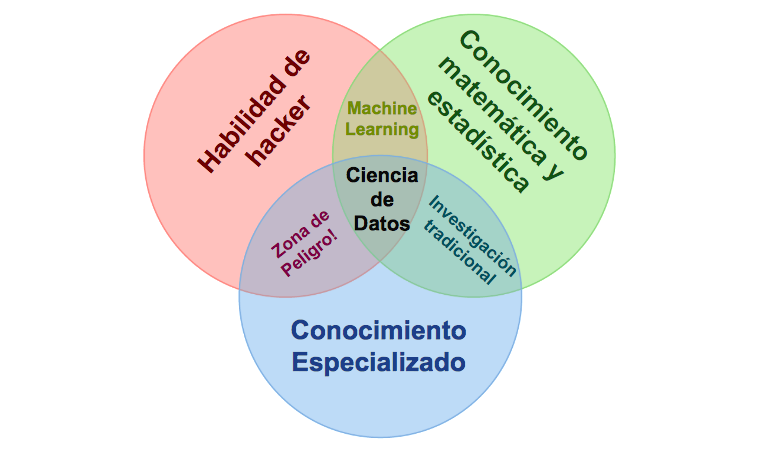
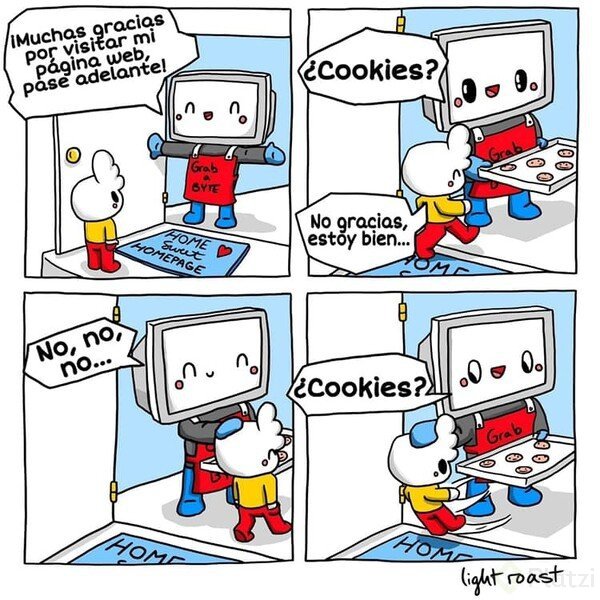
**¿Qué es y qué hace el Científico de datos o Data Scientist?**  
Científico de datos: Persona que sabe más de estadística que cualquier programador y que a la vez sabe más de programación que cualquier estadístico.  
Profesional dedicado a analizar e interpretar grandes bases de datos.

■Recopilar grandes cantidades de datos revueltos y transformarlos a un formato más utilizable.  
■Resolver problemas relacionados con negocios empleando técnicas basadas en datos.  
■Trabajar con diversos lenguajes de programación, incluidos SAS, R y Python.  
■Dominar técnicas analíticas como el machine learning, deep learning y analítica de texto.  
■Buscar orden y patrones en datos, además de detectar tendencias que puedan ayudar a la base de operación de una empresa.  

**¿Qué tipo de información podemos analizar?**

Las cookies incluye los terminos y condiciones. Estan advirtiendo de todo el registro y informacion que van a recoger.



Tipología de datos:

* **Personas**: gustos, likes o preferencias, lugares que has visitado, personas con las que estás más tiempo.
* **Transacciones**:

1. Monetarias: pagos, cuanto pagas, en qué establecimiento, que producto o servicio obtienes.
2. No monetarias: por ejemplos tus llamadas telefónicas, cuánto duran, a quién le marcas, etc.

* **Navegación web**: Quién accede a las páginas web, de dónde son, clicks que das, cuánto tiempo duras en la página.
* **Machine 2 Machine**: conexión de una máquina a otra máquina. Por ejemplo el GPS de tu teléfono para ubicar al conductor de Didi más cercano.
* **Biométricos**: datos que te identifican como persona única, huella dactilar, tipo de sangre, tu voz, tu rostro para desbloquear tu smartphone, iris.

