

EoCode

Clusterización

Aplicado a un caso de estudio **ficticio**



Empresa Patito

Caso de estudio



La compañía **Patito** quiere iniciar **operaciones en México**, lanzará una nueva línea de **equipos celulares** por lo que requiere de hacer un **análisis de mercado** para tomar decisiones que le ayuden a competir

Necesitamos **construir una fuente de datos**
que pueda usar la empresa para tomar
decisiones al entrar al mercado

?



Datos

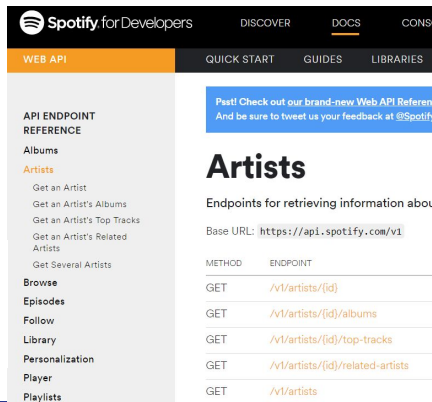
Estructurados

- Bases de datos
- DatawareHouses

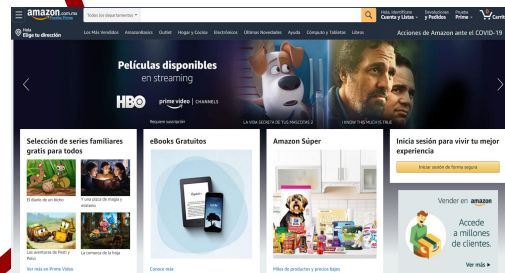


Semi estructurados

- APIs
- Tabulares



No estructurados.



Hipótesis

Creemos que podemos **encontrar valor** de mercado analizando la sección de celulares de un e-commerce como **Amazon**



¿Qué queremos conocer?

Productos

- Marca
- Producto
- S.O
- Procesador
- Memoria
- Peso
- Dimensiones
- Resolución

- Color
- Almacenamiento
- Gama
- Precio

Clientes

- Calificaciones
- Comentarios
- Preguntas



¿Cómo recopilamos la información?

?

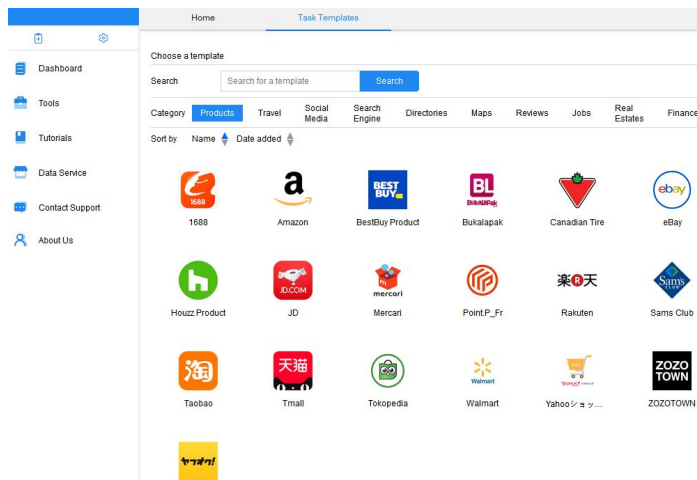


Web scraping es una técnica para **recopilar información** utilizando la **estructura de HTML** de una página web. Con esta técnica podemos automatizar procesos para **obtener datos y llenarlos en una base de datos**.

Herramientas para procesar datos no estructurados (HTML)

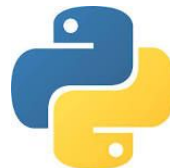
Soluciones comerciales

[Octoparse](#)
[import.io](#)
[Mozenda](#)
[Parsehub](#)
[Crawlmonster](#)
[Connotate](#)
[Common Crawl](#)



Programando

Python



- BS4
- Selenium
- Scrapy



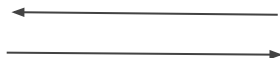
- Rvest

Experiencia con R

- Base estadística
- Soporta Data frames nativos
- R Markdown, R Notebooks
- Persistencia de variables
- Enfocado al análisis



