



UTPL
La Universidad Católica de Loja

Modalidad Abierta y a Distancia



Itinerario 1: Medios Data Journalism

Guía didáctica

Facultad de Ciencias Sociales, Educación y Humanidades

Departamento de Ciencias de la Comunicación

Itinerario 1: Medios Data Journalism

Guía didáctica

Carrera	PAO Nivel
▪ Comunicación	VIII

Autora:

Coronel Salas Gabriela de las Mercedes



P E R I _ 5 0 5 8

Asesoría virtual
www.utpl.edu.ec

Universidad Técnica Particular de Loja

Itinerario 1:Medios Data Journalism

Guía didáctica

Coronel Salas Gabriela de las Mercedes

Diagramación y diseño digital:

Ediloja Cía. Ltda.

Telefax: 593-7-2611418.

San Cayetano Alto s/n.

www.ediloja.com.ec

edilojacialtda@ediloja.com.ec

Loja-Ecuador

ISBN digital -978-9942-39-390-6



Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)

Usted acepta y acuerda estar obligado por los términos y condiciones de esta Licencia, por lo que, si existe el incumplimiento de algunas de estas condiciones, no se autoriza el uso de ningún contenido.

Los contenidos de este trabajo están sujetos a una licencia internacional Creative Commons **Reconocimiento-NoComercial-CompartirIgual 4.0 (CC BY-NC-SA 4.0)**. Usted es libre de **Compartir – copiar y redistribuir el material en cualquier medio o formato**. **Adaptar – remezclar, transformar y construir a partir del material citando la fuente, bajo los siguientes términos: Reconocimiento - debe dar crédito de manera adecuada, brindar un enlace a la licencia, e indicar si se han realizado cambios.** Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante. **No Comercial-no puede hacer uso del material con propósitos comerciales. Compartir igual-Si remezcla, transforma o crea a partir del material, debe distribuir su contribución bajo la misma licencia del original.** No puede aplicar términos legales ni medidas tecnológicas que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia. <https://creativecommons.org/licenses/by-nc-sa/4.0/>

Índice

1. Datos de información.....	8
1.1. Presentación de la asignatura	8
1.2. Competencias genéricas de la UTPL.....	8
1.3. Competencias específicas de la carrera	9
1.4. Problemática que aborda la asignatura	9
2. Metodología de aprendizaje.....	10
3. Orientaciones didácticas por resultados de aprendizaje.....	11
Primer bimestre	11
Resultado de aprendizaje 1.....	11
Contenidos, recursos y actividades de aprendizaje recomendada.....	11
Semana 1	11
 Unidad 1. Introducción a Data Journalism (DJ)	11
1.1. Antecedentes	11
1.2. Inicio del trabajo con datos.....	14
Actividades de aprendizaje recomendadas:	15
Autoevaluación 1	17
 Semana 2	19
 Unidad 2. ¿Qué es Data Journalism o Periodismo de Datos?.....	19
2.1. Definición	19
2.2. Fórmula	21
2.3. Cómo ser un periodista de datos	22
2.4. Trabajo con datos. Caso: The Guardian	24
Actividades de aprendizaje recomendadas	26
Autoevaluación 2	27
 Semana 3	30
 Unidad 3. Trabajo con datos masivos	30
3.1. ¿Por qué los periodistas deben trabajar con datos masivos?	30
3.2. ¿Por qué los datos son el rey o reina del contenido?	34
Actividades de aprendizaje recomendadas	36

Autoevaluación 3	37
Resultado de aprendizaje 2.....	40
Contenidos, recursos y actividades de aprendizaje recomendadas	40
Semana 4	40
Unidad 4. Productos basados en datos	40
4.1. Artículos	41
4.2. Visualizaciones	42
4.3. Datasets	44
4.4. News Apps	48
Actividades de aprendizaje recomendadas	52
Autoevaluación 4.....	53
Semana 5	55
Unidad 5. Transparencia y acceso a la información.....	55
5.1. Agenda de Desarrollo Sostenible.....	56
5.2. Protección de datos personales	58
5.3. Ecuador: acceso a la información	61
5.4. Datos sensibles.....	66
Actividades de aprendizaje recomendadas	68
Autoevaluación 5	69
Semana 6	72
Unidad 6. Open Data	72
6.1. ¿Qué es Open - Abierto?.....	72
Autoevaluación 6	82
Semana 7	84
Unidad 7. Big data	84
7.1. El poder de los datos	85
7.2. . Las Uves (V) de Big data	88
7.3. Fuentes de big data	91
Actividades de aprendizaje recomendadas	96
Autoevaluación 7	97
Actividades finales del bimestre.....	99

Semana 8	99
Segundo bimestre	100
Resultado de aprendizaje 3.....	100
Contenidos, recursos y actividades de aprendizaje recomendadas	100
Semana 9	100
Unidad 8. Data mining o minería de datos	100
8.1. ¿Qué es data mining?	100
8.2. ¿Quién mina?.....	102
8.3. Aplicación de data mining.....	103
8.4. Big data Vs. Data mining	104
8.5. Cómo el periodismo se beneficia del data mining	105
Actividades de aprendizaje recomendadas	106
Autoevaluación 8.....	108
Semana 10	111
Unidad 9. Localización de datos	111
9.1. Ubicación o localización.....	111
9.2. Fuentes de datos	115
Autoevaluación 9	121
Semana 11	124
Unidad 10. Extracción.....	124
10.1.¿Qué comprende la extracción de datos?	124
10.2.Herramientas	126
Actividades de aprendizaje recomendadas	135
Autoevaluación 10.....	137
Semana 12	139
Unidad 11. Limpieza	139
11.1.Importancia de la limpieza de datos.....	139
11.2. Flujo de trabajo	141
11.3.Herramientas	145
Autoevaluación 11	149

Semana 13	151
 Unidad 12. Análisis e interpretación	151
12.1.Importancia	151
12.2.Tres preguntas para interpretar datos.....	154
12.3.Beneficios de la interpretación de datos	156
Actividades de aprendizaje recomendadas	157
Autoevaluación 12.....	158
 Resultado de aprendizaje 4.....	160
 Contenidos, recursos y actividades de aprendizaje recomendadas	160
Semana 14	161
 Unidad 13. Visualización	161
13.1.Importancia del dataviz	161
13.2.Categorías de gráficos.....	162
13.3.Etapas para crear una visualización.....	167
13.4.Catálogo de visualización	167
13.5.Herramientas	169
Autoevaluación 13.....	174
Semana 15	176
 Unidad 14. Productos periodísticos	176
14.1.Contar historias con datos	176
14.2.Storytelling	178
14.3.Tipos de usuarios (público).....	179
14.4.Ejemplos de productos periodísticos basados en datos	179
Autoevaluación 14.....	200
Actividades finales del bimestre.....	202
Semana 16	202
4. Solucionario	203
5. Glosario.....	219
6. Referencias bibliográficas	220
7. Anexos	224



1. Datos de información

1.1. Presentación de la asignatura



1.2. Competencias genéricas de la UTPL

Competencias genéricas, son aquellas capacidades (actitudes, habilidades y conocimientos) comunes a todas las profesiones que oferta la UTPL para la sociedad:

- Vivencia de los valores universales del Humanismo de Cristo.
- Comunicación oral y escrita.
- **Orientación a la innovación y a la investigación.**
- Pensamiento crítico y reflexivo.
- Trabajo en equipo.
- Comunicación en inglés.
- Compromiso e implicación social.
- Comportamiento ético.
- Organización y planificación del tiempo.

1.3. Competencias específicas de la carrera

La competencia específica de la Carrera de Comunicación para la asignatura es:

- Gestiona información para ser publicada en medios de comunicación masiva con el fin de informar a la sociedad.

1.4. Problemática que aborda la asignatura

El eje central de la asignatura es darles sentido a los datos (convertidos en información), los cuales, en su mayoría, provienen de la Red de redes, Internet, a través de la construcción de productos periodísticos. La asignatura de *Data Journalism* (DJ) tiene como objetivo central extraer, limpiar, analizar y visualizar información proveniente de altos volúmenes de datos disponibles en diversos entornos digitales y de esta manera crear productos periodísticos para la web, considerando *open data* y transparencia como eje de la información en la era digital. Se destaca, además el uso de la estadística, los géneros periodísticos y el periodismo de investigación que permitirán coadyuvar a la interpretación de datos y la consolidación en la construcción y narración de historias.



2. Metodología de aprendizaje

La metodología por utilizarse en el desarrollo de la asignatura de *Data Journalism* (DJ) se basa en “*learning by doing*”, la cual se centra en una conexión e hibridación entre contenidos teóricos y sobre todo prácticos, cuyo fin es el desarrollo de productos periodísticos empleando herramientas que se encuentran disponibles en la web. Además, de saber contar historias enfocadas en grandes volúmenes de datos. Y, usando como guía la taxonomía de Bloom, conocerá, comprenderá, analizará, creará, aplicará y evaluará, a través de las experiencias de los grandes medios de comunicación cómo incluir datos en la narración de historias informativas.

Para conocer más sobre la metodología le invito a revisar los siguientes artículos:

1. [Learning by doing.](#)
2. [Taxonomía de Bloom.](#)



3. Orientaciones didácticas por resultados de aprendizaje



Primer bimestre

Resultado de aprendizaje 1

- Conoce la importancia del periodismo de datos para la difusión de grandes bases de datos.

Contenidos, recursos y actividades de aprendizaje recomendada

A través del estudio de la asignatura se dará respuestas a las interrogantes sobre el periodismo de datos: ¿Qué es? ¿Cómo nace? ¿Dónde se desarrolla? ¿Por qué es necesario? ¿Cuáles son sus oportunidades y limitaciones?, etc. Y de esta manera se conocerá la importancia que tiene esta especialidad en el periodismo actual; de igual manera, las unidades y actividades propuestas le permitirán profundizar cómo la difusión de grandes bases de datos ha marcado un antes y un después en la emisión informativa.



Semana 1

Unidad 1. Introducción a Data Journalism (DJ)

1.1. Antecedentes

Sin duda el acontecimiento que permitió el arranque de este periodismo fue la llegada de la Internet, la cual derrumbó el monopolio de la producción de noticias en manos de los medios impresos, radio y televisión. Sin embargo, esta historia tiene sus raíces. El surgimiento de la Red fue y es el mayor impulsor del periodismo de datos en el mundo al permitir la formación de inmensas bases de información con fácil acceso e interpretación.

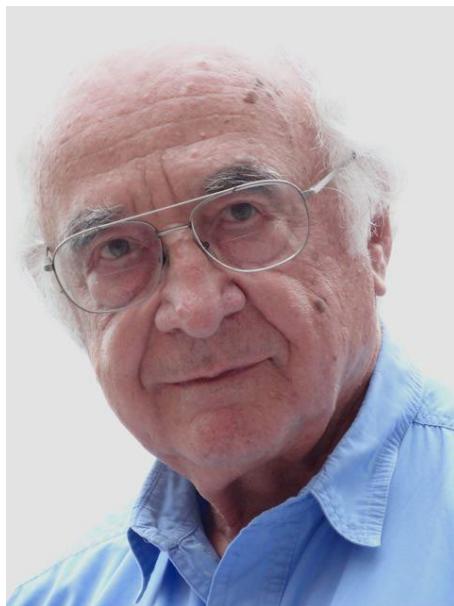
Como todo, este término tiene sus antecedentes, por ejemplo, sus bases están en el periodismo de precisión y de investigación. Por ello, le invito a revisar los siguientes artículos para más detalles:

1. [¿Qué es el Periodismo de Precisión?](#)
2. [¿Qué es el Periodismo de Investigación?](#)

Estos dos artículos le permitirán profundizar en el periodismo de precisión y periodismo de investigación y cómo incide directamente en el periodismo de datos. Entre los puntos más destacables del artículo están: concepto de periodismo de precisión y de investigación, función del periodismo de precisión y de investigación, la estadística aplicada al periodismo, el profesional en periodismo y las técnicas aplicadas en estas dos especialidades.

Figura 1.

Philip Meyer, creador del concepto de periodismo de precisión.



Nota. <https://es.wikipedia.org/wiki/PhilipMeyer>

Philip Meyer, estableció una tendencia nueva: el uso por parte de los reporteros de técnicas de investigación en ciencias sociales para aumentar la profundidad y precisión de las historias importantes. Creador del libro "The New Precision Journalism", muestra a los periodistas y estudiantes cómo utilizar las nuevas tecnologías para analizar datos y proporcionar

información más precisa y de forma más fácil entender, como por ejemplo: elementos de análisis de datos; el uso de estadísticas, computadoras, encuestas y experimentos; aplicaciones de bases de datos; y la política del periodismo de precisión. (Casal, 2010)

Tal como se describe en el Manual de Periodismo de Datos, Meyer trabajaba en «[The Miami Herald](#) y eligió investigar gracias a una denuncia sobre lo costoso que resultaban los seguros escolares contra incendios y huracanes que eran administrados por un consejo de funcionarios» (2021), que fueron elegidos por la comunidad. Consiguió los nombres de los ejecutivos y, con lápiz y papel en mano, demostró que 65 por ciento de las aportaciones de la campaña del presidente tenían origen en las compañías de seguros. Y, otros dos miembros del consejo conseguían más de la mitad de sus recursos de la misma fuente.

Por la iniciativa de los periodistas: Aron Pilhofer (*NYT*); Rich Gordon (Northwestern University) y Burton Herman (*AP*), se crea una organización para fomentar encuentros entre expertos en tecnología y reporteros para intercambiar ideas y encontrar puntos de unión entre las dos profesiones. (Manual de Periodismo de Datos y Ávila, 2021)

1.1.1. ¿Cómo llega el periodismo de datos a Latinoamérica?

A raíz de la publicación del libro de Philip Meyer en la década de los 90s, el Periodismo de Datos (PD) toma un leve significado teórico, pero no estaría respaldado por la gestión administrativa y sobre todo, periodística de organismos y medios de comunicación en Latinoamérica. Lo contrario sucedía en Estados Unidos, donde los *mass media* incluyeron en sus redacciones a perfiles mucho más técnicos como grafistas, estadistas, matemáticos, etc., lo que les permitió complementar y hacer equipo con la planta periodística.

1.1.2. ¿De dónde viene el término Data Journalism?

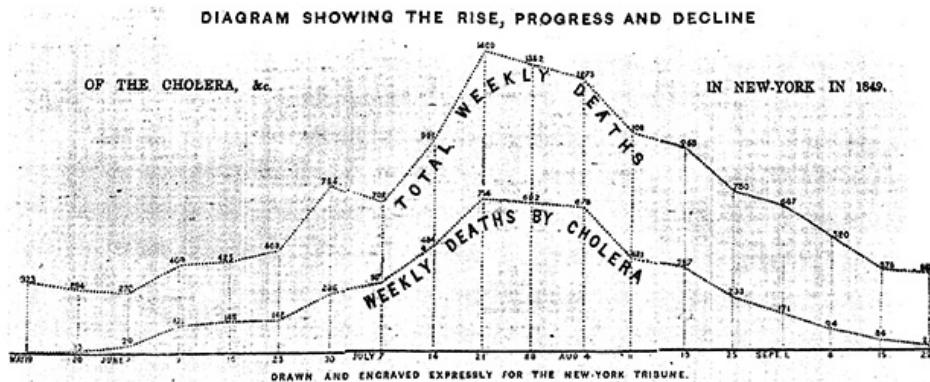
El blog *Rock Content* (Rock Content, 2019), menciona que «*Data-driven Journalism*» comenzó a utilizarse en Estados Unidos en el año 2009, lo que permitía nombrar la generación de productos periodísticos a partir de datos estructurados, adaptándose luego a los diferentes idiomas. Para conocer un poco más el origen del *Data Journalism* (DJ), le invito a leer el siguiente post, el cual permitirá obtener una visión histórica de su procedencia: <https://rockcontent.com/es/blog/periodismo-de-datos/>.

1.2. Inicio del trabajo con datos

Existen algunas iniciativas, pero entre los primeros indicios del trabajo con datos constan en los archivos del diario *The New York Daily Tribune*, que publicó el 29 de septiembre de 1849 un gráfico sobre la epidemia de cólera. A continuación, el gráfico:

Figura 2.

Diagrama que muestra la curva de evolución del cólera en Nueva York, 1849.



The above Diagram, or Graphic Table, (for which we are indebted to Professor GILLETT of Union College,) represents to the eye, in a very striking manner, the rise, progress, and decline of the Cholera, and other diseases in our City, during the last four months.

Each half-inch along the bottom line represents a week. The dates are placed under each. At the end of each half-inch, or week, are upright dotted lines, whose various lengths indicate the number of deaths by Cholera, and other causes, during that week; each inch on these upright lines corresponding to 500 deaths. The numbers are placed at the top of each. The zig-zag lines, which join the ends of these lines, show by their upward or downward slopes, whether the deaths during those weeks have increased or decreased, rapidly or slowly.

Some curious circumstances are apparent, on an inspection of this figure. We see that during the first two weeks of the Cholera, while it was increasing, as shown by

the line sloping upward, the total deaths were decreasing, as shown by the line sloping downward. This was probably caused by the great care in diet, &c. then practiced, on the first alarm. In the week ending July 7, while the Cholera line goes up the "Total deaths" line unaccountably goes down. Thence both lines go on ascending, and the corresponding number of deaths increasing, till in the week ending July 21, they reach their culminating point. Thence there is a constant and quite regular descent, excepting a sudden fall, from Aug. 4 to 11. This may have arisen from the diminution of population, consequent upon the custom of many persons to leave the city at the beginning of that month. The descent keeps on, till it brings up to our present normal condition.

If the average temperature, moisture, electrical state, &c. during these weeks were represented in the same manner, and added to this diagram, their comparison would show at a glance whether there has been any connection between them.

Nota. [Propublica.org](#). *The New York Daily Tribune*, septiembre 29, 1849.

El gráfico mostraba la proporción de muertes por cólera en relación con el número total de muertes en la ciudad, evidenciando un problema municipal de gran interés público. Aunque la medicina no había avanzado lo suficiente para comprender las causas de la enfermedad, el medio alertó a la población sobre su carácter letal y la necesidad de movilizar esfuerzos para prevenir un número masivo de muertes.

¡Lecturas recomendadas!

Para conocer más sobre esta evolución de DJ, le invito a revisar el enlace donde [Propublica.org](#) ([Propublica.org](#), 2016) hace mención a este acontecimiento: <https://www.propublica.org/herds/infographics-in-new-york-city-during-the-1849-cholera-epidemic>

[the-time-of-cholera](#), además a descargar el REA que contiene “*The Data Journalism Handbook*”: <https://www.oercommons.org/courses/the-data-journalism-handbook>. Estos recursos le permitirán acceder a información complementaria en cuanto a la evolución del periodismo de datos y, además, emplear estos conocimientos en las actividades prácticas que se plantean en la asignatura.

Según el blog Rock Content (Rock Content, 2019), otro medio que empleó datos fue el ganador de 17 premios Pulitzer, *The Philadelphia Inquirer*, quien en 1975 contrató a un reportero dedicado a analizar exhaustivamente los números del censo y así crear artículos exclusivos basados en datos numéricos. En perspectiva, netamente una parte de la macro data solo comenzó a ser usada por los medios de comunicación en los años 2000, sobre todo, por la dificultad analizar información estadística. En la actualidad esto ha ido cambiando gracias al avance de las tecnologías, desarrollando sistemas mucho más intuitivos que permiten extraer, analizar y visualizar información.

Estimado estudiante, hemos realizado una revisión introductoria al recorrido que el periodismo de precisión, junto al periodismo de investigación, han permitido desarrollar lo que hoy conocemos como periodismo de datos. A continuación, conozcamos su definición.

¡Video recomendado!



Actividades de aprendizaje recomendadas:

Es muy importante que se detenga y realice esta actividad que se propone, hacerla constituye un refuerzo a esta unidad estudiada. Observe el video “[History of Data Journalism at The Guardian](#)” e identifique cuándo, cómo y dónde nace el periodismo de datos según uno de los medios más importantes y pioneros en el periodismo tradicional, digital y en la actualidad, de datos. A la vez, realice un breve resumen en 280 caracteres, tipo tuit.

Consulte a su profesor:

De esta forma terminamos la unidad. Si tiene alguna duda por favor comuníquese con su profesor a través del EVA.

A continuación, es importante desarrollar la autoevaluación, a fin de determinar su grado de comprensión de los contenidos analizados.

Toda unidad de su guía incluye preguntas de autoevaluación que le recomiendo realizar para reforzar sus conocimientos aprendidos.



Autoevaluación 1

Señale la respuesta correcta:

1. ¿Qué aspecto permitió el arranque del Data Journalism?

- a. Internet.
- b. Periodismo.
- c. Científicos.

2. El periodismo de datos tiene como antecedente el periodismo:

- a. Tradicional
- b. Precisión.
- c. Digital.

3. El periodismo de datos tiene como raíz:

- a. Periodismo online.
- b. Periodismo de precisión e investigación.
- c. Periodismo tradicional.

4. ¿Quién estableció la nueva tendencia del Data Journalism?

- a. Sandra Crucianelli.
- b. Philip Meyer.
- c. Lluís Codina.

5. ¿Cómo llega el periodismo de datos a LATAM?

- a. Por iniciativa de los medios.
- b. Por pedido de los periodistas.
- c. Por la publicación del libro "The New Precision Journalism".

6. El término data journalism proviene del anglo:

- a. *Data Journalism*.
- b. *Driven Journalism*.
- c. *Data-driven Journalism*.

- 7. El medio que empieza a trabajar con grandes volúmenes de datos es:**
- a. *The New York Daily Tribune.*
 - b. *The New York Times.*
 - c. *The Philadelphia Inquirer.*
- 8. ¿Sobre qué tema fue la primera visualización de datos?**
- a. Córera.
 - b. Sika.
 - c. Influenza.
- 9. ¿Qué hizo el diario “*The Philadelphia Inquirer*”, para fortalecer su trabajo con datos?**
- a. Adquirió software para el análisis de datos.
 - b. Contrató a un reportero para analizar los datos.
 - c. Contrató una empresa especializada en analítica de datos.
- 10. Los datos masivos se empiezan a usarse de forma recurrente desde:**
- a. 1975
 - b. 1985
 - c. 2000

[Ir al solucionario](#)

Desarrolle las siguientes preguntas en su blog de notas:

- A. **¿Cuáles son los antecedentes del PD?**
- B. **¿Cómo podría “popularizarse” el periodismo de datos en LATAM y Ecuador?**
- C. **Analice el primer gráfico empleado sobre la evolución del cólera ¿Es comprensible?**

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.



Semana 2

Unidad 2. ¿Qué es Data Journalism o Periodismo de Datos?

Ahora bien, continuemos con el contenido de la asignatura.

“El periodismo de datos es la práctica de encontrar noticias en los números y usar los números para narrar noticias.” Meredith Broussard. (2018)

2.1. Definición

Es así como, para conocer qué es *data journalism* (*DJ*) se hará través de diversos autores y organizaciones que llevan la consigna del manejo informativo con datos de una manera más consciente y profunda.

La Real Academia Española define (Real Academia Española (RAE), 2021), **periodismo y datos** de la siguiente manera:

- **Periodismo:** «Actividad profesional que consiste en la obtención, tratamiento, interpretación y difusión de informaciones a través de cualquier medio escrito, oral, visual o gráfico.»
- **Datos:** «Información sobre algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho.»

Entre los expertos del DJ se encuentra Sandra Crucianelli quién define a esta especialidad cómo:

«Esta disciplina periodística se nutre de otras muchas: de investigación, en profundidad, de precisión, asistido por computadora y analítico. En ella, se trabaja con grandes volúmenes de datos, se aprovecha al máximo la visualización interactiva y se incorpora al programador al equipo periodístico.» (Crucianelli, 2013)

Para Érico Mafra (2019), *Data Journalism*, «es una modalidad de producción digital de noticias que utiliza grandes bases de datos para elaborar piezas de contenido enfocadas en la correlación de información y que, además, utiliza recursos gráficos e interactivos para hacer más agradable la experiencia de visualización del usuario consumidor de noticias.»

Mediante el anexo que se indica a continuación, propuesta por Lluís Codina (2019), se podrá apreciar cómo la evolución del término *Data Journalism* ha caminado por una transición de varias especialidades hasta convertirse hoy en lo que se conoce como periodismo de datos:

Términos para referirse al DJ.

En este sentido, se observa cuán relacionado está el periodismo de datos (PD) o periodismo de bases de datos con el periodismo de investigación, de profundidad, de precisión, asistido por computadora y analítico. A decir de Crucianelli (Crucianelli, 2013), son un todo a la vez, por ello se pregunta ¿periodismo de datos es..?:

- Periodismo de investigación: se usan técnicas propias del periodismo investigativo.
- Periodismo en profundidad: dependiendo del tema que se esté analizando.
- Periodismo de precisión: en ocasiones hay que recurrir a los métodos de investigación social definidos en la década de los sesenta por Philip Meyer para su aplicación en el campo del periodismo.
- Periodismo asistido por computadora: se debe analizar datos a través de hojas de cálculo (por ejemplo, Excel o un CSV).

- Es periodismo analítico: generalmente se usan métodos analíticos para llevar los datos a sistemas de información geográfica que permitan visualizar mejor el comportamiento de las variables.

Sandra Crucianelli (2013), acota que toda esta combinación de técnicas hacen el periodismo de datos, a las cuáles, además suma:

1. Grandes volúmenes de datos, que muchas veces ni siquiera podían alojarse en una hoja de cálculo de Excel.
2. Visualización interactiva.
3. Incorporación del programador al equipo periodístico para desarrollar tareas que van desde la extracción de datos y la depuración de los sets de datos hasta el diseño de aplicaciones de noticias (news apps).

2.2. Fórmula

Miremos a través de la experta en periodismo de datos, Sandra Crucianelli (2013) la fórmula que compone todo el entramado de esta especialidad:

$$\textit{PI} + \textit{PP} + \textit{PPr} + \textit{PA} + \textit{PAC} + \textit{volumen de datos} + \textit{visualización interactiva} + \textit{programación} = \textit{PdD o PBD}$$

- Periodismo de investigación (PI): tres son las condiciones necesarias para que un reportaje se precie de ser investigativo: 1) que sea un tema de relevancia social, 2) que alguien pretenda mantenerlo en secreto y, 3) se requiere un trabajo adicional del periodista, algunos medios destinan un profesional dedicado con exclusividad al tema.
- Periodismo de profundidad (PP): según la experta, es lo mismo que el periodismo de investigación, con la diferencia de que nadie intenta mantener en secreto el proceso de investigación periodística. Por ejemplo, un estudio sobre la salud pública con datos del Ministerio de Salud y otros organismos.
- Periodismo de precisión (PPr): hace uso de la investigación social empleando metodologías de índole cualitativa y cuantitativa.

- Periodismo analítico: (PA): Su trabajo se enfoca en ofrecer pruebas basadas en interpretaciones de la realidad, explicándolas de un modo comprensible. “Contextualiza un tema de fondo, mediante la descripción de detalles históricos y datos estadísticos. Utiliza como pilar de la visualización los llamados sistemas de información geográfica (SIG).”, explica Crucianelli.
- Periodismo asistido por computadora (PAC): conocido como *Computer Assisted Reporting* (CAR) y lo que pretende es utilizar un computador para la extracción y análisis de datos. Se caracteriza por el uso del software Excel.

2.3. Cómo ser un periodista de datos

La experiencia periodística es indispensable en esta especialidad, para ello, mediante la voz de Paul Bradshaw se identificarán los puntos más valiosos de cómo ser un periodista de datos en el siglo XXI.

Paul Bradshaw (2010), es líder del curso de la Maestría en Periodismo de Datos y Maestría en Periodismo Móvil y Multiplataforma en Birmingham City University, Reino Unido. También es consultor de periodismo de datos en la “Unidad de Datos Compartidos” de la BBC. Además, es cofundador de la galardonada red de periodismo de investigación: www.helpmeinvestigate.com.

Él plantea en su artículo publicado en *The Guardian* y al cual tituló: “[How to be a data journalism](#)”, una guía cómo iniciarse en el periodismo de datos, desde que estos llegan a la sala de redacción o de nuestros hogares (pandemia covid-19) hasta su visualización.

Para Bradshaw, el periodismo de datos es gigante. Y se refiere a gigante no por su volumen sino por la compleja e incomprendible amalgama de datos, que representa la convergencia de una serie de campos que son importantes como la investigación, las estadísticas, el diseño y la programación. En este sentido, la idea de combinar esas habilidades para contar historias importantes y poderosas, pero también intimidantes.

El autor propone 5 aspectos de cómo iniciar en el mundo de los datos:

1. Búsqueda de datos: la búsqueda de datos puede implicar cualquier cosa, desde tener conocimientos y contactos con expertos hasta poder utilizar habilidades de informes asistidos por computadora o, para algunos, habilidades técnicas específicamente como [MySQL](#) o [Python](#) que permiten recopilar nuestros propios datos.
2. Interrogar datos: interrogar bien los datos significa que debe tener una buena comprensión de la jerga y el contexto más amplio dentro del cual se encuentran los datos, además de las estadísticas, estar familiarizado con las hojas de cálculo puede ayudar a ahorrar mucho tiempo.
3. Visualización de datos: históricamente, visualizar y combinar datos ha sido responsabilidad de diseñadores y programadores, pero un número cada vez mayor de personas con experiencia editorial está probando ambas cosas, en parte debido a una mayor conciencia de lo que es posible y también a la reducción de las barreras, es decir experimentar con ellos.
4. Mezcla de datos: combinar los datos y cruzarlos es parte de la esencia del periodismo de datos, más aún si se involucran herramientas; tales como "[ManyEyes](#)"¹ para visualización y "[Yahoo! Pipes for mashups](#)"², las cuales le permitieron al autor lograr que los estudiantes de periodismo se involucren rápidamente con las posibilidades, y muchos se contagian del periodismo de datos.
5. Cómo empezar: Entonces, ¿dónde comienza un periodista de datos en ciernes? Una respuesta obvia sería «con los datos», pero también hay una segunda respuesta: «Con una pregunta».

Y es así como, desde la formulación de una pregunta o el tratar de resolver un problema pueden dar pie a iniciar con la preparación de una nota basadas en datos masivos, es decir, a iniciar a contar historias.

A través del siguiente cuadro se observan cuáles son las competencias y habilidades del periodista de datos:

Tabla 1.

Cuadro de competencias del periodista de datos.

1. A noviembre 2021, la web no existe. Se coloca un enlace de referencia.

Etapas	Acción	Técnicas	Competencias	Profesional necesario
Encontrar	Encontrar los datos	Evaluar las variables implicadas	Búsqueda avanzada	Periodista
Extraer	Extracción de datos	OCR, programación, scraping	<ul style="list-style-type: none"> ▪ Detectar el tipo formato de los datos para utilizar una buena técnica. ▪ Técnicas de extracción. ▪ Conocimientos básicos en programación 	Programador
Filtrar / limpiar	Detectar los errores y variables y encontrar los datos mal formateados	Uso de filtros o tablas dinámicas	Programa de análisis de datos (uso de Excel)	Programador
Contextualizar	Comparación y análisis	Entrevistar a los datos	<ul style="list-style-type: none"> ▪ Estadísticas ▪ Búsqueda de datos ▪ Observación 	Periodista
Comunicación	Comunicar los datos	<ul style="list-style-type: none"> ▪ Visualización ▪ Narrativa ▪ Periodismo 	<ul style="list-style-type: none"> ▪ Competencias narrativas ▪ Conocer las herramientas de creación de visualización 	Programador Desarrollador Diseñador Periodista

Nota. (Grassler, 2017)

2.4. Trabajo con datos. Caso: The Guardian

Mediante este caso se indaga cuál es el proceso que tiene el diario *The Guardian* cuando de publicar datos masivos se trata. El medio divide sus procesos a fin de contar con información que tenga valor y validez al momento de su difusión:

1. Compartir datos
 - a. Datos enviados
 - b. Eventos recurrentes
 - c. Teorías por explorar
 - d. Últimas noticias
2. Planillas de cálculos
 - a. Datos en formato equivocado
 - i. Celdas fusionadas
 - b. Datos medidos en unidades diferentes
 - i. Columnas de datos innecesarios
3. Hacer cálculos con los datos
 - a. Recalcular si fuera necesario
 - b. Verificar resultados
4. Visualización
 - a. Equipo de gráficos
 - b. Wordle
 - c. Many eyes
 - d. Google Explore Data
 - e. Google Fusión Tables

En el siguiente recurso, se podrá observar el flujo de trabajo que *The Guardian* realiza previo a la publicación de la información, tal como se observa el proceso está lleno de pasos que deben ser contrastados y verificados y, sobre todo, compartidos entre todos los miembros del equipo. Cabe destacar que este procedimiento lo realizan para cada tema que pretenden publicar.

[Trabajo con datos de The Guardian.](#)



Actividades de aprendizaje recomendadas

¡Video recomendado!

Estimado estudiante, mediante el siguiente video Usted podrá conocer cómo entre las historias, el contexto y la calidad se desenvuelve correctamente el periodismo de datos. Es muy importante que se detenga y realice esta actividad que se propone, hacerla constituye un refuerzo a esta unidad estudiada. Observe el contenido audiovisual denominado: "[Periodismo de datos: historias, contexto y calidad](#)" y realice un breve resumen en 280 caracteres, tipo tuit.

¡Lectura recomendada!

Se preguntará qué sucede en Ecuador con el desarrollo del periodismo de datos, para ello le invito a indagar en los medios nativos y migrantes para conocer si están trabajando en esta especialidad. Para ello, de manera introductoria le invito revisar la investigación: "[Periodismo de datos en Ecuador. Análisis de casos: El Comercio y Plan V](#)" y determinar su estado actual, lo cual le dará una cosmovisión de los medios nacionales.

Consulte a su profesor:

De esta forma terminamos la unidad. Si tiene alguna duda por favor comuníquese con su profesor a través del EVA.

A continuación, es importante desarrollar la autoevaluación, a fin de determinar su grado de comprensión de los contenidos analizados.



Autoevaluación 2

Toda unidad de su guía incluye preguntas de autoevaluación que le recomiendo realizar para reforzar sus conocimientos aprendidos.

Señale la respuesta correcta:

1. El periodismo de datos según Sandra Crucianelli es:

- a. "... es una modalidad de producción digital de noticias que utiliza grandes bases de datos es una modalidad de producción digital de noticias que utiliza grandes bases de datos..."
- b. "... en el PD trabaja con grandes volúmenes de datos, se aprovecha al máximo la visualización interactiva y se incorpora al programador al equipo periodístico.».
- c. «...la gestión de la información mediante el cual el contenido de las noticias se organiza en torno a datos estructurados, en lugar de noticias.»

2. El periodismo de datos según Érico Mafra es:

- a. "... es una modalidad de producción digital de noticias que utiliza grandes bases de datos es una modalidad de producción digital de noticias que utiliza grandes bases de datos..."
- b. Un periodismo basado en la explotación e interpretación de colecciones de datos, generalmente de origen público, aunque también pueden proceder de filtraciones...»
- c. «...la gestión de la información mediante el cual el contenido de las noticias se organiza en torno a datos estructurados, en lugar de noticias.»

- 3. La fórmula que propone Sandra Crucianelli para el trabajar el periodismo de datos es:**
- a. PI + PP + PPr + PA + PAC + volumen de datos + visualización interactiva = PdD o PBD.
 - b. PI + PP + PPr + PA + PAC + volumen de datos + visualización = PdD o PBD.
 - c. PI + PP + PPr + PA + PAC + volumen de datos + visualización interactiva + programación = PdD o PBD.
- 4. Lo que hoy conocemos como periodismo de datos proviene de términos como:**
- a. *Database journalism.*
 - b. *Analytics journalism.*
 - c. *MOJO journalism.*
- 5. Lo que hoy conocemos como periodismo de datos proviene de términos como:**
- a. *Investigative journalism.*
 - b. *Computational journalism.*
 - c. *Software journalism.*
- 6. Se conoce que el periodismo de precisión es parte del periodismo de datos, en esta perspectiva ¿de qué hace uso?**
- a. De la investigación social.
 - b. De la necesidad de informar.
 - c. Del uso de softwares.
- 7. Entre los aspectos que Paul Bradshaw menciona en su artículo “*How to be a data journalism*”, consta:**
- a. Cómo empezar.
 - b. Qué decir.
 - c. Qué hacer.

- 8. Entre los aspectos que Paul Bradshaw menciona en su artículo: “How to be a data journalism”, constan:**
 - a. Búsqueda de datos.
 - b. Extracción de datos.
 - c. Limpieza de datos.
- 9. El proceso del trabajo con datos de “The Guardian” inicia con:**
 - a. Compartir datos.
 - b. Verificar resultados.
 - c. Teorías básicas.
- 10. Para el proceso de visualización de datos, “The Guardian” requiere incluir un equipo de:**
 - a. Reporteros.
 - b. Grafistas.
 - c. Estadistas.

[Ir al solucionario](#)

Desarrolle las siguientes preguntas en su blog de notas:

- A. **¿Qué es el periodismo de datos?**
- B. **¿Qué potencial tiene DJ?**
- C. **¿Cuáles son las características del tratamiento de datos que tiene The Guardian?**

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.



Semana 3

Unidad 3. Trabajo con datos masivos

3.1. ¿Por qué los periodistas deben trabajar con datos masivos?

En la Unidad 3. Trabajo con datos masivos, se enfocará sobre por qué los periodistas deberían trabajar con datos masivos para la publicación de noticias. ¿Es realmente necesario? ¿Qué necesitamos para su publicación? ¿Las empresas periodísticas están interesadas en esta especialidad? ¿Los periodistas les temen a los números?

"El periodismo está situado. En el pasado, como sector, nos basábamos en ser los únicos que operábamos una tecnología para multiplicar y distribuir lo que había pasado de un día al otro. La imprenta servía como puerta de entrada. Cualquiera que quisiera llegar a la gente de una ciudad o una región a la mañana siguiente, recurría a los diarios. Hoy las noticias fluyen al mismo tiempo que suceden, a través de múltiples fuentes, testigos presenciales, blogs y social media, y lo que ha sucedido es filtrado a través de una vasta red de conexiones sociales, se jerarquiza, se comenta y muy a menudo se ignora." (European Journalism Centre y Google News Initiative, 2021)

- *Usar datos en el periodismo: porqué*

En *The Data Journalism Handbook 2* (2021), dos de sus colaboradores, Mirko Lorenz y Deutsche Welle explican a detalle por qué el periodismo y sobre todo los periodistas deben emplear datos masivos en el ejercicio de la profesión. Entre los puntos se destacan:

1. Porque **amplía un nuevo campo de especialización**.
2. Porque **los usuarios requieren otro tipo de contenido**.
3. **Reunir, filtrar y visualizar lo que sucede más allá** de lo que nos muestran nuestros ojos.
4. El lenguaje de esta red es el de los datos: **pequeños puntos de información** que a menudo son irrelevantes como instancia individual, pero **enormemente importantes cuando se los ve desde el ángulo correcto**.
5. Usando datos, la tarea de los periodistas pasa de centrarse en ser los primeros en informar, **a ser los que nos dicen lo que un proceso podría significar realmente**.
6. Usar datos para **transformar algo abstracto en algo que todos pueden entender y con lo que pueden relacionarse**.
7. Pueden **crear herramientas de cálculo personalizadas** para ayudar a la gente a **tomar decisiones**, se trate de comprar un auto o una casa, decidir un rumbo educativo o profesional en su vida, o hacer un control de costos para no incurrir en deudas.
8. **Pueden analizar la dinámica de una situación compleja** como disturbios o un debate político, mostrar falacias y ayudar a todos a **encontrar posibles soluciones** para problemas complejos.
9. Formarse en la búsqueda, depuración y visualización de datos es transformador para la profesión de reunir información. **Los periodistas que dominen esto descubrirán que apoyar sus artículos en datos y la visión que aportan es un alivio**. Menos adivinar, menos buscar citas; en vez de ello, **un periodista puede crear una posición fuerte apoyada en datos y esto puede afectar mucho el rol del periodismo**.

10. Introducirse en el periodismo de datos ofrece una perspectiva para el futuro. **Pero los periodistas de datos o los científicos de datos ya son un grupo de profesionales muy solicitados, no solo por los medios.** Las empresas e instituciones de todo el mundo buscan “**gente que encuentre sentido a las cosas**”, y **profesionales que sepan cómo revisar datos y convertirlos en algo tangible.**
11. **Los datos representan una promesa, y esto es lo que entusiasma a las redacciones, haciéndolas buscar un nuevo tipo de periodista.** Para la gente que trabaja por su cuenta, manejar datos ofrece un camino para obtener nuevas oportunidades y un salario estable también.
12. Trabajar con datos es **como introducirse en un territorio vasto y desconocido.** A primera vista los datos crudos resultan inteligibles para los ojos y la mente. Tales datos son inmanejables. Es difícil ordenarlos correctamente para su visualización. Se necesita periodistas experimentados, que tengan la **energía como para analizar datos crudos a menudo confusos o aburridos y “ver” las historias ocultas allí.**
13. **No responden a intereses particulares.** Es fundamental que la información provenga de fuentes confiables y que no estén influenciadas por intereses comerciales o ideológicos. En este sentido, los números garantizan su veracidad.
14. Aumentan la **calidad de las producciones.** El uso de herramientas que permiten la visualización de datos proporciona a las audiencias otro tipo de experiencias, en algunos casos mucho más interactivas e inmersivas.

A decir de Lluís Codina (2019), como periodista de datos, podemos añadir, además:

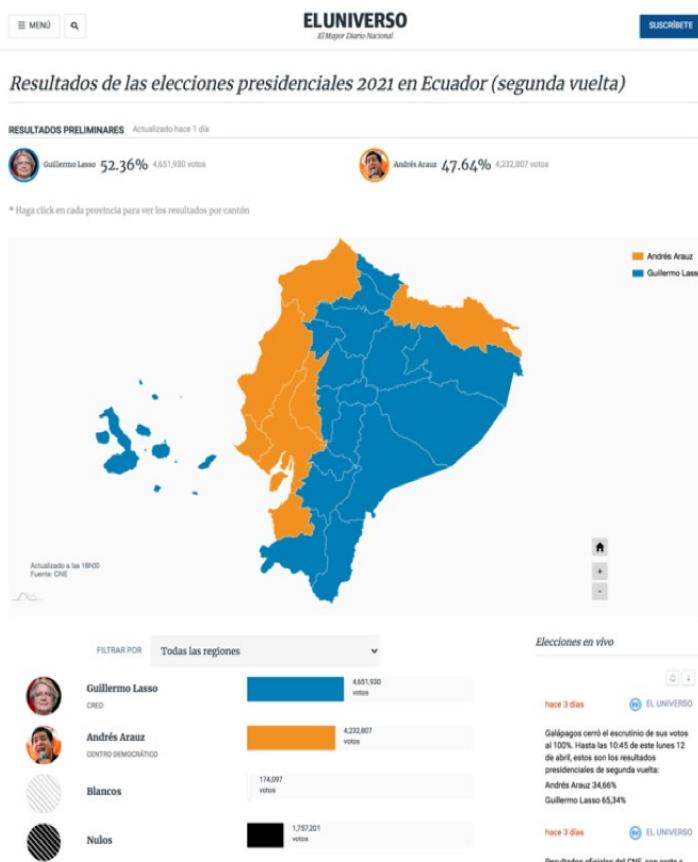
15. **Saber qué se ha publicado antes sobre un tema** o una historia o secuencia de hechos noticiales determinado.
16. **Saber cómo ha sido tratado este tema.**
17. **Identificar actores sociales, entidades y posibles fuentes** de información para un determinado tema o línea de temas.

18. **Encontrar antecedentes o datos biográficos** que hayan generado noticias con anterioridad de un determinado personaje de actualidad: un político, un artista, un empresario, etc., al que queremos entrevistar
19. **Derivar ideas, enfoques y nuevas dimensiones de un ámbito noticioso** que el periodista se haya propuesto tratar, o al que le hayan asignado en el medio para el que trabaja.
20. **Datos certeros no matan el relato, dan credibilidad** a las historias que se pretende difundir.

Estos son 20 puntos que proponen los expertos, no olvidemos que lo que busca en esencia *data journalism* es ubicar, recopilar y analizar grandes cantidades de datos para hacerlos comprensibles a los usuarios a través de artículos, visualizaciones o apps.

