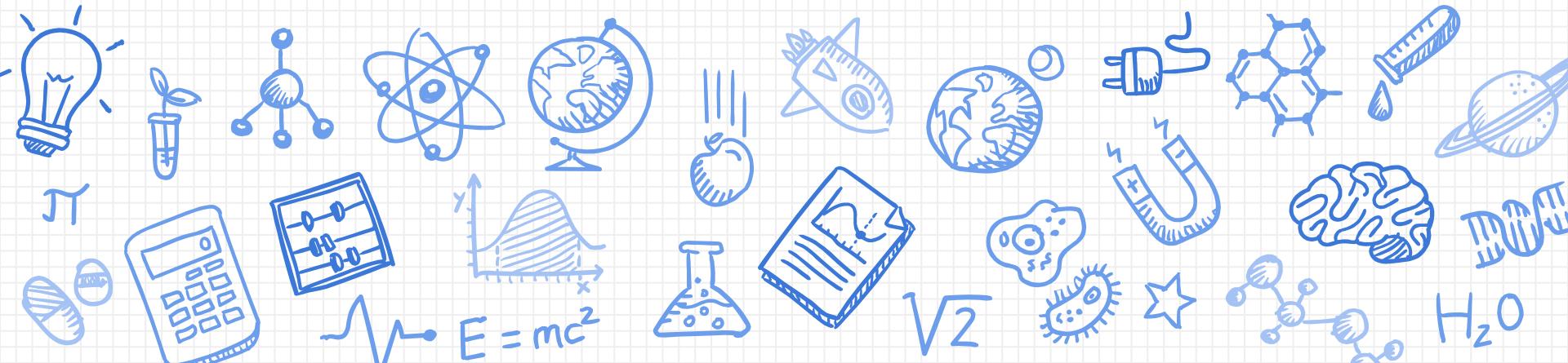


# Data Science en la Industria



# Soy Cristhian Boujon

- Ingeniero en Sistemas de Información.
- Tech Lead en Mercado Libre.
- Profesor de Inteligencia Artificial en Tec. en Diseño y Programación de Videojuegos



# Sobre la Charla

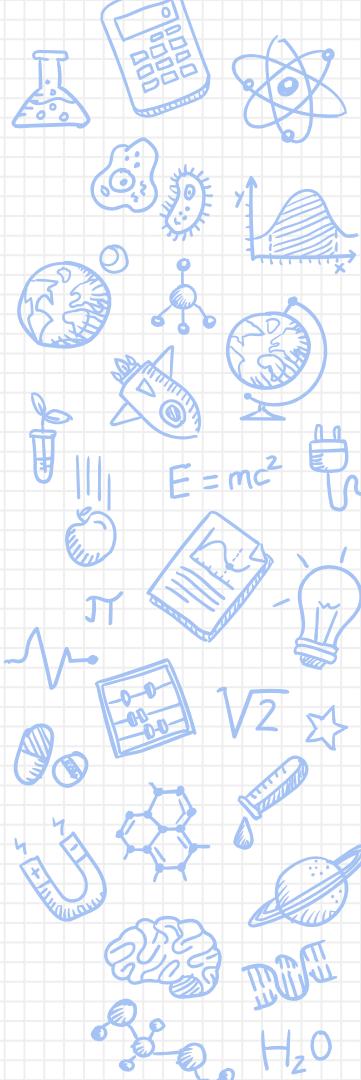
---

Lo que **sí**:

1. Análisis de datos.
2. Machine Learning.
3. Ingeniería de datos.
4. Jobs & Skills

Lo que **no**:

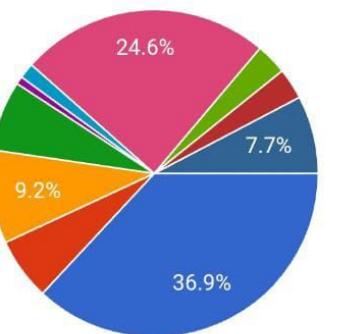
1. Un taller de programación.
2. Profundizar en conceptos técnicos



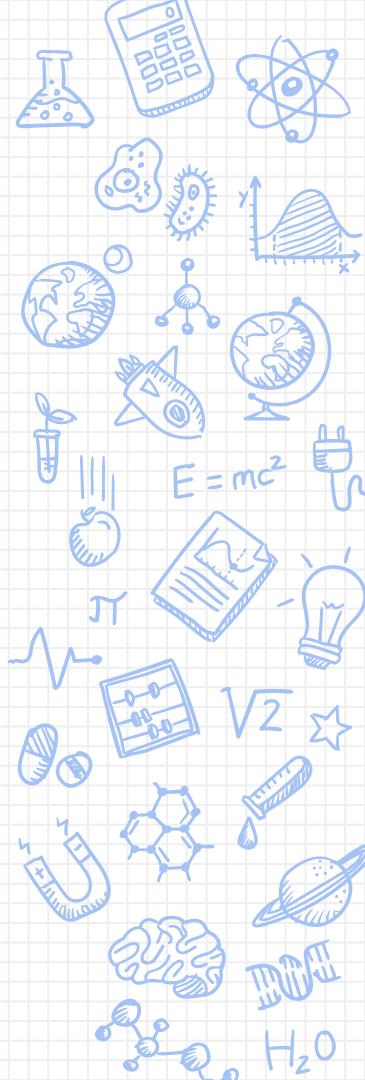
# Sobre la Charla

¿Cuál de los siguientes perfiles te define mejor?

130 respuestas



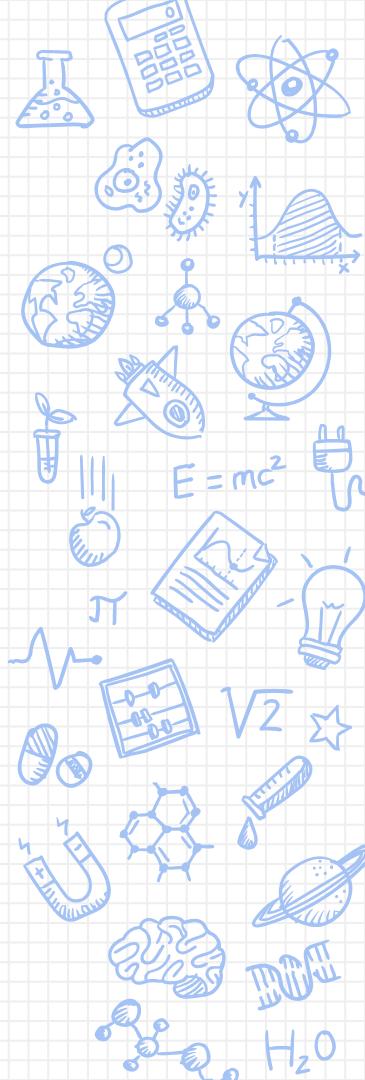
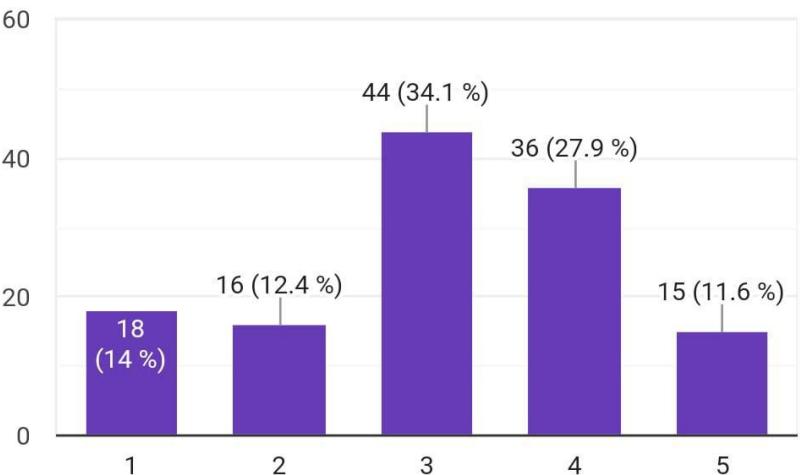
- Profesional del área informática
- Profesional del área de ciencias económicas
- Profesional de otras disciplinas
- Docente carreras de informática
- Docente carreras de ciencias económicas
- Docente de otras carreras
- Estudiante carreras de informática
- Estudiante carreras de ciencias económicas



# Sobre la Charla

¿Cuál es el nivel de conocimiento que tenes de Programación?

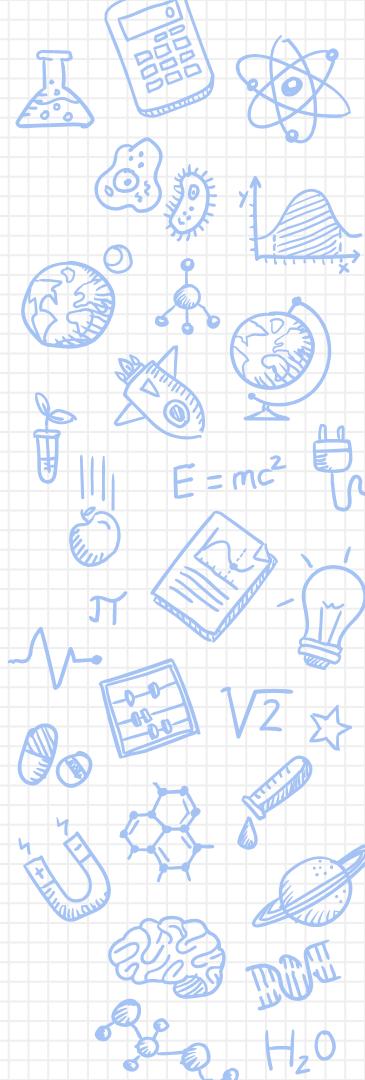
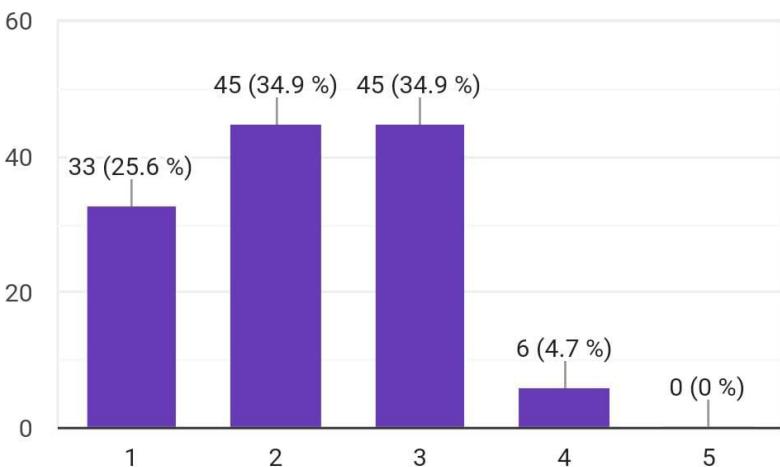
129 respuestas



# Sobre la Charla

¿Cuál es el nivel de conocimiento que tenes de Ciencia de Datos?

