Brainstorm

* Mercado segunda mano y vintage:
  1. Revalorización de prendas de ediciones limitadas pasadas
  2. Conciencia de un consumo más responsable “ya está fabricado, no generemos más prendas”, economía circular
  3. Búsqueda de la calidad de antaño (“la ropa antes duraba más, con más lavados”)
  4. Búsqueda de la exclusividad no a través de “*tengo lo último de Zara/H&M*”, etc. Sino a través de lo único que es tener una prenda de hace XX años *que ya no se fabrica*
* Generar una herramienta que prediga qué prendas, accesorios, van a revalorizarse en base a:
  1. La antigüedad (*este año se cumplen 20 años del bolso XX*)
  2. La exclusividad (*colecciones edición limitada marcas de lujo o colaboraciones con grandes casas tipo las que hace H&M*)
  3. Que esa tendencia esté volviendo a estar de moda (*p ej: prendas amplias, prendas deportivas años 80s*)
     + Se habla de ello en RRSS (#hashtag)
  4. “Joyas ocultas” de vendedores coleccionistas. Y si, además, están a buen precio, su adquisición resulta mucho más satisfactoria para el cliente

Diseño del DAaaS

### Definición de la estrategia del DAaaS

*Definir el catálogo de servicios que proporcionará la plataforma DAaaS, que incluye incorporación de datos, limpieza de datos, transformación de datos, datapedias, bibliotecas de herramientas analíticas y otros.*

La estrategia consiste en ofrecer un servicio de búsqueda (catálogo) de prendas vintage que estén revalorizadas y, además, un servicio de predicción de nuevas tendencias vintage en las que el usuario puede invertir. El catálogo cubrirá todos los sectores de la moda, desde las prendas más básicas y accesibles hasta las grandes marcas.

Es una buena estrategia porque vivimos en un mundo cada vez más consciente en términos de sostenibilidad, en una sociedad que demanda más responsabilidad a las empresas de la industria textil y que está dispuesta a invertir un poco más en las prendas si con ello se garantiza la durabilidad. Todo esto lo aúna el mercado de ropa vintage.

Por otro lado, la estrategia de marketing irá dirigida a la *exclusividad* que aporta al usuario el tener una prenda u accesorio con varios años de antigüedad, que está en perfecto estado y *ya no se fabrica.*

Los datos que incorporar son:

* Publicaciones en Instagram (sus *hashtags* y número de *likes*) relacionadas con ciertas marcas y tendencias de moda.
* Artículos de revistas de moda que hagan referencia a tendencias de prendas y accesorios en general y de moda vintage en particular.
* Páginas web de mercado de segunda mano/vintage de lujo.
* Páginas web de venta de ropa de segunda mano o vintage, mercado medio.

### Arquitectura DAaaS

*Definir la selección de componentes, la definición de procesos de ingeniería y el diseño de interfaces de usuario. Diseño y ejecución de Proofs-of-Concept (PoC) para demostrar la viabilidad del enfoque DAaaS.*

Listado de componentes:

* Scrappers y crawlers para obtener los datasets.
* Subida y almacenamiento de los datasets (tanto en origen como después del procesamiento): Google Cloud Storage.
* Procesamiento de datos: una máquina virtual tipo Google DataProc con Hadoop y HIVE.
* Modelo de predicción: en resumen, es generar una query en HIVE que nos da las prendas más de moda ahora mismo y la predicción de tendencias y revalorizaciones de prendas y accesorios.
* BBDD (Cloud SQL) para alojar el catálogo de la web, en base a los datos arrojados por el Modelo de Predicción.
* Server para la página web (máquina virtual para alojar el servidor web) que se alimenta de ese Cloud SQL.

### DAaaS Operating Model Design and Rollout

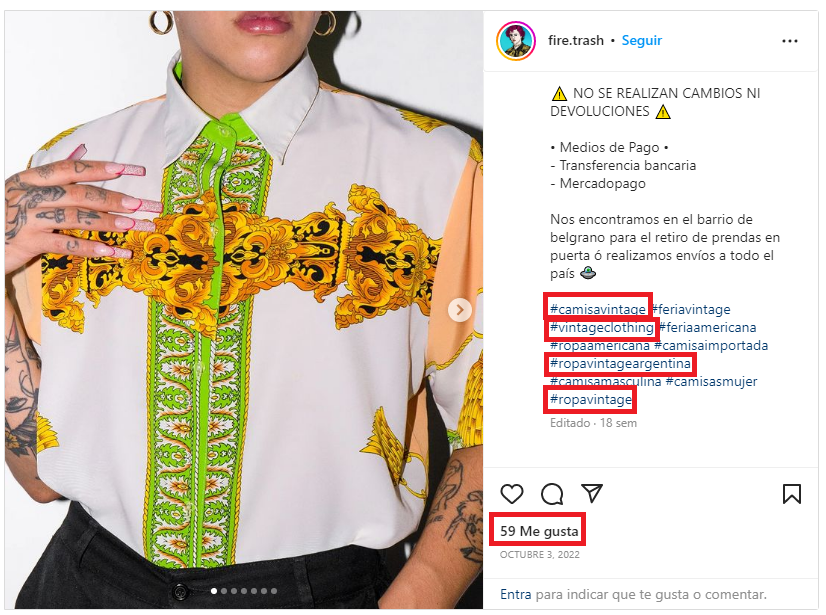
*Personalizar los modelos operativos DAaaS para cumplir con los procesos, la estructura organizacional, las reglas y el gobierno de los clientes individuales. Realizar seguimiento de consumo y mecanismos de informe.*

