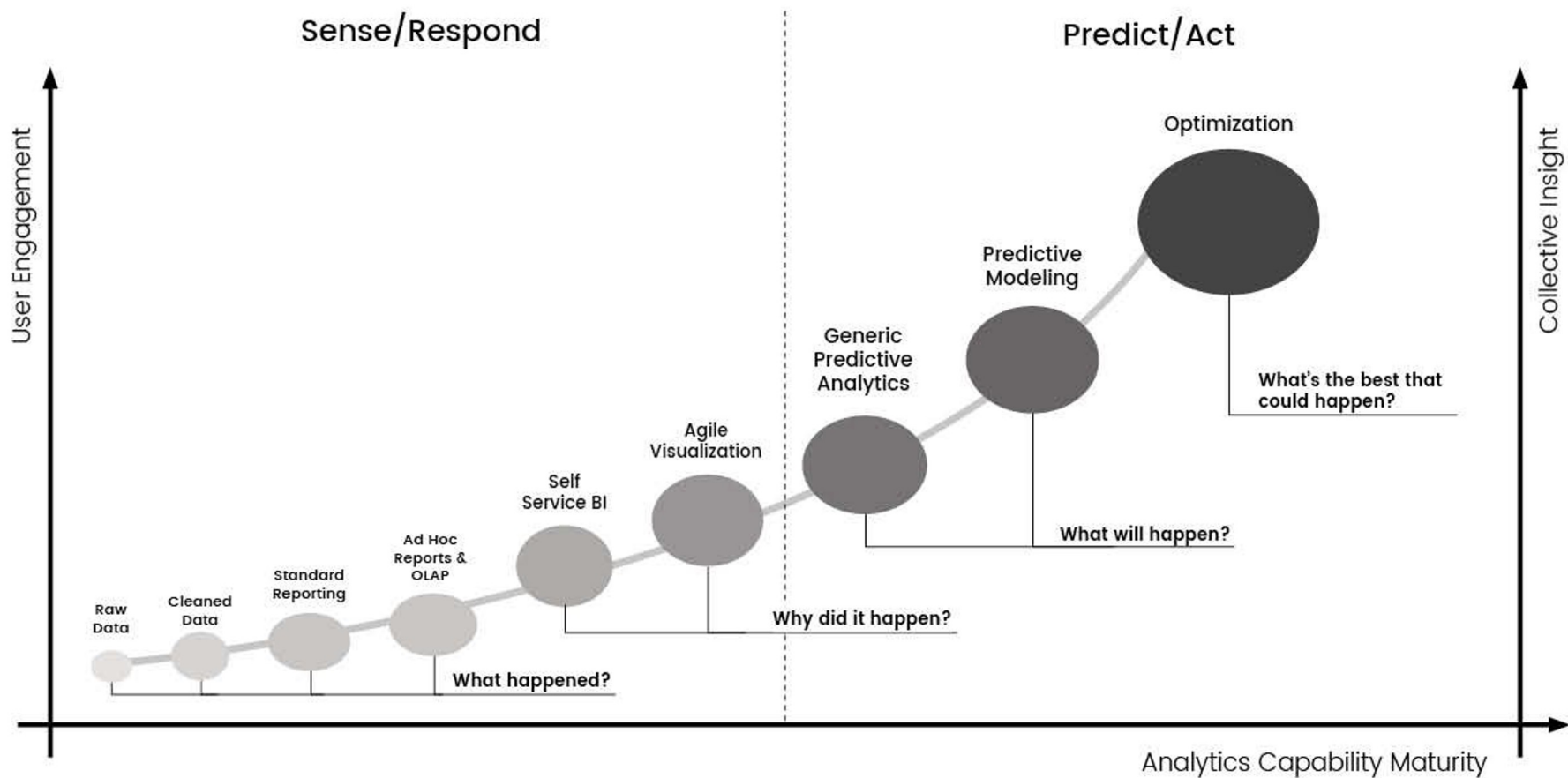
No quiero ser yo quien te sobrecargue de información el primer día. Pero una vez que estés instalada, hay muchas iniciativas con las que me gustaría que comenzaras. DSMarket siempre ha tenido una forma muy "tradicional" de hacer las cosas, pero nunca es demasiado tarde para unirse a la fiesta de la transformación digital. Es posible que estemos en una etapa muy temprana en la curva de madurez analítica, pero con tu ayuda, estoy seguro de que comenzaremos a progresar a un ritmo más rápido.