129 respuestas



# Desarrollo web hace 10 años atrás...

# HTTP

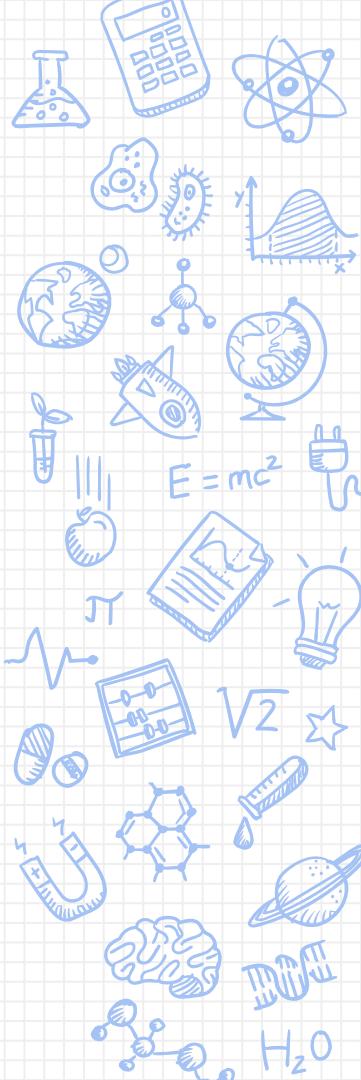
# HTML

CSS

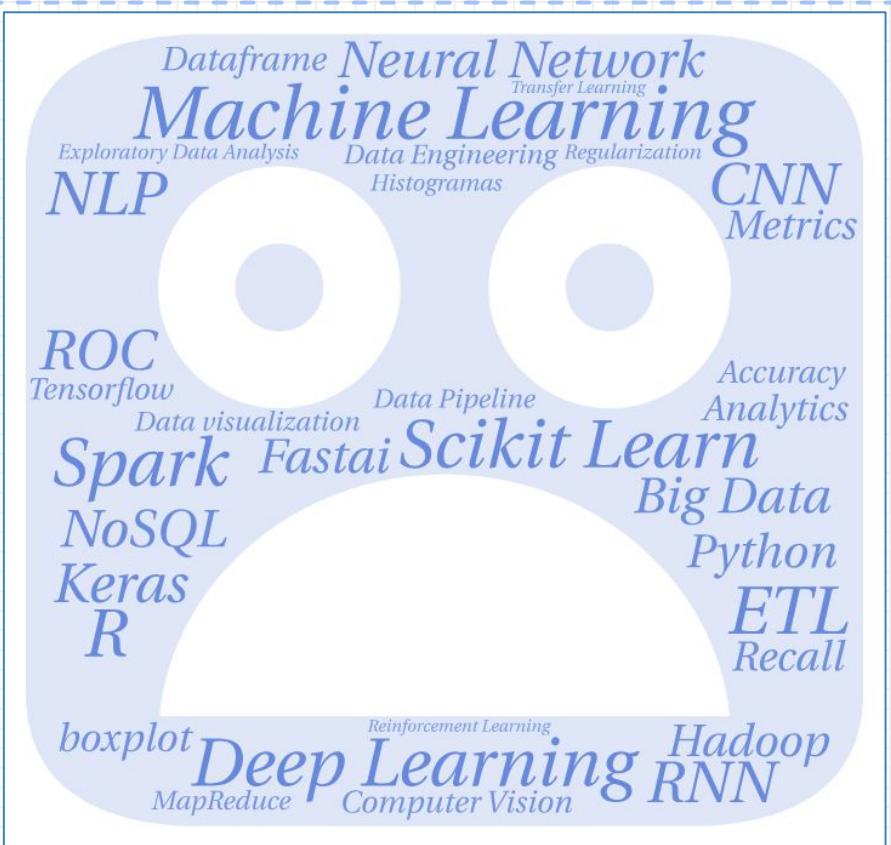
# Javascript

# PHP

# MySQL

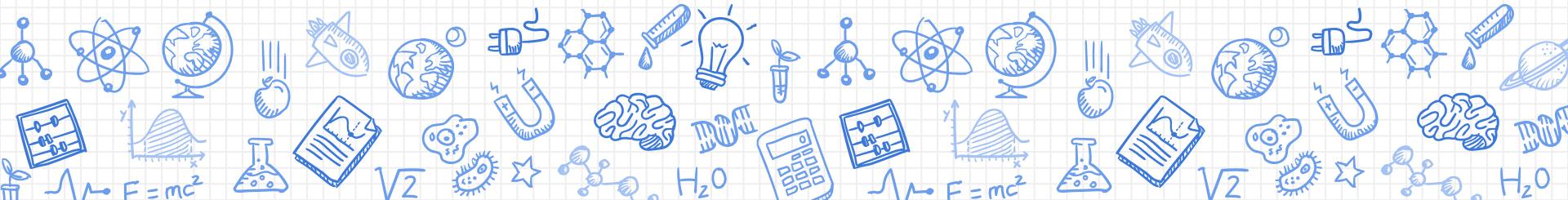


# Data Science



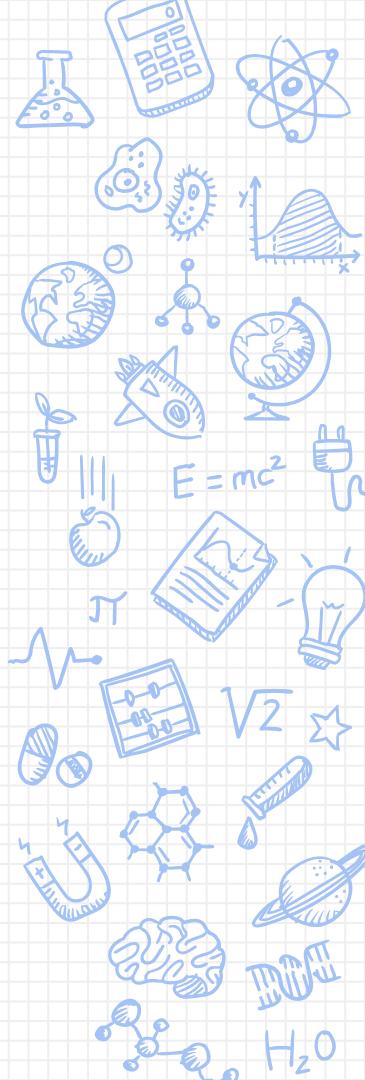
# Análisis de Datos

Se exploran y analizan los datos y así sacar conclusiones sobre la información.



# Análisis de Datos – Conocimiento de Dominio

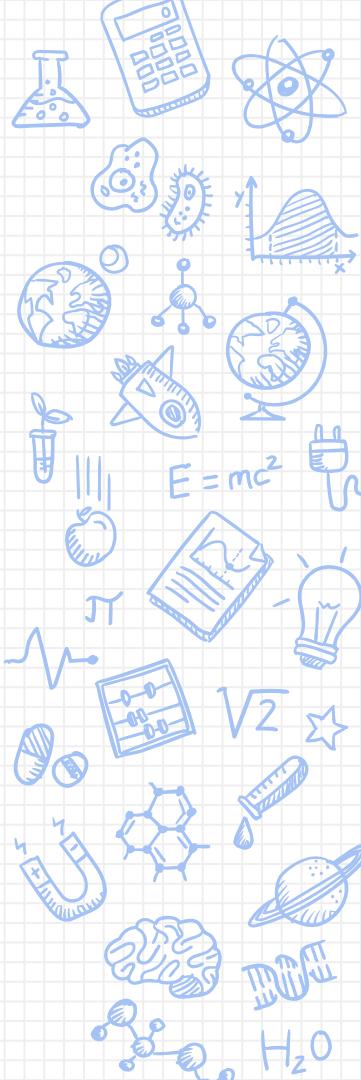
- Conocimientos de Negocio.
  - Necesidad del negocio → Preguntas
  - Planteo de Hipótesis

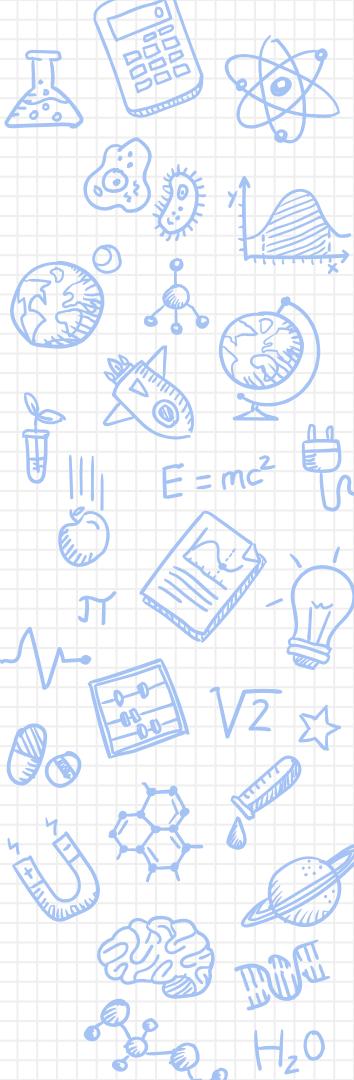


# Análisis de Datos – Exploración

Consiste en entender los datos con los que contamos.

- ¿Qué datos tengo?
  - ¿Cuántos datos tengo?
  - ¿Faltan datos?
  - ¿Hay datos inconsistentes?
  - ¿Hay valores atípicos?
  - ¿Qué distribución tienen los datos?

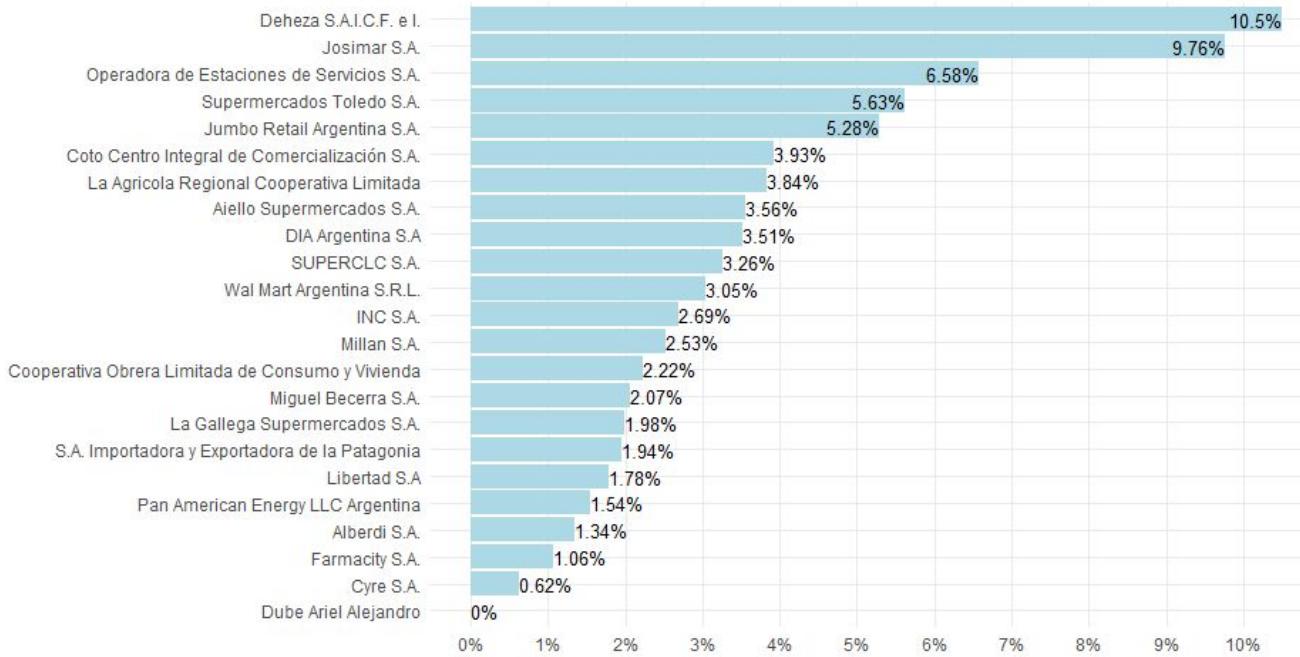




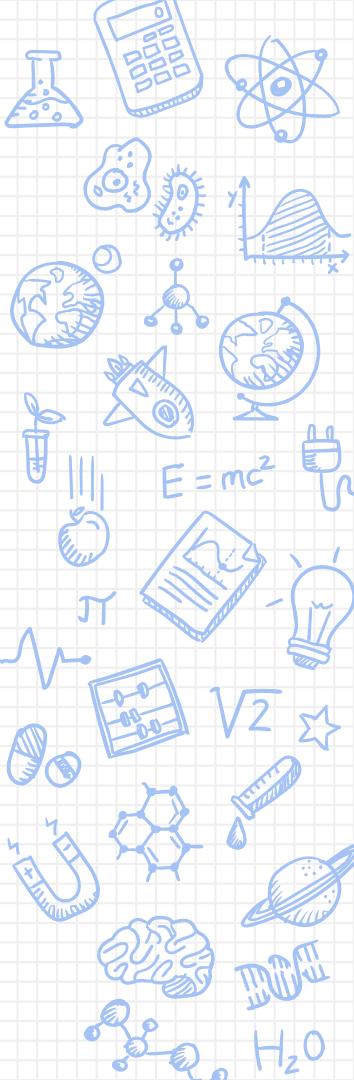
# Análisis de Datos – Análisis

## Aumento por Empresa (Abr-Mayo 2018)

Aumento promedio de productos por empresas dueñas de cadenas de super.



