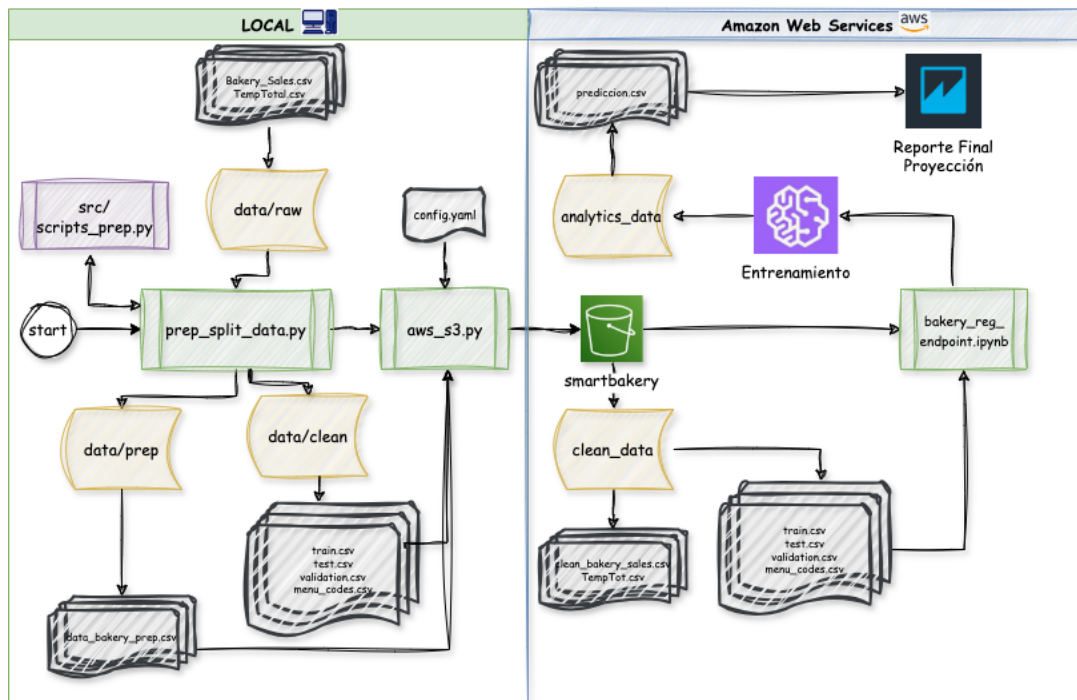


## Diagrama de la Arquitectura de SmartBakery



La arquitectura presentada en el diagrama detalla el flujo de trabajo completo de SmartBakery, diseñado para optimizar la gestión del inventario en panaderías utilizando técnicas avanzadas de ciencia de datos y aprendizaje automático. La estructura se divide en dos entornos principales: el entorno local y el entorno de Amazon Web Services (AWS), cada uno desempeñando roles específicos en la preparación, procesamiento y análisis de los datos.

### Entorno Local

El flujo de trabajo comienza en el entorno local con la recopilación de datos crudos. Estos datos, contenidos en archivos como **Bakery\_Sales.csv** y **TempTotal.csv**, proporcionan información histórica de ventas y Temperatura promedio en Loire-Atlantique, Francia. El script ubicado en **src/scripts\_prep.py** contiene funciones para limpieza y separación de los datos, es utilizado por el siguiente script y toma los datos del directorio **data/raw**.

El siguiente paso es llevado a cabo por el archivo **prep\_split\_data.py**, que realiza una preparación y división más detallada de los datos. Este script organiza los datos en dos directorios: **data/prep** para los datos limpios y ordenados y **data/clean** para los datos separados en entrenamiento, validación y prueba. La transferencia de estos datos a la nube se hace mediante el script **aws\_s3.py**, que utiliza las configuraciones

especificadas en **config.yaml** para subir los datos al servicio de almacenamiento Amazon S3.

### **Entorno en la Nube (AWS)**

Una vez que los datos están en Amazon S3, comienza la fase de procesamiento en la nube. Los datos limpios almacenados en el bucket **smartbakery** son utilizados para el entrenamiento y análisis. Los conjuntos de datos son organizados en **clean\_data**, y contienen los fundamentales para el entrenamiento del modelo de aprendizaje automático.

Estos datos incluyen archivos esenciales como **train.csv**, **test.csv**, **validation.csv** y **menu\_codes.csv**, que son utilizados para desarrollar y ajustar el modelo predictivo. El entrenamiento del modelo se lleva a cabo mediante el script **bakery\_reg\_endpoint.ipynb**, que proporciona predicciones sobre la demanda de productos.

Una vez entrenado, el modelo genera predicciones que se almacenan en **analytics\_data**, produciendo un archivo de predicciones llamado **prediccion.csv**. Estos resultados se analizan y se visualizan en un dashboard de QuickSight.

Esta arquitectura asegura que SmartBakery puede manejar grandes volúmenes de datos de manera eficiente, aprovechando las capacidades de escalabilidad y procesamiento en la nube de AWS. La combinación de procesamiento local y en la nube permite una preparación de datos flexible y un análisis robusto, garantizando que las panaderías puedan optimizar su inventario, reducir desperdicios y mejorar la satisfacción del cliente mediante decisiones informadas y estratégicas basadas en datos precisos y actualizados.