



Julius

Julius.ia : Herramienta IA generativa para el análisis y visualización de datos

Curso: Herramientas de inteligencia Artificial para apoyar el proceso de Investigación

Cursos de perfeccionamiento profesional – receso académico 2025

Organizado por: **Dirección de investigación - Universidad Tecnológica de Panamá**

Autores:

Danny Murillo González

Centro de Investigación, Desarrollo e Innovación en Tecnologías de la Información y las Comunicaciones - CIDITIC

<https://orcid.org/0000-0003-0297-7213>

<https://scholar.google.es/citations?user=YNx08l0AAAAJ&hl=es>

Sucel López

Centro de Investigación, Desarrollo e Innovación en Tecnologías de la Información y las Comunicaciones - CIDITIC

<https://orcid.org/0000-0002-7154-5265>

<https://scholar.google.es/citations?user=nXDOuN4AAAAJ&hl=es>

Febrero 2025



Esta obra está bajo licencia internacional Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0.

Contenido

- Introducción a Julius.AI..... 6
 - Características clave de Julius AI 6
- Guía de formatos de datos 6
 - Formato de datos adecuado 7
- Pasos para el análisis de datos..... 8
 - Estructura de datos..... 8
 - Evaluar calidad de los datos – Limpieza de datos 8
 - Exploración de estructura de datos 8
 - Exploración de datos..... 9
- Acceso a Julius.ai 10
 - Cargar datos..... 11
 - Chatear con los datos cargados 11
- Visualización de datos y análisis exploratorio con Julius..... 13
 - Ejemplo 1: Resumen de datos..... 13
 - ¿Qué deseo hacer?..... 13
 - Prompt: 13
 - Resultados 13
 - Ejemplo 2: Resumen de datos categóricos..... 15
 - ¿Qué deseo hacer?..... 15
 - Prompt: 15
 - Resultados 15
- Análisis exploratorio univariante..... 15
 - Análisis exploratorio univariante de variables categóricas..... 15
 - Ejemplo 3a: Gráfico de sectores (pastel) 16
 - ¿Qué deseo hacer?..... 16
 - Prompt: 16
 - Resultados 16
 - Interpretación: 16
 - Ejemplo 3b: Gráfico de sectores (valores) 17
 - ¿Qué deseo hacer?..... 17

| | |
|--|----|
| Prompt: | 17 |
| Resultados | 17 |
| Ejemplo 3c: Gráfico de sectores (ordenados) | 17 |
| ¿Qué deseo hacer?..... | 17 |
| Prompt: | 18 |
| Resultados | 18 |
| Gráfico de barras - frecuencia..... | 18 |
| Ejemplo 4a: Gráfico de barras | 19 |
| ¿Qué deseo hacer?..... | 19 |
| Prompt: | 19 |
| Resultado | 19 |
| Interpretación (Julius): | 19 |
| Ejemplo 4b: Gráfico de barras ordenada | 20 |
| ¿Qué deseo hacer?..... | 20 |
| Prompt: | 20 |
| Resultado | 20 |
| Ejemplo 4c: Gráfico de barras ordenada con énfasis..... | 21 |
| ¿Qué deseo hacer?..... | 21 |
| Prompt: | 21 |
| Resultado | 21 |
| Interpretación (Julius): | 22 |
| Análisis exploratorio univariante, variable cuantitativa | 22 |
| Gráfico de histograma - frecuencia..... | 22 |
| Ejemplo 5: Gráfico de histograma..... | 22 |
| ¿Qué deseo hacer?..... | 22 |
| Prompt: | 22 |
| Resultado | 23 |
| Interpretación (Julius): | 23 |
| El gráfico de cajas y bigotes (boxplot) | 23 |
| Ejemplo 6: Gráfico de boxplot | 24 |
| ¿Qué deseo hacer?..... | 24 |
| Prompt: | 24 |
| Resultado | 24 |

| | |
|--|----|
| Interpretación 1 (Julius): | 24 |
| Interpretación 2 (Julius): | 25 |
| Análisis exploratorio bivalente..... | 25 |
| Análisis exploratorio bivariantes de variables categóricas | 25 |
| Ejemplo 7: Gráfico de barras con frecuencias conjuntas | 25 |
| ¿Qué deseo hacer?..... | 25 |
| Prompt: | 25 |
| Resultado | 26 |
| Interpretación (Julius): | 26 |
| Ejemplo 8: Gráfico de barras con frecuencias conjuntas apiladas..... | 27 |
| ¿Qué deseo hacer?..... | 27 |
| Prompt: | 27 |
| Resultado | 27 |
| Análisis exploratorio bivariantes e variables cuantitativas | 27 |
| Ejemplo 9: Gráfico de barras con frecuencias conjuntas | 27 |
| ¿Qué deseo hacer?..... | 27 |
| Prompt: | 28 |
| Resultado | 28 |
| Interpretación: | 28 |
| Gráfico de dispersión (Gráfico de puntos) | 29 |
| Ejemplo 10: Gráfico de barras con frecuencias conjuntas | 29 |
| ¿Qué deseo hacer?..... | 29 |
| Prompt: | 29 |
| Resultado | 30 |
| Ejemplo 11: Gráfico de barras con frecuencias conjuntas | 30 |
| ¿Qué deseo hacer?..... | 30 |
| Prompt: | 30 |
| Resultado | 31 |
| Análisis de Clustering | 32 |
| Ejemplo 12: Análisis de clustering (calculo del número de clusters)..... | 32 |
| ¿Qué deseo hacer?..... | 32 |
| Prompt: | 32 |
| Resultado | 32 |

| | |
|---|----|
| Interpretación: | 32 |
| Ejemplo 13 : Análisis de clustering | 33 |
| ¿Qué deseo hacer?..... | 33 |
| Prompt: | 33 |
| Resultado | 33 |
| Ejemplo 14 : Biplot | 34 |
| ¿Qué deseo hacer?..... | 34 |
| Prompt: | 34 |
| Resultado | 34 |
| Interpretación: | 34 |

Introducción a Julius.AI

Julius.ai (<https://julius.ai/>) es una herramienta de análisis de datos basada en IA, impulsada por los mismos grandes modelos de lenguaje y funciona como un chatbot similar a ChatGPT. Su poder reside en su capacidad de hacer que el análisis de datos sea accesible y práctico, incluso para aquellos que no son científicos de datos ni estadísticos.

Proporciona a los usuarios una plataforma para analizar y visualizar datos de hojas de cálculo en formatos como, (.xls, .xlsx, .xlsm, .csv, .xlsb), Google Sheets y bases de datos PostgreSQL (a través de un plugin).

La plataforma ofrece funciones avanzadas para tareas complejas como modelado, proyecciones, regresión lineal y creación de animaciones de datos. También puede analizar hojas de cálculo con varias pestañas y vincular fuentes de datos directamente al sistema de inteligencia artificial. Una vez que vincula una fuente de datos al sistema, puede utilizar una interfaz de procesamiento de lenguaje natural para pedirle a Julius información sobre los datos.

Esta herramienta tiene diferentes opciones de registro, la opción gratuita que es restrictiva, permite escasamente 15 Prompts o prompt durante un mes. Tomando en cuenta que cada consulta que se realice corresponde a un Prompt. Luego está la versión de pago, que tiene tres opciones: **Lite**, 250 Prompts al mes, **Estándar**, Prompts ilimitados y aumento de memoria en el análisis, **Teams**, Prompts ilimitados, añadir otros usuarios y controles de administración de usuarios.

Características clave de Julius AI

Para ayudarle a comprender más claramente de qué es capaz Julius, estas son sus características claves:

1. Chatea con tus datos
2. Crear gráficos y representaciones visuales
3. Crear animaciones y convertir visualizaciones en GIF
4. Motor de respuesta de IA
5. Construir modelos de pronóstico
6. Realizar análisis avanzado

Los resultados de cada análisis de datos y visualizaciones en Julius.ai. muestran el código de programación en el lenguaje Python o R (según selección). Este código puede ser editado o copiado y llevarse a otra plataforma para ejecutarlo. En este documento mostraremos el resultado y solo en el primer ejemplo el código generado en el lenguaje R.

Guía de formatos de datos

En el análisis de datos, es importante conocer que muchas herramientas solo pueden interpretar datos que tenga un formato ordenado o conocido como tidy data, el cual requiere que la estructura de datos tenga un formato específico.

Es posible que Julius no pueda analizar ni transformar hojas de cálculo compuestas o datos con formato irregular. A continuación, se indican varios problemas comunes que se deben evitar (imagen 1):

- **Varias secciones.** Su hoja de cálculo no puede tener varias secciones.
- **Varias tablas.** De manera similar al punto anterior, no debería haber varias tablas en la hoja cargada.
- **Filas vacías.** Asegúrese de que no haya filas (o columnas) vacías.

| Balance Sheet | | | |
|---------------------------------------|------------------|------------------|--|
| Date: | | | |
| Assets | 2020 | 2019 | |
| Current Assets | | | |
| Cash | 11,874 | 12,498 | |
| Accounts receivable | | | |
| Inventory | | | |
| Prepaid expenses | | | |
| Short-term investments | | | |
| Total current assets | \$ 11,874 | \$ 12,498 | |
| Fixed (Long-Term) Assets | | | |
| Long-term investments | 1,208 | 1,340 | |
| Property, plant, and equipment | 15,340 | 15,993 | |
| (Less accumulated depreciation) | (2,200) | (2,046) | |
| Intangible assets | | | |
| Total fixed assets | \$ 14,348 | \$ 15,287 | |
| Other Assets | | | |
| Deferred income tax | | | |
| Other | | | |
| Total Other Assets | \$ - | \$ - | |
| Total Assets | \$ 26,222 | \$ 27,785 | |
| Liabilities and Owner's Equity | | | |
| Current Liabilities | | | |
| Accounts payable | 8,060 | 8,103 | |
| Short-term loans | | | |
| Income taxes payable | 3,145 | 3,368 | |
| Accrued salaries and wages | | | |
| Unearned revenue | | | |

Imagen 1. Datos tabulares con múltiples secciones por página y múltiples hojas

Formato de datos adecuado

Los datos bien estructurados y depurados permiten a Julius realizar análisis y visualizaciones de datos precisos y sin errores. A continuación, se indican las prácticas recomendadas que se deben seguir (imagen 2):

- **Encabezados de columnas.** Los encabezados de columnas descriptivos deben estar en la primera fila.
- **Tabulación.** Los datos deben estar en forma de filas de registros debajo de los encabezados.
- **Universalidad.** Se recomienda utilizar un lenguaje sencillo y de fácil comprensión para los nombres de las columnas, en lugar de utilizar jerga específica de la industria o de la empresa.

