A

**Agenda:** Lista de citas programadas

**Alcance del trabajo (SOW):** Esquema acordado de las tareas a realizar durante un proyecto

**Algoritmo:** Proceso o conjunto de reglas a seguir para realizar una tarea específica

**Análisis computacional de datos:** La ciencia de los datos

**Análisis de datos:** Recopilación, transformación y organización de los datos para sacar conclusiones, hacer predicciones e impulsar una toma de decisiones fundamentada

**Análisis de déficits:** Método para examinar y evaluar el estado actual de un proceso con el fin de identificar las oportunidades de mejora en el futuro

**Analista de datos:** Persona que recopila, transforma y organiza los datos para sacar conclusiones, hacer predicciones e impulsar la toma de decisiones fundamentada

**Anonimización de datos:** Proceso de protección de los datos privados o confidenciales de las personas mediante la eliminación de información que pueda asociarse a ellas

**Apertura:** Aspecto de la ética de los datos que promueve el acceso libre, el uso y el uso compartido de los datos

**Archivo CSV (valores separados por coma):** Archivo de texto delimitado que utiliza una coma para separar valores

**Archivo de audio:** Almacenamiento en audio digitalizado generalmente en MP3, AAC u otro formato comprimido

**Archivo de video:** Conjunto de imágenes, archivos de audio y otros datos generalmente codificados en un formato comprimido como por ejemplo MP4, MV4, MOV, AVI o FLV

**Atributo**: Característica o cualidad de los datos que se usa para etiquetar una columna en una tabla

**AVERAGE:** Función de una hoja de cálculo que muestra el resultado de un promedio de los valores de un rango seleccionado

B

**Base de datos normalizada:** Base de datos en la que solo se almacenan datos relacionados en cada tabla

**Base de datos relacional:** Base de datos que contiene una serie de tablas que se pueden conectar para formar relaciones

**Base de datos:** Recopilación de datos almacenados en un sistema informático

**Bordes**: Líneas que se pueden agregar alrededor de dos o más celdas en una hoja de cálculo

C

**Cadena de texto:** Grupo de caracteres en una celda, mayormente compuesto por letras

**Campo:** Información de una fila o columna de una hoja de cálculo; en una tabla de datos, suele ser una columna de la tabla

**Causa raíz:** Razón por la que ocurre un problema

**Ciclo de vida de los datos:** Secuencia de etapas por las que pasan los datos, que incluye planificar, capturar, gestionar, analizar, archivar y destruir

**Ciencia de datos:** Campo de estudio que utiliza datos sin procesar para crear nuevas formas de modelar y entender lo desconocido

**Clave externa:** Campo en una tabla de una base de datos que es una clave primaria en otra tabla (Ver clave primaria)

**Clave primaria:** Identificador en una base de datos que hace referencia a una columna en la que cada valor es único (Ver clave externa)

**Coherencia:** Grado de repetibilidad de los datos desde diferentes puntos de entrada o recopilación

**Compatibilidad:** Qué tan bien dos o más conjuntos de datos pueden trabajar juntos

**CONCATENATE:** Función de una hoja de cálculo que une dos o más cadenas de texto

**Conjunto de datos:** Recopilación de datos que pueden ser manipulados o analizados como una unidad

**Consentimiento:** Aspecto de la ética de datos que presupone el derecho de una persona a conocer cómo y por qué se utilizarán sus datos personales antes de aceptar proporcionarlos

**Consulta:** Solicitud de datos o información de una base de datos

**Contexto:** Condición en la que algo existe o sucede

**Control de acceso:** Funciones como la protección de contraseñas, permisos de usuario y cifrado que se usan para proteger una hoja de cálculo

**Controlador de relleno:** Cuadro en la esquina inferior derecha de una celda seleccionada de una hoja de cálculo que se puede arrastrar a través de las celdas vecinas para seguir una instrucción