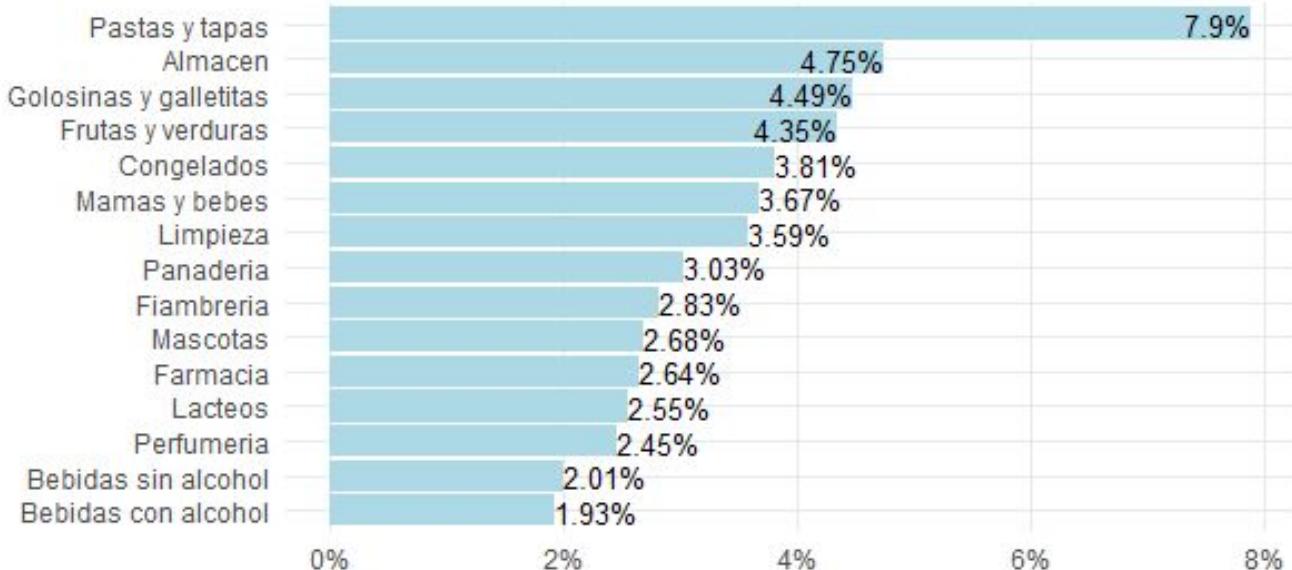
Fuente: Open Data Córdoba en base a relevamiento propio del programa Precios Claros.



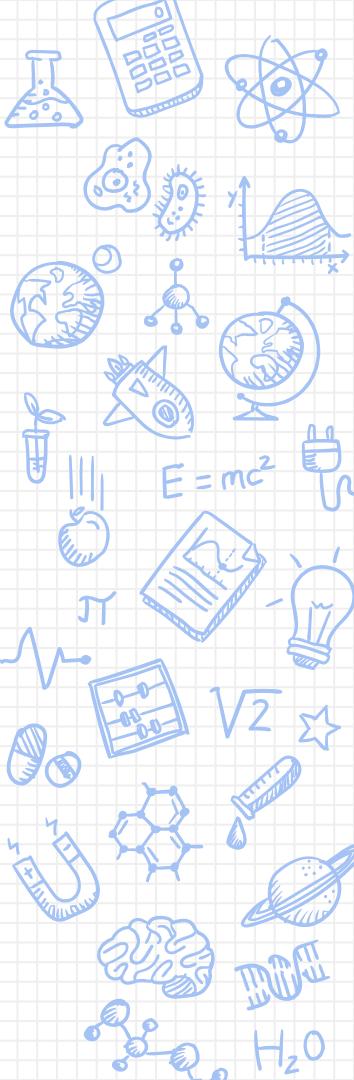
# Análisis de Datos – Análisis

## Aumento por Categoría (Abr-Mayo 2018)

Aumento promedio de productos por Categoría del Producto.



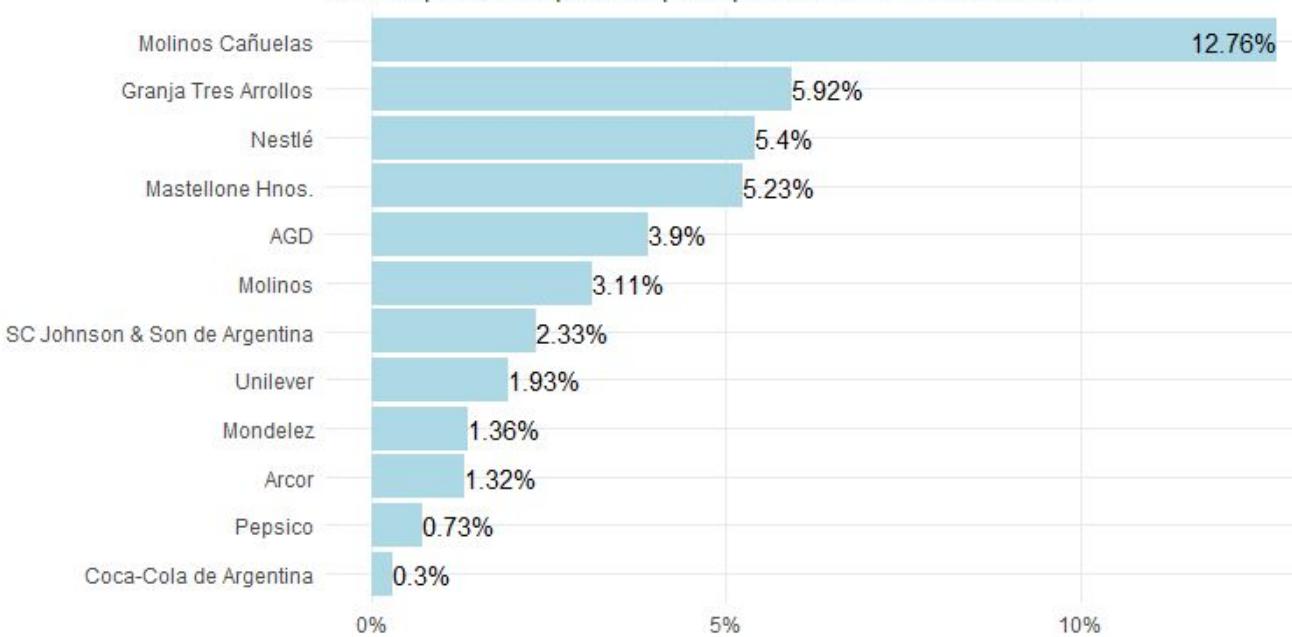
Fuente: Open Data Córdoba en base a relevamiento propio del programa Precios Claros.



# Análisis de Datos – Análisis

## Aumento por Fabricante

Aumento promedio de productos por empresa fabricante. Provincia Córdoba.

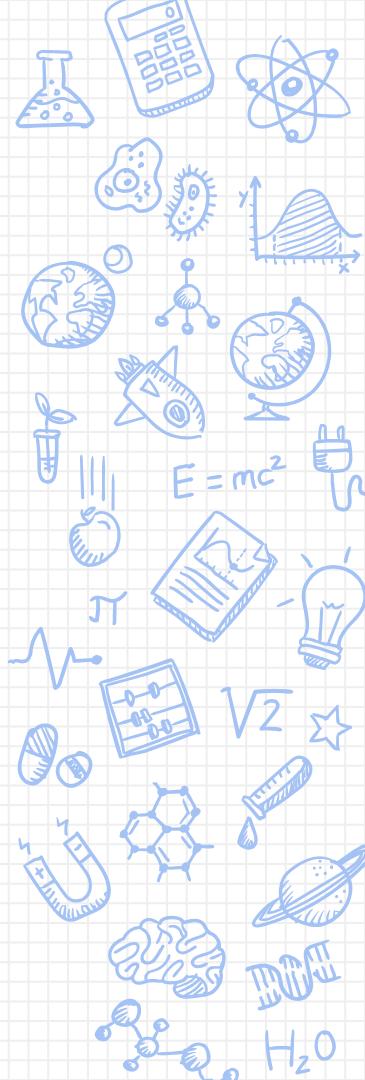


Fuente: Open Data Córdoba en base a relevamiento propio del programa Precios Claros.

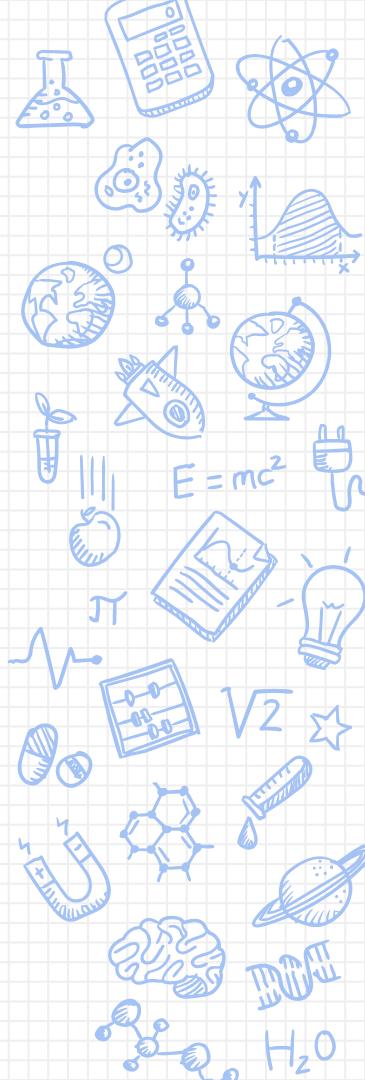
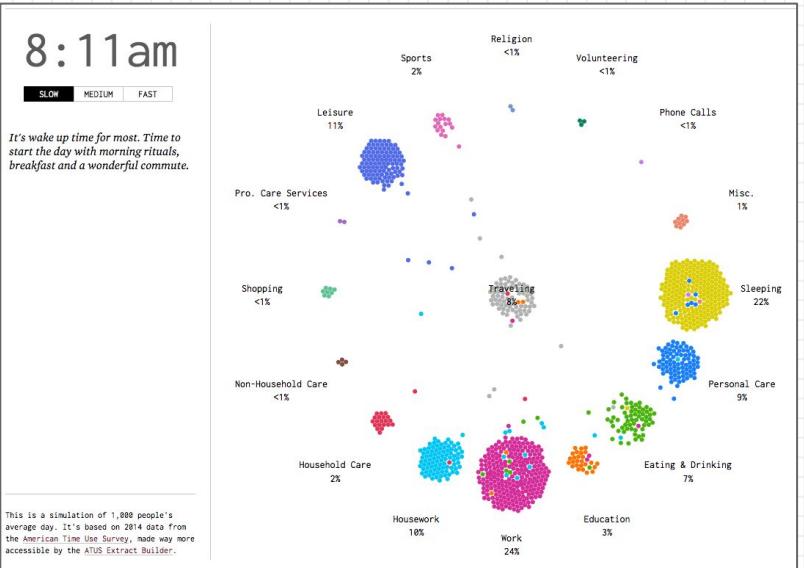
# Análisis de Datos – Interpretación

Dar significado a los datos para que aporten valor al negocio.