| | A | B | C | D | E | F | G | H | I | J | K | L |
|----|------------|--------|---------------|---------|------------|--------|--------------|------------------|-----------------|---------------------|---------------------|---------------------|
| 1 | customerID | gender | SeniorCitizen | Partner | Dependents | tenure | PhoneService | MultipleLines | InternetService | OnlineSecurity | OnlineBackup | DeviceProtection |
| 2 | 7590-VHVEG | Female | 0 | Yes | No | 1 | No | No phone service | DSL | No | Yes | No |
| 3 | 5575-GNVDE | Male | 0 | No | No | 34 | Yes | No | DSL | Yes | No | Yes |
| 4 | 3668-QPYBK | Male | 0 | No | No | 2 | Yes | No | DSL | Yes | Yes | No |
| 5 | 7795-CFOCW | Male | 0 | No | No | 45 | No | No phone service | DSL | Yes | No | Yes |
| 6 | 9237-HQITU | Female | 0 | No | No | 2 | Yes | No | Fiber optic | No | No | No |
| 7 | 9305-CDSKC | Female | 0 | No | No | 8 | Yes | Yes | Fiber optic | No | No | Yes |
| 8 | 1452-KIOVK | Male | 0 | No | Yes | 22 | Yes | Yes | Fiber optic | No | Yes | No |
| 9 | 6713-OKOMC | Female | 0 | No | No | 10 | No | No phone service | DSL | Yes | No | No |
| 10 | 7892-POOKP | Female | 0 | Yes | No | 28 | Yes | Yes | Fiber optic | No | No | Yes |
| 11 | 6388-TABGU | Male | 0 | No | Yes | 62 | Yes | No | DSL | Yes | Yes | No |
| 12 | 9763-GRSKD | Male | 0 | Yes | Yes | 13 | Yes | No | DSL | Yes | No | No |
| 13 | 7469-LKBCI | Male | 0 | No | No | 16 | Yes | No | No | No internet service | No internet service | No internet service |
| 14 | 8091-TTVAX | Male | 0 | Yes | No | 58 | Yes | Yes | Fiber optic | No | No | Yes |
| 15 | 0280-XJGEX | Male | 0 | No | No | 49 | Yes | Yes | Fiber optic | No | Yes | Yes |
| 16 | 5129-JLPIS | Male | 0 | No | No | 25 | Yes | No | Fiber optic | Yes | No | Yes |
| 17 | 3655-SNQYZ | Female | 0 | Yes | Yes | 69 | Yes | Yes | Fiber optic | Yes | Yes | Yes |
| 18 | 8191-XWSZG | Female | 0 | No | No | 52 | Yes | No | No | No internet service | No internet service | No internet service |
| 19 | 9959-WOFKT | Male | 0 | No | Yes | 71 | Yes | Yes | Fiber optic | Yes | No | Yes |
| 20 | 4190-MFLUW | Female | 0 | Yes | Yes | 10 | Yes | No | DSL | No | No | Yes |
| 21 | 4183-MYFRB | Female | 0 | No | No | 21 | Yes | No | Fiber optic | No | Yes | Yes |

Imagen 2. Datos estructurados con encabezado de variables

Pasos para el análisis de datos

A continuación, mostramos un listado de algunos pasos genéricos a realizar en el análisis de datos, que también incluye el tema de visualización de datos según tipo de variable, con el cual generamos algunos ejemplos utilizando **JULIUS.AI**.

Estructura de datos

- Homogenizar los nombres de las variables
- Homogenizar el formato de los datos (cuantitativos / cualitativos)

Evaluar calidad de los datos – Limpieza de datos

- Identificar variables con menos de 20% de datos
- Identificar datos duplicados
- Identificar datos faltantes
- Evaluar imputación de datos (rellenar datos faltantes con medidas de centralidad, media, mediana, moda)
- Identificar datos atípicos

Exploración de estructura de datos

- Mostrar un resumen de la estructura de datos
- Generar un diccionario de datos (opcional)
- Cantidad de datos y variables
- Identificar tipos de variables según su tipo: numéricas, categóricas y ordinales.
 - Variables numéricas

- Numéricas continuas: tienen decimales.
- Numéricas discretas: no tienen decimales.
- Variables categóricas
 - Dicotómicas: identifican 2 grupos.
 - Politémicas: expresan muchos grupos.
- Mostrar un resumen de los datos (10 registros)

Exploración de datos

- Realizar un análisis exploratorio de datos (EDA) (1977, *John W. Tukey*)
- Realizar un análisis de datos utilizando estadística descriptiva
- Realizar un análisis de datos utilizando estadística inferencial

Acceso a Julius.ai

Para utilizar Julius.ai se debe acceder el enlace <https://julius.ai/>.

A continuación, debe darle click al botón **LOG IN** en la parte superior. Se mostrará la ventana de la imagen 3, donde puede acceder de tres maneras, utilizando su correo personal o a través de una cuenta de Gmail o Apple.

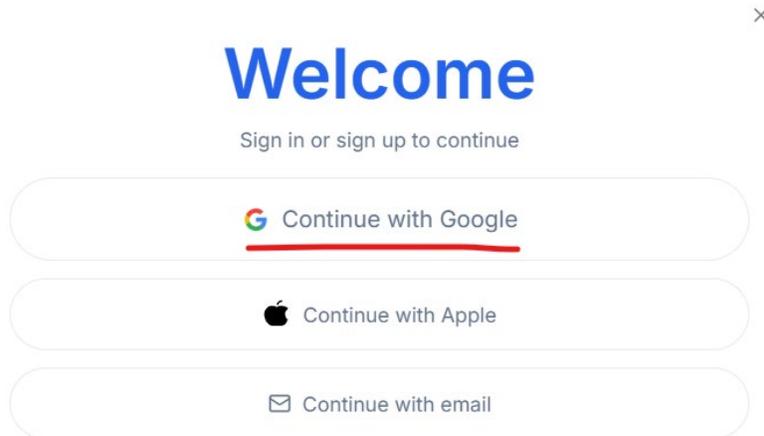


Imagen 3. Acceso a Julius.ai

A continuación, se mostrará la interfaz del programa, donde muestra como elementos principales como:

New Thread: corresponde a un chat o conversación sobre un tema o datos.

My Files: donde permite cargar los archivos a utilizar, como también mostrar el listado de documentos que se han cargados.

En la ventana principal se presenta el espacio para escribir el prompt, con la pregunta, que vas a analizar hoy=

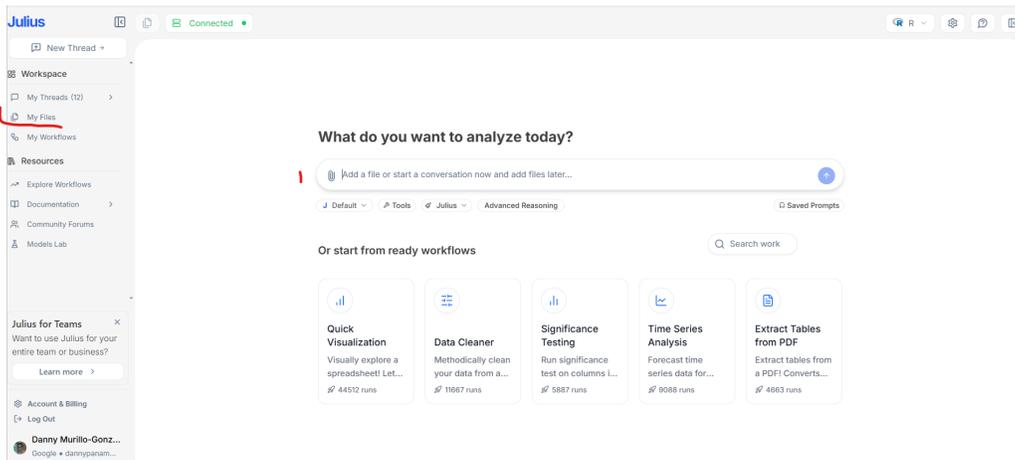


Imagen 4. Interfaz de Julius.ai

Cargar datos

Para cargar los datos en Julius que vamos a analizar, debemos dar click al menú principal en la opción de My files. Se mostrará la ventana de carga de documentos (imagen 5) que permite cargar los documentos a través del botón UPLOAD FILE. Esta opción me permite buscar un documento en mi computador y cargar el documento. También puedo cargar un documento de través de un enlace en Google sheet.

Los documentos cargados se mostrarán en la parte inferior, indicando el tamaño del archivo y la fecha de carga. Cada archivo mostrará la opción para chatear con el documento, descargar y eliminar. En la práctica el documento que cargaremos será el llamado **datos_ia_generativa-tic-19-02-2025.csv**, que son datos sintéticos creados con una IA Generativa.

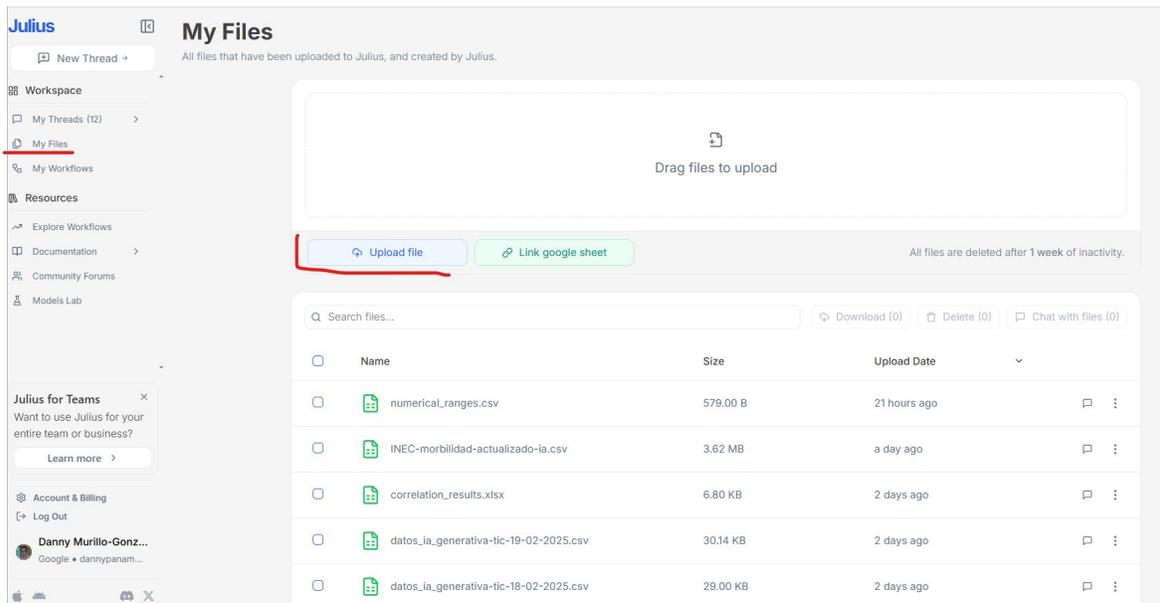


Imagen 5. Interfaz de carga de archivos de datos

Chatear con los datos cargados

Para chatear con los datos de prueba cargados debemos darle click al botón **New Thread**, que permitirá generar un nuevo espacio de chat. En esa ventana debemos darle click al botón con el icono que según imagen 6, se señala con un círculo rojo. Este me muestra diferentes opciones para cargar los documentos en el chat, en este caso seleccionaremos **MY Files**.