**Convenciones de nomenclatura:** Pautas uniformes para el nombre de un archivo que describen el contenido, la fecha de creación y la versión

**Cookie:** Pequeño archivo almacenado en una computadora que contiene información acerca de sus usuarios

**COUNT**: Función de la hoja de cálculo que cuenta el número de celdas en un rango que cumplen con un criterio especificado

**COUNTIF:** Función de la hoja de cálculo que muestra el número de celdas que coinciden con un valor especificado

D

**DATEDIF:** Función de una hoja de cálculo que calcula el número de días, meses o años entre dos fechas

**Datos abiertos:** Datos que están disponibles para el público

**Datos booleanos:** Tipo de datos con solo dos valores posibles, generalmente verdadero o falso

**Datos continuos:** Datos que se miden y que pueden tener casi cualquier valor numérico

**Datos cualitativos:** Medida subjetiva y explicativa de una cualidad o característica

**Datos cuantitativos:** Medida específica y objetiva, como un número, cantidad o rango

**Datos de primera fuente:** Datos recopilados por una persona o por un grupo por medio de sus propios recursos

**Datos de segunda fuente:** Datos recopilados por un grupo directamente de su audiencia y que luego se venden

**Datos de terceros:** Datos proporcionados por fuentes externas que no recopilaron de forma directa

**Datos desactualizados:** Cualquier dato que haya sido reemplazado por información más nueva y más precisa

**Datos discretos:** Datos que se cuentan y tienen un número limitado de valores

**Datos duplicados:** Cualquier registro que inadvertidamente comparte datos con otro registro

**Datos en formato ancho:** Conjunto de datos en el que cada tema tiene una sola fila con varias columnas para retener los valores de los distintos atributos de dicho tema

**Datos en formato largo:** Conjunto de datos en el que cada fila es un punto de tiempo por tema; por lo tanto, cada tema tiene datos en varias filas

**Datos estructurados:** Datos organizados en un cierto formato, por ejemplo, filas y columnas

**Datos externos:** Datos que se alojan y generan fuera de una organización

**Datos incoherentes:** Datos que usan diferentes formatos para representar lo mismo

**Datos incompletos:** Datos que carecen de campos importantes

**Datos incorrectos/inexactos:** Datos que son completos pero inexactos

**Datos internos:** Datos alojados en los sistemas propios de una empresa

**Datos limpios:** Datos que están completos, correctos y que son pertinentes para el problema que se está resolviendo

**Datos no estructurados:** Datos que no se organizan de manera fácilmente identificable

**Datos nominales:** Tipo de datos cualitativos que se categorizan sin un orden establecido

**Datos ordinales:** Datos cualitativos con un orden o escala establecidos

**Datos sucios:** Datos que están incompletos o son incorrectos o irrelevantes para el problema a resolver

**Datos:** Recopilación de hechos

**Delimitador:** Carácter que indica el principio o el fin de un elemento de datos

**Destrezas analíticas:** Cualidades y características asociadas al uso de hechos para resolver problemas

**Diseño de datos:** Cómo se organiza la información

**Dominio del problema:** Área de análisis que abarca cada actividad que afecta a un problema o se ve afectada por él

E

**Ecosistema de datos:** Los distintos elementos que interactúan entre sí para producir, gestionar, almacenar, organizar, analizar y compartir datos

**Ecuación:** Cálculo que implica suma, resta, multiplicación o división (también se denomina “expresión matemática”)

**Elemento de datos:** Información en un conjunto de datos

**Encabezado:** Primera fila en una hoja de cálculo que hace referencia al tipo de datos en cada columna

**Equidad:** Cualidad del análisis de datos que no genera sesgos ni los reafirma

**Especialista en almacenamiento de datos:** Profesional que desarrolla procesos y procedimientos para almacenar y organizar datos efectivamente