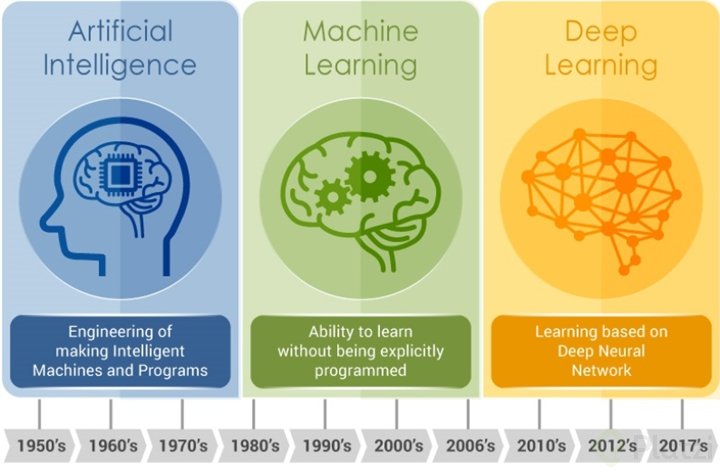
1. **UBER**  
   Datos personales: *para el análisis los clientes, edad, país.*  
   Machine to Machine: *ubicación vía GPS de el cliente y vehículo.*  
   Transacciones no monetarias, *cantidad de viajes que un cliente realiza y su destino.*
2. **FACEBOOK**  
   Datos personales: *registrados para ver la cantidad de usuarios y sus características.*  
   Datos de navegación en la red social: *cantidad de reacciones y comentarios en las publicaciones.*
3. **AMAZON PRIME VIDEO**  
   Datos personales: evaluar quienes son sus clientes y sus preferencias.  
   Transacciones no monetarias: películas y series favoritas en los usuarios.
4. UNICEF: Información de mensajes en facebook y uso de datos móviles, para identificar los focos y la población afectada por el virus, y de esta manera enfocar los esfuerzos hacia los lugares y personas afectados.  
   <https://www.unicef.org/innovation/>
5. GLOBAL FOREST WATCH: thousands of satellite images, to count trees in order to monitor world’s forests, allowing to track illegal deforestation and more.  
   <https://blog.globalforestwatch.org/data-and-research/>
6. STARBUCKS: ingreso de la zona (barrio media de 60.000 de ingreso), edad de la población, kind of neighborgs (offices, universities, etc), traffic (vehicles per day, location (corners, access points), commute side, proximity to other business, parking, crime rates, smartphone ownership.  
   <https://medium.com/locale-ai/site-planning-using-location-data-ae7814973521>

# ¿Cómo crear empresas y culturas data-driven?

**Empresas data-driven**  
**1.Crear una cultura de datos**-> entenderlo, hacer que los empleados tomen decisiones basadas en datos.  
**2.Recolectar información**-> entender toda la información que tenemos.  
**3.Medir todo**->Entendemos lo que tenemos y cómo se comporta.  
**4.Datos relevantes y precisos**->Basarnos en datos que nos sirvan, sea el tiempo, cliente, y sus tipos.  
**5.Testear y crear hipótesis**-> Partir de una hipótesis, plantear una pregunta específica para estudiar lo que vas a testear.  
**6.Desde los insights de datos a las acciones**-> teniendo los insights(visiones), tomamos acciones.  
**7.Cumplir las regulaciones de datos**-> Cumplir los reglamentos  
**8.Automatizar**->Automatizar no más…

# ¿Qué es inteligencia artificial y machine learning?

**Inteligencia Artificial** es la capacidad de las computadoras de mostrar un comportamiento “inteligente”. Mientras que **Machine Learning** es una de las técnica que se utiliza para crear y mejorar dicho comportamiento. Esto mediante entrenamientos automáticos basados en la exposición a datos.  
Al utilizar redes neuronales de gran profundidad, el aprendizaje automático se convierte en aprendizaje profundo.  
**Inteligencia artificial**: capacidad de jugar ajedrez.  
**Machine Learning**: técnica de entrenamiento para enseñar a jugar ajedrez.