Figura 3.

Ejemplo uso de datos. Elección presidencial de Ecuador, 2021.



Nota. (El Universo, 2021)<https://www.eluniverso.com/resultados-elecciones-ecuador-presidente-2021/>

3.2. ¿Por qué los datos son el rey o reina del contenido?

3.2.1. Why journalists should use data

A continuación, se responden las interrogantes ¿por qué los datos son el rey o reina del contenido? y ¿por qué los periodistas deberían usar datos?, a través de la mirada de quiénes han visto en el periodismo de datos una especialidad que se moldeó, gracias al periodismo de precisión y de investigación y llegó lo que hoy en día se conoce como DJ:

Tabla 2.

El porqué es importante DJ. Fuente: (*Manual de Periodismo de Datos 1.0, 2014*).

Nº	Su porqué	Expertos
1.	Filtrar el flujo de datos.	Philip Meyer, University of North Carolina.
2.	Nuevos enfoques para narrar historias.	Aron Pilhofer, <i>The New York Times</i> .
3.	Como periodismo fotográfico con una laptop.	Brian Boyer, <i>Chicago Tribune</i> .
4.	El periodismo de datos es el futuro.	Tim Berners-Lee, fundador de la World Wide Web, WWW.
5.	El procesamiento de cifras se une al pulido del lenguaje.	David Anderton, periodista freelance.
6.	Actualizar sus capacidades.	Jerry Vermanen, NU.nl .
7.	Un remedio para la asimetría de la información.	Tom Fries, Bertelsmann Foundation.
8.	Una respuesta a las relaciones públicas basadas en datos.	Nicolas Kayser-Bril, <i>Journalism++</i> .
9.	Proveer interpretaciones independientes de información oficial.	Isao Matsunami, <i>Tokyo Shimbun</i> .
10.	Manejar el diluvio de datos.	Alex Howard, O'Reilly Media.
11.	Nuestras vidas son datos.	Chris Taggart, Open Corporates.
12.	Una manera de ahorrar tiempo.	Pedro Markun, Transparency Hackers.
13.	Una parte esencial del herramiental del periodista.	Cynthia O'Murchu, <i>Financial Times</i> .
14.	Adaptarse a cambios en nuestro ambiente de información.	César Viana, University of Goiás.
15.	Una manera de ver cosas que de otro modo podría no ver.	Cheryl Phillips, <i>The Seattle Times</i> .
16.	Una manera de enriquecer los artículos.	Sarah Slobin, <i>Wall Street Journal</i> .
17.	Una manera de enriquecer los artículos.	Jonathan Stray, The Overview Project.

Estimado estudiante, para profundizar en estos 17 puntos que resaltan los expertos, le invito a leer el capítulo "Por qué deberían usar datos los periodistas", contenido que consta en el "Manual de Periodismo de Datos 1.0": https://interactivos.lanacion.com.ar/manual-data/introducci%C3%B3n_2.html. Este capítulo le permitirá complementar el contenido expuesto en la tabla 2 y poseer más argumentos enfocados en porqué los periodistas están llamados a emplear datos masivos para su trabajo periodístico.

Ahora bien, le invito a continuar con la revisión de la asignatura a través de la Unidad 4.



Actividades de aprendizaje recomendadas

Es muy importante que se detenga y realice esta actividad que le proponemos, hacerla constituye un refuerzo a esta unidad estudiada. Observe el video "[La importancia de los datos](#)", aquí se menciona que la estadística es fundamental, dado que a través de datos certeros se puede comprender el mundo cambiante en el que vivimos, tal como lo ha demostrado la pandemia por coronavirus. Ahora bien, le invito a realizar un breve resumen en 280 caracteres, tipo tuit y compartirlo en las tutorías.

Consulte a su profesor:

De esta forma terminamos la unidad. Si tiene alguna duda por favor comuníquese con su profesor a través del EVA.

A continuación, es importante desarrollar la autoevaluación, a fin de determinar su grado de comprensión de los contenidos analizados.

Toda unidad de su guía incluye preguntas de autoevaluación que le recomiendo realizar para reforzar sus conocimientos aprendidos.



Autoevaluación 3

Señale la respuesta correcta:

1. Por qué usar datos?

- a. Porque amplía un nuevo campo de especialización.
- b. Porque amplía la producción periodística.
- c. Porque amplía los formatos en la web.

2. ¿Por qué usar datos?

- a. Porque los usuarios requieren contextualización.
- b. Porque los usuarios requieren otro tipo de contenido.
- c. Porque los usuarios requieren ser escuchados.

3. ¿Para qué usar datos?

- a. Reunir, y visualizar lo que sucede más allá de lo que nos muestran nuestros ojos.
- b. Filtrar y visualizar lo que sucede más allá de lo que nos muestran nuestros ojos.
- c. Reunir, filtrar y visualizar lo que sucede más allá de lo que nos muestran nuestros ojos.

4. Mediante el uso datos, la tarea de los periodistas pasa de centrarse en ser los primeros en:

- a. Informar a través de redes.
- b. Tener la primicia sin importar el cómo.
- c. Informar, a ser los que nos dicen lo que un proceso podría significar realmente.

5. Los datos representan:

- a. Una promesa, y esto es lo que entusiasma a las redacciones, haciéndolas buscar un nuevo tipo de periodista.
- b. Una oportunidad, y esto es lo que entusiasma a las redacciones, haciéndolas buscar a nuevos especialistas.
- c. Una ventaja, y esto es lo que entusiasma a las redacciones, haciéndolas buscar nuevas oportunidades.

6. Los datos no responden a:

- a. Intereses particulares.
- b. Intereses políticos.
- c. Intereses ideológicos.

7. ¿Qué permiten los datos masivos?

- a. Aumentar la calidad de la investigación.
- b. Aumentar la calidad de las producciones.
- c. Aumentar el *engagement* del producto.

8. ¿Qué experto considera que los datos permiten identificar actores sociales y posibles fuentes?

- a. Mirko Lorenz.
- b. Deutsche Welle.
- c. Lluís Codina.

9. ¿Qué experto hace hincapié en que los datos permiten: “Nuevos enfoques para narrar historias.”

- a. Philip Meyer.
- b. Aron Pilhofer.
- c. Brian Boyer.

10. Complete: Para el experto _____ del medio _____, los datos permiten proveer interpretaciones independientes de información oficial”.

- a. Isao Matsunami, *Tokyo Shimbun*.
- b. Cynthia O'Murchu, *Financial Times*.
- c. Sarah Slobin, *Wall Street Journal*.

[Ir al solucionario](#)

Desarrolle las siguientes preguntas en su blog de notas:

- A. A su criterio ¿por qué el periodista debe usar datos masivos?**
- B. ¿Qué otros puntos sumaría al porqué usar datos en el periodismo?**

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.

Resultado de aprendizaje 2 ▪ Emplea una correcta gestión de datos en entornos digitales.

Contenidos, recursos y actividades de aprendizaje recomendadas

Mediante el estudio conocerá cómo emplear correctamente los datos en la red, y a través de las siguientes interrogantes: ¿Qué producto deseo obtener? ¿El usuario lo podrá comprender? ¿Estoy pensando en cómo será receptado por mi audiencia?, se dará respuesta a fin de conocer e identificar los productos periodísticos que se puedan obtener gracias al uso de datos masivos.



Semana 4

Unidad 4. Productos basados en datos

Cuatro son los productos que son empleados para narrar historias a partir de un volumen considerable de datos:

1. Artículos
2. Visualizaciones
3. Datasets
4. News Apps

Estimado estudiante, a continuación, le invito a revisar cada una de ellas.

4.1. Artículos

A decir de Crucianelli (2013), este producto son artículos que nacen a partir de un volumen considerable de datos y que nos permiten construir historias; sobre todo, son artículos de corta extensión, y que en su mayoría emplean texto detallando cifras, valores, números, porcentajes, entre otros.

A continuación se detallan ejemplos de artículos periodísticos (enlaces directos) que permiten leer, de una forma no lineal cómo los datos masivos (abundantes) permiten construir historias visualmente atractivas y cuyos datos de base son densos de interpretar y que gracias al trabajo del equipo de datos se pueden transformar en:

- **La Nación:** “*Los millones de la APE: cómo se repartió en 2011 la caja que era de Moyano*”. <https://www.lanacion.com.ar/politica/los-millones-de-la-ape-como-se-repartio-en-2011-la-caja-que-era-de-moyano-nid1484852/> (Jastreblansky, 2012)
- **La Nación:** “*Pandora Papers: una offshore suma misterio a la desaparición de US\$7 millones en la empresa de jugos Baggio*”. <https://www.lanacion.com.ar/politica/pandora-papers-una-offshore-suma-misterio-a-la-desaparicion-de-los-us7-millones-en-la-empresa-de-nid01112021/> (Alconada et al., 2021)
- **Primicias.ec:** “*Más de 159.000 personas dejaron de tener un empleo adecuado en octubre*”. <https://www.primicias.ec/noticias/economia/deterioro-empleo-adecuador-ecuador-octubre/> (Coba, 2021)

Mediante las siguientes capturas de pantalla y enlaces directos a los artículos periodísticos de *La Nación* y *Primicias.ec*, podrá dimensionar el trabajo y proceso de redacción (textos largos o *longform*) de los temas tratados por los medios que llegan a través de bases de datos.

A continuación, se presenta ejemplos de artículos con datos masivos:

[Ejemplo de artículo con datos masivos.](#)

4.2. Visualizaciones

La visualización de datos o *data visualization* (dataviz), permite a través de imágenes, gráficos, infografías, pictogramas y de más elementos, mostrar información obtenida a través grandes volúmenes de datos. A decir de Crucianelli (2013), pueden estar complementados por artículos cortos o de gran extensión, y en algunos casos ir solo acompañados con un titular y una explicación para que los usuarios puedan analizarlos y leerlos.

A continuación, se detallan ejemplos de artículos periodísticos que permiten descubrir y entender a través de visualizaciones cómo con los datos masivos se cuentan historias visualmente atractivas:

- **La Nación:** "La evolución de la pandemia en la Argentina" <https://www.lanacion.com.ar/sociedad/en-detalle-infectados-fallecidos-coronavirus-argentina-nid2350330/#/> (La Nación, 2021)
- **Primicias.ec:** "Más de 159.000 personas dejaron de tener un empleo adecuado en octubre". <https://www.primicias.ec/noticias/economia/deterioro-empleo-adecuador-ecuador-octubre/> (Coba, 2021)

Por medio de las siguientes capturas de pantalla y enlaces directos a los artículos periodísticos de *La Nación* y *Primicias.ec*, podrá analizar el proceso en relación al trabajo de visualización que han realizado los especialistas en el área, como pueden ser los llamados infografías.

En los artículos se destacan el uso de gráficos estáticos y animados para contar una realidad, en el caso de *Primicias.ec*, sobre el mercado laboral en Ecuador, y *La Nación*, la evolución del coronavirus en Argentina. Como se ha visto, esto tiene de fondo un manejo de bases de datos que, para hacerlos más accesibles y digeribles para el usuario, se crean este tipo de visualización.

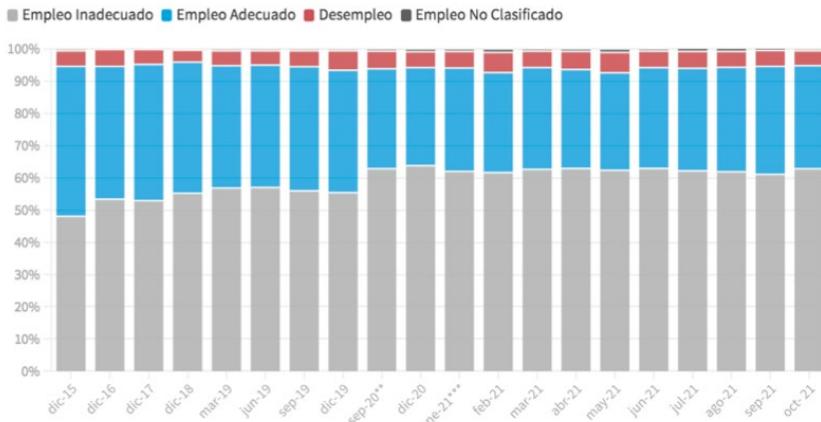
Figura 4.

Ejemplo de artículo con datos masivos.

Un segmento del mercado laboral que se ha mantenido es el desempleo, con 384.204 personas.

Mercado laboral ecuatoriano

Mueva el cursor sobre el gráfico para visualizar las cifras



Fuente: INEC • Gráfico: Eduardo Cobos- Primicias

**Entre septiembre y diciembre del 2020 se delimitó el tamaño de la muestra

***Existieron cambios metodológicos asociados al tamaño y distribución de la muestra, y representatividad estadística, por tanto, las cifras no son estrictamente comparables con los períodos anteriores

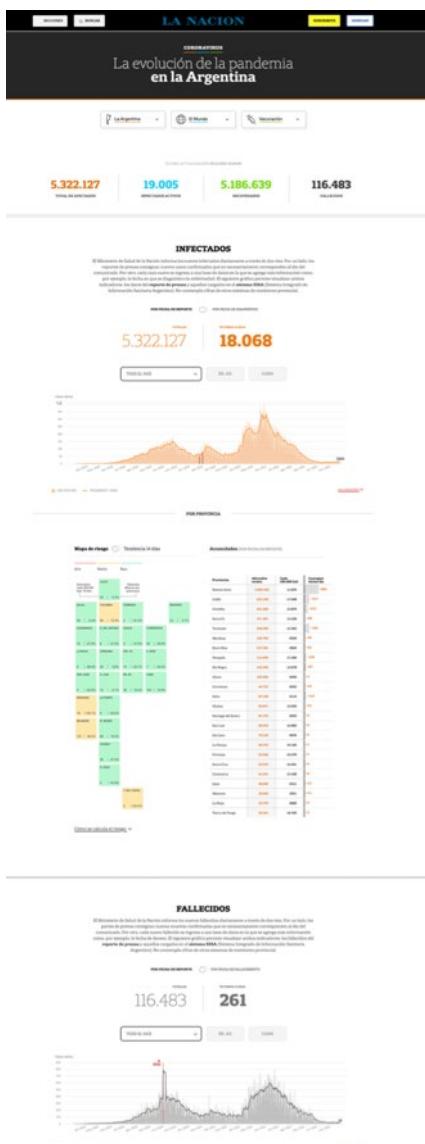
Empleo inadecuado incluye: subempleo, empleo no remunerado y otro empleo no pleno.

A Flourish chart

Nota. [Primicias.ec, 2021. <https://www.primicias.ec/noticias/economia/deterioro-empleo-inadecuado-ecuador-octubre/>](https://www.primicias.ec/noticias/economia/deterioro-empleo-inadecuado-ecuador-octubre/)

Figura 5.

Ejemplo de artículo con datos masivos.



Nota. La Nación, 2021 <https://www.lanacion.com.ar/sociedad/en-detalle-infectados-fallecidos-coronavirus-argentina-nid2350330/#/>

4.3. Datasets

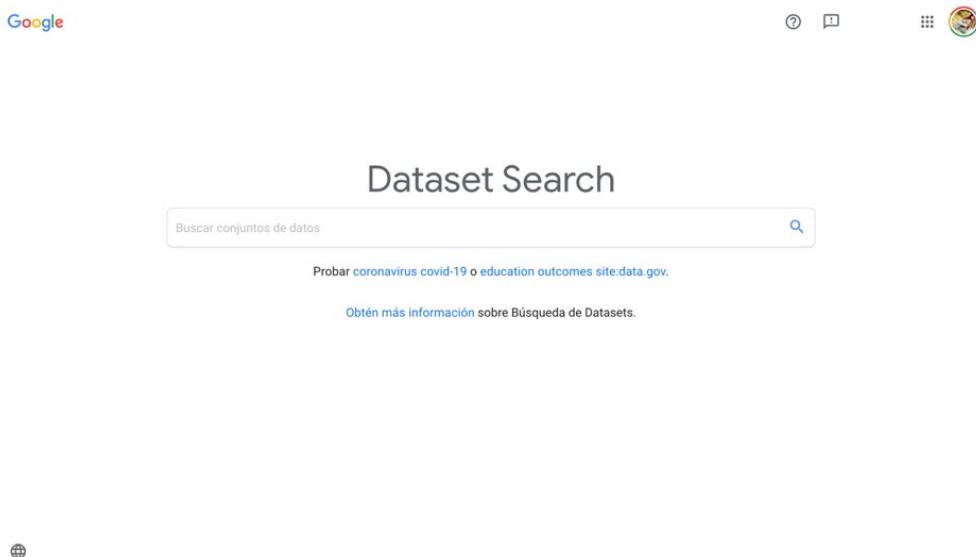
Como su nombre lo indica se trata de una colección de datos, conjunto de datos, o dataset (término en inglés) que permite acceder a la información

para su procesamiento. En cierto sentido suele estar ligado a la transparencia de las organizaciones. En Internet, se puede acceder a bases de datos conforme la temática que se busque, pero esto dependerá de quién los provea y su transparencia ante la comunidad, esto es lo que se considera “open data”.

En este apartado se detalla más la plataforma “Google Dataset Search”. Para acceder al motor de búsqueda enfocado en conjunto de datos, puede realizarlo en el siguiente enlace: <https://datasetsearch.research.google.com/>

Figura 6.

Motor de búsqueda “Google Dataset Search”.



Nota. GDS <https://datasetsearch.research.google.com/>

Figura 7.

Motor de búsqueda “Google Dataset Search”, ejemplo “COVID-19”.

The screenshot shows the Google Dataset Search interface. At the top, there's a search bar with "covid-19". Below it, a navigation bar includes "Última actualización", "Formato de descarga", "Derechos de uso", "Tema", "Gratis", and "Conjuntos de datos guardados". The main search results are displayed in cards:

- [DEPRECATED] Datos del coronavirus de la enfermedad de COVID-19** (data.europa.eu):
 - Logo: eodc
 - Description: [DEPRECATED] Datos del coronavirus de la enfermedad de COVID-19
 - Actions: Explorar en data.europa.eu, csv, excel xlsx, html +3
 - Details: Última actualización: Mar 6, 2020
- Coronavirus Disease (COVID-19) – the data** (ourworldindata.org):
 - Logo: Our World In Data
 - Description: Coronavirus Disease (COVID-19) – the data
 - Actions: csv
 - Details: Última actualización: Apr 16, 2020
- kaggle World Health Organization COVID-19 data** (www.kaggle.com):
 - Logo: kaggle
 - Description: World Health Organization COVID-19 data
 - Actions: zip
 - Details: Última actualización: Apr 16, 2020
- WHO Coronavirus disease (COVID-19) situation reports** (WHO):
 - Logo: WHO
 - Description: WHO Coronavirus disease (COVID-19) situation reports
 - Actions: Replaced by <http://data.europa.eu/88u/dataset/covid-19-coronavirus-data-daily-up-to-14-december-2020>

Nota. Google Data Search: <https://datasetsearch.research.google.com/>

Como se observa en la Figura 7, los datos pueden descargarse en varios formatos (*excel xlsx, html, xml, csv, rss feed, json*, etc.) a fin de organizarlos y procesarlos de una forma eficaz. Si bien Google muestra sets actualizados también indica de dónde provienen.

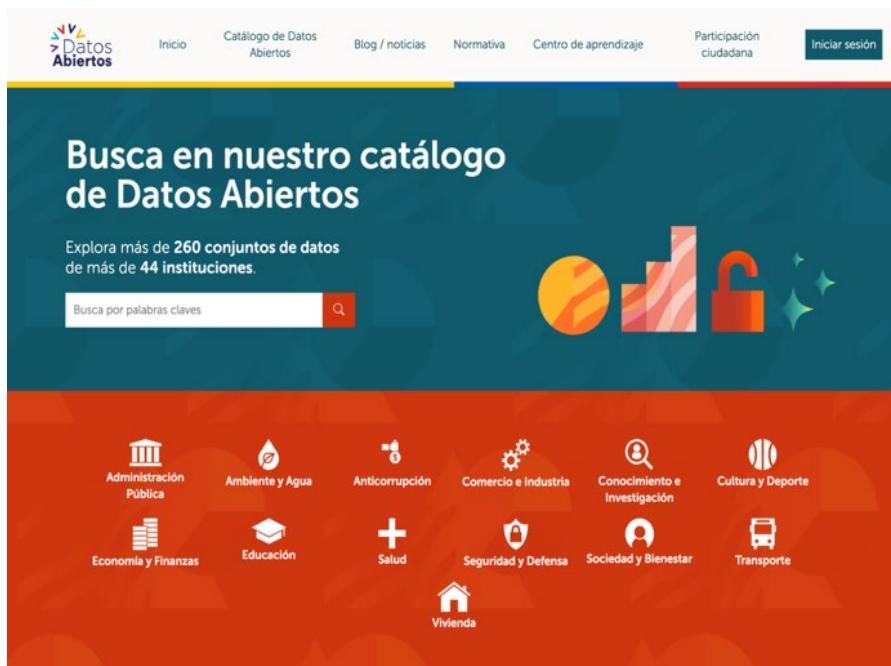
Lo adecuado e ideal es buscar la fuente de datos primarias y descargar sus datos, esto en el caso de que sean libres. Algunas organizaciones como la [Organización Mundial de Salud \(OMS\)](#) tiene su propio catálogo con conjuntos de datos a los cuales podemos acceder sin restricciones.

Otro ejemplo de conjunto de datos lo encontramos en diario La Nación (Argentina) quién creó un departamento exclusivo para contar historias a través de grandes bases de datos. O, de la plataforma Tableau, que a más de visualizar datos suele entregar datasets, tales como: <https://www.tableau.com/es-es/learn/articles/free-public-data-sets>.

En Ecuador existe la iniciativa de “Datos abiertos”: <https://datosabiertos.planificacion.gob.ec/>

Figura 8.

Web de Datos Abiertos, Ecuador.



Nota. Datos Abiertos <https://datosabiertos.planificacion.gob.ec/>

Lo mismo ocurre en Argentina, conjuntos de datos proporcionados por el gobierno nacional, para ello lo invito a revisar la web oficial, en donde se expone una serie de set de datos enfocados en temas que van desde salud, cultura, política, educación, justicia, etc., todos ellos son de carácter libre para quien desee descargarlos: <https://datos.gob.ar/dataset>

Figura 9.
Web de datos abiertos, Argentina.

The screenshot shows the homepage of the datos.gob.ar website. At the top, there's a navigation bar with links to 'Datasets', 'Series', 'Organizaciones', 'APIs', and 'Acerca de'. The main title 'Datasets' is displayed prominently, along with the number '1099 DATASETS'. Below this, there's a search bar labeled '¿Qué dataset buscas?' and a dropdown menu set to 'Última modificación'. On the left side, there's a sidebar with sections for 'Temas' (Icons for Transport, Environment, Science, Culture, Defense, etc.) and 'Organizaciones' (List of government ministries). The main content area lists several datasets with their names, descriptions, and small icons:

- Sistema Integrado de Transporte Automotor (SISTAU)
Descripción: Dirección de Acceso a la Información Pública y Transparencia. Subsidios y Compensaciones Transporte Automotor de Pasajeros. Son aquellas transferencias realizadas por la Adm...
- Sistema Ferroviario Integrado (SIFER)
Descripción: Dirección de Acceso a la Información Pública y Transparencia. Subsidios y Compensaciones Transporte Ferroviario Son aquellas transferencias realizadas por la Adm. Pública a...
- Aterrizajes y despegues procesados por la Administración Nacional de Aviación Civil...
Descripción: Ministerio de Transporte. Secretaría de Articulación Interjurisdiccional.. Detalle de aterrizajes y despegues procesados por la Administración Nacional de Aviación Civil
- PMyE - Indicadores de contexto - Infraestructura Hídrica
Descripción: Dirección Nacional de Transparencia. Ministerio de Obras Públicas. Indicadores de contexto - Infraestructura Hídrica

Nota. Datos abiertos, Argentina <https://datos.gob.ar/dataset>

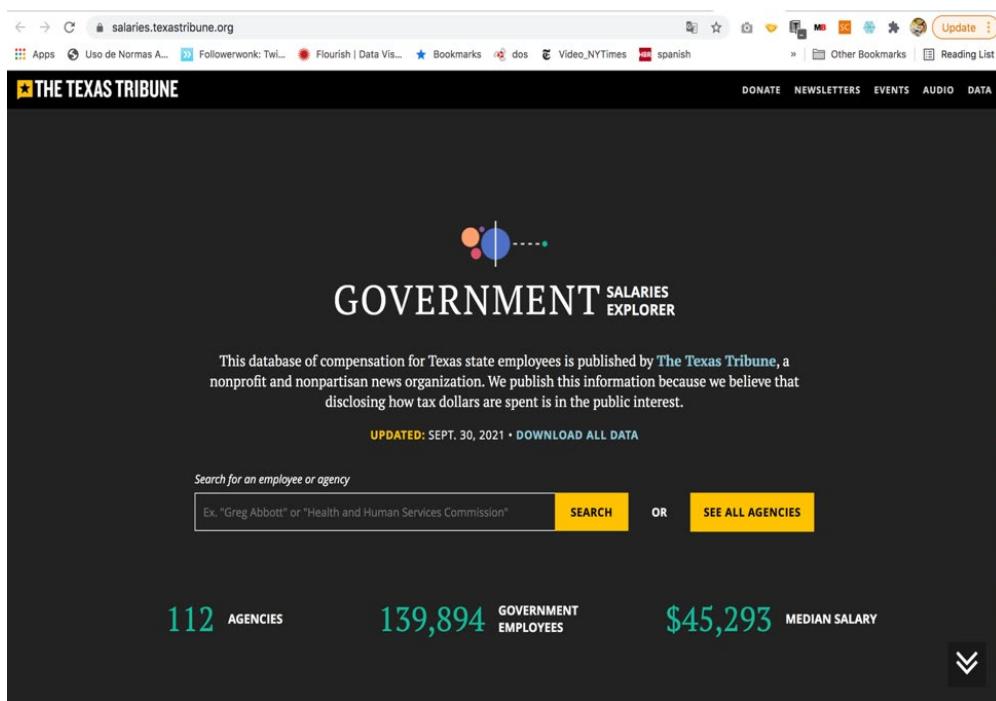
4.4. News Apps

Las aplicaciones de noticias son conocidas como “*news apps*” y permiten concentrar y centrar una historia en un solo espacio, a fin de contarla mucho más allá de un solo artículo. Para Crucianelli (2013), “muchas veces, el volumen de datos es tan grande que resulta imposible encontrar una noticia si no se diseña una aplicación que permita agrupar y analizar variables, por ejemplo, por geolocalización, por fecha, por nombre de compañía, etc.”

Entre los ejemplos de aplicaciones de noticias, se encuentran:

- *The Texas Tribune*: aplicación a modo de “buscador” que permite mostrar cómo se gastan los impuestos por parte de entes de gobierno y los cuales son de interés público: <https://salaries.texastribune.org/>

Figura 10.
Ejemplo de news app.



Nota. The Texas Tribune: <https://salaries.texastribune.org/>

- *Propublica:* muestra a través de esta aplicación el valor que los doctores y hospitales de EE. UU reciben de la industria farmacéutica, esto por ley: <https://projects.propublica.org/docdollars/>

Figura 11.
Ejemplo de news app.

PROPUBLICA | See more at VITAL SIGNS

Dollars for Docs

How Industry Dollars Reached Your Doctors

By Mike Tigas, Ryann Grochowski Jones, Charles Ornstein, and Lena Groeger, ProPublica. Updated October 17, 2019

Pharmaceutical and medical device companies are required by law to release details of their payments to a variety of doctors and U.S. teaching hospitals for promotional talks, research and consulting, among other categories. Use this tool to search for general payments (excluding research and ownership interests) made from August 2013 to December 2018. [[About the Dollars for Docs Data](#)] [Read the entire series](#)

Has Your Doctor Received Drug or Device Company Money?

For example: Andrew Jones, Boston, 10013

All States

Archive: Search our older data for payments made by 17 drug companies between 2009 and 2013.

Get Updates

Sign up to be notified when Dollars for Docs is updated, and get more of ProPublica's stories in your inbox.

Email address

This site is protected by reCAPTCHA and the Google [Privacy Policy](#) and [Terms of Service](#) apply.

[Download the Data](#) | [Patients, Take Action](#) | [Reporters, Use Our Data](#)

 **\$12B** in disclosed payments

 **1,036,163** doctors

 **1,249** teaching hospitals

 **2,191** companies

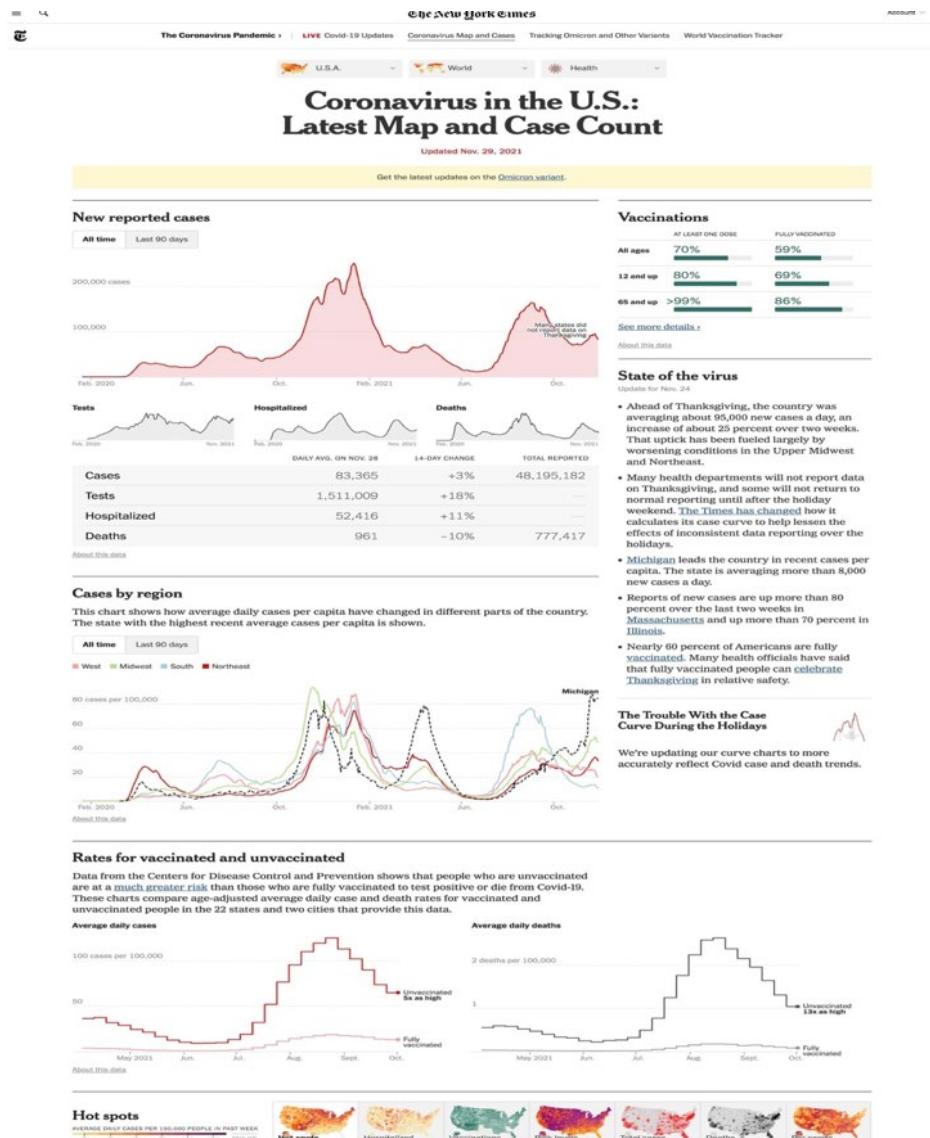
Nota. Propublica.org: <https://projects.propublica.org/docdollars/>

Para conocer cómo el medio accedió a la información de este set de datos le invito a leer el “tras cámara o tras los datos masivos” del trabajo periodístico: <https://www.propublica.org/article/about-the-dollars-for-docs-data>

El medio *The New York Times*, tiene destinadas ciertas secciones para contar información basadas en datos, un ejemplo de ello son los temas de salud, caso específico “covid-19”. En la siguiente figura podrá apreciar su forma de narrar este tipo de acontecimientos, que se caracterizan por el uso de gráficos, diagramas, tablas y mapas, que claramente permiten dimensionar y visualizar cómo evoluciona la enfermedad en ciertos momentos y lugares de Estados Unidos.

Figura 12.

Contador de casos y mapas sobre Coronavirus en EE. UU.



Nota. The New York Times: <https://www.nytimes.com/interactive/2021/us/covid-cases.html>

Para este tipo de productos periodísticos se recurre no solo al periodista sino a otro tipo de perfiles, tales como: programadores, desarrolladores, diseñadores, estadistas y más.



Actividades de aprendizaje recomendadas

Es muy importante que se detenga y realice esta actividad que se propone, hacerla constituye un refuerzo a esta unidad estudiada. El siguiente material audiovisual permite conocer de la mano de la periodista de datos, Mar Cabra, cuál es la importancia que tienen la visualización de datos a través de la aplicación de noticias, *apps* que le dan el poder al usuario para interactuar directamente con los datos masivos, además, se destaca la importancia de los profesionales o perfiles que programan este tipo de aplicaciones: informáticos de la mano de periodistas. Para profundizar en ello, le invito a observar el video "[News Apps periodísticas](#)" y realizar un breve resumen en 280 caracteres, tipo tuit.

Consulte a su profesor:

De esta forma terminamos la unidad. Si tiene alguna duda por favor comuníquese con su profesor a través del EVA.

A continuación, es importante desarrollar la autoevaluación, a fin de determinar su grado de comprensión de los contenidos analizados.

Toda unidad de su guía incluye preguntas de autoevaluación que le recomiendo realizar para reforzar sus conocimientos aprendidos.



Autoevaluación 4

Señale la respuesta correcta:

- 1. Uno de los productos periodísticos basado en datos son las aplicaciones de:**
 - a. Noticias.
 - b. Juegos.
 - c. Datos.

- 2. Uno de los productos periodísticos basado en datos es el conjunto de:**
 - a. Apps.
 - b. Artículos.
 - c. Datos.

- 3. Uno de los productos periodísticos basado en datos es:**
 - a. DataApp.
 - b. DataViz.
 - c. DataWeb.

- 4. Uno de los productos periodísticos basado en datos es:**
 - a. Artículos.
 - b. Aplicaciones.
 - c. Documentos.

- 5. Los artículos, en su extensión, pueden ser en su mayoría:**
 - a. Cortos.
 - b. Largos.
 - c. No hay límite.

- 6. Para Sandra Crucianelli un artículo para PD es aquel que nace de:**
 - a. "...un volumen corto datos y permite contar historias".
 - b. "...datos y permite contar historias".
 - c. "...un volumen considerable de datos y permite contar historias".

- 7. El medio que narra la historia en un artículo de datos y a la cual denominó “Más de 159.000 personas dejaron de tener un empleo adecuado en octubre”, fue:**
- a. La Nación.
 - b. Primicias.ec
 - c. El Comercio.
- 8. El medio que narra la historia en un artículo de datos y a la cual denominó: “Pandora Papers: una offshore suma misterio a la desaparición de US\$7 millones en la empresa de jugos Baggio”, fue:**
- a. La Nación.
 - b. Primicias.ec
 - c. El Comercio.
- 9. Para Sandra Crucianelli la visualización de datos puede:**
- a. “...estar complementada por artículos acompañados con un titular y una explicación para que los usuarios puedan analizarlos y leerlos.”
 - b. “...ir solo acompañados con un titular y una explicación para que los usuarios puedan analizarlos y leerlos.”
 - c. “...estar complementada por artículos cortos o de gran extensión y en algunos casos ir solo acompañados con un titular y una explicación para que los usuarios puedan analizarlos y leerlos.”
- 10. Los datasets son conjuntos de:**
- a. Datos: que nos permite acceder a la información para su procesamiento.
 - b. Apps: que nos permite acceder a la información para su procesamiento.
 - c. Dataviz: que nos permite acceder a la información para su procesamiento.

[Ir al solucionario](#)

Desarrolle las siguientes preguntas en su blog de notas:

- A. ¿Cuáles son los productos del periodismo de datos?**
- B. ¿Qué producto le parece más relevante?**
- C. Coloque 5 ejemplos de los distintos productos de datos que encuentra en la Red. Señale el medio de comunicación que los emplea.**

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.



Semana 5

Unidad 5. Transparencia y acceso a la información

Para iniciar esta Unidad 5. Transparencia y acceso a la información es importante mencionar que, a través de la Declaración de los Derechos Humanos proclamada por la Asamblea General de las Naciones Unidas en París, el 10 de diciembre de 1948 en su Resolución 217 A (III), como un ideal común para todos los pueblos y naciones; elaborada por representantes de todas las regiones del mundo con diferentes antecedentes jurídicos y culturales, se hace mención a la libertad de opinión y expresión.

5.1. Agenda de Desarrollo Sostenible

Figura 13.

Objetivos de Desarrollo Sostenible, ODS.



Nota. (Organización de las Naciones Unidas, 2015)

Ya se indica en uno de los artículos de la DDH: "Todo individuo tiene derecho a la libertad de opinión y de expresión; este derecho incluye el de no ser molestado a causa de sus opiniones, el de investigar y recibir informaciones y opiniones, y el de difundirlas, sin limitación de fronteras, por cualquier medio de expresión." (Naciones Unidas (ONU), 1948, Artículo 19)

Con este antecedente ¿Por qué es importante el acceso a la información? Porque le da al ciudadano la llave para conocer, ahondar y profundizar sobre hechos que son de importancia para sí mismo y su entorno.

"La Asamblea General de la Organización de Naciones Unidas en septiembre de 2015 adoptó la Agenda 2030 para el Desarrollo Sostenible, que consiste en un "plan de acción a favor de las personas, el planeta y la prosperidad, que también tiene la intención de fortalecer la paz universal y el acceso a la justicia"; la agenda que plantea 17 Objetivos de Desarrollo Sostenible (ODS) y 169 metas de

carácter integrado e indivisible en los sectores económico, social y ambiental." (Organización de las Naciones Unidas, 2015)

El ODS que nos compete en esta Unidad es el **Objetivo 16. Paz, justicia e instituciones sólidas**. Cuyo indicador es el *16.10 Garantizar el acceso público a la información y proteger las libertades fundamentales, de conformidad con las leyes nacionales y los acuerdos*.

En este marco, la Organización para la Cooperación y el Desarrollo Económico (OECD, p.9), cataloga datos como el requisito previo para cumplir la Agenda 2030 de las Naciones Unidas para el Desarrollo Sostenible:

"Los datos son un requisito previo para cumplir la Agenda 2030 para el Desarrollo Sostenible, asegurando que nadie se queda atrás. Sin embargo, simplemente producir más datos no es suficiente: los datos deben ser transformados, analizados y para ser útiles en la formulación de políticas, el seguimiento y la rendición de cuentas. (...) la calidad, los datos oportunos y desglosados que son cruciales para lograr los objetivos finales del desarrollo - mejorar el bienestar de las personas y combatir la pobreza. La inversión en sistemas estadísticos debe convertirse en una prioridad estratégica para los países en desarrollo y proveedores de cooperación para el desarrollo por igual. Un fuerte liderazgo político en los países en desarrollo es necesario para promover la causa de los datos para el desarrollo y garantizar que los datos se produzcan con alta calidad, normas, protegiendo la privacidad y confidencialidad. La cooperación al desarrollo puede ayudar a los países en desarrollo a producir y usar más y mejores datos de una manera responsable y transparente para obtener buenos resultados políticos." (CEPAL, 2019, p.9)

Las leyes de transparencia a nivel mundial se alinean tanto con el artículo 19 de la Declaración Universal de los Derechos Humanos como con el ODS 16.10, en América Latina se ha buscado participar de leyes y sistemas que garanticen el acceso a la información, y "aunque los detalles técnicos de una ley de acceso a la información son sumamente importantes, lo que importa sobre todo es la implementación e impacto social de la ley, la cual depende de su uso cotidiano por la prensa, la sociedad civil, y los ciudadanos comunes" (Orme, 2017, p. 8) (CEPAL, 2019)

5.2. Protección de datos personales

La protección de datos debe ser considerado como un derecho de los usuarios y que debe ser cumplido por las organizaciones, por ello, como introducción al tema, le invito a mirar el video: [¿Qué es GDPR y cómo nos afecta?](#), en el cual se comenta sobre la *General Data Protection Regulation* (GDPR) aplicada en la Unión Europea a partir del 25 de mayo de 2018, donde el usuario debe dar su consentimiento inequívoco para que las empresas puedan usar datos.

5.2.1. Principios

Como se tiene derecho al acceso a la información también se posee derecho a la privacidad, por ello se debe considerar que los datos que se están exponiendo, incluso, los que son de nuestra propiedad tengan una protección. En la actualidad, como es conocido, muchos usuarios han sido contactados por empresas particulares para adquirir algún bien o servicio; cuyos datos (nombre, cédula, teléfono, email, etc.) no han sido entregados particularmente por los ciudadanos, sino que ha sido otorgados a través de terceros: por una venta de bases de datos.

A continuación se detallan los Principios sobre la Protección de la Privacidad y los Datos Personales que la OEA (2019) establece:

1. **Propósitos legítimos y justos:** los datos personales deben ser recopilados solamente para fines legítimos y por medios justos y legales.
2. **Claridad y consentimiento:** se deben especificar los fines para los cuales se recopilan los datos personales en el momento de proceder a su recopilación. Como regla general, los datos personales solamente deben ser recopilados con el consentimiento de la persona a que se refieran.
3. **Pertinencia y necesidad:** los datos deben ser verídicos, pertinentes y necesarios para los fines expresos de su recopilación.
4. **Uso limitado y retención:** los datos personales deben ser mantenidos y utilizados solamente de manera legítima no incompatible con el fin o fines para los cuales se recopilaron. No deberán mantenerse más

del tiempo necesario para su propósito o propósitos y de conformidad con la legislación nacional correspondiente.

5. **Deber de confidencialidad:** los datos personales no deben divulgarse, ponerse a disposición de terceros ni emplearse para otros propósitos que no sean aquellos para los cuales se obtuvieron, excepto con el conocimiento o consentimiento de la persona en cuestión o bajo la habilitación de la ley.
6. **Protección y seguridad:** los datos personales deben ser protegidos mediante salvaguardias razonables y adecuadas contra accesos no autorizados, pérdida, destrucción, uso, modificación o divulgación.
7. **Fidelidad de los datos:** los datos personales deben mantenerse fieles y actualizados hasta donde sea necesario para los propósitos de su uso.
8. **Acceso y corrección:** Se deben establecer métodos razonables para permitir que aquellas personas cuyos datos personales han sido recopilados puedan solicitar al controlador de datos que los modifique, corrija o elimine. En caso de disponerse la restricción a dicho acceso o corrección deberán fundarse los motivos de cualquiera de estas restricciones de conformidad a la legislación nacional.
9. **Datos personales sensibles:** Algunos tipos de datos personales, teniendo en cuenta su sensibilidad en contextos particulares, son especialmente susceptibles de causar daños considerables a las personas si se hace mal uso de ellos. Los controladores de datos deberían adoptar medidas de privacidad y de seguridad que sean acordes con la sensibilidad de los datos y su capacidad de hacer daño a los individuos sujetos de la información.
10. **Responsabilidad:** Los controladores de datos adoptarán e implementarán las medidas correspondientes para el cumplimiento de estos principios.
11. **Flujo transfronterizo de datos y responsabilidad:** Los Estados Miembros cooperarán entre sí en la creación de mecanismos y procedimientos que aseguren que los controladores de datos que operen en más de una jurisdicción puedan ser efectivamente responsabilizados por el incumplimiento de estos principios.

12. **Publicidad de las excepciones:** Cuando las autoridades nacionales establezcan excepciones a estos principios por motivos relacionados con la soberanía nacional, la seguridad interna o externa, el combate a la criminalidad, el cumplimiento de normativas u otras prerrogativas de orden público, deben poner en conocimiento del público dichas excepciones.
13. **Anonimización:** entendiendo como la aplicación de medidas de cualquier naturaleza dirigidas a impedir la identificación o reidentificación de una persona física.