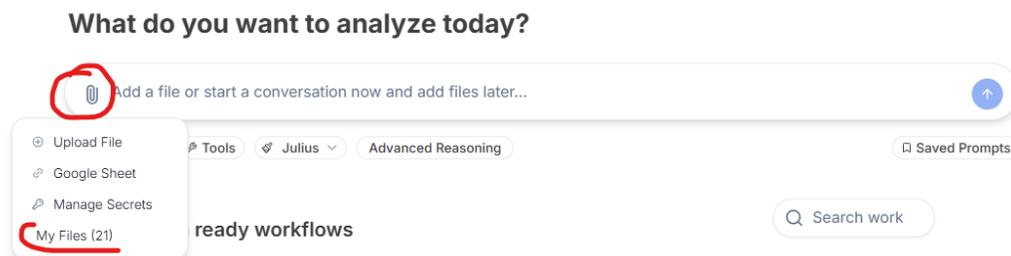


Imagen 6. Opción para seleccionar datos a evaluar

Se mostrará a continuación una ventana con los documentos cargados, seleccionamos el documento llamado **datos_ia_generativa-tic-19-02-2025.csv**, dando click sobre el nombre del documento.

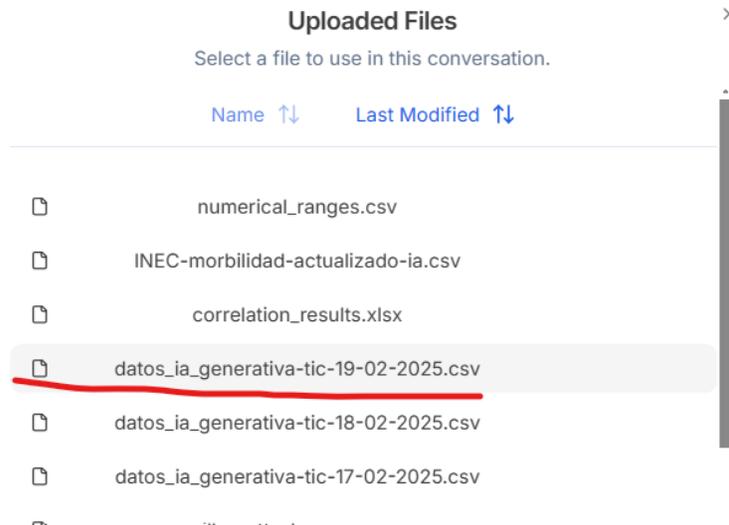


Imagen 7. Listado de documentos cargados en Julius.ia

Al cargar los datos se mostrará una ventana que muestra 20 registros de la tabla y todas sus variables (imagen 8), además de que, en la parte de la derecha, Julius muestra una explicación de lo que la herramienta interpretó de los datos:

El archivo, datos_ia_generativa-tic-19-02-2025.csv contiene datos acerca del conocimiento de las personas relacionado con IA Generativa y Tecnología de información y comunicación...

En la parte superior aparecen algunas preguntas que la herramienta recomienda se puede hacer a los datos, como, por ejemplo, **¿cómo la variable Nivel_conocimiento_TIC varía por país?**, entre otras.

| ID | Pais | Universidad | Sexo | Nivel_Conocimiento_TIC | Cursos_IA_Enseñados | Horas_Estudio | Nivel_Conocimiento_IA | Libros |
|----|------|-------------|-----------------------------|------------------------|---------------------|---------------|-----------------------|--------|
| 0 | 1 | Costa Rica | Universidad Centroamericana | Masculino | 2.1165588871 | 6.509736209 | 6.6572375687 | 3 |
| 1 | 2 | Nicaragua | Universidad de El Salvador | Femenino | 5 | 22.0552265538 | 15.3699729203 | 5 |
| 2 | 3 | Costa Rica | Universidad Autónoma | Masculino | 1.7866161095 | 18.9560808437 | 1.3185779079 | 1 |
| 3 | 4 | Costa Rica | Universidad Tecnológica | Femenino | 1.7398406897 | 6.0933672714 | 17.0437909541 | 2 |
| 4 | 5 | El Salvador | Universidad Nacional | Femenino | 3.0866920314 | 13.4445288907 | 14.0084986831 | 5 |
| 5 | 6 | Panamá | Universidad Nacional | Femenino | 2.0471927003 | 11.4671040127 | 6.8787269075 | 4 |
| 6 | 7 | El Salvador | Universidad de El Salvador | Femenino | 2.8175731735 | 14.988613465 | -0.5976633559 | 5 |
| 7 | 8 | El Salvador | Universidad del Valle | Masculino | 1 | 9.2411923647 | -1.7193417186 | 3 |
| 8 | 9 | Costa Rica | Universidad de El Salvador | Masculino | 3.3769811019 | 12.8335208189 | 10.9151871213 | 4 |

Imagen 8. Carga de datos y resumen de explicación de datos.

Visualización de datos y análisis exploratorio con Julius

El análisis exploratorio de datos, y en particular su visualización, es el primer análisis que se debe hacer sobre cualquier conjunto de datos antes de abordar otras técnicas estadísticas, sean sencillas o complejas.

A continuación, realizaremos algunas consultas a los datos en Julius, mostrando el resultado que genera la herramienta, en el espacio del prompt de Julius, como se ve en la imagen 9.

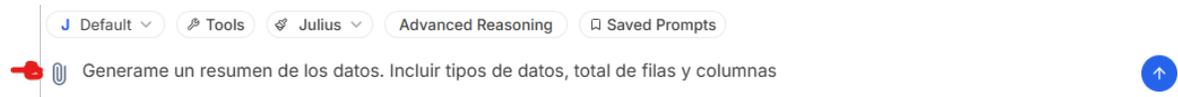


Imagen 9. Insertar Prompt o prompt de consulta de datos

Ejemplo 1: Resumen de datos

¿Qué deseo hacer?

Conocer la estructura de datos de la tabla `datos_ia_generativa-tic-19-02-2025.csv`, para identificar el nombre de las variables, el tipo de variable y el total de observaciones.

Prompt:

Genérame un resumen de los datos. Incluir tipos de datos, total de filas y columnas

Resultados

Julius inicialmente generará el código en R (lenguaje seleccionado) de cómo se realiza el resumen de datos (imagen 10). En caso de necesitar otros resultados, generará otros códigos en R. En este documento solo incluiremos el código de R en este ejemplo. La respuesta solicitada se muestra en la imagen 11,12 y 13.

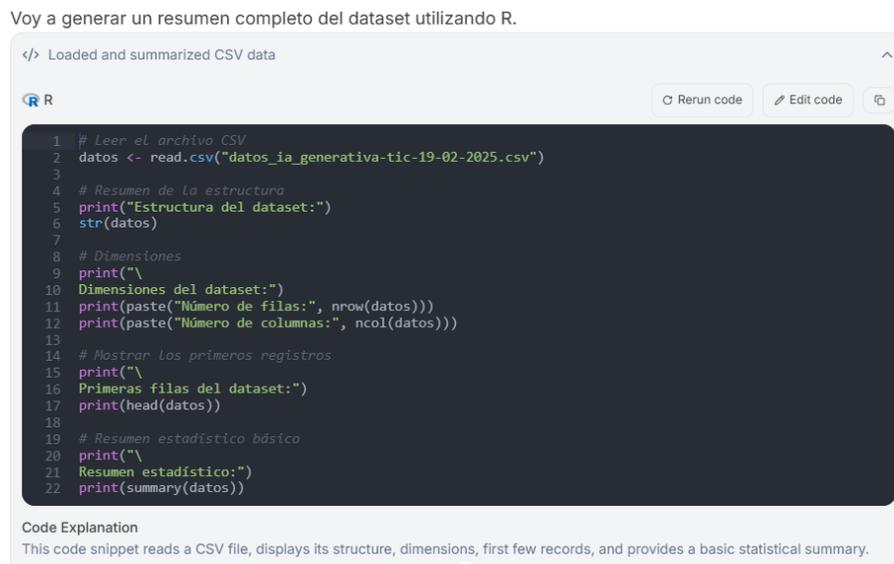


Imagen 10. Código en R relacionado con resumen de variables y sus tipos

```

[1] "Estructura del dataset:"
'data.frame': 260 obs. of 14 variables:
 $ ID          : int 1 2 3 4 5 6 7 8 9 10 ...
 $ Pais        : chr "Costa Rica" "Nicaragua" "Costa Rica" "Costa Rica" ...
 $ Universidad  : chr "Universidad Centroamericana" "Universidad de El Salvador" "Universidad
 Autónoma" "Universidad Tecnológica" ...
 $ Sexo        : chr "Masculino" "Femenino" "Masculino" "Femenino" ...
 $ Nivel_Conocimiento_TIC: num 2.12 5 1.79 1.74 3.09 ...
 $ Cursos_IA_Enseñados  : num 6.51 22.06 18.96 6.09 13.44 ...
 $ Horas_Estudio   : num 6.66 15.37 1.32 17.04 14.01 ...
 $ Nivel_Conocimiento_IA : int 3 5 1 2 5 4 5 3 4 3 ...
 $ Libros_IA_Leidos   : int 11 12 12 10 11 11 8 14 11 6 ...
 $ Asistencia_Eventos : int 4 5 5 6 5 4 5 5 4 10 ...
 $ Proyectos_IA      : int 2 2 3 3 3 2 3 2 1 4 ...
 $ Articulos_Publicados : int 1 1 1 1 0 1 0 1 0 0 ...
 $ Cursos_IA_Tomados  : int 19 25 40 20 24 38 34 24 47 18 ...
 $ horas_ia          : int NA NA 7 NA NA NA NA NA NA NA ...
[1] "\nDimensiones del dataset:"
[1] "Número de filas: 260"
[1] "Número de columnas: 14"

