

Introducción a la Inteligencia Artificial – Ventana Contextual

Impacto en la analítica de datos

Objetivos de la sesión (ajustados)

1 Comprender el concepto de ventana contextual

Analizar qué significa, cómo varía entre diferentes modelos de IA y por qué determina el alcance del análisis de datos que se puede realizar.

3 Explorar el funcionamiento de Copilot en Excel y Microsoft 365

Conocer cómo se integra en el flujo de trabajo empresarial, cómo aprovecha las estructuras de datos ya disponibles y qué ventajas ofrece frente a herramientas externas.

2 Identificar limitaciones y fortalezas del uso de IA en analítica de datos

Reconocer cuándo confiar en Copilot, ChatGPT o Claude, y cuáles son los riesgos más comunes (sesgo, pérdida de contexto, restricciones de datos corporativos).

4 Aplicar casos prácticos de Copilot y Copilot Chat en escenarios de negocio reales

Desde limpieza y transformación de datos, hasta generación de dashboards y validación de hipótesis.

Competencias a desarrollar

Análisis crítico

Identificar qué tipo de IA conviene según el problema empresarial.

Competencia digital aplicada

Usar Copilot en Excel y Copilot Chat como asistentes de análisis de datos.

Pensamiento sistémico

Integrar resultados de la IA en procesos de toma de decisiones empresariales.

Comunicación de resultados

Transformar datos en insights accionables para equipos directivos.

Ventana Contextual: concepto y relevancia

La **ventana contextual** es la **capacidad de memoria activa de un modelo de IA**: representa la cantidad máxima de información (texto, datos, instrucciones) que el modelo puede procesar y tener "presente" al mismo tiempo para generar una respuesta coherente. Se mide en **tokens**, que son fragmentos de palabras o números.

Ejemplo sencillo:

Si un modelo tiene una ventana contextual de 4.000 tokens (\approx 3.000 palabras), solo podrá "recordar" y razonar con ese rango de información dentro de una consulta. Todo lo que quede fuera de esa ventana queda "olvidado" para el análisis inmediato.

¿Por qué es importante conocer este concepto?

Límites en el análisis de datos

Aunque un modelo parezca "inteligente", si el dataset excede su ventana contextual, tendrá que trabajar con fragmentos. Esto puede llevar a **pérdida de contexto** (ej. un análisis parcial de las ventas que no considere todos los trimestres).

Precisión y confianza en los resultados

Entender esta limitación permite evitar **conclusiones erróneas**. Por ejemplo, pedir un resumen de un archivo muy grande a ChatGPT o Claude puede dar una respuesta incompleta si el archivo excede la ventana.

Define una estrategia antes

Cuando un archivo de datos supera el tamaño de la ventana contextual del modelo, ocurren las siguientes consecuencias:

- **Pérdida de Información Crítica:** El modelo no puede "ver" todo el archivo a la vez. Esto significa que si le pides un resumen o un análisis que requiere comprender la totalidad de los datos, la respuesta será incompleta o incorrecta porque se basará únicamente en el fragmento que cabe en su "memoria". Por ejemplo, al analizar un registro de ventas anual que excede la ventana, el modelo podría ignorar las ventas de los primeros meses.
- **Fragmentación del Análisis:** Para procesar el archivo, es necesario dividirlo en partes más pequeñas. Esto obliga al modelo a trabajar con fragmentos aislados, perdiendo la relación y el contexto entre ellos. Un análisis de tendencias o la identificación de patrones a lo largo de todo el conjunto de datos se vuelve prácticamente imposible.

Entiende tu set de datos

Un conjunto de datos desordenado, con inconsistencias, errores o falta de estructura, impacta negativamente la ventana contextual de las siguientes maneras:

- **"Contaminación" de la Ventana Contextual:** Información irrelevante, mal formateada o errónea ocupa un espacio valioso en la ventana contextual. Esto es como intentar tener una conversación coherente en una habitación llena de ruido. El modelo tiene que esforzarse más para discernir los datos útiles de la "basura", lo que aumenta la probabilidad de generar respuestas incorrectas o sin sentido.
- **Dificultad para Identificar Patrones:** Los modelos de lenguaje son excelentes para encontrar relaciones en datos bien estructurados. Si los datos están desordenados (por ejemplo, fechas en diferentes formatos, nombres con errores tipográficos, columnas desalineadas), el modelo gastará gran parte de su capacidad de procesamiento tratando de entender la estructura en lugar de analizar el contenido. Esto reduce la calidad del análisis.