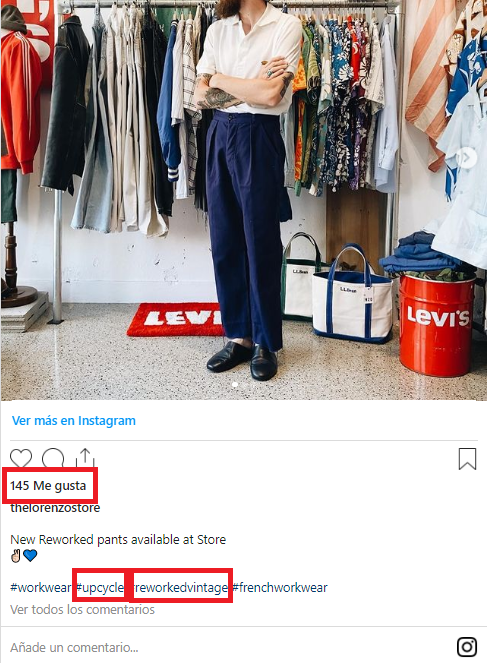
1. [*Google Compute Engine*] Fase extracción de datos:
   1. Scrapping de Instagram 🡪 un empleado lanza el script de Python, a diario
   2. Scrapping y Crawling de las webs mencionadas 🡪 un empleado lanza un script de Python para cada web, a diario
2. [*Google Cloud Storage*] En la fase de extracción se obtienen, gracias al script de Python, varios ficheros .csv, los cuales se almacenan en un Google Bucket 🡪 el script mencionado lo hace, a diario.
3. [*Google DataProc]* Mediante una instancia de Google DataProc, con módulo HIVE incluido, se hacen las “matemáticas” de valoración y de predicción partiendo de los ficheros los datos de los .csv que se han volcado a tablas HIVE🡪 automáticamente con un script bash, a diario.
4. [*Google Cloud Storage]* El fichero de salida de la Query de DataProc (.csv) se almacena en un Bucket 🡪 automáticamente con un script bash, a diario.
5. [*Google Cloud SQL*] Este .csv pasa a alimentar a la BBDD Cloud SQL a través de [*Google Compute Engine*] un script de conversión de .csv a .sql 🡪 un empleado lanza este script de Python, a diario.
6. *[Google Compute Engine*] Esta BBDD es la que alimenta el servidor web (servicio final) con el catálogo de prendas y accesorios 🡪 web se actualiza cada día según la BBDD.

### Desarrollo de la plataforma DAaaS

*Construcción iterativa de todas las capacidades de la plataforma, incluido el diseño, desarrollo e integración,* ***pruebas****, carga de datos, metadatos y población de catálogos, y despliegue.*

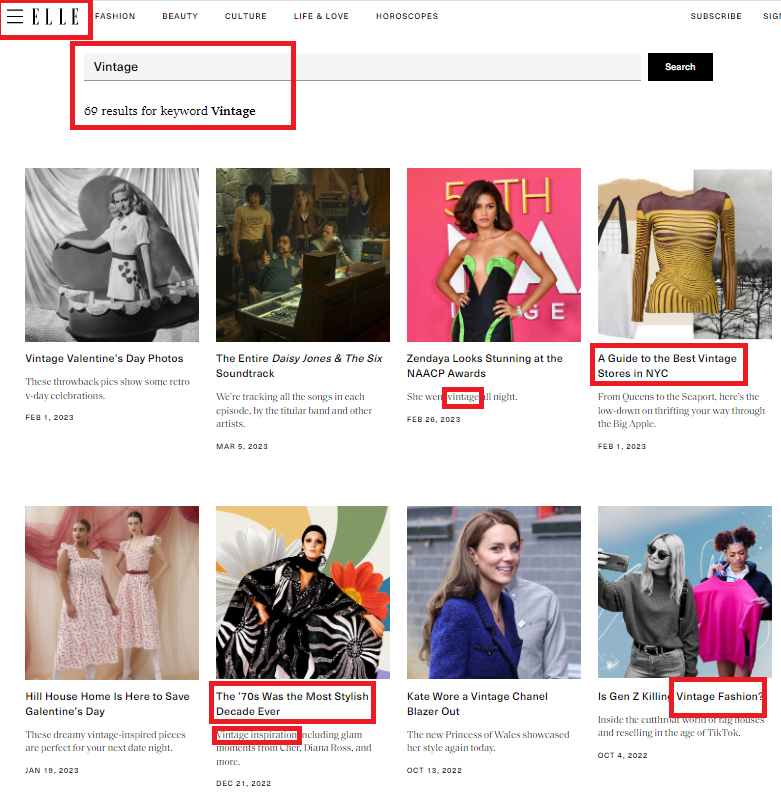
* Scrapper de hashtags en Instagram. Se seleccionarán hashtags relacionados con la moda, la moda vintage, estilismos, *outfits*, etc. También se captará el número de “me gusta” que tiene la publicación





* Crawler + Scrapper de Vogue.com (Vogue EEUU, Vogue Francia, Vogue UK, Vogue España). También de búsquedas en Google de las palabras: *vintage Vogue*
* Crawler + Scrapper de <https://www.elle.com/es/search/?q=vintage#location-choice>, <https://www.elle.com/search/?q=Vintage>,





* Crawler + Scrapper para vestiairecollective.com, whatgoesaroundnyc.com, etsy.com, thredup.com (este mercado sólo disponible en EEUU, pero sirve como marcador de tendencias)
* Scrapper para obtener información (marca, talla, precio…) y fotografías de los artículos que estarán en el catálogo

Link a Diagrama:

<https://github.com/najli-ye/BC-BigDataArchitecture/diagrama_practica.drawio>