Al tratarse de una Declaración de la OEA, se colocan los 13 puntos de forma íntegra y tal como corresponde.

¡Lectura recomendada!

Para profundizar más en el tema de la protección de datos lo invito a revisar el informe presentado por Carlos Mata Prates, representante de la OEA, a través del siguiente enlace: http://www.oas.org/es/sla/cji/docs/CJI_doc_582-19.pdf, le cual detalla la importancia en la relación a las normas de los derechos internos de los diferentes Estados. En él se señala que en algunos países el Derecho a la Protección de Datos Personales se encuentra recogido en la norma de mayor eficacia formal, es decir la Constitución, de manera expresa o derivada de la interpretación evolutiva de la misma y, en otros, se establece el mismo mediante normas legislativas. A su vez, cabe destacar que en la mayoría de los países latinoamericanos el dictado de leyes sobre el Derecho a la Protección de Datos Personales fue realizadas en la última década.

¡Video recomendado!

¿Qué sucede en Europa con relación a Latinoamérica con la regulación de datos personales? Veámoslo en el siguiente video: "Reglamento Europeo de Protección de Datos": <https://youtu.be/heKapvVLjng>. En este vídeo se explica qué es el Reglamento Europeo de Protección de Datos y porqué es importante que las empresas revisen y adecúen sus procedimientos para recopilar y procesar datos personales de residentes en la Unión Europea.

5.3. Ecuador: acceso a la información

Ahora bien, miremos qué sucede con el acceso a la información en nuestro país, Ecuador. A pesar de las iniciativas entre todos los Estados a través de los ODS, no podemos afirmar que existe una corriente a favor de la transparencia completa de los organismos públicos.

Recordemos que el numeral 12 del artículo 66 de la Constitución de la República del Ecuador, establece: "Se reconoce y garantizará a las personas: (...) 19. El derecho a la protección de datos de carácter personal, que incluye el acceso y la decisión sobre información y datos de este carácter, así como su correspondiente protección. La recolección, archivo, procesamiento, distribución o difusión de estos datos o información requerirán la autorización del titular o el mandato de la ley". (Constitución de la República del Ecuador, 2008).

5.3.1. Ley Orgánica de Protección de Datos Personales

En Ecuador, el 10 de mayo de 2021 a través de la Asamblea Nacional aprobó con 118 votos y 1 abstención el proyecto de Ley de Protección de Datos Personales.

Figura 14.

Tuit del expresidente de Ecuador, Lenín Moreno anunciando la aprobación de la Ley.

The screenshot shows a tweet from Lenín Moreno (@Lenin). The tweet reads: "Otra buena noticia: la Asamblea Nacional, por unanimidad (118 de 119 asambleístas), votó a favor de la Ley de Protección de Datos Personales que presentamos con @caanmichelena en septiembre 2019. Este cuerpo legal blindará más y mejor la información privada de los ecuatorianos." The tweet was posted at 2:48 p.m. on May 10, 2021. It has 325 likes and 272 retweets. There are options to copy the link or view more details.

Nota. Twitter- Lenín Moreno [https://twitter.com/Lenin/
status/1391842525489606659?s=20](https://twitter.com/Lenin/status/1391842525489606659?s=20)

A decir de la Dirección Nacional de Registro Público (Dinarp) (2021), "con la Ley de Protección de Datos Personales, se busca cuidar a las personas

titulares de los datos, para que ellas puedan decidir a quién entregar su información personal porque confían en los proveedores de servicios digitales."

Para comprender el contexto de la ley es importante dar un vistazo a los principales puntos:

Ley de Protección de Datos Personales

1. Protección de derechos

A través de la infografía que muestra la Figura 15 se detalla la importancia de la protección de las personas y garantizando sus derechos jurídicos.

Figura 15.

Cuadro que define la protección de derechos.



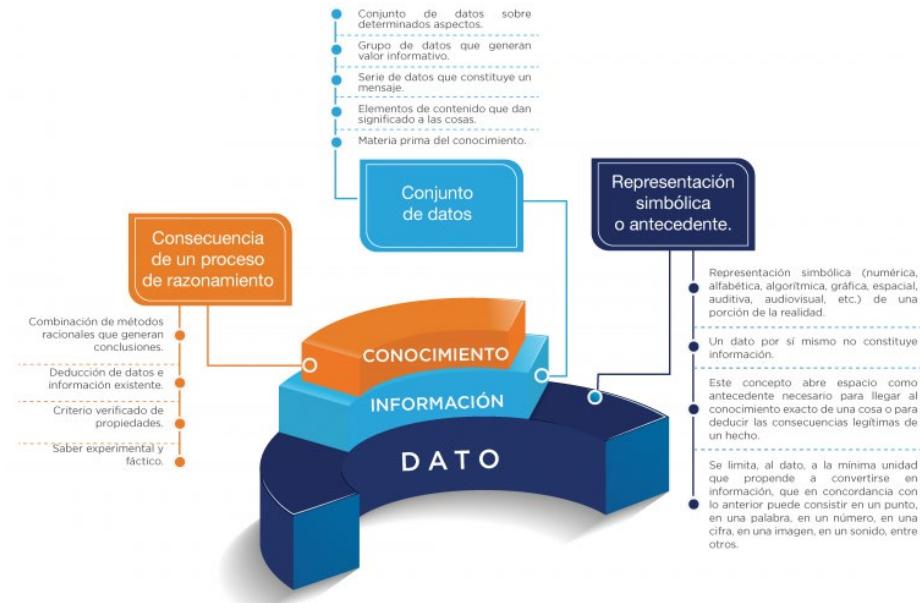
Nota. Dirección Nacional de Registro Público: <https://www.dinardap.gob.ec/protege-a-las-personas-y-garantiza-derechos/>

2. Dato

Se observa la diferencia entre: dato, información y conocimiento, para conocer detalles acceder al enlace y descargar la infografía de la Figura 16.

Figura 16.

Cuadro de dato.



Nota. Dirección Nacional de Registro Público: <https://www.dinardap.gob.ec/dato/>

3. Dato personal

Elementos de contenido que permiten identificar o hacer identifiable a una persona natural, tales como: dato inocuo (forma de vestir) y metadato (última hora de conexión).

Figura 17.
Cuadro de dato personal



Nota. Dirección Nacional de Registro Público: <https://www.dinardap.gob.ec/datos-personales/>

4. Dato personal registrable

Este apartado se divide entre dato personal registrable y dato de información pública. El primero trata de aquellos datos personales que por disposición de la ley se encuentran en registros públicos, y, el segundo, se enfoca en representaciones simbólicas, antecedentes o elementos de contenido creados u obtenidos por instituciones públicas o con recursos del Estado.

Figura 18.

Cuadro de dato personal registrable.



Nota. Dirección Nacional de Registro Público: <https://www.dinardap.gob.ec/dato-publico/>

5. Tratamiento de datos

La siguiente infografía muestra cómo debe enfocarse el tratamiento de datos, iniciando desde la operación del dato, su ubicación, extracción, procesamiento, valoración hasta llegar a la transferencia de los datos personales.

Figura 19.

Cuadro del tratamiento de datos.



Nota. Dirección Nacional de Registro Público: <https://www.dinardap.gob.ec/tratamiento/>

¡Lectura recomendada!

Si desea profundizar en el tema de protección de datos y sobre todo, como parte de su acervo cultural y conocimiento jurídico para su profesión periodística, comparto el enlace con la Ley de Protección de Datos Personales de Ecuador, Resolución Nº. 009-NG-DINARDAP-2021: https://www.dinardap.gob.ec/wp-content/uploads/downloads/2021/05/resolucion_No_009-ng-dinardap-2021-signed1.pdf, y/o el archivo final que consta en el Registro oficial: <https://www.registrooficial.gob.ec/index.php/registro-oficial-web/publicaciones/suplementos/item/14857-quinto-suplemento-al-registro-oficial-no-459>. Cabe aclarar que es un documento extenso dado que se trata de una ley que cuenta con 18 artículos y disposiciones generales, además de anexos.

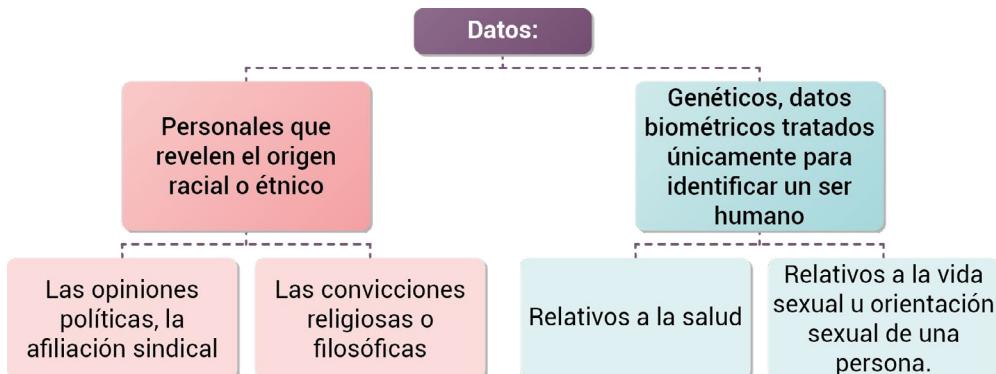
5.4. Datos sensibles

Entendemos por "datos sensibles", aquellos datos que deben ser cuidados por las empresas, tales como: datos personales (nombres, cédula, teléfono,

dirección), datos de salud, género, etnia, religión, etc. De forma la Comisión Europea (2021) menciona siguientes:

Figura 20.

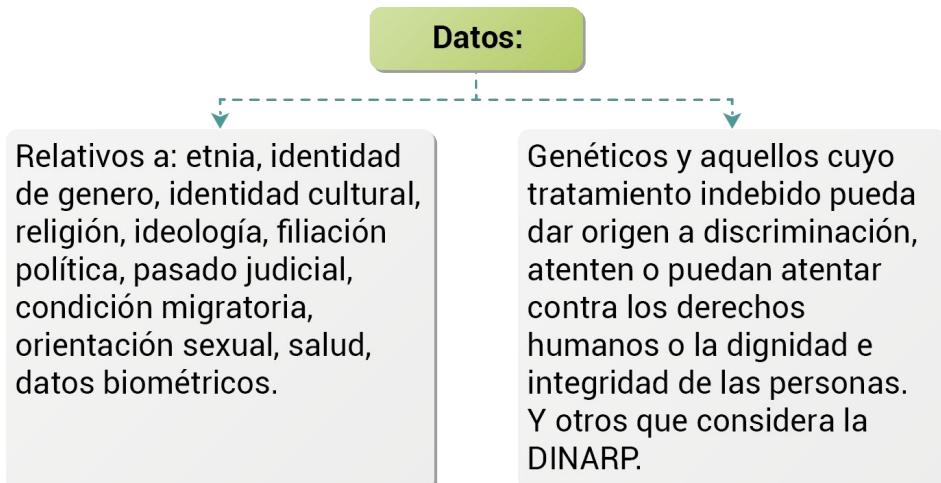
Cuadro que representa los datos sensibles.



Nota. (Comisión Europea, 2021)

Figura 21.

Cuadro que representa los datos sensibles de la Ley de Protección de Datos, Ecuador.



Nota. (Dirección Nacional de Registros Públicos (Dinarp), 2021)



Actividades de aprendizaje recomendadas

¡Video recomendado!

Para conocer la realidad en cuanto al acceso de la información de las organizaciones del Gobierno de Ecuador, le invito a observar denominado "[En Ecuador la Ley de Transparencia y Acceso a la Información Pública es invisible](#)", producido por Fundamedios, el cual se centra en cuán complicado y difícil es acceder a la información pública, de 11 instituciones consultadas solo 3 proporcionaron información requerida por la organización, esto puede deberse, según el Defensor Público, Harold Burbano, porque en Ecuador existe una "cultura" de ocultar información. Observe el video y realice un breve resumen en 280 caracteres, tipo tuit. Recuerde, es muy importante que se detenga y realice esta actividad que le proponemos, hacerla constituye un refuerzo a esta unidad estudiada.

Consulte a su profesor:

De esta forma terminamos la unidad. Si tiene alguna duda por favor comuníquese con su profesor a través del EVA.

A continuación, es importante desarrollar la autoevaluación, a fin de determinar su grado de comprensión de los contenidos analizados.

Toda unidad de su guía incluye preguntas de autoevaluación que le recomiendo realizar para reforzar sus conocimientos aprendidos.



Autoevaluación 5

Señale la respuesta correcta:

- 1. Para hablar de acceso a la información debemos recurrir primer a:**
 - a. La Declaración de los Derechos Humanos.
 - b. Aplicaciones de juegos.
 - c. Aplicaciones de datos.
- 2. La Agenda de Desarrollo Sostenible es un plan de:**
 - a. Trabajo al que deben someterse todos los países.
 - b. Acción a favor de las personas, el planeta y la prosperidad.
 - c. Trabajo para las organizaciones educativas.
- 3. Qué ODS garantizar el acceso público a la información:**
 - a. Objetivo 16. Paz, justicia e instituciones sólidas y transparentes.
 - b. Objetivo 17. Paz, justicia e instituciones transparentes.
 - c. Objetivo 18. Paz, justicia e instituciones sólidas.
- 4. El indicador del ODS que a garantizar el acceso público a la información es:**
 - a. 16.10
 - b. 16.11
 - c. 16.12
- 5. Qué garantiza el Objetivo 16:**
 - a. "Acceso público a la información y proteger las libertades fundamentales, de conformidad con las leyes nacionales y los acuerdo."
 - b. "Acceso público a la información."
 - c. "Proteger las libertades fundamentales, de conformidad con las leyes nacionales y los acuerdo."

6. Las leyes de transparencia a nivel mundial se alinean:

- a. Con el artículo 29 de la DDH y el ODS 16.10.
- b. Con el artículo 19 de la DDH y el ODS 16.11.
- c. Con el artículo 39 de la DDH y el ODS 16.12.

7. Entre los principios sobre la Protección de la Privacidad y los Datos Personales que la OEA establece constan:

- a. Datos limpios.
- b. Fidelidad de los datos.
- c. Confidencialidad de registros.

8. Se entiende por: datos personales sensibles:

- a. A aquellos susceptibles de causar daños considerables a las personas si se hace mal uso de ellos.
- b. A aquellos que pueden ser empleados por cualquier persona sin restricciones.
- c. A aquellos que pueden ser empleados por cualquier organización y empresa de tecnología.

9. En Ecuador, la Ley Orgánica de Protección de Datos Personales se aprueba el:

- a. 10 de marzo de 2020.
- b. 10 de mayo de 2021.
- c. 10 de mayo de 2022.

- 10. Uno de los puntos a destacar en la Ley Orgánica de Protección de Datos Personales de Ecuador se menciona:**
- a. Proteger y almacenar los datos de todos los ciudadanos del país.
 - b. Otorgar protección a las empresas para el correcto uso datos de los usuarios.
 - c. Garantizar la preservación de la dignidad e integridad frente al tratamiento de datos personales.

[Ir al solucionario](#)

Desarrolle las siguientes preguntas en su blog de notas:

- A. ¿Qué entiende por “protección de datos personales?**
- B. Señale los 10 puntos más importantes de la Ley de Protección de Datos de Ecuador.**
- C. Para usted ¿qué es un dato sensible?**

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.



Semana 6

Unidad 6. Open Data

Estimado estudiante, en la Unidad 6. Open data, se tratará sobre la tendencia *open data* o datos abiertos que va de la mano con lo que hemos visto en la Unidad 5. Transparencia y acceso a la información. En este sentido citaré las dudas que nos plantea la web de: Open Data Handbook: "¿Sabemos cuánto del dinero de nuestros impuestos es destinado en iluminación de las calles o en estudios sobre el cáncer? ¿Cuál es el camino más corto, seguro y pintoresco para ir en bicicleta de casa al trabajo? ¿Dónde podemos encontrar la mejor oportunidad laboral y el número más alto de árboles nativos per cápita? ¿Cuándo puedes influir en decisiones sobre tópicos que te preocupan profundamente, y con quién debes hablarlo?" (2021), estas preguntas y más nos vienen a la cabeza diariamente y es ahí dónde los datos abiertos toman sentido.

6.1. ¿Qué es Open - Abierto?

La definición de "Open" engloba todos aquellos datos, información y en general, contenido que tiene un carácter de "abierto". Los *datos abiertos* son los componentes básicos del conocimiento. El conocimiento abierto es en lo que se convierten los datos abiertos cuando son útiles, utilizables y utilizados por los usuarios.

Ahora bien, miremos la repercusión que tienen los datos en el ámbito gubernamental.

Especificamente los datos que competen a los gobiernos tales como información pública, a la que cualquier persona es libre de acceder y utilizar para cualquier propósito, ha estado presente por varios años. Desde 2009 países como Estados Unidos, Canadá, Gran Bretaña y Nueva Zelanda, anunciaron su iniciativa hacia la apertura de su información de índole pública. (Open Data HandBook, 2021)

Existen áreas de los llamados "gobiernos abiertos" en las cuales se están creando un valor agregado, por ejemplo:

- Participación ciudadana.
- Innovación.
- Transparencia y control democrático.
- Mejora o creación de nuevos productos y servicios.
- Mejora en la eficiencia de los servicios ofrecidos por el gobierno.
- Mejora en la eficacia de los servicios ofrecidos por el gobierno.
- Medición del impacto de políticas públicas.
- Nuevos conocimientos a partir de fuentes de datos combinadas y patrones en grandes volúmenes de datos.
- Entre otros.

¡Video recomendado!

Estimado estudiante, le recomiendo observar el siguiente material audiovisual el cual le permitirá descubrir brevemente [¿Qué es Open data?](#) y como su filosofía de acceso a la información es sinónimo de transparencia para las administraciones públicas, lo que le permite al ciudadano saber cómo se está gestionando y administrando un gobierno local o nacional. Encontrará ejemplos de Suecia, EE. UU y España.

A continuación, se presentan algunas características del open data:

6.1.1. Características clave del open access

Características del Open data.

Entre las características que encontramos para el acceso abierto de datos están:



Disponibilidad y acceso: los datos deben estar disponibles en su totalidad y a un costo de reproducción razonable, preferiblemente mediante descarga a través de Internet; y, estar disponibles en una forma conveniente y modificable.



Reutilización y redistribución: los datos deben proporcionarse en condiciones que permitan la reutilización y la redistribución, incluida la mezcla con otros conjuntos de datos.



Participación universal: todos debemos poder usar, reutilizar y redistribuir la información; no debe existir discriminación contra personas o grupos. Por ejemplo, no se permiten restricciones "no comerciales" que impidan el uso "comercial" o restricciones de uso para determinados fines, por ejemplo, sólo para uso exclusivo y de índole educativo.

Estimado estudiante, existen características y tipo de datos abiertos que permiten dimensionar la importancia que tiene este tema, es así que para continuar con la planificación, lo invito a revisar la web: <https://opendatahandbook.org/>, la cual es fuente primaria en esta unidad, proveyéndonos de información sobre: guías, estudios de casos y recursos para gobiernos y la sociedad civil sobre el «qué, por qué y cómo» de los datos abiertos.

6.1.2. Tipos de datos abiertos

Existen tipos de datos abiertos que tienen usos y aplicaciones potenciales, tales como:

- **Cultura:** datos sobre obras artísticas: títulos y autores; galerías, bibliotecas, archivos y museos.
- **Ciencia:** datos que se generan como parte de la investigación científica desde la astronomía hasta la zoología.
- **Finanzas:** datos como cuentas gubernamentales (gastos e ingresos) e información sobre los mercados financieros (acciones, bonos, bitcoins, etc.).
- **Estadísticas:** datos producidos por oficinas de estadística como el censo poblacional, los indicadores socioeconómicos y de empleo.
- **Clima:** información que se utiliza para comprender y predecir el tiempo y el clima.
- **Medio ambiente:** Información relacionada con el medio ambiente como el nivel de contaminación, la calidad de los ríos y mares, entre otros.

6.1.3. Crear datos abiertos



Elegir el dato o conjunto de datos, escoger un conjunto de datos que deseamos hacer público.



Utilizar una licencia abierta. 1) Establecer los derechos de propiedad intelectual. 2) Emplear licencia abierta, como por ejemplo, *Creative Commons*. Para conocer más sobre este tipo de licencias puede visitar: <https://creativecommons.org/>



Hacer disponible la información, en conjunto de datos y en un formato útil, tales como: *excel xlsx, html, xml, csv, rss feed, json*, etc. Porque si los tenemos en imágenes nos demoraremos en su procesamiento.

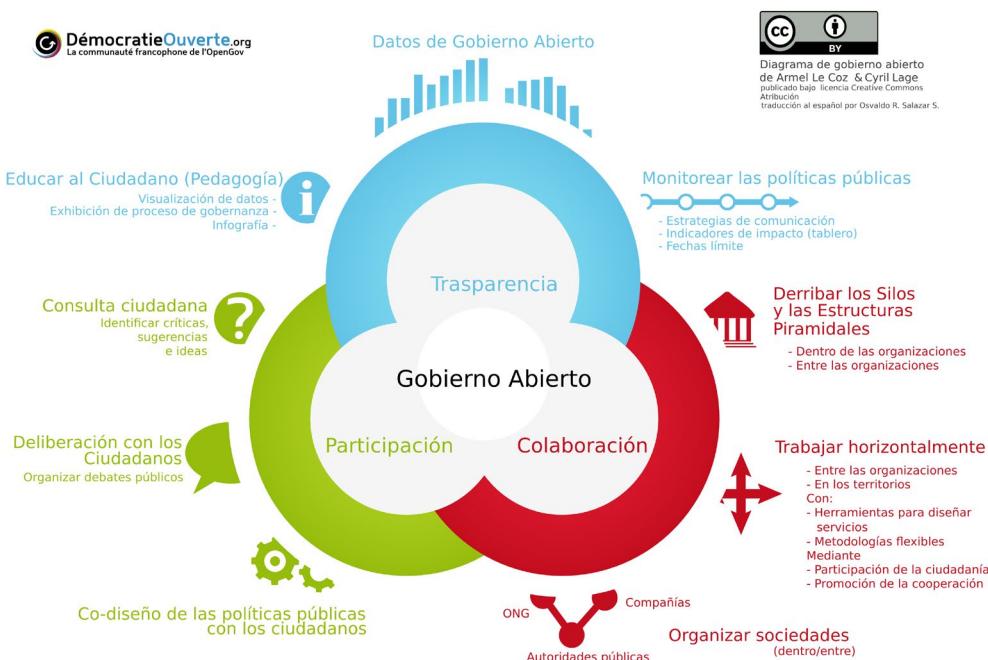


Hacerla visible, publicarla en un sitio web y organizar un catálogo que enliste los datos abiertos.

Miremos a continuación la infografía (Democratie Ouverte, 2019) que permite visualizar la trilogía de los datos abiertos: transparencia, participación y colaboración.

Figura 22.

Infografía que representa los datos de gobierno abierto.



Nota. Wikipedia: <https://www.democratieouverte.org/>

Aquí se presenta un ejemplo de cómo el índice de datos abiertos es empleado por el gobierno de Guatemala.

Figura 23.

¿Qué es el índice de datos abiertos?

El índice de Datos Abiertos basa su calificación en las respuestas obtenidas en una batería de preguntas que incluye cuestiones sobre el formato, el nivel de apertura, la actualización o periodicidad de publicación de los datos. La puntuación de cada una de ellas se suma para crear un ranking final donde se visualiza, además, el nivel de apertura de los conjuntos de datos (datasets) a nivel nacional.



#DatosAbiertosGT



Nota. Red Ciudadana: <https://redciudadana.org/presentacion-del-indice-de-datos-abiertos-2019>

6.1.4. ¿Qué sucede en Ecuador con los datos abiertos?

Como hemos visto en la Unidad 5. Transparencia y acceso a la información, la transparencia y las leyes que se aplican a su alrededor le permiten a los ciudadanos acceder a la información, sin embargo esto no es una constante. Para conocer qué sucede en Ecuador le invito a observar el video enfocado en "gobierno abierto": "[CONOCE MÁS DE GOBIERNO ABIERTO I Programa 13: Elaboración de la política y guía de datos abiertos](#)", el cual detalla elaboración de la política y guía de datos abiertos en Ecuador a través de sus voceros oficiales y la importancia de la anonimización de los datos, a la par, cuentan con más de 100 datasets en su portal.

Le invito a que juntos descubramos cuál es el índice de datos abiertos que tiene nuestro país. El índice es un “listado de calificación que presenta el estado del uso y adopción de datos abiertos en los 100 municipios de cantones con mayor población en el país.” (Datos Abiertos Ecuador y Fundación de Ayuda por Internet (FUNDAPI), 2021).

A decir FUNDAPI (2021), para realizar la revisión de los datos abiertos, seleccionaron los 100 cantones con mayor población en Ecuador y si cumplen con las siguientes características básicas:

Figura 24.

Características de los datos abiertos en Ecuador.



Nota. FUNDAPI (2021)

Entre los indicadores que se emplearon para determinar si un gobierno local es abierto o no, se emplearon los siguientes:

Figura 25.

Indicadores.

Funcionarios públicos	Lista de funcionarios que laboran en la institución.
Ordenanzas	Documentos con las ordenanzas.
Ejecución presupuestaria	Célula de liquidación mensual del presupuesto anual de la entidad.
Presupuesto	Presupuesto anual de la entidad actualizado.
Contrataciones	Contrataciones realizadas en el sistema de contratación pública.
Lugares públicos	Ubicación georeferenciada de parques, plazas, oficinas públicas y sitios para atención a la ciudadanía.
Límites administrativos	Mapa georeferenciado de los límites cantonales.
Transporte público	Georeferenciación de las rutas de transporte y horarios de buses urbanos.
Indicadores sociales	Información sobre proyectos sociales, beneficiarios, indicadores de impacto que lleva la entidad.
Medio ambiente	Datos medio ambientales del cantón, calidad de aire, agua, ruido, áreas protegidas, inventario verde, entre otros.

Nota. Datos Abiertos Ecuador y Fundación de Ayuda por Internet, FUNDAPI, (2021).

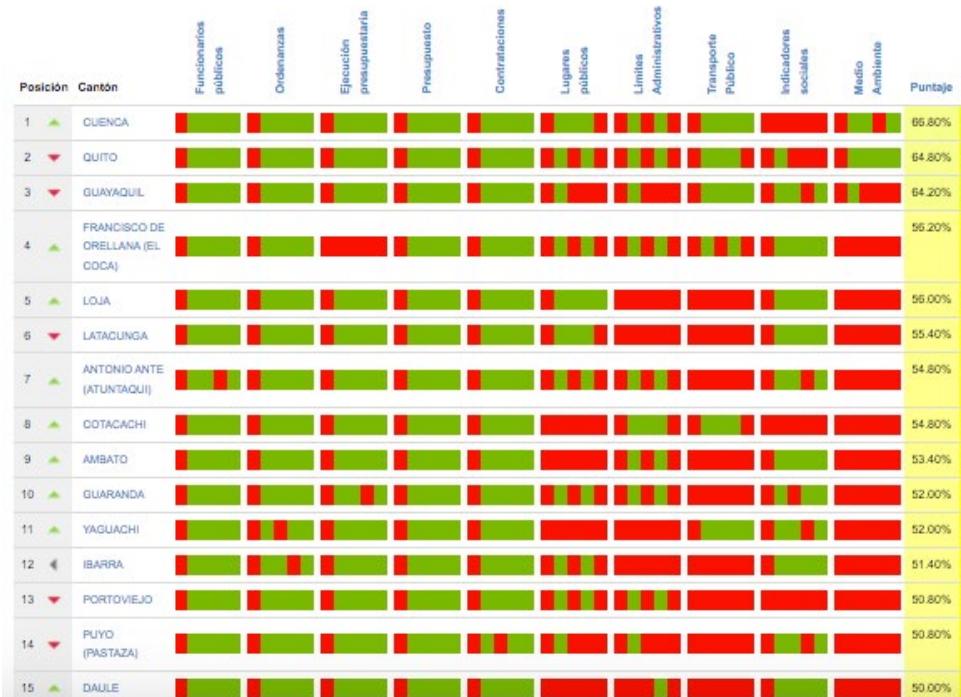
Figura 26.

Índice de open data, Ecuador.



El Índice de Datos Abiertos de **Ciudades** de Ecuador es un listado de calificación elaborado por Fundapi, que presenta el estado del uso y adopción de datos abiertos en los 100 municipios de cantones con mayor población en el país. Ver más sobre la metodología utilizada.

Cantones: **100** Datasets revisados: **1000** Datasets abiertos: **0** Promedio de **apertura** de entidades: **34.86%**



Nota. Datos Abiertos, <https://datosabiertos.ec/>

Para conocer la metodología en su totalidad que la FUNDAPI aplica, puede visitar el enlace: <https://datosabiertos.ec/acerca.php>, sitio que tiene el fin de contar con una línea base relacionada a los datos abiertos en los municipios y prefecturas de Ecuador y además motivar a estas entidades a fortalecer su uso y adopción de estadísticas, indicadores, mapas, entre otros, para analizar o resolver problemáticas identificadas. Si de su parte existe más interés en este tema, puede seguir las cuentas oficiales de Twitter:

- Gobierno Abierto de Ecuador: <https://twitter.com/ECGobAbierto>
- Datos Abiertos: <https://twitter.com/planificacionec/>
- FUNDAPI: <https://twitter.com/fundapi>

Consulte a su profesor:

De esta forma terminamos la Unidad 6. Si tiene alguna duda por favor comuníquese con su profesor a través del EVA.

A continuación, es importante desarrollar la autoevaluación, a fin de determinar su grado de comprensión de los contenidos analizados.

Toda unidad de su guía incluye preguntas de autoevaluación que le recomiendo realizar para reforzar sus conocimientos aprendidos.



Autoevaluación 6

Señale la respuesta correcta:

1. Open Data se relaciona con datos abiertos para fines:

- a. Políticos.
- b. Ciudadanos.
- c. Económicos.

2. Los datos abiertos se convierten en conocimiento cuando son:

- a. Útiles, utilizables y utilizados.
- b. Útiles, aplicables y usables.
- c. Útiles, utilizados y aplicables.

3. Los primeros gobiernos que decidieron apostar por open data fueron:

- a. Estados Unidos, Colombia, Gran Bretaña y Suiza.
- b. Estados Unidos, Canadá, Gran Bretaña y Nueva Escocia.
- c. Estados Unidos, Canadá, Gran Bretaña y Nueva Zelanda.

4. Existen áreas de los “gobiernos abiertos” en las cuales se están creando un valor agregado, por ejemplo:

- a. Participación ciudadana.
- b. Participación política.
- c. Participación digital.

5. Una de las características del open access es:

- a. Medición del impacto político.
- b. Reutilización y redistribución.
- c. Creación de productos.

6. Una de las características del open access es:

- a. Disponibilidad y acceso.
- b. Medición del impacto político.
- c. Innovación.

7. Una de las características del open access es:

- a. Participación ciudadana
- b. Participación universal.
- c. Eficiencia en los servicios.

8. Entre los tipos de datos abiertos, consta:

- a. Tecnología.
- b. Cultura.
- c. Educación.

9. Para crear datos abiertos se requiere:

- a. Elegir el conjunto de datos.
- b. Pedir información a una entidad.
- c. Procesar los datos.

10. Entre los formatos útiles para el procesamiento de datos están:

- a. excel xlsx, html, xml, csv, rss feed, jpg...
- b. excel xlsx, html, png, csv, rss feed, json...
- c. excel xlsx, html, xml, csv, rss feed, json...

[Ir al solucionario](#)



Unidad 7. Big data

Para continuar el proceso de enseñanza y aprendizaje a través de la Unidad 7. Big data se hará un repaso por uno de los temas que más hemos escuchado hablar en el último tiempo: *big data* y su poder para la toma de decisiones. Por ello, en esta unidad se compartirá parte del capítulo del libro sobre "Big data, augmented data y computación cognitiva en la era del millenial", que he publicado con mi colega, Félix Ortega de la Universidad de Salamanca. Espero que sea de utilidad y provecho para su estudio: <https://dialnet.unirioja.es/servlet/articulo?codigo=7115934>

Para usted **¿qué son los datos del nuevo petróleo?** Le invito a que lo descubramos juntos.

Figura 27.
Portada de la revista Time.



Nota. Time, <http://content.time.com/time/covers/0,16641,20110321,00.html>

Le invito a revisar, *big data* y su poder para la toma de decisiones:

7.1. El poder de los datos

En principio y cómo concepto básico, que nos dice Wikipedia (2021) sobre los macrodatos: "hace referencia a conjuntos de datos grandes y complejos que precisan de aplicaciones informáticas no tradicionales de procesamiento de datos para tratarlos adecuadamente", Además, "tiende a referirse al análisis del comportamiento del usuario, extrayendo valor de los datos almacenados, y formulando predicciones a través de los patrones observados."

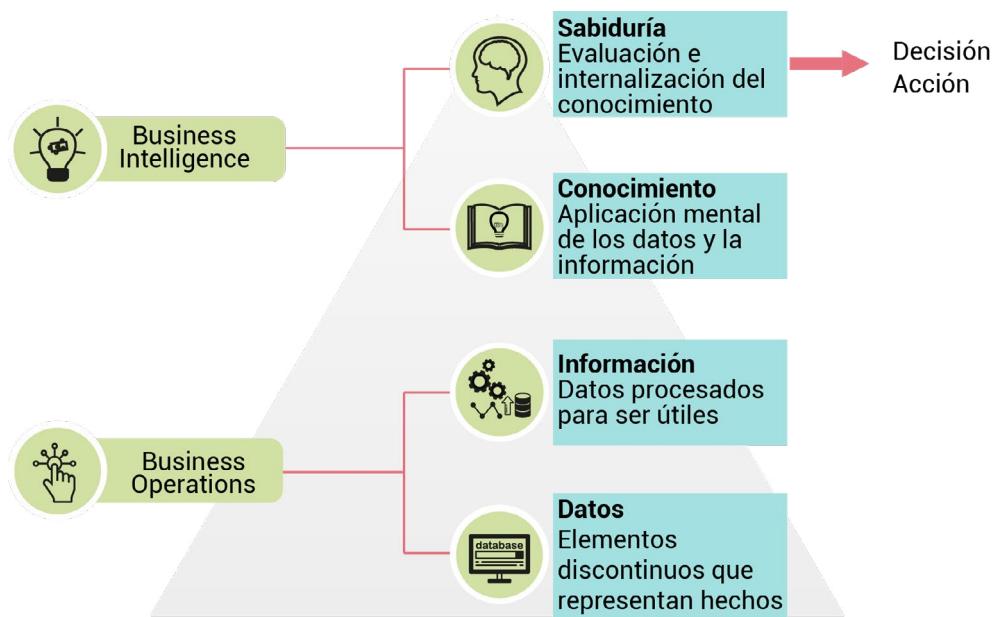
Big data es definido tradicionalmente como el conjunto de "agrupaciones de datos" de un tamaño tan voluminoso y complejo que no pueden ser procesados por el software tradicional de análisis estándar. Los retos que afronta la computación de estos datos masivos se asocian a la captura de la información, su almacenamiento, el análisis, la búsqueda y consulta, la transferencia, la visualización y compartición, la actualización, la privacidad de la información y las fuentes de origen de la información, entre otros.

(Ortega Mohedano y Coronel-Salas, 2019, p. 825)

De los datos a la sabiduría.

Figura 28.

Pirámide de la información “Data, Information, Knowledge, Wisdom (DIKW)”.



Nota. The Metaphorical Link.

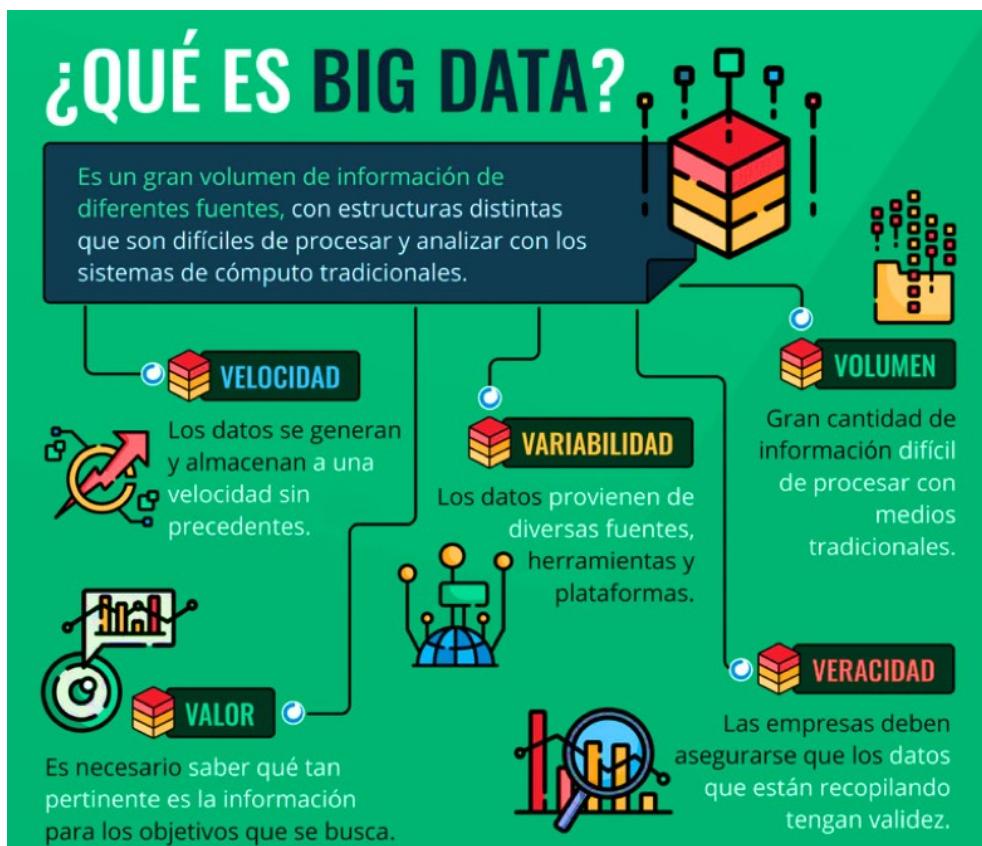
Con el fin de llegar a la etapa de la sabiduría, las compañías deben aprovechar las oportunidades de una estrategia de grandes volúmenes de datos, invertir en tecnologías y formación, con el fin de aprovechar la inversión y utilizar esta “nueva sabiduría” (obtenida a través del proceso de recolección, depuración y visualización de datos) para mejorar los procesos, tales como servicio al cliente y productos.

Existen aplicativos y plataformas en línea que permite a las organización saber cuántos usuarios ingresar a ver sus perfiles y también, cuando deciden no contratar sus bienes o servicios e indicándoles quien sí lo hace, permitiéndoles de esta forma, identificar a su competencia y mejorar la atención al cliente.

El costo de implementar *big data* en un negocio puede no estar en los planes de sus directivos, pero se resalta que grandes empresas suelen recopilar información de forma general y que pueden ser tomados como referencia para una toma de decisiones, incluso suelen ser, como hemos visto en unidades anteriores, *open access*: visualizar, descargar los datos, cruzarlos, usarlos e incluso publicarlos.

Figura 29.

¿Qué es Big data? Fuente: EDteam.



Nota.

Todos los objetos capaces de soportar un chip, una antena y un sensor estarán conectados entre sí y con Internet y serán más inteligentes. Podremos comunicarnos con ellos. Dialogar. Recibir la información que generan y darles órdenes para que, en teoría, nuestro día a día sea cada vez más fácil. (...) Al fenómeno se lo conoce como "Internet de las cosas" (en inglés, IOT), o "Internet de todo", "Machine to Machine" (M2M) o "Device to Device" (D2D). (Zanoni, 2014, pp. 18-19, citado en Ortega Mohedano y Coronel-Salas, 2019)

La premisa central de *big data* es la de tomar decisiones adecuadas y acertadas, y a su vez, permiten tener un rol más activo en la sociedad. Por ello, le invito a revisar el siguiente recursos (REA) sobre *big data*:

[https://socialsci.libretexts.org/Bookshelves/Communication/Journalism_and_Mass_Communications/Book%3A_Information_Strategies_for_Communicators_\(Hansen_and_Paul\)/14%3A_Information_Resources-_](https://socialsci.libretexts.org/Bookshelves/Communication/Journalism_and_Mass_Communications/Book%3A_Information_Strategies_for_Communicators_(Hansen_and_Paul)/14%3A_Information_Resources-_)

[Data_and_Statistics/14.10%3A_Big_Data](#), el cual evidencia lo difícil que es hablar sobre el uso de datos por parte de los comunicadores sin escuchar rápidamente el término «big data». Para ello introduce el tema con una definición: «*Big data* generalmente incluye conjuntos de datos con tamaños más allá de la capacidad de las herramientas de software de uso común para capturar, seleccionar, administrar y procesar los datos dentro de un tiempo transcurrido tolerable...»

7.2. . Las Uves (V) de Big data

Los principios básicos asociados a la calidad final de un correcto análisis de *big data* se componen de las seis "uves":

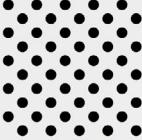
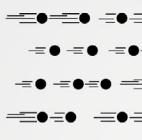
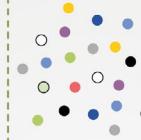
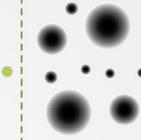
1. **Volumen** "big = grande" de información para ser considerada como tal;
 2. **Variedad** de la naturaleza de los datos recabados –ya sean texto, imágenes, audio o video– y su potencial complementariedad.
 3. **Velocidad** a la que los datos son capturados y procesados –con cada vez más frecuencia la captura y visualización de informes se realiza en casi "tiempo real"–.
 4. **Veracidad** en la calidad de los datos procesados;
 5. **Valor** real para el objeto de análisis o contraste; y
 6. **Validez** final de los datos.
7. Otros expertos han sumado a estas "uves", la **visualización** y otras más.

Estimado estudiante, a continuación, miraremos varios tipos de infografías que hacen referencia a las "uves", ya sean 4, 5, 6, 7 u 8. ¿Qué de común tienen estas? Observará que a través de las figuras (36-40) que se presentan permiten identificar a través de varios autores la diversidad de "uves" que se toman en cuenta dentro de la definición de big data, como se observa, conforme cada autor existe una variación. Es decir, no hay una respuesta correcta, todo está en cómo marcó una estrategia de big data para el trabajo periodístico.

The Four V's of Big Data.

Figura 30.

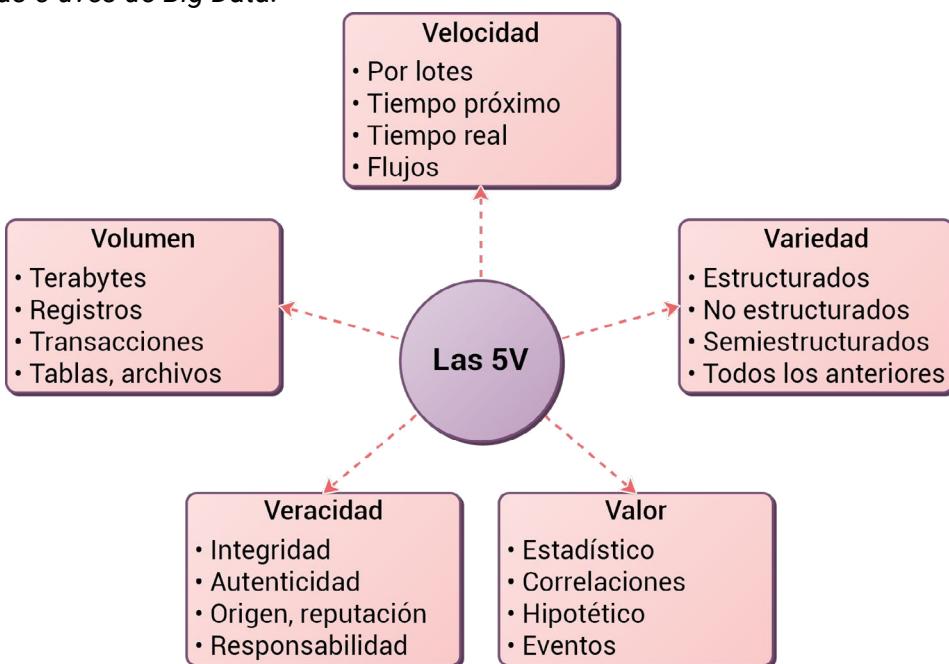
Five V's of Big Data.