Retrocompatibilidad

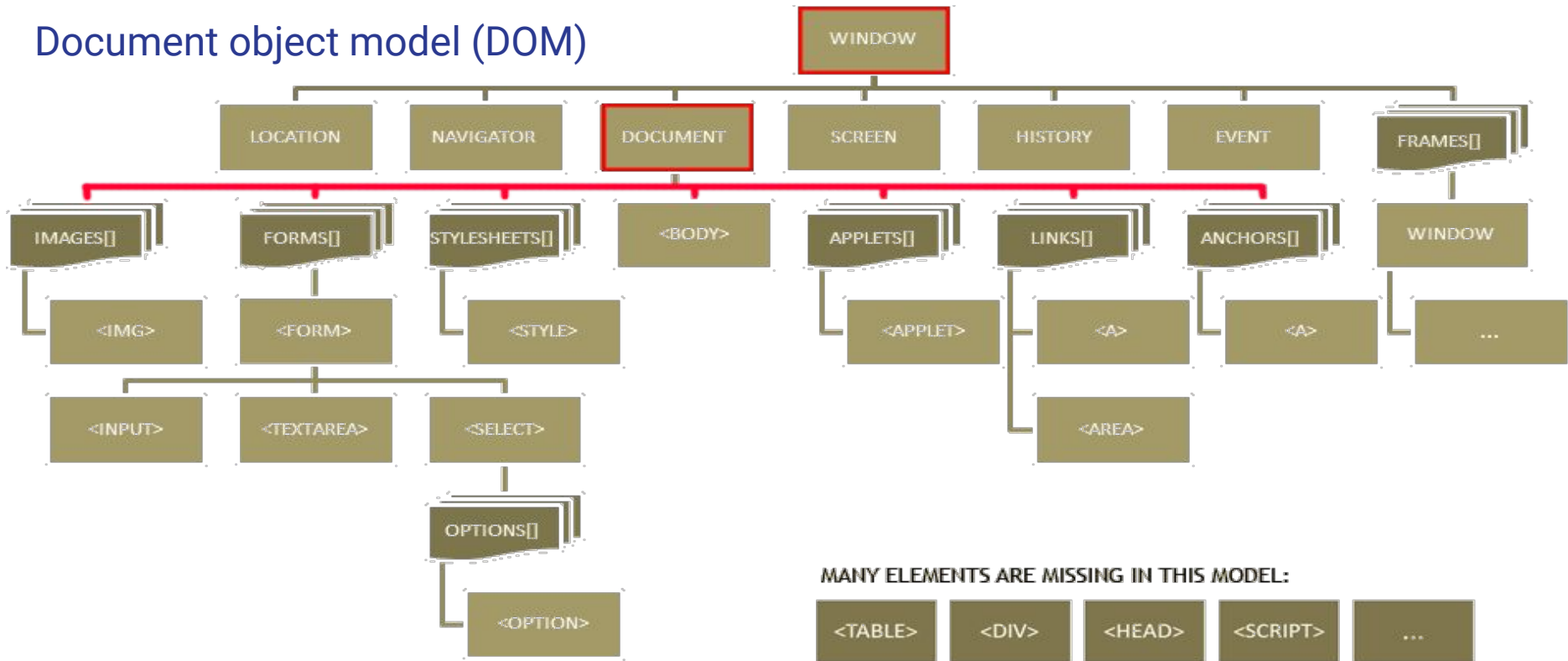


Experiencia con Python

- Muchas librerías que complementan al lenguaje y lo hacen robusto
- Modular y escalable
- Adaptable a necesidades de todo tipo
- Herramientas Jupyter Notebook, Anaconda, Pycharm



Document object model (DOM)



El árbol de nodos de una página web

Para hacer Web Scraping

Consideremos que cualquier sitio web

CONTIENE

Estructura



Datos

Identificadores

Clases

Etiquetas

Recursos multimedia



Contenido en Texto
que representa algo



```
<!doctype html>
<html lang="en">
<head>
  <!-- Required meta tags -->
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width, initial-scale=1">

  <!-- Bootstrap CSS -->
  <link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/

  <title>Hello, world!</title>
</head>
<body>
  <div id="carouselExampleSlidesOnly" class="carousel slide" data-ride="carousel">
    <div class="carousel-inner">
      <div class="carousel-item active">
        
      </div>
      <div class="carousel-item">
        
      </div>
      <div class="carousel-item">
        
      </div>
    </div>
  </div>

  <div class="jumbotron">
    <h1 class="display-4">Hello, world!</h1>
    <p class="lead">This is a simple hero unit, a simple jumbotron
    <hr class="my-4">
    <p>It uses utility classes for typography and spacing to space
    <a class="btn btn-primary btn-lg" href="#" role="button">Learn
  </div>

  <section>
    <div class="card">
      <div class="card-body">
        This is some text within a card body.
      </div>
    </div>
  </section>
</body>
</html>
```