```

Imagen 11. Resumen de variables numéricas

[1] "\nPrimeras filas del dataset:"

| Optional: Click on columns or rows to focus on specific data | | | | | | | | Export | Info |
|--|----|-------------|-----------------------------|-----------|------------------------|---------------------|------------------|--------|------|
| | ID | Pais | Universidad | Sexo | Nivel_Conocimiento_TIC | Cursos_IA_Enseñados | Horas_Estudio | | |
| 1 | 1 | Costa Rica | Universidad Centroamericana | Masculino | 2.11655888707421 | 6.50973620895081 | 6.6572375687207 | | |
| 2 | 2 | Nicaragua | Universidad de El Salvador | Femenino | 5 | 22.055226553843 | 15.369972920328 | | |
| 3 | 3 | Costa Rica | Universidad Autónoma | Masculino | 1.78661610949244 | 18.9560808436629 | 1.31857790785667 | | |
| 4 | 4 | Costa Rica | Universidad Tecnológica | Femenino | 1.73984068971088 | 6.09336727136079 | 17.0437909540776 | | |
| 5 | 5 | El Salvador | Universidad Nacional | Femenino | 3.08669203144834 | 13.4445288907237 | 14.008498683076 | | |
| 6 | 6 | Panamá | Universidad Nacional | Femenino | 2.04719270030514 | 11.4671040127009 | 6.8787269074696 | | |

Imagen 12. Primeras filas del dataset

Resumen de los datos:

- **Número de filas:** 260
- **Número de columnas:** 14
- **Tipos de datos:**
 - Variables categóricas: Pais, Universidad, Sexo
 - Variables numéricas: Nivel_Conocimiento_TIC, Cursos_IA_Enseñados, Horas_Estudio, Nivel_Conocimiento_IA, Libros_IA_Leidos, Asistencia_Eventos, Proyectos_IA, Articulos_Publicados, Cursos_IA_Tomados, horas_ia

Imagen 13- Resumen de los datos

Ejemplo 2: Resumen de datos categóricos

¿Qué deseo hacer?

Conocer que países se utilizan en los datos y cuantas observaciones hay por país

Prompt:

Realiza un resumen de la variable categórica País, indicando el número de filas por categorías.

Resultados

[1] "Resumen de la variable categórica País:"

Summary ^

| Costa Rica | El Salvador | Guatemala | Honduras | Nicaragua | Panamá |
|------------|-------------|-----------|----------|-----------|--------|
| 52 | 47 | 30 | 44 | 44 | 43 |

El resumen indica el número de filas por categoría de país. Según la tabla:

- Costa Rica: 52 filas
- El Salvador: 47 filas
- Guatemala: 30 filas
- Honduras: 44 filas
- Nicaragua: 44 filas
- Panamá: 43 filas

Imagen 14. Número de observaciones por país

Análisis exploratorio univariante

Es un análisis exploratorio que se hace para conocer las características de una sola variable, ya sea cualitativa o cuantitativa. En este espacio realizaremos diferentes análisis a través de la visualización de los datos.

Análisis exploratorio univariante de variables categóricas

Cuando nuestra variable no se expresa con números, sino con etiquetas o categorías de una determinada característica, el resumen numérico que utilizamos es la tabla de frecuencias.

La frecuencia de los datos cualitativos se puede representar gráficamente de dos maneras, a través de un gráfico de barras o un gráfico de sectores (pastel).

Gráfico de sectores – frecuencia

Los gráficos de sectores también pueden representar variables cualitativas, se deben incluir los valores (frecuencias o porcentajes). Aunque no se recomiendan utilizar debido a que el ojo humano no es tan bueno distinguiendo ángulos como alturas.

Ejemplo 3a: Gráfico de sectores (pastel)

¿Qué deseo hacer?

Conocer la distribución de los datos por universidad

Prompt:

Crear un gráfico de sectores utilizando la variable universidad, incluir leyenda del gráfico

Resultados

Distribución de estudiantes por universidad

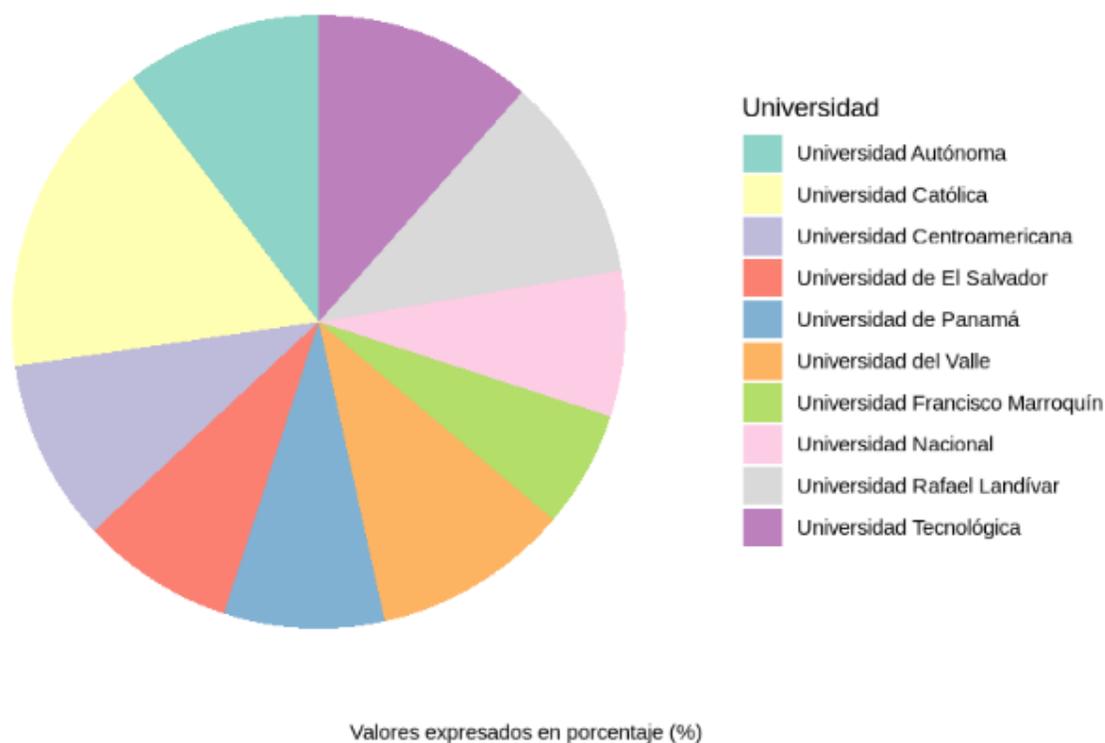


Imagen 15- Distribución de docentes por universidad

Interpretación:

- El gráfico de sectores permite visualizar la proporción relativa de docentes en cada universidad.
- Se aprecia que algunas universidades tienen una mayor representación (por ejemplo, la que tiene 44 docentes y 16.9% del total) en comparación con otras.
- La leyenda facilita identificar cada sector y se muestra el porcentaje relativo, lo que ayuda a comprender la magnitud de cada parte respecto al conjunto total.

Ejemplo 3b: Gráfico de sectores (valores)

¿Qué deseo hacer?

Conocer la distribución de los datos por universidad

Prompt:

Crear un gráfico de sectores utilizando la variable universidad, incluir leyenda del gráfico

Resultados

Distribución de estudiantes por universidad

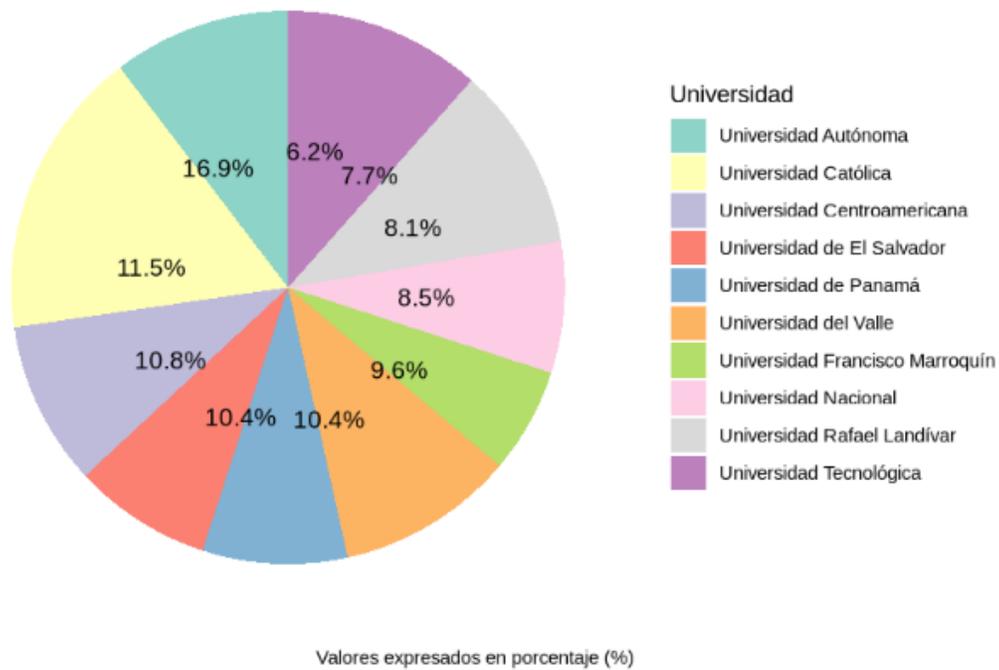


Imagen 16. Distribución de docentes por universidad

Ejemplo 3c: Gráfico de sectores (ordenados)

¿Qué deseo hacer?

Conocer la distribución de los datos por universidad con un máximo de 6 sectores,

Prompt:

seleccionar las 5 universidades con mayor porcentaje, con el resto de los sectores crear un sector llamado otros. colocar solo los valores, centrados e incluir la leyenda. Ordenar los sectores en sentido de las manecillas del reloj. Mostrar gráfico.

