**ANALISIS DE DATOS CON PYHTON (PARTE 1)**

El proceso de ciencia de datos es un enfoque sistemático para resolver problemas utilizando datos. Implica los siguientes pasos:

* Comprenda el problema: asegúrese de que el trabajo con datos esté alineado con los objetivos comerciales.
* Recopilar datos: recopilar datos relevantes
* Limpiar y preparar datos: identificar y corregir errores, inconsistencias e imprecisiones.
* Explorar datos: utilice visualizaciones y estadísticas de resumen para comprender la distribución y las relaciones de los datos.
* Construir modelos: crear modelos predictivos.
* Comunicar los resultados: compartir los hallazgos de una manera que sea fácil de entender para las partes interesadas.

**CONSEJOS PARA EL PROCESO DE DATOS:**

* **Utilice el análisis exploratorio de datos (EDA):** el EDA es un paso fundamental para comprender los datos.
* **Iterar y perfeccionar los modelos:** seguir mejorando los modelos en función del rendimiento y la retroalimentación.
* **Utilice métodos estadísticos:** estos métodos son cruciales para sacar conclusiones de los datos.

**COMO ANALIZAR DATOS PASO A PASO**

1. **EXTRACCION:**
   1. SQL
   2. CSV
   3. EXCEL
   4. JSON
   5. APIs
   6. Web Scrapping
   7. XML
   8. NonSQL
   9. OTRAS
2. **DEPURACIÓN Y PREPROCESAMIENTO**
   1. Valores faltantes.
   2. Imputar valores.
   3. Arreglar tipo de datos
   4. Valores inválidos
   5. Valores repetidos
   6. Outliers
   7. Eliminación de datos no significativos
3. **ANÁLISIS EXPLORATORIO (EDA)**
   1. Importancia de características
   2. Gestión de categóricas
   3. Nuevas características
   4. Índices
   5. Subconjuntos de datos
   6. Análisis estadístico
   7. Visualización
   8. Correción y causalidad
   9. Hipótesis

**QUÉ HERRAMIENTAS UTILIZAR?**

**EXCEL:**

* Cálculos simples.
* Pocos datos.
* Funciones tradicionales.

**PYTHON:**

* Trabajos complejos.
* Grandes volúmenes
* Librerías específicas

**Elegimos Python por lo siguiente:**

* Fácil de usar.
* Open Source
* Rápido y eficiente
* Gran comunidad online de soporte y desarrollo de librerías específicas

**ENTORNOS:**

* GOOGLE COLAB
* SUITE ANACONDA
* JUPYTERLAB DESKTOP

1. GOOGLE COLAB
   1. Google tiene una plataforma para trabajar con Python sin necesidad de descargar ningún software.
   2. Se trata de la plataforma Google Colab, en la cual podemos escribir y ejecutar celdas de código y celdas de texto enriquecido en un documento llamado “Notebook”.
   3. Celdas de código y texto enriquecido

Las celdas de código son espacios en los que se escribe código ejecutable. Se pueden encontrar en diferentes programas, como Databricks, Google Colab y ArcGIS Enterprise.

El texto enriquecido es un formato de texto que permite aplicar opciones de formato, como negrita, cursiva, subrayado, alineación, sangría, viñetas, tablas, imágenes, hipervínculos y fuentes.

Se utiliza para crear documentos visualmente atractivos como informes, boletines, folletos o currículums.

Los archivos de texto enriquecido tienen extensiones como .rtf, .docx o .pdf y requieren software o aplicaciones específicas para abrirlos y editarlos.