

---

# Google Data Analysis Course

Un libro de apuntes por Amber Costa

# Índice de Contenidos

# 1. Conceptos Clave

- **Datos :**
  - Conjunto de hechos
  - Centro de la toma de decisiones, por lo que se deben reducir sesgos, lagunas y errores
- **Ecosistema de Datos :** Todos los elementos que interactúan entre si para producir, administrar, almacenar, organizar, analizar y compartir datos
  - Esto incluye las herramientas de *Hardware* y *Software*
- **Data-driven decision making :** Utilización de datos para guiar la Estrategia de Negocios
  - Los datos nunca serán más poderosos que su combinación con experiencia humana, observación e intuición
- **Expertos en la Materia :** Profesionales del área de estudio que pueden aportar conocimiento experto al análisis
- **Stakeholder :** Persona que invierte tiempo y recurso en el proyecto, que le interesa el resultado del estudio
- **Análisis de Negocios :** Matemática y estadística enfocada a recolectar, analizar e interpretar datos para realizar decisiones de negocio. Se identifican cuatro tipos según la [Escuela de Negocios de Harvard](#):
  - **Descriptivo :** Identificar tendencias y patrones
  - **Predictivo :** Predicción de resultados en base a tendencias y patrones
  - **Diagnóstico :** Evaluación del problema, búsqueda del problema raíz
  - **Prescriptivo :** Test y pruebas para encontrar el mejor resultado
- **Análisis de Personas :** Análisis de datos enfocado en los empleados y su vida laboral, con tal de definir y crear un lugar de trabajo más productivo y empoderador
- **La Nube :** Ubicación virtual donde pueden almacenarse datos, en vez de utilizar un harddrive de computador

## 1.1. Diferencias entre Análisis y Ciencia de Datos

- **Análisis de Datos :**
  - Colecta, transformación, organización de datos, con el fin de sacar conclusiones, realizar predicciones y tomar decisiones informadas basadas en hechos

- Responder preguntas ya existentes, al proponer nuevas perspectivas en base a datos y sus fuentes
- **Analista de Datos** : Persona que recolecta, transforma, organiza datos para ayudar a realizar decisiones informadas
- **Análisis Computacional de Datos** :
  - *Ciencia de los Datos*
  - Creación de modelos, exploración de lo desconocido usando datos sin editar
  - Creación de preguntas y problemas usando datos
  - **Ciencia de Datos** contiene a los *Ecosistemas de Datos* y al *Análisis de Datos*

## 2. Ciclo de Vida de los Datos

*No confundir con análisis, solamente datos*

- **Planning - Planificación :**

- Características de los datos a tomar: qué tipo, cómo obtenerlos, quién los administra, resultados óptimos
- Qué tipos de datos se necesitan? Cómo se gestionarán? Quién será el responsable de ellos

- **Capture - Captura :**

- Recolección de datos desde distintas fuentes como Bases de Datos, Archivos, Encuestas...

- **Manage - Mantención :**

- Cuidado de los datos, almacenamiento y seguridad
- Cómo y dónde se almacenan? Qué herramientas se usan para hacerlo?
- *No corresponde limpieza, ya que es una tarea del analista y no una parte del ciclo de vida*

- **Analyze - Análisis :**

- Resolución del problema utilizando los datos
- Uso de los datos para respaldar objetivos empresariales y tomar decisiones

- **Archive - Archivar :**

- Almacenar los datos en un lugar donde estén disponible para referencias futuras y a largo plazo
- Pueden o no ser usados de nuevo

- **Destroy - Destrucción :**

- Eliminar los datos y las copias compartidas
- Se destruye la documentación en papel, o se realiza limpieza de discos duros, servidores y bases de datos