**MACHINE LEARNING**  
■Detección de fraudes  
■Búsqueda web  
■Anuncios a tiempo real  
■Análisis de textos  
■Next best action

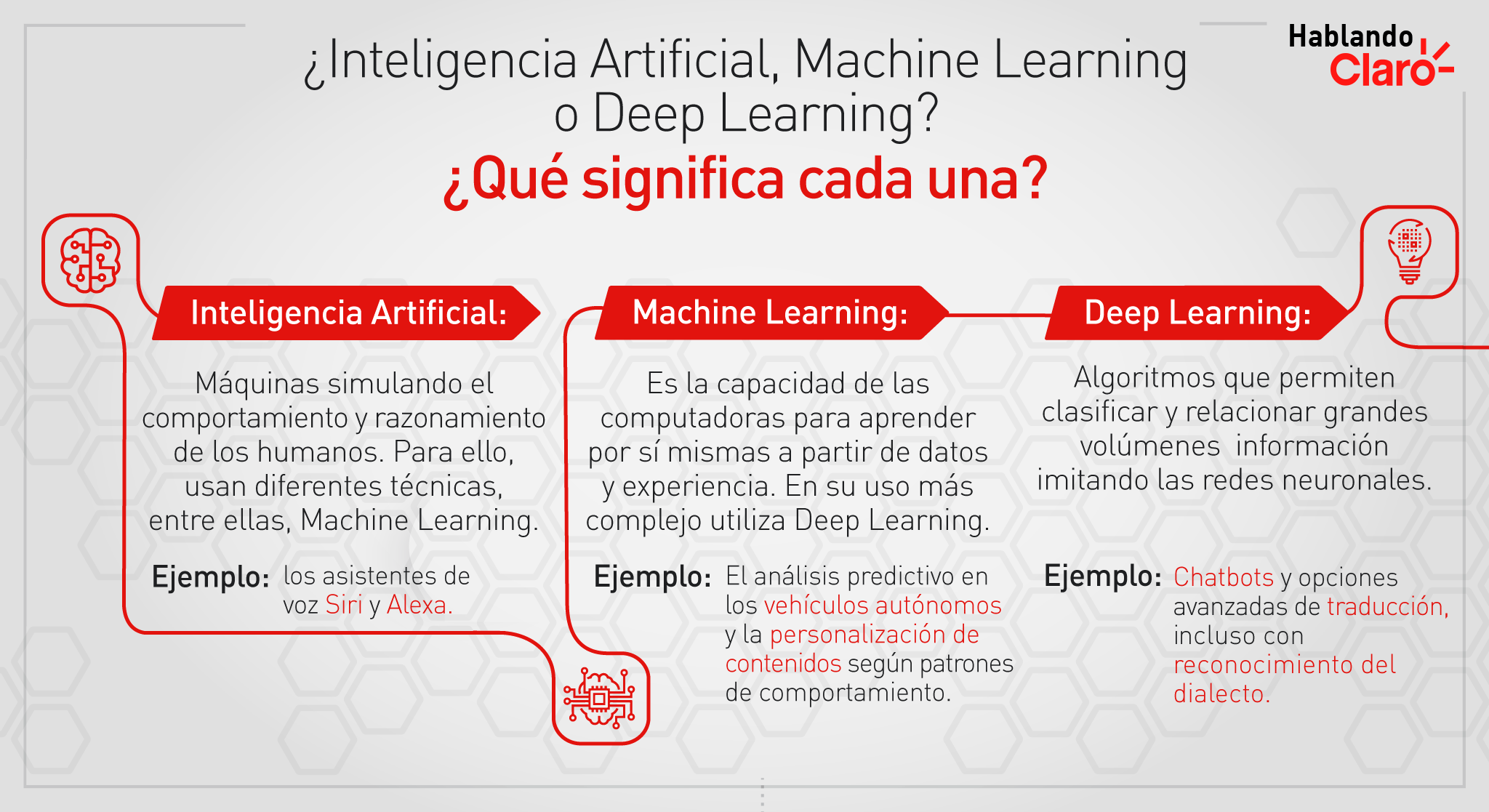
Inteligencia artificial: Máquina simulando el comportamiento y razonamiento de los humanos. Para ello usan diferentes técnicas como Machine learning

