**Informe de Consumo de Tokens y Costos en el Proceso de Conversión de Reportes**

**Flujo del proceso**

1. **Reporte -> Imágenes:**
   * Cada reporte contiene entre 5 y 15 páginas, y cada página se convierte en una imagen.
   * El proceso de transformación de cada imagen a JSON consume aproximadamente **40,000 tokens**.
2. **Imágenes -> JSON:**
   * Cada imagen convertida a JSON utiliza un total de **40,000 tokens**.
     + De estos, **37,192 tokens** corresponden a los **tokens de entrada** (prompt).
     + **982 tokens** corresponden a los **tokens de salida** (respuesta o completion).
3. **JSONs -> Informe:**
   * Una vez obtenidos los JSONs, se generan informes.
   * Cada informe tiene un texto generado de 180 palabras por imagen, lo que equivale a **aproximadamente 240 tokens de salida** por imagen (considerando que 1 palabra en inglés equivale a unos 1.33 tokens).

**Desglose de Precios**

* **Modelo:** GPT-4o-mini.
* **Costo de tokens de entrada (input tokens):** $0.150 por cada millón de tokens.
* **Costo de tokens cacheados (cached input tokens):** $0.075 por cada millón de tokens.
* **Costo de tokens de salida (output tokens):** $0.600 por cada millón de tokens.

**Nota sobre "cached input tokens":** Los tokens cacheados se refieren a los tokens que se han utilizado en solicitudes repetidas, es decir, cuando el mismo input se envía varias veces y el modelo no necesita procesarlo de nuevo desde cero. El precio reducido se aplica porque el procesamiento para estos tokens es menos intensivo en recursos.

**Cálculo detallado del consumo de tokens y costos**

El consumo de tokens y el costo dependen de la cantidad de páginas (imágenes) en cada reporte, que puede variar entre 5 y 15. A continuación, se detalla el consumo y el costo para diferentes cantidades de imágenes:

1. **Consumo por imagen:**
   * Tokens de entrada (prompt): **~37,192 tokens**.
   * Tokens de salida (respuesta/completion): **~982 tokens**.
   * **Total por imagen (entrada + respuesta): ~40,000 tokens**.
2. **Consumo por texto del informe:**
   * Tokens por informe de 180 palabras por imagen: **~240 tokens de salida**.
   * Para un reporte con nnn imágenes, el total de tokens para el informe sería: n×240n \times 240n×240.
3. **Consumo total por reporte:**
   * Si un reporte tiene nnn imágenes, el total de tokens consumidos será:

Total de tokens de entrada=n×(37,192+982) x 0.15$/M

Total de tokens de salida=n×(982+240) x 0.6$/M