- Reportes
  - Insights
  - Data storytelling
  - Simulaciones
  - Infografías



# Análisis de Datos – Simulación



# Análisis de Datos – Infografías

## 3 de cada 10 usuarios **ADICTOS** a las **REDES SOCIALES**

### ADICTOS A LAS REDES SOCIALES

3 de cada **10** usuarios  
Entre 15 y 25 años

**54%** Hombres



TIPOS DE CLASE ECONÓMICA  
**A, B, C Y D+.**  
85% de los adictos

**ntmx** Diseño y Arte: Oscar Aguilar  
Investigación y Redacción: Mónica Fuentes

Fuente: Notimex,  
Sociedad Mexicana de Autores de las Artes Plásticas (Somaap)

Cada vez más personas están siendo atrapadas por las redes sociales, ya que tres de cada 10 usuarios de entre 15 a 25 años que usan Twitter o LinkedIn confesaron ser adictos a éstas, de acuerdo con la Sociedad Mexicana de Autores de las Artes Plásticas (Somaap)



### SÍNTOMAS DEL ADICTO

Pérdida de control del uso de redes sociales

Episodios de uso compulsivo con gran inversión de tiempo y energía

Daño o progresivo de la calidad de vida

Postergan labores por estar en las redes

La usan pese al daño familiar o así mismo

Gran ansiedad si no se encuentra conectado

Fantasías o sueños acerca de las redes

### RAZONES PARA ESTAR EN LAS REDES

Más de 6 horas

**56%**  
Porque les gusta

**30%**  
Para tener amigos

**17%**  
Para enterarse de lo que pasa

### PREFIEREN REDES A RELACIONES INTERPERSONALES

**80%**  
"Es más barato"

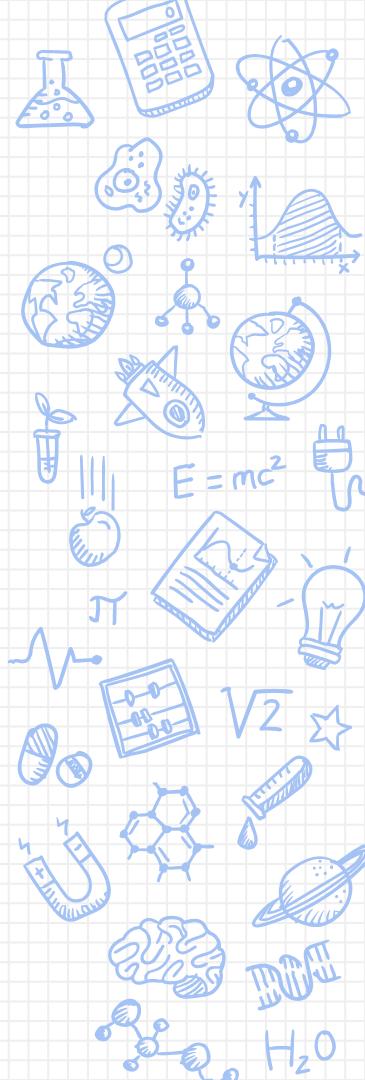
**78%**  
"Resulta ser más cómodo"

**30%**  
"Es seguro"

**14%**  
"Presupone más autenticidad y libertad"

### NOTA IMPORTANTE

En abril pasado, la Asociación dio a conocer que 54 por ciento de los hombres establecieron relaciones humanas virtuales, de las cuales sólo 11 por ciento derivó en relaciones personales



# Tools

---



matplotlib



R Studio®

tableau  
SOFTWARE

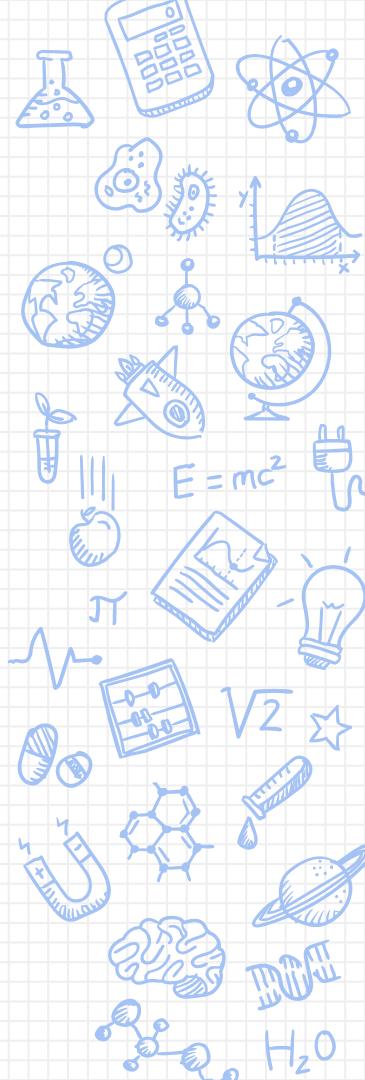
plotly

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

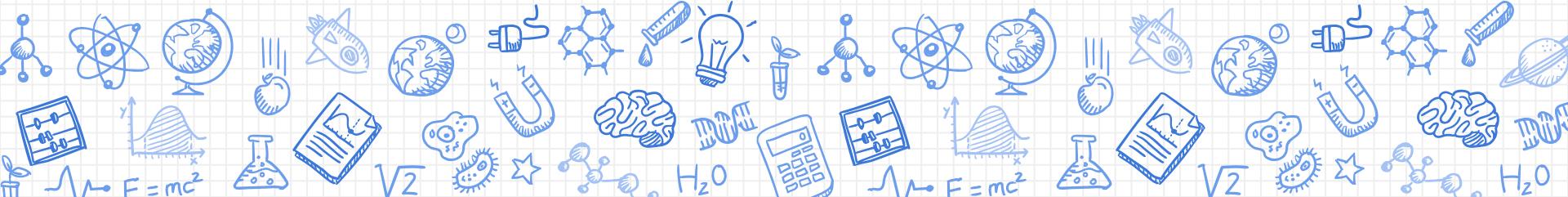


co



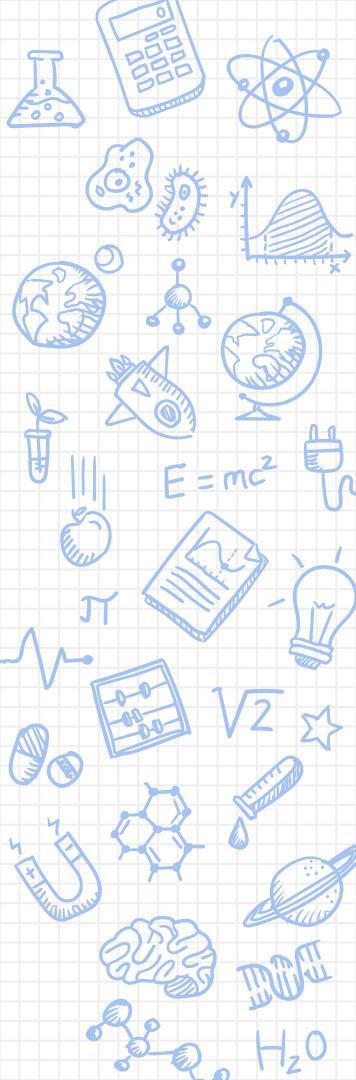
# Machine Learning

Creación de modelos de Machine Learning



# Machine Learning – ¿Para qué sirve?

- Clasificación/Regresión de Datos Estructurados.
  - Series de Tiempo.
  - Procesar el Lenguaje Natural.
  - Computer Vision.
  - Sistemas de Recomendación.
  - Y mucho más!

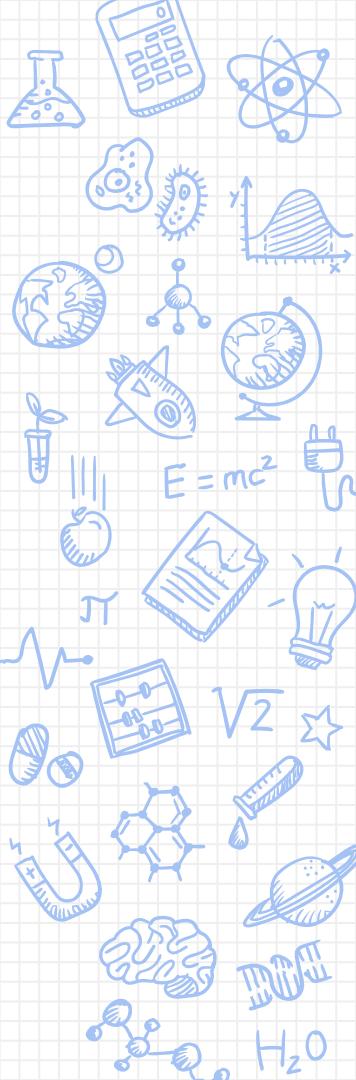


# Machine Learning – Datos Estructurados

# bedrooms	# bathrooms	# surface_to...	# surface_co...	# price	▲ currency	▲ operation_...
1	1			62000	ARS	Alquiler
1	1	45	48	10000	ARS	Alquiler
1	1	40	40	27500	ARS	Alquiler
1	1	35	35	6500	ARS	Alquiler
1	1	47	41	7000	ARS	Alquiler
1	1	40	40	110000	USD	Venta
1	1	48	42	105000	USD	Venta
1	1	48	45	75000	USD	Venta

# age	sex	# bmi	# children	✓ smoker	region	# expenses
19	female	27.9	0	yes	southwest	16884.92
18	male	33.8	1	no	southeast	1725.55
28	male	33.0	3	no	southeast	4449.46
33	male	22.7	0	no	northwest	21984.47
32	male	28.9	0	no	northwest	3866.86
31	female	25.7	0	no	southeast	3756.62
46	female	33.4	1	no	southeast	8240.59

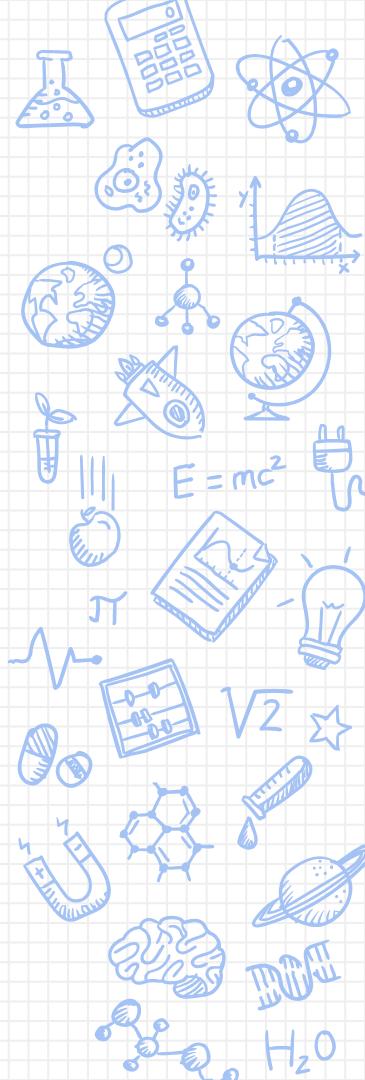
Fuente: <https://www.kaggle.com/noordeen/insurance-premium-prediction>



