

Ciencia de datos en la escuela: Herramientas y Estrategias aplicadas en experiencias en La Plata

Sofía Martín¹, Claudia Banchoff¹, Paula Venosa¹, Liliana Hurtado²

¹LINTI - Laboratorio de Investigación en Nuevas Tecnologías Informáticas.

Facultad de Informática. Universidad Nacional de La Plata

Calle 50 esq. 120, 2do Piso. Tel: +54 221 4223528

{smartin, cbanchoff, pvenosa}@info.unlp.edu.ar}

²Facultad de Ciencias Económicas, Jurídicas y Sociales. Universidad Nacional de Salta

lilianahur@hotmail.com

Resumen

Existen numerosas experiencias enfocadas en la integración de la Ciencia de la Computación en las escuelas. Esto da lugar al trabajo con otras temáticas relacionadas que han cobrado mucha relevancia en la actualidad, como es la Ciencia de Datos. Esta disciplina permite integrar, en modo transversal, aspectos de programación de computadoras. aplicado al análisis de datos.

El objetivo de este trabajo es presentar los aspectos más destacados de una primera experiencia llevada a cabo en la ciudad de La Plata, en donde se trabajaron aspectos básicos de la Ciencia de Datos con docentes y estudiantes de escuelas. Se describe la metodología llevada a cabo, junto con las herramientas utilizadas y los aprendizajes obtenidos de la experiencia.

Palabras clave: Ciencia de Datos, nivel secundario, Python, Jupyter notebook, Streamlit

1. Introducción

En los últimos años se ha trabajado intensamente, en muchos países, para integrar diversos aspectos de la Ciencia de la Computación en las escuelas. Se han escrito numerosos trabajos sobre los beneficios de incorporar el pensamiento computacional desde edades tempranas[1] y es campo de investigación sobre las distintas estrategias didácticas sobre lo mismo[2]. La Ciencia de Datos es una disciplina de gran relevancia en la actualidad que se encuentra relacionada a la Ciencia de la Computación ya que comparten

contenidos y habilidades que brindan en las escuelas. Esta disciplina tiene una estrecha relación con la recopilación y análisis de datos en diversos ámbitos, como la industria, los dispositivos digitales, la salud y la educación. Las aplicaciones digitales juegan un papel importante en la generación de grandes cantidades de datos, dejando un rastro de preferencias de usuarios que es crucial tener en cuenta, especialmente entre los jóvenes. Estos datos, disponibles en fuentes públicas, ofrecen oportunidades para explorar temas curriculares y fomentar el pensamiento crítico y computacional a través de proyectos educativos. Existen distintos portales en los cuales es posible descargar conjuntos de datos sobre casi cualquier tema: desde información de salud como nivel de nutrición o enfermedades, hasta opiniones sobre videojuegos, películas o interacciones en posteos públicos de las redes sociales.

Hay varias herramientas con las que se puede procesar estos conjuntos de datos, pero es posible plantear actividades educativas aprovechando las distintas iniciativas enfocadas principalmente en la enseñanza de pensamiento computacional y programación. La enseñanza de la programación en las escuelas está enfocada principalmente en el uso de lenguajes visuales basados en bloques como Scratch o MIT APP Inventor, o Python, como un ejemplo de los lenguajes textuales que se utilizan[3].

Las dos alternativas más utilizadas, que permiten la recuperación, análisis y presentación de resultados que se basan en el uso de los lenguajes de licencias libres, son Python y R. En ambos casos se cuenta con aplicaciones como Jupyter Notebook⁶ en el

⁶ <https://jupyter.org/>

caso de Python o R-STUDIO⁷ en el caso de R, que permiten el trabajo dinámico e interactivo con los datos.

Teniendo en cuenta las experiencias en las escuelas, se pensó en abordar el trabajo con datos desde la programación en Python utilizando librerías como pandas⁸, para el acceso y análisis de los mismos, y Matplotlib⁹ para la visualización de los resultados. Esto permite acceder a los datos en los formatos abiertos más utilizados, como ser CSV y JSON y elegir entre los múltiples tipos de visualizaciones que provee la librería Matplotlib[4].

Respecto a las estrategias didácticas para llevar la Ciencia de Datos a la escuela secundaria, se indagaron experiencias similares y se realizó una propuesta en la cual se trabajó con dos escuelas de la región de La Plata, provincia de Buenos Aires, Argentina durante el año 2022. Asistieron a los talleres docentes de quinto y sexto año y estudiantes de 15 a 17 años de escuelas técnicas. En las siguientes secciones de este trabajo se detallan las experiencias obtenidas y lo aprendido de las mismas.