- Marca
- ☐ Samsung
 - ☐ Motorola
 - ☐ Gadgets México
 - ☐ GADGETS MEXICO
 - ☐ tech21
 - ☐ Apple
 - ☐ Incipio
 - ☐ mophie
 - ☐ Plantronics
 - ☐ BINDEN
 - ☐ Lanix

Órdenes Recurrentes

- ☐ Planea y Ahorra

Tamaño de la Pantalla del Teléfono

- ☐ 3.9 Pulgadas y Menos
- ☐ 4 a 4.4 Pulgadas
- ☐ 4.5 a 4.9 Pulgadas
- ☐ 5 a 5.4 Pulgadas
- ☐ 5.5 Pulgadas y Más

1 - Marcas

2 - Catálogo de celulares

Cámara Trasera 50 MP, 8 GB RAM + 256 GB ROM...
✓prime



Huawei P30 Lite Smartphone, 128 GB, Desbloqueado, 6.15 pulgadas, Color Azul Orquidea

★★★★★ v 2,099

\$5,399⁰⁰ \$6,999.00

✓prime Recíbelo el jueves, 14 de mayo
Envío GRATIS por Amazon México

Leica Quad Camera, 256 GB / 8 GB, Desbloquead...
★★★★★ 864
✓prime



Xiaomi Redmi Note 8 (4gb 64gb) 6.3" Neptune Blue

★★★★★ v 7,658

\$3,999⁰⁰ \$5,700.00

✓prime Envío GRATIS por Amazon México

Curva de 6.53" (Kirin 990, 8 + 256 GB, Cuádruple...
✓prime

Amazon's Choice



Xiaomi Redmi 8A (32GB, 2GB RAM) 6.22" visualización HD, Snapdragon 439, Dual SIM gsm Desbloqueado - US & Global 4G LTE versión Internacional...

★★★★★ v 635

\$2,988⁰⁰ \$3,500.00

✓prime Recíbelo el jueves, 14 de mayo
Envío GRATIS por Amazon México

Más vendido



Xiaomi Redmi Note 8 (4gb 64gb) 6.3" Negro meteoro

★★★★★ v 6,783

\$3,898⁰⁰ \$5,700.00

✓prime Recíbelo el jueves, 14 de mayo
Envío GRATIS por Amazon México

3. Código HTML

```

::before
<div id="s-skipLinkTargetForMainSearchResults" tabindex="-1"></div>
<span data-component-type="s-top-slot" class="rush-component" data-component-id="5">
  <div data-uuid="7ebf39ca-3115-4a15-b9d8-237a92dced67">
    <div class="a-section s-border-bottom">
      <span cel_widget_id="TOP-SAFE_FRAME" class="celwidget slot=TOP template=SAFE_FRAME pd_rd_w=HIqst widget=loom-desktop-top-slot_hsa-5581007350301 pf_rd_p=4f101435-4272-4c6b-82a6-59afb7a5a5ad pd_rd_wg=69UDR pf_rd_r=94947HRA6SN55ZS0G72F pd_rd_r=0b7140fb-6859-45ef-980d-789ba58832c4 widgetId=loom-desktop-top-slot_hsa-5581007350301" data-cel-widget="TOP-SAFE_FRAME">
        <style></style>
html body #a-page #search div div div span div div span #32d55a1e-1b4a-4a23-955a-1e1b4afa232c div div div a div.sb_3pR2krbd div.sb_275e6iZL span.sb_34-gYlW1 span.sb_3XafbQbXsb_25yic0YU.sb_CyFsQ0hU

```

Styles

```

:hov .cls
element.s
style {
  -webkit-
  line-
  clamp
  : 2;
  -webkit-

