

**TECNICATURA SUPERIOR EN  
Ciencia de Datos e Inteligencia Artificial**

**EduDatáalisis**  
**Análisis del rendimiento académico estudiantil**

**Tercera Evidencia**  
**Módulo: Analista de Datos**

**Integrantes:**

- Florencia Andrada
- Federico Córdoba
- Melania Ligorria
- Guadalupe Mendoza
- Miguel Rojas Medina
- Noelia Soto

**Docentes Guía:**

- Nahuel Pratta
- Marcos Ugarte

AÑO: 2025

## Proyecto para el Módulo Analista de Datos

Primer Cuatrimestre 2025.

Docentes:

- Marcos Ugarte
- Nahuel Pratta

Integrantes:

Nombre y Apellido	Usuario GitHub
<b>Federico Córdoba</b>	fedecordoba-ispc
<b>Florencia Andrada</b>	Flor3ncia-Andr4d4
<b>Melania Ligorria</b>	mel-ligorria
<b>Guadalupe Mendoza</b>	Guadamendoza
<b>Miguel Eduardo Rojas Medina</b>	Migueerm
<b>Noelia Soto</b>	NoeliaSoto

**Link Repositorio Github:**

 [https://github.com/adriel1364/Analisis\\_25](https://github.com/adriel1364/Analisis_25)

## 1. NOMBRE DEL PROYECTO:

**EduDatálisis** — Análisis del rendimiento académico estudiantil a través de técnicas de ciencia de datos. Este proyecto se desarrolla en el segundo cuatrimestre del año 2025, utilizando un dataset de desempeño escolar con el objetivo de identificar patrones y generar información útil para la toma de decisiones educativas.

## 2. TIPO DE PROYECTO:

Es un proyecto de investigación con la finalidad de aplicar los conceptos aprendidos en el espacio curricular.

## 3. ESPACIO CURRICULAR O ESPACIOS PARTICIPANTES EN EL MÓDULO:

Modulo Analista de datos, integrado por las materias:

- Procesamiento de Datos.
- Estadística y Exploración de datos 1.

## 4. EJES TEMÁTICOS/RED DE CONCEPTOS:

Los ejes temáticos de este proyecto integran conocimientos de estadística descriptiva, programación en Python y técnicas de visualización de datos, vinculados al análisis del rendimiento académico estudiantil. Desde el enfoque interdisciplinario del módulo, se articulan los siguientes conceptos clave:

- Limpieza y procesamiento de datos (Pandas, NumPy).
- Estadística descriptiva aplicada (medias, modas, varianzas).
- Visualización (Matplotlib, Seaborn).
- Técnicas de análisis exploratorio y modelado predictivo simple.

Estas temáticas fortalecen competencias propias del perfil profesional de un Analista de Datos, como:

- Interpretación y comunicación de hallazgos basados en datos.
- Aplicación de lenguajes de programación para resolver problemas reales.

- Trabajo colaborativo en entornos técnicos.
- Pensamiento analítico orientado a la mejora continua en contextos sociales.

## **5. PROBLEMÁTICAS/NECESIDADES:**

Analizar el desempeño académico de los alumnos en función de diversas variables socioeconómicas, demográficas y académicas presentes en un conjunto de datos reales. Existe una necesidad creciente en las instituciones educativas de comprender qué factores inciden en el rendimiento académico para diseñar políticas de intervención eficaces y personalizadas.

## **6. FUNDAMENTACIÓN:**

La elección de este problema responde al interés del equipo en aplicar herramientas de ciencia de datos a un ámbito cercano y significativo: el desempeño académico estudiantil. Esta problemática permite no solo afianzar contenidos técnicos del módulo, sino también desarrollar una mirada crítica sobre los factores que influyen en el aprendizaje.

El potencial del proyecto radica en su capacidad para generar información relevante que podría ser utilizada por instituciones educativas para diseñar estrategias de acompañamiento o mejora. A su vez, la utilización de datos reales fomenta el desarrollo de habilidades clave del perfil profesional, como la capacidad de análisis, la ética en el manejo de información y la comunicación efectiva de resultados.

Este proyecto tiene un impacto comunitario al ofrecer un modelo replicable para que otras escuelas, centros de formación o docentes puedan analizar sus propios datos, potenciando una toma de decisiones más informada.