2. La Ciencia de Datos y la importancia de los datos abiertos

La Ciencia de Datos, como campo de estudio, permite planificar, adquirir y extraer conocimientos significativos a partir de los datos[5]. Este campo interdisciplinario está estrechamente relacionado con la Ciencia de la Computación y las Estadísticas, situación que facilita la aplicación de técnicas de programación en el análisis de datos en diferentes áreas de estudio. La evolución de la Ciencia de Datos, asociado inicialmente al campo de la estadística[6] ha respondido a la creciente necesidad de integrar eficazmente los tres campos disciplinares que se entrecruzan y poniendo en evidencia la importancia de la interdisciplinariedad[7]:

- La Ciencia de la Computación, la cual proporciona las habilidades y técnicas para procesar y manipular eficientemente grandes conjuntos de datos.

- Las Estadísticas, que brinda los fundamentos teóricos para aplicar métodos estadísticos y algoritmos de aprendizaje automático, permitiéndonos extraer información significativa.
- La Disciplina específica que corresponde al conocimiento especializado que describe el dominio del problema. Estos campos pueden corresponder con ciencias especializadas como también los compartimentos disciplinares de las cajas curriculares.

El crecimiento exponencial en la magnitud de datos disponibles, que corresponde principalmente al uso masivo de dispositivos digitales y el Internet de las Cosas, ha impulsado la necesidad de automatizar los procesos y comprender de manera ágil los análisis requeridos. Diversos estudios han realizado estimaciones y proyecciones sobre este crecimiento en los próximos años. Un ejemplo de ello es el informe de Visual Capitalist[8], que estima que para el año 2025 la cantidad de datos generada diariamente llegaría a la impresionante cifra de 474,112TB. La enorme cantidad de datos disponibles requiere enfoques más eficientes y avanzados para extraer información significativa. De este modo, la realidad manifestada por este crecimiento, destaca la automatización como una herramienta clave en este proceso.

Esta expansión de la Ciencia de Datos ha tenido un impacto significativo en diversos ámbitos, como el comercial, la salud, el gubernamental, el educativo, entre otros. Se ha generado una mayor demanda de trabajos relacionados con estas habilidades debido a su versatilidad de aplicación y los beneficios que ofrecen en la toma de decisiones y la mejora de políticas, la industria, la salud, la educación, el cuidado del ambiente, gobierno y en las economías regionales[9]. Algunas de las aplicaciones concretas del análisis de datos en los ámbitos más comunes son:

- En el ámbito comercial permite conocer las preferencias de los consumidores y, como consecuencia, personalizar estrategias de mercado.

7 <https://www.r-studio.com/es/>

8 <https://pandas.pydata.org/>

9 <https://matplotlib.org/>

- En el campo de la salud favorece identificar patrones que respaldan el diagnóstico temprano de enfermedades, redundando en una atención médica más eficiente y personalizada.
- En el ámbito gubernamental, el uso de datos colabora en políticas públicas que mejoren la calidad de vida de los ciudadanos.
- En el ámbito educativo se ha centrado en el estudio del rendimiento estudiantil, lo que ha permitido identificar áreas de mejora y desarrollar estrategias educativas más efectivas. Al favorecer la comprensión de los factores que afectan el éxito académico, es posible implementar intervenciones tempranas y personalizadas para ayudar a los estudiantes a alcanzar su máximo potencial.

La disponibilidad de datos abiertos ha revolucionado la forma en que se abordan problemas sociales, económicos y ambientales, proporcionando una base sólida para la toma de decisiones fundamentadas y contribuyendo a un mundo más sostenible y equitativo.

En muchos casos, estos desafíos están relacionados con la identificación de problemáticas complejas y la generación de soluciones en línea con los Objetivos de Desarrollo Sostenible (ODS) propuestos por las Naciones Unidas[10].

A modo de ejemplo, podemos mencionar el portal de datos abiertos de la República Argentina¹⁰, o el del Banco Mundial¹¹ o el portal Kaggle¹² en el cual se encuentran numerosos conjuntos de datos relacionados al mundo del entretenimiento, deporte, política, entre otros.

2.1 El proceso del análisis de datos

El análisis de datos en el ámbito de la Ciencia de Datos es un proceso fundamental que convierte datos en conocimiento, crucial para la toma de decisiones informadas y la generación de valor a partir de la información. Para realizar un análisis crítico y asegurar la inclusión de datos válidos y significativos, es esencial

comprender los pasos que subyacen en este proceso.