Volumen	Velocidad	Variedad	Veracidad	Valor
 Datos en reposo Terabytes a exabytes de datos existentes para procesar.	 Datos en movimiento Transmisión de datos, que requieren de milisegundos a segundos para responder.	 Datos en muchas formas Estructurada, no estructurada, texto, multimedia,...	 Datos en duda Incertidumbre debido a la inconsistencia e incompletitud de los datos, ambigüedades, latencia, engaño, aproximaciones de modelos.	 Data into money Los modelos de negocio se pueden asociar a los datos.

Nota. Michael Walker.

Figura 31.

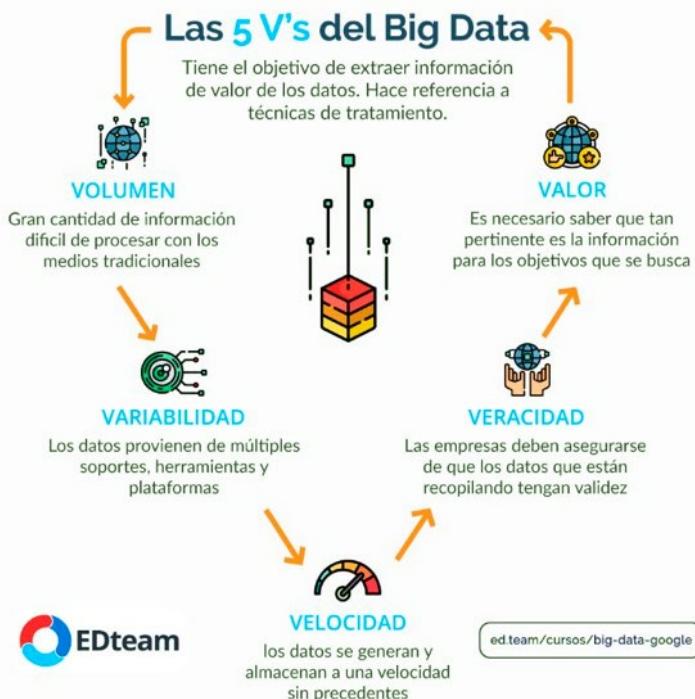
Las 5 uves de Big Data.



Nota. Cravero, Saldaña, Espinosa y Antileo, 2018.

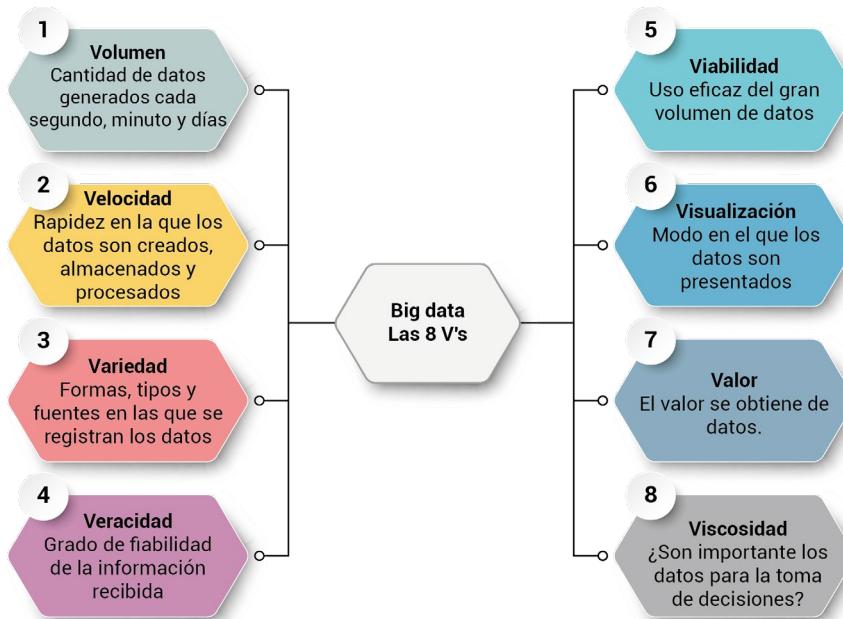
Figura 32.

Las 5 V's del Big data.



Nota. EDteam.

Figura 33.
Las 8 V's de Big data.



Nota. <https://www.juanbarrios.com/la-importancia-de-los-datos-en-salud/>

Las figuras (30-33) que se presentan permiten identificar a través de varios autores la diversidad de "uves" que se toman en cuenta dentro de la definición de *big data*, como se observa, conforme cada autor existe una variación. Es decir, no hay una respuesta correcta, todo está en cómo marcó una estrategia de *big data* para el trabajo periodístico.

7.3. Fuentes de big data

Entre las fuentes de datos podemos detallar las siguientes:

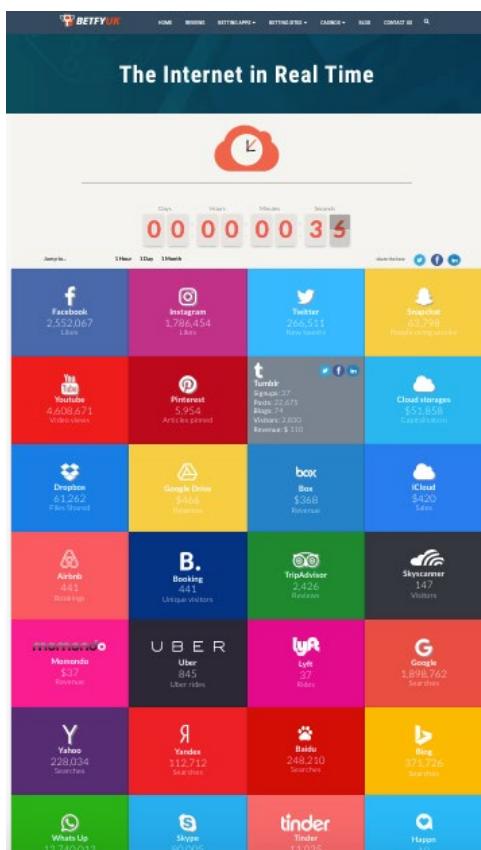
- **Públicas**: normalmente abiertas como los sitios gubernamentales, por ejemplo: <https://data.europa.eu/>
- **No visibles**: datos que entran en el sistema a través de la red de dispositivos que están conectados.
- **Web social**: datos tomados de medios sociales como: Facebook, Instagram, Twitter, YouTube, TikTok, entre otros.

- **Propias:** se suman los datos que el usuario o empresa desee crear de forma manual o automática, generados desde cero.

Recordemos que los datos que se obtienen deben ser protegidos y sobre todo, evitar pagar por ellos, ser vendidos o comercializados, todo ello tienen que ver con la ética en el tratamiento de la información.

La infografía que se encuentra a continuación representa la cantidad de datos que son creados, subidos, compartidos y redistribuidos por la Red de redes, Internet. Puede acceder a esta aplicación para conocer cómo información se incrementa segundo a segundo, todo ello en tiempo real:
<https://betfy.co.uk/internet-realtime/>

Figura 34.
Internet real time.



Note. <https://betfy.co.uk/internet-realtime/>

7.3.1. Tipos de datos

Existe una diversidad de datos que se pueden crear, ubicar o extraer de internet, entre los que encontramos:

- **No estructurados:** documentos, videos, audios, imágenes, etc.
- **Semi-estructurados:** software, hojas de cálculo, informes, reportes.
- ***Estructurados:** conjunto de datos listos para ser procesados.

Según Tim O'Reilly (2014), sólo el 20% de información es estructurada y eso puede provocar muchos errores si no acometemos un proyecto de calidad de datos.

Más allá del gran volumen de datos es qué se hace con ellos. En principio, preguntémonos ¿cuáles son sus ventajas?, y miremos cómo puede ser empleado en el ámbito empresarial, incluyendo la industria mass media.

Figura 35.

Ventajas del uso de datos en las empresas.



A más de las empresas, se expone un ejemplo en el ámbito académico de cómo funciona el proceso de análisis de *big data*:

Figura 36.

Proceso *big data* en la educación.

BIG DATA



The Flipped Classroom

Adaptado de: <http://www.harbus.org/2014/big-data-should-be-eating-education/>

<http://www.theflippedclassroom.es/>

Nota. The flipped classroom <https://www.theflippedclassroom.es/sabes-lo-que-es-el-big-data/>

No olvidemos que, de forma tradicional, un medio a través de su equipo periodístico se introducía en la búsqueda de datos que le permitan verificar y comprobar sus hipótesis. En la actualidad el proceso "tradicional" se mantiene, pero se suma la tecnología a través del big data (incluyendo inteligencia artificial AI), para el análisis, desarrollo y evaluación de la información. La producción en el ámbito periodístico, el uso de métodos de ponderar y recolectar la información ha evolucionado: las noticias están surgiendo desde los números, y no al revés. Sin dejar de lado que gracias a este tipo de tecnologías se han creado nuevos roles, tales como "data science", arquitecto de datos, consultor *big data*, analista de datos, científico de datos, entre otros. Para conocer más le invito a leer el siguiente post denominado [¿Qué es y qué hace un Analista de Datos o Big Data Analyst?](#).

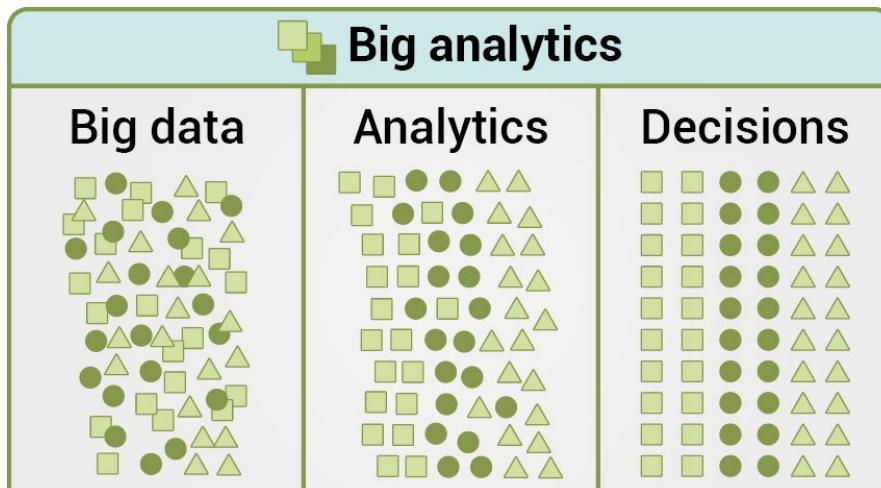
- **Documentales**

He compartido varios textos, pero al estar enfocados en temas de comunicación y sus diversos medios, le recomiendo un listado de documentales sobre *big data* para que en su tiempo libre pueda mirarlos y apreciar el potencial que tiene esta tecnología. Algunos están disponibles en YouTube y Netflix.

- *Big data: el valor de nuestra información:* <https://youtu.be/xDlV1jCW7n0>
- *Prediction by the Numbers:* <https://youtu.be/jq5SLhBQWEw>
- *The Great Hack o Nada es privado:* https://youtu.be/-bLGb_cr1i8
- *The Power of Privacy:* <https://youtu.be/KGX-c5BJNFk>
- *Ojo con tus datos:* <https://youtu.be/uVlqRe02H4s>
- *The Human Face of Big Data:* <https://youtu.be/WAo72Se-yPs>
- *The Wall Street Code:* <https://youtu.be/kFQJNeQDDHA>

Figura 37.

Big Data Analytic



Nota.

¡Video recomendado!

Para profundizar en el maravilloso mundo de los datos le invito a mirar la charla de TedX de Kenneth Cukier, a la cuál llamó: "["Big data is better data"](#)", cuyo speaker narra cómo automóviles auto-dirigidos fueron solo el comienzo de toda esta tecnología. ¿Cuál es el futuro de la tecnología y el diseño impulsado por los datos masivos? En esta apasionante charla se

observa lo que esto supone para el aprendizaje automático, y, por ende, para el conocimiento humano.

¡Lectura recomendada!

Estimado estudiante, para conocer cómo los algoritmos, creados por seres humanos, pueden otorgarle al usuario la capacidad de acceder a la información y mejorar los procesos de comunicación, para ello le invito a revisar el artículo titulado: "[Algoritmos, aplicaciones y Big data, nuevos paradigmas en el proceso de comunicación y de enseñanza-aprendizaje del periodismo de datos.](#)", cuyo *abstracto* detalla que las tecnologías disruptivas y su impacto en el periodismo y comunicación obligan a asumir retos en el aprendizaje de nuevas técnicas para el procesamiento de datos e información. El conocimiento interdisciplinario se pone de manifiesto en la formación de nuevos perfiles profesionales. El periodismo de datos es una muestra de ello, por lo que la inmersión hacia una cultura de datos debe estar precedida por una concienciación en el aprendizaje de aplicaciones de noticias, algoritmos o el tratamiento del Big data, elementos que configuran nuevos paradigmas entre los periodistas de los medios en Internet.



Actividades de aprendizaje recomendadas

Es muy importante que se detenga y realice esta actividad que le proponemos, hacerla constituye un refuerzo a esta unidad estudiada. Conforme el video "[7 ejemplos de Big data](#)", aquí se muestra la acción e impacto de *big data* que podría ser difícil de imaginar tanto sus aplicaciones y beneficios a corto, mediano y largo plazo. En el video se verán ejemplos de: Netflix, McDonald's, Amazon, ZARA, Walmart, Spotify y Hopper.

Consulte a su profesor:

De esta forma terminamos la unidad. Si tiene alguna duda por favor comuníquese con su profesor a través del EVA.

A continuación, es importante desarrollar la autoevaluación, a fin de determinar su grado de comprensión de los contenidos analizados.

Toda unidad de su guía incluye preguntas de autoevaluación que le recomiendo realizar para reforzar sus conocimientos aprendidos.



Autoevaluación 7

Señale la respuesta correcta:

1. **Los macrodatos son:**

- a. Datos.
- b. Conjunto de datos.
- c. Conjunto de datos masivos.

2. **Big Data, se puede referir a:**

- a. Análisis del comportamiento del usuario.
- b. Análisis de resultados electorales.
- c. Análisis del comportamiento ambiental.

3. **En la pirámide de la información, la siguiente descripción “elementos discontinuos que representan hecho”, corresponden a:**

- a. Datos.
- b. Información.
- c. Conocimiento.

4. **En la pirámide de la información, la siguiente descripción “datos procesados para ser útiles”, corresponden a:**

- a. Datos.
- b. Información.
- c. Conocimiento.

5. **En la pirámide de la información, la siguiente descripción “aplicación mental de los datos y la información”, corresponden a:**

- a. Información.
- b. Sabiduría.
- c. Conocimiento.

6. Con el fin de llegar a la “sabiduría”, las empresas deben:

- a. Aprovechar la información.
- a. Invertir en tecnología y formación.
- b. Aumentar su presupuesto.

7. Complete: Qué V’s de Big Data representa. “_____ naturaleza de los datos recabados –ya sean texto, imágenes, audio o video– y su potencial complementariedad”.

- a. Volumen.
- b. Variedad.
- c. Validez.

8. Complete: Qué V’s de Big Data representa. “_____ en la calidad de los datos procesados”.

- a. Variedad.
- b. Validez.
- c. Veracidad.

9. Entre las fuentes de Big Data encontramos:

- a. Privadas.
- b. Gubernamentales.
- c. Públicas.

10. Los grandes volúmenes de datos permiten:

- a. Ser analizados.
- b. Tomar decisiones.
- c. Ser vendidos y comercializados.

[Ir al solucionario](#)

Desarrolle las siguientes preguntas en su blog de notas:

- A. **¿Cuáles son las ventajas del Open data?**
- B. **¿Qué características destaca de los datos abiertos?**
- C. **¿Cómo big data puede tener presencia en los procesos de producción periodística?**

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.



Actividades finales del bimestre



Semana 8

Actividad final.

Se destina la semana 8 para finalizar y entregar las actividades propuestas dentro de la planificación del primer bimestre. Además de realizar la actividad suplementaria para quienes no pudieron asistir a la actividad síncrona.

Con esto se culmina la revisión teórica de la asignatura *Data Journalism*, y se da inicio al tratamiento de la información con fines periodísticos, dado que para entender de datos se debe trabajar con ellos.

¡Estimado estudiante gracias por llegar hasta aquí! ¡Ánimo en el cierre del primer bimestre y nos vemos en el 2BM!



Segundo bimestre

Resultado de aprendizaje 3

- Aplica técnicas y herramientas para la extracción, limpieza, refinado y análisis de datos.

Contenidos, recursos y actividades de aprendizaje recomendadas

El estudio de la asignatura permitirá, de forma práctica, a través de herramientas la gestión para: la extracción, limpieza, análisis y visualización de grandes volúmenes de datos, en conjunto con ejemplos de cómo los medios de comunicación más importantes del mundo cuentan historias y crean productos periodísticos. En principio, se enfocará en delimitar cuál es el rol que tienen la minería de datos o *data mining* en los procesos informativos.



Semana 9

Unidad 8. Data mining o minería de datos

8.1. ¿Qué es data mining?

La minería de datos o *data mining* (DM) dentro del mundo *big data*, lo convierte en un término general para una variedad de estrategias y tácticas que involucran conjuntos de datos masivos y tecnologías, que tiene sentido a partir de la generación de una avalancha de datos. La tendencia de *big data* ha impactado a todas las industrias, incluida la industria de los medios, ya que las tecnologías se están desarrollando para automatizar y simplificar el proceso de análisis de datos, mostrando como multitudes de analistas están siendo capacitados y contratados para satisfacer la demanda en los diversos sectores.

A decir de Ortega Mohedano y Coronel-Salas (2019, p. 826), la minería de datos es una técnica de análisis de datos que focaliza su objeto en el

modelado y el descubrimiento de conocimiento con fines predictivos más que puramente descriptivos. En sí, DM examina grandes volúmenes de datos de manera (semi)automática, cuyo objetivo es encontrar patrones que expliquen su comportamiento en diversos sectores.

Son diversos aspectos que pueden ser tomados para implementar, gestionar y optimizar *big data* en los procesos de minería de datos, los cuales se enfocan en:

1. Escuchar y pensar como lo hacen los usuarios.
2. Atraer y fidelizar al usuario.
3. Anticiparse a las necesidades del sector.
4. Mejorar el impacto de las acciones de difusión.
5. Nuevas oportunidades.
6. Toma de decisiones basadas en datos certeros.

Para visualizar estos aspectos puede revisar la Figura 38 que muestra, visualmente, cual es el ciclo de vida del dato:

Figura 38.

Fases del ciclo de vida del dato.

FASES DEL CICLO DE VIDA DEL DATO. EP



Nota. Juan Barrios <https://www.juanbarrios.com/la-importancia-de-los-datos-en-salud/>

Las estrategias del *data mining* pueden incluir:

- Análisis para permitir una mejor comprensión y orientación de los usuarios;
- Herramientas para entender lo público y lo privado.
- Bases de datos para contar historias.
- Herramientas para administrar y buscar contenido en redes sociales.
- Herramientas para orientar las campañas de comunicación, publicidad o marketing.
- Herramientas para automatizar la producción de texto e historias de video.
- Herramientas para identificar datos relevantes y permitir ser eficaz a la hora de acceder a datos.

En tal sentido, la importancia de *data mining* en el periodismo de datos es conocer las formas de analizar volúmenes de datos y organizar la información que facilite la toma de decisiones al momento de narrar historias. Para profundizar le invito a revisar y analizar el siguiente REA sobre *data mining*, que profundiza en las aplicaciones para realizar experimentos prácticos con algoritmos para la extracción de datos utilizando software: <https://www.oercommons.org/authoring/7053-data-mining>

¡Video recomendado!

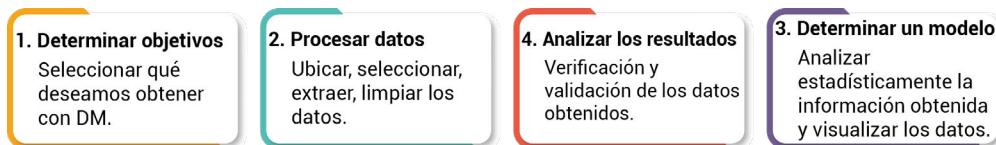
Le invito a mirar el video denominado [*¿Qué es Data Mining?*](#), en el cual se menciona que «el reto de las empresas de hoy en día es extraer información a raíz de la gran cantidad de datos que recopilan de usuarios y clientes».

8.2. ¿Quién mina?

La persona (minero) que realiza la minería de datos (tal como ocurre con los mineros en las minas de oro, cobre, etc.), deberá realizar las siguientes etapas: 1. Determinar objetivos. 2. Procesamiento de datos. 3. Determinación de modelo y 4. Análisis de resultados.

Figura 39.

Pasos que realizan los exploradores de datos.

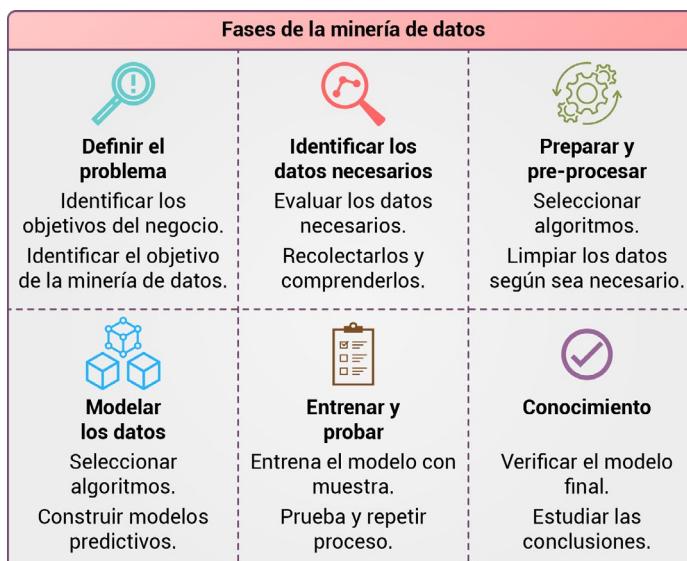


Nota. (IEBS, 2018).

Entre las fases que acompañan las etapas del DM están:

Figura 40.

Fases de data mining.



Nota. Unayta: <https://unayta.es/data-mining-big-data/>

8.3. Aplicación de data mining

Gracias a Internet y al estar rodeados de información, la minería de datos se aplica en varios sectores, entre los más destacados están:

- Ventas online
- Marketing
- Educación
- Indumentaria

- Transporte
- Tecnología
- Finanzas
- Salud
- Mass Media
- Entre otros

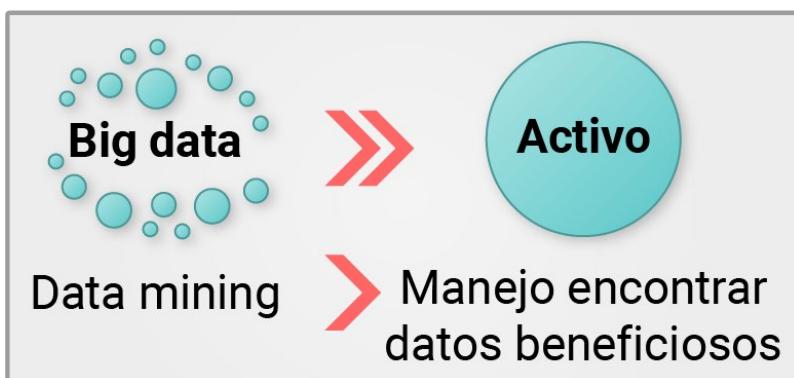
Uno de los sectores desde dónde se pueden recabar datos es la Web a través de diversas empresas enfocadas en social media: Facebook, Twitter, YouTube, Instagram y empresas de *retail* como Amazon.

8.4. Big data Vs. Data mining

Recordemos que donde hay **datos**, **hay información**, donde hay información, **hay conocimiento** y este lo podemos obtener a través del proceso de *big data* en conjunto con *data mining*.

Figura 41.

Roles del big data y data mining.



Nota. Elaboración propia

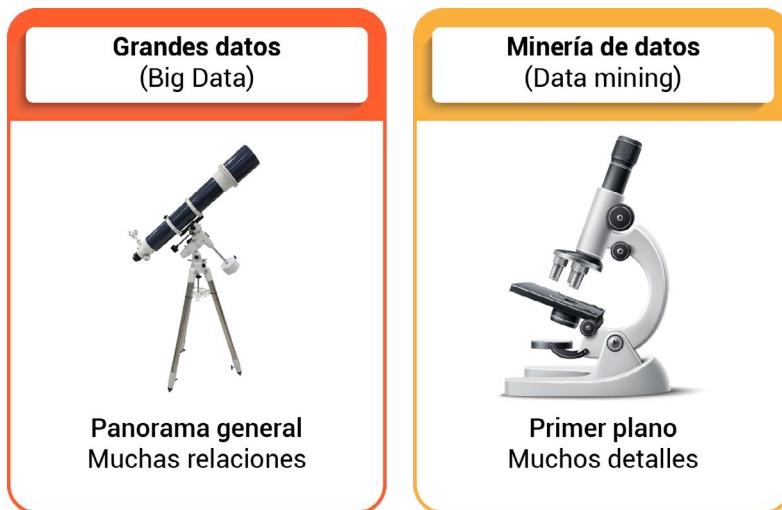
A continuación, miremos las diferencias entre estas dos tecnologías, sobre todo centrarnos en qué Data Mining permite identificar y extraer la información relevante de grandes conjuntos de datos y Big data, se enfoca en la recolección y almacenamiento de grandes cantidades de datos:

Data Mining Vs Big data.

Aquí, un ejemplo mucho más gráfico del rol que tiene tanto BD como DM. De un gran volumen de datos podemos, a través de la minería, encontrar información útil para el cumplimiento de nuestros objetivos.

Figura 42.

Diferencia entre *Big data* y *Data mining*.



Nota. <https://www.tuataratech.com/2015/06/mineria-de-datos-data-mining-vs-grandes.html>

8.5. Cómo el periodismo se beneficia del data mining

Como lo hemos analizado en todas las unidades, los datos nos permiten hacer conexiones y darles un contexto, más aún si lo que pretendemos es publicar nuestros productos periodísticos. Para ello, seguiremos los siguientes pasos:

Figura 43.

Pasos minería de datos.



Nota. Elaboración propia.

Es conocido que el periodismo es consumidor de datos: los recopila, procesa (en diversos géneros) y los archiva o documenta (uso de bases). Y, si el periodismo es un gran consumidor, también es un productor de información, sobre todo de la que más valor posee y con la cual puede relacionar, cruzar y validar con otra. Para ello, le invito a revisar la infografía sobre Data Mining que está a continuación:

[Infografía a Data Mining](#)



Actividades de aprendizaje recomendadas

¡Video recomendado!

Previo a iniciar con la Unidad 9, dejo una pregunta: ¿puede el periodismo beneficiarse de data mining? Por ello, le invito a mirar el video denominado “[¿Quieres predecir el futuro? Usa datos](#)”, en donde el speaker de la charla TED, Nicolás Loeff, menciona que los “nuevos algoritmos que procesan grandes cantidades de datos revolucionaron nuestra habilidad de modelar el mundo. Estos pasaron de curiosidades científicas a herramientas imprescindibles que cambiaron industrias enteras en poco tiempo. Así,

los algoritmos predicen qué vamos a comprar antes que lo hagamos, dónde ocurrirá un crimen antes que suceda, y hasta aspectos de nuestra personalidad mejor que nuestras parejas."

Ahora bien, continuemos con los pasos a seguir para la recolección de datos con fines periodísticos.

Consulte a su profesor

De esta forma terminamos la unidad. Si tiene alguna duda por favor comuníquese con su profesor a través del EVA. A continuación, es importante desarrollar la autoevaluación, a fin de determinar su grado de comprensión de los contenidos analizados.

Toda unidad de su guía incluye preguntas de autoevaluación que le recomiendo realizar para reforzar sus conocimientos aprendidos.



Autoevaluación 8

1. Complete: La tendencia ____ ha impactado a todas las industrias, incluida la industria de los medios, ya que las tecnologías se están desarrollando para automatizar y simplificar el proceso de análisis de datos.
 - a. Big Data
 - b. Data Mining
 - c. Open data
2. Complete: _____ es una técnica de análisis de datos que focaliza su objeto en el modelado y el descubrimiento de conocimiento con fines predictivos más que puramente descriptivos.
 - a. Big Data
 - b. Data Mining
 - c. Open data
3. Las fases del ciclo del dato son:
 - a. Datos, captura, almacenamiento, procesamiento, uso del dato, publicación y archivado.
 - b. Datos, captura, almacenamiento, limpieza, uso del dato, publicación y archivo.
 - c. Datos, captura, extracción, procesamiento, uso del dato, publicación y archivado.
4. Complete, entre las estrategias de Data Mining están: _____ para permitir una mejor comprensión y orientación de los usuarios.
 - a. Segmentación
 - b. Gestión
 - c. Análisis

5. Complete, entre las estrategias de Data Mining están: _____ para contar historias.
- a. Bases de datos
 - b. Información
 - c. Conocimiento
6. ¿Quién mina la información?
- a. Ingeniero.
 - b. Minero.
 - c. Computador.
7. El proceso de Big Data y Data Mining se compone por:
- a. Datos, información y conocimiento.
 - b. Información, datos y conocimiento.
 - c. Conocimiento, datos e información.
8. Complete: la siguiente característica/definición: “*Se refiere a la recolección y almacenamiento de grandes cantidades de datos*”, corresponde a:
- a. Data Mining.
 - b. Big Data.
 - c. Open Data.
9. Complete: la siguiente característica/definición: “*Utiliza diferentes técnicas basadas en la estadística y la inteligencia artificial*”, corresponde a:
- a. Data Mining.
 - b. Big Data.
 - c. Open Data.

10. Complete: la siguiente característica/definición: “Necesita herramientas especiales para capturar, gestionar y tratar la información”, corresponde a:

- a. Open Data.
- b. Data Mining.
- c. Big Data.

[Ir al solucionario](#)

Desarrolle la siguiente pregunta:

A. ¿Cómo el periodismo puede hacer uso efectivo de data mining?

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.

Su participación en el proceso de enseñanza y aprendizaje es esencial en nuestra asignatura. ¡Ahora, continuemos con la Unidad 9!



Semana 10

Unidad 9. Localización de datos

9.1. Ubicación o localización



Ubicación – Localización

La característica principal de este primer paso es la obtención de datos a través de diversas fuentes, entre ellas Internet. Todo dato es susceptible de localizar si este se encuentra disponible y en formatos adecuados. Existen maneras de obtener estos, por ejemplo de: propios, de APIs, documentos (tipo PDF), *screen scraping* o *web scraping*.

Miremos a detalle cada uno:

- 1. Obtener datos de API:** interfaces de programación de aplicaciones online, tales como interfaces provistas por bases de datos y muchas aplicaciones (incluyendo Twitter, Facebook y otras). Esta es una manera fantástica de acceder a datos oficiales o comerciales, así como datos de medios sociales.
- 2. Extraer datos de PDF:** esto es complicado dado que es un lenguaje para impresoras y no retiene mucha información sobre la estructura de los datos presentados en el documento. Para extraer información de PDF hay algunas herramientas y tutoriales (gracias a YouTube o la

web de Tabula). De estar en imágenes (hecho .PDF) será mucho más complicado extraer la información.

3. **Usar “screen scraping” (extracción) para obtener datos de sitios de la Red:** Se trata de extraer contenido estructurado de una página web con la ayuda de un programa de recuperación de información o escribiendo una pequeña pieza de software (para personas con conocimiento en programación). Si bien este método es muy poderoso y puede ser usado en muchos lugares, se requiere comprender un poco cómo funciona.

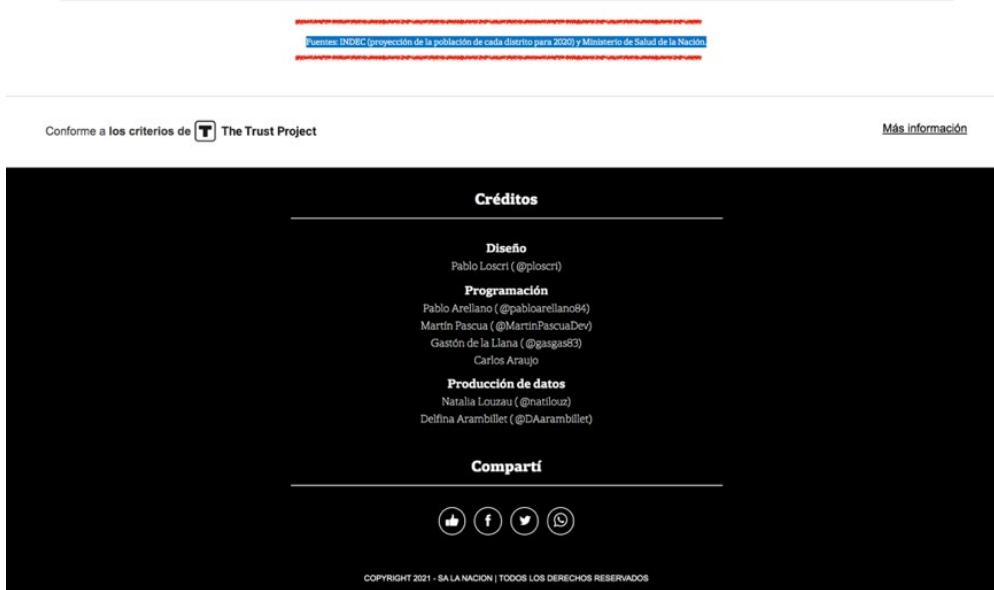
Con todas esas opciones técnicas, no debemos olvidar las opciones simples: a menudo vale la pena invertir un poco de tiempo en buscar un archivo con datos que pueden ser interpretados por una computadora o llamar a la institución/empresa/persona que tiene los datos que se quiere.

Presento a continuación algunas imágenes y enlaces que muestran la localización de las fuentes de datos en medios de comunicación y organismos públicos. Por ejemplo, diario *La Nación* toma como fuente para su artículo “Vacunación en la Argentina en tiempo real” al:

- Instituto Nacional de Estadísticas y Censos de la República Argentina, INDEC: <https://www.indec.gob.ar/indec/web/Nivel4-Tema-2-24-119>
- Ministerio de Salud de Argentina: <https://www.argentina.gob.ar/salud>
- *La Nación (Argentina)*: [https://www.lanacion.com.ar/sociedad/la-vacunacion-en-la-argentina-en-tiempo-real-nid02022021/#/](https://www.lanacion.com.ar/sociedad/la-vacunacion-en-la-argentina-en-tiempo-real-nid02022021/#/https://www.lanacion.com.ar/sociedad/la-vacunacion-en-la-argentina-en-tiempo-real-nid02022021/#/)

Figura 44.

Ejemplos del uso de datos externos al medio.



Nota. Diario La Nación.

Lo mismo ocurre con el artículo "Vacunas: lo que hay que saber sobre todas las que se aplican en el mundo": <https://www.lanacion.com.ar/sociedad/vacunas-lo-que-hay-que-saber-sobre-todas-las-que-se-aplican-en-el-mundo-nid01062021/>. Al ser un tema de carácter global (COVID-19) el interés es general y la información se vuelve relevante.

Figura 45.

Ejemplos del uso de datos externos al medio.

ASTRAZENECA / COVISHIELD	
<small>Para que la desarrolló: Reino Unido</small>	
Eficacia para prevenir casos sintomáticos	70,4% (en estudios clínicos) 70,4% (en estudios clínicos) 70,4% (en estudios clínicos)
Eficacia para prevenir casos graves	80,0% (en estudios clínicos) 80,0% (en estudios clínicos) 80,0% (en estudios clínicos)
Nº de estudios	6
¿Se aplica en la Argentina?	Sí
Instituciones	Universidad de Oxford, AstraZeneca
Marcos científicos	AZD1222
Datos de priorización	Argentina, Azerbaiyán, Bielorrusia, Brasil, Canadá, Costa Rica, Colombia, Egipto, Eslovaquia, Eslovenia, Francia, Hong Kong (China), India, Japón, Malta, Perú, Reino Unido, Rusia, Sudáfrica
Fábricas productoras	Allemánia, Argentina, Bélgica, Corea de Sur, Japón, Reino Unido, Tailandia, México, Países Bajos, Polonia, Unión
Tecnología que utiliza	Virus viral no replicativo
Efectos adversos más comunes	Dolor espontáneo en el sitio de la inyección, fiebre, malestar, náuseas, malestar general, dolor de cabeza, dolor articular y muscular. También se han señalado casos raros de reacciones anafilácticas con plaquetas bajas.
Precio por dosis	Entre 2,2 y 5 dólares
Temperatura de almacenamiento	2°C - 8°C
Dosis	2
Intervalo mínimo entre dosis	28 días
Póker de autorizaciones	Internacional

PRIMERAS CONCLUSIONES EN LA POBLACIÓN

Los primeros análisis publicados en la literatura sobre un intervalo de dosis más largo del recomendado presentan. La vacuna resultó tener un 81,9% de eficacia cuando se administró en un intervalo de 12 semanas o más. Estos resultados fueron confirmados por estudios que midieron los niveles de anticuerpos de los voluntarios. De esta forma, tanto Australia, Australia la Argentina, optaron por extensiones de recomendación basadas en estos resultados. Un artículo llevado a cabo por el Departamento de Salud de Reino Unido, al no recomendar por la comunidad científica, concluyó que tanto la vacuna de Oxford como la de Pfizer reducen la probabilidad de传染 70% y 95,3% respectivamente. La eficacia de la vacuna en un régimen de dos dosis tiene una efectividad de 67,1% para los casos sintomáticos producidos por la variante Delta, detectada en Reino Unido, y de 95,8% para los casos de la variante Delta, detectada en India. Un tercer estudio, realizado en México, concluyó que la vacuna resultó ser 88% efectiva para disminuir los riesgos de longitudes trascendentes por Covid-19 dentro los 28

Los resultados de un estudio del Ministerio de Salud de la Nación, que evalió el funcionamiento de las vacunas en grupos de 60 años, indican que la efectividad de la vacuna es de 79,3% en una sola dosis, y de 88,8% con un régimen de dosis completa.

DÓNDE SE APlica LA VACUNA DE ASTRAZENECA



Nota. Diario La Nación.

Lo más certero en la localización de datos es identificar, en principio, el tema (por ejemplo: salud → coronavirus → vacuna), luego ir a la fuente de datos primaria (por ejemplo: ministerio de salud); como vemos es esencial y así lo hace el diario:

Lo más certero en la localización de datos es:

Conforme a los criterios de  The Trust Project

Más información

CRÉDITOS

COORDINACIÓN Y EDICIÓN PERIODÍSTICA: **Florencia Fernández Blanco** @fiorfb
PRODUCCIÓN Y TEXTO: **Juana Copello** @CopelloJuana
EDICIÓN VISUAL: **Pablo Loscri** @ploscri, **Giselle Ferro** @giselleferrod, **Gabriel Podestá** @gabrielpodesta, **Alejandra Bliffeld** @aleblif, **Mariana Trigo Viera** @mariantv
ANÁLISIS DE DATOS: **Delfina Arambillet** @DArambillet, **Natalia Louzau** @nataliouz
PROGRAMACIÓN: **Carlos Araujo**, **Gaston de la Llana** @gasgas83, **Nicolás Rivera** @dudeneico, **Nicolás Caffarini** @caffarini, **Martín Pascua** @MartinPascuaDev
FUENTES: **Ministerio de Salud de la Nación**, **Organización Mundial de la Salud**, **Unicef**, **Moderna**, **Instituto Jenner de Oxford**, **Pfizer**, **Johnson & Johnson**,
Instituto de Gamaleya, **Sinovac Biotech**, **Bharat Biotech**, **Instituto Finlay de Vacunas**.



COMPARTIR



COPYRIGHT 2021 - SA LA NACION | TODOS LOS DERECHOS RESERVADOS

9.2. Fuentes de datos

Para mantener una organización en la recolección de información, recomiendo crear una cuenta en la web RainDrop (<http://raindrop.io>) que permite guardar nuestros enlaces por temáticas y así mantener organizadas nuestras posibles fuentes de datos.

Entre las plataformas, webs y datasets abiertos que podemos emplear para localizar información están:

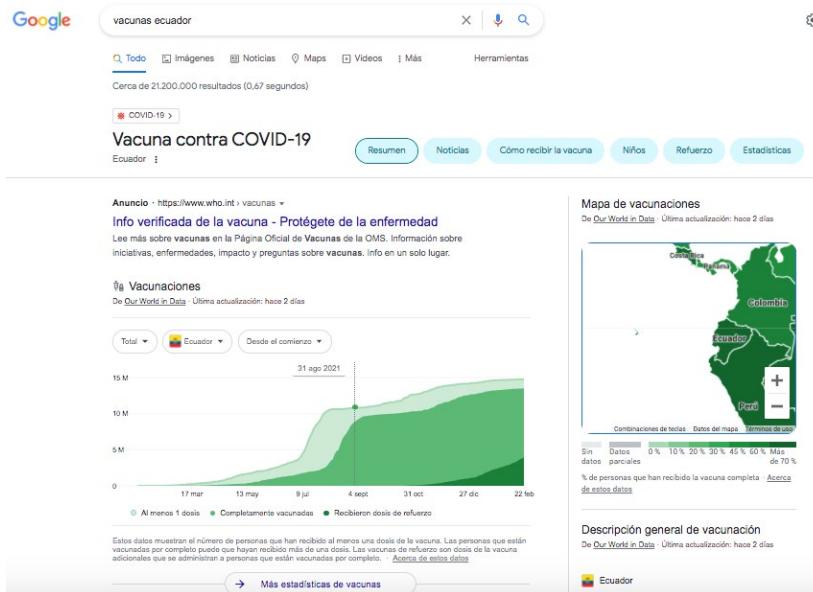
- *Instituto Nacional de Estadísticas y Censos (INEC)*: <https://www.ecuadorencifras.gob.ec/institucional/home/>
- *Catálogo de datos abiertos, Ecuador*: <https://datosabiertos.gob.ec/dataset/>
- *Red ecuatoriana de datos abiertos y metadatos*: <https://datosabiertosecuador.org/>
- *Google Data Search*: <https://toolbox.google.com/datasetsearch>
- *Google Public Data Explorer*: <https://www.google.com/publicdata/directory>
- *Google Trends*: <https://trends.google.com/trends/explore>

- *Catálogo de datasets del European Union Open Data Portal*: <http://data.europa.eu/euodp/es/data>
- *Open Addresses*: <http://openaddresses.io/>
- *Open Knowledge Foundation*: <https://okfn.org/>
- *PANGAEA*: www.pangaea.de/
- *UK Data Archive*: <http://data-archive.ac.uk/>
- *World Bank Open Data*: <http://datos.bancomundial.org/>
- *Datahub*: <http://datahub.io/>
- *National Climatic Data Center*: <https://www.ncdc.noaa.gov/data-access/quick-links>
- *Global Health Observatory data*: www.who.int/gho/en/
- *Data.gov.sg*: <https://data.gov.sg/>
- *Earthdata*: <https://earthdata.nasa.gov/>
- *Amazon Web Services Open Data Registry*: <https://registry.opendata.aws/>
- *Pew Internet*: www.pewinternet.org/datasets/
- *Datos.gob.s, España*: <https://datos.gob.es/>
- *Catálogo de datos, Uruguay*: <https://catalogodatos.gub.uy/>
- *Our World Data*: <https://ourworldindata.org/coronavirus-data>
- *GitHub - coronavirus*: <https://github.com/CSSEGISandData/COVID-19>
- *World Health Organization (WHO)*: <https://www.who.int/>
- *European Centre for Disease Prevention and Control (ECDC)*: <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>
- *DXY.cn. Pneumonia*: <http://3g.dxy.cn/newh5/view/pneumonia>
- *US CDC*: <https://www.cdc.gov/coronavirus/2019-ncov/index.html>

- *BNO News*: <https://bnonews.com/index.php/2020/02/the-latest-coronavirus-cases/>
- *Worldometers*: <https://www.worldometers.info/coronavirus/>
- *1Point3Arces*: <https://coronavirus.1point3acres.com/en>
- *COVID Tracking Project*: <https://covidtracking.com/data>.
- *Los Angeles Times*: <https://www.latimes.com/projects/california-coronavirus-cases-tracking-outbreak/> y <https://github.com/datadesk/california-coronavirus-data>
- *The Mercury News*: <https://www.mercurynews.com/tag/coronavirus/>

Incluso el propio Google (búsqueda simple o avanzada) nos pueda dar información dirigida hacia bases de datos, aquí un ejemplo de las "vacunas en Ecuador", donde al final de los gráficos entrega la fuente.