## **7. VISIÓN DEL PROYECTO:**

Este proyecto busca consolidar el aprendizaje de herramientas de análisis de datos en un contexto interdisciplinario, promoviendo el pensamiento crítico, la autonomía en la investigación y la colaboración. Se espera que al finalizar el proyecto, los/as estudiantes sean capaces de integrar conocimientos técnicos con una perspectiva analítica para generar conocimiento útil en el ámbito educativo.

## DISEÑO DE LOS OBJETIVOS

- **Objetivo general:** Desarrollar un análisis del rendimiento académico estudiantil utilizando técnicas de ciencia de datos, con el fin de identificar patrones significativos que contribuyan a la toma de decisiones en el ámbito educativo.
- **Objetivos específicos:**
  1. Aplicar herramientas de procesamiento y exploración de datos con Python para analizar un dataset educativo.
  2. Describir estadísticamente las variables del conjunto de datos mediante medidas de tendencia central y dispersión.
  3. Identificar relaciones y patrones en los datos que puedan estar asociados con el rendimiento académico.
  4. Presentar los hallazgos mediante visualizaciones y reportes colaborativos en Jupyter Notebook.
- **Metas:**
  1. Analizar al menos 100 registros del dataset educativo.
  2. Generar 3 visualizaciones clave con insights sobre el rendimiento estudiantil.
  3. Entregar un informe de análisis con una extensión máxima de 10 páginas en un plazo de 6 semanas.
  4. Documentar el proyecto y el equipo de trabajo en Markdown.
  5. Participación activa del 100% de los integrantes del equipo en la elaboración del proyecto.



INSTITUTO SUPERIOR  
POLITÉCNICO CÓRDOBA

## 8. SELECCIÓN DE ACCIONES

OBJETIVO ESPECÍFICO	ACCIONES
1. Aplicar herramientas de procesamiento de datos con Python sobre un dataset educativo.	<ul style="list-style-type: none"> <li>- Cargar el dataset.</li> <li>- Limpiar y preparar datos.</li> <li>- Estandarizar variables.</li> </ul>
2. Describir y analizar estadísticamente el rendimiento estudiantil.	<ul style="list-style-type: none"> <li>- Calcular indicadores básicos (media, mediana, desvío).</li> <li>- Detectar correlaciones.</li> <li>- Agrupar datos por variables sociodemográficas.</li> </ul>
3. Representar visualmente la información para facilitar su comprensión.	<ul style="list-style-type: none"> <li>- Crear gráficos con Matplotlib/Seaborn.</li> <li>- Comparar grupos mediante boxplots, gráficos de barras, histogramas y gráficos de torta.</li> </ul>
4. Documentar y comunicar los hallazgos de manera clara y profesional.	<ul style="list-style-type: none"> <li>- Redactar conclusiones.</li> <li>- Elaborar un notebook explicativo.</li> <li>- Presentar informe final en equipo.</li> </ul>

## 9. CRONOGRAMA

CRONOGRAMA	SEMANA 1	SEMANA 2	SEMANA 3	SEMANA 4
<b>Objetivo 1</b>  Aplicar herramientas de procesamiento de datos con Python sobre un dataset educativo	Conformación de equipos Revisión del dataset Carga inicial de datos	Limpieza de datos Filtrado de registros Estandarización de variables	----	----
<b>Objetivo 2</b>  Describir y analizar estadísticamente el rendimiento estudiantil.	----	----	Cálculo de indicadores estadísticos Análisis de correlaciones.	----
<b>Objetivo 3</b>  Representar visualmente la información para facilitar su comprensión.	----	----	----	Creación de gráficos Comparación de grupos mediante visualizaciones.

<b>Objetivo 4</b>	----	----	----	Redacción completa del notebook explicativo Armado del informe final.
Documentar y comunicar los hallazgos de manera clara y profesional.				

## 10. PRODUCTO FINAL:

### Cumplimiento de los objetivos

A lo largo del análisis de datos, se han aplicado herramientas de ciencia de datos para examinar el rendimiento académico estudiantil. Los objetivos planteados se han cumplido de la siguiente manera:

- **Procesamiento y exploración de datos en notebook con Python:** se realizó una limpieza exhaustiva del dataset, eliminando valores erróneos y asegurando la integridad de los datos.
- **Análisis estadístico de las variables:** se calcularon medidas de tendencia central y dispersión, identificando patrones clave en asistencia, GPA, hábitos de sueño y participación extracurricular.
- **Identificación de relaciones y patrones:** se analizó la correlación entre variables, destacando la influencia de las horas de estudio en el GPA, la variabilidad en asistencia y el impacto del descanso en el rendimiento académico.
- **Visualización y reporte colaborativo:** se generaron gráficos descriptivos y análisis interpretativos en el notebook, permitiendo visualizar tendencias y facilitar la toma de decisiones educativas.