El proceso para generar informes o visualizar situaciones que no son transparentes a simple vista se compone de una serie de pasos fundamentales, que forman un ciclo continuo de trabajo. Estas etapas incluyen: establecer un objetivo claro, obtener los datos necesarios, preparar y analizar los datos, y finalmente, comunicar los resultados de manera clara y efectiva mediante gráficos y visualizaciones apropiadas. La Fig. 1 muestra las etapas que involucra este proceso, el cual puede sintetizarse en forma simple en 4 fases:

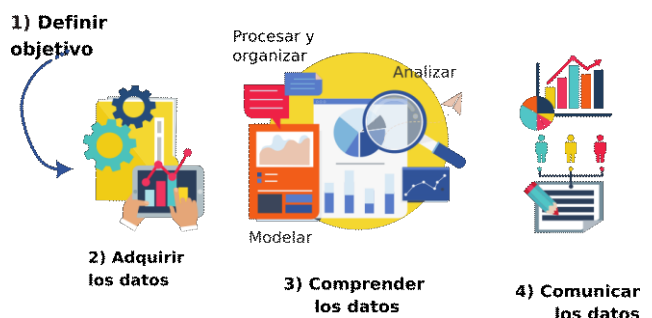


Fig. 1. Proceso del análisis de datos

Aunque el proceso puede variar en complejidad según el ámbito de aplicación específico, es esencial comprender y practicar cada uno de estos pasos para abordar de manera coherente preguntas de investigación apropiadas y lograr un análisis válido y significativo que facilite la comprensión de los resultados[11].

3. El pensamiento computacional y el análisis crítico como foco para trabajar CD en la escuela

La Ciencia de Datos ha emergido con características distintivas que han impulsado su integración en la educación secundaria. Esta tendencia se justifica no solo por la relevancia de comprender los conceptos que componen la CD, sino también por la oportunidad que brinda a los estudiantes para desarrollar habilidades valiosas y conectar su aprendizaje con el mundo real. Al enfrentarse a preguntas y desafíos del mundo actual utilizando datos, los estudiantes experimentan un mayor compromiso, motivación y la posibilidad de obtener

10 <https://datos.gob.ar>

11 <https://databank.worldbank.org>

12 <https://www.kaggle.com>

resultados significativos y tangibles. Entonces, dicho proceso fomenta el pensamiento crítico y las habilidades de resolución de problemas en un entorno digital impulsado por los datos[12]. En los últimos años, nuestro país ha avanzado en la integración de contenidos de la Ciencia de la Computación en diversos proyectos gubernamentales y propuestas pedagógicas para los planes de estudio[13]. Estas iniciativas han demostrado ser de gran relevancia, ya que permiten a los estudiantes adentrarse en el conocimiento de la Ciencia de Datos, un campo que guarda estrecha relación con los conceptos previamente abordados.

Uno de los aspectos destacados de estas propuestas ha sido la adopción del lenguaje de programación Python como una herramienta para enseñar los fundamentos de la Ciencia de la Computación¹³. Python ha ganado una amplia aceptación en la comunidad científica y se ha convertido en uno de los lenguajes más utilizados en este campo. Su sintaxis sencilla y legible, combinada con su amplia gama de bibliotecas especializadas, lo convierten en una opción ideal para introducir a los estudiantes en la Ciencia de Datos. Además, Python cuenta con una comunidad de desarrolladores activa y una gran cantidad de recursos en línea, facilitando el acceso a materiales educativos y el apoyo necesario para el aprendizaje.

3.1 Habilidades en juego en el proceso del análisis

En los procesos de aprendizaje significativo, se destaca la importancia de cómo los estudiantes interiorizan el conocimiento y establecen relaciones, construyendo interés hacia él. La construcción, producción y aplicación del conocimiento implica un diálogo continuo que requiere esfuerzos tanto individuales como colectivos. Esta dinámica dialógica promueve un entorno colaborativo de aprendizaje, transformando tanto los roles de estudiantes como de docentes. En este contexto, implica captar y reinterpretar lo que está sucediendo, siendo una parte esencial del proceso de asimilación del conocimiento, donde cada estudiante posee la capacidad de elegir

voluntariamente su grado de involucramiento en este proceso[14].

La integración de problemas de la vida real en la enseñanza recurriendo a la Ciencia de Datos, proporciona el escenario propicio para llevar a cabo la transposición didáctica, lo cual implica interacciones entre docentes, estudiantes y el contenido de estudio[15]. El aprendizaje significativo se nutre en la posibilidad de cuestionar ideas preconcebidas y examinar evidencias. El proceso así descrito contribuye a evitar la mirada ingenua, familiar y simple al momento de estudiar un tema de la currícula.

Además de desafiar los conocimientos previos y actuales de los estudiantes, la transposición didáctica fomenta la adquisición de habilidades fundamentales ofrecidas por el marco teórico-conceptual de la Ciencia de Datos. Esto incluye el pensamiento computacional y el análisis interdisciplinario y colaborativo en grupo, facilitando una comprensión más profunda para crear soluciones mediante la aplicación significativa de algoritmos y estructuras de datos [16].