**Esquema:** Forma de describir cómo se organiza algo, por ejemplo, los datos

**Estrategia de datos:** Gestión de las personas, los procesos y las herramientas que se usan en el análisis de datos

**Ética de los datos:** Normas justificadas respecto de lo que está bien y lo que está mal a la hora de recopilar, compartir y usar datos

**Ética:** Normas justificadas respecto de lo que está bien y lo que está mal. Por lo general, presuponen lo que deben hacer los seres humanos, usualmente en términos de derechos, obligaciones, beneficios para la sociedad, equidad o virtudes específicas

**Exactitud:** Grado de conformidad de los datos con respecto a la entidad real que se mide o describe

**Exhaustividad:** Grado en que los datos contienen todas las medidas o componentes deseados

**Expresión matemática:** Cálculo que implica suma, resta, multiplicación o división (también se denomina ecuación)

**Expresión regular (RegEx):** Regla que establece que los valores de una tabla deben coincidir con un patrón prescrito

F

**Filtrado:** Proceso que muestra solo los datos que cumplen con un criterio específico mientras oculta el resto

**Formato condicional:** Herramienta de una hoja de cálculo que cambia la forma en que aparecen las celdas cuando los valores cumplen con una condición específica

**Fórmula:** Conjunto de instrucciones que se utilizan para realizar un cálculo utilizando los datos de una hoja de cálculo

**Foto digital:** Imagen electrónica o computarizada generalmente en formato BMP o JPG

**FROM:** Sección de una consulta que indica de qué tabla(s) extraer los datos

**Fuente de datos correctos:** Fuente de datos que es confiable, original, integral, actual y citada (ROCCC)

**Fuente de datos erróneos:** Fuente de datos que no es confiable, original, integral, actual ni citada (ROCCC)

**Función matemática**: Función que se utiliza como parte de una fórmula matemática

**Función:** Comando preestablecido que realiza automáticamente un proceso o tarea especificado al usar los datos de una hoja de cálculo

**Fusión de datos:** Proceso de combinar dos o más conjuntos de datos en un único conjunto de datos

**Fusión:** Acuerdo que une dos organizaciones en una organización nueva

G

**Geolocalización:** Ubicación geográfica de una persona o dispositivo mediante información digital

**Gobierno de datos:** Proceso para garantizar la gestión formal de los recursos de datos de una empresa

**Gráfico dinámico:** Gráfico creado a partir de los campos en una tabla dinámica

H

**Hacer contactos:** Construir relaciones con otros tanto en persona como en línea

**Hoja de cálculo:** Hoja de cálculo digital

I

**Informe:** Recopilación estática de datos que se entrega periódicamente a los interesados

**Ingeniero de datos:** Profesional que transforma los datos en un formato útil para su análisis y les da una estructura confiable

**Ingresos:** Cantidad total de ingresos generados por la venta de mercaderías o servicios

**Integridad de datos:** Exactitud, exhaustividad, coherencia y confiabilidad de los datos a lo largo de su ciclo de vida

**Interesados:** Personas que invierten tiempo y recursos en un proyecto y se interesan por su resultado

**Interoperabilidad de los datos:** Capacidad de integrar datos de varias fuentes y un factor clave que conduzca al uso satisfactorio de los datos abiertos entre las empresas y los gobiernos

**Intervalo de confianza:** Rango de valores que transmite qué probabilidad hay de que una estimación estadística refleje la población

L

**LEFT:** Función que devuelve un número establecido de caracteres a la izquierda de una cadena de texto

**LEN:** Función que indica la longitud de una cadena de texto al contar el número de caracteres que contiene

**Lenguaje de consulta estructurado:** Lenguaje de programación informática usado para comunicarse con una base de datos

**Lenguaje de consulta:** Lenguaje de programación informática usado para comunicarse con una base de datos

**Limitaciones de datos:** Criterio que determina si un dato está limpio y es válido

**Longitud de campo:** Herramienta para determinar cuántos caracteres pueden incluirse en el campo de una hoja de cálculo