## **Estadísticas Descriptivas**

Antes de adentrarnos en la identificación de patrones y correlaciones, se realizó un análisis estadístico general del conjunto de datos con el fin de comprender mejor las características globales de los estudiantes evaluados. Este diagnóstico inicial permitió establecer un punto de partida sólido para los análisis posteriores.

### ❖ **Variables numéricas**

Se examinaron cinco variables continuas: edad, GPA, tasa de asistencia, horas de estudio por semana y horas de sueño por noche. A través del cálculo de medidas de tendencia central y dispersión, se identificaron los siguientes valores representativos:

- **Edad:**

Media: 29 años | Mediana: 28.5 | DE: 6.80

Distribución relativamente simétrica, representando un grupo de adultos jóvenes típicos en educación superior.

- **GPA (Promedio Académico):**

Media: 1.99 | Mediana: 1.91 | DE: 1.19

Distribución ligeramente variable, con diferencias entre carreras.

- **Tasa de asistencia:**

Media: 73.0% | Mediana: 71% | DE: 14.60

Se observa variabilidad significativa en la asistencia, lo que sugiere que algunos estudiantes no asisten regularmente.

- **Horas de estudio semanales:**

Media: 21.3 | Mediana: 21 | DE: 11.89

Existe un rango óptimo de estudio en 21-25 horas, asociado con mejor rendimiento.

- **Horas de sueño por noche:**

Media: 6.83 | Mediana: 7 | DE: 3.22

Algunos estudiantes duermen solo 1-2 horas, lo que puede afectar su desempeño.

### ❖ **Asimetría y distribución**

El análisis de skewness reveló los siguientes comportamientos:

- El GPA mantiene una distribución equilibrada.
- La asistencia y las horas de estudio presentan sesgos moderados, indicando diferencias en los hábitos académicos.
- Las horas de sueño tienen una dispersión alta, reflejando patrones de descanso desiguales entre estudiantes.

#### ❖ **Variables categóricas**

Se incluyeron variables cualitativas para entender mejor la composición del grupo estudiado:

- **Género:**  
Female (44.90%) | Male (44.27%) | Other (10.83%)
- **Carrera:**  
Predominan Engineering (15.21%), Architecture (15.10%), Law (14.02%), Psychology (13.70%), Medicine (13.16%) y Administration (13.05%).
- **Estado de salud mental reportado:**  
Good (31.46%) | Average (30.31%) | Bad (28.96%) | Sin especificar (9.27%)
- **Becas:**  
True (50.38%) | False (49.62%)
- **Acceso a Internet:**  
True (51.21%) | False (48.79%)

El análisis de estas variables permitió identificar un grupo estudiantil diverso. La representación equilibrada sugiere una distribución inclusiva. Se detecta que más del 50% cuenta con becas o acceso a internet, aunque una proporción significativa no dispone de estos recursos, lo que podría impactar en su desempeño académico.

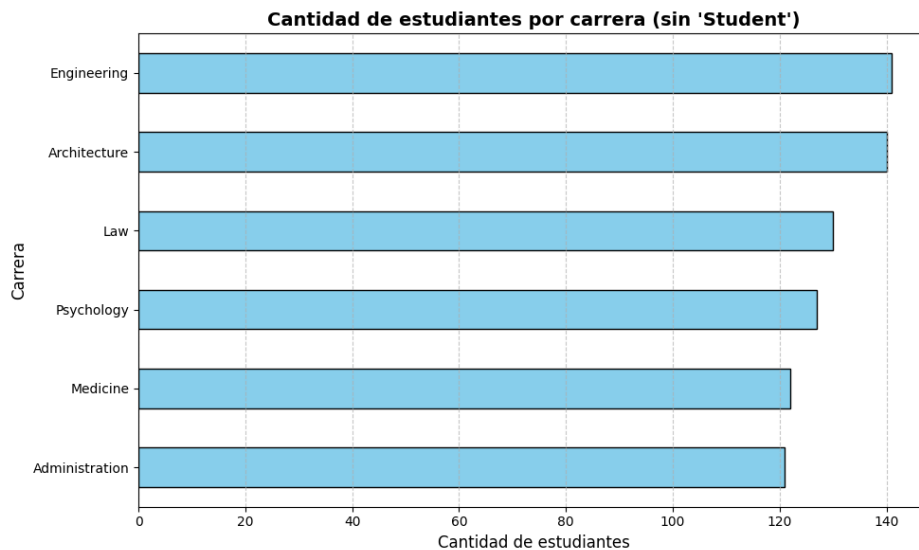
El estado de salud mental no mostró una relación clara con el GPA, indicando la influencia de múltiples factores sobre el rendimiento estudiantil.