```

Creamos un DataSet en R

DataSet Extraído de Amazon



Empresa Patito

	title	details	edition	color	seller	price
zon.com.mx/Smartphone-Huawei-P30-Lite...	...	Tamaño de la pantalla: 6,15 pulga...	International Version	Azul Orquídea	Huawei	\$5,399.00
zon.com.mx/Moto-E6-Plus-extra%C3%ADb...	...	Modelo internacional. Compatibi...	NA	Grafito	Motorola	\$4,230.00
zon.com.mx/Xiaomi-Redmi-Note-64gb-Dr...	...	Pantalla: Grande 6,3 pulgadas, 10...	NA	NA	Xiaomi	\$4,199.00
zon.com.mx/Xiaomi-Redmi-visualizaci%C3...	...	visualización: pantalla táctil) capa...	NA	NA	Xiaomi	\$2,989.00
zon.com.mx/Huawei-FullView-Smartphone...	...	Pantalla FullView de 5,45" 3, 020...	NA	Negro	Huawei	\$1,849.00
zon.com.mx/Xiaomi-Redmi-Note-Negro-m...	...	Pantalla: Grande 6,3 pulgadas, 10...	NA	NA	Xiaomi	\$4,097.00
zon.com.mx/Samsung-Galaxy-desbloquea...	...	Pantalla: 164,0 mm (6,5"), Pantall...	NA	Negro	Samsung	\$7,163.99
zon.com.mx/Samsung-pantalla-pulgadas-1...	...	Compatible con la mayoría de op...	NA	NA	Samsung	\$3,399.00
zon.com.mx/SAMSUNG-4203040-NEGRO-...	...	Modelo internacional. Compatibi...	NA	Negro	Samsung	\$4,169.00
zon.com.mx/pantalla-pulgadas-c%C3%A1...	...	Visualización: LTPS IPS LCD 6,59 ...	NA	NA	Huawei	NA
zon.com.mx/Motorola-KT1922-5-Pantalla-...	...	UMTS/HSPA+ (B1, 2, 4, 5, 8) - 4G ...	NA	Índigo profundo	Motorola	\$2,999.00
zon.com.mx/Huawei-JKM-LX3-Smartphone...	...	Color Negro MicroSD expandible...	NA	Negro	Huawei	\$4,499.00
zon.com.mx/Samsung-Galaxy-desbloquea...	...	Pantalla: 164,0 mm (6,5"), Pantall...	NA	NA	Samsung Electronics	\$7,198.00
zon.com.mx/Tel%C3%A9fono-desbloquead...	...	Batería de Iones de litio de 3800 ...	NA	Azul	Cobcob cell phone	\$791.88
zon.com.mx/Huawei-DBU-LX3-Smartphon...	...	Color Azul MicroSD expandible ...	NA	Azul	Huawei	\$3,250.00
zon.com.mx/gp/video/ssoredirect/?ru=htt...	NA	NA	NA	NA	NA	NA
zon.com.mx/Xiaomi-Redmi-128GB-Interest...	...	Qualcomm Snapdragon 720G, Pr...	NA	NA	Xiaomi	\$5,878.00
zon.com.mx/Samsung-Galaxy-Triple-Camar...	...	Pantalla Infinity-V Super AMOLE...	NA	Bianco	Samsung	\$5,299.00
zon.com.mx/Huawei-Y5-NEO-Smartphone...	...	Pantalla FullView de 5,45" 3,020 ...	NA	NA	Huawei	\$1,899.00
zon.com.mx/Xiaomi-pulgadas-Smartphone...	...	HSDPA 650/900/2100 - LTE B1 (2...	NA	Azul (Neptune Blue)	Xiaomi	\$4,499.00
zon.com.mx/Xiaomi-Redmi-64GB-Dual-Sl...	...	Doble Sim	NA	NA	Xiaomi	\$3,850.00
zon.com.mx/Motorola-Moto-plus-1965-Ro...	...	Cámara 16+5 MP Dual LED Flash...	NA	Deep Indigo	Motorola	NA