Adquirir estos conocimientos y habilidades no solo implica dominar herramientas actuales, sino también desarrollar la capacidad de enfrentar desafíos mediante la experimentación con soluciones novedosas, incluso en contextos aún no explorados.

La colaboración y el trabajo en equipo adquieren una posición central en el desarrollo de habilidades en el campo de la Ciencia de Datos. Al abordar proyectos integrados con diversos grados de complejidad, se crea un entorno propicio donde cada estudiante aporta habilidades y perspectivas únicas al equipo. De este modo, se promueve la creación creativa de soluciones innovadoras y eficientes a problemas planteados, reconociendo que cada individuo puede encontrar su propio rol dentro del equipo, maximizando su contribución y participación en el trabajo grupal, además de sumar autoestima y seguridad.

Estas habilidades que se fomentan permiten reforzar las etapas del proceso cognitivo como recordar, entender, aplicar, analizar, evaluar y crear. Cada una de las etapas del proceso de

análisis implica acciones específicas que se alinean con las etapas del proceso cognitivo mencionadas. A medida que se avanza en el ciclo, se activan distintas habilidades cognitivas, que enriquecen el proceso de análisis y facilita una comprensión más profunda de los datos y por lo tanto del contenido curricular[7].

El proceso de análisis de datos incluye etapas que involucran tomar decisiones éticas en cuanto a la privacidad, integridad e imparcialidad al trabajar con información sensible, conduciendo a abordar el análisis de datos y la comunicación de los resultados de manera responsable y ética, reforzando las etapas del proceso cognitivo y facilitando una comprensión más profunda del contenido curricular. Es crucial desarrollar recursos educativos adaptados y promover el trabajo colaborativo entre docentes para llevar a cabo proyectos transversales que permitan a los estudiantes aplicar los conocimientos adquiridos en diferentes áreas de estudio.

3.2 Experiencias aplicadas en escuelas

En los últimos años, las políticas educativas se han centrado en estrategias de acceso a la tecnología, con diferentes enfoques, en algunos casos mediante la distribución de laptops o tablets en modelos 1 a 1. Estas iniciativas han cobrado relevancia en la última década y han sido implementadas en varios países.

En un contexto global, países como Estonia y Finlandia han desarrollado proyectos de enseñanza de programación que abarcan desde la educación primaria hasta la secundaria. Por otro lado, en lugares como Inglaterra, la enseñanza de la programación y las Ciencias de la Computación se ha convertido en una materia específica y fundamental dentro de los planes de estudio[17].

En relación al contenido de Ciencia de Datos, se encuentra todavía en desarrollo para la enseñanza en escuelas secundarias y, en algunas experiencias, implementan propuestas simplificadas. Por ejemplo, la exploración de conjuntos de datos ya cargados en herramientas disponibles. Estas herramientas facilitan el análisis a través de la generación de gráficos

preestablecidos con la posibilidad de ajustar los parámetros. Algunas de estas herramientas han sido desarrolladas por el grupo Concord Consortium¹⁴, que ofrece propuestas pedagógicas para respaldar el uso en diferentes modalidades. Estas herramientas pueden ser instaladas localmente o utilizadas en línea.

En la región de Latinoamérica se observa, que a pesar de los avances significativos en la digitalización, todavía persisten brechas importantes que se ven influenciadas tanto por aspectos económicos como por cambios en las políticas educativas de cada país. Entre los programas gubernamentales inicialmente orientados a la reducción de la brecha digital se destacan el Programa Nacional de Informática Educativa (PRONIE MEP-FOD) en Costa Rica, el Plan Ceibal en Uruguay, Conectar Igualdad en Argentina, entre otros. En muchos casos, con problemáticas de acceso a Internet, lo cual dificulta la implementación de los proyectos tecnológicos educativos.

Si bien los contenidos trabajados en estos programas y en proyectos relacionados han ido evolucionando, el enfoque principal se centra en el desarrollo del pensamiento crítico en el uso de la tecnología, a través de la introducción de contenidos relacionados con el pensamiento computacional, la programación y la robótica.

A nivel nacional, en Argentina, como hemos mencionado anteriormente, se han implementado planes gubernamentales a través de la entrega de dispositivos tecnológicos que incluyen propuestas educativas difundidas a través del sitio Educar¹⁵. Estas propuestas han incorporado el uso del lenguaje Python en diversas instancias educativas.

No obstante, es importante señalar que, a pesar de las iniciativas a nivel nacional, no todas las provincias han incluido los contenidos de Ciencia de la Computación en sus planes curriculares. Un ejemplo de una provincia que ha avanzado en la incorporación de estos contenidos es Neuquén, donde se han desarrollado cambios curriculares específicos[18].