### **Resumen de hallazgos clave**

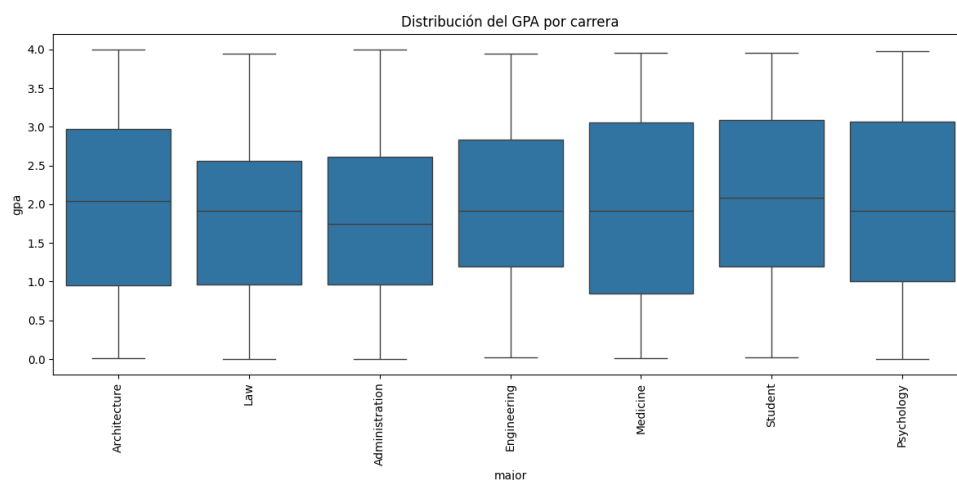
A través del análisis exploratorio y descriptivo de los datos, se han identificado varias tendencias relevantes:

- **Distribución por carrera:** la distribución de estudiantes por carrera se mantiene equilibrada, con Engineering (141), Architecture (140) y Law (130) como las

áreas con mayor cantidad de alumnos, seguidas de Psychology (127), Medicine (122) y Administration (121).

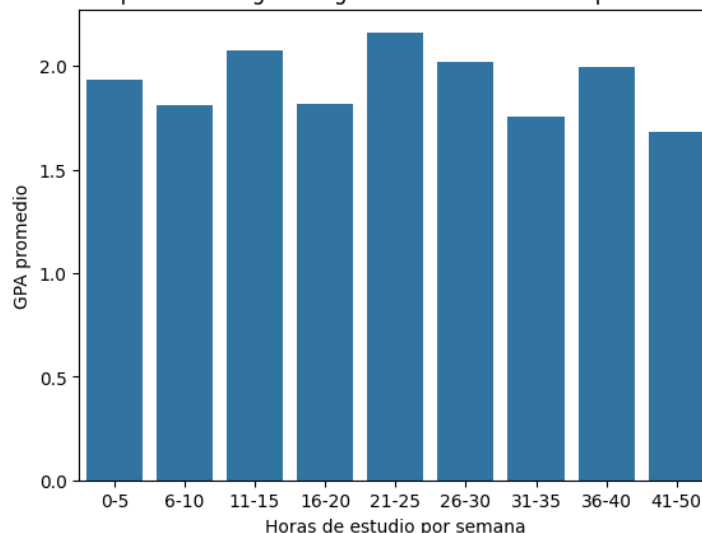


- **Distribución del GPA:** la mayoría de los estudiantes presentan un GPA moderado, oscilando entre 1.80 y 2.10, con diferencias por carrera.