Resultados

Distribución de estudiantes por universidad

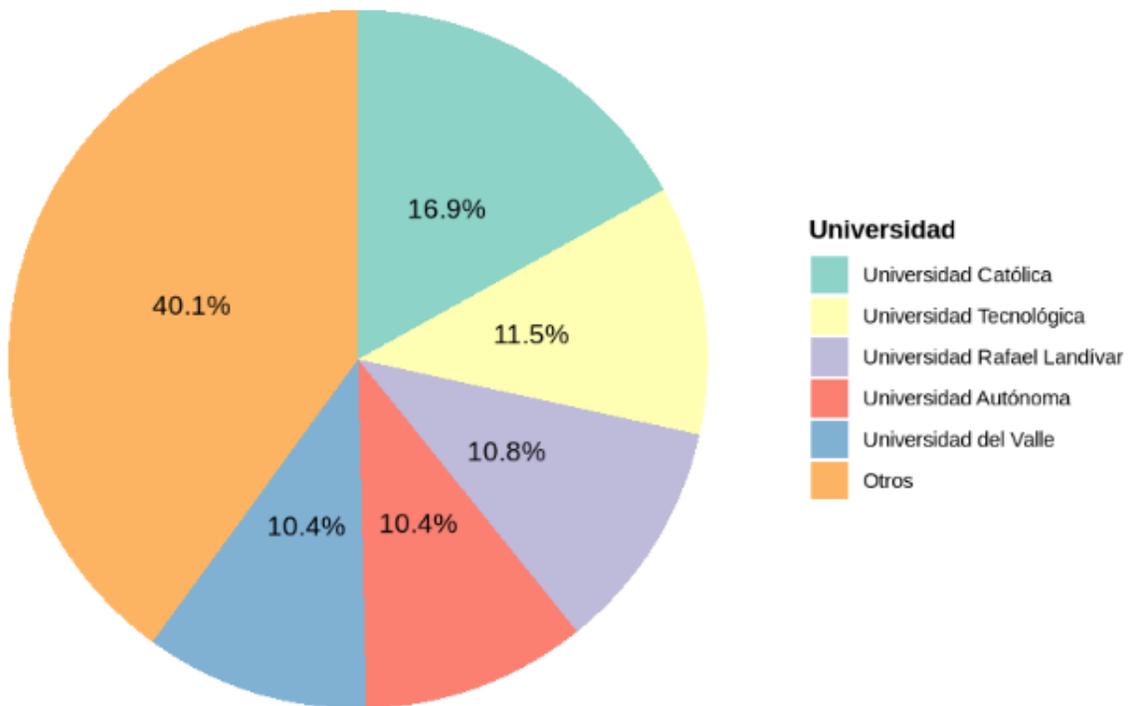


Imagen 17. Distribución de docentes por universidad

Gráfico de barras - frecuencia

En el gráfico de barras, se representan las categorías en el eje horizontal (X) y las frecuencias en el eje vertical (Y), y representamos las barras cuya altura representa la frecuencia. El gráfico de barras se puede representar también invirtiendo los ejes (a veces mejora la visualización de las etiquetas), representando líneas en vez de barras, u ordenando las barras según la frecuencia.

Ejemplo 4a: Gráfico de barras

¿Qué deseo hacer?

Conocer la distribución de los datos por universidad

Prompt:

Generar un gráfico de barras de la variable universidad

Resultado

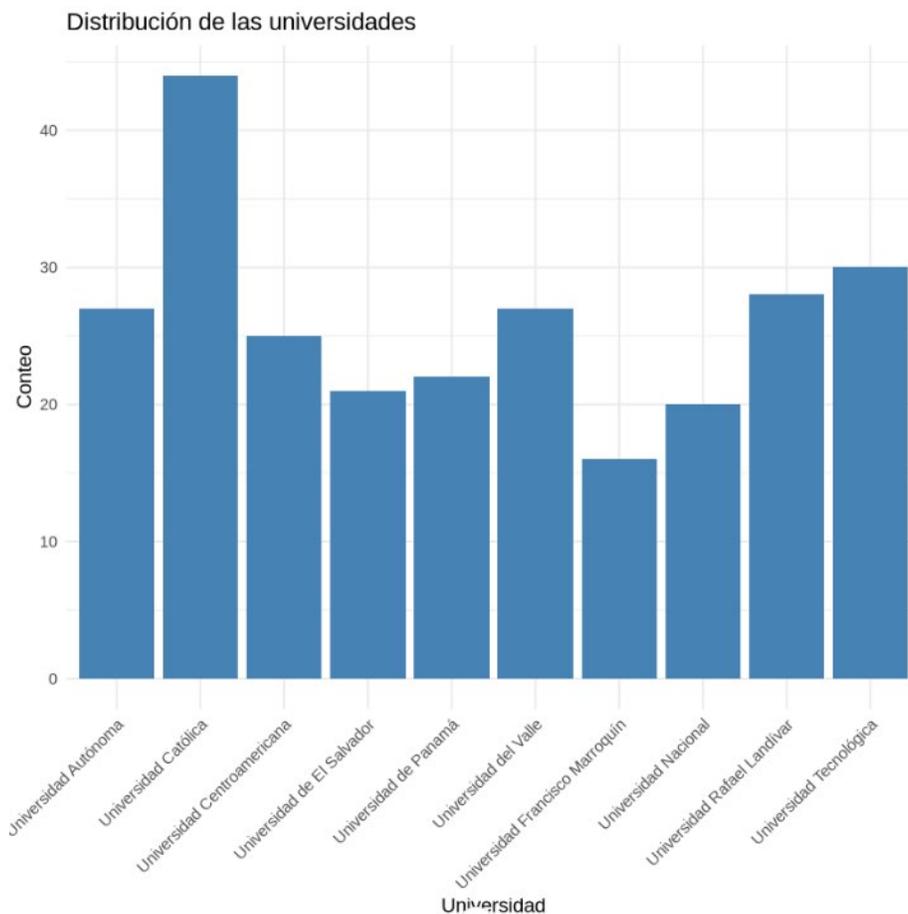


Imagen 18. Frecuencia de datos de la variable País

Interpretación (Julius):

El gráfico muestra la frecuencia de docentes en cada universidad. Se puede observar que hay una distribución relativamente uniforme entre las diferentes instituciones, con algunas variaciones en el número de docentes por universidad.

Ejemplo 4b: Gráfico de barras ordenada

¿Qué deseo hacer?

Conocer la distribución de los datos por universidad

Prompt:

Intercambiar los ejes del gráfico de barra de universidades, ordenar las barras de mayor a menor y colocar el valor en cada barra

Resultado

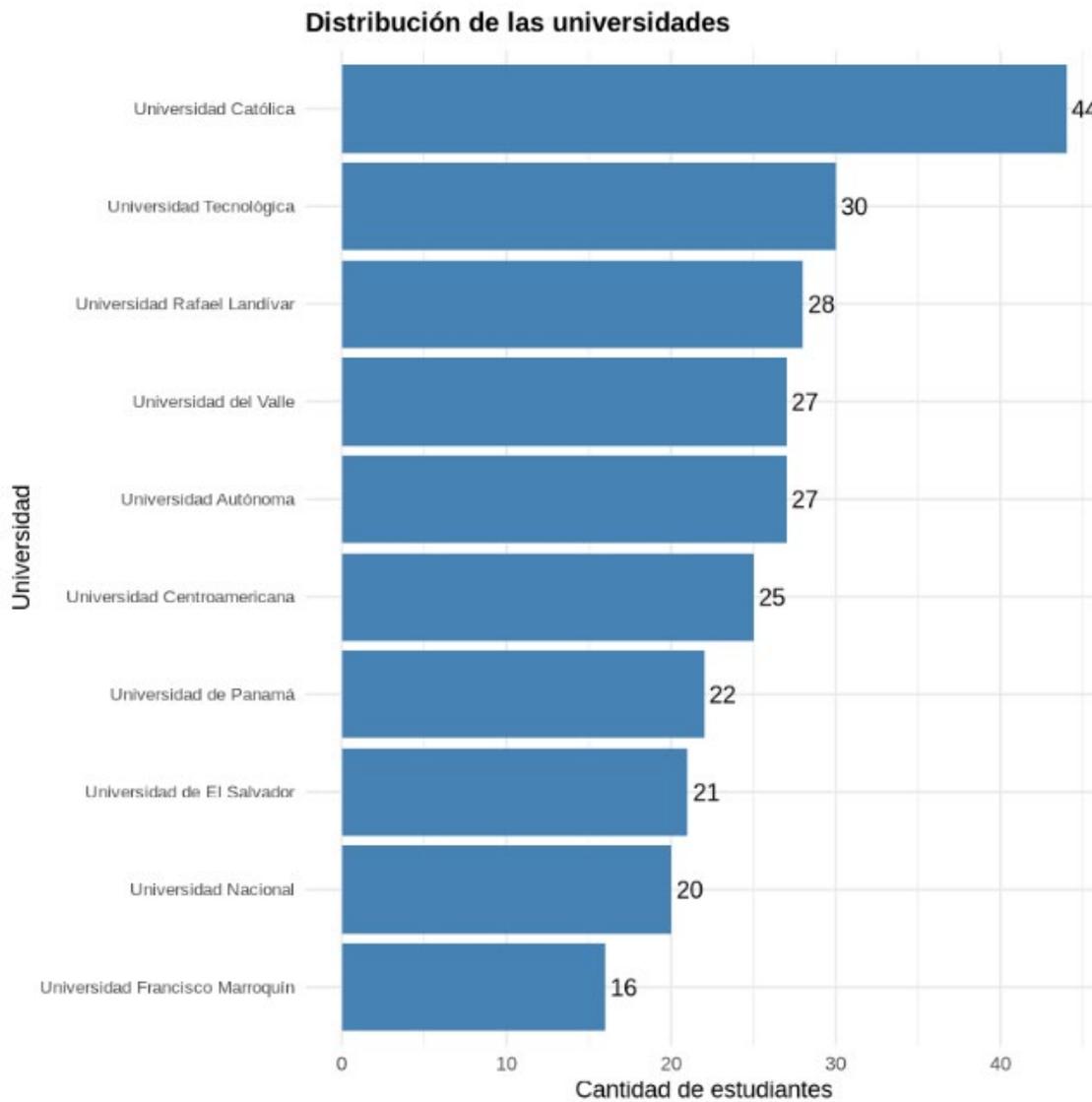


Imagen 19. Frecuencia de datos de la variable País

Ejemplo 4c: Gráfico de barras ordenada con énfasis

¿Qué deseo hacer?

Conocer la distribución de los datos con énfasis en una universidad

Prompt:

Haz un énfasis de color (naranja) en la barra de la Universidad Tecnológica. ¿Qué interpretas del gráfico?