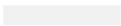
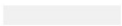
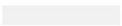
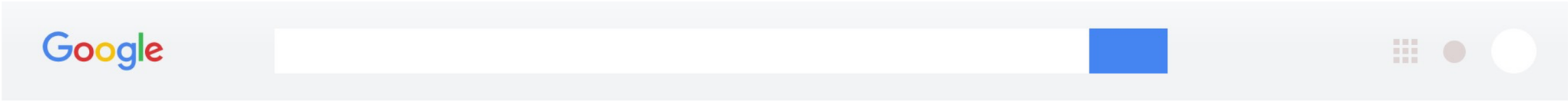
Ya le he pedido a Martin (nuestro Technology Lead) que te dé acceso a una muestra de las principales fuentes de datos de la empresa. Tu trabajo durante estos primeros meses se centrará en nuestras tiendas de 3 grandes ciudades (NY, Boston y Filadelfia). Comencemos a probar tu magia allí, y nos expandiremos al resto del país en una etapa posterior.

Estoy seguro de que tu contribución será uno de los principales pilares de nuestro viaje. Tengo muchas ganas de empezar a trabajar contigo.

M

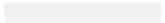
Advanced Analytics Maturity Curve



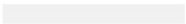
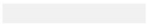
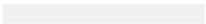
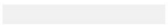


From: Martin (Technology Lead)

September 21th, 2021



To: Nicole



Hola Nicole,

Bienvenida al equipo.

Michelle ha solicitado que te demos acceso a algunas de las fuentes de datos de la empresa. Hemos decidido volcar las tablas necesarias en una carpeta compartida por el momento. Eso debería permitirte comenzar a trabajar de inmediato. No estoy particularmente a favor de otorgar más permisos de los necesarios, y parece que durante tus primeros meses te centrarás en una muestra definida de tiendas. Programaremos un par de sesiones de training sobre los sistemas de información de la empresa, pero no creo que sea una prioridad en este momento.

He adjuntado un archivo con toda la información relevante que necesitarás para entender las tablas de la carpeta compartida (principalmente descripciones de columnas).

Saludos,

Martin

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

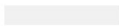
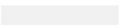
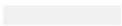
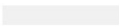
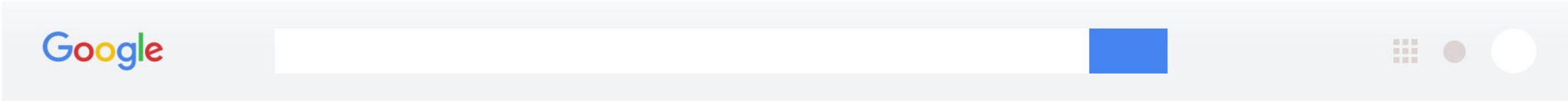
<u>Name</u>	<u>Table</u>	<u>Description</u>
item	prices	product id
category	prices	product category
store_code	prices	alphanumeric code of the store
yearweek	prices	date period for the price (year-week format)
sell_price	prices	price for the product “item” for the period in “yearweek”. Prices are provided 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

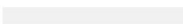
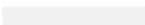
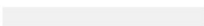
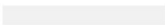
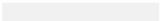
Tarea 1: Análisis





From: Michelle (Chief Digital Officer)
To: Nicole
CC: Paul (Finance Director)

September 22th, 2021



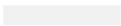
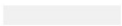
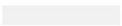
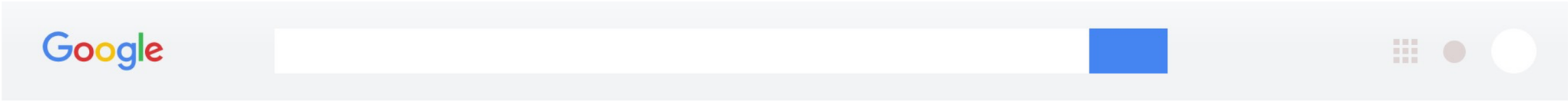
Hola Nicole,

Espero que te estés adaptando bien al nuevo rol. Martin me ha comentado que ya has podido acceder a nuestras bases de datos. ¡Me alegra escuchar eso!

Desde que me uní a DSMarket, he querido analizar en profundidad el estado actual de la empresa. Hasta ahora solo he podido ir echando un ojo a las tendencias de ventas globales que me han ido pasando, pero me gustaría evaluar desde ya cada ángulo de nuestra actividad. Me gustaría que me ayudaras con eso. Te agradecería que comenzaras a mirar los datos de Nueva York, Boston y Filadelfia. Mi intuición me dice que probablemente tenemos productos que ya no son tan populares, y es probable que los productos más populares varíen entre ciudades, o incluso entre tiendas (que también pueden variar en precios). Nuestras acciones de marketing tendrían que explotar esas diferencias. ¡Necesitamos entender cada detalle del negocio! Confío en ti para eso 😊.