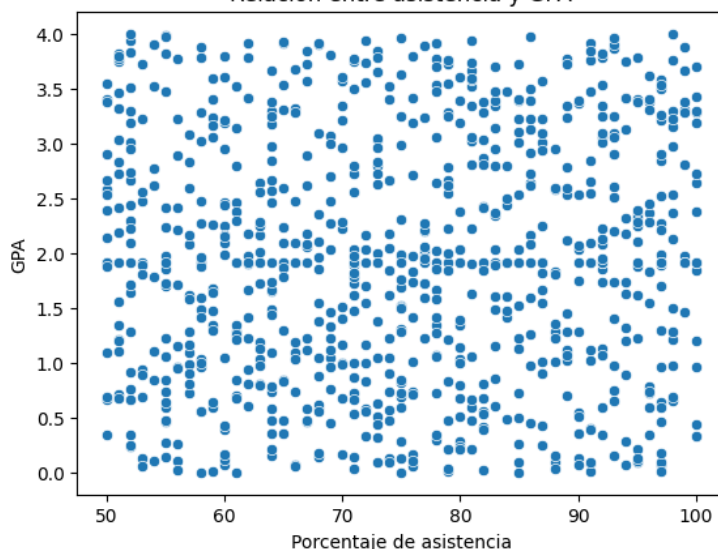
- **Horas de estudio y rendimiento académico:** se detectó un rango óptimo (21-25 horas semanales) asociado con mejores resultados. Sin embargo, exceder las horas de estudio no garantiza un mejor desempeño.

GPA promedio según rangos de horas de estudio por semana



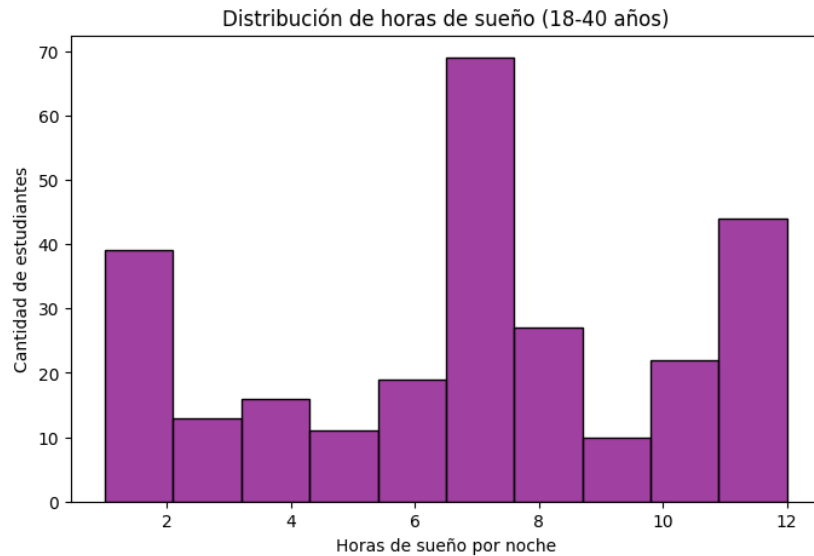
- **Asistencia a clases:** aunque los estudiantes con mayor asistencia tienden a tener un GPA más alto, la relación no es completamente lineal, lo que indica que otros factores influyen en el rendimiento.

Relación entre asistencia y GPA

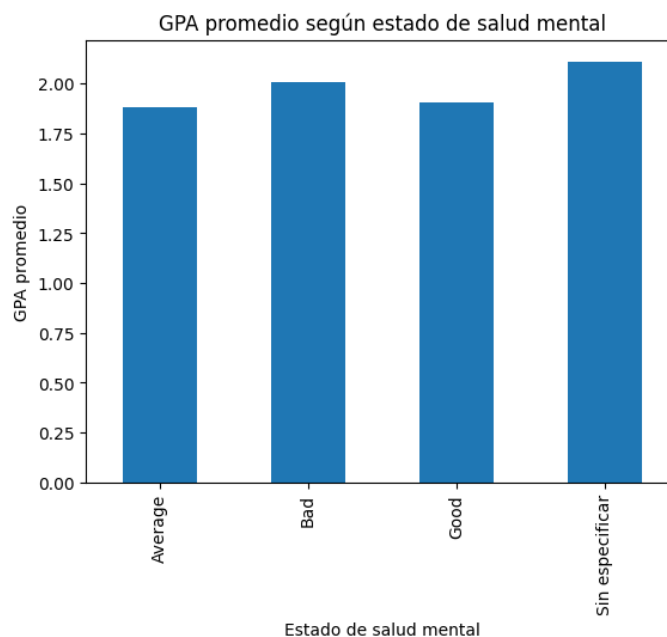


- **Hábitos de sueño:** la mayoría de los estudiantes duerme entre 5 y 9 horas por noche, con un promedio de 6.83 horas, ligeramente por debajo del rango recomendado de 7-9 horas. La dispersión en los hábitos de sueño es considerable (DE: 3.22 horas), con valores que van desde 1 hasta 12 horas, reflejando una alta heterogeneidad en los patrones de descanso. El valor más frecuente es 7 horas, seguido de 8, 10 y 12 horas, mientras que un grupo significativo reporta dormir solo 1-2 horas, lo que podría deberse a errores en el registro, hábitos poco saludables o situaciones particulares. Aunque la mayoría mantiene un patrón de sueño dentro del rango saludable, la variabilidad

observada sugiere que el descanso puede influir en el bienestar y el rendimiento académico.