En este contexto, surgen desafíos significativos, especialmente en lo que respecta a la formación de docentes. La naturaleza cambiante y

14 <https://concord.org/our-work/focus-areas/data-science-education>

15 <https://www.educ.ar/buscador?q=python>

especializada de estos contenidos ha planteado dificultades en su incorporación en las escuelas. Esto no se limita únicamente a la inclusión de los contenidos en el currículo, sino que también involucra la necesidad de mantenerlos actualizados en un entorno tecnológico en constante evolución.

Contar con especialistas que puedan brindar apoyo y orientación durante estos procesos es igualmente crucial. Además, se requiere simplificar las propuestas y herramientas educativas relacionadas con estos contenidos. Esto simplificará su adopción en el entorno educativo y permitirá una integración más efectiva. La transversalidad es un enfoque valioso, pues permite abordar problemas reales que están relacionados con múltiples disciplinas, enriqueciendo así la experiencia de aprendizaje.

4. Ciencia de datos en la escuela secundaria argentina: una primera experiencia

Como se mencionó en la sección previa, en la República Argentina se viene trabajando en la incorporación a la currícula escolar de temas de Ciencias de la Computación, aunque en la mayoría de los casos, se centran en aspectos básicos de programación. Esta temática se aborda desde distintos enfoques y, dependiendo del nivel escolar y tipo de establecimiento solo se centran en aspectos muy básicos utilizando mayormente programación en lenguajes visuales basados en bloques o en la programación con lenguaje Python[3]

Aprovechando estas iniciativas y los contactos establecidos con instituciones de la región de La Plata y Gran La Plata, en la convocatoria de proyectos de extensión universitaria 2021, “Universidad, Cultura y Territorio” de la Secretaría de Políticas Universitarias (SPU), Ministerio de Educación de la Nación Argentina, se presentó un primer proyecto para trabajar aspectos básicos de la Ciencia de Datos en la escuela secundaria.

El objetivo principal del proyecto fue implementar un conjunto de propuestas que permitieran la incorporación de esta temática en forma transversal a los contenidos, eligiendo algunas áreas específicas de acuerdo a la escuela, y proponiendo el análisis de datos

como una forma distinta de aplicar la programación.

Para esto se planificó, en primer lugar, una capacitación con los docentes de las dos escuelas participantes del proyecto, en donde se abordó el uso de las herramientas informáticas comúnmente utilizadas, como Jupyter Notebook, y la programación en Python.

En esta experiencia participaron, en el armado y dictado de los talleres, docentes y estudiantes de la Facultad de Informática de la Universidad Nacional de La Plata, docentes y estudiantes de dos escuelas secundarias de la región: las escuelas EEST N°9 y la EES N°4 de la ciudad de La Plata.

Esta experiencia permitió, no solo establecer un punto de partida para la incorporación de esta temática en el aula de la escuela secundaria, sino también sirvió de puntapié inicial para un trabajo de investigación sobre herramientas y didácticas asociadas a la enseñanza de esta temática en el nivel medio.

5.1. Las actividades realizadas

En el marco del proyecto se elaboraron contenidos y se implementó una capacitación destinada a los docentes del nivel secundario, previo a acordar un cronograma y los contenidos a abordar, de acuerdo a las asignaturas dictadas por cada docente. Esta capacitación se planteó en una modalidad de taller para los docentes, que involucró cuatro encuentros: dos dictados en modalidad virtual y los dos últimos en forma presencial. En estos encuentros se presentó la herramienta Jupyter Notebook y se trabajó con algunos aspectos básicos de la programación en Python y las librerías de pandas y Matplotlib que usamos en las actividades propuestas.

En estos encuentros participaron siete docentes de las dos escuelas participantes.

En este punto, cabe aclarar que la EEST N°9 dicta materias de programación de computadoras, por lo que sus docentes ya tienen formación en el tema, mientras que la EES N°4 no. Esto resultó ser un factor preponderante a la hora de analizar de qué manera es posible incorporar la Ciencia de Datos en el nivel medio. En el primer encuentro de la capacitación realizada se presentaron las herramientas y los objetivos del proyecto y se relevaron los posibles ejes temáticos donde introducir

conceptos de la Ciencia de Datos de acuerdo a las asignaturas dictadas por los docentes participantes. Entre los temas surgidos se seleccionaron algunos con los que se siguió trabajando a lo largo de los siguientes encuentros:

- Salud escolar y nacimientos en Argentina: se analizaron los datos disponibles en el portal de datos abiertos del Ministerio de Salud de la República Argentina correspondientes a la 3° Encuesta Mundial de Salud Escolar (EM-SE) 12 y el dataset correspondiente a las personas nacidas vivas registradas en la República Argentina entre los años 2005-2021.
- Información sobre conectividad en Argentina: se trabajó con información obtenida en el año 2019 sobre conectividad a Internet en la Argentina disponibles en el repositorio de datos abiertos de la nación.
- Ubicación de rampas en CABA: este ejemplo se propuso para plantear una forma distinta de visualizar los resultados de un análisis. En este caso se mostró sobre un mapa la ubicación y estado de rampas de accesibilidad relevado en la Ciudad de Buenos Aires en el año 2016.