Múltiples Hojas de Excel

Trabajar con datos distribuidos en varias hojas de cálculo introduce una capa adicional de complejidad que afecta la ventana contextual:

- **Incapacidad para Correlacionar Datos Automáticamente:** Un modelo de lenguaje no puede, por defecto, "ver" y cruzar información entre diferentes hojas de un archivo Excel de forma simultánea si la información combinada excede la ventana contextual. El usuario debe guiar explícitamente al modelo, a menudo cargando los datos de cada hoja por separado.
- **Sobrecarga Contextual con Información Redundante:** Para analizar datos entre hojas, es necesario explicarle al modelo la estructura de cada una y cómo se relacionan. Esta "meta-información" (los nombres de las columnas de cada hoja, las claves de relación, etc.) también consume espacio en la ventana contextual, dejando menos lugar para los datos reales que se van a analizar.
- **Aumento del Riesgo de Errores por Mala Interpretación:** El modelo puede confundirse fácilmente sobre qué hoja contiene qué información, especialmente si las columnas tienen nombres similares. Esto puede llevar a que se realicen cálculos incorrectos o se extraigan conclusiones erróneas al mezclar datos de diferentes contextos (hojas). Es necesario un esfuerzo adicional y una descripción muy clara por parte del usuario para evitar estas confusiones.

Tabla comparativa de IA para análisis de datos

Herramienta	Volumen de datos soportado	Integración nativa con Excel/365	Capacidades de análisis estadístico	Fortalezas clave	Limitaciones clave
Copilot Chat	Datos pequeños a medianos (~100k filas)	Alta (Excel, Word, PowerPoint, Teams)	Bueno en análisis descriptivo	Integración M365, seguridad corporativa	Depende de calidad del dataset
Copilot 365	Similar a Chat, con contexto organizacional	Muy alta (ecosistema completo)	Descriptivo y predictivo ligero	Contexto empresarial, políticas de seguridad	Menos flexible que LLMs abiertos
ChatGPT (GPT-4o)	Alto (datasets resumidos en texto)	Limitada (requiere plugins/APIs)	Avanzado (estadística y ML vía prompts)	Versátil, creativo, fuerte en ML	No acceso directo a Excel/365
Claude (Opus)	Muy alto (hasta 1M tokens)	Limitada (requiere API/Notion)	Excelente en análisis textual	Gran capacidad de contexto y documentos largos	Menos afinado en cálculos avanzados

Comparativa de modelos por ventana contextual

La ventana contextual determina cuánta información puede manejar un modelo en una sola conversación o análisis sin perder coherencia. Esto incluye texto, datos tabulares, instrucciones, y documentos largos.

Modelo	Capacidad contextual estimada	Fortalezas en contexto largo	Debilidades en contexto largo
Claude (Opus)	 Hasta 1 millón de tokens	Excelente para documentos extensos, múltiples archivos, y análisis narrativo profundo	Menos preciso en cálculos numéricos complejos o fórmulas técnicas
ChatGPT (GPT-4o)	 ~128k tokens (con plugins)	Muy bueno en razonamiento estadístico y ML con contexto medio	Requiere estructuración previa si el input es muy largo o desordenado
Copilot Chat	 ~100k filas en Excel (equivalente a ~20k tokens)	Ideal para tareas estructuradas en Excel, Word, PowerPoint	No diseñado para mantener múltiples documentos o narrativas extensas
Copilot 365	 Similar a Copilot Chat, pero con contexto organizacional	Puede usar contexto empresarial (correo, Teams, Planner) para enriquecer respuestas	No retiene grandes volúmenes de texto técnico o histórico en una sola sesión

Ranking por ventana contextual

1

Más fuerte: Claude Opus

Capacidad masiva de tokens, ideal para análisis de documentos largos, múltiples PDFs, y síntesis de información compleja.

2

Intermedio: ChatGPT (GPT-4o)

Muy competente en razonamiento técnico y estadístico, pero con límites si se le sobrecarga sin estructura.

3

Copilot Chat y Copilot 365

Aunque muy útiles en entornos Microsoft, su ventana contextual está optimizada para tareas específicas, no para análisis narrativos extensos.

Comparativa de modelos por ventana contextual y tamaño de archivo

Modelo	Capacidad contextual estimada	Tamaño de archivo aprox. (MB)	Fortalezas en contexto largo	Debilidades en contexto largo
Claude (Opus)	 Hasta 1 millón de tokens	 ~500–700 MB (texto plano)	Excelente para múltiples documentos, síntesis profunda y análisis narrativo	Menos preciso en cálculos técnicos complejos
Grok 4	 256,000 tokens	 ~120–150 MB	Muy fuerte en razonamiento estructurado y funciones empresariales	No admite parámetros como reasoning_effort; menos flexible en ajustes avanzados
ChatGPT (GPT-4o)	 ~128k tokens	 ~60–80 MB	Muy bueno en ML, estadística y razonamiento técnico	Requiere estructura clara si el input es largo o desordenado
Copilot Chat	 100k filas en Excel (20k tokens)	 ~15–20 MB	Ideal para tareas estructuradas en Excel, Word, PowerPoint	No diseñado para mantener múltiples documentos o narrativas extensas
Copilot 365	 Similar a Copilot Chat, pero con contexto organizacional	 ~15–25 MB	Usa contexto empresarial (correo, Teams, Planner) para enriquecer respuestas	Limitado en análisis técnico profundo o documentos largos

Notas técnicas:

- Los tamaños están basados en texto plano (UTF-8). Archivos con imágenes o formatos enriquecidos pueden pesar más sin aportar tokens útiles.
- Grok 4 es un modelo de razonamiento puro, lo que lo hace ideal para tareas empresariales complejas, pero menos configurable en ciertos parámetros.