	Servicio de Pesca y Vida Silvestre (USA)	Servicio Geológico de EEUU (USGS)	Instituciones Financieras	Escuela de Negocios de Harvard (HBS)
Ciclo de Vida	<ul style="list-style-type: none"><li>• Planificar</li><li>• Adquirir</li><li>• Mantener</li><li>• Acceder</li><li>• Evaluar</li><li>• Archivar</li></ul>	<ul style="list-style-type: none"><li>• Planificar</li><li>• Adquirir</li><li>• Procesar</li><li>• Analizar</li><li>• Preservar</li><li>• Publicar/ Compartir</li></ul>	<ul style="list-style-type: none"><li>• Capturar</li><li>• Calificar</li><li>• Transformar</li><li>• Utilizar</li><li>• Informar</li><li>• Archivar</li><li>• Depurar</li></ul>	<ul style="list-style-type: none"><li>• Generación</li><li>• Recolección</li><li>• Procesamiento</li><li>• Almacenamiento</li><li>• Gestión</li><li>• Análisis</li><li>• Visualización</li><li>• Interpretación</li></ul>

USGS realiza otras actividades transversales, como Describir (metadatos y documentación), y Gestionar calidad y copias de seguridad

### 3. Destrezas del Analista de Datos

**Habilidades Analíticas** : Cualidades y características asociadas con la resolución de problemas usando hechos:

- **Curiosidad** : Querer aprender algo
- **Entender Contexto** : Comprensión del ambiente en que se genera, produce o utiliza una idea
- **Mentalidad Técnica** : Desglosar cosas en pasos o partes pequeñas, y trabajar con ellas de forma lógica y ordenada
- **Diseño de Datos** : Organización de la información por formatos, tipos o almacenamientos
- **Estrategia de datos** : Gestión de la gente, procesos y herramientas usadas en el Análisis de Datos

**Pensamiento Analítico** : Identificar y definir un problema, para luego resolverlo usando datos de forma organizada, paso a paso. Presenta 5 aspectos principales:

- **Visualización** : Representación gráfica de la información
- **Ser Estratégico** : Crear planes para resolver un problema dado ciertos datos, manteniendo el enfoque, calidad y utilidad
- **Pensamiento Orientado a Problemas** : Identificar problema, y mantenerlo como motor principal del proyecto completo
- **Correlación** : Relación entre dos variables. *Correlación  $\neq$  causalidad*
- Los siguientes aspectos debieran considerarse al mismo tiempo :
  - **Big Picture Thinking** : Visión del panorama completo
  - **Detail Oriented Picture** : Estudio de las partes en detalle

## 4. Ciclo de Vida del Análisis de Datos

Segun Google Data Analytics

### 4.1. Act - Preguntar

- Definición del desafío/objetivo empresarial, junto a los lineamientos para lo que se considera exitoso
- Comprender las expectativas de los Stakeholders
- Colaboración con Expertos en la Materia y otros profesionales
- Aplicación del Pensamiento Analítico/Estructural
- **Preguntas efectivas :**
  - Cuál es la principal causa del problema?
    - **Five Why's :** Preguntar "por qué?" cinco veces para llegar a la raíz del problema
  - Cuáles son los déficits de nuestro proyecto?
    - **Análisis de Déficit - Gap Analysis :** Método de evaluación y examinación de un proceso y cómo funciona actualmente, para determinar cómo llegar a la meta a futuro
  - Qué no consideramos previamente?
    - Observar el estado actual, e identificar qué lo diferencia del estado ideal
    - Hay datos previos para revisar?
    - Hay preconcepciones? Cuáles?
  - Qué intentamos lograr?

