

Nombre: Celic Gabriel Hernández Archundia.		Matrícula: 2877240
Nombre del curso: Fundamentos de programación de Big Data.		Nombre del profesor: Sergio Arturo Damián Sandoval.
Módulo I		Actividad: Tarea 2
Fecha: 24/02/2023		
Bibliografía: <p>-Unir. (2020). “R vs Python: ¿cuál es mejor para análisis de datos?”. Recuperado el 4 de febrero del 2023 de: https://www.unir.net/ingenieria/revista/r-vs-python/#:~:text=R%20vs%20Python%3A%20%C2%BFcu%C3%A1l%20es%20mejor%20para%20el,principal%20ventaja%20es%20la%20facilidad%20en%20la%20</p> <p>-Edx. (2021). “9 principales lenguajes de programación para la ciencia de datos”. Recuperado el 4 de febrero del 2023 de: https://blog.edx.org/es/9-principales-lenguajes-de-programacion-para-la-ciencia-de-datos#:~:text=9%20principales%20lenguajes%20de%20programaci%C3%B3n%20de%20ciencia%20de,Java%20...%208%208.%20C%2FC%2B%2B%20...%20M%C3%A1s%20elementos</p> <p>-Platzi. (s/f). “Curso de Manipulación y Análisis de Datos con Pandas y Python”. Recuperado el 4 de febrero del 2023 de: https://platzi.com/clases/1794-pandas/27914-que-es-pandas/#:~:text=%C2%BFQu%C3%A9%20es%20pandas%3F%20En%20Computaci%C3%B3n%20y%20Ciencia%20de,operaciones%20para%20manipular%20tablas%20num%C3%A9ricas%20y%20series%20temporales.</p> <p>-Rondón, I. (2022). “¿Qué es Anaconda?”. Recuperado el 4 de febrero del 2023 de: https://eiposgrados.com/blog-python/que-es-anaconda/</p> <p>-Sas. (s/f). “Hadoop. ¿Qué es y por qué es importante?”. Recuperado el 4 de febrero del 2023 de: https://www.sas.com/es_mx/insights/big-data/hadoop.html</p> <p>-Anónimo. (2022). “Anaconda User Reviews”. Recuperado el 4 de febrero del 2023 de: https://slashdot.org/software/p/Anaconda/</p> <p>-Developer.mozilla.org. (s/f). “SQL”. Recuperado el 4 de febrero del 2023 de: https://developer.mozilla.org/es/docs/Glossary/SQL</p> <p>-Google. (s/f). “¿Qué es Colaboratory?”. Recuperado el 4 de febrero del 2023 de: https://research.google.com/colaboratory/intl/es/faq.html#:~:text=Colaboratory%2C%20</p>		

[o%20%22Colab%22%20para%20abreviar%2C%20es%20un%20producto,de%20aprendizaje%20autom%C3%A1tico%2C%20an%C3%A1lisis%20de%20datos%20y%20educaci%C3%B3n.](#)

-Rodledano, A. (2019). “Qué es Python: Características, evolución y futuro”. Recuperado el 4 de febrero del 2023 de: <https://openwebinars.net/blog/que-es-python/>

-Unir. (2019). “¿Conoces el lenguaje R? Te contamos los detalles de R y por qué se ha extendido tanto en las áreas de análisis de datos y big data.”. Recuperado el 4 de febrero del 2023 de: <https://www.unir.net/ingenieria/revista/lenguaje-r-big-data/>

1. Investiga y describe qué son y para qué se utilizan cada una de las siguientes tecnologías:

A. SQL: SQL (Structured Query Language) es un lenguaje de programación que sirve para actualizar, obtener, y calcular información en bases de datos relacionales.

B. Python: Es un lenguaje de programación de código abierto y legible que es ideal para trabajar con grandes volúmenes de datos.

C. R: Es un lenguaje de programación interpretado, o sea, que ejecuta las instrucciones directamente sin una previa compilación del programa a instrucciones máquina. Este software es usado comúnmente para la computación estadística y gráfica. Sirve tanto para Windows, como MacOS y plataformas Linux.

D. Anaconda: Es una distribución de los lenguajes de programación Python y R para las aplicaciones de Machine Learning, análisis predictivo, ciencia de datos, procesamiento de datos a gran escala,

etc.; cuya ventaja es simplificar la gestión e implementación de paquetes.

- E. Google Colaboratory: Es un producto de Google Research que permite a cualquier usuario escribir y ejecutar código arbitrario de Python en el navegador. Muy adecuado para análisis de datos y educación, y tareas de aprendizaje automático.

- F. Pandas: Pandas es una librería de datos creada por Wes McKinney en el 2008. El nombre “Pandas” es una función de “Pan” (Paneles) y “Da” (Datos); Paneles de datos. Estos paneles son estructuras de datos organizados por categorías en las que se pueden encontrar variables de tipo texto, booleano, y numérico. Es útil para reducir líneas de código, análisis de matrices de datos en 2D, y posee una API fácil y concisa.

- G. Hadoop: Es una estructura de software de código abierto que sirve para almacenar datos y ejecutar aplicaciones en clústeres de hardware comercial (grupo de ordenadores independientes que se conectan entre sí para trabajar juntos como un único sistema). Este software también proporciona almacenamiento masivo para cualquier tipo de datos.

H. Apache Spark: Es un framework de código abierto que admite el procesamiento en memoria para mejorar el rendimiento de las aplicaciones que analizan macrodatos. Sirve principalmente para el análisis de grandes conjuntos de datos (en terabytes o petabytes).

2. Contesta las siguientes preguntas:

A. ¿En qué casos es más recomendable utilizar Anaconda y en qué casos es más recomendable usar Google Colaboratory?

Google Colab sirve para trabajar con proyectos pequeños o medianos con una gran facilidad de compartir el contenido con los demás. Sin embargo, para proyectos más serios probablemente se necesite de un ambiente más estable con recursos garantizados para trabajar. Por lo que Anaconda, a pesar de ser un software pesado, sería una buena opción para hacer proyectos serios con potencial de desarrollo.

B. ¿En qué casos preferirías utilizar Python en vez de R y viceversa?

Python es un lenguaje de programación multipropósito, de fácil legibilidad de código, y para personas no tan familiarizadas con el ámbito de las matemáticas. En cambio, el lenguaje R está orientado más hacia este enfoque; es por eso que es más usado por profesionales matemáticos e informáticos. Por ejemplo: Si se desea realizar un análisis estadístico complejo, R es más adecuado que Python. Mientras que para implementaciones de código más generales (o que no involucren tantas matemáticas), Python es la opción más recomendable por la sencillez de desarrollo.

Al final, todo depende del contexto en el que la persona se encuentre.

C. ¿Qué otros lenguajes de programación son normalmente utilizados para la ciencia de datos y el Big Data?

1. Scala
2. Julia
3. Java
4. C++
5. MATLAB
6. JavaScript