**Longitud:** Número de caracteres en una cadena de texto

M

**Macrodatos:** Conjuntos de datos grandes y complejos que, generalmente, se recopilan durante largos períodos y que permiten que los analistas de datos aborden los problemas comerciales de gran alcance

**Manipulación de datos:** Proceso para cambiar los datos, de manera que estén más organizados y sean más fáciles de leer

**Mapeo de datos:** Proceso de hacer coincidir campos entre una fuente de datos y otra

**Margen de error:** Cantidad máxima que se espera que los resultados de la muestra difieran de los de la población real

**MAX:** Función de la hoja de cálculo que muestra el valor numérico más alto de un rango de celdas

**Mentalidad técnica:** Capacidad de dividir las cosas en pasos o piezas más pequeñas y trabajar con ellas de forma ordenada y lógica

**Mentor:** Persona que comparte su conocimiento, sus habilidades y su experiencia para ayudar a otras personas a crecer, tanto en el campo profesional como el personal

**Metadato administrativo:** Metadato que indica el origen técnico de un recurso digital

**Metadatos descriptivos:** Metadatos que describen datos y que se pueden utilizar para identificarlos más adelante

**Metadatos estructurales:** Metadatos que indican cómo se organizan ciertos datos y si forman parte de una recopilación de datos o de varias

**Metadatos:** Datos sobre los datos

**Metodología SMART:** Herramienta para determinar la eficacia de una pregunta basándose en si es específica, medible, orientada a la acción, relevante y con plazos determinados

**Métrica:** Tipo único y cuantificable de datos que pueden utilizarse para medición

**Microdatos:** Puntos de datos pequeños, específicos, que generalmente implican un breve período y que son útiles para tomar decisiones diarias

**MID:** Función que extrae un segmento desde el medio de una cadena de texto

**MIN**: Función de la hoja de cálculo que muestra el valor numérico menor de un rango de celdas

**Modelo de datos:** Herramienta para organizar los elementos de los datos y la forma en que se relacionan entre ellos

**Muestra:** En el análisis computacional de datos, un segmento de una población representativo de toda la población

**Muestreo aleatorio:** Forma de seleccionar una muestra de una población para que todos los tipos posibles de la muestra tengan la misma oportunidad de ser elegidos

**Muestreo imparcial:** Cuando la muestra de la población que se está midiendo es representativa de la población como un todo

N

**Nivel de confianza:** Probabilidad de que el tamaño de una muestra refleje con precisión a la porción más grande de la población

**Notebook:** Entorno de programación interactivo y editable para generar informes de datos y mostrar destrezas en el uso de datos

**Nube:** Lugar para mantener los datos en línea, en vez de guardarlos en el disco duro de una computadora

**Nulo:** Indicación de que un valor no existe en un conjunto de datos

O

**Objetivo métrico:** Objetivo medible establecido por una empresa y evaluado mediante métricas

**Obligatorio:** Valor de datos que no puede quedar en blanco ni vacío

**Observación:** Atributos que describen los datos contenidos en la fila de una tabla

**Oficina del Censo de los Estados Unidos:** Agencia del Departamento de Comercio de los Estados Unidos que funciona como proveedor principal de la nación de datos de calidad sobre las personas y la economía

**Operador:** Símbolo que designa la operación o cálculo a realizarse

**Orden de las operaciones:** Uso de paréntesis para agrupar los valores de la hoja de cálculo a fin de aclarar el orden en el que deben realizarse las operaciones

**Ordenación:** Proceso de organizar los datos en un sistema de clasificación significativo para que sean más fáciles de entender, analizar y visualizar

**Organización Mundial de la Salud:** Organización cuya función principal es la de conducir y coordinar la salud internacional dentro del sistema de las Naciones Unidas

P

**Panel:** Herramienta que monitorea los datos entrantes en vivo

**Patrocinador:** Profesional que se compromete a hacer progresar la carrera profesional de otra persona

**Pensamiento analítico:** Proceso de identificar y definir un problema para luego resolverlo mediante el uso de datos de manera organizada, paso a paso

