**Informe: Evaluación de Calidad de Datos Públicos para Forecast de Demanda**

**Curso:** Gestión y Gobierno de Datos  
**Estudiante:** Neftali James Valdivia Vara

**1. Objetivo del Informe**

El presente informe tiene como finalidad documentar el proceso de extracción y evaluación de datos públicos relevantes (noticias, clima y tipo de cambio), aplicables al caso de estudio del curso: *forecast* de demanda y estrategia de precios ante escasez de inventario.

**2. Proceso de Extracción de Datos Públicos**

Se emplearon tres fuentes abiertas con acceso vía API o feeds, seleccionadas por su relevancia y accesibilidad:

| **Categoría** | **Fuente/API Utilizada** | **Variables Obtenidas Principales** |
| --- | --- | --- |
| Noticias | Feeds RSS: El Comercio, La República, Gestión | Título, resumen, fecha de publicación, enlace, fuente |
| Clima | Open-Meteo API | Temperatura, velocidad del viento, código de clima, ciudad |
| Tipo de Cambio | e-api.net.pe | Compra, venta, fecha, origen del dato |

🔒 *Nota:* Las credenciales de acceso (API Keys) no han sido incluidas en el repositorio, conforme a las buenas prácticas de seguridad.

**3. Indicadores de Calidad Evaluados**

Se eligieron cuatro indicadores de calidad de datos alineados con la finalidad del proyecto:

| **Indicador** | **Definición** | **Relevancia en el Proyecto de Forecast** |
| --- | --- | --- |
| Completitud | Verifica que los datos no contengan campos nulos o vacíos. | Datos incompletos afectan la precisión de los modelos. |
| Frescura | Evalúa qué tan actualizados están los datos. | Información desactualizada puede llevar a decisiones erróneas. |
| Relevancia | Determina si los datos se relacionan con los objetivos. | Permite trabajar con variables que realmente impactan la demanda. |
| Confiabilidad | Mide si la fuente es confiable, válida y consistente. | Asegura resultados sólidos para decisiones estratégicas. |

**4. Resultados de la Evaluación de Calidad**

**a) Noticias (RSS – Medios Peruanos)**

* **Completitud:** Baja. Algunos campos como el resumen aparecen truncados.
* **Frescura:** Alta. Noticias actualizadas varias veces al día.
* **Relevancia:** Media. Se filtraron por palabras clave (economía, mercado, etc.).
* **Confiabilidad:** Alta. Medios reconocidos.

**b) Clima (Open-Meteo)**

* **Completitud:** Alta. Información estructurada por ciudad y variable.
* **Frescura:** Alta. Datos en tiempo real.
* **Relevancia:** Alta. Temperaturas altas pueden influir en productos estacionales.
* **Confiabilidad:** Alta. Fuente especializada.

**c) Tipo de Cambio (e-api.net.pe)**

* **Completitud:** Alta.
* **Frescura:** Alta. Actualización diaria.
* **Relevancia:** Alta. Impacto directo en precios e importaciones.
* **Confiabilidad:** Alta. Fuente oficial validada (SBS/SUNAT).

**5. Conclusiones y Recomendaciones**

* Las fuentes de **clima** y **tipo de cambio** presentan excelente calidad para proyectos de analítica de demanda y precios.
* La fuente de **noticias** requiere preprocesamiento adicional (por ejemplo, técnicas NLP) para mejorar su completitud y relevancia.
* Se recomienda establecer un sistema de monitoreo continuo de indicadores clave (especialmente frescura y completitud).
* La metodología aplicada puede extenderse a otras áreas como logística, ventas y distribución.

**6. Justificación de los Indicadores de Calidad**

Cada indicador fue elegido considerando su aporte directo a la confiabilidad de los modelos analíticos:

* **Completitud:** Minimiza el riesgo de errores por ausencia de datos críticos.
* **Frescura:** Asegura que las decisiones se tomen con información actual.
* **Relevancia:** Permite enfocar los análisis en variables significativas.
* **Confiabilidad:** Refuerza la confianza en los resultados obtenidos.

**7. Evaluación Técnica Complementaria**

| **Pregunta** | **Respuesta** |
| --- | --- |
| ¿Cómo evalúas la calidad de datos? | Mediante validaciones de campos nulos, fechas de actualización y reglas de negocio. |
| ¿Cómo sabes que es confiable? | Uso de fuentes oficiales y respuesta con estructura válida (JSON/XML, HTTP 200). |
| ¿Cuáles son los indicadores usados? | Completitud, frescura, relevancia y confiabilidad. |
| ¿Qué tan actualizada está la información? | Noticias: cada hora. Clima: en tiempo real. Tipo de cambio: diario. |
| ¿Cuál es el periodo de actualización? | Noticias: horas. Clima: por consulta. Tipo de cambio: 24 horas. |
| ¿Qué tan relevante es para el negocio? | Todas las fuentes tienen relación directa con demanda, clima y precios. |
| ¿Cómo mides la relevancia? | Reglas de negocio (ej. palabras clave, umbrales de temperatura, etc.). |
| ¿Es una buena fuente para proyectos de Data Science? | Sí. Son accesibles, estructuradas y compatibles con Python/Pandas. |
| ¿Puedes gobernarlas correctamente? | Sí. Mediante automatización, validación y trazabilidad. |
| ¿La API es correcta? | Sí. Cumple con estructura, respuesta esperada y no requiere autenticación compleja. |