Figura 46.
Buscador Google.



Nota.[Google](#)

Aunque pueden no ser siempre fáciles de encontrar, muchas bases de datos en la Red están indexadas por motores de búsqueda:

- Cuando busquemos datos debemos incluir tantos términos de búsqueda relacionados con el contenido de los datos que tratamos de encontrar. Google y otros motores de búsqueda le permiten buscar por tipo de archivo.
- Por ejemplo, puede buscar solo planillas de cálculo (agregando a su búsqueda "filetype:XLS filetype:CSV"), datos geográficos ("filetype:shp"), o extractos de bases de datos ("filetype:MDB, filetype:SQL, filetype:DB"), incluso puede buscar PDF ("filetype:pdf").
- También podemos buscar con una parte de una URL. Hacer una búsqueda en Google de "inurl:downloads filetype:xls" o "inurl:descargas filetype:xls" significa buscar todos los archivos Excel que tienen "*downloads*" o "*descargas*" en su dirección de la red (si encuentra una sola descarga, a menudo vale la pena simplemente verificar qué otros resultados existen para la misma carpeta en el servidor de la red). También puede limitar su búsqueda sólo a aquellos resultados dentro de un solo nombre de dominio, buscando "site:agency.gov".

Tabla 3.

Guía de búsqueda de datos propuesto en el Manual de Periodismo de Datos

Tipo de búsqueda	Detalle de la búsqueda	Ejemplos
Búsqueda en la red	Ajustar la búsqueda en Google: utilizar términos relacionados con el contenido de datos, buscar por lugar de disponibilidad	hoja de cálculo: "filetype: XLS filetype:CSV; datos geográficos ("filetype:MDB, filetype:SQL, filetype:DB); PDF (filetype:pdf") URL: "inurl: downloads filetype:xlx" o "inurl: descargas filetype:xls "sitio: lugar+datos buscados"
	Explorar sitios y servicios de datos	Portales oficiales Datos de investigación Portales estadísticos
	Conocer a tecnología informática de los entes oficiales	Buscar las grandes bases de datos los cuadros organizativos, departamentos/unidades
	Busque nuevamente	Cambiar frases y conjunto de palabras
Búsqueda directa	Ir directamente a la fuente	Ir a la persona que maneja los datos de la organización en cuestión
	Pregunte en un foro	Get The Data o Quora
	Pregunte en lista de correo	Data-Driven Journalism List, NICAR-L, Open Knowledge Foundation, Proyecto Wombat
	Pregunte a un experto	Especialistas
	Súmase a Hacks/Hackers	Organización periodística internacional
	Pedido de acceso a la información	Cuando la entidad pública tiene los datos buscados
Búsqueda específica	Obtener datos de archivos en papel	Requiere el uso de un software de reconocimiento de texto
	Investigar registros de llamadas telefónicas	Datos que se pueden cruzar para obtener una historia

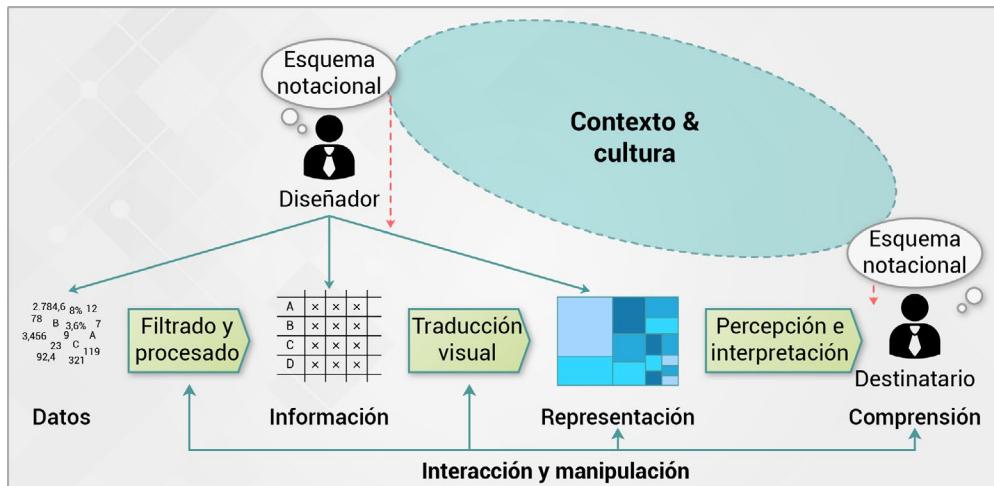
Nota. Adaptación de "Una guía para trabajos de campo de 5 minutos" de Boyer, B., Keefe, J. Lindenberg, F., Park, J. Y Wu, C., 2012, recuperado de [Http://interactivos.lanacion.com.ar/manual-data/obtener_datos_0.html](http://interactivos.lanacion.com.ar/manual-data/obtener_datos_0.html)

Es esencial definir el tema a investigar para así destinar los esfuerzos en la búsqueda de bases de datos (en sitios abiertos) que permitan acompañar

nuestro producto periodístico. Le recuerdo, estimado estudiante, que las bases de datos también se pueden construir desde cero y eligiendo las variables pertinentes conforme a nuestro objetivo y preguntas que surgen previo a la ejecución de una noticia o reportaje.

Figura 47.

Diagrama de Infovis.



Nota. JC. Dürsteler y Y. Engelhardt.

Tal como muestra la figura 47, este diagrama permite conocer el proceso que obtiene el dato hasta su comprensión.

Consulte con su profesor:

De esta forma terminamos la unidad. Si tiene alguna duda por favor comuníquese con su profesor a través del EVA. A continuación, es importante desarrollar la autoevaluación, a fin de determinar su grado de comprensión de los contenidos analizados.

Toda unidad de su guía incluye preguntas de autoevaluación que le recomiendo realizar para reforzar sus conocimientos aprendidos.



Autoevaluación 9

Señale la respuesta correcta:

1. **Complete: en la localización, la siguiente descripción: “interfaces de programación de aplicaciones online, tales como interfaces provistas por bases de datos y muchas aplicaciones (incluyendo Twitter, Facebook y otras)”, corresponde a:**
 - a. Obtener datos de API.
 - b. Extraer datos de PDF.
 - c. Uso “screen scraping”.
2. **Complete: en la localización, la siguiente descripción: “esto complicado dado que es un lenguaje para impresoras y no retiene mucha información sobre la estructura de los datos presentados en el documento”, corresponde a:**
 - a. Obtener datos de API.
 - b. Extraer datos de PDF.
 - c. Uso “screen scraping”.
3. **Complete: en la localización, la siguiente descripción: “Se trata de extraer contenido estructurado de una página web con la ayuda de un programa de recuperación de información o escribiendo una pequeña pieza de software”, corresponde a:**
 - a. Obtener datos de API.
 - b. Extraer datos de PDF.
 - c. Uso “screen scraping”.
4. **Entre los ejemplos de la Unidad 9 (Ubicación y localización de datos), ¿qué medio emplea fuentes externas para la producción de productos periodísticos enfocados en las vacunas contra covid-19?**
 - a. *La Nación*.
 - b. *El País*.
 - c. *Primicias.ec*.

- 5. Entre los sitios para realizar búsquedas de datos están:**
- a. Google.
 - b. Presidencia Ecuador.
 - c. Banco Central .
- 6. Según la guía de datos propuesta en el Manual de Periodismo, “ajustar la búsqueda en Google: usar términos relacionados con el contenido de datos...”, corresponde a búsqueda:**
- a. Directa.
 - b. Específica.
 - c. En la red.
- 7. Según la guía de datos propuesta en el Manual de Periodismo, “ir directamente a la fuente”, corresponde a búsqueda:**
- a. Directa.
 - b. Específica.
 - c. En la red.
- 8. Según la guía de datos propuesta en el Manual de Periodismo, “obtener datos de archivo en papel”, corresponde a búsqueda:**
- a. Directa.
 - b. Específica.
 - c. En la red.
- 9. Según la guía de datos propuesta en el Manual de Periodismo, “conocer la tecnología informática de los entes oficiales”, corresponde a búsqueda:**
- a. Directa.
 - b. Específica.
 - c. En la red.

10. Según la guía de datos propuesta en el Manual de Periodismo, “investigar registros de llamadas telefónicas”, corresponde a búsqueda:

- a. Directa.
- b. Específica.
- c. En la red.

[Ir al solucionario](#)

Desarrolle la siguiente pregunta

- A. ¿Cuáles son las mejores alternativas para la localización y ubicación de datos enfocados en el ámbito periodístico?

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.

Su participación en el proceso de enseñanza y aprendizaje es esencial en nuestra asignatura. ¡Ahora, continuemos con la Unidad 10!



Semana 11

Unidad 10. Extracción

10.1. ¿Qué comprende la extracción de datos?



Extracción

¿Qué sucede cuando no existe información accesible para elaborar un producto periodístico? En estos casos se recurre a herramientas que permitan extraer el contenido de páginas web que posean, sobre todo, información valiosa.

Aclaro que no todos los sitios nos entregan datos, por ejemplo, si se trata de sitios corporativos donde lo que más se destaca es "quiénes somos, su función", etc.

- *Primer paso: la captura*

Identificado el tema y el sitio web, debemos buscar inputs de datos de calidad.

Nuestros datos deben ser:

- Completos.
- Únicos.

- Precisos.
- Consistentes.

Asegurémonos de:

Volumen

- Nunca es suficiente: Debemos pensar siempre en nuevas opciones que complementen nuestros inputs de datos.
- Volumen argumentado: datos necesarios y legítimos (fuente y objetivo).

Velocidad

- Nuevos escenarios: información en tiempo real. Si los procesos de depuración no son adecuados, para cuando se verifica su calidad ya han perdido valor.
- Pensar en trabajar con *samples* (muestras) y no con *datasets* completos en ocasiones que lo permitan.

Variedad

- Diferentes fuentes.
- Estructuras de datos diferentes.
- Métricas de calidad diferenciadas.
- Fuentes: bases de datos (propias o de terceros), datos de CRM de la empresa, conversación social, facturación, ideas/creatividad, memoria humana.

Veracidad

- Evitar sesgos.
- Evitar imprecisión.
- Buscar la conexión entre fuentes y usuarios finales.

Valor

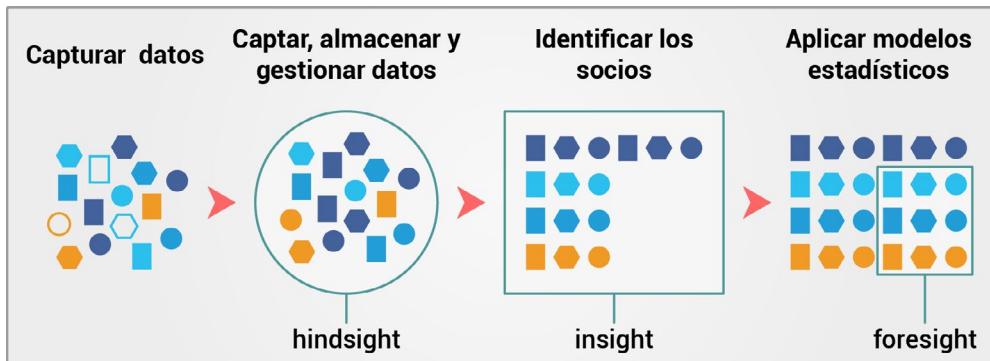
- Perspectiva coste - beneficio.
- Consecución de objetivos

10.2. Herramientas

Cuando capturamos datos muchos de los formatos suelen estar en forma de imágenes o en el mejor de los casos en .PDF editable. La figura 48 ayuda a ejemplificar el proceso de gestión de datos capturados.

Figura 48.

Proceso de gestión de datos capturados.



Nota. Amadeus.com

- **Tabula**

Para ello recomiendo descargarse el aplicativo "Tabula", cuyo fin es extraer la información, específicamente "tablas" que estén dentro de estos archivos: <https://tabula.technology/>

Figura 49.

Tabula.

Tabula



Tabula is a tool for liberating data tables locked inside PDF files.

[View the Project on GitHub](#)
tabulapdf/tabula

[Download for Windows](#) [Download for Mac](#)

[View source on GitHub](#) Current Version: 1.2.1 Other Versions: pre-releases & archives

Need help? Open an [Issue on Github](#).

Donate: Help support this project by [backing us on OpenCollective](#).

We'd love to hear from you! Say hi on Twitter at [@TabulaPDF](#)

Latest Version: Tabula 1.2.1

June 4, 2018

Tabula 1.2.1 fixes several bugs in the user interface and processing backend. (You can read about all the changes in the [release notes](#).)

Download Tabula below, or on the [release notes page](#).

Special thanks to our OpenCollective backers for supporting our work on Tabula; if you find Tabula useful in your work, please consider a one-time or monthly donation.

How Can Tabula Help Me?

If you've ever tried to do anything with data provided to you in PDFs, you know how painful it is — there's no easy way to copy-and-paste rows of data out of PDF files. Tabula allows you to extract that data into a CSV or Microsoft Excel spreadsheet using a simple, easy-to-use interface. Tabula works on Mac, Windows and Linux.

Who Uses Tabula?

Tabula is used to power investigative reporting at news organizations of all sizes, including ProPublica, The Times of London, Foreign Policy, La Nación (Argentina), The New York Times and the St. Paul (MN) Pioneer Press.

Grassroots organizations like SchoolCuts.org rely on Tabula to turn clunky documents into human-friendly public resources.

And researchers of all kinds use Tabula to turn PDF reports into Excel spreadsheets, CSVs, and JSON files for use in analysis and database applications.

Nota. Tabula Technology <https://tabula.technology/>

¿Cómo usar Tabula?

- Subir un archivo PDF que contenga una tabla de datos.
- Busque la página que desee, luego seleccione la tabla haciendo clic y arrastrando para dibujar un cuadro alrededor de la tabla.
- Haga clic en “vista previa y exportación de datos extraídos”. Tabula intentará extraer los datos y mostrar una vista previa. Inspeccione los datos para asegurarse de que se vean correctos. Si faltan datos, puede volver atrás para ajustar su selección.
- Haga clic en el botón “Exportar”.
- Ahora podremos trabajar con los datos como un archivo de texto o una hoja de cálculo en lugar de un PDF.

A continuación podrá observar dos tutoriales de la herramienta:

- <https://youtu.be/2JHsAYV5eaw>

- <https://youtu.be/XwGjZK37Qt0>



- **Web Scraper**

La herramienta permite a través de un complemento en el navegador Chrome, "raspar, escarbar ... en sí, extraer datos de la web."

¿Para qué se puede utilizar Web Scraper?

- Generación de clientes potenciales: correo electrónico, número de teléfono, otros detalles de contacto relacionados con la extracción de datos de varios sitios web.
- Comercio electrónico: extracción de datos de productos, de precios de productos, descripción, de URL, recuperación de imágenes, etc.
- Rastreo de contenido de sitios web: extrae información de portales de noticias, blogs, foros, etc.
- Supervisión minorista: supervisa el rendimiento del producto, las existencias y los precios de los competidores o proveedores, etc.
- Monitoreo de marca: revisión de productos, rastreo de contenido social para análisis de sentimientos;
- Inteligencia comercial: recopilar datos para decisiones comerciales clave.
- Extracción de *big data* para aprendizaje automático, marketing, desarrollo de estrategias comerciales, investigaciones.

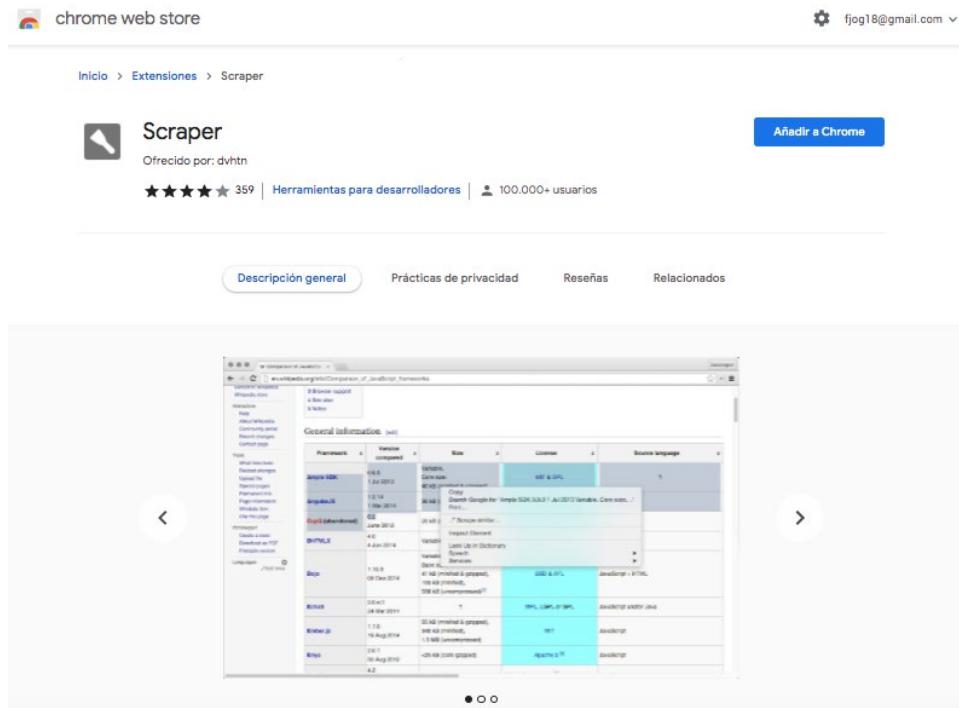
Antes de iniciar, le invito a mirar el video en el cual cuenta, brevemente, qué hace la herramienta: <https://youtu.be/aViWT-WpzYI>

- Ingrese al sitio web: <https://webscraper.io/>
- Clic en "*Install*". Se dirigirá hacia la *web* de complementos de Google Chrome.
- El primer paso es instalarlo, para ello ingrese en:
<https://chrome.google.com/webstore/detail/scrap...mbigbapnjcgaffohmbkdlcacepngjd?hl=es>

Le aparecerá una imagen como la siguiente y deberá hacer clic en "instalar".

Figura 50.

Extensión Scraper



Nota.

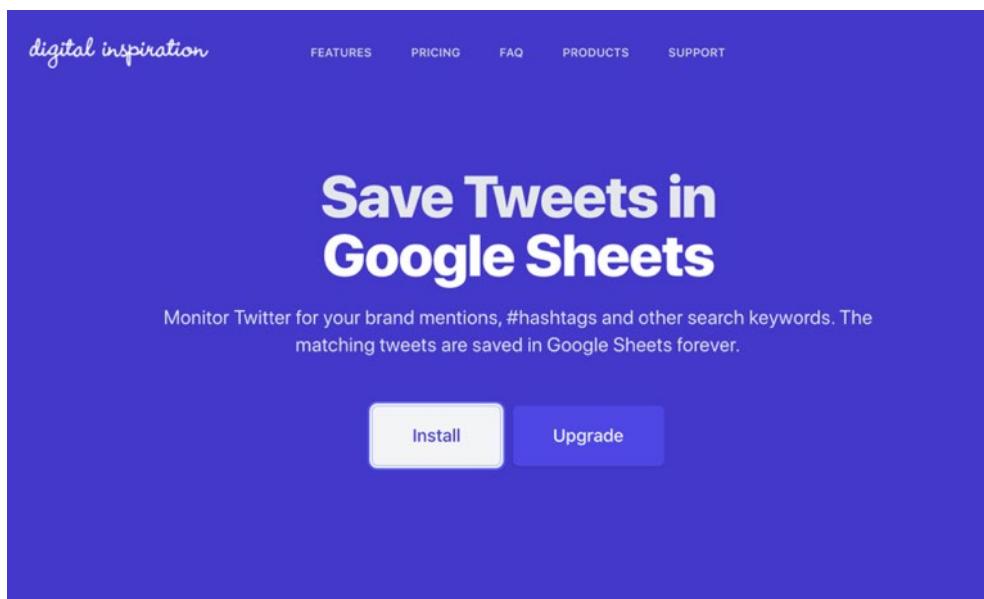
Aquí un tutorial introductorio, por favor, no deje de mirarlo, será de apoyo hasta encontrarnos en la tutoría donde aprenderemos a manejar web scraper: https://youtu.be/n7fob_XVsbY

▪ Twitter Archiver

Permite monitorear, a través de Twitter, las menciones de personas o marcas, sus hashtags (#) y otras palabras clave de búsqueda. Los tweets se guardan en una hoja de cálculo de Google.

Figura 51.

Web de Digital inspiration.



Nota. Twitter Archiver <https://digitalinspiration.com/product/twitter-archiver>

Para iniciar:

- Instale el complemento en su cuenta de [Gmail.com](https://gmail.com) y siga los pasos para conectar la hoja de cálculo con su cuenta de Twitter.
- Especifique su consulta de búsqueda de Twitter o utilice operadores de búsqueda avanzada, y los tuits coincidentes se guardarán automáticamente.
- Cierre la hoja ya que la aplicación se ejecutará en segundo plano y descargará automáticamente los nuevos tweets coincidentes.
- Clic para mirar una introducción: <https://youtu.be/MGU7azCYFpw>
- **InstaBro**

Esta herramienta permite descargar todas las fotos y videos desde cualquier cuenta, hashtag o ubicación de Instagram. Netamente está disponible para el sistema operativo "macOS", la comparto en caso sea de su utilidad.

Figura 52.

InstaBro.

The screenshot shows the InstaBro website for Mac OS X. At the top, there's a large circular logo with a downward-pointing arrow icon. Below it is a blue button labeled "Download v5.3.7, macOS 10.12+". To the right, there's a "View Demo" button and a "Buy Now!" button. The main content area has a heading "#1. Users, Tags & Places. Add as much sources as you need" and a sidebar with sections for "USERS", "HASHTAGS", and "PLACES", each listing several Instagram handles. The main window shows a grid of Instagram photos with their file names and download dates. A status bar at the bottom indicates "300 items | 300 total".

Nota. <https://instabro.app/>

A continuación, el video tutorial: <https://youtu.be/V1aH9y1lcwU>

▪ **NodeXL**

Otra de las herramientas que permiten extraer y visualizar datos de redes sociales es "NodeXL", está habilitado para el sistema operativo de Windows. Le invito a conocerla: <https://nodexl.com/>

Figura 53.

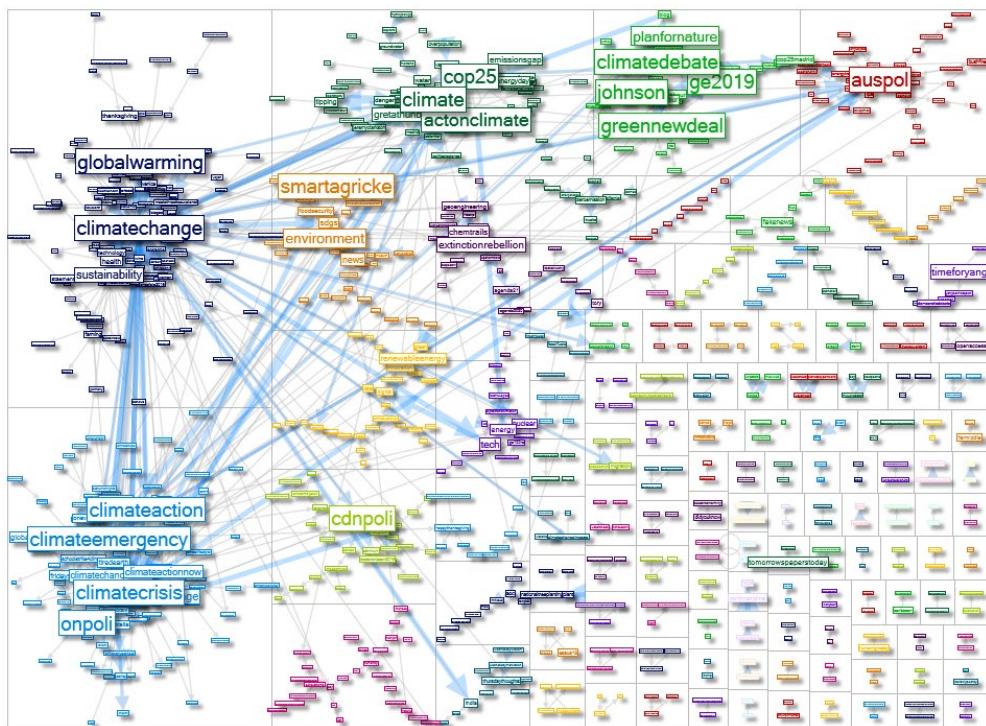
NodeXL



Nota. <https://nodexl.com/>

Figura 54.

Grafos obtenidos por NodeXL.



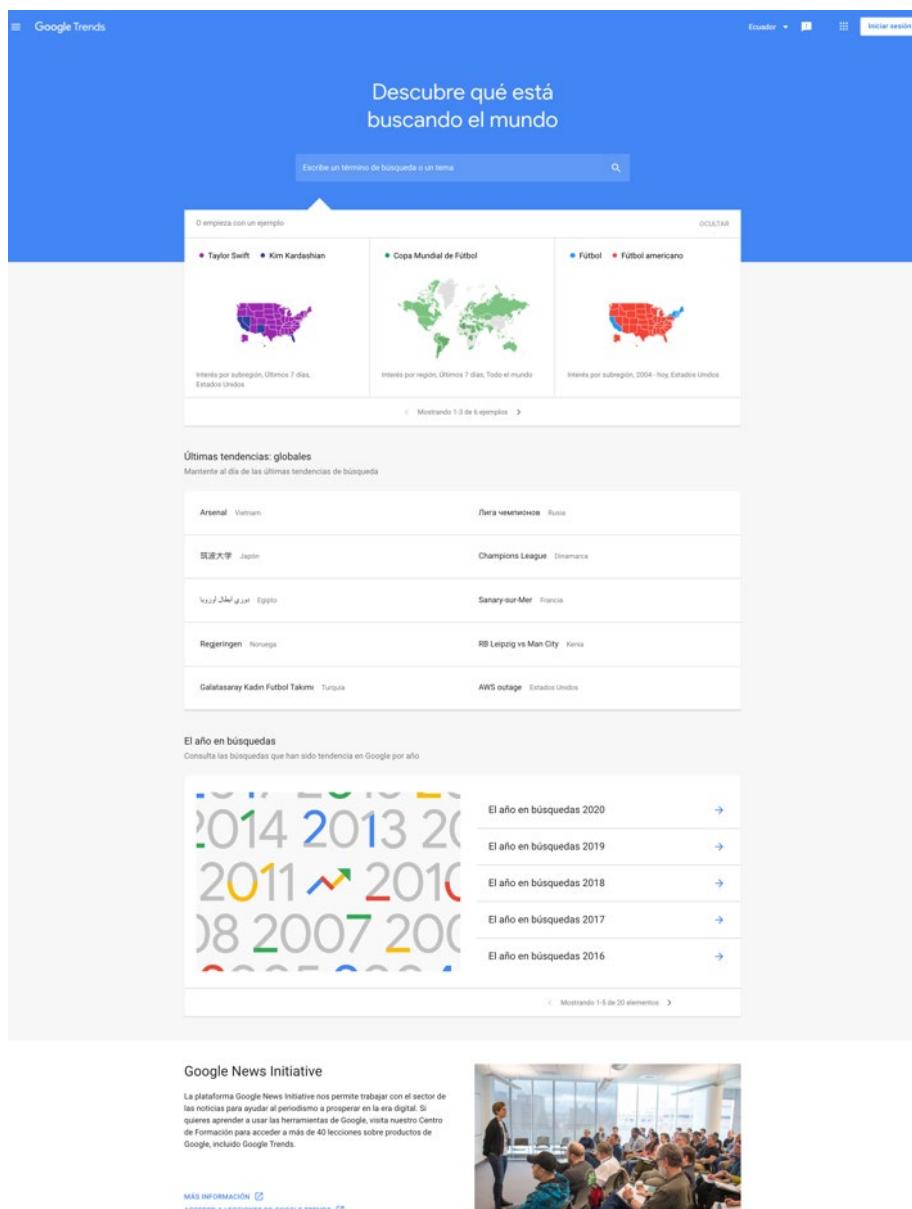
Nota. NodeXL <https://www.smrfoundation.org/nodexl/tutorials/semantic-networks/>

En este sitio web podrá encontrar varios tutoriales: <https://www.smrfoundation.org/nodexl/tutorials/>

- **Google Trends**

Es otra aplicación que permite conocer tendencias y compararlas, incluso determinar su búsqueda por geolocalización.

Figura 55.
Sitio web de Google Trends



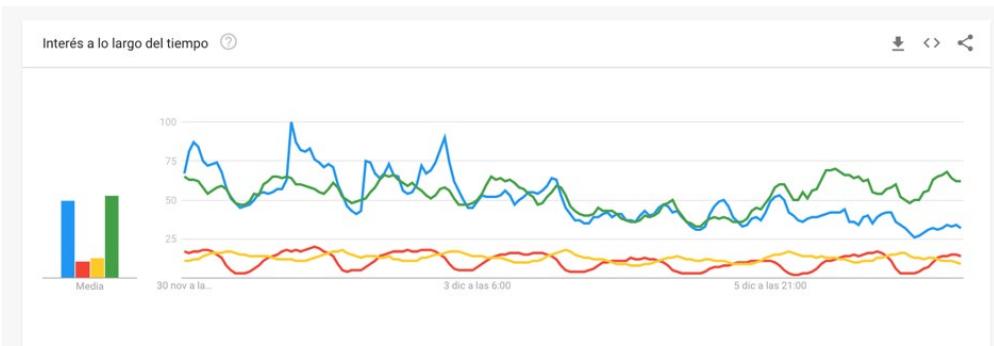
Nota. Google Trends <https://trends.google.es/trends/>

Aquí el resultado de la búsqueda en todo el mundo, en los últimos 7 días de los siguientes términos: Búsqueda de Tendencias Ómicron, covid-19 y vacuna.

Búsqueda de Tendencias Ómicron

Es una aplicación versátil porque podemos descargar los datos en imágenes y en un documento en formato ".csv", por ejemplo: con la primera categoría "Interés a lo largo del tiempo", además de visualizar cuál es el resultado de esas búsquedas.

Figura 56.
Búsqueda de tendencias.



Nota. Google Trends <https://trends.google.es/trends/explore/TIMESERIES/1638906600?hl=es&tz=300&date=now+7-d&q=Omicron,vacuna,covid-19,Hospital&sni=3>

A continuación, se analiza la tabla de descarga desde GT.

[Tabla que se descarga desde GT.](#)

Como se ve en la tabla, hay celdas vacías, las cuales debemos limpiar de forma manual o automática, a fin de consolidar nuestros datos y poder analizarlos sin inconvenientes. Por ello, le invito a continuar en la revisión de nuestra siguiente unidad.



Actividades de aprendizaje recomendadas

[**¡Video recomendado!**](#)

Le invito a mirar el video "[Formatos abiertos y cerrados](#)", en el cual, la experta en datos Sandra Crucianelli, realiza una aproximación a cómo encontrar y procesar grandes volúmenes de información.

Consulte con su profesor:

De esta forma terminamos la unidad. Si tiene alguna duda por favor comuníquese con su profesor a través del EVA. A continuación, es importante desarrollar la autoevaluación, a fin de determinar su grado de comprensión de los contenidos analizados.

Toda unidad de su guía incluye preguntas de autoevaluación que le recomiendo realizar para reforzar sus conocimientos aprendidos.



Autoevaluación 10

Señale la respuesta correcta:

1. La única alternativa para extraer datos es:

- a. Pedirlas a las organizaciones de forma directa.
- b. Extraer datos desde un sitio web.
- c. Capturarlos desde Google.

2. Para la extracción debemos considerar que nuestros datos sean:

- a. Completos.
- b. Perfectos.
- c. Vendibles.

3. Para la extracción debemos considerar que nuestros datos sean:

- a. Correctos.
- b. Únicos.
- c. Claros.

4. Para la extracción debemos considerar que nuestros datos sean:

- a. Precisos.
- b. Veloces.
- c. Suficientes.

5. Para la extracción debemos considerar que nuestros datos sean:

- a. Capturados.
- b. Diferentes.
- c. Consistentes.

6. Las características “Evitar sesgos, imprecisiones y conexión de fuente”, corresponden a:

- a. Veracidad.
- b. Variedad.
- c. Valor.

- 7. Las características “Nuca es suficiente y volumen argumentado”, corresponden a:**
- a. Veracidad.
 - b. Variedad.
 - c. Volumen.
- 8. Las características “Nuevos escenarios y pensar en trabajar con muestras”, corresponden a:**
- a. Variedad.
 - b. Velocidad.
 - c. Valor.
- 9. Las características “consecución de objetivos y perspectiva de coste – beneficio”, corresponden a:**
- a. Volumen.
 - b. Variedad.
 - c. Valor.
- 10. Las características “métricas de calidad diferenciadas” corresponden a:**
- a. Valor.
 - b. Variedad.
 - c. Volumen.

[Ir al solucionario](#)

Desarrolle las siguientes preguntas

- A. ¿Qué herramientas de extracción han sido empleadas por usted?
- B. Realice una búsqueda en la Red y ubique 5 herramientas para la extracción de datos:

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.

Su participación en el proceso de enseñanza y aprendizaje es esencial en nuestra asignatura. ¡Ahora, continuemos con la Unidad 11!



Semana 12

Unidad 11. Limpieza

11.1. Importancia de la limpieza de datos



Limpieza

La limpieza o *data cleaning* de los resultados puede hacerse de dos maneras:

- Automática
- Manual

Con la primera opción (automática) podemos emplear fórmulas sencillas que el propio Excel nos da. Además, existe la posibilidad de trabajar los datos en línea a través de las hojas de cálculo de Google Sheets.

En la Unidad anterior vimos una de las tablas que logramos descargar de Google Trends, como se observó necesita un filtrado y una limpieza. Al ser un documento con poca información se puede recurrir a la opción 2 (manual), es decir, a limpiar de forma manual los errores y ruidos que encontramos: acentos, columnas o filas en blanco, etc. Aquellos errores que son visibles a nuestra mirada.

Figura 57.

Infografía del proceso de minería de datos.



Fuente: (IEBS, 2018)

¿Cuáles son los beneficios?

Son algunos los beneficios que podemos obtener al limpiar nuestros datos, como por ejemplo citamos los siguientes:

- Elimina los principales errores e inconsistencias que son inevitables cuando se extraen múltiples fuentes de datos en un conjunto de datos.
- El uso de herramientas para limpiar los datos hará que todos sean más eficientes, ya que podrán obtener rápidamente lo que necesitan de los datos.
- Menos errores es igual a más eficiencia en los procesos.
- La capacidad de mapear las diferentes funciones y lo que sus datos están destinados a hacer y de dónde provienen sus datos.

11.2. Flujo de trabajo

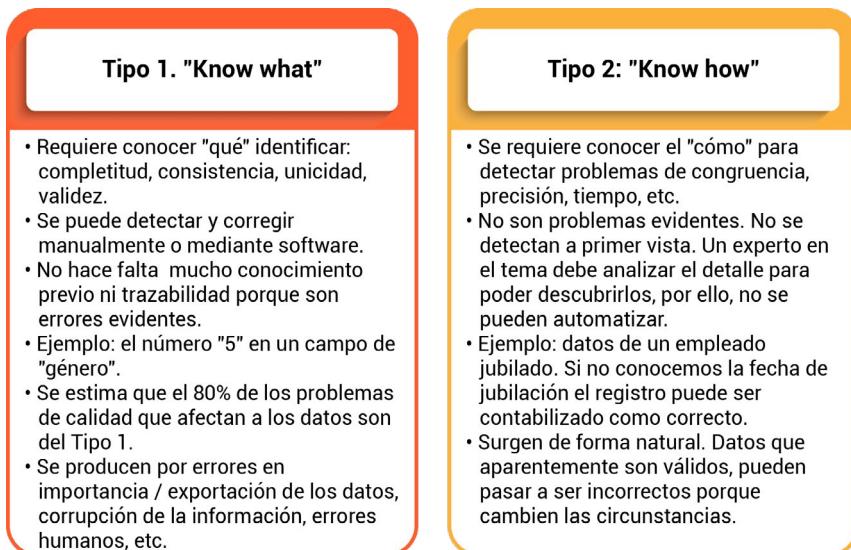
Es importante tener presente respetar el proceso de trabajo que realizaremos en nuestra limpieza de datos, el cual iniciamos con la inspección hasta llegar al informe:

1. Inspección: detecta datos inesperados, incorrectos e inconsistentes.
2. Limpieza: corrige o elimina las anomalías descubiertas.
3. Verificación: después de la limpieza, los resultados se inspeccionan para verificar la corrección.
4. Informes: se registra un informe sobre los cambios realizados y la calidad de los datos almacenados actualmente.

Miremos lo que Jim Barker (citado por Recuero de los Santos, 2017) ha considerado "el problema de los datos":

Figura 58.

Clasificación del tipo de problemas encontrados al limpiar datos.



Nota. Jim Barker.

En esencia, problemas del tipo 1: Fáciles, sencillos y económicos de resolver. Tipo 2: Complejos de resolver. Requieren, por norma, de ayuda externa.

Entre las tareas que debemos cumplir para la limpieza de datos están:

- Descartar/implementar campos.
- Normalizar e igualar formatos.
- Revisar ortografía.
- A formatos de fecha.
- Eliminar duplicidades en columnas y filas.
- Eliminar ruido (registros no útiles).

Además, debemos asegurarnos de:

- **Validez**
 - Restricciones de tipo de datos: los valores en una columna en particular deben ser de un tipo de datos en particular, por ejemplo, booleano (lógica binaria "si - no"), numérico, fecha, etc.)
 - Restricciones de rango: generalmente, los números o las fechas deben estar dentro de un cierto rango.
 - Restricciones obligatorias: ciertas columnas no pueden estar vacías.
 - Restricciones únicas: un campo, o una combinación de campos, debe ser único en un conjunto de datos.
 - Restricciones de conjunto de miembros: los valores de una columna provienen de un conjunto de valores discretos, por ejemplo, valores de enumeración, el género de una persona puede ser masculino o femenino.
 - Restricciones de clave externa: como en las bases de datos relacionales, una columna de clave externa no puede tener un valor que no existe en la clave primaria referenciada.
 - Patrones de expresión regular: campos de texto que tienen que estar en un cierto patrón. Por ejemplo, se puede requerir que los números de teléfono tengan el patrón (593) 999–9999.

- Validación de campo cruzado: se deben cumplir ciertas condiciones que abarcan varios campos. Por ejemplo, la fecha de alta del paciente del hospital no puede ser anterior a la fecha de ingreso.
- **Exactitud:**
 - El grado en que los datos están cerca de los valores verdaderos.
 - Datos completos: El grado en que se conocen todos los datos requeridos.
- **Consistencia:**
 - El grado en que los datos son consistentes (coherentes), dentro del mismo conjunto de datos o en múltiples conjuntos de datos.
 - La inconsistencia ocurre cuando dos valores en el conjunto de datos se contradicen entre sí.
- **Uniformidad:**
 - El grado en que se especifican los datos utilizando la misma unidad de medida.
 - Usar kilos o libras; kilómetros o millas, por ejemplo, sería una decisión por tomar.

Tabla 4.Consejos de *limpieza de base de datos*.

Consejos de limpieza	Práctica específica	Ventajas
No cambiar las tablas ni los archivos originales	Etiquetar para no confundir los archivos	Possibilidad de volver a empezar en caso de errores Permite la comparación de los dos archivos
		Mantener los datos que no se necesitan almacenados en la base original
Revisar una muestra de la base	Proceder a una verificación aleatoria de los datos	Detectar si hay algo fuera de lo normal Ayudar a conocer mejor los datos
Revisar si hay valores extraños	Revisar la lista de valores	Detectar si hay algo fuera de lo normal
	Revisar los espacios en blanco y los caracteres ajenos que se infiltran en las celdas	
Preocuparse por el tipo de columnas		Especialmente importante si los datos serán utilizados para alimentar una aplicación
		Evitar el desorden
Practicar antes de cambiar las cosas	Asegurarse con una pequeña muestra que lo que se quiere hacer funcionará	Asegurarse un buen resultado
Limpio es mejor que ingenioso	Realizar operaciones sencillas y trabajar por etapas	Ganar tiempo
Dejar rastro de lo que se hace	Mantener un registro de lo que se hace con los datos	Resolver problemas Corregir errores Repetir procesos

Nota. Adaptación de "Evita errores al momento de limpiar tus datos", de Escuela de datos, 2012, recuperado de <http://es.schoolofdata.org/>

11.3. Herramientas

Entre las herramientas para la limpieza de datos, constan:

- Tableau (con guía): <https://www.analyticsvidhya.com/learning-paths-data-science-business-analytics-business-intelligence-big-data/tableau-learning-path/>
- Cloud Dataprep by Trifacta: <https://cloud.google.com/dataprep/>
- Open Refine <https://openrefine.org/>

Figura 59.

Web de OR.

The screenshot shows the homepage of the OpenRefine website. On the left, there's a sidebar with links to Home, Community, Documentation, Download, Data Privacy, Contact Us, and Blog. The main content area has a dark header with the text "Welcome!" and a brief introduction about OpenRefine being a powerful tool for working with messy data. Below this, there's a section about data privacy and a link to the Code for Science & Society. The main part of the page features a large video player titled "Google Refine 2.0 - Introduct...". The video player shows a screenshot of the OpenRefine interface, which is a grid-based editor for cleaning and transforming data. A play button is visible in the center of the video frame.

Fuente: Open refine <https://openrefine.org/>

En el periodismo de datos existe un amigo incondicional (si estamos muy alejados de esa "amistad" es hora de retomar contacto) y son las hojas de cálculo de Excel. Le invito a conocerlo un poco más y sobre todo, aprovechar todas sus ventajas más si lo que deseamos es contar historias basadas en macro(micro)datos.

- Excel y Google Sheets: de las más sencillas para eliminar errores y ruidos en nuestra base de datos.
 - Guía en Excel: <https://www.digitalresponse.es/blog/8-formas-de-limpiar-datos-en-excel-parte-1>
 - Guía en Google Sheets: <https://support.google.com/a/users/answer/9604139?hl=es>

Figura 60.

Funcionalidades de las hojas de cálculo de Google.

Lista de funciones de Hojas de cálculo de Google

Hojas de cálculo de Google admite las fórmulas de celdas que se suelen encontrar en la mayoría de los paquetes de hojas de cálculo para computadoras de escritorio. Estas fórmulas se pueden usar para crear funciones que manejen datos y calculen strings y números.

Aquí te ofrecemos una lista de todas las funciones disponibles en cada categoría. Cuando las utilices, no olvides agregar comillas alrededor de todos los componentes de las funciones que se compongan de caracteres alfabéticos y no referieran a celdas ni columnas.

Puedes cambiar el idioma de las funciones de Hojas de cálculo de Google del inglés a otros 21 idiomas.

Filtrar con algunas palabras clave...

Restringir por...

Web

- [Lista de funciones de Hojas de cálculo de Google](#)
- [HIPERVÍNCULO \(HYPERLINK\)](#)
- [IMPORTDATA](#)
- [IMPORTFEED](#)
- [IMPORTHTML](#)
- [IMPORTRANGE](#)
- [IMPORTXML](#)
- [ISURL](#)

Tipo	Nombre	Sintaxis	Descripción
Analizador	TO_DOLLARS	TO_DOLLARS(valor)	Convierte un número en un valor en dólares. Más información
Analizador	TO_PURE_NUMBER	TO_PURE_NUMBER(valor)	Convierte una fecha, hora, porcentaje, moneda a algún otro valor numérico con formato en un número puro sin formato. Más información
Analizador	CONVERT	CONVERT(valor, unidad_de_inicio, unidad_de_finalización)	Convierte un valor numérico a una unidad de medida diferente. Más información
Analizador	TO_DATE	TO_DATE(valor)	Convierte un número especificado en una fecha. Más información
Analizador	TO_PERCENT	TO_PERCENT(valor)	Convierte un número en un porcentaje. Más información
Analizador	TO_TEXT	TO_TEXT(valor)	Convierte un valor numérico en un valor de texto. Más información
Arreglo	ARRAY_CONSTRAIN	ARRAY_CONSTRAIN(rango_entrada, num_filas, num_columnas)	Restringe un resultado de arreglo a un tamaño especificado. Más información

Nota. Google Sheets https://support.google.com/docs/table/25273?hl=es-419&ref_topic=9199554

Comparto, además, uno tutoriales para:

- la limpieza, ordenamiento <https://youtu.be/QCt1L40cR9I>;
- normalización de bases de datos en Excel <https://youtu.be/fwEnUFhtZuw> y,
- creación de tablas dinámicas https://youtu.be/EQ3mX-_hros.