# Machine Learning – Datos Estructurados

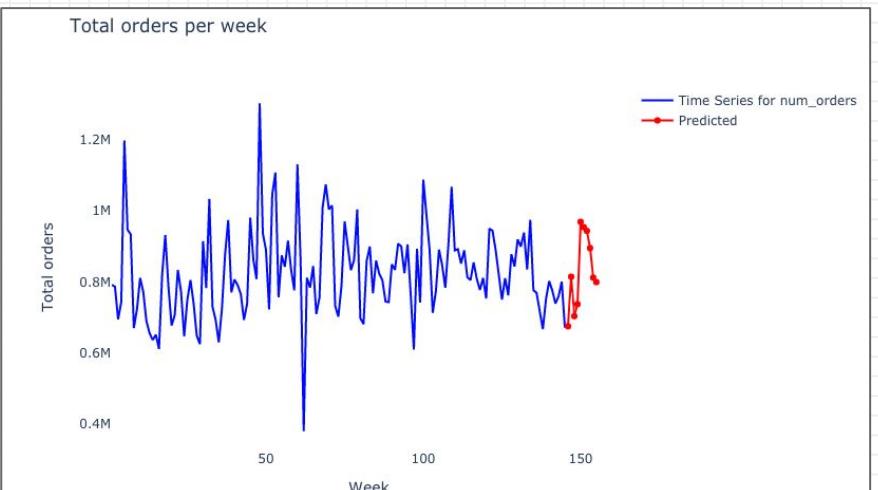
# type	# amount	# oldbalance... oldbalance... newbalance... oldbalance... newbalance... isFraud
PAYOUT	9839.64	170136.0 160296.36 0.0 0.0 0
PAYOUT	1864.28	21249.0 19384.72 0.0 0.0 0
TRANSFER	181.0	181.0 0.0 0.0 0.0 1
CASH_OUT	181.0	181.0 0.0 21182.0 0.0 1
PAYOUT	11668.14	41554.0 29885.86 0.0 0.0 0
PAYOUT	7817.71	53860.0 46042.29 0.0 0.0 0
PAYOUT	7107.77	183195.0 176087.23 0.0 0.0 0
PAYOUT	7861.64	176087.23 168225.59 0.0 0.0 0

# age	# sex	# cp	# chol	# restecg	# thalach	# oldpeak	# target
77	1	0	304	0	162	0	0
76	0	2	197	2	116	1.1	1
74	0	1	269	0	121	0.2	1
71	0	1	302	1	162	0.4	1
71	0	2	265	0	130	0	1
71	0	0	149	1	125	1.6	1
70	1	1	245	0	143	0	1
70	1	0	174	1	125	2.6	0

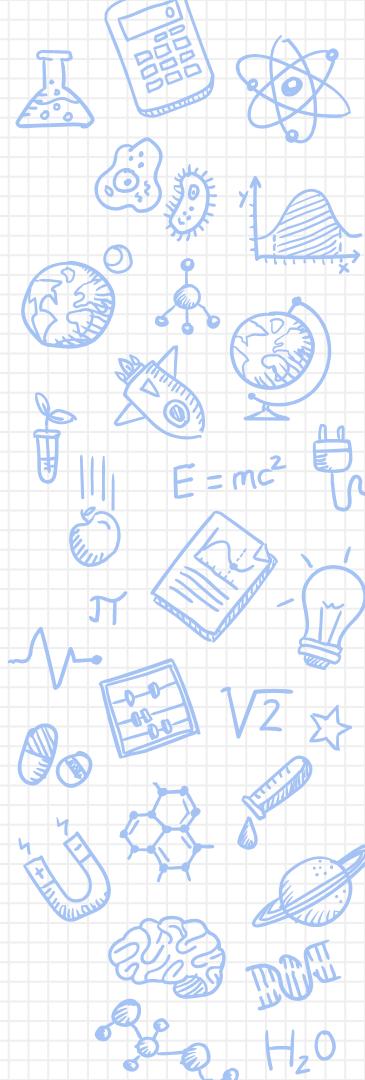


# Machine Learning – Series de Tiempo

# week	center_id	meal_id	# checkout...	# base_price	#emailer_fo...	# homepage...	# num_orders
1	55	1885	136.83	152.29	0	0	177
1	55	1993	136.83	135.83	0	0	278
1	55	2539	134.86	135.86	0	0	189
1	55	2139	339.5	437.53	0	0	54
1	55	2631	243.5	242.5	0	0	40



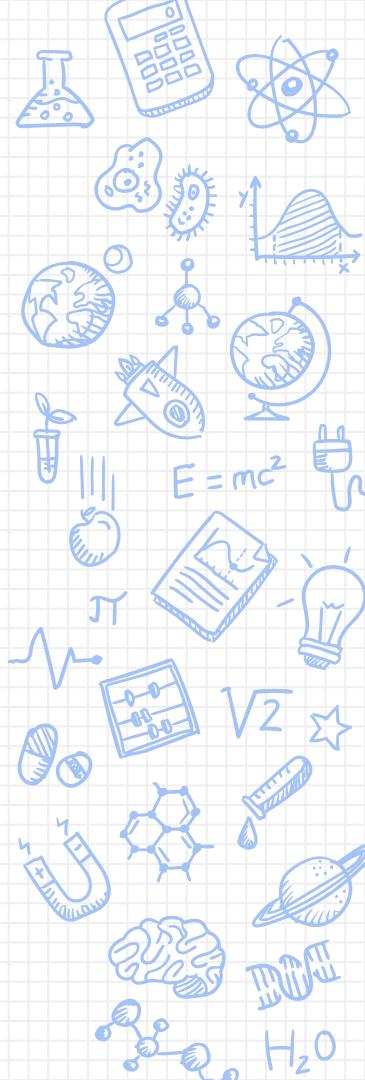
Fuente: <https://www.kaggle.com/ghoshaptarshi/av-genpact-hack-dec2018>



# Machine Learning – Procesamiento de Lenguaje Natural

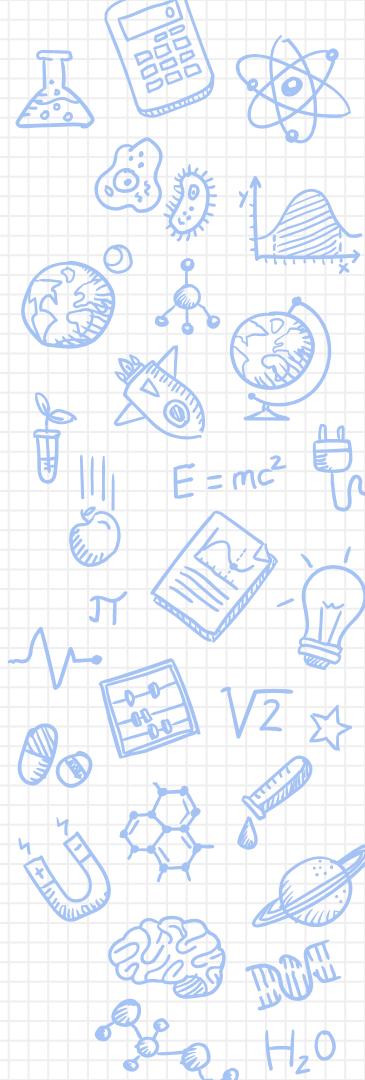
---

- Clasificar preguntas en el Congreso de la Nación
- Análisis de sentimiento en Redes Sociales.
- Generación automática de noticias.
- Traducciones de idiomas.
- Automatización de atención al cliente.



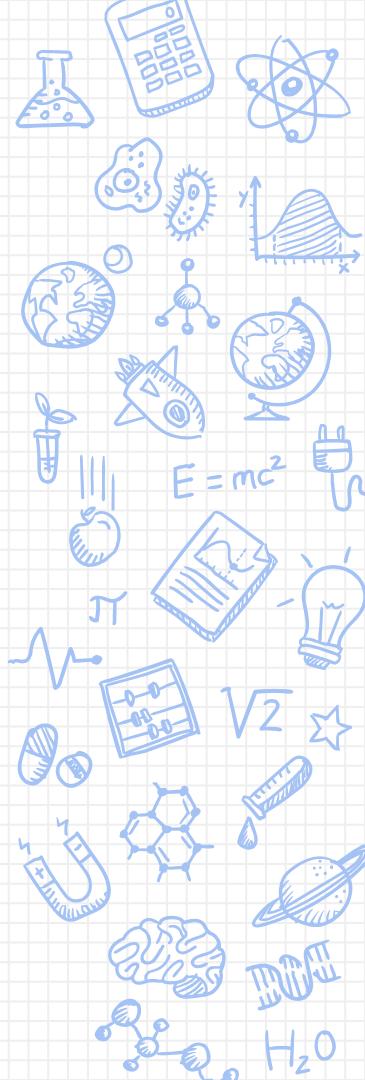
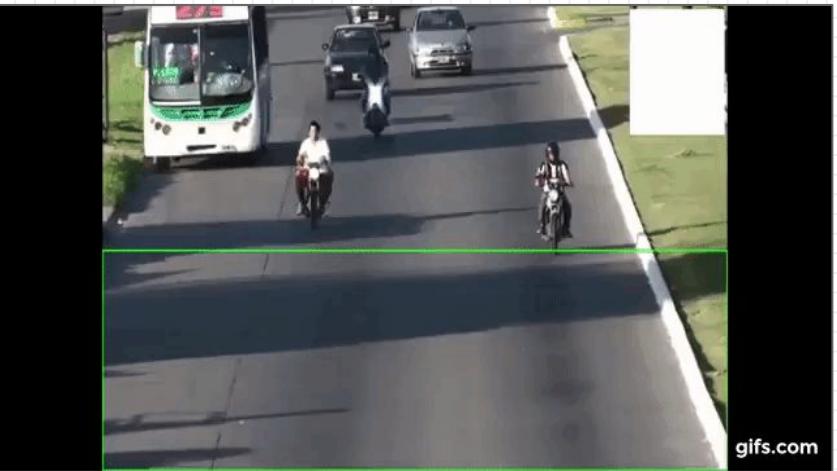
# Machine Learning – Computer Vision

## Reconocimiento Facial.



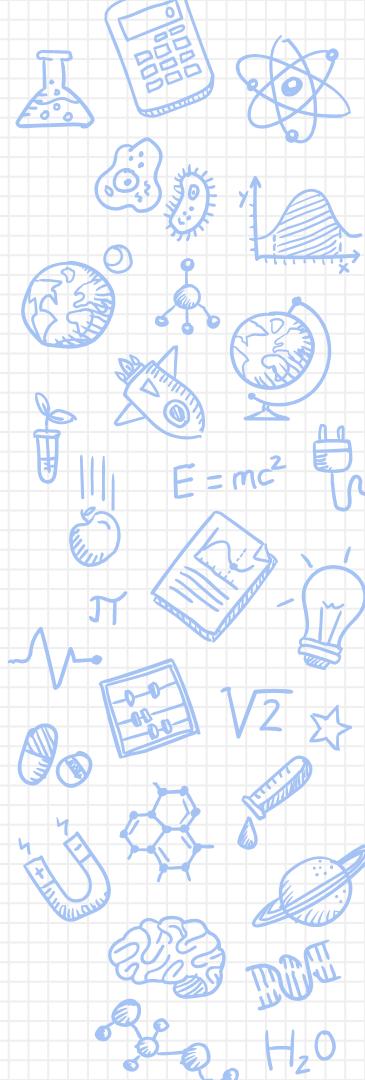
# Machine Learning – Computer Vision

## Seguridad Vial



# Machine Learning – Computer Vision

## Búsquedas visuales



# Machine Learning – Sistemas de Recomendación

**Basadas en lo escuchado recientemente**

Inspirada en tu actividad reciente.

