

Primer Parcial - Parte Teórica

¿Qué es Python y por qué es popular en el análisis de datos?

Python es un lenguaje de programación de alto nivel que se caracteriza por su sintaxis sencilla y legible.

La popularidad de Python en el análisis de datos se debe a varias razones:

- ☐ **Facilidad de uso:** Su naturaleza dinámica y su capacidad para escribir código conciso lo convierten en una opción atractiva para tareas de análisis de datos.
- ☐ **Amplia variedad de bibliotecas:** Python cuenta con una gran cantidad de bibliotecas especializadas en análisis de datos, como NumPy, pandas, matplotlib y scikit-learn, entre otras. Estas bibliotecas ofrecen herramientas poderosas para manipular datos, realizar visualizaciones, realizar estadísticas y realizar tareas de aprendizaje automático.
- ☐ **Comunidad activa:** Esto significa que hay una abundancia de recursos, tutoriales y documentación disponibles en línea para ayudar a los usuarios a aprender y resolver problemas.
- ☐ **Integración con otras tecnologías:** Python se integra fácilmente con otras tecnologías y herramientas utilizadas en el análisis de datos, como bases de datos, herramientas de visualización y entornos de desarrollo integrado (IDE). Esto facilita la creación de flujos de trabajo completos para el análisis de datos.

¿Para qué se usa la librería de Numpy? ¿Y qué ventaja nos ofrece?

NumPy se utiliza para trabajar con vectores, matrices y arreglos n-dimensionales de manera eficiente.

La ventaja principal que presenta Numpy es su orden, ya que, a diferencia de las listas en Python (que crecen de manera dinámica), este tipo de matrices ya tienen un tamaño fijo en su creación. Además, las matrices de Numpy son mucho mejores que las listas de Python, ya que se almacenan en un lugar continuo en la memoria, por tanto, durante las transacciones es mucho más rápido acceder a ellas y manipularlas. Numpy permite crear operaciones matemáticas de alta complejidad, ecuaciones lineales de álgebra lineal o estructuras de datos, entre otras operaciones.

¿Para qué nos sirve Pandas? ¿Y qué tipos de objetos maneja?

Pandas nos sirve para el análisis de datos y la manipulación de estructuras de datos tabulares con Python.

Los objetos que maneja Pandas son los siguientes:

- ☐ **DataFrame** es una estructura bidimensional similar a una tabla.
- ☐ **Serie** es una estructura unidimensional similar a un array.

¿Qué es Matplotlib y cual es su uso?. Mencione un escenario real en donde lo usaría.

Matplotlib es una librería de visualización de datos con Python y su uso radica en obtener gráficos que representen a la información con la que se está trabajando.

Casos reales de uso:

La cantidad de casos en los cuales se puede usar esta librería es muy extensa, ya que siempre que se esté trabajando con Python y se desee mostrar gráficos es muy conveniente su uso, aún así mencionare algunos ejemplos:

- ☐ Comparación mes a mes de ganancias totales de una empresa en un determinado año (Gráfico de barras).
- ☐ Rangos de edades que solicitan en mayor medida la Beca progresar (Histograma).
- ☐ Productos que mayor porcentaje de venta abarcan en una empresa respecto al resto de productos vendidos (Gráfico de tortas).
- ☐ Tendencia de goles marcados por un jugador durante el año (Polígono de frecuencias).