Consulte con su profesor:

De esta forma terminamos la unidad. Si tiene alguna duda por favor comuníquese con su profesor a través del EVA. A continuación, es importante desarrollar la autoevaluación, a fin de determinar su grado de comprensión de los contenidos analizados.

Toda unidad de su guía incluye preguntas de autoevaluación que le recomiendo realizar para reforzar sus conocimientos aprendidos.



Autoevaluación 11

Señale la respuesta correcta:

- 1. La limpieza de datos es conocida como:**
 - a. Estandarización.
 - b. *data cleaning.*
 - c. automatización.

- 2. Complete: “_____ los principales errores e inconsistencias que son inevitables cuando se extraen múltiples fuentes de datos en un conjunto de datos.**
 - a. Eliminar
 - b. Filtrar
 - c. Gestionar

- 3. Complete: “El uso de _____ para limpiar los datos hará que todos sean más eficientes, ya que podrán obtener rápidamente lo que necesitan de los datos.”**
 - a. artefactos
 - b. recursos
 - c. herramientas

- 4. Complete: “____ errores es igual a____ eficiencia en los procesos.”**
 - a. Más – mejor
 - b. Menos - más
 - c. Mejor - más

- 5. Complete: “La capacidad de _____ las diferentes funciones y lo que sus datos están destinados a hacer y de dónde provienen sus datos.”**
 - a. mapear
 - b. visualizar
 - c. limpiar

- 6. El flujo de trabajo de la limpieza de datos es el siguiente:**
- a. Inspección, limpieza, validación, informes.
 - b. Inspección, limpieza, verificación, innovación.
 - c. Inspección, limpieza, verificación, informes.
- 7. La clasificación de problemas encontrados al limpiar datos del Tipo 1, según Jim Barker son:**
- a. Tipo 1: Complejos de resolver.
 - b. Tipo 1: Fáciles, sencillos y económicos de resolver.
 - c. Tipo 1: Requieren, por norma, de ayuda externa.
- 8. La clasificación de problemas encontrados al limpiar datos del Tipo 2, según Jim Barker son:**
- a. Tipo 2: Complejos de resolver.
 - b. Tipo 2: Fáciles, sencillos y económicos de resolver.
 - c. Tipo 2: Requieren ser estandarizados.
- 9. Entre las herramientas para limpiar datos están:**
- a. Tabula.
 - b. Open Refine.
 - c. Datacleaning.
- 10. Entre los consejos para la limpieza de datos se destaca:**
- a. No cambiar las tablas ni los archivos originales.
 - b. Cambiar los archivos originales para evitar errores.
 - c. Trabajar sobre los archivos originales.

[Ir al solucionario](#)

Desarrolle la siguiente pregunta

A. ¿En qué beneficia al periodismo la limpieza de datos?

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.

Su participación en el proceso de enseñanza y aprendizaje es esencial en nuestra asignatura. ¡Ahora, continuemos con la Unidad 12!



Semana 13

Unidad 12. Análisis e interpretación

12.1. Importancia



Análisis e interpretación

El análisis e interpretación de los datos está sujeto a la correcta extracción y limpieza de nuestra base. Se debe recalcar que es importante saber interpretar los datos previos a su visualización.

Figura 61.

Imagen que emula la limpieza



Nota. phromdechalfreepik.com

Una vez que tenemos datos: ¿Qué hacer con ellos? ¿Qué se debe buscar? ¿Qué herramientas debemos usar? ¿Cómo los debemos analizar e interpretar?

A partir de la definición que nos proporciona [QuestionPro.com](#) (2021), al afirmar que "el análisis de datos es la ciencia que se encarga de examinar un conjunto de datos con el propósito de sacar conclusiones sobre la información para poder tomar decisiones, o simplemente ampliar los conocimientos sobre diversos temas.", podemos determinar que es una de las fases más importantes dentro del proceso de producción periodística que se verá representada en los productos informativos que creemos.

Nuestro análisis de datos puede ser de índole cualitativo (entrevistas, grupos focales, observación, etc.) o cuantitativo (información numérica, por ejemplo, encuestas), lo que nos puede permitir:

- Tomar decisiones rápidas e informadas.
- Visualización, lo que permite tomar decisiones certeras.
- Proporciona una ventaja competitiva frente a otros.

En la siguiente figura consta cómo utilizar Excel:

Tabla 5.*Usar Excel para hacer periodismo de datos*

Tarea	Definiciones
Clasificar	Consiste en ordenar los datos de manera alfabética o destacando valores extremos (el más grande, el más pequeño, etc.)"
Filtrar	Resulta necesario cuando se requiere estudiar un dato en particular a partir de una gran colección de datos
Matemáticas y texto	Uso de fórmulas (cálculos en hojas de cálculo)
Tablas dinámicas	Permite adicionar datos que estén en categorías

Nota. (Grassler, 2017, p. 88)

Retomemos el estudio para recordar que el periodismo de profundidad, interpretativo y análisis, se orientan a explicar el porqué de los acontecimientos que se producen en la vida diaria. Interpretar consiste en buscar el sentido a los hechos noticiosos que llegan en forma aislada, situarlos en un contexto, darles un sentido y entregárselo a un usuario no especializado. Además, Alex Gongora (2014), sostiene que, por exigencia profesional, esta interpretación debe tratar de prescindir de opiniones personales, debe basarse en hechos concretos, pertinentes y presentada en forma amena y atractiva.

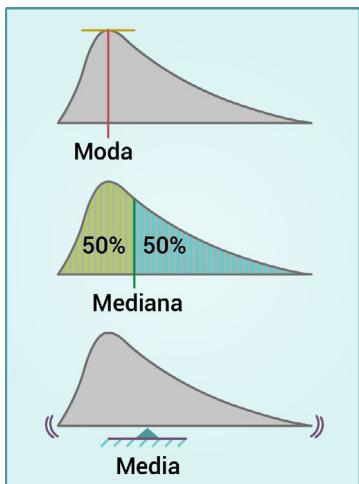
"Los periodistas no sólo reproducen lo que ven y oyen, ejercen también una investigación sobre lo acontecido porque los hechos no se producen descontextualizados de una situación concreta. Los hechos no surgen aislados de una realidad más amplia, se insertan en ella." (Gongora, 2014)

Qué (no) busca el periodismo interpretativo:

- No busca convencer (a través de la opinión) sino informar.
 - Lo que se VE es noticia.
 - Lo que se SABE es antecedente.
 - Lo que se SIENTE es opinión.
- Busca aclarar las noticias aparentemente dispersas.
- Contar el qué ocurrieron las cosas y no describir sólo qué ocurrió.

Figura 62.

Gráfico que detalla la “moda, mediana y media”, bases para el análisis de datos (estadística).



Nota.Twitter [https://twitter.com/GabrielGruber/
status/1384926207121166338?s=20](https://twitter.com/GabrielGruber/status/1384926207121166338?s=20)

¿Cómo pasar de estadista a intérprete de datos?

Así como la alfabetización refiere a “la capacidad de leer para conocer, escribir de modo coherente y pensar críticamente acerca de material impreso, la alfabetización en materia de datos es la capacidad de manejar datos para conocer, producir coherentemente y pensar críticamente acerca de datos.

Una periodista por cierto no necesita tener un título en estadística para ser más eficiente en el manejo de los datos. Enfrentada a las cifras, unos pocos trucos simples pueden ayudarla a armar una historia mucho mejor.

12.2. Tres preguntas para interpretar datos

Para Gerd Gigerenzer (2014) del Instituto Max Planck. “Tener mejores herramientas no permitirá hacer mejor periodismo si éstas no son utilizadas con visión. Aunque no tenga ningún conocimiento de matemática o estadísticas, puede convertirse fácilmente en una periodista de datos experimentada haciendo 3 preguntas muy simples.” Miremos a continuación cuáles son:

1. ¿Cómo se obtuvieron los datos?

Cuando dudamos de la credibilidad de una cifra, verifiquemos, contrastemos y si no estamos seguros (re)preguntemos.

Por ejemplo:

- Para los datos policiales, los sociólogos a menudo realizan estudios de victimización, en los que preguntan a la gente si es víctima de crímenes. Estos estudios son mucho menos volátiles que los datos policiales. Quizás ese sea el motivo por el que no se los destaca en los medios.

Otros test / leyes /fórmulas permiten evaluar la credibilidad de los datos, pero ninguno de ellos suplanta su pensamiento crítico.

2. ¿Qué se puede aprender de ellos?

Ejemplo:

- "En promedio, 1 de cada 15 europeos es totalmente analfabeto"
 - Puede asustar, pero es un titular verídico. Entre los 500 millones de europeos, 36 millones probablemente no saben leer. Agregar que 36 millones tienen menos de 7 años. Datos de Eurostat. (Manual de Periodismo de Datos 1.0, 2014)
 - Cuando se escribe sobre un promedio, pensar:
 - ¿Un promedio de qué?
 - ¿La población de referencia es homogénea?

Informar con base a frecuencias naturales (1 de cada 100) es mucho más fácil de entender para los lectores que usar porcentuales (1%).

3. ¿En qué medida es confiable la información?

Ejemplo:

- "Tomar té reduce el riesgo de infarto"
 - Si los investigadores no toman en cuenta los factores de estilo de vida en sus estudios sobre el té, no nos dicen más que "los

ricos son más sanos y probablemente toman té". (Manual de Periodismo de Datos 1.0, 2014)

Tal como se describe en el ejemplo que propone el Manual de Periodismo de Datos, si los investigadores no buscan correlaciones (tomar té se correlaciona con hacer ejercicio), sus resultados son de escaso valor. Como periodista, tiene poco sentido cuestionar los resultados numéricos de un estudio, tales como el tamaño de la muestra, a menos que haya serias dudas al respecto. Sin embargo, es fácil de ver si los investigadores no tomaron en cuenta elementos relevantes de información.

12.3. Beneficios de la interpretación de datos

El análisis permite llegar a conclusiones válidas de forma a partir de los datos que se han recolectado y limpiado. La interpretación y análisis puede incluir las siguientes características:

- Identificación y explicación.
- Comparación y contraste
- Identificación de datos extraños.
- Predicciones.

El objetivo de la recopilación e interpretación es adquirir información útil, utilizable y tomar las decisiones más informadas posibles. El análisis e interpretación de datos, al final, ayuda a mejorar los procesos e identificar problemas.

¡Lectura recomendada!

Le invito a leer el post publicado por [QuestionPro.com "Análisis de datos"](#), puede acceder a través del siguiente enlace y el capítulo "Entender los datos" que propone el [Manual de Periodismo de Datos "Aprenda a manejar datos con 3 pasos simples"](#): los cuales le darán una aproximación al análisis de datos y cómo pueden ser entendidos por parte del equipo periodístico.



Actividades de aprendizaje recomendadas

Le invito a tomarse un tiempo libre el fin de semana y observar la película "MoneyBall": <https://youtu.be/ou-bhZM7eBo>, la cual le dará más insumos para comprender porqué los datos permiten tomar decisiones.

Consulte con su profesor:

De esta forma terminamos la unidad. Si tiene alguna duda por favor comuníquese con su profesor a través del EVA. A continuación, es importante desarrollar la autoevaluación, a fin de determinar su grado de comprensión de los contenidos analizados.

Toda unidad de su guía incluye preguntas de autoevaluación que le recomiendo realizar para reforzar sus conocimientos aprendidos.



Autoevaluación 12

Señale la respuesta correcta:

1. Complete: El ____ e interpretación de los datos está sujeto a la ____ extracción y limpieza de nuestra____. Debemos recalcar que es importarte saber interpretar los datos previos a su visualización.
 - a. análisis – correcta – base.
 - b. análisis – gestión – data.
 - c. análisis – correcto – data.
2. El análisis de datos puede ser de índole:
 - a. Comprensivo.
 - b. Cuantitativo.
 - c. Consistente.
3. El análisis de datos puede ser de índole:
 - a. Comprensivo.
 - b. Consistente.
 - c. Cualitativo.
4. Identifique a qué tarea de Excel corresponde la siguiente descripción: “Consiste en ordenar los datos de manera alfabética o destacando valores extremos (el más grande, el más pequeño, etc.)”
 - a. Clasificar.
 - b. Filtrar.
 - c. Tablas dinámicas.
5. Identifique a qué tarea de Excel corresponde la siguiente descripción: “Permite adicionar datos que estén en categorías”.
 - a. Clasificar.
 - b. Filtrar.
 - c. Tablas dinámicas.

- 6. Identifique a qué tarea de Excel corresponde la siguiente descripción: "Resulta necesario cuando se requiere estudiar un dato en particular a partir de una gran colección de datos".**
- a. Matemáticas y texto.
 - b. Filtrar.
 - c. Tablas dinámicas.
- 7. Identifique a qué tarea de Excel corresponde la siguiente descripción: "Uso de fórmulas (cálculos en hojas de cálculo)".**
- a. Matemáticas y texto.
 - b. Clasificar.
 - c. Tablas dinámicas.
- 8. Complete: "Lo que se VE es ____".**
- a. noticia
 - b. antecedente
 - c. opinión
- 9. Un beneficio de la interpretación de datos es:**
- a. La segmentación.
 - b. La estandarización.
 - c. La predicción.
- 10. Complete: "El objetivo de la recopilación e interpretación es adquirir información _____, _____, _____":**
- a. útil, utilizable y tomar las acciones...
 - b. útil, utilizable y tomar las decisiones...
 - c. útil, reutilizable y tomar las decisiones...

[Ir al solucionario](#)

Contestadas las preguntas verifique las respuestas correctas en el solucionario que encontrará al final de la guía. Si usted acertó al menos el 70% de respuestas correctas pase a la siguiente unidad, de lo contrario retome aquello que no está completamente asimilado.

Desarrolle la siguiente pregunta

A. ¿Por qué analizar e interpretar datos masivos?

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.

Su participación en el proceso de enseñanza y aprendizaje es esencial en nuestra asignatura. ¡Ahora, continuemos con la Unidad 13!

Resultado de aprendizaje 4

- Aplica técnicas y herramientas de visualización de datos enfocados en productos de información y comunicación.

El estudio tiene como eje central resolver las interrogantes: ¿Es importante la visualización? ¿Cómo visualizamos correctamente? ¿Qué tipo de herramienta puede emplear la visualización datos periodísticos? ¿Debo pensar como usuario al momento de la visualización?, que tendrán respuesta a través de la aplicación de técnicas y herramientas de visualización de datos, lo que permitirá, realizar productos enfocados y publicarlos en entornos digitales.

Contenidos, recursos y actividades de aprendizaje recomendadas



Unidad 13. Visualización

13.1. Importancia del dataviz



Para iniciar esta unidad le pido, por favor, que observe el siguiente video:
<https://youtu.be/jbkSRLYSOjo>.

Ahora bien, con la introducción al maravilloso mundo del *dataviz* gracias a Hans Rosling, empezamos.

La visualización o *data visualization* (*dataviz* en inglés) es el proceso de traducir grandes conjuntos de datos y métricas en cuadros, gráficos, pictogramas y otros elementos visuales. El resultado de la visualización de los datos facilita la identificación y el intercambio de tendencias en tiempo real, valores atípicos y nuevos conocimientos sobre la información representada en los datos.

A medida que aumenta el volumen de big data, más personas utilizan herramientas de visualización para acceder a información en su computadora y dispositivos móviles. Los empresarios, los analistas de datos, los científicos de datos y los periodistas de datos la utilizan para tomar decisiones y entregar al usuario información visualmente digerible y entendible. No es lo mismo publicar una tabla de Excel con cientos de celdas que publicar una infografía contextualizando la información de ese archivo.

La visualización de datos es importante:

- Tomar decisiones y actuar rápidamente sobre las decisiones que se tomaron.

- Podemos ver una imagen o panorama completo, descubrir conocimientos y ver patrones dentro de datos complejos sin depender de un científico de datos.
- Presentar datos significativos y compartir conocimientos con otros en una forma fácil de entender.
- Democratizar los datos al proporcionar una fuente de verdad y transparencia ante los usuarios.
- Capital a la hora de conocer a los públicos, segmentar públicos/audiencias es necesario "dibujar" su perfil a través de la visualización de datos.
- Contribuye y ayuda a visualizar tendencias en conjuntos de datos.
- Ayuda a absorber información de manera más rápida a diferentes niveles.
- Usa el lenguaje visual al que la sociedad (usuarios) está acostumbrada.
- Los patrones y relaciones se identifican de manera más rápida usando visualizaciones.

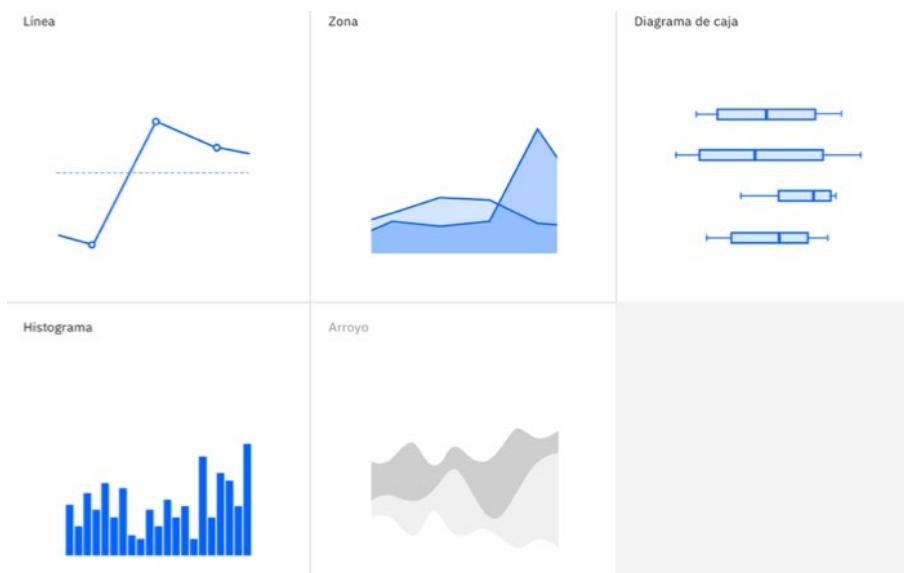
13.2. Categorías de gráficos

Los gráficos se dividen en categorías según sus objetivos, estética o características visuales (IBM, 2021). Dado que los gráficos pueden ser versátiles y usarse de diferentes maneras, los detalles y características de estas categorías se presentan a continuación:

- **Tendencias:** los gráficos de tendencias representan datos junto con la dimensión del tiempo. Se usan principalmente para realizar un seguimiento de los cambios durante períodos de tiempo de duración y escala variables. Los ejemplos incluyen gráficos de líneas, gráficos de áreas, histogramas y gráficos de corrientes.

Figura 63.

Ejemplos de gráficos para presentar tendencias

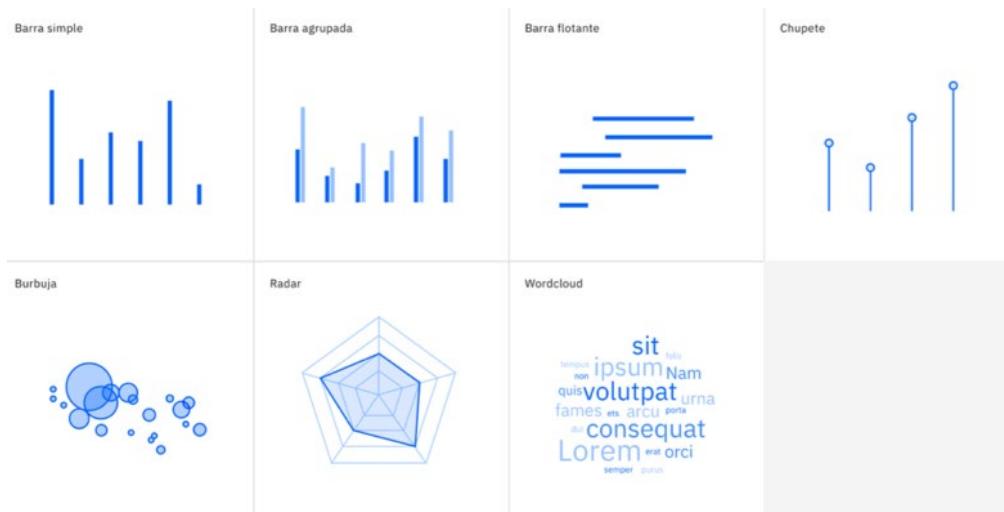


Nota. IBM, 2021.

- **Comparaciones:** Los gráficos diseñados para la comparación tienen como objetivo visualizar las diferencias entre los elementos. La mayoría de las veces, las comparaciones se basan en la capacidad del ojo humano para identificar formas más largas o más grandes con muy poco o ningún esfuerzo. El posicionamiento lado a lado y la alineación de los elementos visuales facilitan aún más las comparaciones. Estos gráficos se utilizan para datos basados en tiempo, por ejemplo, unidades vendidas por día u horas trabajadas por mes, ingresos por mercado o unidades vendidas.

Figura 64.

Ejemplos de gráficos para presentar comparaciones

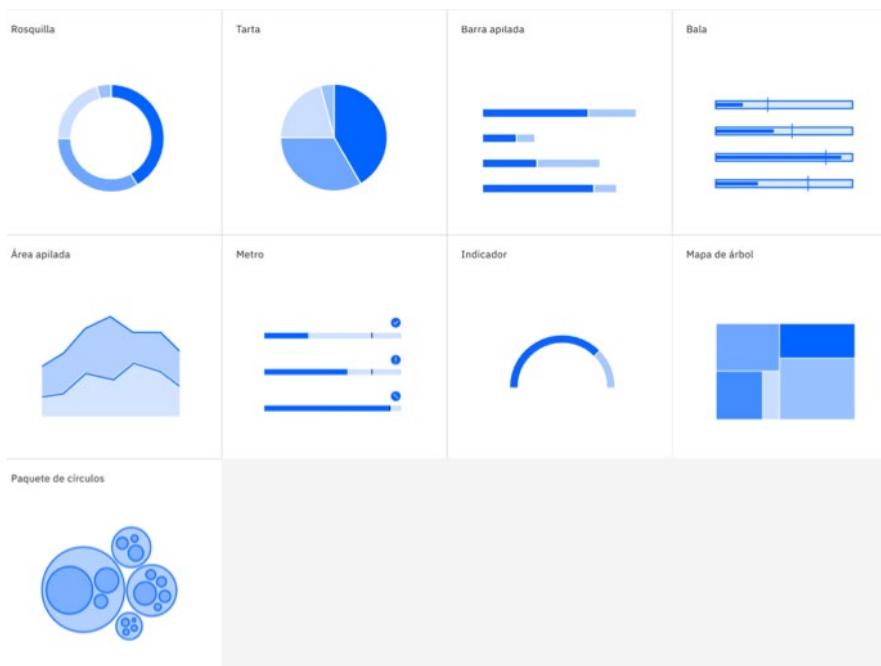


Nota. IBM, 2021.

- **Parte de un todo:** El objetivo de estos gráficos es mostrar la subdivisión interna de un valor entre diferentes categorías o grupos. Se utilizan principalmente para representar porcentajes, pero también se pueden utilizar para valores absolutos. Su función no depende de las formas gráficas utilizadas, como circular, pastel/tarta, cuadrado, etc. Estos gráficos se utilizan para datos categorizados, por ejemplo, subdivisión de ingresos por producto o porcentaje de usuarios por navegador.

Figura 65.

Ejemplos de gráficos para presentar la parte de un todo



Nota. IBM, 2021.

- **Correlaciones:** Estos gráficos son más adecuados para resaltar la posible correlación entre dos o más indicadores y cómo podrían afectarse entre sí. Los gráficos de correlación tienen el objetivo final de facilitar al ojo humano la detección de comportamientos combinados. Estos gráficos se utilizan para datos multidimensionales, por ejemplo, la correlación entre la duración de la llamada telefónica y la satisfacción del cliente.

Figura 66.

Ejemplos de gráficos para presentar correlaciones

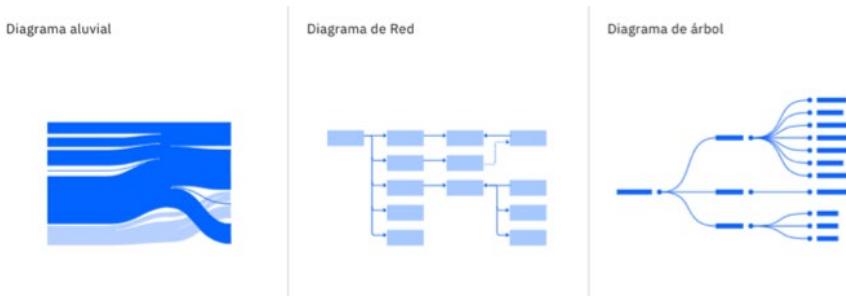


Nota. IBM, 2021.

- **Relaciones y conexiones:** Los gráficos incluidos en esta categoría representan jerarquías. La intención es explicar el papel de un elemento dentro de un ecosistema u observar la naturaleza interna de un sujeto en diferentes fases y estados de un proceso. Estos gráficos se utilizan para datos categorizados, por ejemplo, país de origen del solicitante de asilo y género. También se utilizan para datos multidimensionales, por ejemplo, número de usuarios activos por fase de prueba del cibermedio.

Figura 67.

Ejemplos de gráficos para presentar relaciones y conexiones



Nota. IBM, 2021.

- **Mapas:** Los mapas son la forma más fácil e inmediata de comunicar información geolocalizada. Los mapas permiten al usuario reconocer áreas y lugares, comprender el contexto geográfico del tema e identificar patrones, todo basándose en la posición de los elementos. Estos gráficos se utilizan para datos geográficos, por ejemplo, votantes por país, provincia, ciudad o salario promedio por sectores o barrios.

Figura 68.

Ejemplos de gráficos para presentar mapas



Nota. IBM, 2021.

Continuemos con las diversas etapas para la construcción de una visualización.

13.3. Etapas para crear una visualización

Entre las etapas que [Alberto Cairo](#) (2014), experto en visualización de datos, constan y lo podrán observar en el siguiente recurso:

[Etapas para crear una visualización.](#)

13.4. Catálogo de visualización

Para conocer qué gráfico aplicar conforme al objetivo que nos hemos planteado, le invito a revisar los siguientes catálogos, los cuales muestran claramente el gráfico y la herramienta en dónde se puede elaborar la visualización de las series más vistas:

[Ejemplo de visualización de las series más vistas.](#)

Estimado estudiante, le invito a leer el post sobre la siguiente figura el cual está basado en los datos de las recomendaciones de un sitio de recomendaciones: <https://fortune.com/2019/09/05/map-favorite-tv-shows-by-country-imdb/>

A continuación se muestra el mapa de los programas de TV favoritos en EEUU.

Mapa de los programas de TV favoritos en EE.UU.

Entre los catálogos más conocidos están:

- *The Data Visualisation Catalogue* (Catálogo de Visualización de Datos) <https://datavizcatalogue.com/>
- Depic Data Studio: <https://depicdatastudio.com/charts/>
- Data Viz Project: <https://datavizproject.com/>

A través de la siguiente imagen miremos cómo "*The Data Visualisation Catalogue*", presenta: descripción, funciones, similitud con otros gráficos, su anatomía y las herramientas (*tools*) para la elaboración del gráfico: <https://datavizcatalogue.com/>

Figura 69.

Ejemplo de la anatomía del gráfico



Nota. DVC, <https://datavizcatalogue.com/>. Elaboración propia.

13.5. Herramientas

Entre las herramientas más empleadas para la visualización de datos y que también son empleadas en equipos editoriales, detallamos:

- Datawrapper: <https://app.datawrapper.de/>
- Onodo: <https://onodo.org/>
- Infogram: <https://infogram.com/>
- Flourish: <https://app.flourish.studio/>
- Tableau Public: <https://public.tableau.com/>
- Power Bi: <https://powerbi.microsoft.com/>

Para todas las herramientas que empleará en su proyecto de visualización de datos periodísticos deberá crear una cuenta y/o registro. Si bien son de acceso gratuito algunas funcionalidades solo se habilitan cuando los usuarios pagan una suscripción.

Gracias a la Web Social (Web 2.0), este tipo de herramientas son completamente intuitivas, donde lo que se precisa es estandarizar la base de datos y acoplarlas a las diversas plantillas (al estilo: canva.com o wix.com). Es por ello que, le invito a revisarlas, a perder el miedo y sobre todo, a realizar una práctica recurrente.

- A continuación un ejemplo con Datawrapper:

Lo importante en este tipo de visualización es elegir la fuente de datos (ejemplo: base de datos de la Asamblea Nacional de Ecuador), limpiarlos en caso exista alguna inconsistencia y luego ingresarlos en la herramienta de visualización que consideremos oportuna. Podemos, además, embeberlos con el código HTML directamente en nuestro sitio web y así puedan ser vistos por los usuarios de forma dinámica, generando así interacción.

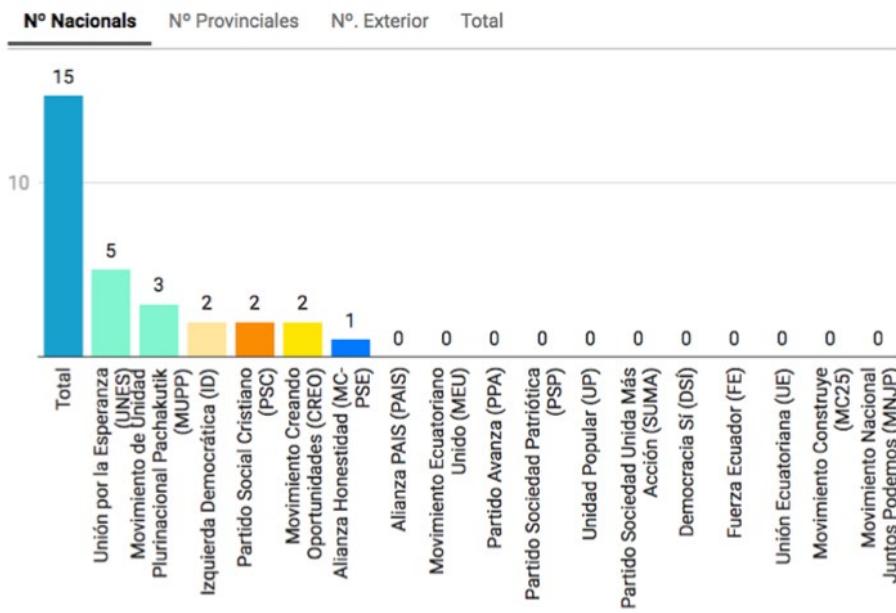
[Los 4 pasos del proceso de elaboración de un gráfico con la herramienta Datawrapper.](#)

Figura 70.

Gráfico empleando la herramienta Datawrapper

Asamblea Ecuador,

Composición de la Asamblea Nacional de Ecuador según partidos y/o movimientos políticos, periodo 2021 - 2025.



Datos tomados de Wikipedia y validados con Asamblea

Gráfico: Gabriela Coronel Salas • Fuente: [Asamblea Ecuador](#); • Creado con [Datawrapper](#)

Nota. Asamblea Nacional de Ecuador, 2021. Elaboración propia.

Aquí, otros ejemplos con Flourish:

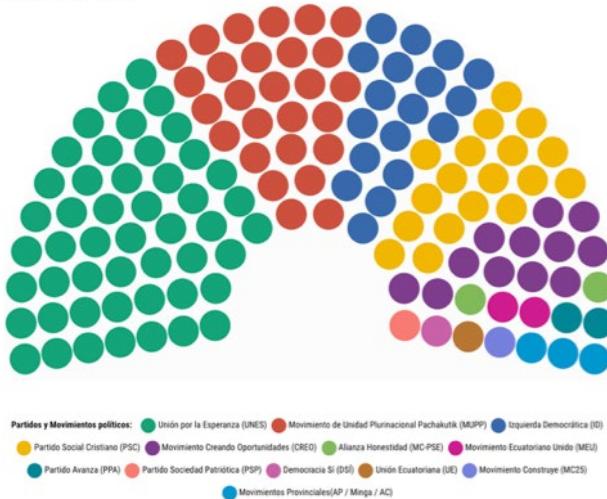
Figura 71.

Gráfico empleando la herramienta *Flourish*

Asamblea Ecuador

Composición de la Asamblea Nacional de Ecuador según partidos y/o movimientos políticos, periodo 2021 - 2025.

El Pleno es el máximo órgano de decisión de la Asamblea Nacional. Está integrado por la totalidad de las y los asambleístas. Para la instalación y funcionamiento del Pleno se requerirá la presencia de la mayoría absoluta de las y los miembros de la Asamblea Nacional.

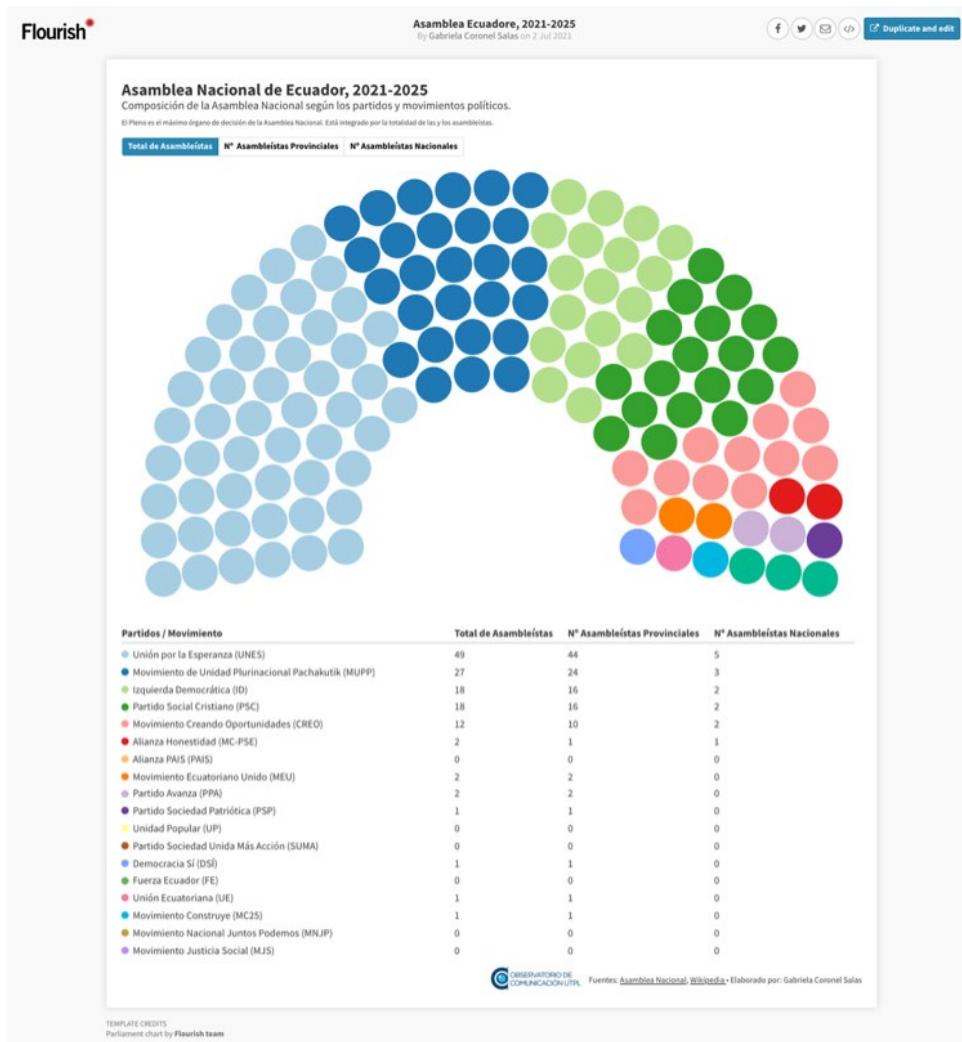


Fuente: [Asamblea Ecuador, Wikipedia](#)
Realizado por: Gabriela Coronel Salas

Nota. Asamblea Nacional de Ecuador, 2021. Elaboración propia.

Figura 72.

Gráfico empleando la herramienta *Flourish*



Nota. Asamblea Nacional de Ecuador, 2021. Elaboración propia. <https://public.flourish.studio/visualisation/6617639/>

¡Recursos recomendados!

Comparto e invito a revisar los tutoriales y ejemplos para la elaboración de gráficos:

- Datawrapper: <https://blog.datawrapper.de/datawrapper-en-espanol/>
- Flourish:
 - <https://flourish.studio/examples/>
 - <https://youtu.be/0Vg8rBvEgss>

- Storybench: <https://www.storybench.org/>

En nuestro espacio de tutorías realizamos ejercicios con este tipo de herramientas. No olvide conectarse en el día y horario establecido para la asignatura, esta información estará en el EVA.

Consulte con su profesor:

De esta forma terminamos la unidad. Si tiene alguna duda por favor comuníquese con su profesor a través del EVA. A continuación, es importante desarrollar la autoevaluación, a fin de determinar su grado de comprensión de los contenidos analizados.

Toda unidad de su guía incluye preguntas de autoevaluación que le recomiendo realizar para reforzar sus conocimientos aprendidos.



Autoevaluación 13

Señale la respuesta correcta:

- 1. Se conoce la visualización de datos, como:**
 - a. *Dataviz.*
 - b. *Datatools.*
 - c. *Datavill.*

- 2. Complete. La visualización de datos es importante: “_____ y actuar rápidamente sobre...”**
 - a. Tomar decisiones
 - b. Gestionar los datos
 - c. Filtrar

- 3. Complete. La visualización de datos es importante: “_____al que la sociedad (usuarios) está acostumbrada.**
 - a. Usa el lenguaje verbal
 - b. Usa el lenguaje web
 - c. Usa el lenguaje visual

- 4. Entre las categorías de gráficos, tenemos:**
 - a. Comprensiones.
 - b. Tendencias.
 - c. Conexiones.

- 5. Entre las categorías de gráficos, tenemos:**
 - a. Correlaciones.
 - b. Jerarquías.
 - c. Categorías.

- 6. Entre las categorías de gráficos, tenemos:**
 - a. Geografía.
 - b. Barras.
 - c. Mapas.

7. Un ejemplo de catálogo de datos es:

- a. Tabula.
- b. Data Viz Project.
- c. Catálogo.ec.

8. Una de las herramientas para visualización de datos es:

- a. Onodo: <https://onodo.org/>
- b. Depic Data Studio: <https://depicdatastudio.com/charts/>
- c. Open Refine <https://openrefine.org/>

9. Una de las herramientas para visualización de datos es:

- a. Excel: <https://excel.com>
- b. Google Trends: <http://trends.google.com>
- c. Flourish: <https://app.flourish.studio/>

10. Qué autor propone 7 etapas para crear una visualización:

- a. Annie Ker.
- b. Alberto Cairo.
- c. Alberto Emery.

[Ir al solucionario](#)

Desarrolle la siguiente pregunta

- A. ¿Cuáles son para Usted las ventajas de visualizar datos masivos en el periodismo?

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.

Su participación en el proceso de enseñanza y aprendizaje es esencial en nuestra asignatura. ¡Ahora, continuemos con la Unidad 14!



Semana 15

Unidad 14. Productos periodísticos



Toma de decisiones – Conocimiento

Para cerrar esta Unidad, y por ende la revisión del bimestre, le invito a conocer de las manos de los expertos cómo presentar datos a los usuarios y más aún, enfocados en el ámbito periodístico. No olvide que se encuentra en la etapa de toma de decisiones, en sí, se ha transitado desde la ubicación de los datos, luego su limpieza, su análisis y por supuesto, su visualización, hasta llegar al conocimiento. Ante esto, le invito a reflexionar sobre las siguientes interrogantes: ¿Cuándo necesitamos visualizar una historia? ¿Los datos nos pueden hacer cambiar de opinión? ¿Cuándo necesitamos visualizar una historia?

14.1. Contar historias con datos

*Debemos saber QUÉ queremos contar.
Para luego pensar el CÓMO vamos a hacerlo.
Y mediante QUÉ HERRAMIENTAS lo haremos.*

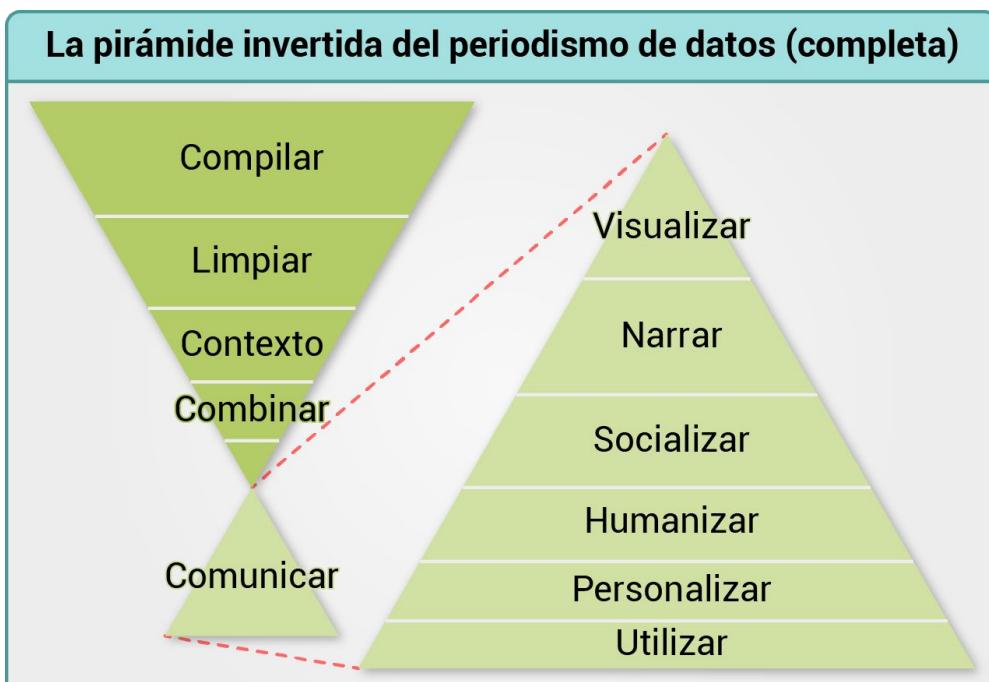
Recuerde que al inicio del primer bimestre (Unidad 4) se hizo mención del tipo de productos basados en datos, entre los que tenemos:

- Artículos basados en datos.
- Visualizaciones interactivas.
- Set de datos.
- Aplicaciones o *news apps*.

Este tipo de productos nos permiten llegar a los usuarios de una manera más cercana e interactiva. Para ello, retome el conocimiento sobre la escritura periodística y a su pirámide invertida. En el caso expreso del periodismo de datos existe una forma de contar historias en la web a través de ésta, por ejemplo, tal como consta en la figura 71:

Figura 73.

Pirámide invertida del periodismo de datos



Nota. Paul Bradshaw citado en (Grassler, 2017, p. 84)

Para John Bones del [Verdens Gang](#), muchas de las ocasiones no se necesita visualizar una historia, incluso, a veces no es necesario pero los medios lo suelen hacer para que los lectores o usuarios puedan dimensionar un hecho. Si bien las historias que contienen un gran volumen de datos a menudo necesitan de una visualización. Pero debemos ser bastante críticos al elegir qué tipo de datos vamos a presentar. Conocemos todo tipo de cosas cuando informamos sobre algo, pero, ¿qué necesita saber realmente el lector sobre la historia? Quizás baste una tabla, o un

gráfico simple que muestra un proceso que va del año A al año C. Cuando se trabaja con periodismo de datos, el objetivo no es necesariamente presentar grandes cantidades de datos. Se trata de periodismo.