Limpiamos la información

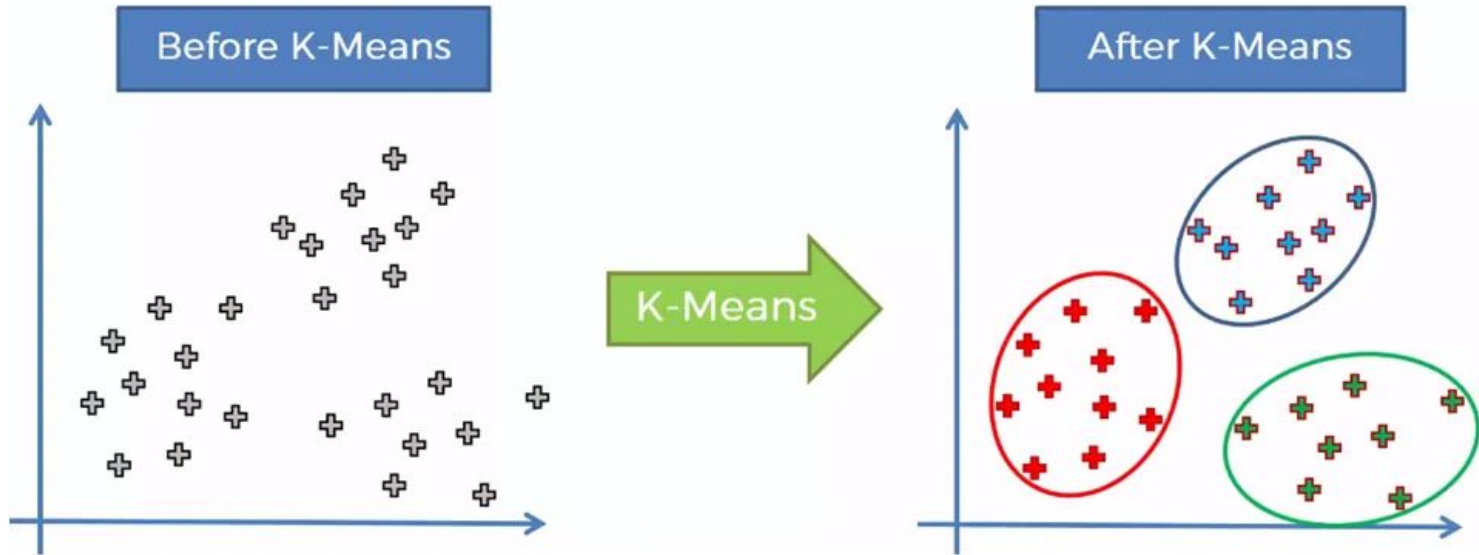
Aplicamos transformaciones para limpiar los datos