Resultado

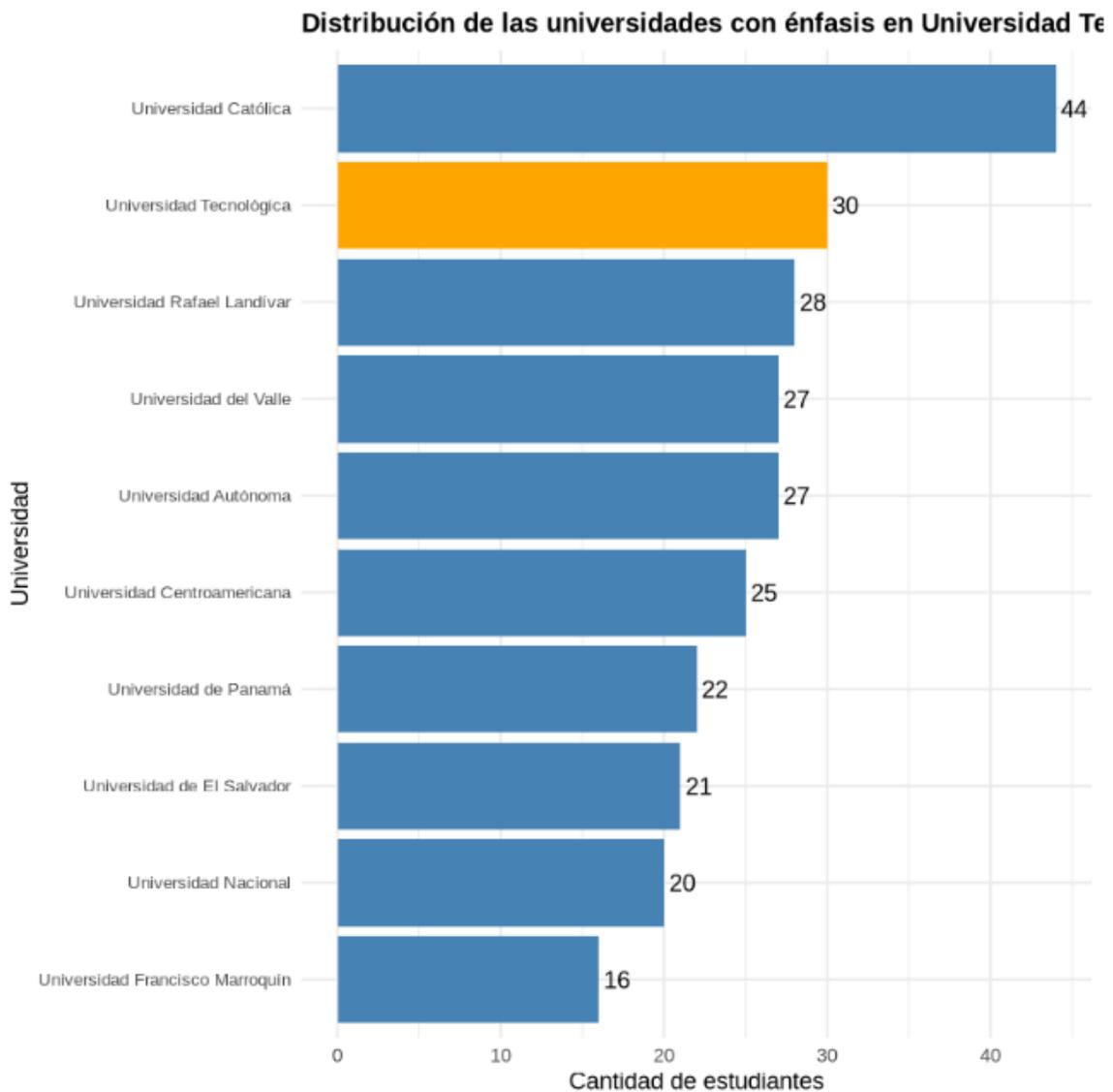


Imagen 20. Frecuencia de datos de la variable País

Interpretación (Julius):

La Universidad Tecnológica tiene una representación notable, pero no dominante, en comparación con la Universidad Católica.

Análisis exploratorio univariante, variable cuantitativa

Cuando nuestra variable se expresa con números, debemos evaluar si los datos son discretos o continuos. Si son discretos podemos utilizar como representación gráfica para las frecuencias absolutas, el gráfico de barras. En el caso de que la variable sea continua o también discreta, podemos utilizar el gráfico de histograma, densidad o diagrama de caja.

Gráfico de histograma - frecuencia

La representación gráfica de la tabla de frecuencias por intervalos de una variable numérica es el **Histograma**. En el eje Y se representan las frecuencias (absolutas o relativas) de cada clase. La geometría serán barras, en este caso sin espacio entre ellas para representar la continuidad.

Ejemplo 5: Gráfico de histograma

¿Qué deseo hacer?

Conocer la distribución de los datos de la variable cursos_IA_enseñados

Prompt:

Generar una distribución de datos de la variable Cursos_IA_Enseñados, incluir línea de densidad. Incluir gráfico e interpretación.

Resultado

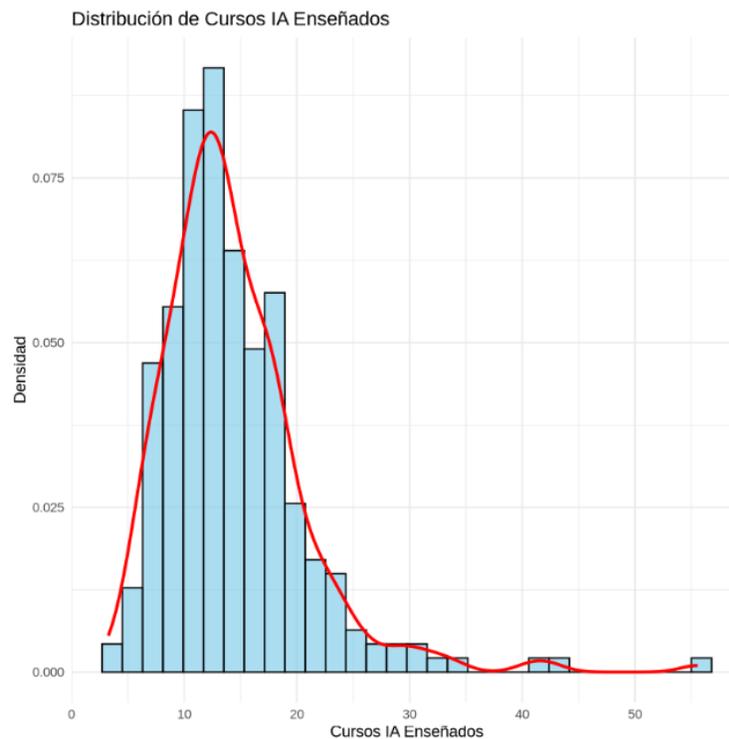


Imagen 21. Distribución de docentes por curso de ia enseñados

Interpretación (Julius):

- **Histograma:** Representa la frecuencia de los valores de la variable **Cursos_IA_Enseñados**. Cada barra indica cuántas observaciones caen dentro de un rango específico.
- **Línea de densidad:** Suaviza la distribución de los datos, mostrando la tendencia general de la variable. La línea roja permite identificar picos (modas) y la forma general de la distribución (simétrica, sesgada, etc.).
- **Ejes:** El eje X muestra los valores de **Cursos_IA_Enseñados**, mientras que el eje Y representa la densidad relativa.

El gráfico sugiere cómo se distribuyen los cursos enseñados, permitiendo identificar patrones como concentración en ciertos valores o dispersión.

El gráfico de cajas y bigotes (boxplot)

Este gráfico representa generalmente en el eje vertical, los siguientes estadísticos: El mínimo (extremo bigote inferior), el primer cuartil (borde inferior de la caja), la mediana (línea cruzando caja), el tercer cuartil (borde superior de la caja) y el máximo (extremo bigote superior).

Ejemplo 6: Gráfico de boxplot

¿Qué deseo hacer?

Conocer la distribución de los datos de la variable cursos_IA_enseñados

Prompt:

Generar una distribución de datos de la variable cursos ia enseñados utilizando un boxplot . Generar interpretación. Muestra gráfico, rotar ejes.

Resultado

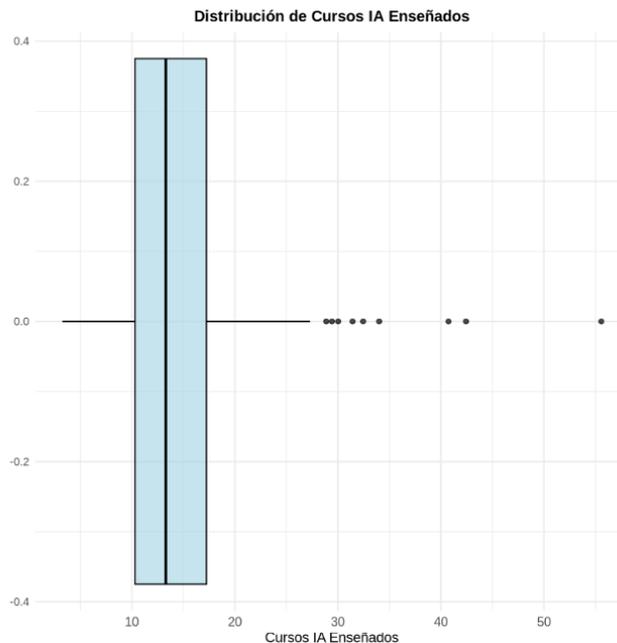


Imagen 22. Distribución de docentes por cursos IA enseñados

Interpretación 1 (Julius):

- **Tendencia Central:** La mediana de cursos IA enseñados es de 13.29, con una media de 14.33, lo que sugiere una ligera asimetría hacia valores más altos.
- **Dispersión:**
 - El rango intercuartílico (IQR) es de 6.93 cursos
 - El 50% central de los datos se encuentra entre 10.31 (Q1) y 17.24 (Q3) cursos
 - Los valores se extienden desde un mínimo de 3.26 hasta un máximo de 55.57 cursos
- **Valores Atípicos:**
 - Se identificaron 9 valores atípicos por encima del límite superior (27.64)
 - No hay valores por debajo del límite inferior (-0.09)
 - Estos valores atípicos representan profesores con un número inusualmente alto de cursos enseñados

Interpretación 2 (Julius):

Desde el punto de vista de la variable, cursos ya enseñados por docentes, ¿que indica el gráfico?

La mayoría de los docentes imparten un número moderado de cursos, unos pocos enseñan significativamente más, lo que podría reflejar diferencias en la especialización o la carga de trabajo.

Análisis exploratorio bivalente

El análisis bivalente o multivalente es muy útil para describir dos o más características de la población en estudio.

Análisis exploratorio bivariantes de variables categóricas

Al igual que en el caso univariante, la mejor representación de tablas de frecuencias para variables cualitativas y para variables numéricas discretas sigue siendo el gráfico de barras.

Ejemplo 7: Gráfico de barras con frecuencias conjuntas

¿Qué deseo hacer?