Una buena visualización es como una buena imagen. Se entiende de qué se trata con solo mirar uno o dos instantes. Cuanto más se mira la visualización, más se ve. La visualización es mala cuando el lector no sabe por dónde empezar o dónde termina, y cuando la visualización está sobrecargada de detalles. En este caso, quizás una pieza de texto sería mejor, concluye Bones. (Manual de Periodismo de Datos 1.0, 2014)

Los datos son valiosos dado que el acceso a ellos tiene el potencial de clarificar cuestiones de un modo que genere resultados. Pero el mal manejo de los datos puede ubicar los hechos en una estructura opaca que no comunica nada. Si no promueven la discusión o aportan una comprensión en contexto, los datos pueden ser de limitado valor para el público.

14.2. Storytelling

Storytelling se traduce como "el arte de contar historias", porque es importante conectar emocionalmente a través de una historia, dejar que te lleguen, te toquen el corazón, la mente, el cuerpo y el espíritu: lo racional y lo instintivo. Una buena historia crea una emoción positiva que inspira a la gente a tomar una acción. No se trata de mostrar una marca, se trata de contar qué experiencia sentirá el usuario con ESE producto/servicio. Para ello se hace imprescindible "pensar como usuario".

A lo largo de la carrera ha escuchado y estoy segura de que conoce sobre este término, así que me permito compartir el siguiente post que puede ser de utilidad para reforzar su conocimiento sobre el tema: <https://rockcontent.com/es/blog/que-es-storytelling/>

14.2.1. Plataformas para crear contenido

Para todo ello se detallan una serie de herramientas y plataformas que permitirán crear contenido digital:

[Plataformas para la creación de contenido.](#)

Y para aplicar el *storytelling* a través de las diversas plataformas (sumadas las de visualización) es importante conocer a la audiencia a la cual se

destinará nuestro producto periodístico; por ello, miremos a continuación los tipos de usuarios.

14.3. Tipos de usuarios (público)

Para crear nuestros productos periodísticos basados en datos debemos pensar en el destinatario final, es decir en nuestro público, en nuestra audiencia, en sí, en nuestro usuario, por ello existe la siguiente clasificación:

- **Usuarios ocasionales:** son usuarios que quieren información de modo simple y rápido. Les interesa tener una idea de los datos, no un análisis detallado. Podemos dirigirnos a ellos vía tweet o gráficos interactivos.
- **Usuarios activos:** usuarios que estimulan el debate y usan los datos para incrementar su conocimiento de un área determinada o cuestionan los supuestos de los datos. A estos usuarios, hay que proveerles mecanismos de retroalimentación y la posibilidad de compartir su visión con sus pares vía las redes sociales.
- **Acaparadores de datos:** estos usuarios quieren datos en crudo para hacer visualizaciones o análisis. Simplemente les damos los datos para sus propósitos.

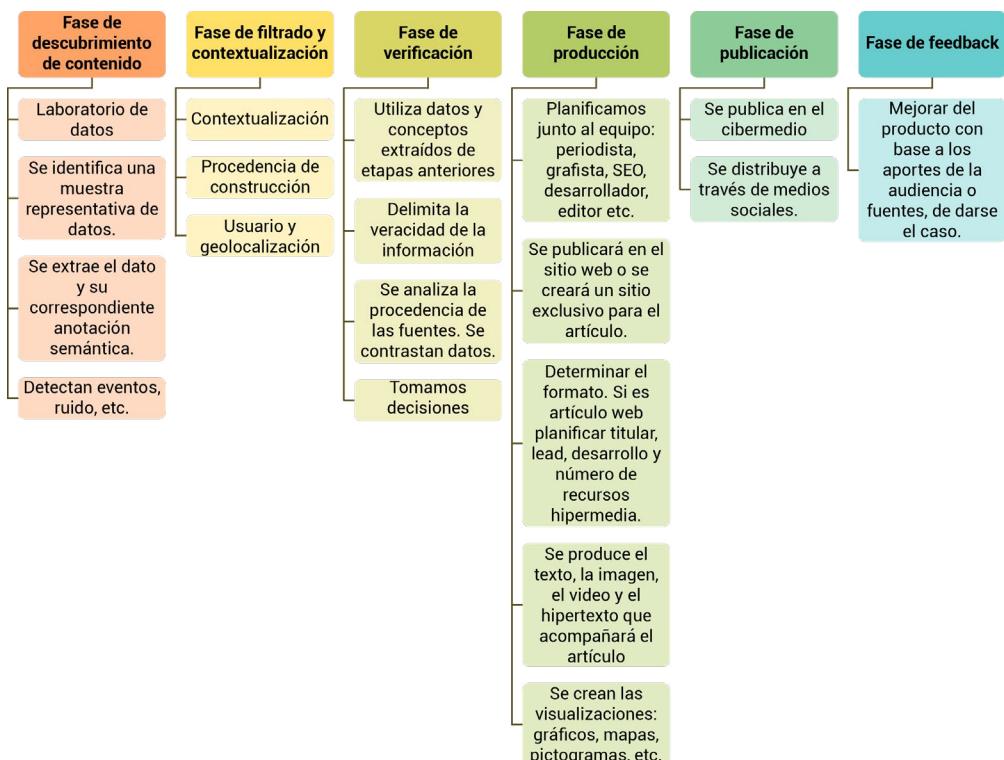
¡Video recomendado!

Le invito a observar la charla TED, "["Datos, arte y nuevas formas de comprender el mundo"](#)", en la cual, Pau García, menciona que "en tiempos de Big Data es cada vez más importante pensar en cómo se muestra la información. Gráficos y tablas de datos no transmiten la importancia ni la historia detrás de estos datos. Mediante varios ejemplos impactantes destaca la importancia de la forma de visualización para que tenga el impacto deseado en el espectador."

14.4. Ejemplos de productos periodísticos basados en datos

Veamos las fases de trabajo para la implementación de un producto con datos.

Figura 74.
Fases de trabajo con datos



Nota. Ideas de autores especializados en datos y aporte de la autora de la guía,
Elaboración propia.

Algunos medios internacionales han creado un laboratorio (Labs) de innovación donde incluyen el desarrollo de nuevos productos enfocados en una narrativa hipermedia. Como se aprecia en la figura, es indispensable un trabajo en conjunto con varios perfiles profesionales del medio y sobre todo, cumplir con los preceptos básicos del periodismo: verificar y contrastar la información, más aún si nos enfocamos en grandes volúmenes de datos captados desde internet.

Estimado estudiante, en conjunto, le invito a observar algunos ejemplos de productos periodísticos de importantes medios de comunicación que han visto en el periodismo de datos un valor agregado a la hora de llegar a sus audiencias.

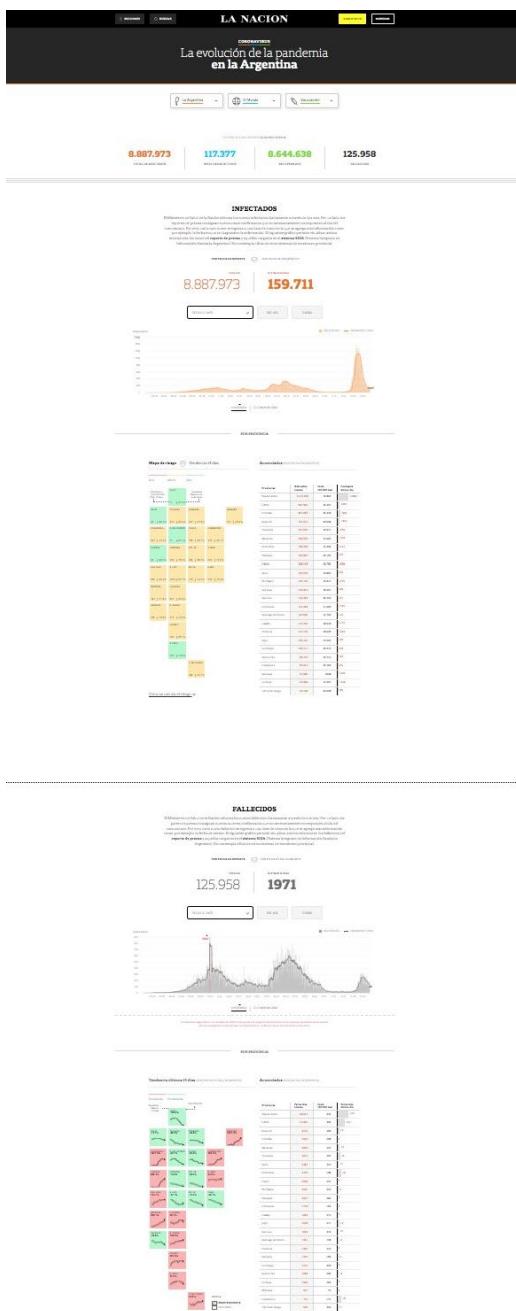
Diario *La Nación* es un referente iberoamericano en el periodismo de datos, hace algunos años decidió crear su propio laboratorio y sección

denominada "La Nación Data", la cual les invito a explorar: <https://www.lanacion.com.ar/data/>

- ***La Nación La evolución de la pandemia en la Argentina*** <https://www.lanacion.com.ar/sociedad/en-detalle-infectados-fallecidos-coronavirus-argentina-nid2350330/>

Figura 75.

Ejemplo de aplicación de noticias con datos



Nota. La Nación La evolución de la pandemia en la Argentina <https://www.lanacion.com.ar/sociedad/en-detalle-infectados-fallecidos-coronavirus-argentina-nid2350330/>

Es indispensable que navegue (haciendo clic en los enlaces) en los artículos periodísticos que se han colocado a modo de ejemplo en la asignatura.

- **La Nación: Subsidio transporte** <https://www.lanacion.com.ar/economia/por-dia-el-subsidio-al-transporte-llego-a-154-millones-nid1857603/>

Figura 76.
Ejemplos de artículos con datos



Por Diego Cabot

Conforme a los criterios de  The Trust Project

Conoce The Trust Project

Nota. La Nación. Subsidio transporte <https://www.lanacion.com.ar/economia/por-dia-el-subsidio-al-transporte-llego-a-154-millones-nid1857603/>

Diario *El País* tiene una trayectoria en el trabajo con datos y salud, por ejemplo en 2021 el equipo de Nuevas Narrativas, en manos de sus periodistas Mariano Zafra y Javier Salas ganaron el premio “Malofiej”,

llamados los "Pulitzer de la infografía", justamente por el artículo que a continuación veremos.

Figura 77.

Tuit de Malofiej anunciando a los ganadores

 **Malofiej**
@malofiej

En respuesta a @malofiej

Best of Show Digital for El País (Spain) for the project 'A room, a bar and a classroom: how the coronavirus is spread through the air' #malofiej29
english.elpais.com/society/2020-1...

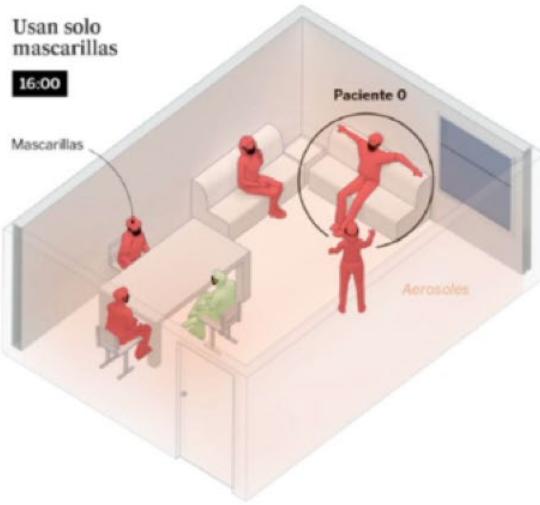
Traducir Tweet

Malofiej 29

'A room, a bar and a classroom: how the coronavirus is spread through the air'

EL PAÍS


Best of Show
Digital



10:45 a. m. • 17 jun. 2021 • Twitter Web App

Nota. Twitter: <https://twitter.com/malofiej/status/1405552095160373248>

- **El País: Un salón, un bar y una clase: así contagia el coronavirus en el aire** <https://elpais.com/especiales/coronavirus-covid-19/un-salon-un-bar-y-una-clase-asi-contagia-el-coronavirus-en-el-aire/>

Figura 78.

Ejemplos de artículos con datos

≡ EL PAÍS Materia LA CRISIS DEL CORONAVIRUS / SUSCRÍBETE INICIAR SESIÓN

Un salón, un bar y una clase: así contagia el coronavirus en el aire

Los interiores son más peligrosos, pero es posible minimizar los riesgos si se ponen en juego todas las medidas disponibles para combatir el contagio por aerosoles. Estas son las probabilidades de infección en estos tres escenarios cotidianos dependiendo de la ventilación, las mascarillas y la duración del encuentro

Reunión social en un salón

12:00

Paciente 0

20 m²

Nota. El País: Un salón, un bar y una clase: así contagia el coronavirus en el aire
<https://elpais.com/especiales/coronavirus-covid-19/un-salon-un-bar-y-una-clase-asi-contagia-el-coronavirus-en-el-aire/>

- **El País: Mapa de la renta de los españoles:** <https://elpais.com/economia/2021-04-29/el-mapa-de-la-renta-de-los-espanoles-calle-a-calle.html>

Figura 79.

Ejemplos de artículos con datos



El mapa de la renta parte España en dos. Los datos por secciones censales, zonas donde viven menos de 1.500 personas, revelan que hay 250 vecindarios donde la mitad de la gente gana más de 36.750 euros y otros 150 donde la mayoría no llega a 7.500. Entre unos y otros están las más de 35.000 secciones censales del resto del territorio (que incluye Navarra y las Islas Baleares excluidas). De los 250 más ricos, 180 están en las provincias. Los 250 más ricos se encuentran, sin excepciones, en Madrid y Barcelona y hoy que llegar hasta la posición 150 de esta clasificación para encontrar un sitio fuera de estas provincias. Es una de las secciones de Pta Del Río, un distrito de viviendas altas en la ciudad de Valencia.

Los datos de cada sección, relativos a 2018, representan la renta neta por persona (o lo que se conoce como unidad de consumo). Esta dato se calcula sumando todas las rentas de un hogar y dividéndolas por el número de personas que conviven en él. Una persona adulta es considerada como un adulto cuanta como 1, el segundo como 0,5 y los menores como 0,3. Es una forma estandarizada de conseguir la renta de un hogar por su tamaño. El INE ofrece datos de renta media y mediana, pero la segunda es más robusta y más fácil de entender: es la renta de la persona que queda en medio cuando ordenamos a todos los vecinos de cada población a media renta.

En el mapa se muestra en qué sección se encuentra cada sección, entre todas las de España y las de su misma comunidad. Así es posible ver algunas zonas que siendo las más ricas a nivel nacional no destacan especialmente dentro de su comunidad, como algunos vecindarios del distrito Centro de Madrid.

Los grandes contrastes del mapa

El INE, para respetar el secreto estadístico, no ofrece información por encima de 36.750 euros. En ese umbral se encuentran varias secciones del norte de Barcelona, en los arrabales de Les Corts, Sant Gervasi, y de Espàgues de Llobregat, y en el distrito de Sarrià. En Madrid, se sitúan las principales urbanizaciones de La Línea de la Moraleja, en Alcalá de Henares o Madrid y varios vecindarios de distritos de la capital, como Salamanca o Chamberí. Todas estas secciones han visto crecer sus rentas medianas alrededor de un 22% con respecto a 2015, en línea con la media española.

En el lado opuesto se encuentran los vecindarios más pobres de España, donde la renta no alcanza los 6.000 euros anuales. Son secciones en distritos de Pinares (Granada), Puertollano (Ciudad Real), Melilla y Ceuta o Almería, pero también en ciudades más grandes como Málaga o Sevilla.

La mejor forma de identificar estos contrastes es navegar por el mapa que encabeza esta noticia. Estos son algunos de los más llamativos.

Madrid. En Madrid destacan las diferencias que marca la M-30, que se agudizan en el sur de la capital. Pero también en numerosos barrios dentro de la conurbación. Tetuán (en el norte) y Lavapiés, se componen en su mayoría de secciones que se encuentran en el 25%-30% más pobre de la comunidad.

Fuente: El País, Mapa de la renta de los españoles: <https://elpais.com/economia/2021-04-29/el-mapa-de-la-renta-de-los-espanoles-calle-a-calle.html>

Recuerde, es indispensable que navegue (haciendo clic en los enlaces) en los artículos periodísticos que se han colocado a modo de ejemplo en la asignatura.

- **El País: Escuela, conciliación y pobreza** <https://elpais.com/espana/2021-07-03/escuela-conciliacion-y-pobreza-los-problemas-de-las-familias-jovenes-en-espana.html>

Figura 80.

Ejemplos de artículos con datos



Nota. El País, Mapa de la renta de los españoles: <https://elpais.com/espana/2021-07-03/escuela-conciliacion-y-pobreza-los-problemas-de-las-familias-jovenes-en-espana.html>

- ***El País: ¿Quién gana la Eurocopa?*** https://elpais.com/politica/2021/06/25/actualidad/1624630452_675971.html

Figura 81.

Ejemplos de artículos con datos



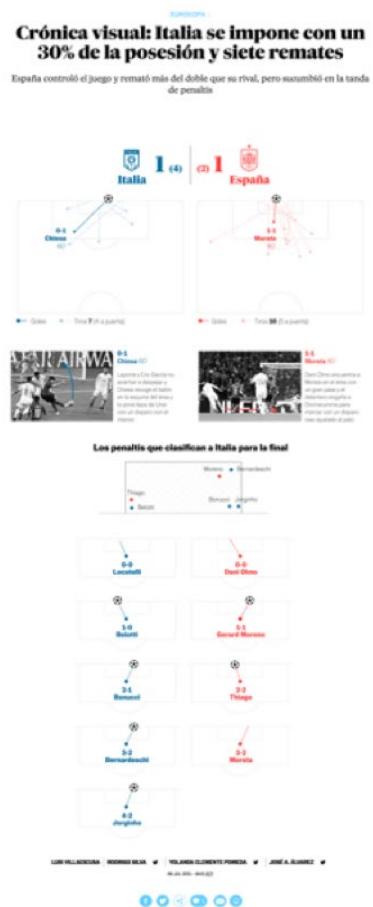
Nota. El País: ¿Quién gana la Eurocopa? https://elpais.com/politica/2021/06/25/actualidad/1624630452_675971.html

Es indispensable que navegue (haciendo clic en los enlaces) en los artículos periodísticos que se han colocado a modo de ejemplo en la asignatura.

- **El País: Crónica visual** <https://elpais.com/deportes/eurocopa-futbol/2021-07-06/cronica-visual-italia-se-impone-con-un-30-de-la-posesion-y-siete-remates.html>

Figura 82.

Ejemplos de artículos con datos



La posesión

Nota. El País: Crónica visual <https://elpais.com/deportes/eurocopa-futbol/2021-07-06/cronica-visual-italia-se-impone-con-un-30-de-la-posesion-y-siete-remates.html>

- **El País: No respires el aire de otro** <https://elpais.com/ciencia/2021-03-28/no-respires-el-aire-de-otro-como-esquivar-el-coronavirus-en-interiores.html>

Figura 83.

Ejemplo de artículos con datos

≡ EL PAÍS

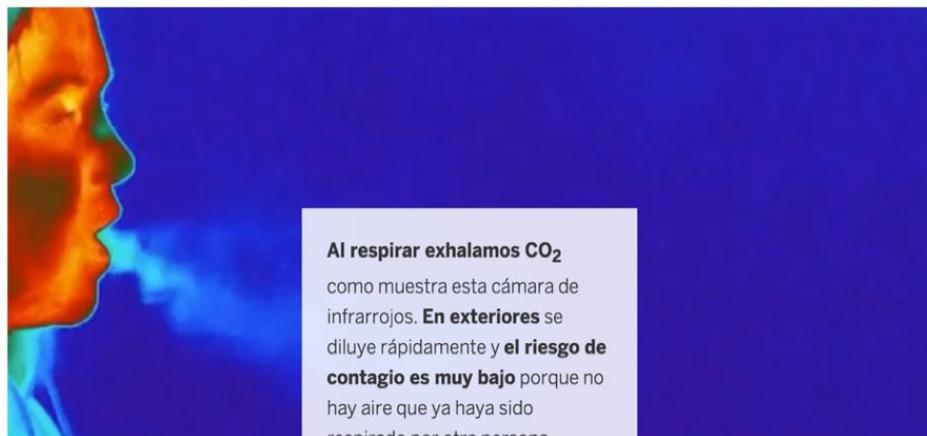
Materia

E+ SUSCRÍBETE

INICIAR SESIÓN

No respires el aire de otro: cómo esquivar el coronavirus en interiores

Ventilación continua y control permanente del CO₂: dos claves para evitar el contagio en espacios cerrados al diluir la presencia de partículas contagiosas en suspensión



Nota. El País: No respires el aire de otro <https://elpais.com/ciencia/2021-03-28/no-respires-el-aire-de-otro-como-esquivar-el-coronavirus-en-interiores.html>

No olvide, es indispensable que navegue (haciendo clic en los enlaces) en los artículos periodísticos que se han colocado a modo de ejemplo en la asignatura.

- **El País: Las doce letras que cambiaron el mundo** <https://elpais.com/especiales/coronavirus-covid-19/el-genoma-del-nuevo-coronavirus/>

Figura 84.

Ejemplo de artículo con datos

CORONAVIRUS

ccu cgg cgg gca

Las doce letras que cambiaron el mundo

El genoma del nuevo coronavirus esconde una brevíssima secuencia sospechosa de ser la principal culpable de su insólita capacidad de contagio y de su virulencia

MANUEL ANSEDE | ARTUR GALOCHA | MARIANO ZAFRA
11 MAY 2020 - 01:42 ECT

El temible enemigo que ha obligado a miles de personas a esconderse en sus casas es una **minúscula pelotita de unas 70 milonésimas de milímetro**. El nuevo coronavirus, llamado SARS-CoV-2 por los científicos, es tan pequeño respecto a un ser humano como una gallina respecto a todo el planeta Tierra. Ese es el gran adversario de la humanidad. El virus es apenas un brevíssimo mensaje escrito con combinaciones de las mismas cuatro letras. Cada una de ellas es la inicial de un compuesto químico con diferentes cantidades de carbono, hidrógeno, nitrógeno y oxígeno. Con estas cuatro letras (a, u, g, c) está escrito el texto que ha matado a más de 275.000 personas desde que se detectó su existencia hace poco más de cuatro meses.



SARS-CoV-2

C5H5N5
Adenina

C4H4N2O2
Uracilo

C6H6N5O
Citosina

Este material genético que unos virus heredan de otros, denominado ARN, funciona como un libro de instrucciones para fabricar las proteínas que forman el SARS-CoV-2.



CCU UGG CGG GCA



C5H5N5O
Guanina

C

Nota. El País: Las doce letras que cambiaron el mundo. https://elpais.com/elpais/2020/05/09/ciencia/1589059080_203445.html

Le invito a identificar a los periodistas de datos, por ejemplo, por ello, comparto cuatro de los periodistas de diario *El País* a los que puede seguir en sus cuentas de Twitter:



Kiko Llaneras ✅
@kikollan
Doctor en ingeniería que escribe en @el_pais. Datos, gráficos y encuestas. De @Politikon_es. ¡Con newsletter! 📰
plus.elpais.com/newsletters/in...

Relevant people



daniele grasso
@danielegrasso
Italian in Madrid. Data journalism at @el_pais. @ICIJorg member. 🇮🇹
dgrasso@elpais.es



Mariano Zafra
@Mariano_Zafra
Following
Redactor jefe de Narrativas Visuales en @el_pais. Me gustan los #mapas y #gráficos que ayudan a entender cómo funciona el mundo. @WSJ, @univision @elmundoes

- Francisco (Kiko) Llaneras: <https://elpais.com/autor/francisco-llaneras-estrada/>
- Daniele Grasso: <https://elpais.com/autor/daniele-grasso/>
- Álvaro Sánchez: <https://elpais.com/autor/alvaro-sanchez-lopez/>
- Mariano Zafra: <https://elpais.com/autor/mariano-zafra/.>

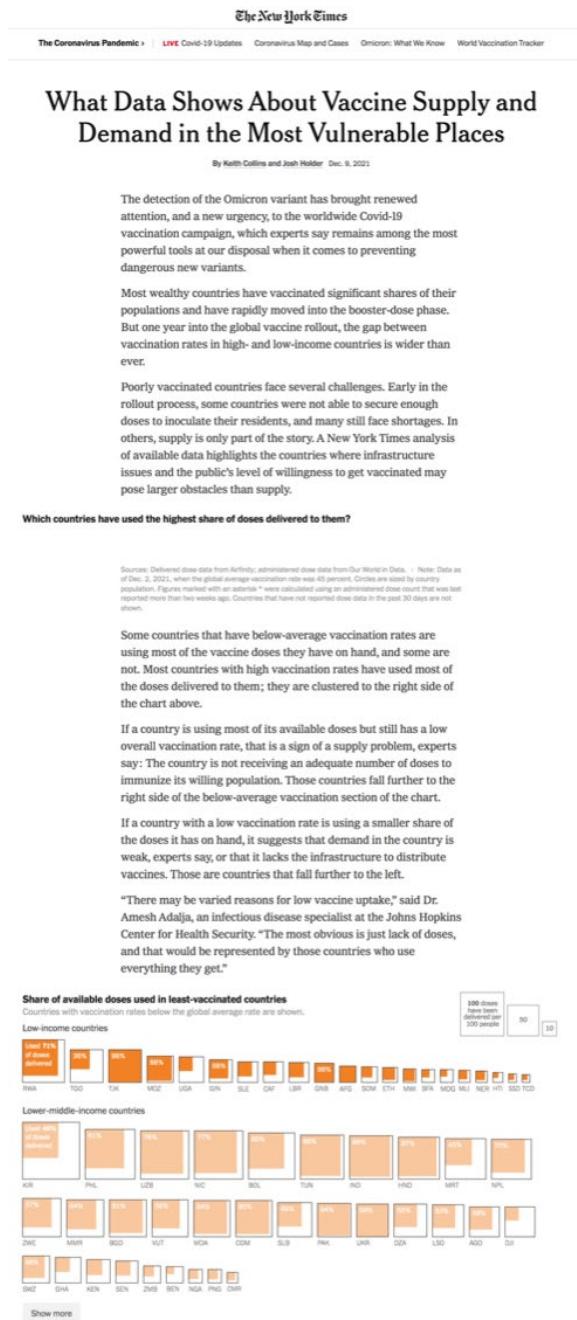
A través de la siguiente lectura podrá conocer más sobre “Estándares de calidad en el periodismo de datos: fuentes, narrativas y visualizaciones en los Data Journalism Awards 2019.” <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2020.may.28>

Otro referente y modelo por seguir en la construcción de nuevas narrativas digitales es *The New York Times*.

- *NYTimes: What Data Shows About Vaccine Supply and Demand in the Most Vulnerable Places* <https://www.nytimes.com/interactive/2021/12/09/world/vaccine-inequity-supply.html>

Figura 85.

Ejemplo de artículo con datos

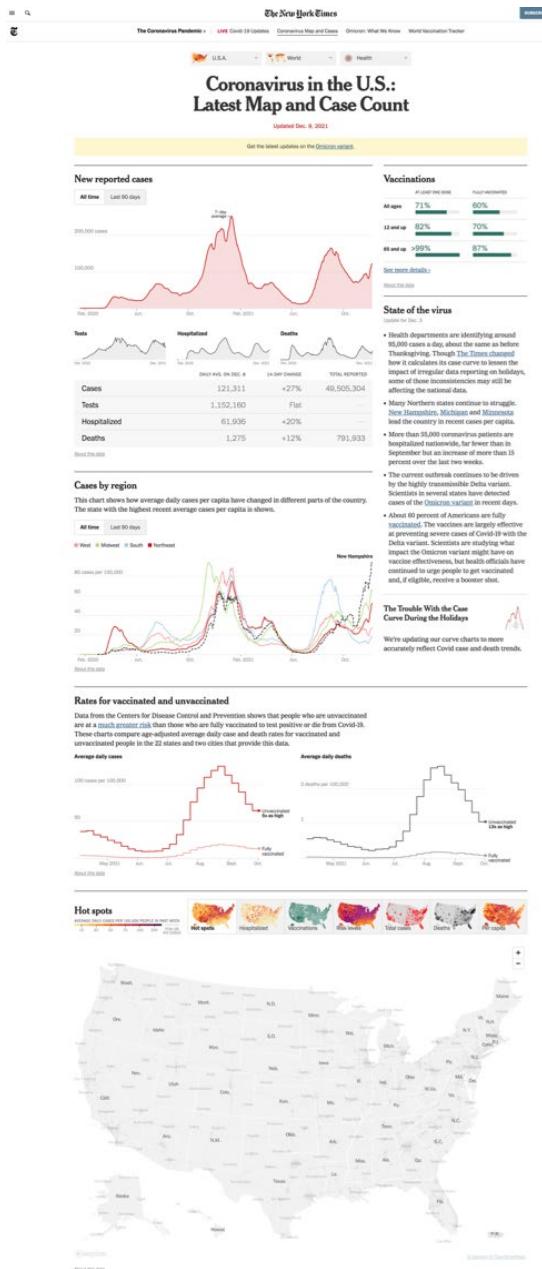


Nota. NYTimes: What Data Shows About Vaccine Supply and Demand in the Most Vulnerable Places <https://www.nytimes.com/interactive/2021/12/09/world/vaccine-inequity-supply.html>

- NYTimes: Coronavirus in the U.S.: Latest Map and Case Count <https://www.nytimes.com/interactive/2021/us/covid-cases.html>

Figura 86.

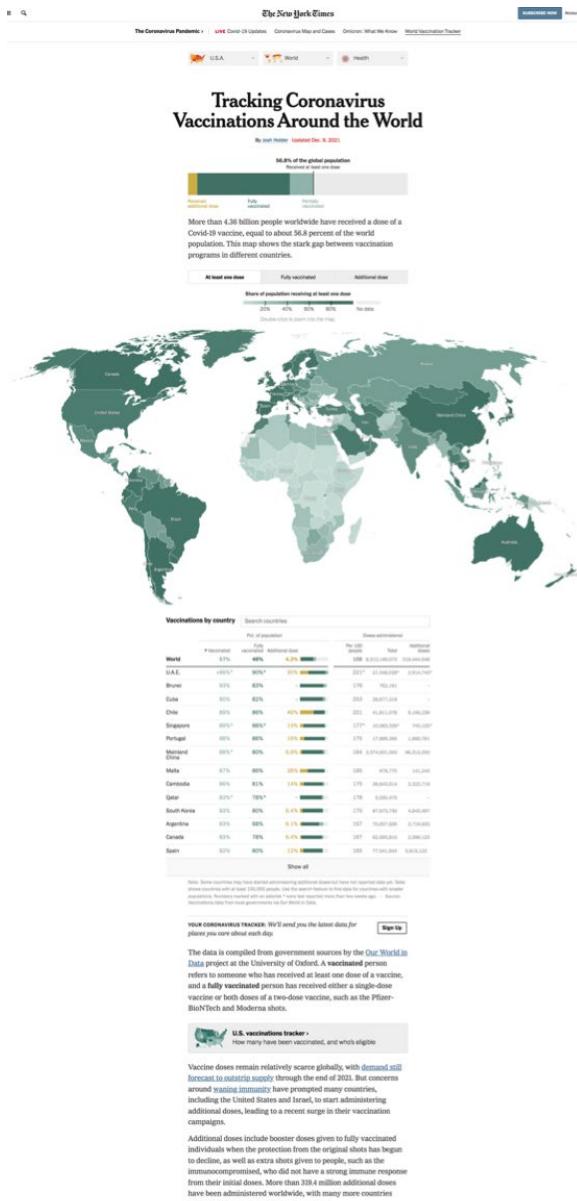
Ejemplo de artículo con datos



Nota. NYTimes: Coronavirus in the U.S.: Latest Map and Case Count <https://www.nytimes.com/interactive/2021/us/covid-cases.html>

- NYTimes: Tracking Coronavirus Vaccinations Around the World <https://www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html>

Figura 87.
Ejemplo de artículo con datos



Nota. NYTimes: Tracking Coronavirus Vaccinations Around the World <https://www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html>

Le invito a revisar otros artículos más en la página del Times: <https://www.nytimes.com/> y en especial el sitio creado para contar los hechos que marcaron el 2020 (los del 2021 están en construcción) <https://www.nytimes.com/interactive/2020/12/30/us/2020-year-in-graphics.html>

¡Video recomendado!

Le invito a analizar el [REA: "The beauty of data visualization"](#) de David McCandless, que convierte conjuntos de datos complejos, como el gasto militar mundial, el rumor de los medios y las actualizaciones de estado de Facebook, en diagramas hermosos y simples que revelan patrones y conexiones invisibles. Un buen diseño, sugiere, es la mejor manera de navegar por el exceso de información, y puede cambiar la forma en que vemos el mundo.

Además, comparto un listado de medios de comunicación que se enfocan en periodismo de datos:

- *The Guardian*: www.theguardian.com/data
 - <https://twitter.com/GuardianData>
- *The Washington Post*:
 - <https://twitter.com/PostGraphics>
- *The New York Times*:
 - <https://twitter.com/nytgraphics>
 - <https://twitter.com/UpshotNYT>
- *La Nación*: <https://www.lanacion.com.ar/data/>
 - <https://twitter.com/LNdata>
- *The Wall Street Journal*
 - <https://twitter.com/WSJGraphics>
- *The Financial Times*
 - <https://twitter.com/ftdata>

Para finalizar, le invito a leer "[Presentar datos al público](#)" (Manual de Periodismo de Datos 1.0, 2014), en donde desde la propia voz de los expertos, contarán su punto de vista sobre esta posibilidad que presenta el mundo de los datos en el periodismo tanto tradicional como digital.

Consulte con su profesor:

De esta forma terminamos la unidad. Si tiene alguna duda por favor comuníquese con su profesor a través del EVA. A continuación, es importante desarrollar la autoevaluación, a fin de determinar su grado de comprensión de los contenidos analizados.

Toda unidad de su guía incluye preguntas de autoevaluación que le recomiendo realizar para reforzar sus conocimientos aprendidos.



Autoevaluación 14

Señale la respuesta correcta:

1. **¿Qué autor propone la pirámide invertida en el periodismo de datos?**
 - a. Paul Bradshaw.
 - b. Alberto Cairo.
 - c. Ann Emery.

2. **Complete: "Una buena visualización es como una _____"**
 - a. buena fotografía.
 - b. buena gráfica.
 - c. buena imagen.

3. **Complete: "Los datos son valiosos dado que el acceso a ellos tiene el potencial de clarificar cuestiones de un modo que genere ____."**
 - a. dinero
 - b. resultados
 - c. interacción

4. **Entre los tipos de usuarios están:**
 - a. Ocasionales.
 - b. Únicos.
 - c. Prosumidores.

5. **Entre los tipos de usuarios están:**
 - a. Únicos.
 - b. Activos.
 - c. Ocasionales.

6. **Entre los tipos de usuarios están:**
 - a. Consumidores.
 - b. Ocasionales.
 - c. Acaparadores de datos.

7. Una de las fases de trabajo con datos es:

- a. Fase de descubrimiento.
- b. Fase de construcción.
- c. Fase de decisiones.

8. Una de las fases de trabajo con datos es:

- a. Fase de migración.
- b. Fase de eliminación.
- c. Fase de producción.

9. Uno de los medios pioneros en periodismo de datos en Iberoamérica es:

- a. *Primicias.ec*.
- b. *La Nación*.
- c. *Miami Herald*.

10. El medio de comunicación que ganó el considerado premio Pulitzer de la infografía: "Malofiej", fue:

- a. El País.
- b. El Confidencial.
- c. El Diario.es.

[Ir al solucionario](#)

Desarrolle la siguiente pregunta

En liste 10 ejemplos de productos periodísticos de índole nacional e internacional.

Nota: conteste las actividades en un cuaderno de apuntes o en un documento Word.

Su participación en el proceso de enseñanza y aprendizaje es esencial en nuestra asignatura.

¡Gracias por su atención y participación en el componente de Data Journalism!



Actividades finales del bimestre



Semana 16

Actividad final:

Se destina la semana 16 para finalizar y entregar las actividades propuestas dentro de la planificación del segundo bimestre. Además de realizar la actividad suplementaria para quienes no pudieron asistir a la actividad síncrona.

Con esto hemos culminado la revisión de nuestra asignatura *Data Journalism*, gracias por llegar hasta aquí. *Ánimo en el cierre del bimestre.*



4. Solucionario

Autoevaluación 1		
Pregunta	Respuesta	Retroalimentación
1	a	Sin duda el acontecimiento que permitió el arranque de este periodismo fue la llegada de Internet, la cual derrumbó el monopolio de la producción de noticias en manos de los medios impresos, la radio y televisión.
2	b	Correcto, el periodismo de precisión junto al de investigación son quienes le han permitido ser en la actualidad una especialización periodística.
3	b	Como todo, este término tiene sus antecedentes, sí, los tiene, por ejemplo tiene sus raíces en el periodismo de precisión y de investigación.
4	b	Philip Meyer estableció una tendencia nueva: el uso por parte de los reporteros de técnicas de investigación en ciencias sociales para aumentar la profundidad y precisión de las historias importantes.
5	c	A raíz de la publicación del libro de Philip Meyer en la década de los 90s, el periodismo de datos toma un leve significado teórico, pero no estaría respaldado por la gestión administrativa y sobre todo, periodística de los medios de comunicación o instituciones a fines en Latinoamérica.
6	c	«Data-driven Journalism» comenzó a utilizarse en Estados Unidos en el año 2009, lo que permitía nombrar la generación de productos periodísticos a partir de datos estructurados, adaptándose luego a los diferentes idiomas.
7	a	Este medio publicó el 29 de septiembre de 1849 un gráfico sobre la epidemia de cólera.
8	a	The New York Daily Tribune, publicó el 29 de septiembre de 1849 un gráfico sobre la epidemia de cólera.
9	b	The Philadelphia Inquirer, quien en 1975 contrató a un reportero dedicado a analizar exhaustivamente los números del censo y así crear artículos exclusivos basados en datos numéricos.

Autoevaluación 1

Pregunta	Respuesta	Retroalimentación
10	c	En perspectiva, netamente una parte de la macro data solo comenzó a ser usarse por los medios en los años 2000, sobre todo, por la dificultad analizar información estadística. En perspectiva, netamente una parte de la macro data solamente comenzó a ser utilizarse por los medios en los años 2000, sobre todo, por la dificultad analizar información estadística.

[Ir a la
autoevaluación](#)

Autoevaluación 2		
Pregunta	Respuesta	Retroalimentación
1.	b	«Esta disciplina periodística se nutre de otras muchas: de investigación, en profundidad, de precisión, asistido por computadora y analítico. En ella, se trabaja con grandes volúmenes de datos, se aprovecha al máximo la visualización interactiva y se incorpora al programador al equipo periodístico.» (Crucianelli, 2013)
2.	a	Para Érico Mafra, Data Journalism, «es una modalidad de producción digital de noticias que utiliza grandes bases de datos para elaborar piezas de contenido enfocadas en la correlación de información y que, además, utiliza recursos gráficos e interactivos para hacer más agradable la experiencia de visualización del usuario consumidor de noticias.» (Rock Content, 2019)
3.	c	Correcto, es la fórmula propuesta por la autora.
4.	a	Sí, de Database journalism. <i>“Database Journalism or structured journalism is a principle in information management whereby news content is organized around structured pieces of data, as opposed to news stories”</i>
5.	b	Incluye todos los aspectos que relacionan procesamiento de la información y periodismo. El término se hizo popular hace unos pocos años con motivo del así llamado «robot periodista» (software que ensambla informaciones de procedencia distinta y componen notas de prensa).
6.	a	Periodismo de precisión (PPr): hace uso de la investigación social empleando metodologías de índole cualitativa y cuantitativa.
7.	a	Cómo empezar: Entonces, ¿dónde comienza un periodista de datos en cierres? Una respuesta obvia sería “con los datos”, pero también hay una segunda respuesta: “Con una pregunta”.
8.	a	Búsqueda de datos: La búsqueda de datos puede implicar cualquier cosa, desde tener conocimientos y contactos con expertos hasta poder utilizar habilidades de informes asistidos por computadora o, para algunos, habilidades técnicas específicamente como MySQL o Python que permiten recopilar nuestros propios datos.
9.	a	Correcto: Compartir datos Datos enviados Eventos recurrentes Teorías a explorar Últimas noticias

Autoevaluación 2

Pregunta | Respuesta | Retroalimentación

10. b Visualización

Equipo de gráficos

Wordle

Many eyes

Google Explore Data

Google Fusión Tables

Ir a la
autoevaluación

Autoevaluación 3

Pregunta	Respuesta	Retroalimentación
1.	a	Sí, porque amplía un nuevo campo de especialización.
2.	b	Sí, porque los usuarios requieren otro tipo de contenido.
3.	c	Sí, para reunir, filtrar y visualizar lo que sucede más allá de lo que nos muestran nuestros ojos.
4.	c	Usando datos, la tarea de los periodistas pasa de centrarse en ser los primeros en informar, a ser los que nos dicen lo que un proceso podría significar realmente.
5.	a	Los datos representan una promesa, y esto es lo que entusiasma a las redacciones, haciéndolas buscar un nuevo tipo de periodista. Para la gente que trabaja por su cuenta, manejar datos ofrece un camino para obtener nuevas oportunidades y un salario estable también.
6.	a	No responden a intereses particulares. Es fundamental que la información provenga de fuentes confiables y que no estén influenciadas por intereses comerciales o ideológicos. En este sentido, los números garantizan su veracidad.
7.	b	Aumentan la calidad de las producciones. El uso de herramientas que permiten la visualización de datos proporcionan a las audiencias otro tipo de experiencias, en algunos casos mucho más interactivas e inmersivas.
8.	c	Correcto. Lluís Codina comenta que se puede identificar actores sociales, entidades y posibles fuentes de información para un determinado tema o línea de temas.
9.	b	Correcto, el periodista de The New York Times, Aron Pilhofer.
10.	a	Correcto, Isao Matsunami del medio Tokyo Shimbun.

**Ir a la
autoevaluación**

Autoevaluación 4		
Pregunta	Respuesta	Retroalimentación
1.	a	Correcto, es aplicaciones de noticias o new apps.
2.	c	Correcto, conjunto de datos o datasets.
3.	b	Correcto, DataViz o visualización de datos.
4.	a	Correcto, artículos noticiosos.
5.	a	A decir de Crucianelli (2013), este producto son artículos que nacen a partir de un volumen considerable de datos y que nos permiten construir historias; sobre todo, son artículos de corta extensión, y que en su mayoría emplean texto detallando cifras, valores, números, porcentajes, entre otros.
6.	c	A decir de Crucianelli (2013), este producto son artículos que nacen a partir de un volumen considerable de datos y que nos permiten construir historias; sobre todo, son artículos de corta extensión, y que en su mayoría emplean texto detallando cifras, valores, números, porcentajes, entre otros.
7.	b	Correcto. Primicias.ec: "Más de 159.000 personas dejaron de tener un empleo adecuado en octubre". https://www.primicias.ec/noticias/economia/deterioro-empleo-adecuador-ecuador-octubre/ (Coba, 2021)
8.	a	Correcto. La Nación: "Pandora Papers: una offshore suma misterio a la desaparición de US\$7 millones en la empresa de jugos Baggio". https://www.lanacion.com.ar/politica/pandora-papers-una-offshore-suma-misterio-a-la-desaparicion-de-los-us7-millones-en-la-empresa-de-nid01112021/
9	c	La visualización de datos o data visualization (dataviz), permite a través de imágenes, gráficos, infografías, pictogramas y demás elementos, mostrar información obtenida a través grandes volúmenes de datos. A decir de Crucianelli (2013), pueden estar complementados por artículos cortos o de gran extensión, y en algunos casos ir solo acompañados con un titular y una explicación para que los usuarios puedan analizarlos y leerlos.
10	a	Como su nombre lo indica se trata de una colección de datos, conjunto de datos, o dataset (término en inglés) que nos permite acceder a la información para su procesamiento. En cierto sentido suele estar ligado a la transparencia de las organizaciones. En Internet podemos acceder a bases de datos conforme la temática que busquemos pero esto dependerá de quién los provea y su transparencia ante la comunidad, lo que se considera "open data".