url	title	details	edition	color	seller	price	rating
https://www.amazon.com.mx/Smartphone-Huawei-P30-Lite...	Huawei P30 Lite Smartphone, 128 GB, Desbloqueado, 6.15 p...	Tamaño de la pantalla: 6.15 pulgadas MicroSD expandible a ...	International Version	Azul Orquidea	Huawei	5399.00	4.2
https://www.amazon.com.mx/Moto-E6-Plus-extra%C3%ADb...	Motorola Moto E6 Plus (64 GB, 4 GB) 6.1 pulgadas, batería e...	Modelo internacional. Compatible con la mayoría de operad...	NA	Grafito	Motorola	4230.00	4.1
https://www.amazon.com.mx/Xiaomi-Redmi-Note-64gb-Dr...	Xiaomi Redmi Note 8 (4gb 64gb) 6.3" Neptune Blue	Pantalla: Grande 6,3 pulgadas, 1080 x 2340 píxeles Chipset: ...	NA	NA	Xiaomi	4199.00	4.1
https://www.amazon.com.mx/Xiaomi-Redmi-visualizaci%C3...	Xiaomi Redmi 8A (32GB, 2GB RAM) 6.22" visualización HD, S...	visualización: pantalla táctil capacitiva IPS LCD de 6,22 pulga...	NA	NA	Xiaomi	2989.00	4.0
https://www.amazon.com.mx/Huawei-FullView-Smartphone...	HUAWEI Y5 2018 Smartphone 4.5" HD, 16GB, 3020mAh Batt...	Pantalla FullView de 5.45" 3, 020 mAh GRAN BATERÍA 1 GB ...	NA	Negro	Huawei	1849.00	4.1
https://www.amazon.com.mx/Xiaomi-Redmi-Note-Negro-m...	Xiaomi Redmi Note 8 (4gb 64gb) 6.3" Negro meteoro	Pantalla: Grande 6,3 pulgadas, 1080 x 2340 píxeles Chipset: ...	NA	NA	Xiaomi	4097.00	4.1
https://www.amazon.com.mx/Samsung-Galaxy-desbloquee...	Samsung Galaxy A51 GSM desbloqueado, Prisma Crush Ne...	Pantalla: 164.0 mm (6.5"), Pantalla Infinity-O, Super AMOLED...	NA	Negro	Samsung	7163.99	4.1
https://www.amazon.com.mx/Samsung-pantalla-pulgadas-l...	Samsung Galaxy A10S A107M 32GB Unlocked GSM DUOS P...	Compatible con la mayoría de operadores GSM como T-Mo...	NA	NA	Samsung	3399.00	4.0
https://www.amazon.com.mx/SAMSUNG-4203040-NEGRO-...	SAMSUNG 4203040-NEGRO Galaxy A20s, 32GB, Dual Sim, 3...	Modelo internacional. Compatible con la mayoría de operad...	NA	Negro	Samsung	4169.00	4.2
https://www.amazon.com.mx/pantalla-pulgadas-c%C3%A11...	Huawei Y9 Prime 2019 (128 GB, 4 GB RAM) pantalla de 6.59 ...	Visualización: LTPS IPS LCD 6.59 pulgadas 1080 x 2340 píxel...	NA	NA	Huawei	4156.55	4.2
https://www.amazon.com.mx/Motorola-XT1922-5-Pantalla-...	Motorola Moto G6 Play Dual Sim 32GB XT1922-5 Pantalla ...	UMTS/HSPA+ (B1, 2, 4, 5, 8) - 4G LTE LTE B1(2100), B2(1900),...	NA	Índigo profundo	Motorola	2999.00	4.1
https://www.amazon.com.mx/Huawei-JKM-LX3-Smartphone...	Huawei Y9 2019 - Smartphone 6.5" FHD+, 64GB, 4000mAh ...	Color Negro MicroSD expandible a 400 GB Lector de huella ...	NA	Negro	Huawei	4499.00	4.1
https://www.amazon.com.mx/Samsung-Galaxy-desbloquee...	Samsung Galaxy A51 GSM desbloqueado, 128 GB, Prisma Cr...	Pantalla: 164.0 mm (6.5"), Pantalla Infinity-O, Super AMOLED...	NA	NA	Samsung Electronics	7198.00	4.2
https://www.amazon.com.mx/Tel%C3%A9fono-desbloquead...	Teléfono celular desbloqueado 2019, ultradelgado Android ...	Batería de iones de litio de 3800 mAh para toneladas de tie...	NA	Azul	Cobcob cell phone	791.88	2.0
https://www.amazon.com.mx/Huawei-DBU-LX3-Smartphon...	Huawei DBU-LX3 Y7 - Smartphone de 6.26", Azul	Color Azul MicroSD expandible a 512 GB Lector de huella	NA	Azul	Huawei	3250.00	4.1
https://www.amazon.com.mx/gp/video/ssoredirect/?ru=htt...	NA	NA	NA	NA	NA	4156.55	4.2
https://www.amazon.com.mx/Xiaomi-Redmi-128GB-Interest...	Xiaomi Redmi Note 9S 128GB Gris Interestelar Desbloqueado	Qualcomm Snapdragon 720G, Procesador de 8 nm con hast...	NA	NA	Xiaomi	5878.00	4.2
https://www.amazon.com.mx/Samsung-Galaxy-Triple-Camar...	Samsung Galaxy A30S Teléfono GSM desbloqueado, 64GB + ...	Pantalla Infinity-V Super AMOLED de 6.4 pulgadas, 720 x 15...	NA	Bianco	Samsung	5299.00	4.1
https://www.amazon.com.mx/Huawei-Y5-NEO-Smartphone...	Huawei Y5 NEO - Smartphone 5.45" HD, 16GB, 3020mAh Ba...	Pantalla FullView de 5.45" 3,020 mAh GRAN BATERÍA 1 GB R...	NA	NA	Huawei	1899.00	4.1
https://www.amazon.com.mx/Xiaomi-pulgadas-Smartphone...	Xiaomi Redmi Note 8 64 GB + 4 GB de RAM, 6.3 pulgadas L...	HSDPA 850/900/2100 - LTE B1 (2100) B2 (1900 PCS) B3 (180...	NA	Azul (Neptune Blue)	Xiaomi	4499.00	4.1
https://www.amazon.com.mx/Xiaomi-Redmi-64GB-Dual-Si...	Xiaomi Redmi 8 4G 64GB 4GB RAM Dual-SIM Blue EU	Doble Sim	NA	NA	Xiaomi	3850.00	4.2
https://www.amazon.com.mx/Motorola-Moto-plus-1965-Ro...	Motorola, Moto G7 plus XT-1965 Moto G7 plus, Deep Indigo	Cámara 16+5 MP Dual LED Flash, Cámara Frontal 12 MP + F...	NA	Deep Indigo	Motorola	4156.55	4.1

Algoritmo K-Means (Algoritmo de Aprendizaje no supervisado)