En todos los casos, la propuesta de trabajo implicó, como primer paso, definir claramente el objetivo a analizar, es decir, qué es lo que se requiere estudiar y conocer. Luego se fue iterando en el proceso del análisis introduciendo las distintas herramientas y funcionalidades y abordando las dificultades que fueron apareciendo con posibles estrategias que podrían ser replicadas en el aula si surgiera la misma situación.

Cabe destacar que todo el material de estas actividades y las propuestas de trabajo se encuentran disponibles en un sitio público generado en GitLab¹⁶ de manera tal que puedan ser utilizados en experiencias similares.

Independientemente de la temática abordada, este trabajo permitió también destacar la importancia de los datos abiertos y de las distintas formas de visualización de acuerdo al objetivo propuesto. En todos los casos se accedieron a los portales de datos abiertos gubernamentales, pero también se hicieron

algunas actividades con otros portales que contienen una gran variedad de datos para procesar, como ser Kaggle.

Respecto a las formas de visualización, se trató de mostrar no solo los típicos gráficos de barras y tortas, sino también algunos otros como gráficos de dispersión y líneas. En el caso del análisis de rampas en la Ciudad de Buenos Aires, se utilizó la librería Geopandas¹⁷ que permite el trabajo con mapas y, más específicamente, se hizo uso de los mapas abiertos del proyecto OpenStreetMap¹⁸, dando lugar nuevamente a las ventajas de que este tipo de información sea de libre disponibilidad.

Luego de la capacitación realizada con los docentes, se realizó un taller con 30 estudiantes de la EEST N°9. En esta ocasión se planteó un taller introductorio, que se dictó con la presencia de los docentes del establecimiento y en el cual se abordaron los conceptos básicos y se dejaron planteadas las actividades para su continuación en el ámbito de la escuela. Es importante destacar que esto pudo realizarse en un único encuentro debido a que este grupo ya contaba con experiencia en programación en Python.

5.2. Aprendizajes obtenidos de la experiencia

Como se mencionó anteriormente, este proyecto permitió establecer las bases de trabajo conjunto con docentes de escuelas secundarias para la incorporación de esta temática. Esto involucra dos áreas:

- Los contenidos o asignaturas que puedan involucrar el trabajo en la temática. En este sentido se vió claramente la dificultad que tienen las escuelas que no cuentan con un soporte para abordar temas como programación, en comparación con la escuela técnica, en donde estos temas se abordan en otras asignaturas. También se notó que la búsqueda de una temática transversal es complicada, debido principalmente a la falta de coordinación entre distintas áreas en la escuela.
- Las herramientas a utilizar. Claramente, introducir el análisis de datos directamente a través de un lenguaje textual como Python, requiere que los docentes a cargo

¹⁶https://ciencia_datos_escuela.gitlab.io

¹⁷ <https://geopandas.org/en/stable/>

¹⁸ <https://www.openstreetmap.org>

también tengan nociones de programación en dicho lenguaje, aspecto que no en todos los establecimientos se pueda considerar. Incluso, en aquellos casos donde se abordan estos temas de manera previa, se notó la necesidad de iniciar las actividades con algunas otras herramientas más simples o intuitivas.

Este proyecto también dejó algunos datos interesantes al interior del equipo de trabajo. Por un lado, se notó un alto grado de interés de los estudiantes integrantes del proyecto, quienes son estudiantes de las carreras de Informática, sobre el área del análisis de datos. La mayoría de los estudiantes convocados habían cursado la asignatura de Seminario de Lenguajes, opción Python, que se dicta en la Facultad de Informática de la UNLP, y no contaban con experiencia laboral previa. En dicha asignatura, en una encuesta inicial que se realiza al comienzo de la cursada, aproximadamente un 50% de los estudiantes mencionan al análisis de datos como uno de los motivos por los cuales les interesó cursar el seminario. Esto permitió revisar los contenidos de este seminario incorporando algunos temas adicionales.

Por otro lado, quedó en evidencia que es necesario utilizar otro tipo de herramientas para la introducción de la temática, especialmente en aquellas escuelas cuyos docentes no cuentan con una formación en nociones básicas de programación.

Esto dio lugar a un trabajo conjunto con otro grupo de becarios que investigaban herramientas educativas de software libre.

