**TAREA 46: DICCIONARIO**

* **HERRAMIENTAS DE BUSINESS INTELLIGENCE:** Se denomina business intelligence al conjunto de estrategias, aplicaciones, datos, productos, tecnologías y arquitectura técnicas, los cuales están enfocados a la administración y creación de conocimiento sobre el medio, a través del análisis de los datos existentes en una organización o empresa.

Las herramientas de inteligencia empresarial se basas en la utilización de un sistema de información de inteligencia que se forma con distintos datos extraídos de la producción, con información relacionada con la empresa o sus ámbitos, y con datos económicos.

Mediante las herramientas y técnicas ETL (Extract, transform, load), se extraen los datos de distintas fuentes, se depuran y preparan (homogeneización de los datos), para luego cargarlos en un almacén de datos.

Además, las herramientas de inteligencia analítica posibilitan el modelado de las representaciones basadas en consultas para crear un cuadro de mando integral, que sirve de base para la presentación de informes.

* **CALIDAD DEL DATO:** Se refiere a los procesos, técnicas, algoritmos y operaciones encaminados a mejorar la calidad de los datos existentes en empresas y organismos. Hay varias definiciones de calidad en datos pero lo más común es que sirven su propósito. La limpieza de datos tal vez se requiere para asegurar la calidad de dichos datos.

Los principales beneficios son:

* Ahorrar costes directos: evitando tener información duplicada.
* Potenciar las acciones de marketing y la gestión: la normalización de archivos mejora el análisis de datos y permite segmentaciones precisas para que sus acciones de mercadotecnia y su gestión ganen en precisión y eficacia.
* Optimizar la captación y la fidelización de clientes
* Mejorar la imagen corporativa
* Mejorar el servicio
* **ANALISIS EXPLORATORIO:** Forma de analizar datos definido por John W. Turkey (EDA: Exploratory data análisis) y es el tratamiento estadístico al que someten las muestras recogidas durante un proceso de investigación en cualquier campo científico. Para mayor rapidez y precisión, todo el proceso suele realizarse por medios informáticos, con aplicaciones especificas para el tratamiento estadístico. Los EDA, no necesariamente, se llevan a cabo con una base de datos al uso, no con una hoja de calculo convencional; no obstante el programa SPSS y R (lenguaje de programación) son las aplicaciones mas utilizadas, aunque no las únicas.

Los pasos seguidos en el EDA, son básicamente dos:

* **Medición y descripción** de los datos tecnológicos – tipológicos- y dimensiones, por medio de la estadística descriptiva. Aquí tenemos, por un lado, las medidas de tendencia central (promedios que, en una sola cifra, resumen todos los valores de una muestra: media, mediana y moda son las más habituales) y, por otro, las medidas de dispersión (que calculan hasta que punto la muestra se agrupa o no en torno a esos promedios). Dentro de este apartado, se ha de procurar, además, calibrar la confianza de las muestras a través de tres estadímetros básicos: la desviación estándar de la muestra, la curtosis y la asimetría.
* **Comparación** de los caracteres de una muestra, o de varias muestras diferentes por medio de la Estadística inferencial. Las pruebas mas frecuentemente utilizadas comienzan por las más sencillas comparaciones visuales – a través de graficas como la campana de Gauss, nubes de dispersión o diagramas de caja y arbotantes-, pasando por las socorridas tablas de contingencia (incluido la prueba del X2), y por los típicos análisis de varianza (que no es más que una confrontación muy precisa de los promedios de varias muestras), hasta llegar a los más complejos análisis multivariantes de conglomerados.

Los cálculos estadísticos orientan sobre la fiabilidad de las muestras usadas, aunque no son infalibles, e indican si los resultados obtenidos al calcular las pruebas inferenciales son aceptables, es lo que llamamos nivel de confianza (se debe procurar que este nunca sea inferior al 95% = 0,95).

* **ESTADISTICA DESCRIPTIVA:** La estadística descriptiva es la técnica matemática que obtiene, organiza, presenta y describe un conjunto de datos con el propósito de facilitas el uso, generalmente con el apoyo de tablas, medidas numéricas o graficas. Estas técnicas son utilizadas en el proceso de investigación, en la etapa donde el investigador necesita procesar y analizar los datos recolectados en dicho estudio.
* **HISTOGRAMA:** Representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados. Sirven para obtener una “primera vista” general, o panorama, de la distribución de la población, o de la muestra, respecto a una característica, cuantitativa y continua. De esta manera ofrece una visión de grupo permitiendo observar una preferencia, o tendencia, por parte de la muestra o población por ubicarse hacia una determinada región de valores dentro del espectro de valores posibles (sean infinitos o no) que pueda adquirir la característica. Así podemos evidenciar comportamientos, observar el grado de homogeneidad, acuerdo o concisión entre los valores de todas las partes que componen la población o la muestra, o , en contraposición, poder observar el grado de variabilidad, y por ende, la dispersión de todos los valores que toman las partes, también es posible no evidenciar ninguna tendencia y obtener que cada miembro de la población toma por su lado y adquiere un valor de la característica aleatoriamente sin mostrar ninguna preferencia o tendencia, entre otras cosas.

Se utilizan para relacionar variables cuantitativas continuas. Para variables cuantitativas discretas las barras se dibujan separadas y el grafico se llama diagrama de frecuencias, porque la variable representada en el eje horizontal ya no representa un espectro continuo de valores, sino valores cuantitativos específicos, igual que ocurre en un diagrama de barras, usado para representar una característica cualitativa o categórica. Su utilidad se hace mas evidente cuando se cuenta con un gran numero de datos cuantitativos y que se han agrupado en intervalos de clase. Ejemplos de su uso es la representación de edades o estaturas de una población.

