Brainstorm

* Formula 1
  + Predecir quién va a ganar la temporada en diferentes puntos de esta
  + Cuántas veces gana la temporada quien va lider en la primera mitad…
  + Métricas para predecir qué tipo de neumáticos usa el ganador de la temporada
  + Paquetes de viajes a diferentes Grandes Premios, predecir a cuales será más conveniente ir en función de dónde vives o del estado de la temporada…
  + ( o a parques de atracciones del mundo)
* Twitter
  + Predecir qué pasará con la bolsa y las acciones de Tesla cuando Elon Musk publica un tweet polémico, positivo, negativo…
  + Predecir qué pasará con el clima de Albacete según el tweet…
  + O cómo influye el clima en sus tweets
* Servicio predicción literaria
  + Predictor de éxito/interés para recomendaciones de novedades literarias en función de datos históricos de editoriales y/o librerías

Diseño del DAaaS

### Definición la estrategia del DAaaS

Se ofrecerá un servicio de predicción de éxito en novedades literarias, basado en diferentes variables como autor, género y subgéneros o si es basado en contenido de plataformas de streaming o temática en tendencia en internet.  
Puede ser monetizable para librerías online, para mejorar la gestión del catálogo de novedades mediante la recomendación de adquisición de nuevos productos, así como también puede existir una versión para recomendaciones literarias a usuarios finales via web o email.

Qué información se requerirá:

* Datos históricos de éxitos de las editoriales (en una primer versión, grandes grupos editoriales como Grupo Planeta o Penguin Random House, para después ir incorporando datasets de otras más pequeñas)
* Datos históricos de éxitos de ventas de librerías (en una primera versión, se trabajará con datos de La Casa del Libro, FNAC y Santos Ochoa, para después incluir otras).

### Arquitectura DAaaS

Componentes a utilizar:

* [Cloud function (Python)] Crawlers de las siguientes páginas:
  + <https://www.planetadelibros.com/>
  + <https://www.penguinlibros.com/es/>
  + <https://www.casadellibro.com/>
  + <https://www.santosochoa.es/>
  + <https://www.fnac.es/home/book.aspx>
* [Google Cloud Storage] Bucket de staging de datos
* [VM Instancia Elastic] Elastic para procesar datos (o valorar si Dataproc Hive y conectar con instancia ES).
* [Kibana] Visualizaciones tendencias/predicciones que se mostrarán en el visualizador de búsquedas.
* [Python] API y visualizador de busquedas

### DAaaS Operating Model Design and Rollout

1. [Cloud Scheduler] La actualización de información de novedades en la base de datos se realizará de forma semanal, por lo que se programará la llamada a los crawlers en Cloud Function todos los lunes a las 23:59, ya que los lunes es cuando habitualmente las editoriales cargan novedades.
2. Todos los martes 8am se realizará limpieza de los nuevos datos.
3. Tras la limpieza se ejecutará script de actualización de datos de Google Storage a Elastic (o Hive)
4. A continuación se crearán nuevas visualizaciones en Kibana de los datos obtenidos en la última semana y que serán mostrados de forma estática junto a las búsquedas dinámicas de los usuarios.
5. Carga de los datos estáticos en el visualizador de datos

### Desarrollo de la plataforma DAaaS.

1. Creación nuevo proyecto Google Cloud
2. Creación buckets Google Storage
3. Creación cluster Hive (si utilizamos hive con ES)
4. Creación instancia Elastic
5. Desarrollo Crawlers con librería Scrapy en Python
6. Pruebas crawlers
7. Carga scripts python en Cloud Functions
8. Carga primeros datos historicos obtenidos de los crawlers en Google Storage
9. [Proceso de entrenamiento de algoritmo con los datos actuales de Storage]
10. Con el modelo de predicción listo, desarrollo API python.
11. Desarrollo frontend HTML, Javascript, PHP
12. Carga iterativa de nuevos datos de forma semanal

Link a Diagrama:

https://drive.google.com/file/d/1Da13paeeAP8r5\_QBE40D4ejuVKxK8lfo/view?usp=sharing