Ir a la
autoevaluación

Autoevaluación 5		
Pregunta	Respuesta	Retroalimentación
1.	a	Correcto, la Declaración de los Derechos Humanos proclamada por la Asamblea General de las Naciones Unidas en París, el 10 de diciembre de 1948 en su Resolución 217 A (III), como un ideal común para todos los pueblos y naciones.
2.	b	Correcto es un plan de acción a favor de las personas, el planeta y la prosperidad...
3.	c	Correcto. Objetivo 16. Paz, justicia e instituciones sólidas.
4.	a	Correcto. 16.10 Garantizar el acceso público a la información y proteger las libertades fundamentales, de conformidad con las leyes nacionales y los acuerdo.
5.	a	Correcto. 16.10 Garantizar el acceso público a la información y proteger las libertades fundamentales, de conformidad con las leyes nacionales y los acuerdo.
6.	b	Las leyes de transparencia a nivel mundial se alinean tanto con el artículo 19 de la Declaración Universal de los Derechos Humanos como con el ODS 16.10.
7.	b	Correcto: fidelidad de los datos
8.	a	Correcto. Datos personales sensibles: Algunos tipos de datos personales, teniendo en cuenta su sensibilidad en contextos particulares, son especialmente susceptibles de causar daños considerables a las personas si se hace mal uso de ellos.
9.	b	En Ecuador, el 10 de mayo de 2021 a través de la Asamblea Nacional aprobó con 118 votos y 1 abstención el proyecto de Ley de Protección de Datos Personales.
10.	c	Sí, Proteger a las personas y garantizar derechos.

Ir a la
autoevaluación

Autoevaluación 6

Pregunta	Respuesta	Retroalimentación
1.	b	Los datos abiertos tienen una relación directa con la transparencia hacia los ciudadanos, dado que son el componente básico del conocimiento.
2.	a	Sí, el conocimiento abierto es en lo que se convierten los datos abiertos cuando son útiles, utilizables y utilizados por los usuarios.
3.	c	Correcto, fueron: Estados Unidos, Canadá, Gran Bretaña y Nueva Zelanda.
4.	a	Sí, sobre todo en la participación
5.	b	Correcto. Reutilización y redistribución: los datos deben proporcionarse en condiciones que permitan la reutilización y la redistribución, incluida la mezcla con otros conjuntos de datos.
6.	a	Correcto. Los datos deben estar disponibles en su totalidad y a un costo de reproducción razonable.
7.	b	Correcto. Todos debemos poder usar, reutilizar y redistribuir la información; no debe existir discriminación contra personas o grupos.
8.	b	Correcto. Cultura: datos sobre obras artísticas: títulos y autores; galerías, bibliotecas, archivos y museos.
9.	a	Correcto. Escoger un conjunto de datos que deseamos hacer público.
10.	c	Correcto: excel xlsx, html, xml, csv, rss feed, json, etc., dado que en formatos planos como las imágenes o un PDF se tornará más compleja la extracción de la información.

Ir a la
autoevaluación

Autoevaluación 7

Pregunta	Respuesta	Retroalimentación
1.	c	Correcto, trata del volumen de datos.
2.	a	Correcto, "tiende a referirse al análisis del comportamiento del usuario, extrayendo valor de los datos almacenados, y formulando predicciones a través de los patrones observados."
3.	a	Correcto, a los datos.
4.	b	Correcto, a la información.
5.	c	Correcto, al conocimiento.
6.	b	Correcto. Con el fin de llegar a la etapa de la sabiduría, las compañías deben invertir en tecnologías y formación, con el fin de aprovechar la inversión y utilizar esta "nueva sabiduría".
7.	b	Correcto, variedad.
8.	c	Correcto, veracidad.
9.	c	Correcto, Públicas: normalmente abiertas como los sitios gubernamentales.
10.	b	La premisa central de Big Data es la toma decisiones adecuadas y acertadas, y a su vez, tener nos permiten tener un rol más activo en la sociedad.

Ir a la
autoevaluación

Autoevaluación 8

Pregunta	Respuesta	Retroalimentación
1.	a	Correcto, Big Data.
2.	b	Correcto, la minería de datos o Data Mining.
3.	a	Correcto: Datos, captura, almacenamiento, procesamiento, uso del dato, publicación y archivado.
4.	c	Correcto, análisis para permitir una mejor comprensión y orientación de los usuarios.
5.	a	Complete, entre las estrategias de Data Mining están:
6.	b	Correcto, la persona que mina la información se denomina "minero".
7.	a	Correcto, Recordemos que donde hay datos hay información, donde hay información hay conocimiento y este lo podemos obtener a través del proceso de Big Data en conjunto con Data Mining.
8.	b	Así es, corresponde a Big Data. Infografía: Data Mining Vs. Big Data.
9.	a	Correcto, corresponde a Data Mining. Infografía: Data Mining Vs. Big Data.
10.	c	Así es, corresponde a Big Data. Infografía: Data Mining Vs. Big Data.

Ir a la
autoevaluación

Autoevaluación 9		
Pregunta	Respuesta	Retroalimentación
1.	a	Claramente, corresponde a "obtener datos de APIs".
2.	b	Correcto, a la extracción de datos de .PDF.
3.	c	Correcto, se trata del screen scraping"
4.	a	Así es, diario La Nación con su nota La Nación (Argentina): https://www.lanacion.com.ar/sociedad/la-vacunacion-en-la-argentina-en-tiempo-real-nid02022021/#
5.	a	Correcto, Google (búsqueda simple o avanzada) funciona como fuente básica para localizar información
6.	c	Sí, corresponde a la búsqueda en la red. Información en la figura sobre Guía de búsqueda de datos propuesto en el Manual de Periodismo de Datos.
7.	a	Sí, debemos acudir primero a la fuente y se trata de una búsqueda directa. Información en la figura sobre Guía de búsqueda de datos propuesto en el Manual de Periodismo de Datos.
8.	b	Correcto, se trata de una búsqueda específica. Información en la figura sobre Guía de búsqueda de datos propuesto en el Manual de Periodismo de Datos.
9.	c	Sí, corresponde a la búsqueda en la red. Información en la figura sobre Guía de búsqueda de datos propuesto en el Manual de Periodismo de Datos.
10.	b	Correcto, se trata de una búsqueda específica. Información en la figura sobre Guía de búsqueda de datos propuesto en el Manual de Periodismo de Datos.

Ir a la
autoevaluación

Autoevaluación 10

Pregunta	Respuesta	Retroalimentación
1.	b	Correcto, podemos extraerlos desde un sitio web.
2.	a	Correcto, nuestros datos deben ser completos.
3.	b	Correcto, nuestros datos deben ser únicos.
4.	a	Correcto, nuestros datos deben ser precisos.
5.	c	Correcto, nuestros datos deben ser consistentes.
6.	a	Así es, corresponde a la veracidad.
7.	c	Así es, corresponde a el volumen.
8.	b	Correcto, corresponde a velocidad.
9.	c	Sí, se refiere al valor.
10.	b	Correcto, se trata de variedad.

Ir a la
autoevaluación

Autoevaluación 11

Pregunta	Respuesta	Retroalimentación
1.	b	Correcto, también es conocido como "data cleaning".
2.	a	Correcto: Elimina los principales errores e inconsistencias que son inevitables cuando se extraen múltiples fuentes de datos en un conjunto de datos.
3.	c	Correcto: El uso de herramientas para limpiar los datos hará que todos sean más eficientes, ya que podrán obtener rápidamente lo que necesitan de los datos.
4.	b	Correcto: Menos errores es igual a más eficiencia en los procesos.
5.	a	Correcto: La capacidad de mapear las diferentes funciones y lo que sus datos están destinados a hacer y de dónde provienen sus datos.
6.	c	Correcto: Inspección, limpieza, verificación, informes.
7.	b	Correcto, son del Tipo 1: Fáciles, sencillos y económicos de resolver.
8.	a	Muy bien. Tipo 2: Complejos de resolver. Requieren, por norma, de ayuda externa.
9.	b	Correcto, Open Refine https://openrefine.org/
10.	a	Correcto, no debemos cambiar los archivos originales.

Ir a la
autoevaluación

Autoevaluación 12

Pregunta	Respuesta	Retroalimentación
1.	a	Así es, correcto. El análisis e interpretación de los datos está sujeto a la correcta extracción y limpieza de nuestra base. Debemos recalcar que es importante saber interpretar los datos previo a su visualización.
2.	b	Nuestro análisis de datos puede ser de la índole cuantitativa (información numérica, por ejemplo, encuestas), lo que nos puede permitir:
3.	c	Nuestro análisis de datos puede ser de la índole cualitativa (entrevistas, grupos focales, observación, etc.)
4.	a	Sí, corresponde a la clasificación.
5.	c	Correcto, corresponde a tablas dinámicas.
6.	b	Sí, corresponde a filtrar.
7.	a	Correcto, matemáticas y textos.
8.	a	Sí, Lo que se VE es noticia.
9.	c	Correcto, son las predicciones
10.	b	El objetivo de la recopilación e interpretación es adquirir información útil, utilizable y tomar las decisiones más informadas posibles.

Ir a la
autoevaluación

Autoevaluación 13

Pregunta	Respuesta	Retroalimentación
1.	a	Correcto, como dataviz o data visualization
2.	a	Correcto: Tomar decisiones y actuar rápidamente sobre las decisiones que se tomaron.
3.	c	Sí, usa el lenguaje visual al que la sociedad (usuarios) está acostumbrada.
4.	b	Correcto, tendencias.
5.	a	Correcto, correlaciones
6.	c	Así es, mapas
7.	b	Correcto, Data Viz Project: https://datavizproject.com/
8.	a	Correcto, Onodo: https://onodo.org/ que permite realizar conexiones.
9.	c	Exacto, Flourish: https://app.flourish.studio/
10.	b	Correcto, el experto en dataviz, Alberto Cairo.

Ir a la
autoevaluación

Autoevaluación 14

Pregunta	Respuesta	Retroalimentación
1.	a	Correcto, Paul Bradshaw.
2.	c	Así es, una buena visualización es como una buena imagen.
3.	b	Los datos son valiosos dado que el acceso a ellos tiene el potencial de clarificar cuestiones de un modo que genere resultados.
4.	a	Usuarios ocasionales: son usuarios que quieren información de modo simple y rápido. Les interesa tener una idea de los datos, no un análisis detallado. Podemos dirigirnos a ellos vía tweet o gráficos interactivos.
5.	b	Usuarios activos: usuarios que estimulan el debate y usan los datos para incrementar su conocimiento de un área determinada o cuestionan los supuestos de los datos.
6.	c	Acaparadores de datos: estos usuarios quieren datos en crudo para hacer visualizaciones o análisis. Simplemente, les damos los datos para sus propósitos.
7.	a	Correcto, la fase de descubrimiento es esencial y la primera.
8.	c	Correcto, la fase de producción
9.	b	Correcto, La Nación ha sido pionera con el trabajo con datos y fue uno de los primeros medios en crear un área de esta especialidad denominada: LaNaciónData.
10.	a	Diario El País tiene una trayectoria en el trabajo con datos y salud, por ejemplo en 2021 el equipo de Nuevas Narrativas, en manos de sus periodistas Mariano Zafra y Javier Salas ganaron el premio "Malofiej", llamados los "Pulitzer de la infografía", justamente por el artículo que a continuación veremos.

[Ir a la autoevaluación](#)



5. Glosario

Brevemente y a través de la Real Academia de la Lengua, RAE, (2021) y Fundéu conoceremos los principales términos que hemos empleado en nuestra asignatura:

- **Algoritmo:** es un conjunto ordenado de instrucciones, pasos o procesos que permiten desarrollar una tarea determinada. Ejemplo: algoritmo que Google usa para su buscador.
- **Base de datos:** conjunto de datos organizado de tal modo que permita obtener con rapidez diversos tipos de información.
- **Big data o macrodatos:** conjunto de datos que, por su volumen y variedad y por la velocidad a la que necesitan ser procesados, supera las capacidades de los sistemas informáticos habituales.
- **Data mining o minería de datos:** conjunto de técnicas y tecnologías que se emplea para explorar y extraer datos de soportes digitales, descubriendo patrones de comportamiento para la toma de decisiones.
- **Inteligencia Artificial (AI):** se trata de enseñar a las máquinas a desempeñar labores complejas que solo podían hacer humanos.
- **Machine learning:** implica procesos en los que las propias máquinas son las que crean sus reglas (algoritmos) y predicciones basándose en datos que le suministran los humanos.
- **Open data:** busca la disponibilidad y acceso a datos e información sin restricciones para los usuarios.



6. Referencias bibliográficas

- Alconada, H., Jastreblansky, M., y Brom, R. (2021, Noviembre 1). Pandora Papers: una offshore suma misterio a la desaparición de US\$7 millones en la empresa de jugos Baggio. <https://www.lanacion.com.ar/politica/pandora-papers-una-offshore-suma-misterio-a-la-desaparicion-de-los-us7-millones-en-la-empresa-de-nid01112021/>
- Asamblea General de la Organización de los Estados Americanos, O. (2019, February 18). La actualización de los principios sobre la privacidad y protección de datos personales. Http://Www.Oas.Org/Es/Sla/Cji/Docs/CJI_doc_582-19.Pdf
- Bradshaw, P. (2010, Octubre 10). *How to be a data journalist*. Https://Www.Theguardian.Com/News/Datablog/2010/Oct/01/Data-Journalism-How-to-Guide?CMP=twt_gu.
- Casal, F. (2010). *El periodismo de investigación asistido por ordenador y de precisión aplicado a la prensa local y regional: el contraste entre estados unidos y un modelo de futuro para Galicia*.
- CEPAL. (2019). *La Agenda 2030 y los Objetivos de Desarrollo Sostenible: una oportunidad para América Latina y el Caribe. Objetivos, metas e indicadores mundiales*.
- Coba, G. (2021, Noviembre 5). Más de 159.000 personas dejaron de tener un empleo adecuado en octubre *Primicias.ec*: <https://www.primicias.ec/noticias/economia/deterioro-empleo-adecuador-ecuador-octubre/>. <https://www.primicias.ec/noticias/economia/deterioro-empleo-adecuador-ecuador-octubre/>.
- Codina, L. (2019, Abril 10). *Periodismo en bases de datos*. <Https://Www.Lluiscodina.Com/Periodismo-Documentacion-Medios/Periodismo-Bases-Datos/>.
- Comisión Europea. (2021, Diciembre 1). *¿Qué datos personales se consideran sensibles?* <Https://Ec.Europa.Eu/Info/Law/Law-Topic/>

Data-Protection/Reform/Rules-Business-and-Organisations/Legal-Grounds-Processing-Data/Sensitive-Data/What-Personal-Data-Considered-Sensitive_es.

Constitución de la República del Ecuador. (2008). Constitución de la República del Ecuador. In Asamblea Nacional de la República del Ecuador (Ed.), https://www.asambleanacional.gob.ec/sites/default/files/documents/old/constitucion_de_bolsillo.pdf (RO: 449, Vol. 449). Asamblea Nacional.

Crucianelli, S. (2013). ¿Qué es el periodismo de datos? *Cuadernos de Periodistas*, 26.

Datos Abiertos Ecuador, y Fundación de Ayuda por Internet (FUNDAPI). (2021, Octubre 3). *Datos Abiertos, Ecuador*. <Https://Datosabiertos.Ec/>.

Democratice Ouverte. (2019, Mayo 15). *Démocratice Ouverte*. <Https://Www.Democraticeouverte.Org/> <https://www.democraticeouverte.org/>

Dirección Nacional de Registros Públicos (Dinarp). (2021, Noviembre 9). *Ley de Protección de Datos Personales*. <Https://Www.Dinardap.Gob.Ec/Programas-Servicios/Servicios/Proyecto-de-Ley-de-Proteccion-de-Datos/>.

El Universo. (2021, Abril 12). Resultados de las elecciones presidenciales 2021 en Ecuador (segunda vuelta). <Https://Www.Eluniverso.Com/Resultados-Elecciones-Ecuador-Presidente-2021/>.

European Journalism Centre, y Google News Initiative. (2021, Noviembre 26). *The Data Journalism Handbook 2*. <Https://Datajournalism.Com/Read/Handbook/Two.>

Gongora, A. (2014, Febrero 2). *Definición de periodismo interpretativo*. <Https://Alexrgongora.Wordpress.Com/2014/02/02/Definicion-de-Periodismo-Interpretativo/>.

Grassler, M. (2017). *El rol del periodista de datos en el proceso de los sistemas de gestión y decisión pública y en la recuperación de la confianza entre el ciudadano y las instituciones públicas*.

Howard, A. (2018). *El arte y la ciencia del periodismo de datos*.

IBM. (2021, Diciembre 8). *Data visualization*. <Https://Www.Ibm.Com/Analytics/Data-Visualization>.

IEBS. (2018, Enero 8). ¿Qué es el Data Mining o minería de datos? <Https://Www.lebschool.Com/Blog/Data-Mining-Mineria-Datos-Big-Data/>.

Jastreblansky, M. (2012, June 25). Los millones de la APE: cómo se repartió en 2011 la caja que era de Moyano. <https://www.lanacion.com.ar/politica/los-millones-de-la-ape-como-se-repartio-en-2011-la-caja-que-era-de-moyano-nid1484852/>

La Nación. (2021, Noviembre 25). La evolución de la pandemia en la Argentina. *La Evolución de La Pandemia En La Argentina*.

Manual de Periodismo de Datos 1.0. (2014, Abril 28). *Manual de Periodismo de Datos 1.0*. <https://interactivos.lanacion.com.ar/manual-data/>

Manual de Periodismo de Datos, y Ávila, A. M. (2021, Octubre 12). *Manual de Periodismo de Datos Iberoamericano*. <http://manual.periodismodedatos.org/>

Naciones Unidas (ONU). (1948, Diciembre 10). *La Declaración Universal de Derechos Humanos*. https://www.un.org/es/documents/udhr/UDHR_booklet_SP_web.pdf

Open Data HandBook. (2021, Noviembre 30). *Open Data HandBook*. <Https://Opendatahandbook.Org/>.

Organización de las Naciones Unidas. (2015, Septiembre 25). *Objetivos de Desarrollo Sostenible*. <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>.

Ortega Mohedano, F., y Coronel-Salas, G. (2019). Big data, augmented data y computación cognitiva en la era del millenial. In L. M. L. M. Romero-Rodríguez y D. Rivera-Rogel (Eds.), *La comunicación en el escenario digital. Actualidad, retos y prospectivas* (Primera, Vol. 01, pp. 825–853). Pearson.

Propublica.org. (2016, Marzo 6). Infographics in the time of cholera. <Https://Www.Propublica.Org/Nerds/Infographics-in-the-Time-of-Cholera>.

QuestionPro. (2021, Diciembre 7). ¿Qué es el análisis de datos? <https://www.questionpro.com/es/analisis-de-datos.html>.

Real Academia Española (RAE). (2021, Octubre 14). Definición periodismo y datos. <https://dle.rae.es/periodismo>.

Recuero de los Santos, P. (2017, June 13). *Tus datos más limpios, casi sin frotar.* <Https://Empresas.Blogthinkbig.Com/Tus-Datos-Mas-Limpios-Casi-Sin-Frotar/>.

Rock Content. (2019, Mayo 6). Periodismo de Datos: transformación digital en la producción y consumo de noticias. <https://rockcontent.com/es/blog/periodismo-de-datos/>

Wikipedia. (2021, Octubre 15). *Macrodatos.* <https://es.wikipedia.org/wiki/Macrodatos>



7. Anexos

Términos para referirse al DJ

Término	Explicación
Periodismo de Bases de datos	Sinónimo de Periodismo de Datos. Es poco empleado.
Periodismo de Datos	
Periodismo de Datos	Un periodismo basado en la explotación e interpretación de colecciones de datos, generalmente de origen público, aunque también pueden proceder de filtraciones, como en el caso <i>WikiLeaks</i> , <i>Panamá Papers</i> o <i>Pandora papers</i> , así como de corporaciones y empresas privadas. Suele presentar una orientación de denuncia, aunque no es un requerimiento imprescindible. Una de sus denominaciones minoritarias es Periodismo de bases de datos.
Database journalism	De acuerdo con la Wikipedia, " <i>Database Journalism or structured journalism is a principle in information management whereby news content is organized around structured pieces of data, as opposed to news stories</i> " ("El periodismo de base de datos o periodismo estructurado es un principio en la gestión de la información mediante el cual el contenido de las noticias se organiza en torno a datos estructurados, en lugar de noticias.")
Structured journalism	Sinónimos de <i>Database Journalism</i>
Computational journalism	Incluye todos los aspectos que relacionan procesamiento de la información y periodismo. El término se hizo popular hace unos pocos años con motivo del así llamado «robot periodista» (software que ensambla informaciones de procedencia distinta y componen notas de prensa). Tal como se utiliza es un término que se refiere a todos los aspectos de la unión de periodismo más computación. Por eso, el Periodismo en Bases de Datos, sería una parte del Periodismo Computacional, junto con el Periodismo de Datos.

Término	Explicación
Periodismo en Bases de Datos	Una derivación muy característica y con una fuerte identidad propia, del periodismo de datos, que se basa en el conocimiento y la explotación de informaciones periodísticas y de actualidad disponibles en hemerotecas digitales y bases de datos de prensa. Utiliza técnicas documentales de análisis de necesidades de información periodística y su conversión en estrategias de búsqueda avanzadas. Puede considerarse una derivación o una dimensión del periodismo de datos que pone en juego competencias informáticas e informacionales especializadas, así como el conocimiento de fuentes de información digitales retrospectivas especializadas, como las mencionadas, pero también bancos de imágenes, repositorios y archivos históricos y otros recursos especializados en información periodística textual o gráfica, retrospectiva.
Fundamentos	El fundamento del Periodismo en Base de Datos es doble: por un lado, requiere de la existencia de bases de datos de prensa y de hemerotecas de calidad; y por el otro, requiere de profesionales del periodismo bien formados en competencias informáticas, informacionales y documentales.

Nota. Lluís Codina, 2019. <https://www.lluiscodina.com/periodismo-documentacion-medios/periodismo-bases-datos/>

Ejemplo de artículo con datos masivos.

The screenshot shows a political article from 'LA NACION' titled 'Los millones de la APE: cómo se repartió en 2011 la caja que era de Moyano'. The article discusses the distribution of millions of pesos between the Autonomous Public Enterprise (APE) and the Central Bank during the Moyano era. It includes several charts and tables, such as a pie chart showing the percentage of APE's budget allocated to different sectors and a table comparing the distribution of funds between the APE and the Central Bank. The page also features a sidebar with related news and a footer with navigation links and a search bar.

Nota. <https://www.lanacion.com.ar/politica/los-millones-de-la-ape-como-se-repartio-en-2011-la-caja-que-era-de-moyano-nid1484852/>

Nota. La Nación, 2021. <https://www.lanacion.com.ar/politica/pandora-papers-una-offshore-suma-misterio-a-la-desaparicion-de-los-us7-millones-en-la-empresa-de-nid01112021/>



Economía

Autor: Gabriela Costa

Actualizada:

25 Nov 2021 • 10:13

Suscríbase

[Desempleo](#) [Aempleo](#) [Aempleo adecuado](#) [AfNAMU](#) [AfNIC](#) [Población económicamente activa](#) [Aduanas](#)

Más de 159.000 personas dejaron de tener un empleo adecuado en octubre

El mercado laboral se deterioró en octubre: el empleo adecuado cayó y el subempleo creció, según el Instituto Nacional de Estadística y Censo (INEC).

Un poco más de 2.6 millones de personas tuvieron un empleo adecuado en octubre de 2021, según la Encuesta Nacional de Empleo, Desempleo y Subempleos (Enred).

Por lo tanto, el empleo adecuado tuvo una contracción de 1,6 puntos porcentuales o lo que es igual a 159.474 personas menos en ese segmento del mercado laboral en comparación con septiembre de 2021.

- Ast es como Ecuador pagará los efectos de la guerra en Ucrania

Si se analiza la situación del empleo adecuado en términos interanuales, este segmento ha crecido 0,7 puntos porcentuales frente a octubre de 2020, año golpeado por la pandemia.

El Instituto Nacional de Estadística y Censo (INEC) considera que una persona tiene un empleo pleno o adecuado cuando:

- Percebe ingresos laborales iguales o superiores al salario mínimo, que es de USD 400. Trabaja 40 o más horas a la semana, independientemente del deseo y la disponibilidad de trabajar horas adicionales.
- Gana un salario mínimo o más por mes, trabaja menos de 40 horas, pero no desea trabajar horas adicionales.

La contracción del empleo adecuado se traduce en un aumento del subempleo y una salida de las personas de la población económicamente activa (PEA), es decir, del mercado laboral.

El subempleo aumentó 1,2 puntos porcentuales o el equivalente a 84.080 personas entre septiembre y octubre.

Se considera que una persona está subempleada cuando percibe ingresos inferiores al salario mínimo y trabaja menos de la jornada legal, a pesar de que tienen el deseo y la disponibilidad de trabajar horas adicionales.

Un segmento del mercado laboral que se ha mantenido es el desempleo, con 984.204 personas.

LO MÁS LEÍDO

01 Reforma al IESS incluirá aumento de edad de jubilación y ajuste al aporte estatal

4024 visitas

02 Así es como Ecuador pagará los efectos de la guerra en Ucrania

3450 visitas

03 Bess analiza alza en tasa de interés para hipotecarios de hasta USD 90.000

1962 visitas

04 El Insgol iniciará acciones colectivas por operaciones migratorias en el país

1205 visitas

05 Bess recuperará maquinaria decomisada en mina ilegal de Yatayapíne

875 visitas

06 Bess invertirá USD 1.000 millones en bonos del Estado en 2022

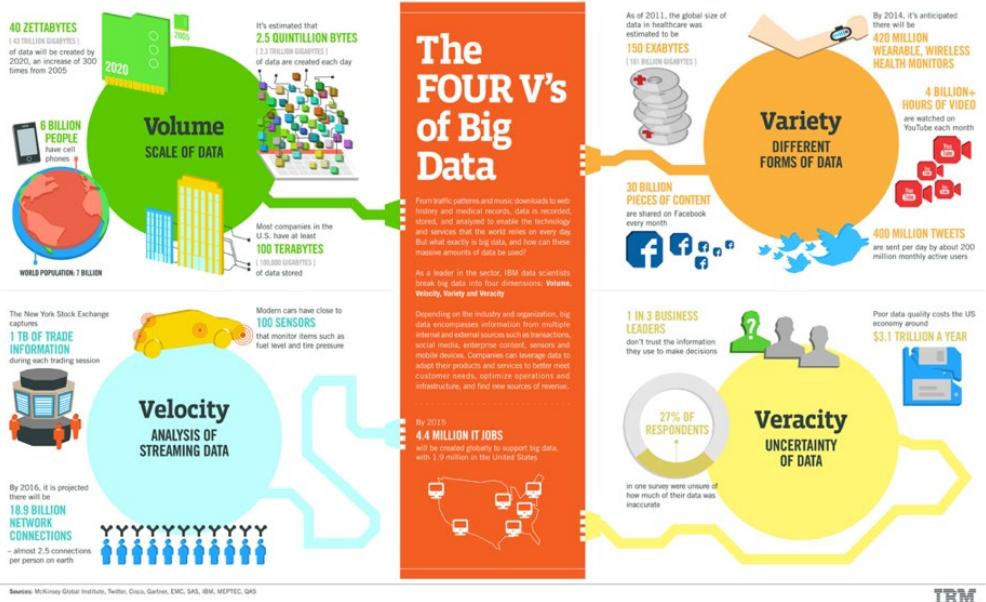
1181 visitas

07 Hay minería ilegal en sedes barriales del cantón San José de Tena

1019 visitas

Nota. Primicias.ec, 2021. <https://www.primicias.ec/noticias/economia/deterioro-empleo-adecuado-ecuador-octubre/>

The Four V's of Big Data



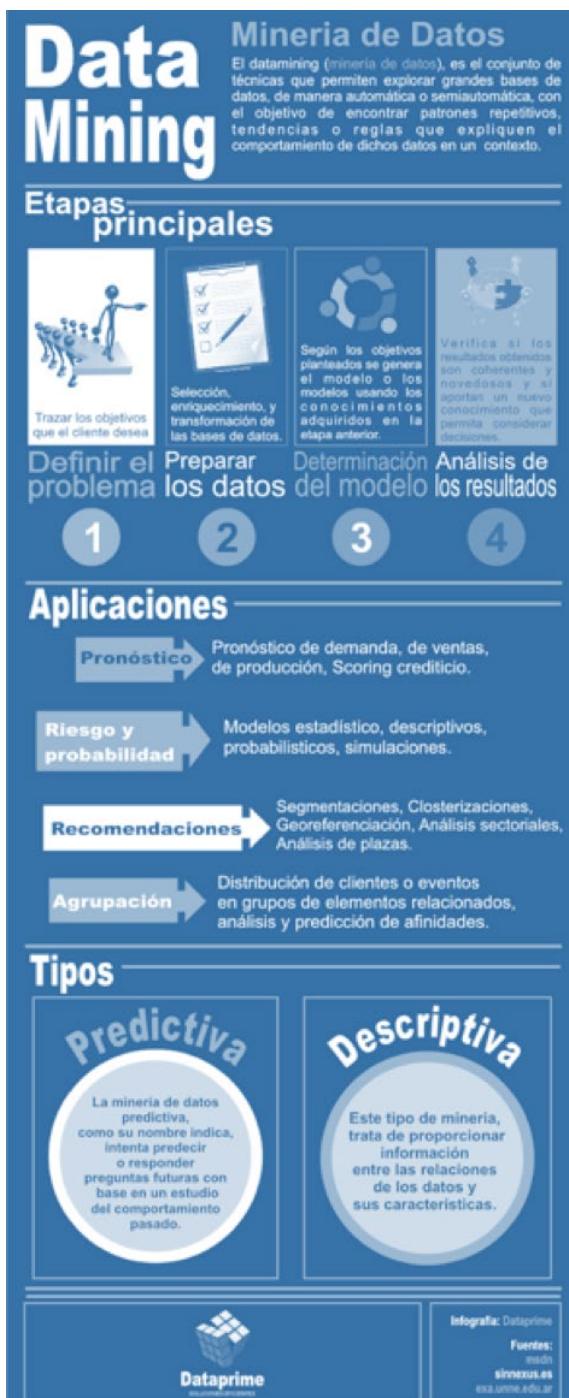
Nota. IBM

Data mining Vs. Big data.

 Data mining	Vs.	 Big data
Identifica y extrae la información relevante de grandes conjuntos de datos.		Se refiere a la recolección y almacenamiento de grandes cantidades de datos.
Utiliza diferentes técnicas basadas en la estadística y la Inteligencia artificial.		Debido a su volumen es imposible procesarlos con software convencional.
Arroja resultados específicos y concretos.		Necesita herramientas especiales para capturar, gestionar y tratar la información.
Crea modelos predictivos, de clasificación o segmentación.		Estos grupos de datos tienen baja densidad de información para realizar predicciones.
Transforma la información en conocimiento.		La calidad de los datos puede variar mucho y afectar a los resultados del análisis.

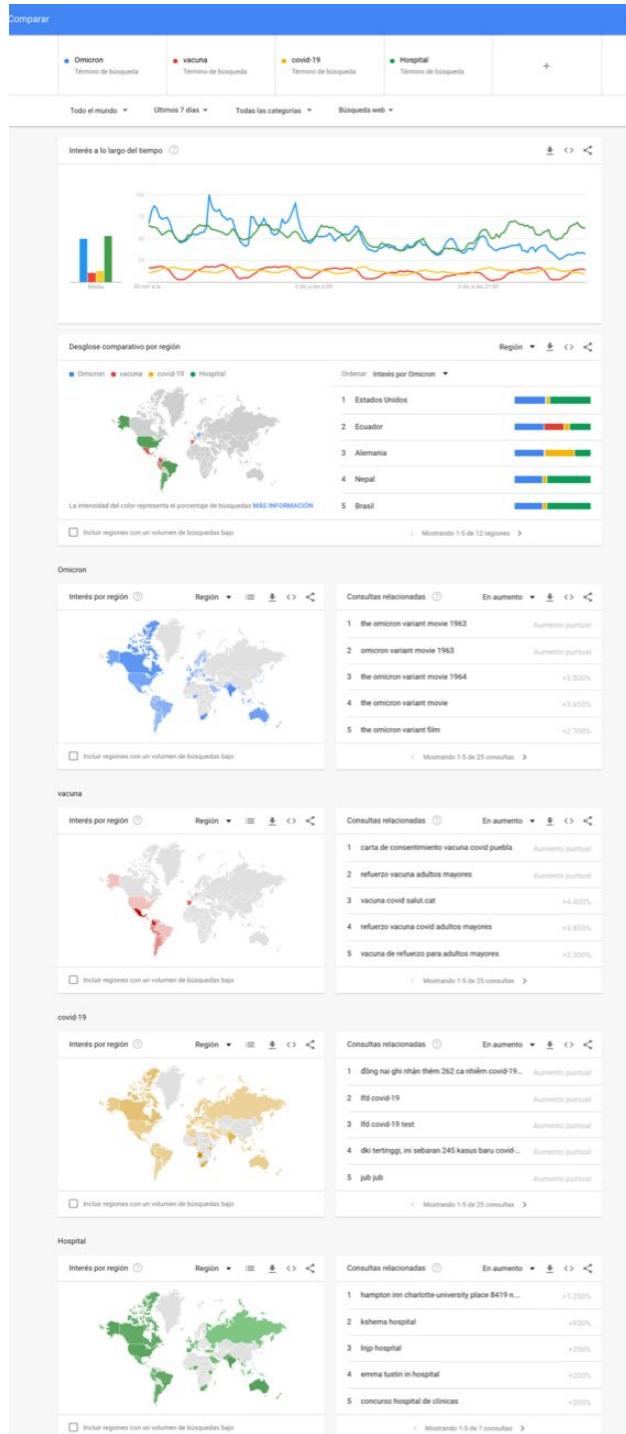
Nota. Iberdrola https://www.iberdrola.com/wcorp/gc/prod/es_ES/comunicacion/docs/Infografia_Data_mining_esp.pdf

Infografía Data Mining.



Nota. Dataprime

Búsqueda de Tendencias Ómicron, covid-19 y vacuna.



Nota. Google Trends

Tabla que se descarga desde GT.

Categoría: Todas las categorías	Valor
covid-19: (30/11/21, 14:42 - 7/12/21, 14:42, Todo el mundo)	
TOP	
covid	100
covid-19 vaccine	71
vaccine	71
covid 19	37
covid-19 test	31
covid-19 certificate	27
certificate	27
covid vaccine	21
covid-19 vaccination	17
vaccination	17
covid-19 registration	16
covid-19 cases	15
covid-19 vaccine certificate	14
vaccine certificate	14
covid-19 download	14
what is covid-19	13
covid-19 symptoms	13
covid-19 omicron	13
covid-19 india	13
covid-19 certificate download	12
covid-19 pandemic	11
coronavirus	10
covid-19 testing	9
covid-19 news	9
cowin	9
RISING	
đồng nai ghi nhận thêm 262 ca nhiễm covid-19 trong cộng đồng	Aumento puntual
nguoiduatin	
lfd covid-19	Aumento puntual
lfd covid-19 test	Aumento puntual
dki tertinggi, ini sebaran 245 kasus baru covid-19 3 desember 2021	Aumento puntual
jub jub	Aumento puntual
arti kata omicron, alfabet yunani yang jadi nama varian baru covid-19	4.200 %
beda gejala covid-19 varian omicron dan delta	2.700 %
ireland covid-19 restrictions	750 %
covid-19 omicron variant symptoms	450 %
zamrażarki do szczepionek pfizera	450 %
zamrażarka na szczepionki covid-19	400 %
cyril ramaphosa	350 %

Categoría: Todas las categorías	Valor
przenośne inkubatory co2	350 %
covid-19 viral test	300 %
covid-19 relief fund application online 2021	250 %
certificazione verde covid-19	200 %
prilly latuconsina	190 %
lyodra	160 %
covid-19 full name	150 %
universal pass covid-19	150 %
heru hidayat	150 %
cowin.gov.in covid-19 vaccine certificate	130 %
boson covid-19 antigenetest	120 %
pencegahan covid-19	110 %
sa health covid-19 update today	90 %

Nota. CSV de Google Trends.

Ejemplo de visualización de las series más vistas.



Nota. IMDB <https://www.imdb.com/>. Elaboración propia.

Mapa de los programas de TV favoritos en EE.UU.



Nota. Fortune Magazine: <https://fortune.com/2019/09/05/map-favorite-tv-shows-by-country-imdb/>

Los 4 pasos del proceso de elaboración de un gráfico con la herramienta Datawrapper.

This chart is in My Charts

1 Upload Data 2 Check & Describe ✓ 3 Visualize ✓ 4 Publish & Embed

How do you want to upload your data?

Copy & paste data table XLS/Csv upload Connect Google Sheet Link external dataset

Copy & paste your data

Select your data (including header row/column) in Excel or LibreOffice and paste it in the text field. You can also upload a CSV or Excel file from your computer.

If you just want to try Datawrapper, here's a list of some example datasets you can use:

Select a sample dataset

Partidos / Movimiento	Asambleistas Ecuador	Nº Nacionales	Nº Provinciales	Nº Exterior	Total
Unión por la Esperanza (UNES)	5	44	4	49	49
Movimiento de Unidad Plurinacional Pachakutik (MUPP)	3	24	1	27	27
Izquierda Democrática (ID)	2	16	0	18	18
Partido Social Cristiano (PSC)	2	16	0	18	18
Movimiento Creando Oportunidades (CREO)	2	18	1	1	12
Alianza Honestidad (MC-PSE)	1	1	0	2	2
Alianza PAIS (PAIS)	0	0	0	0	0
Movimiento Ecuatoriano Unido (MEU)	0	2	0	2	2
Partido Avanza (PPA)	0	2	0	2	2
Partido Sociedad Patriótica (PSP)	0	1	-	1	1

Datawrapper is developed by Datawrapper GmbH.
Academy - Blog - Terms - Privacy Policy - Imprint - Changelog - support@datawrapper.de

Back to top
version 1.25.0

This chart is in My Charts

1 Upload Data ✓ 2 Check & Describe ✓ 3 Visualize ✓ 4 Publish & Embed

Make sure the data looks right

Please make sure that Datawrapper interprets your data correctly. In the table `number` columns should be shown in blue, `dates` in green and `text` in black. A `red` cell indicates a problem in your dataset that needs to be fixed. `grey` cells contain no data.

First row as label

Output locale

Defines decimal and thousand separators as well as translation of month and weekday names.

Español (es-ES)

Click on table header to edit column properties

#	A	B	C	D	E
1	Partidos / Movimiento	Nº Nacionales	Nº Provinciales	Nº Exterior	Total
2	Unión por la Esperanza (UNES)	5	44	4	49
3	Movimiento de Unidad Plurinacional Pachakutik (MUPP)	3	24	1	27
4	Izquierda Democrática (ID)	2	16	0	18
5	Partido Social Cristiano (PSC)	2	16	0	18
6	Movimiento Creando Oportunidades (CREO)	2	18	1	12
7	Alianza Honestidad (MC-PSE)	1	1	0	2
8	Alianza PAIS (PAIS)	0	0	0	0
9	Movimiento Ecuatoriano Unido (MEU)	0	2	0	2
10	Partido Avanza (PPA)	0	2	0	2
11	Partido Sociedad Patriótica (PSP)	0	1	-	1
12	Unidad Popular (UP)	0	0	0	0
13	Partido Sociedad Unida Más Acción (SUMA)	0	0	0	0
14	Democracia Si (DSi)	0	1	0	1
15	Fuerza Ecuador (FE)	0	0	0	0

This chart is in My Charts

1 Upload Data ✓

2 Check & Describe ✓

3 Visualize

4 Publish & Embed

Chart type Refine Annotate Layout

Columns

Select column

Vertical axis

Custom range - Show grid lines Show axis labelsNumber format Custom ticks Position outside insideAlignment left right

Grid labels will always be displayed on the left when Category labels are displayed directly.

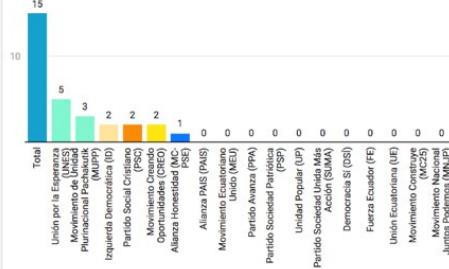
 Use the same scale for all tabs

Horizontal axis

Asamblea Ecuador

Composición de la Asamblea Nacional de Ecuador según partidos y/o movimientos políticos, período 2021 - 2025.

Nº Nacionales Nº Provinciales Nº. Exterior Total



Datos tomados de Wikipedia y validados con Asamblea

Gráfico: Gabriela Coronel Salas - Fuente: Asamblea Ecuador - Creado con Datawrapper

Datawrapper

Dashboard Create new ...

Archive

≡

This chart is in My Charts

1 Upload Data ✓

2 Check & Describe ✓

3 Visualize ✓

4 Publish & Embed

Publish visualization



You'll need to publish this visualization before embedding it on your website or sharing it on social media.

Your published visualization will still only be visible to people who know its URL. We won't share it publicly.

Export or duplicate visualization

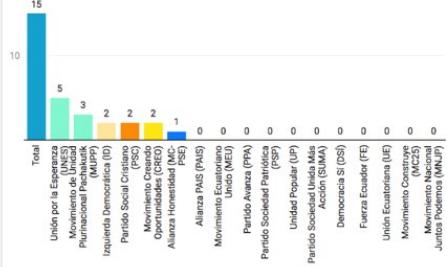
You can duplicate it to start editing a copy of the visualization. Or export it into other formats.



Asamblea Ecuador

Composición de la Asamblea Nacional de Ecuador según partidos y/o movimientos políticos, período 2021 - 2025.

Nº Nacionales Nº Provinciales Nº. Exterior Total



Datos tomados de Wikipedia y validados con Asamblea

Gráfico: Gabriela Coronel Salas - Fuente: Asamblea Ecuador - Creado con Datawrapper

Datawrapper is developed by Datawrapper GmbH.

Academy - Blog - Terms - Privacy Policy - Imprint - Changelog - support@datawrapper.de

Back to top

version 1.25.0

Nota. Elaboración propia.

Plataformas para la creación de contenido.

Sector	Nombre	Enlace	Detalle
Planificación	MockFlow	https://www.mockflow.com/	Planificar o bosquejar interfaces
	Mind master	https://www.mindmeister.com/	Mapas mentales en línea
	Google	https://coggle.it/	
Web	Wix	www.wix.com	Construcción de sitios web, fáciles de usar a través de sus plantillas.
	ReadyMag	https://readymag.com/	
	Webnote	https://www.webnode.es/	
	Strikingly	https://www.strikingly.com/	
	Shorthand	https://shorthand.com/	
	Atavist	https://atavist.com/	
	Google sites	https://sites.google.com/	
Herramientas	Knight Lab	https://knightlab.northwestern.edu/projects	Serie de herramientas propuestas por Knight Lab enfocadas en storytelling
	Juxtapose	https://juxtapose.knightlab.com/	Para comparar imágenes, por ejemplo, antes y después.
	Scene	https://scene.knightlab.com/	Videos
	Soundcite	https://soundcite.knightlab.com/	Sonidos
	Storyline	http://storyline.knightlab.com/	Complementar datos
	StoryMap	http://storymap.knightlab.com/	Mapas
	Timeline	https://timeline.knightlab.com/	Líneas de tiempo.
	Powtoon	https://www.powtoon.com/	animaciones
Diseño	Canva	https://www.canva.com/	Desde post hasta infografías.
Imagen, video, audio			
Infografías	Infogram	https://infogram.com/	Infografías animadas
	Thinglink	https://www.thinglink.com/	
	Genially	https://genial.ly/	
	Visme	https://www.visme.co/	
Audio	Ivoox	https://www.ivoox.com/	Incluir sonidos, audios o podcasts.
	SoundCloud	https://soundcloud.com/	
Mapas	Google Maps	https://mymaps.google.com/	Creación de mapas

Nota. Elaboración propia.