5.2 Dos herramientas visuales para introducir Ciencia de Datos en la escuela

Como se mencionó en la sección previa, la experiencia de trabajo con docentes de escuelas permitió investigar sobre otras herramientas que permitieran la introducción a la temática de una manera más sencilla. Como el propósito también es dar otro sentido a la enseñanza de la programación, se buscaron herramientas de uso simple, para luego continuar con la programación en Python, y no simplemente el uso de una planilla de cálculo, por ejemplo.

De esta manera se exploraron dos herramientas visuales: DataExplore¹⁹ y PandasGUI²⁰. Ambas aplicaciones son de código abierto, permitiendo a estudiantes y docentes usar, distribuir y, de ser necesario, personalizar y adaptar el software.

Tanto PandasGUI como DataExplore son herramientas visuales que se utilizan para trabajar con la librería pandas de Python, pero a través de una interfaz gráfica de usuario, permitiendo así la interacción con los datos de una manera más visual e intuitiva.

Index	Provincia	Partido	Localidad	Poblacion	ADSL	CABLEMODEM
0	BUENOS AIRES	25 de Mayo	25 de Mayo	23408	SI	SI
1	BUENOS AIRES	25 de Mayo	Agustín Maza	280	SI	SI
2	BUENOS AIRES	25 de Mayo	San Vicente	899	SI	SI
3	BUENOS AIRES	25 de Mayo	Ernestina	145	SI	SI
4	BUENOS AIRES	25 de Mayo	Gobernador Ugarte	547	SI	SI
5	BUENOS AIRES	25 de Mayo	Lucas Montemayor	481	SI	SI
6	BUENOS AIRES	25 de Mayo	Norberto de la Nostra	4324	SI	SI
7	BUENOS AIRES	25 de Mayo	Federales	1432	SI	SI
8	BUENOS AIRES	25 de Mayo	San Enrique	279	SI	SI
9	BUENOS AIRES	25 de Mayo	Valdes	279	SI	SI
10	BUENOS AIRES	9 de Julio	12 de Octubre	219	SI	SI
11	BUENOS AIRES	9 de Julio	9 de Julio	5694	SI	SI
12	BUENOS AIRES	9 de Julio	Adolfo Demarelli (Dist. Francisco Quiroga)	1933	SI	SI
13	BUENOS AIRES	Carlos Tejedor	Coronel	2075	SI	SI
14	BUENOS AIRES	Carlos Tejedor	Tinajas	526	SI	SI
15	BUENOS AIRES	Carlos Tejedor	Tres Arroyos (Dist. Cornejo)	5201	SI	SI
16	BUENOS AIRES	Coronel de Arce	Coronel de Arce	1275	SI	SI
17	BUENOS AIRES	Coronel de Arce	Puebla Guich	122	SI	SI
18	BUENOS AIRES	Coronel de Arce	Tres Arroyos	391	SI	SI
19	BUENOS AIRES	Castelli	Castelli	4639	SI	SI
20	BUENOS AIRES	Castelli	Carlos Guillermo	114	SI	SI
21	BUENOS AIRES	Castelli	Cerro de la Gloria	167	SI	SI
22	BUENOS AIRES	Chacabuco	Castillo	669	SI	SI

Fig. 2. Interfaz de PandasGUI

Con estas herramientas, es posible cargar y explorar datos, realizar transformaciones y manipulaciones de datos, crear visualizaciones y analizar estadísticas descriptivas sin necesidad de escribir código. Otra de las ventajas que tienen es que permiten a los usuarios exportar sus datos en varios formatos y compartir su trabajo.

La Figura 2 muestra la interfaz de usuario de PandasGUI, en donde se puede ver un dataset, en este caso corresponde al de conectividad en Argentina.

Puede observarse en la Figura 3 la interfaz de usuario de DataExplore, en donde se puede ver el mismo dataset mostrado en la Figura 2.

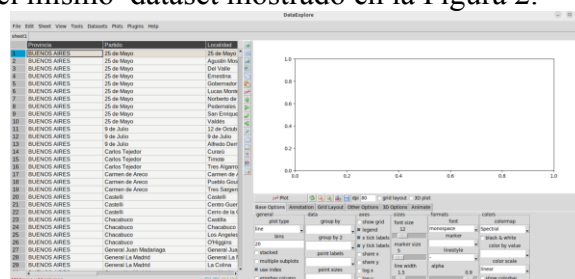


Fig. 3. Interfaz de DataExplore

Actualmente se han realizado actividades con la herramienta Streamlit²¹ la cual es una librería de Python que permite generar una aplicación web interactiva y está especialmente orientada

19 <https://pandastable.readthedocs.io/en/latest/description.html#the-dataexplore-application>

20 <https://github.com/adamero/pandasgui>

21 <https://streamlit.io/>

al trabajo con datos, facilitando su integración de forma sencilla con pandas y Matplotlib. Además de estas características, esta herramienta posibilita la visualización dinámica de mapas con datos geolocalizados.