### 4.2. Prepare - Preparar

- Planificación de datos de acuerdo al proyecto
- Recolección y Captura
- Seleccionar tipo de dato de acuerdo al problema: formatos, tipos y estructuras
- Tomar medidas para evitar sesgos y mantener imparcialidad
- Almacenamiento y administración de dichos datos



### 4.3. Process - Procesar

- Limpieza de datos :
  - Completar datos usando otras fuentes externas
  - Eliminar y manejar errores, imprecisiones y casos extremos
  - Transformación a formatos distintos
- Control de calidad previo al análisis

### 4.4. Analyze - Analizar

- Exploración de los datos ya preparados :
  - Usar herramientas como hojas de cálculo, bases de datos y queries
  - Formatear, transformar, ordenar y filtrar datos
- Identificar patrones y conclusiones
- Realizar predicciones y recomendaciones

### 4.5. Share - Compartir

- Interpretar resultados a través de **Visualización**
- Comunicar resultados :
  - Exclusivo para Stakeholders y personas con acceso autorizado
  - Conversaciones para ayudar a comprender resultados
- Evaluación y toma de decisiones

### 4.6. Act - Actuar

- Transformar resultados en acciones de cambio
- Resolver el desafío empresarial definido al principio del ciclo
- Crear algo nuevo: otro proceso, otro problema, modelos nuevos

## 5. Otros tipos de Ciclo de Vida del Análisis de Datos

### 5.1. EMC/Dell EMC - Proceso cíclico

Descubrimiento - Preprocesamiento de datos - Planificación de modelos - Construcción de Modelos - Comunicación (de resultados) - Puesta en funcionamiento Las fases no son hitos estáticos = dan paso al siguiente y luego se repiten

### 5.2. SAS - Iterativo

Preguntar - Preparar - Explorar - Modelar - Implementar - Actuar - Evaluar [evaluar soluciones para posteriormente volver a preguntar] Producir resultados repetibles, fiables y predictivos

### 5.3. Basado en Proyectos

Identificación del problema - Diseño de requisito de datos - Preprocesamiento de datos - Realización del análisis de datos - Visualización de datos

### 5.4. Ciclo del Big Data

Evaluación del caso de negocio - Identificación de datos - Adquisición y Filtrado de Datos - Extracción de datos - Validación y Limpieza de datos - Agregación y representación de datos - Análisis de datos - Visualización de datos - Utilización de resultados del análisis

## 6. Informes con L<sup>A</sup>T<sub>E</sub>X

### 6.1. Una breve introducción

Este es un párrafo, puede contener múltiples “Expresiones” así como fórmulas o referencias<sup>1</sup> como (??) o (??). A continuación se muestra un ejemplo de inserción de imágenes (como la Figura ??) con el comando `\insertimage`:

---

<sup>1</sup> Las referencias se hacen utilizando la expresión `\label{etiqueta}`.



Figura 1: Where are you? de “Internet”.

A continuación<sup>2</sup> se muestra un ejemplo de inserción de ecuaciones simples con el comando `\insertequation`:

$$a^k = b^k + c^k \quad \forall k > 2 \quad (1)$$

Este template [?] ha sido diseñado para que sea completamente compatible con editores L<sup>A</sup>T<sub>E</sub>X para escritorio y de manera online<sup>[?]</sup>. La compilación es realizada siempre usando las últimas versiones de las librerías, además se incluyen los parches oficiales para corregir eventuales *warnings*.

Este es un nuevo párrafo. Para crear uno basta con usar `\\` en el anterior, lo que fuerza una nueva línea. También se pueden insertar con el comando `\newp` si el compilador de latex arroja una alerta del tipo *Underfull \hbox (badness 10000) in paragraph at lines ...*

## 6.2. Añadiendo tablas

También puedes usar tablas, ¡Crearlas es muy fácil!. Puedes usar el plugin [Excel2Latex](#) [?] de Excel para convertir las tablas a L<sup>A</sup>T<sub>E</sub>X o bien utilizar el “creador de tablas online” [?].

Tabla 1: Ejemplo de tablas.

Columna 1	Columna 2	Columna 3
$\omega$	$\nu$	$\delta$
$\xi$	$\kappa$	$\varpi$

<sup>2</sup> Como se puede observar las funciones `\insert...` añaden un párrafo automáticamente.

## 7. Aquí un nuevo tema

### 7.1. Haciendo informes como un profesional



Figura 2: Apolo flotando a la izquierda.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget

orci sit amet orci dignissim rutrum.