**Pensamiento estructurado:** Proceso de reconocer el problema o la situación actuales, organizar la información disponible, revelar déficits y oportunidades e identificar opciones

**Píxel:** En imágenes digitales, área pequeña de iluminación en una pantalla de visualización que, cuando se combina con otras áreas adyacentes, forma una imagen digital

**Población:** En el análisis computacional de datos, todos los valores posibles en un conjunto de datos

**Poder estadístico:** Probabilidad de que una prueba de importancia reconozca un efecto presente

**Pregunta con límite de tiempo:** Pregunta que especifica un plazo para ser analizada

**Pregunta específica:** Pregunta simple, significativa y enfocada en un solo tema o en algunas ideas estrechamente relacionadas entre sí

**Pregunta injusta:**  Pregunta en la que se hacen suposiciones o que es difícil de responder honestamente

**Pregunta medible:** Pregunta cuyas respuestas se pueden cuantificar y evaluar

**Pregunta orientada a la acción:** Pregunta cuyas respuestas conducen al cambio

**Pregunta principal:** Pregunta que orienta a las personas hacia cierta respuesta

**Pregunta relevante:** Pregunta que tiene importancia para el problema que se debe resolver

**Privacidad de los datos:** Preservación de la información sobre los datos de una persona cada vez que ocurre una transacción de datos

**Proceso de análisis de datos:** Las seis fases de preguntar, preparar, procesar, analizar, compartir y actuar con el propósito de obtener información que propicie la toma de decisiones informada

**Propiedad:** Aspecto de la ética de datos que presupone que cada persona es dueña de los datos sin procesar que proporciona y que tiene control primordial sobre su uso, procesamiento y uso compartido

**Pruebas A/B:** Proceso de probar dos variaciones de la misma página web para determinar qué página es más exitosa para atraer el tráfico de usuarios y generar ingresos

Q

**Quitar duplicados:** Herramienta de una hoja de cálculo que busca y elimina automáticamente las entradas duplicadas de una hoja de cálculo

R

**Rango de datos:** Valores numéricos que se encuentran entre valores máximos y mínimos predefinidos

**Rango:** Conjunto de dos o más celdas en una hoja de cálculo

**Redes sociales:** Sitios web y aplicaciones donde los usuarios crean y comparten contenido o interactúan entre sí

**Redundancia:** Cuando los mismos datos se almacenan en dos o más lugares

**Referencia de celda:** Celda o rango de celdas en una hoja de cálculo que se usa generalmente en fórmulas y funciones

**Reformulación:** Proceso de replantear un problema o desafío, que se redirecciona luego hacia una posible resolución

**Registro:** Conjunto de datos relacionados en una tabla de datos, generalmente sinónimo de fila

**Reglamento General de Protección de Datos de la Unión Europea (RGPD):** Organismo formulador de políticas en la Unión Europea, creado para ayudar a proteger a las personas y sus datos

**Replicación de datos:** Proceso de almacenamiento de datos en varias ubicaciones

**Repositorio de metadatos:** Base de datos creada para almacenar metadatos

**Retorno de la inversión (ROI):** Fórmula que utiliza las métricas de inversión y ganancias para evaluar el éxito de una inversión

**RIGHT:** Función que muestra un número establecido de caracteres a la derecha de una cadena de texto

S

**Seguridad de los datos:** Emplear medidas de seguridad para proteger los datos contra el acceso no autorizado o contra la corrupción

**SELECT:** Sección de una consulta que indica de qué columna(s) extraer los datos

**Sesgo de confirmación:** Tendencia de buscar o interpretar la información de manera que confirme creencias preexistentes

**Sesgo de interpretación:** Tendencia a interpretar situaciones ambiguas de manera positiva o negativa

**Sesgo de los datos:** Cuando una preferencia a favor o en contra de una persona, un grupo de personas o una cosa sesga sistemáticamente los resultados del análisis de datos en una cierta dirección