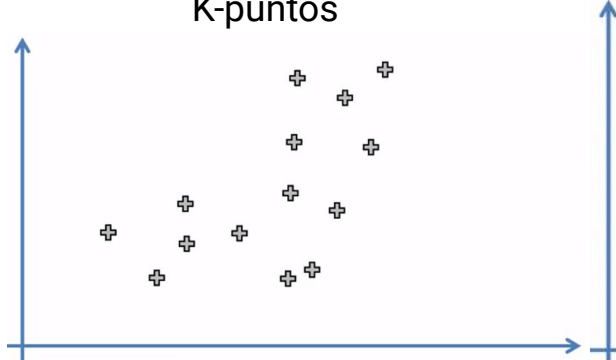
Dónde k , es la cantidad de grupos o clústers que se estarían encontrando

Los círculos son centroides que miden la distancia entre cada uno de los puntos para agrupar

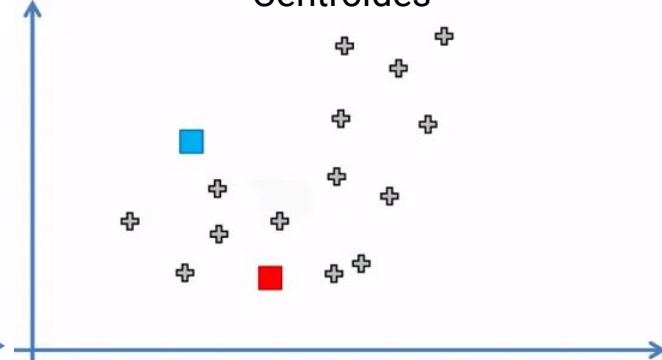


Funcionamiento del algoritmo

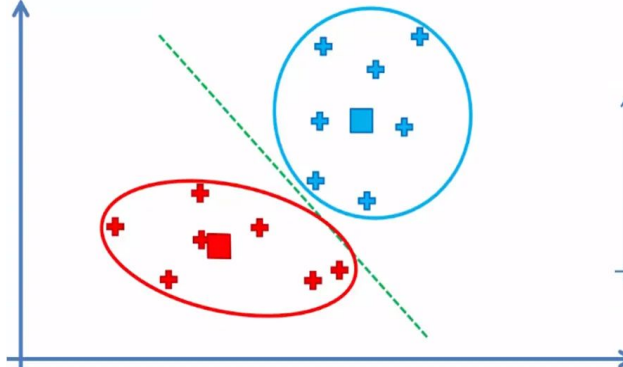
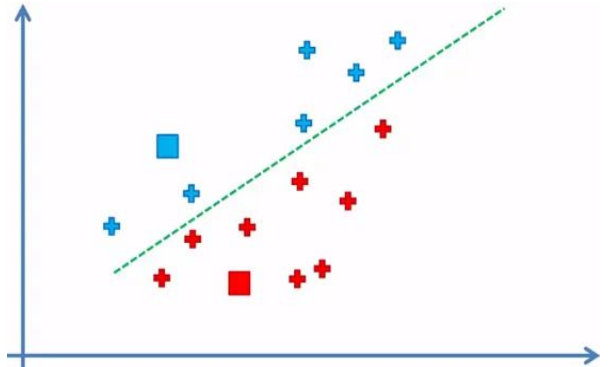
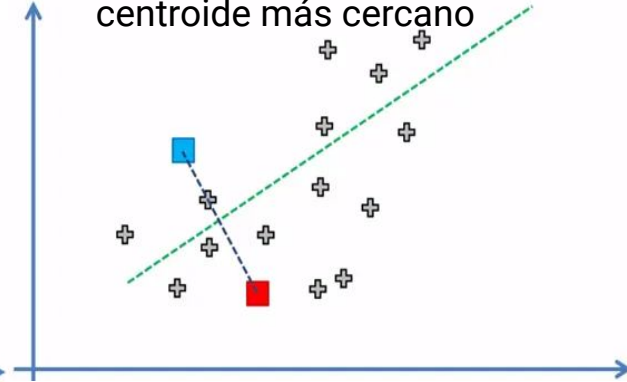
K-puntos