**Los 2000 Argent...**  
100 canciones de la época nostálgica más reciente...

**Musicalizan: Los...**  
Disfruta esta selección musical, de la mano de L...

**Musicalizan: Los...**  
Disfruta esta selección musical, de la mano de L...

**Festival Buena V...**  
Hipódromo de Palermo, Buenos Aires. Sábado 15 ...

**Películas recomendadas**  
de Películas de YouTube

**Ha nacido una estrella**  
2:15:43

**High Plains Drifter**  
1:45:02

**El Angel**  
1:56:28

**¡SCOOBY!**  
1:33:44

**The Social Network**  
2:00:28

**Rampage**  
1:47:12

**Ha nacido una estrella**  
Películas de YouTube Comprar

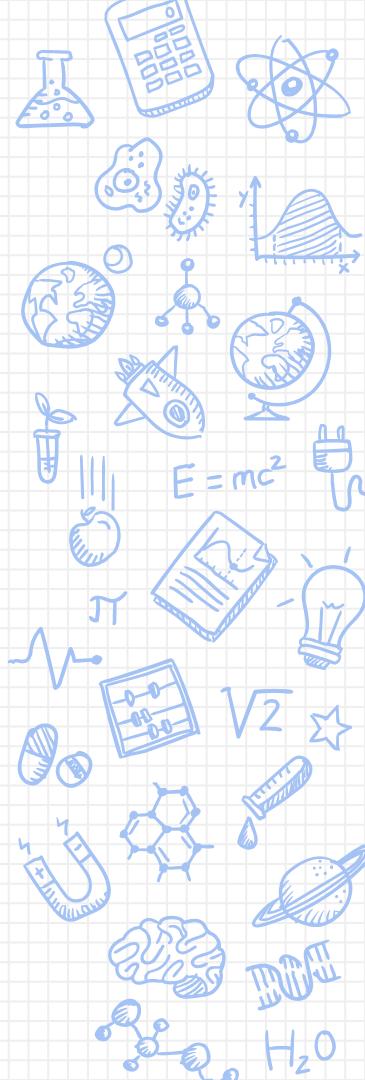
**High Plains Drifter**  
Películas de YouTube Comprar o alquilar

**El Angel**  
Películas de YouTube Comprar o alquilar

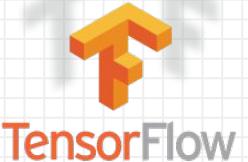
**¡SCOOBY!**  
Películas de YouTube Comprar o alquilar

**The Social Network**  
Películas de YouTube Comprar

**Rampage**  
Películas de YouTube Comprar o alquilar



# Tools

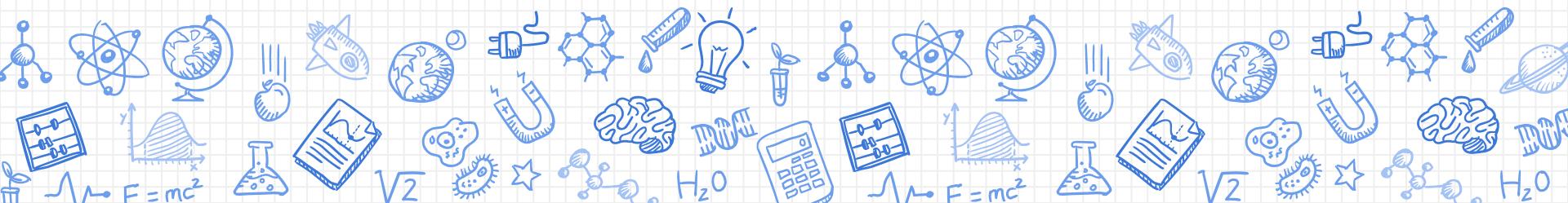


# Natural Language Toolkit (NLTK)



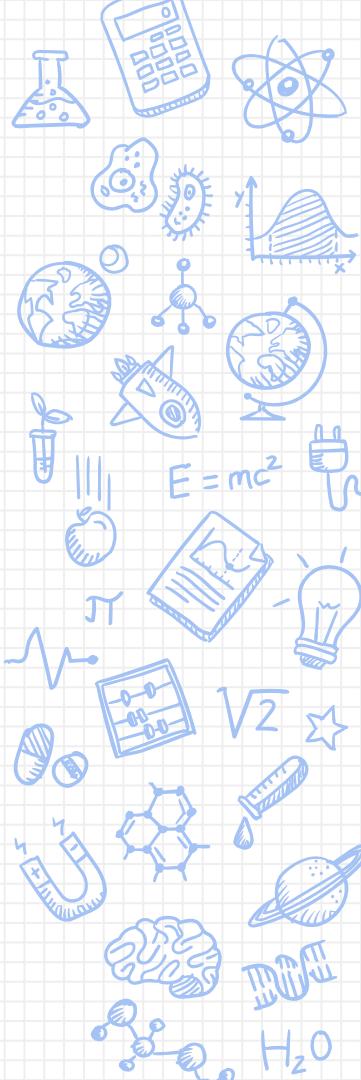
# Ingeniería de Datos

Se encarga de que todo funcione  
eficientemente, estable y escalable.



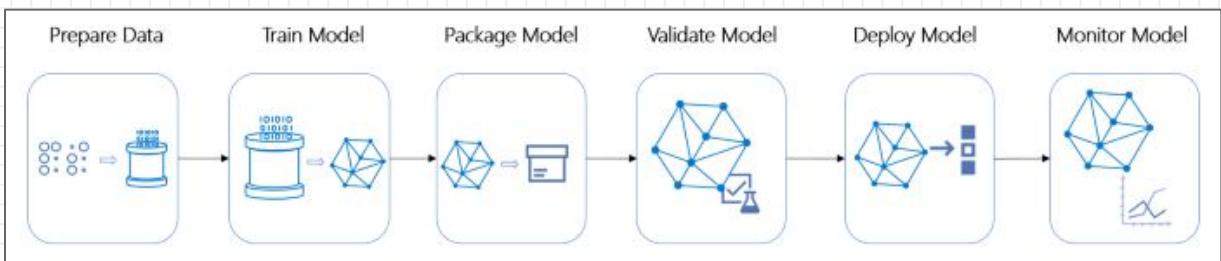
# Gestión de los Datos

- Extracción de datos.
  - Transformación de datos.
  - Almacenamiento de datos.



# Desarrollo de Software

- Data pipelines.
  - Herramientas internas.
  - APIs.
  - Entrenamiento Automático de Modelos.
  - Monitoreo.

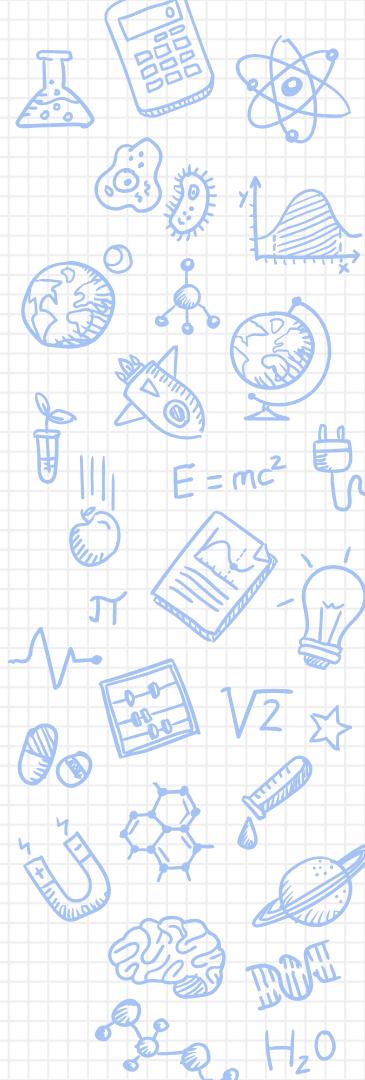


Fuente <https://docs.microsoft.com/en-us/azure/machine-learning/concept-ml-pipelines>

# Ingeniería de Software

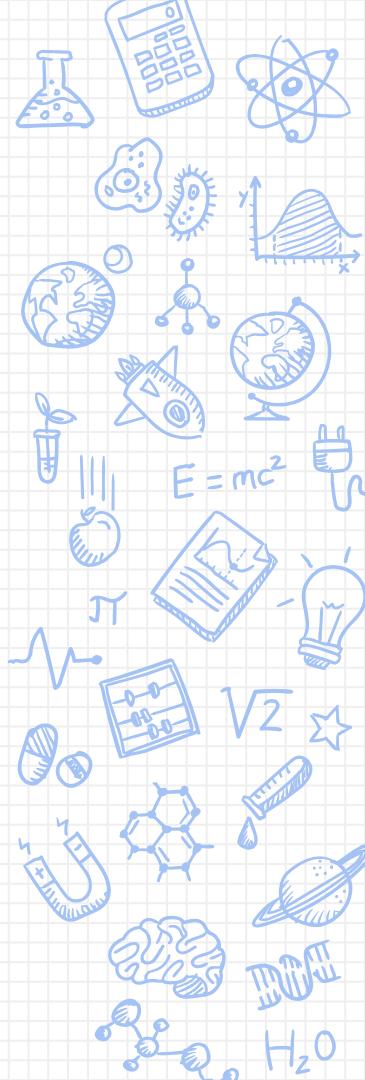
---

- Adopción de Metodologías Ágiles.
- Versionado.
- Mantenibilidad.
- Escalabilidad.
- Modularidad.

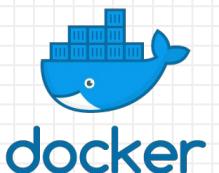


# Ingeniería de Datos – Desafíos

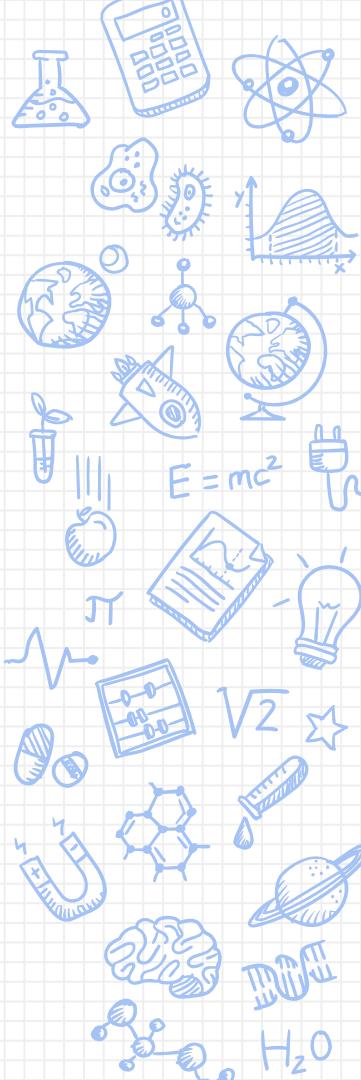
- Procesamiento Real-Time.
  - Multiprocessing y Multithreading.
  - Concurrency y Asincronía.
  - Tolerancia a Fallas.
  - Almacenamiento Distribuido.