Proin sit amet augue. Praesent lacus. Donec a leo. Ut turpis ante, condimentum sed, sagittis a, blandit sit amet, enim. Integer sed elit. In ultricies blandit libero. Proin molestie erat dignissim nulla convallis ultrices. Aliquam in magna. Etiam sollicitudin, eros a sagittis pellentesque, lacus odio volutpat elit, vel tincidunt felis dui vitae lorem. Etiam leo. Nulla et justo.

$$\int_a^b f(x) dx = \frac{\partial^n f(x)}{\partial x^n} \cdot \sum_{x=a}^b f(x) \left(1 + \Delta x\right)^{1 + \frac{\epsilon}{k}} \quad (2)$$

*Ecuación sin sentido.*

**Definición 1** (ver [? ]) *Definición definitiva*

$$\frac{d}{dx} \int_a^x f(y) dy = f(x)$$

Proin sit amet augue. Praesent lacus. Donec a leo. Ut turpis ante, condimentum sed, sagittis a, blandit sit amet, enim. Integer sed elit. In ultricies blandit libero. Proin molestie erat dignissim nulla convallis ultrices. Aliquam in magna. Etiam sollicitudin, eros a sagittis pellentesque, lacus odio volutpat elit, vel tincidunt felis dui vitae lorem. Etiam leo. Nulla et justo.

## 7.2. Otros párrafos más normales

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

$$\Lambda_f = \frac{L \cdot f}{W} \cdot \frac{Q_e^2}{8\pi^2 W^4 g} + \sum_{i=1}^l \frac{f \cdot (M-d)}{l \cdot W} \cdot \underbrace{\frac{(Q_e - i \cdot Q)^2}{8\pi^2 W^4 g}}_{\sim \mathcal{A}} \quad (3)$$

$$Q_e = 2.5Q \cdot \int_0^e V(x) dx + \sin^{-1} \left( 1 + \frac{1}{1-e} \right)$$

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

$$\boxed{f(x) = \frac{\partial^2 u}{\partial t^2}} \quad (4)$$

Fusce mauris. Vestibulum luctus nibh at

lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

## 7.3. Ejemplos de inserción de código fuente

El template permite la inserción de los siguientes lenguajes de programación de forma nativa: ABAP, Ada, Assembler x64, Assembler x86[masm], Awk, Bash, Basic, C, Caml, CMake, Cobol, C++, C#, CSS, CSV, CUDA, Dart, Docker, Elisp, Elixir, Erlang, Fortran, F#, GLSL, Gnuplot, Go, Haskell, HTML, INI, Java, Javascript, JSON, Julia, Kotlin, LaTeX, Lisp, LLVM, Lua, Make, Maple, Mathematica, Matlab, Mercury, Modula-2, Objective-C, Octave, OpenCL, OpenSees, Pascal, Perl, PHP, Texto plano, PostScript, Powershell, Prolog, Promela, Pseudocódigo, Python, Q#, R, Racket, Reil, Ruby, Rust, Scala, Scheme, Scilab, Simula, SPARQL, SQL, Swift, TCL, VBScript, Verilog, VHDL y XML.

Para insertar un código fuente se debe usar el entorno **sourcecode**, o el entorno **sourcecodep** si es que se quiere utilizar parámetros adicionales. A continuación se presenta un ejemplo de inserción de código fuente en Python (Código ??), Java y Matlab:

Código 1: Ejemplo en Python.

```
1 import numpy as np
2
3 def incmatrix(genl1, genl2):
4     m, n = len(genl1), len(genl2)
5     VT = np.zeros((n*m, 1), int) # Comentario
```

Código 2: Ejemplo en Java.

```
1 import javax.servlet.*;
2
3 // Hola mundo
4 public class Hola extends GenericServlet {
5     public void service(ServletRequest request, ServletResponse response)
6         throws ServletException, IOException{
7         PrintWriter pw = response.getWriter();
8         pw.println("Hola, mundo!");
9     }
10 }
```