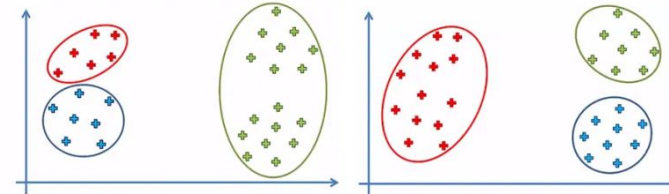
Centroides



A cada punto se asigna el
centroide más cercano



Clasificación encontrada



Resultados del análisis

Servidor Web en R usando Shiny

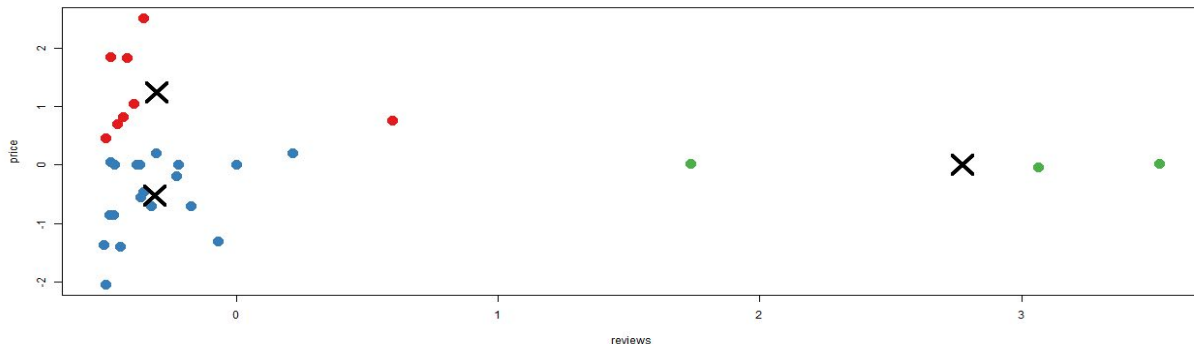
Mediante el algoritmo K-means podemos buscar patrones en las variables y podemos clasificar los datos que tenemos

Phones k-means clustering

X Variable
reviews

Y Variable
price

Cluster count
3

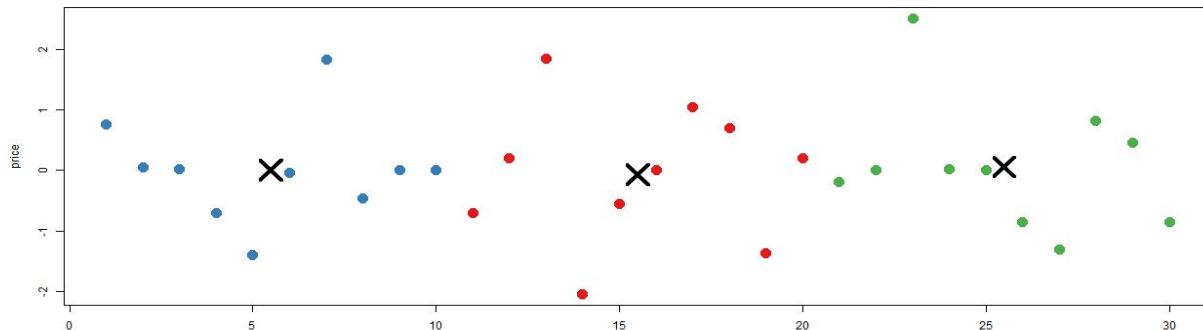


Phones k-means clustering

X Variable
X

Y Variable
price

Cluster count
3



Clasificador RealTime

Se montó una API que ejecuta el modelo con Plumber en R

Mediante los datos de entrada el algoritmo realiza una clasificación de forma automática

The screenshot displays the Swagger UI for the `/getCluster` endpoint. The interface is divided into several sections:

- GET /getCluster**: The endpoint name and method.
- Parameters**: A table of query parameters for the endpoint.
- Code**: The HTTP status code and description.
- Details**: A section containing the response body and headers.

Name	Description
weight string (query)	undefined 1.25
questions string (query)	undefined 32
reviews string (query)	undefined 6
rating string (query)	undefined 2.5
price string (query)	undefined 5000

The **Details** section shows the response body and headers. A red circle highlights the response body, which contains the following JSON:

```
[ 1 ]
```

The response headers are:

```
connection: close  
content-length: 3  
content-type: application/json  
date: Sun, 10 May 2020 08:47:33 GMT, dom., 10 may. 2020 08:47:33 p. m. GMT
```

Conclusión

A mayor cantidad de datos los resultados pueden ser más precisos, sin embargo el proceso de Scraping suele ser tardado, hay información en el detalle de los productos que se puede separar y encontrar o completar campos faltantes. Sin embargo, la muestra tomada dió una perspectiva general para concluir un flujo completo de Data Science con el lenguaje R.

Extracción->Limpieza->Análisis->Entrenamiento->Modelado
->Implementación