Casos de uso en Excel con Copilot

1

Resaltar lo revelante

Prompt: "Resalta el 20% superior de la columna ventas."

1

Transformación de fechas

Prompt: "Convierte las fechas en formato DD/MM/AAAA a AAAA-MM-DD."

1

Análisis descriptivo

Prompt: "Resume las métricas clave: promedio, mediana, desviación estándar de ventas mensuales."

1

Identificación de outliers

Prompt: "Resalta los valores atípicos en la columna de ingresos."

1

Dashboard automático

Prompt: "Crea un dashboard con gráficos de barras, líneas y KPI para este dataset."

1

Modelo de predicción simple

Prompt: "Usa regresión lineal para predecir ventas según inversión en marketing."

1

Simulación de escenarios

Prompt: "Simula el impacto en rentabilidad si aumentamos precios un 10%."

1

Segmentación de clientes

Prompt: "Agrupa clientes por volumen de compra y frecuencia."

1

Análisis de correlación

Prompt: "Calcula la correlación entre precio, cantidad y rentabilidad."

Casos de uso en Copilot Chat aplicado a datos

Explicación de fórmulas complejas

Prompt: "¿Qué hace la fórmula =FORECAST.LINEAR(B2:B13, C2:C13)?"

Conversión de instrucciones en pasos

Prompt: "Quiero calcular el ROI por producto. ¿Qué pasos debo seguir en Excel?"

Generación de insights resumidos

Prompt: "Resume los hallazgos clave de este dataset de ventas."

Hipótesis exploratorias

Prompt: "Genera 5 hipótesis sobre por qué bajaron las ventas en Q2."

Traducción de lenguaje de negocio a fórmulas

Prompt: "Quiero saber cuántos clientes compran más de 3 veces al mes. ¿Qué fórmula uso?"

Generación de código para análisis

Prompt: "Escribe una fórmula para calcular el margen bruto por producto."

Comparación entre periodos

Prompt: "Compara el rendimiento de ventas entre Q1 y Q2."

En el archivo peces

1. Análisis descriptivo básico

- **Variables numéricas:** AÑO, MES, Latitude, Longitude, V_Economico, count, TOTAL.
 - Ejemplo: count tiene mediana 0 y máximo 225; TOTAL tiene mediana 5 y máximo 235.
- **Variables categóricas:** Date, Periodo, Estacion_clima, Station_ID, Station_Name, Data_Freq, Vertiente, Familia, Especies, CCODE, Ambiente, Miller_66_76.
 - Ejemplo: 116 especies distintas, 34 familias, 4 vertientes hidrográficas.

2. Tendencias temporales

- Evolución de **TOTAL** y **count** por año y mes.
- Comparación entre periodos (**2004-2016** vs **2017-2025**) y estaciones climáticas (**LLUVIOSA** vs **SECA**).
- Series de tiempo de abundancia, riqueza y proporción de especies con valor económico.

3. Agrupaciones y comparaciones por categorías

- Por **vertiente**: diferencias en abundancia (TOTAL) y número de ejemplares (count).
- Por **ambiente** (D vs M), **familia**, **especie**, **estación climática**, etc.
- Ejemplo: la vertiente "I" tiene mayor media de TOTAL (18.4) que las demás.

4. Relaciones entre variables

- Correlaciones numéricas: la más fuerte es entre `count` y `TOTAL` (~ 0.21).
- Relación entre `V_Economico` y `TOTAL`, entre especies y estaciones, etc.
- Posibles análisis multivariantes (regresión, clustering, PCA).

5. Visualización geoespacial

- Mapas con coordenadas (Latitude, Longitude) para mostrar:
 - Abundancia por estación.
 - Distribución de especies o familias.
 - Comparación entre ambientes o periodos.

6. Limpieza y tratamiento de datos

- Nulos en Latitude, Longitude, Station_Name, Data_Freq.
- Conversión de Date a formato datetime.
- Codificación de variables categóricas para modelos predictivos.

7. Estudios ecológicos

- **Riqueza de especies** por año, estación, vertiente.
- **Diversidad:** índices de Shannon, Simpson.
- Curvas de acumulación de especies.

8. Modelos predictivos y exploratorios

- Regresión para predecir **TOTAL** o **count**.
- Clustering de estaciones por composición de especies.
- PCA para reducción de dimensionalidad.