Machine learning: Capacidad de las computadoras para aprender por sí mismas a partir de datos y experiencia. en usos complejos usa el Deep learning

Utilidad:

* Detección de fraudes
* Búsqueda web
* Anuncios a tiempo real
* Análisis de textos
* Next best action: Predecir la acción que un usuario va a tomar

Deep learning: Algoritmos que permiten clasificar y relacionar grandes volúmenes de información imitando redes neuronales.



**Blue River Technology-** Agricultura- Desarrolla soluciones basadas en IA y ML para optimizar la forma en la que se esta produciendo a gran escala, por ejemplo: tienen un sistema que identifica las plagas y las malas hierbas y puede rociar automaticamente herbicidas y pesticidas, disminuyendo el desperdicio y ahorrando costos al no tener que rociar los cultivos completos, tambien tienen un sofware para que por medio de drones se recaben datos para gestionar mejor los cultivos y su producción.  
Se prevee que para 2050 el consumo de alimentos aumente un 50% por lo que sera imprescindible hacer uso de la tecnología para poder optimizar la forma en la que producimos nuestros alimentos.

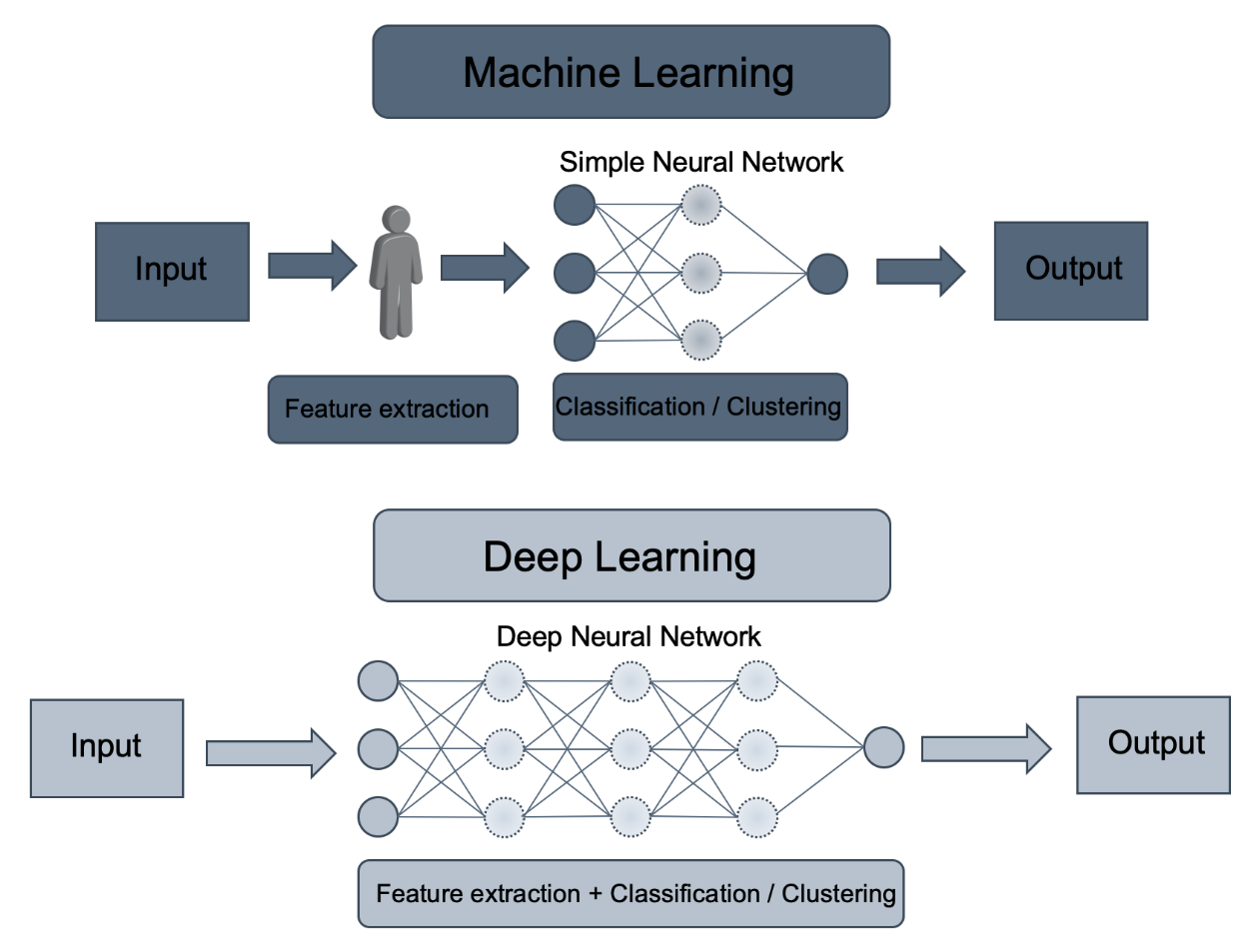
**ITMI**- Biomedicina- Recopila terabytes de información clínica y genómica de distintos pacientes alrededor de decenas de paises para identificar los vínculos genéticos con distintos padecimientos. Para que se den una idea el ADN humano contiene mas de seis mil millones de bits de información. ITMI esta trabajando en reunir la base de datos de secuencias de genoma mas grande del mundo y ya ha ayudado a diagnosticar enfermedades congenitas que no habian sido previamente detectadas en bebes. Me parece muy emocionante imaginar lo mucho que podremos avanzar si empezamos a aprender cada vez mas del genoma humano y al tener una base de datos a nivel mundial, las aplicaciones serian enormes.