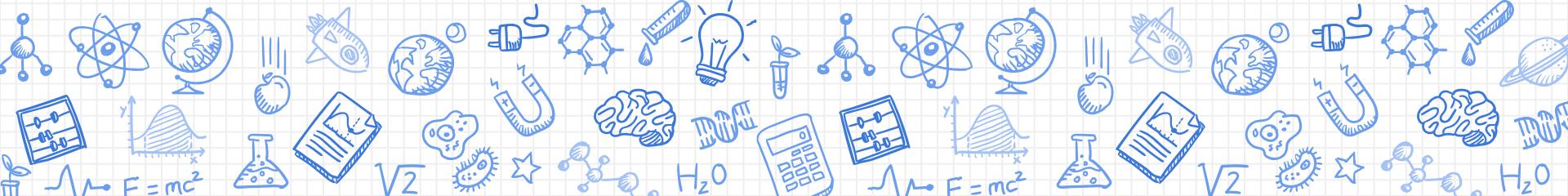
# Tools



# Kubernetes



# Skills & Jobs



7 de abril a las 20:01

Buenas! está buscando recomendaciones.

Es un tema que me interesa mucho!

Hola! Quiere saber si hay administradores de empresas que recomiendan libros o sitios web en los que se habla de negocios y administración.

Puedo ser un analista de ventas? o por ejemplo que no estudio para poder mostrar necesidades. El mismo no tiene conocimientos profundos de una disciplina como la medicina o lo que fuese en los datos que este

Buenas! Económico "sistema" cosa

Hola, me pides recomendaciones. Estoy en busca de un estudiante de sistema de información.

Buen Día a todos! Hoy les comparto un enlace

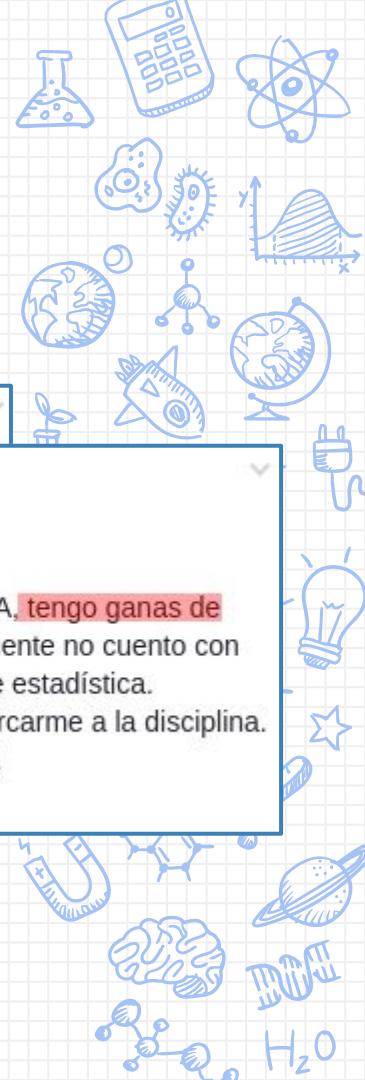
Hola gente! Hoy les comparto un enlace

11 de julio de 2018

Hola como andan?

Me recibí hace unos años en Ciencia Política en la UBA, tengo ganas de incursionar en el mundo de Data science, lamentablemente no cuento con formación en programación y tampoco tengo mucho de estadística. Necesito si alguno puede orientarme desde donde acercarme a la disciplina. Mi idea era hacer un curso en digital house o en EANT. Escucho consejos! Gracias!!

- Estadísticos.
- Informáticos.
- Contadores.
- Economistas.
- etc!



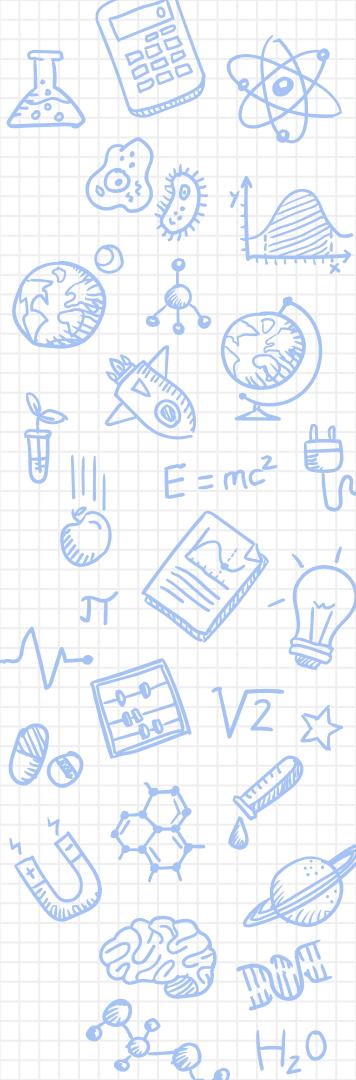
# Jobs - Analista de Datos

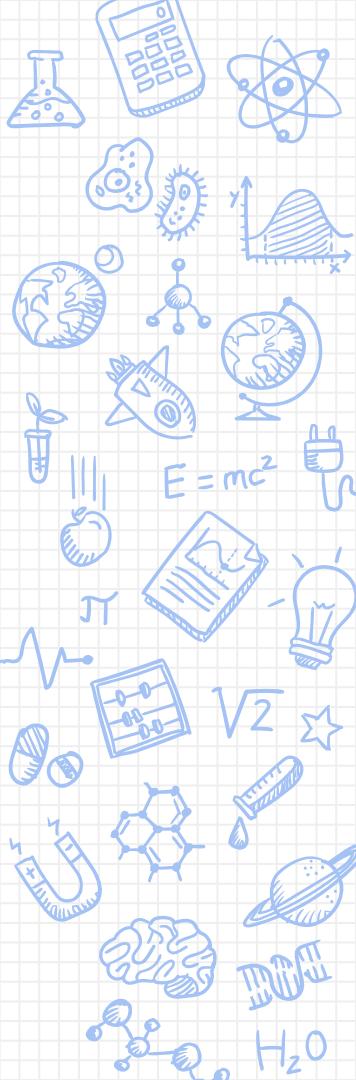
## PRINCIPALES RESPONSABILIDADES:

- Deberá colectar, procesar, y realizar análisis estadísticos sobre big data, para convertirla en descubrimientos y recomendaciones, usando técnicas de minería de datos.
- Deberá tener la habilidad de transmitir y visualizar la historia que cuenta la data.
- Investigará el comportamiento de diferentes segmentaciones de clientes para detectar oportunidades.
- Participará en la construcción y definición de los objetivos de los equipos de trabajo.
- Generará modelos predictivos con el objetivo de optimizar recursos.
- Elaborará estudios analíticos y estadísticos ad-hoc para diferentes áreas de la empresa.

## REQUERIDOS:

- Nivel de Educación: Graduados en Ciencias Económicas, Sistemas o afín.
- Tiene que tener una mente curiosa, intuitiva, analítica y apasionada por descubrir.
- Proactividad y predisposición
- Conocimiento de herramientas de minería de datos como Python, R o SQL.





# Jobs - Analista de Datos

## **RESPONSIBILITIES:**

- Assist in the day-to-day reporting and data analyst functions to support a global performance marketing team. Covering channels including: .com performance, search data, social data, programmatic data, and offline data feeds from Salesforce / Customer feeds
  - Working with various data feeds and assisting with the integration of data across various platforms
  - Helping assess existing marketing reporting framework, and how this can be improved through automation and advanced data-integration techniques
  - Assist the performing marketing teams with ad hoc data & insight questions about markets, channels or specific trends
  - Working with the marketing channel execution leads to help identify key insights in performance, and challenge areas that need to be addressed
  - Role will report into regional performance marketing leads, with dotted line to global data science and BI Director

#### **QUALIFICATIONS:**

- BA/BS, MBA or equivalent work experience in Statistics or other quantitative field
  - 1-2+ years of marketing analysis, or digital marketing experience
  - Extremely strong in Excel, Power BI and data visualization tools, certificates preferred.
  - Familiarity with data query solutions, such as SQL, Python or R.
  - Understanding of Web Analytics tools, particularly Google Analytics
  - Ability to easily transition from high level strategic thinking to detailed execution

# Jobs - Analista de Datos

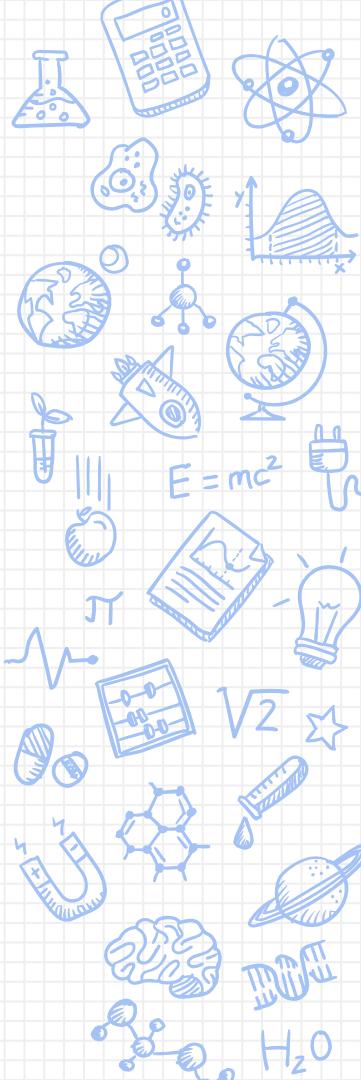
## The Requirements

- Master or PhD in Bioinformatics, Statistics, Engineering or related field
- Have good statistical modelling and machine learning skills
- Hand on experience with end to end development of data analytics solutions
- Good knowledge of big data technologies
- Good with time-series data analytics, anomaly detection, unsupervised learning, hypothesis testing
- Proficient programming in Python, C/C++
- Good research ability and critical thinking skills
- Good communicator

CESVI ARGENTINA S.A. ubicada en Parque Industrial Pilar se encuentra en la búsqueda de un Científico de Datos para sumarse a proyecto que incorpora el concepto de Data Analytics en uno de sus principales productos.

### Requisitos:

- Profesionales o estudiantes avanzados de carreras preferentemente Actuario, Matemático, Estadísticas, Administración de Empresas, u otras carreras afines.
- Conocimiento avanzado en lenguajes como Python y R. para manipular datos y obtener información.
- Manejo de herramientas analíticas.
- Manejo avanzado de bases de datos de grandes volúmenes.



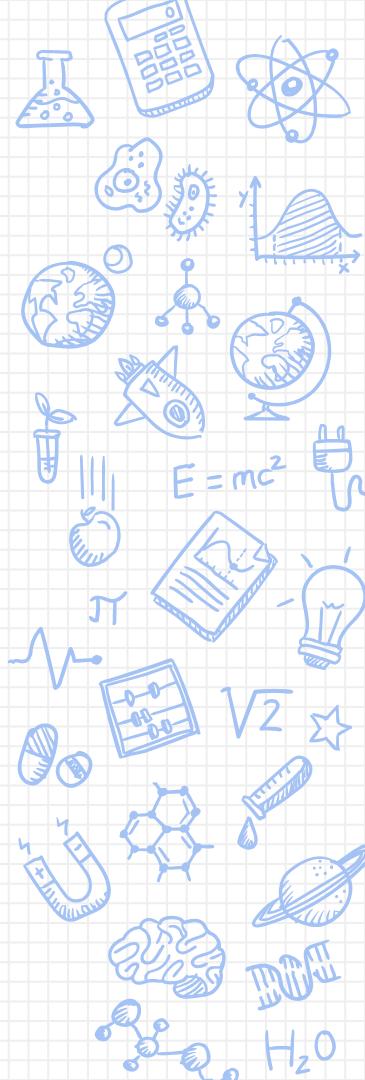
# Jobs - Data Scientist

En Progres buscamos Data Scientists.

