

Memoria Técnica: FoodTok Strategic & Predictive Ecosystem 2025

Autor: Joaquim Serra

Fecha: Enero 2026

Proyecto: Final de Bootcamp (Data Science & AI)

Entorno: EE. UU. y Europa (Occidente)

1. Resumen Ejecutivo del Proyecto

Este ecosistema integral de datos está diseñado para transformar el ruido de las redes sociales en **inteligencia comercial accionable**. El proyecto se desmarca de los análisis convencionales al ofrecer una solución en dos capas:

- 1. Capa Estratégica (BI):** Un análisis retrospectivo de 2025 para entender qué funcionó y por qué.
- 2. Capa Operativa (IA):** Una herramienta de predicción y optimización para el contenido que se publicará hoy.

2. Arquitectura del Proyecto (Dual-Layer)

Capa	Enfoque	Temporalidad	Entregable Principal
 Estratégica	Business Intelligence & Market Trends	Histórico (Todo 2025)	Dashboard en PowerBI
 Operativa	Machine Learning & AI Generation	"Live" (Últimos 7 días)	Webapp en Streamlit

2.1. Parte 1: El Dataset 2025 (Estratégica)

- Origen:** Combinación de datasets de Kaggle con extracciones masivas programadas en Apify desde enero de 2025.
- Objetivo:** Identificar los patrones de éxito del año: ¿Qué categorías de cocina dominaron? ¿Qué gadgets de TikTok Shop tuvieron el mayor ROI? ¿Cómo afectó la estacionalidad a las recetas virales?
- Valor:** Proporciona la "foto fija" necesaria para tomar decisiones de inversión a largo plazo.

2.2. Parte 2: El Dataset "Live" (Técnica/Predictiva)

- Origen:** Scraping dinámico mediante Python de los últimos 7 días.
- Objetivo:** Capturar la tendencia del momento (lo que es viral hoy puede no serlo mañana), predecir la viralidad de un nuevo clip y automatizar la creación de guiones ganadores.

- **Valor:** Una herramienta de trabajo diario para creadores y gestores de e-commerce.

3. Stack Tecnológico Detallado

Para garantizar un rendimiento profesional, se ha seleccionado el siguiente conjunto de herramientas:

Fase	Herramientas	Detalle Técnico y Justificación
 Ingestión	 Python + Apify	Uso de proxies residenciales para extraer metadatos de videos, perfiles y productos de TikTok Shop sin bloqueos.
 Almacenamiento	 DuckDB	Base de datos analítica "in-memory". Permite consultas SQL ultrarrápidas sobre archivos Parquet sin configurar un servidor pesado.
 Procesamiento	 Pandas / Polars	Limpieza de datos (data wrangling), manejo de valores nulos, normalización de monedas (\$ a €) y tratamiento de tipos de datos.
 IA Cognitiva	 Google Gemini 1.5	Extracción de entidades (ingredientes, marcas) de las transcripciones y análisis semántico de los "hooks" iniciales.
 ML Engine	 Scikit-Learn	Entrenamiento de modelos XGBoost para el "Viral Score" basándose en métricas de engagement y variables temporales.
 Visualización	 PowerBI	Creación de un modelo de datos en estrella conectado a archivos Parquet para máxima fluidez en los filtros.
 Frontend	 Streamlit	Interfaz de usuario interactiva basada en Python para desplegar el predictor y el optimizador de guiones.

4. Plan de Ejecución (3 Sprints de Alto Rendimiento)

Semana 1: Ingeniería de Datos (El Cimiento)

- **Día 1-2: Setup & Historical Fetch.** Configuración de los scrapers. Primera gran extracción en Apify (USA/EU) para consolidar el "Bloque 2025".
- **Día 3-4: Data Warehouse Local.** Implementación de la base de datos en DuckDB.
 - `tbl_historical_videos` : Metadatos completos (vistas, likes, hashtags).
 - `tbl_historical_products` : Rendimiento de ventas en TikTok Shop.
- **Día 5-7: PowerBI Development.** Diseño de visualizaciones:
 - *Tendencias:* Evolución de hashtags de cocina mes a mes.
 - *E-commerce:* Correlación entre el tipo de receta y el volumen de ventas de gadgets.

Semana 2: Inteligencia y Modelado (El Cerebro)

- **Día 1-2: Live Scraper.** Script automatizado para la captura semanal de datos frescos.
- **Día 3-4: Feature Engineering.** Generación de variables para el modelo de ML:
 - *Velocity Score:* Velocidad de interacción por hora.
 - *Hook Power:* Análisis mediante Gemini de los primeros 3 segundos del audio/texto.
- **Día 5-7: Machine Learning.** Entrenamiento y testeo de modelos:
 - *Predictor:* Estimación de visualizaciones según la hora y el nicho.
 - *Recomendador:* Detección de la ventana óptima de publicación.

Semana 3: IA Generativa y Despliegue (El Producto)

- **Día 1-2: Gemini Integration.** Desarrollo del "Prompt System" para el refinamiento de guiones.
- **Día 3-5: Streamlit Webapp.** Construcción de la interfaz con 3 secciones:
 1. *Live Trends:* Qué se está cocinando hoy en TikTok.
 2. *Predictor:* Sube tu idea y recibe una puntuación de éxito.
 3. *IA Creator:* Generación de guiones con "hooks" optimizados por Gemini.
- **Día 6-7: Documentación y Demo.** Preparación del README, grabación de la funcionalidad y cierre del proyecto.

5. Gestión de Tareas (Trello Board)

Puedes copiar estas tareas directamente a tus columnas de Trello:

COLUMNA 1: ESTRATÉGICO (PowerBI - 2025)

- [] Extraer Dataset 2025 (Apify)
- [] Normalizar datos de TikTok Shop (Limpieza de moneda y ratings)
- [] Crear medidas DAX (Engagement Rate, % Growth)
- [] Diseñar UI del Dashboard (Dark Mode estilo TikTok)

COLUMNA 2: OPERATIVO (Streamlit - Live)

- [] Programar Script de Scraping Semanal
- [] Pre-procesamiento de datos para ML (One-Hot Encoding, Escalado)
- [] Entrenar Modelo de Clasificación de Viralidad
- [] Conectar API de Google Gemini

COLUMNA 3: PRESENTACIÓN & DOCUMENTACIÓN

- [] Redactar Memoria Técnica final
- [] Crear video de demostración (Demo Reel)
- [] Desplegar App en Streamlit Cloud

💰 6. Presupuesto y Recursos (Máx 30€)

1. 💚 **Apify Credits (~20€)**: Pago puntual para la extracción de los datasets base (Histórico + Semanal).
2. 💚 **Google Cloud / Gemini API (~10€)**: Saldo para el procesamiento de lenguaje natural. Gemini 1.5 Flash es extremadamente eficiente en costes.
3. 💡 **Herramientas Gratuitas**: PowerBI Desktop, Streamlit Cloud, GitHub, DuckDB.

💡 7. Factor Diferencial para el Bootcamp

Al presentar este proyecto, tu discurso debe centrarse en la **utilidad real**:

"No solo presento un análisis de datos; presento un sistema capaz de predecir el futuro del contenido gastronómico. Mientras que el dashboard de PowerBI nos enseña las lecciones de 2025, la aplicación de IA guía al creador paso a paso para que su próximo video de cocina sea un éxito de ventas basándose en los datos capturados hace menos de 24 horas."