Conocer la distribución de los docentes por país y género

Prompt:

Genera un gráfico de barras conjuntas de las variables país y género. Mostrar e interpretar gráfico

Resultado

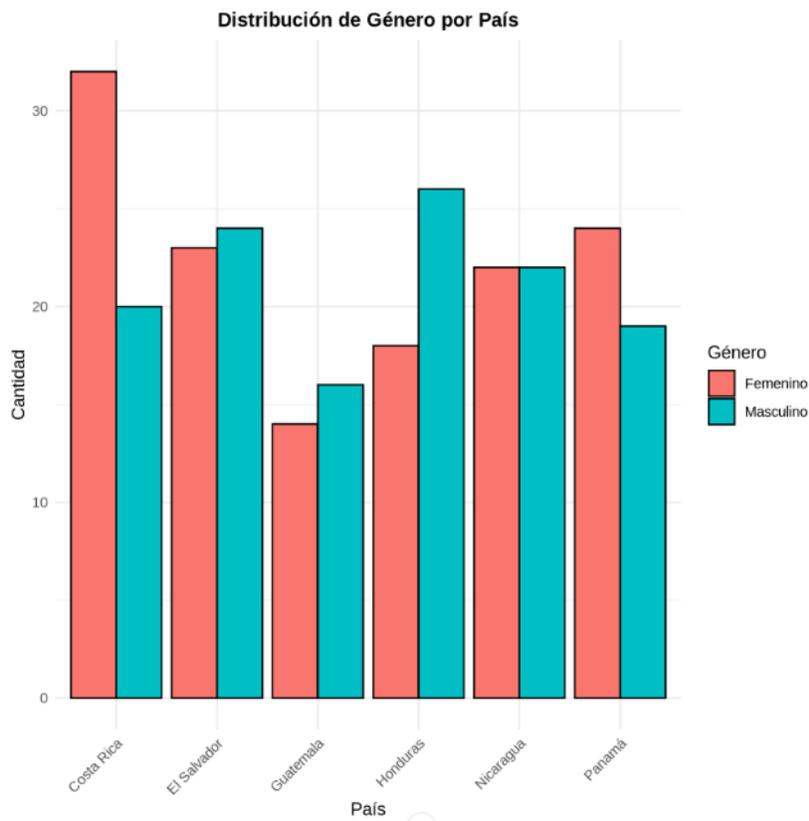


Imagen 23. Distribución de docentes por país y género

Interpretación (Julius):

Basado en el gráfico y los datos presentados, podemos observar:

1. Distribución por País:

- Costa Rica muestra una mayor participación femenina (61,54% vs 38,46%)
- Honduras presenta una mayor participación masculina (59,09% vs 40,91%)
- Nicaragua muestra una distribución perfectamente equilibrada (50% cada género)

2. Patrones Generales:

- La mayoría de los países muestran una distribución relativamente equilibrada entre géneros
- Las diferencias más notables se encuentran en Costa Rica (mayor presencia femenina) y Honduras (mayor presencia masculina)
- Panamá y El Salvador muestran distribuciones cercanas al

3. Participación Total:

- Guatemala muestra la menor cantidad total de participantes
- Costa Rica y Honduras muestran una mayor cantidad total de participantes
- La representación por género varía significativamente entre países.

Ejemplo 8: Gráfico de barras con frecuencias conjuntas apiladas

¿Qué deseo hacer?

Conocer la distribución de los docentes por país y género

Prompt:

Genera un gráfico de barras apiladas de las variables país y género. Mostrar valores. Mostrar e interpretar gráfico

Resultado

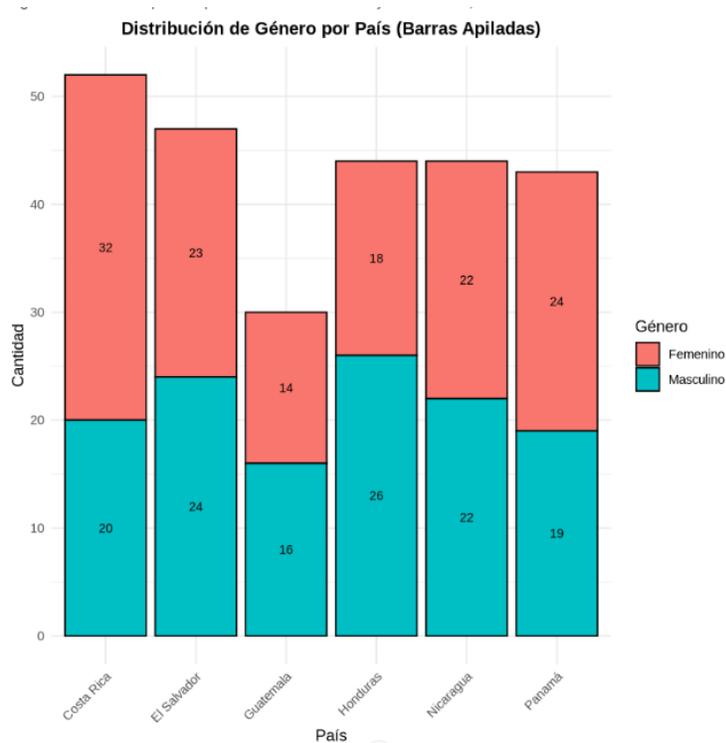


Imagen 24. Distribución de docentes por Género y por País

Análisis exploratorio bivariantes e variables cuantitativas

Ejemplo 9: Gráfico de barras con frecuencias conjuntas

¿Qué deseo hacer?

Conocer la distribución de los docentes por país de los cursos tomados

Prompt:

Generar una distribución de datos utilizando boxplot de las variables, país y cursos ia tomados. Mostrar e interpretar el gráfico.

Resultado

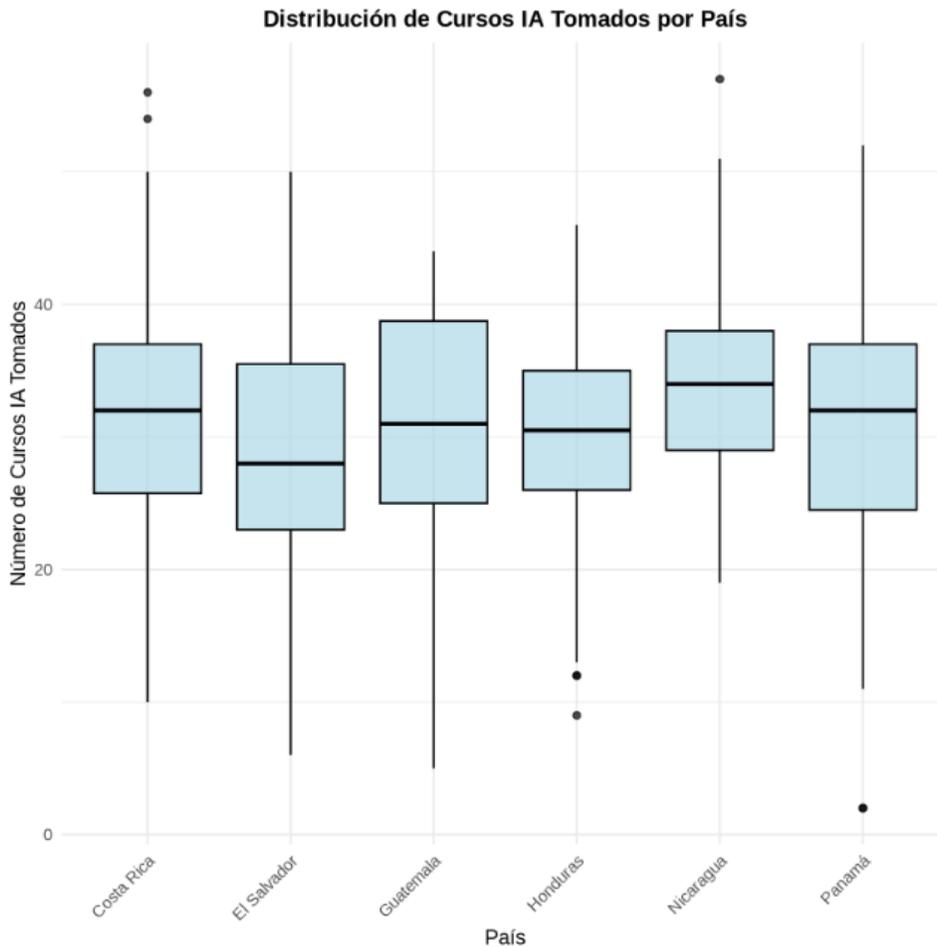


Imagen 25-. Distribución de cursos IA tomado por país

Interpretación:

Cada país muestra una distribución diferente en cuanto al número de cursos tomados en IA. Los boxplots permiten visualizar la mediana, el rango intercuartílico (IQR), y los valores mínimo y máximo de cada grupo.

Comparación entre países:

- **Costa Rica:** Se observa un rango amplio, con una mediana en torno a 32 cursos, lo que indica que la mayoría de los docentes se concentran alrededor de este valor.

- **El Salvador:** La mediana es menor (alrededor de 28), y la distribución evidencia menores valores en comparación con otros países.
- **Guatemala, Honduras, Nicaragua y Panamá:** Muestran distribuciones con medianas en el rango de 30 a 34 cursos, pero también diferencias en cuanto a dispersión; por ejemplo, Nicaragua presenta valores máximos más altos.

Gráfico de dispersión (Gráfico de puntos)

Para dos variables continuas, la representación más adecuada es el gráfico de dispersión. El gráfico de dispersión es una “nube de puntos” representada en unos ejes cartesianos X,Y, que representan las escalas de cada una de las variables.

Ejemplo 10: Gráfico de barras con frecuencias conjuntas

¿Qué deseo hacer?

Conocer la relación entre los cursos tomados y los cursos enseñado

Prompt:

Realizar una gráfica de puntos de las variables cursos ia tomados y cursos ia enseñados. Mostrar e interpretar gráfico. Indicar el valor de la aceleración R.

Resultado

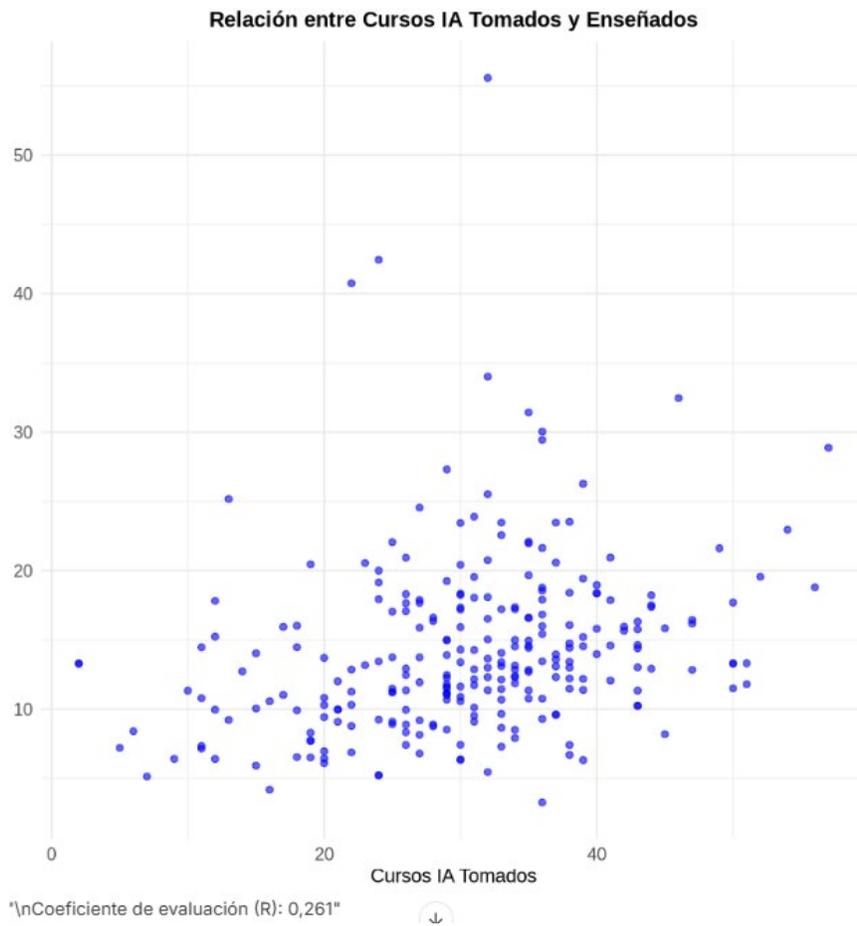


Imagen 26. Relación entre los cursos ia tomados y enseñados.

