**TAREA 46: DICCIONARIO**

* **ESTADISTICA DESCRIPTIVA:** La estadística descriptiva es la técnica matemática que obtiene, organiza, presenta y describe un conjunto de datos con el propósito de facilitar el uso, generalmente con el apoyo de tablas, medidas numéricas o graficas.

Los principales parámetros estadísticos , sabiendo que un parámetro es un numero que se obtiene gracias a una distribución de datos estadísticos y ayuda a organizar la información dada ya sea por una gráfica o una tabla, los principales son:

* Centralización
* Posición
* Dispersión
* **ANALISIS EXPLORATORIO:** El análisis exploratorio de datos es una forma de analizar datos definido por John W. Tukey (E.D.A: Exploratory Data Analysis) es el tratamiento estadístico al que se someten las muestras recogidas durante un proceso de investigación en cualquier campo científico.

El análisis exploratorio tiene como objetivo identificar el modelo teórico mas adecuado para representar la población de la cual proceden los datos muestrales. Dicho análisis se basa en gráficos y estadísticos que permiten explorar la distribución identificando características tales como: valores atípicos o outliers, saltos o discontinuidades, concentraciones de valores, forma de la distribución, etc; Por otra parte, este análisis se puede realizar sobre todos los casos conjuntamente o de forma separada por grupos. En este ultimo caso los gráficos y estadísticos permiten identificar si los datos proceden de una o varias poblaciones, considerando la variable que determina los grupos como factor diferenciador de las poblaciones. También permite comprobar, mediante técnicas gráficas y contrastes no paramétricos, si los datos han sido extraídos de una población con distribución aproximadamente normal.

* **DATOS ESTRUCTURADOS:** Datos organizados perfectamente en filas y columnas bien definidas. Son los que usan de manera habitual en la mayor parte de bases de datos relacionales y las relaciones entre filas y columnas está claramente determinada.
* **DATOS SEMIESTRUCTURADOS:** Son aquellos con un nivel medio de estructuración y rigidez organizativa. Se encuentran a medio camino entre los estructurados y los no estructurados. En ejemplo es un servidor local que almacenará todos los datos de correo electrónico y archivos adjuntos dentro de la base de datos.

Tiene cierto nivel de estructura, jerarquía y organización, aunque carecen de un esquema fijo. En vez de estructuras esquemáticas, tienen más bien estructura de árbol, además de tener propiedades organizativas que facilitan su análisis.

Procesando los datos, se puede lograr organizar como una base de datos estructural, pero no todos los que se colocan en un grupo tienen siempre las mismas propiedades. A veces difieren en tipo y tamaño.

Además contienen metadatos (etiquetas y elementos) que se utilizan para agruparlos y describir como se almacenan, aunque no tantos como los datos estructurados, con lo que su gestión y automatización se ve dificultada.

Entre los datos de este tipo, además de los correos están:

* Archivos comprimidos
* XML y otros lenguajes de marcado
* Ejecutables binarios
* Paquetes TCP/IP
* **DATOS NO ESTRUCTURADOS:** Suponen el 80% de los datos existentes en cualquier organización y su manejo es más dificultoso y no se pueden usar en una base de datos tradicional, ya que no se pueden caracterizar en filas y columnas.

Existen aplicaciones que pueden procesar más de 1000 tipos de formatos de datos no estructurados. Los ejemplos de datos no estructurados más usados son:

* Documentos de oficina en archivos de texto
* Archivos de imágenes
* Archivos PDF
* Archivos de registro y de datos de aplicaciones como .ini o .dll
* Datos de redes sociales o de plataformas como YouTube
* Datos de ubicaciones y mensajería instantánea
* Grabaciones telefónicas, archivos de audio como MP3

Estos tipos de datos tienen estructura interna, pero no están estructurados a través de modelos o esquemas de datos fijos y predefinidos.