Conclusiones

Las experiencias realizadas permitieron relevar que, en algunas escuelas, la temática de análisis de datos se trabaja de forma intuitiva sin profundizar en los conceptos subyacentes. Se suele utilizar planillas de cálculo con actividades asociadas principalmente a contenidos de estadísticas. Esto, al igual que las iniciativas de enseñar a programar con Python, dan un marco propicio para la inclusión de los restantes conceptos asociados a la Ciencia de Datos, adecuando las propuestas a los distintos contextos. Además, estas experiencias evidenciaron la necesidad de incluir herramientas visuales más simples para introducir el tema en aquellos establecimientos en donde no se ha trabajado aspectos de programación. Si bien PandasGUI y DataExplore presentaron otros desafíos en su aplicación, como ser la instalación de las mismas en las salas de computadoras, su análisis y uso permitió establecer otro ángulo de abordaje de la temática.

Este trabajo también demuestra la estrecha vinculación entre los tres pilares de la universidad: la extensión, la investigación y la docencia. Se estableció un nuevo grupo de investigación y, notando el gran interés de los estudiantes universitarios, se adecuaron algunos contenidos de asignaturas para incluir una introducción a esta temática.

Referencias

[1] Wing, J. M. (2006). Computational Thinking. *Communications of the ACM*, vol. 49, 33-35.

[2] Waite J. y Sentance S. Teaching programming in schools: A review of approaches and strategies. Raspberry Pi Foundation. Noviembre 2021. ISSN 2514-586X (19)

[3] Banchoff Tzancoff C., Martín S., Gómez S. y López F. Experiencias en robótica educativa. Diez años trabajando con escuelas argentinas. Claudia Aceptado para su publicación en el

Congreso TEYET 2019. San Luis, Argentina. Julio 2019.

[4] Hunter J., Dale D., Firing E., Droettboom M. Plot types. 2012.

[5] Donoho D.. 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, 26(4):745–766, October 2017. ISSN 1061-8600, 1537-2715.

[6] Cleveland W. S.. Data Science: an Action Plan for Expanding the Technical Areas of the Field of Statistics. *International Statistical Review*, 69(1):21–26, April 2001. ISSN 0306-7734, 1751-5823.

[7] Danyluk A. y Leidig P.. Computing Competencies for Undergraduate Data Science Curricula. ACM Data Science Task Force. (2021). volume ACM Data Science Task Force. Association for Computing Machinery., 2021. ISBN 978-1-4503-9060-6.

[8] Desjardins J.. How Much Data is Generated Each Day?, April 2019.

[9] Cao, L. (2017). Data science: a comprehensive overview. *ACM Computing Surveys (CSUR)*, 50(3), 1-42. doi: 10.1145/3076253.

[10] Gamez, M. J. (s. f.). Objetivos y metas de desarrollo sostenible. *Desarrollo Sostenible*.

[11] Cielen D., Meysman A. y Ali M. (2016). *Introducing data science: big data, machine learning, and more, using Python tools*. Manning Publications, Shelter Island, NY. ISBN 78-1-63343-003-7.

[12] Rao R, Desai Y. y Mishra K.. Data science education through education data: an end-to-end perspective. In 2019 IEEE Integrated STEM Education Conference (ISEC), pages 300–307, Princeton, NJ, USA, March 2019. IEEE. ISBN 978-1-72811-502-3.

[13] Jara, I., Hepp, P., Rodríguez, J., & Claro, M. (2019). Políticas y prácticas para la enseñanza de las Ciencias de la Computación en América Latina. Microsoft América Latina.

[14] Eraña, Á., & Barceló Aspeitia, A. A. (2016). El conocimiento como una actividad colectiva. *Tópicos (México)*, (51), 9-35.

[15] Castiblanco, J. E. M. (2014). La transposición didáctica del saber sabio al saber enseñado. Autor: Yves Chevallard. *Góndola, enseñanza y aprendizaje de las ciencias*, 9(2), 97-100. ISSN 2346-4712.

[16] Hicks, S. C., & Irizarry, R. A. (2018). A guide to teaching data science. *The American*

Statistician, 72(4), 382-391. ISSN 0003-1305, 1537-2731.

[17] UNESCO. Ciencias de la computación en los sistemas educativos de América Latina, 2017.

[18] Branchini, G., Cortez, M. M., Pedemonte, M. D. L. Á., & Rodríguez, J. (2019). Las ciencias de la computación en el currículum escolar. In XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019, Universidad Nacional de San Juan)