- **Estado mental:** no se encontró una relación significativa entre la percepción del estado de salud mental y el rendimiento académico, lo que sugiere que hay otros factores más determinantes en el desempeño.



## **Impacto y aplicación en la comunidad educativa**

El análisis realizado sobre el rendimiento académico estudiantil aporta información valiosa que puede ser aplicada directamente en contextos educativos. Al convertir los datos en conocimiento comprensible y accionable, se abre la posibilidad de mejorar significativamente las prácticas pedagógicas, la gestión institucional y el bienestar del alumnado.

### **Utilidad del análisis en la comunidad educativa:**

Este estudio permite a las instituciones comprender mejor los factores que influyen en el rendimiento académico, como las horas de estudio, la asistencia, el descanso y el acceso a recursos digitales. Gracias a ello, es posible diseñar intervenciones más efectivas, identificar estudiantes en riesgo y fomentar hábitos saludables que impacten positivamente en el aprendizaje.

Acciones que podrían desarrollarse a partir de estos hallazgos:

- **Informe educativo para directivos y docentes:** un documento de síntesis con los hallazgos más relevantes, destinado a guiar decisiones pedagógicas y estrategias de acompañamiento estudiantil.
- **Folleto digital educativo:** material visual y accesible que resume los descubrimientos principales del análisis, pensado para su difusión en redes sociales, plataformas virtuales o carteleras institucionales.
- **Repositorio en GitHub:** publicación del código fuente, la metodología y los resultados del análisis, fomentando la transparencia y permitiendo que otros investigadores o educadores repliquen y amplíen el estudio.
- **Charlas y talleres en instituciones:** espacios de capacitación y reflexión basados en los resultados obtenidos, orientados a docentes, estudiantes y personal educativo, con foco en la mejora del rendimiento académico y la promoción de estilos de vida saludables.

### **Objetivo principal de la aplicación:**

Transformar los datos recopilados en herramientas concretas para la toma de decisiones educativas informadas, con el fin de promover entornos de aprendizaje más equitativos, inclusivos y centrados en el bienestar integral del estudiante.

## 11. BIBLIOGRAFÍA: Normas APA, 7ma edición.

ISPC. (s.f.). *Pandas*.

[https://acceso.ispc.edu.ar/pluginfile.php/217730/mod\\_page/content/4/Pandas.pdf](https://acceso.ispc.edu.ar/pluginfile.php/217730/mod_page/content/4/Pandas.pdf)

ISPC. (s.f.). *Jupyter Notebook*.

[https://acceso.ispc.edu.ar/pluginfile.php/217729/mod\\_page/content/5/jupyterNotebook.pdf](https://acceso.ispc.edu.ar/pluginfile.php/217729/mod_page/content/5/jupyterNotebook.pdf)

ISPC. (s.f.). *Procesamiento de datos*.

[https://acceso.ispc.edu.ar/pluginfile.php/217729/mod\\_page/content/5/Procesamiento%20de%20datos-10-17.pdf](https://acceso.ispc.edu.ar/pluginfile.php/217729/mod_page/content/5/Procesamiento%20de%20datos-10-17.pdf)

ISPC. (s.f.). *Diagrama de Caja y Bigotes*.

[https://acceso.ispc.edu.ar/pluginfile.php/217751/mod\\_page/content/4/Diagrama%20de%20Caja%20y%20Bigotes.pdf](https://acceso.ispc.edu.ar/pluginfile.php/217751/mod_page/content/4/Diagrama%20de%20Caja%20y%20Bigotes.pdf)

Mockaroo. (s.f.). *Mockaroo: Realistic Data Generator*. <https://www.mockaroo.com/>

W3Schools. (s.f.). *Pandas Tutorial*.

<https://www.w3schools.com/python/pandas/default.asp>

W3Schools. (s.f.). *Matplotlib Tutorial*.

[https://www.w3schools.com/python/matplotlib\\_intro.asp](https://www.w3schools.com/python/matplotlib_intro.asp)