Código 3: Ejemplo en Matlab.

```
1 % Se crea gráfico
2 f = figure(1); hold on; movegui(f, 'center');
3
4 fad = ones(1, NDATOS); % Arreglo para el FAD
5 for i = 1:NDATOS
6     [t, u_t, ~, ~] = main(BETA(j), r(i), M, K, F0, 0);
7     fad(i) = max(abs(u_t)) / uf0;
8 end
```

## 7.4. Agregando múltiples imágenes

El template ofrece el entorno `images` que permite insertar múltiples imágenes de una manera muy sencilla. Para crear imágenes múltiples se deben usar las siguientes instrucciones:

```
1 \begin{images}{\label{imagenmultiple}}{Ejemplo de imagen múltiple.}
2   \addimage{\label{ciudadfoto}}{ejemplos/test-image}{width=6.5cm}{Ciudad}
3   \addimageanum{ejemplos/test-image-wrap}{width=5cm}
4   \imagesnewline
5   \addimage{ejemplos/test-image}{width=12cm}{Ciudad más grande}
6 \end{images}
```

Obteniendo así:



(a) Ciudad



(b) Ciudad más grande

Figura 3: Ejemplo de imagen múltiple.

# Más ejemplos

## Listas y Enumeraciones

Hacer listas enumeradas con  $\text{\LaTeX}$  es muy fácil con el `template`<sup>3</sup>, para ello debes usar el comando `\begin{enumerate}`, cada elemento comienza por `\item`, resultando así:

1. Grecia
2. Abracadabra
3. Manzanas

También se puede cambiar el tipo de enumeración, se pueden usar letras, números romanos, entre otros. Esto se logra cambiando el **label** del objeto `enumerate`. A continuación se muestra un ejemplo usando letras con el estilo `\alph`<sup>4</sup>, números romanos con `\roman`<sup>5</sup> o números griegos con `\greek`<sup>6</sup>:

- |             |             |                        |
|-------------|-------------|------------------------|
| a) Peras    | i) Rojo     | $\alpha$ ) Matemáticas |
| b) Manzanas | ii) Café    | $\beta$ ) Lenguaje     |
| c) Naranjas | iii) Morado | $\gamma$ ) Filosofía   |

Para hacer listas sin numerar con  $\text{\LaTeX}$  hay que usar el comando `\begin{itemize}`, cada elemento empieza por `\item`, resultando:

- |            |          |           |
|------------|----------|-----------|
| – Peras    | * Rojo   | • Árboles |
| – Manzanas | * Café   | • Pasto   |
| – Naranjas | * Morado | • Flores  |

## Otros

Recuerda revisar el manual de todas las funciones y configuraciones de este `template` visitando el siguiente link: <https://latex.ppizarror.com/informe>. Si encuentras un error en el `template`, puedes abrir un nuevo issue a través de su página en GitHub.

<sup>3</sup> También puedes revisar el manual de las enumeraciones en <https://latex.ppizarror.com/doc/enumitem.pdf>.

<sup>4</sup> Con `\Alph` las letras aparecen en mayúscula.

<sup>5</sup> Con `\Roman` los números romanos salen en mayúscula.

<sup>6</sup> Una característica propia del `template`, con `\Greek` las letras griegas están escritas en mayúscula.



## Anexo A. Cálculos realizados

### A.1. Metodología

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum. (??)



Figura A.1: Imagen en anexo.

### A.2. Resultados

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetur odio sem sed wisi.

Tabla A.1: Tabla de cálculo.

Elemento	$\epsilon_i$	Valor	Descripción
A	10	$3,14\pi$	Valor muy interesante <sup>a</sup>
B	20	6	Segundo elemento
C	30	7	Tercer elemento <sup>1</sup>
D	150	10	Sin descripción
E	0	0	Cero

<sup>a</sup> Este elemento tiene una descripción debajo de la tabla

<sup>1</sup> Más comentarios