Tipos relacionados:

* **Histograma de frecuencias absolutas:** Representa la frecuencia absoluta mediante la altura de barras y solo cuando los intervalos son iguales.
* **Histograma de frecuencias relativas:** Representa la frecuencia relativa mediante la altura de las barras y solo cuando los intervalos son iguales, ya que en ese caso las alturas y las superficies son proporcionales.
* **Histograma:** Representa la frecuencia relativa mediante la superficie de barras. Aunque esto sea cierto en todos los histogramas, cuando se agrupan los datos en intervalos desiguales hay que atender a la superficie de las barras, que no se corresponderá con la altura como ocurriría en los casos anteriores. Para su elaboración debe introducirse el concepto de altura de histograma, que es un concepto equivalente al de la densidad de probabilidad, y que se calcula dividiendo la frecuencia relativa de ese intervalo (o sea la superficie que queremos darle) entre la anchura del intervalo (la base del rectángulo). Ahora las barras tendrán siempre superficie igual a la frecuencia relativa y la suma de todas esas superficies (de todas las barras) será 1, o sea el 100%.
* **Función densidad:** Representa la probabilidad mediante la superficie de barras. Es un gráfico idéntico al histograma pero aplicado a distribuciones teóricas. El concepto de frecuencia relativa se cambia por el de probabilidad, pero también se representa por superficies y la suma de todas esas superficies (de todas las barras) será 1, como en el histograma, o sea el 100% de probabilidad.
* **Curva acumulativa u ojiva:** Es un gráfico acumulativo (Función de distribución) que representa la frecuencia relativa acumulada hasta cada valor de la variable. Si el rango es finito el primer valor del rango tiene frecuencia acumulada (anterior) cero y el ultimo tiene frecuencia acumulada 1 (100%). Así el eje vertical siempre toma valores de cero a 1 y representa frecuencias relativas (o probabilidades si se trata de distribuciones teóricas). Se utiliza para introducir el concepto y el calculo de la mediana, los cuartiles, los deciles y en general los parámetros llamados de posición. Si el rango es infinito, como suele ocurrir en las distribuciones teóricas (Normal, student, chi-cuadrado, etc.) el cero puede no alcanzarse y será el valor asintótico por la izquierda, si tampoco se alcanza el uno también será el valor asintótico derecho, y en muchos casos no se alcanza ni uno ni otro, teniendo dos asíntotas. Este gráfico es la integral del histograma (cuando trabajamos con distribuciones reales) o de la función densidad (cuando trabajamos con distribuciones teóricas).
* **Curva acumulativa de frecuencias absolutas:** Es un grafico acumulativo que representa la frecuencia absoluta acumulada hasta cada valor de la variable. Realmente no es un grafico relacionado con el histograma, pero es muy parecido a la curva acumulativa y a la función de distribución. Esta curva no ira entre cero y uno sino entre cero y el total de los individuos de la muestra.
* **GRAFICO DE BARRAS:** Un gráfico de barras o grafico de columnas, es una forma de representar gráficamente un conjunto de datos o valores mediante barras rectangulares de longitud proporcional a los valores representados. Los gráficos de barras pueden ser usados para comparar cantidades de una variable en diferentes momentos o diferentes variables para el mismo momento. Las barras pueden orientarse horizontal o verticalmente.

El grafico de barras es una forma de resumir un conjunto de datos por categorías de variable cualitativa y su frecuencia de aparición en una muestra.

* **GRAFICO CIRCULAR:** Recurso estadístico que se utiliza para representar porcentajes y proporciones. El numero de elementos comparados dentro de una grafica circular suele ser de mas de cuatro.
* **GRAFICO DE DISPERSION:** O diagrama de dispersión, es un tipo de diagrama matemático que utiliza las coordenadas cartesianas para mostrar los valores de dos variables para un conjunto de datos. Se utiliza cuando una o varias variables está bajo el control del experimentador. Si existe un parámetro que se incrementa o disminuye de forma sistemática por el experimentador, se le denomina parámetro de control o variable independiente y habitualmente se representa a lo largo del eje horizontal. Si no existe variable dependiente, cualquier variable se puede representar en cada eje y el diagrama de dispersión mostrará el grado de correlación (no causalidad) entre las dos variables.
* **GRAFICO DE CAJA Y BIGOTES:**  O Box plot es un método estandarizado para representar gráficamente una serie de datos numéricos a través de cuartiles. De esta manera, el diagrama de caja muestra a simple vista la mediana y los cuartiles de los datos, pudiendo representar los valores atípicos de estos. Conviene recordar que se utilizan las bisagras de Turkey, y no los cuartiles a la hora de dibujar la caja del grafico, aunque los resultados son semejantes en muestras grandes.

Sus componentes son:

* Rango (sin datos atípicos)
* Datos atípicos
* Rango intercuartil (también conocido como RIC)
* Cuartiles (Denotados como Q1, Q2 y Q3)
* Mediana (Q2)
* Mínimo y máximo
* **CARTOGRAMA:** Mapa o diagrama que muestra datos de cantidad asociados a sus respectivas áreas, mediante la modificación de los tamaños de las unidades de enumeración. La información es aportada mediante la distorsión de las superficies reales, utilizando cada superficie de enumeración como un símbolo proporcional, el cual aumenta o disminuye en función de los valores correspondientes.