Ejemplo 11: Gráfico de barras con frecuencias conjuntas

¿Qué deseo hacer?

Conocer la relación entre los cursos tomados y los cursos enseñados por universidad

Prompt:

Realizar una gráfica de puntos de las variables cursos ia tomados y cursos ia enseñados. Categorizar los puntos con la variable Universidad. Mostrar gráfico

Resultado

relación entre Cursos IA Tomados y Enseñados por Universidad

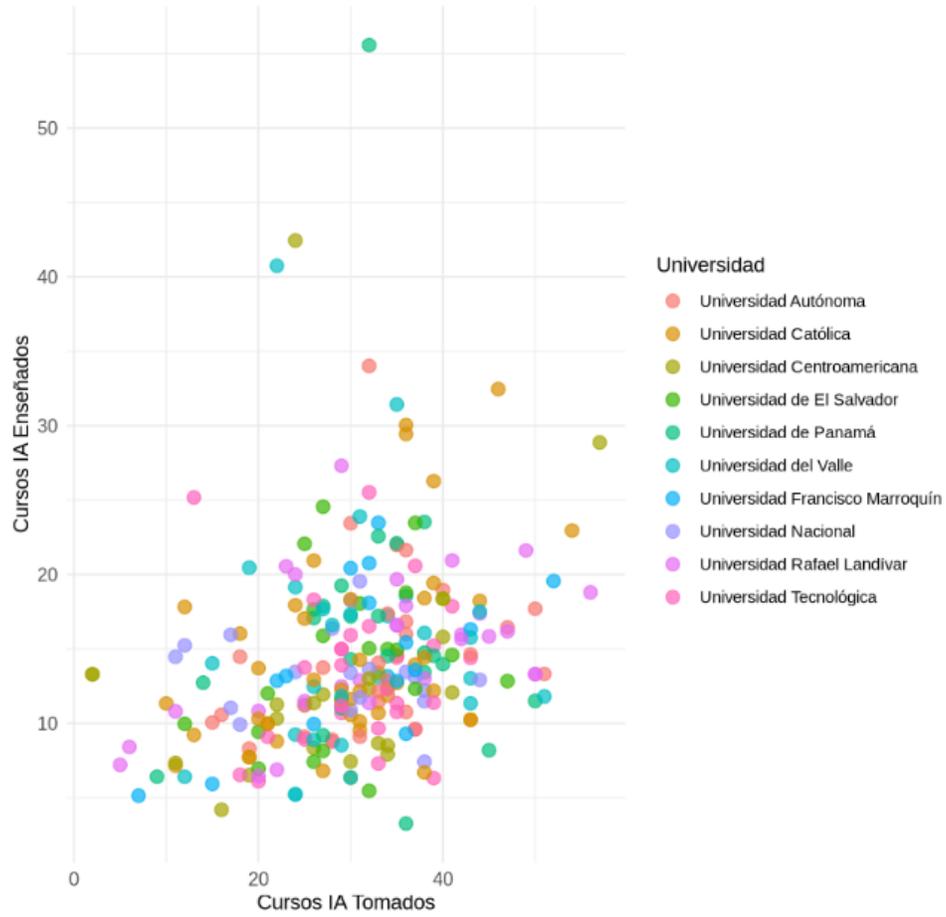


Imagen 27. Relación entre los cursos tomados y los cursos enseñado por universidad

Análisis de Clustering

El análisis de clustering es una técnica de aprendizaje no supervisado que agrupa datos similares en subconjuntos (clusters), de manera que los elementos dentro de un mismo cluster sean más parecidos entre sí que con los de otros clusters. Se utiliza para descubrir patrones ocultos, segmentar datos y simplificar estructuras complejas.

Para determinar el número óptimo de clusters, puedes aplicar varias pruebas y métricas:

1. **Método del codo (Elbow Method)**
2. **Estadístico Gap (Gap Statistic)**
3. **Índice de Silueta (Silhouette Index)**

Ejemplo 12: Análisis de clustering (calculo del número de clusters)

¿Qué deseo hacer?

Conocer el número optimo de cluster para separar los datos

Prompt:

Realizar un Gap estadístico para evaluar el número óptimo de clusters a generar según datos. Indicar el número óptimo de cluster según el gráfico generado.

Resultado

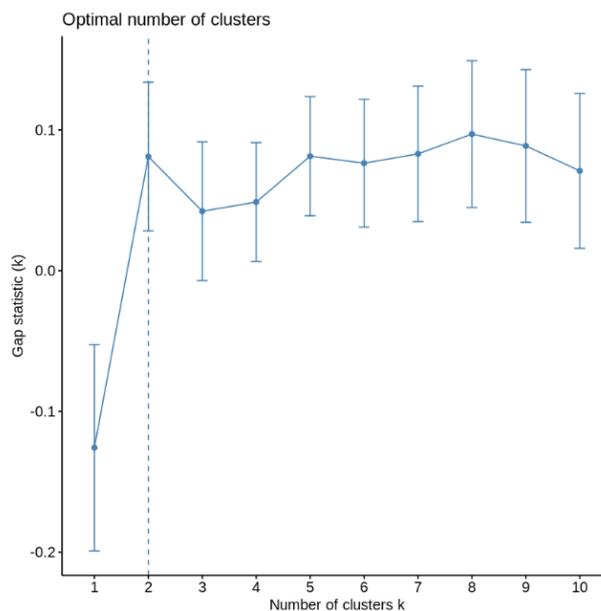


Imagen 28. Cálculo del número de clusters con **Gap Statistic**

Interpretación:

El número óptimo de clústeres para la representatividad de los datos es de 2.

Ejemplo 14 : Biplot

¿Qué deseo hacer?

Conocer el comportamiento y desarrollo profesional de docentes

Prompt:

generar biplot y categorizar puntos con color, el número de clusters es igual a 2

Resultado

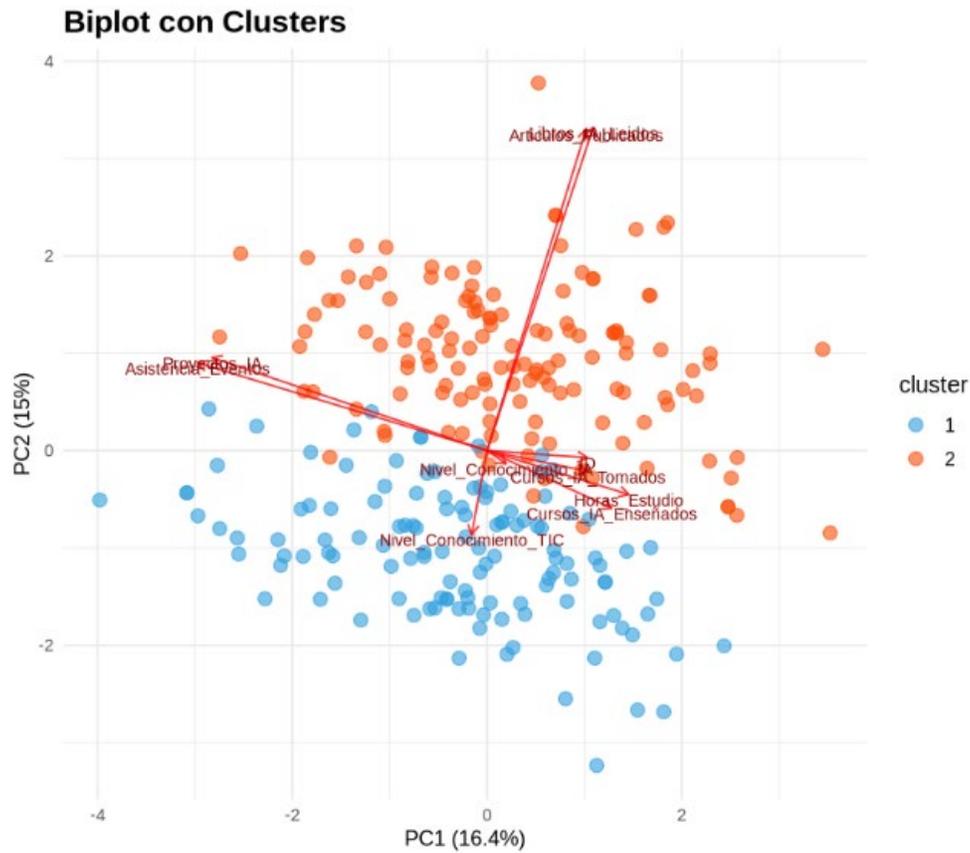


Imagen 30. Biplot con dos clusters

Interpretación:

- **Mayor contribución (flechas más largas):**
- **Correlaciones positivas (flechas en dirección similar):**
 - **Artículos_Publicados** y **Cursos_IA_Tomados** muestran una correlación positiva
 - **Nivel_Conocimiento_IA** y **Proyectos_IA** están positivamente correlacionados
- **Correlaciones negativas (flechas en direcciones opuestas):**

- **Libros_IA_Leidos** muestra una correlación negativa con **Articulos_Publicados**
- **Horas_Estudio** tiene una correlación negativa débil con varias variables
- **Cluster 1 (Azul):**
 - Tiende a tener valores más altos en **Libros_IA_Leidos**
 - Valores más bajos en **Articulos_Publicados**
 - Menor participación en **Cursos_IA_Tomados**
- **Cluster 2 (Naranja):**
 - Valores más altos en **Articulos_Publicados**
 - Mayor participación en **Cursos_IA_Tomados**
 - Valores más bajos en **Libros_IA_Leidos**