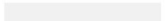
Sería bueno que Paul y tú presentarais vuestros resultados a la capa ejecutiva. ¿Qué fecha sería razonable para programar esa reunión? Muchas gracias Nicole!!

M



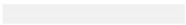
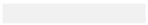
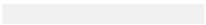
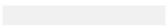
From: Michelle (Chief Digital Officer)

September 23th, 2021



To: Nicole

CC: Paul (Finance Director)

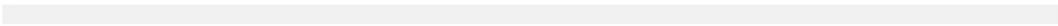
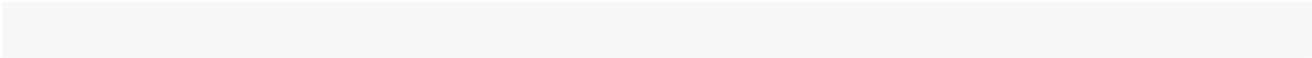


Hola Nicole,

Casi se me olvida. ¿Crees que podrías trabajar también en un servicio de BI que nos permitiera hacer seguimiento de los principales resultados de tu análisis de forma regular? Esto la verdad es que sería muy útil para la capa ejecutiva. Usa la solución para dashboards con la que se sienta más cómoda.

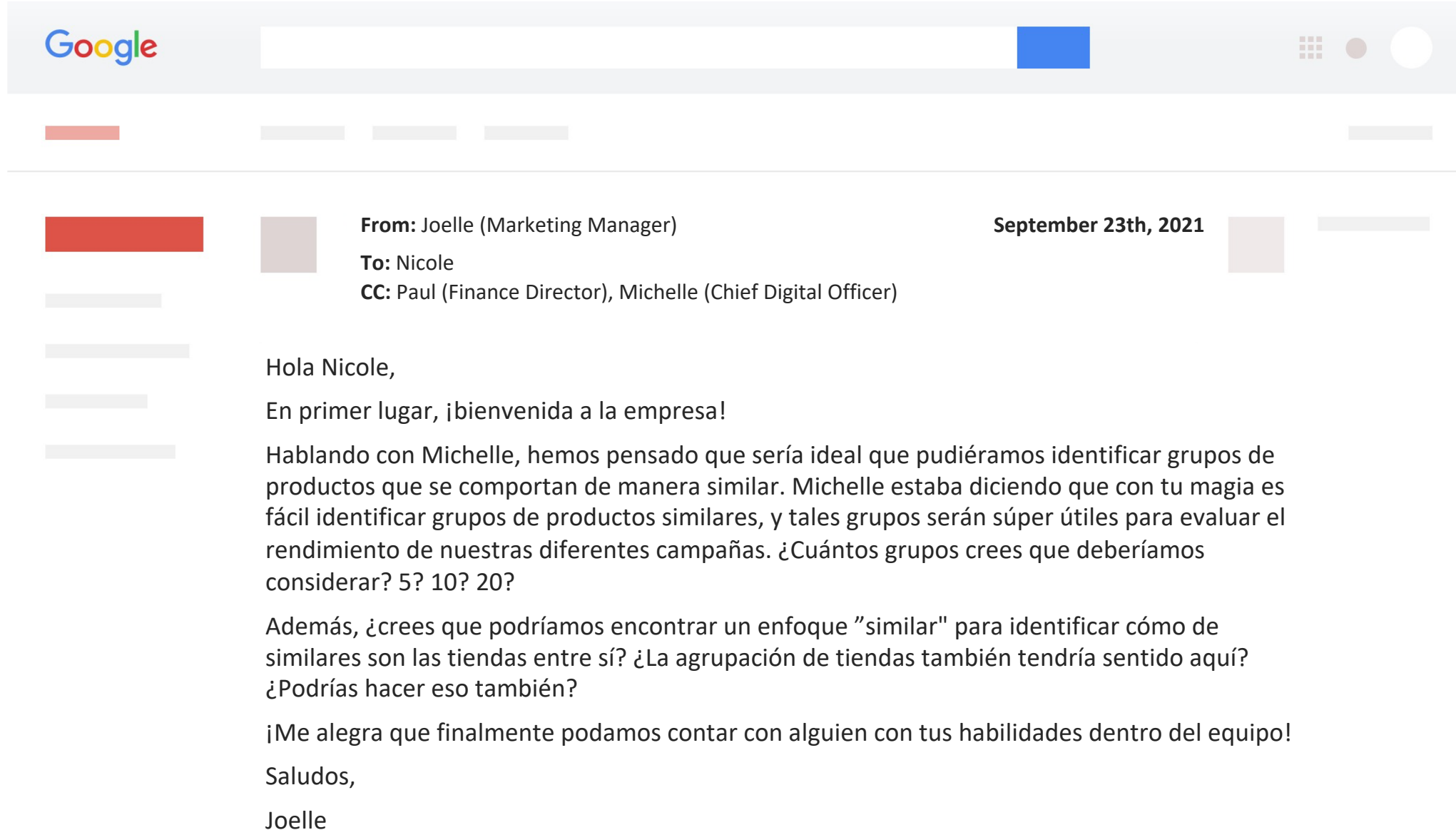
Gracias de nuevo,

M



Tarea 2: Clustering

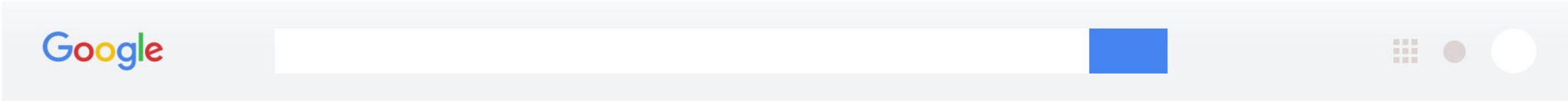




Tarea 3: Predicción de ventas

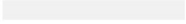
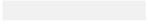
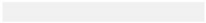
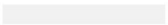
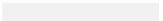






From: Martin (Technology Lead)
To: Paul (Finance Director)
CC: Nicole

September 28th, 2021



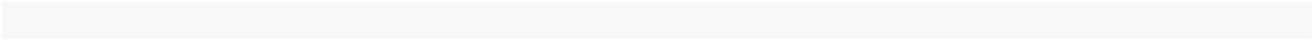
Hola Paul,

Sí, estaremos encantados de participar en la evaluación de los modelos de predicción de ventas. Tenemos datos adicionales disponibles con los que podremos probar el error predictivo del modelo.

@Nicole, ya te enviaremos el formato de salida que necesitaremos para evaluar tus modelos.