**Sesgo del investigador:** Tendencia de distintas personas a observar las cosas de forma diferente (Ver “Sesgo del observador”)

**Sesgo del muestreo:** Representar en mayor o en menor medida a ciertos miembros de una población debido a que se trabaja con una muestra que no representa a la población en su totalidad

**Sesgo del observador:** Tendencia de distintas personas a observar las cosas de forma diferente (también se denomina sesgo del investigador)

**Sesgo:** Preferencia consciente o subconsciente a favor o en contra de una persona, un grupo de personas o una cosa

**Significancia estadística:** Probabilidad de que los resultados de una muestra no se deban a una posibilidad aleatoria

**Sintaxis:** Estructura predeterminada de un lenguaje, que incluye todas las palabras, los símbolos y la puntuación requeridos, así como su correcta ubicación

**SPLIT:** Función que divide el texto en función de un carácter específico y ubica cada fragmento en una nueva celda separada

**SQL:** (Ver Lenguaje de consulta estructurado)

**Subcadena:** Subconjunto de una cadena de texto

**SUM:** Función de una hoja de cálculo que suma los valores de un rango de celdas seleccionadas

T

**Tabla dinámica:** Herramienta de resumen de datos que se utiliza para clasificar, reorganizar, agrupar, contar, totalizar o promediar datos

**Tarea empresarial:** Pregunta o problema que el análisis de datos resuelve para un negocio

**Tasa de respuesta estimada**: Número promedio de personas que suele completar una encuesta

**Tasa de rotación:** Ritmo en el que los empleados abandonan voluntariamente una empresa

**Testeo de hipótesis:** Proceso que se realiza para determinar si una encuesta o un experimento tiene resultados significativos

**Tipo de datos de cadena:** Secuencia de caracteres y puntuación que contiene información textual (también se denomina tipo de datos de texto)

**Tipo de datos de texto:** Secuencia de caracteres y puntuación que contiene información textual (también se denomina tipo de datos de cadena)

**Tipo de datos:** Atributo que describe cierto dato según sus valores, su lenguaje de programación o las operaciones que puede realizar

**Tipos de problemas:** Distintos problemas que encuentra el analista de datos; entre ellos, categorizar elementos, descubrir conexiones, hallar patrones, identificar temas, hacer predicciones y detectar algo inusual

**Toma de decisiones basada en datos:** Uso de datos para guiar la estrategia empresarial

**Toma de decisiones inspirada en datos:** Explorar diferentes fuentes de datos para descubrir qué tienen en común

**Transacciones:** Aspecto de la ética de los datos que considera que las personas deben conocer las transacciones financieras resultantes de sus datos personales y la dimensión de esas transacciones

**Transferencia de datos:** Proceso de copiar datos de un dispositivo de almacenamiento a la memoria de la computadora, o de una computadora a otra

**Transparencia de la transacción:** Aspecto de la ética de datos que presupone que se deben explicar todas las actividades de procesamiento de datos y los algoritmos a la persona que proporciona los datos y que también presupone que esta persona debe comprenderlos

**TRIM:** Función que quita los espacios al principio, al final o los repetidos en los datos

U

**Único:** Valor que no puede tener un duplicado

V

**Validación de campos cruzados:** Proceso que garantiza el cumplimiento de ciertas condiciones para múltiples campos de datos

**Validación de datos:** Herramienta para corroborar la exactitud y la calidad de los datos

**Validez:** Grado de conformidad de los datos con respecto a las restricciones cuando se los ingresa, recopila o crea

**Visualización de datos:** Representación gráfica de los datos

**Visualización:** (Ver Visualización de datos)

**VLOOKUP:** Función de una hoja de cálculo que busca verticalmente cierto valor en una columna y arroja la información correspondiente

W

**WHERE:** Sección de una consulta que especifica los criterios que deben cumplir los datos solicitados