Valoramos formación en carreras de Sistemas, Economía, Estadística, Matemática, Sociología, entre otras.

Nos orientamos a profesionales con experiencia en minería de datos, conocimientos de metodologías ágiles y de la Industria Financiera. Vas a trabajar con MicroStrategy, Teradata, R y Python.

Puesto efectivo, importante empresa financiera en la zona de Parque Patricios.



# Jobs - Analista de Datos

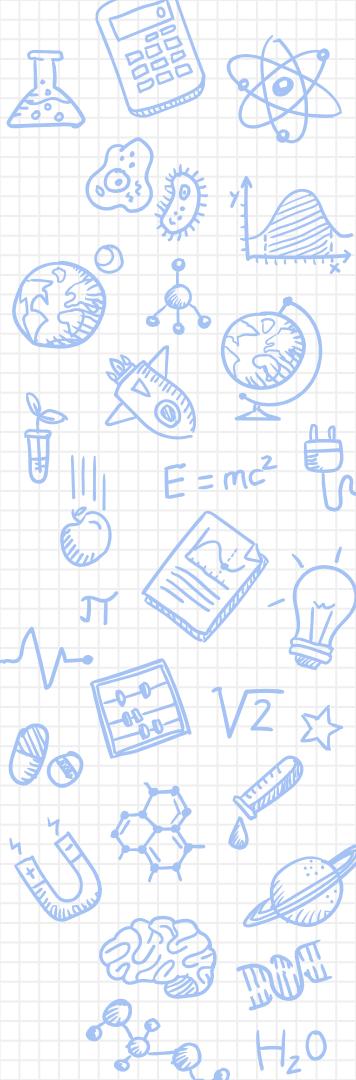
## Requirements And Skills

Licenciatura en Matemáticas, Economía, Ciencias de la Computación, Gestión de la Información o Estadística

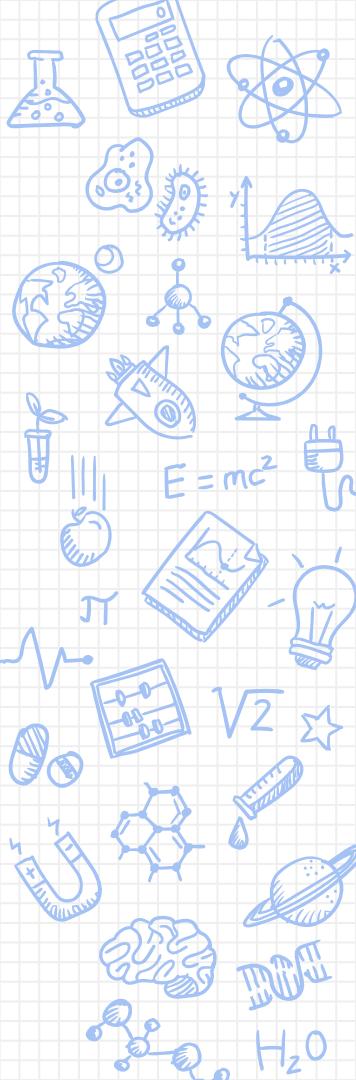
- Experiencia laboral como analista de datos
- Experiencia técnica en modelos de datos, desarrollo de diseño de bases de datos, minería de datos y técnicas de segmentación.
- Conocimiento y experiencia con paquetes de informes, bases de datos (SQL, etc.), programación (XML, Javascript o marcos ETL)
- Conocimiento de estadísticas y experiencia en el uso de paquetes estadísticos para analizar conjuntos de datos (Excel, SPSS, SAS, etc.)
- Fuertes habilidades analíticas con la capacidad de recopilar, organizar, analizar y difundir cantidades significativas de información con atención al detalle y la precisión.
- Experto en consultas, redacción de informes y presentación de hallazgos.

## Additional information

En Cervecería y Maltería Quilmes, nuestro sueño de unir a la gente por un mundo mejor refleja nuestro esfuerzo por cultivar un ambiente inclusivo y diverso en el que la meritocracia y la igualdad de oportunidades prevalezcan por sobre cualquier característica de género, raza, religión o cualquier elección de índole personal.



# Jobs – Machine Learning



## Responsibilities

- Take responsibility for ensuring that Machine Learning code, models and pipelines are deployed successfully into production, and troubleshooting issues that arise.
  - Design solution architectures for applications that will use the machine learning models.
  - Deploy applications to AWS's cloud leveraging the full spectrum of the AWS cloud services.
  - Automate model training** and testing and deployment via machine learning continuous delivery pipelines.
  - Design and implement metrics to verify model and algorithm effectiveness.

## **BASIC QUALIFICATIONS**

- Master's Degree in Computer Science or Bachelor's Degree with 5+ years related work experience
  - 2-4 years of experience building production quality and large scale deployment of applications related to natural language processing or related machine learning applications.
  - Strong programming skills in at least one object oriented programming language (Java, Scala, C++, Python, etc.)

## PREFERRED QUALIFICATIONS

Expert familiarity with a variety of classic and modern machine learning techniques including deep learning, clustering, decision tree, classification, regression and neural networks.

Solid background in algorithms, data structures, and object-oriented programming

Knowledge of mining complex data (including structure and unstructured), identifying patterns, and feature engineering.

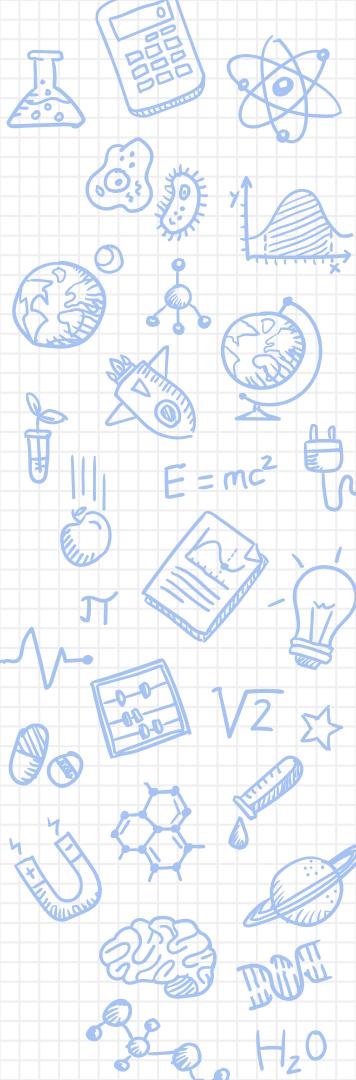
# Jobs – Data Engineering

## Responsibilities

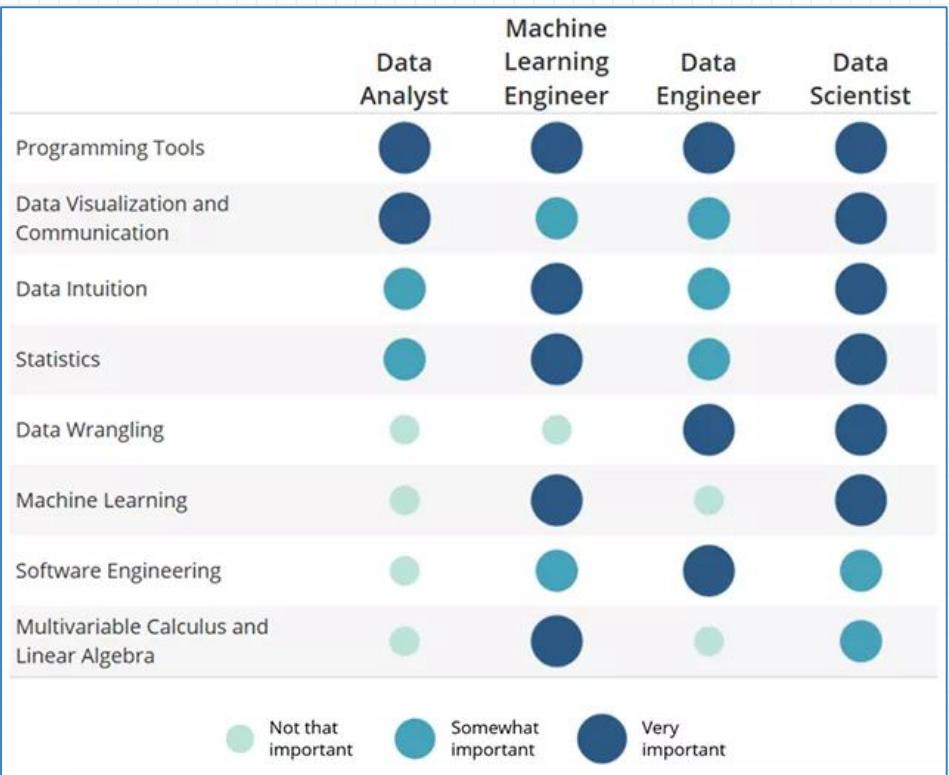
- Understand and implement mathematical approaches to solve complex business problems, utilizing available research at your disposal.
  - Build machine learning applications that execute in an autonomous production environment and fully considers scalability for big data, error handling, exception conditions, etc.
  - Perform unit testing and model validation of statistical routines.
  - Collaborate extensively with stakeholders, program management, and software development team members to ensure that solutions meet business needs, permit valid inferences, and have functional implementability.
  - Document the technical details of your work to improve collaboration with other team members.
  - Guide and mentor other modelers in best practices regarding methodology (prototype formulation and testing, data mining and analysis), as well as implementation (proper documentation, source control, and coding practices).

## Qualifications

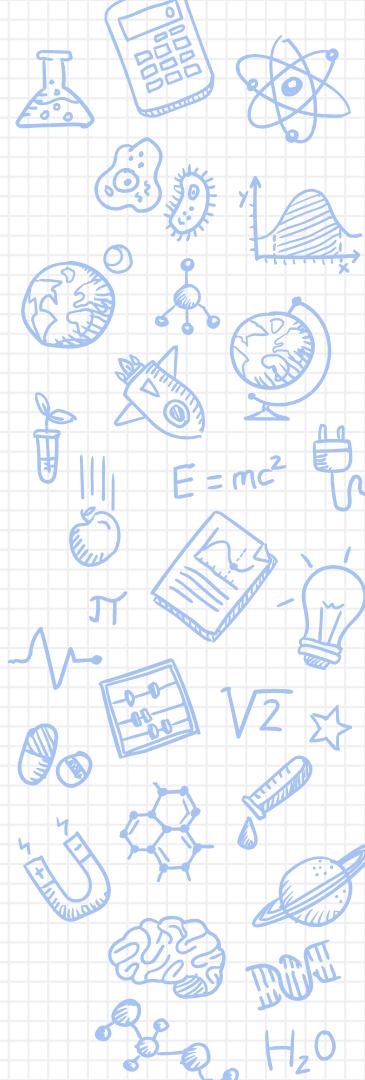
- Degree in computer science, software engineering or applied mathematics
  - Comprehensive understanding of quantitative methods and extreme attention to detail.
  - Proficiency and experience in some data science technologies like Python, R, Scala, Spark.
  - Experience applying machine learning / data science to solve a real life problem and validating it in a productive environment.



# Skills



Fuente: <https://blog.udacity.com/2014/11/data-science-job-skills.html>



# Gracias!

¿Preguntas?