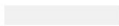
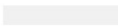
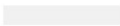
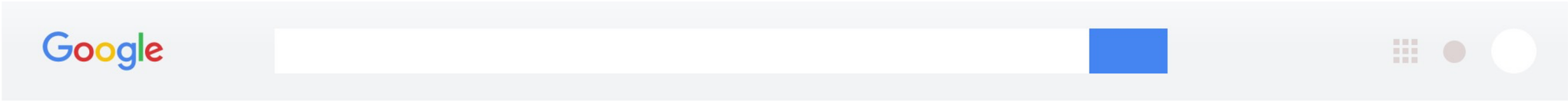
Gracias,

Martin



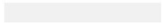
Tarea 4: Caso de uso de abastecimiento de tiendas



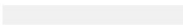
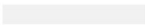
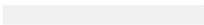
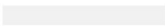


From: Paul (Finance Director)

September 28th, 2021



To: Nicole



Hola Nicole,

¿Cómo van esas predicciones de ventas? El Departamento de Operaciones ya está pensando en aplicar tus modelos predictivos para uno de sus casos de uso más críticos: la reposición en tiendas. El abastecimiento de stock a tiendas se lleva a cabo de manera semanal, al principio de cada semana (aunque si es necesario, el stock de algunos productos se puede reponer diariamente). Probablemente ya estés visualizando la importancia de las predicciones de ventas para eso. Minimizar el stock remanente es el sueño de todo retailer, pero ese sueño es aún más fuerte para los artículos de supermercado.

¿Podrías redactar una propuesta que detalle tu solución para aplicar los modelos de ventas al caso de uso de abastecimiento de tiendas? También tendríamos que especificar las extensiones que los modelos puedan requerir, así como los detalles de la productivización del modelo para el caso de uso. Martin solicitó el despliegue de una API para su ejecución. ¿Es algo con lo que te sentirías cómoda? No te preocupes por la implementación por ahora, centrémonos en ese documento con la solución.

Gracias

Paul

Algunas reglas:

- La práctica 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