**\*\*Centre d’Estudis Montseny \*\*-** Educación- Es un escuela situada en Barcelona, esta implementando por medio de Clickedu y Machine Learning educación completamente personalizada para los alumnos, en la que se enseñen los conocimientos necesarios pero adaptadas a la forma en la que aprende y en los gustos, intereses y habilidades del alumno. Otra idea que me parece increible para llevar a todas las aulas, me parece que existe una brecha muy grande en la educación gracias a la estandarización de la educación, ya que el proceso de aprendizaje es distinto para todos y cuando una persona se adelanta termina por aburrirse y cuando una se atrasa termina frustrandose, y en ambos casos puede terminar en el abandono.  
Gracias por leer XD creo que me extendi pero es que me encanto todo lo que encontre sobre el tema y volo mi imaginación

# ¿Qué es deep learning? Análisis de imagen, audio y video

Deep learning: es un conjunto de algoritmos de aprendizaje automático (en inglés, machine learning) que intenta modelar abstracciones de alto nivel en datos usando arquitecturas computacionales que admiten transformaciones no lineales múltiples e iterativas de datos expresados en forma matricial o sensorial.

* Amazon Prime Video: es capaz de identificar los actores y actrices en cada escena de una serie que estés reproduciendo.
* Google (búsqueda por imágenes): sin indicar un texto de búsqueda sino subiendo una imagen Google es capaz’de hacer una búsqueda con gran probabilidad de acierto.
* CamScanner: Es una aplicación móvil capaz de reconocer un documento desde la cámara del teléfono y le aplica una transformación precisa para obtener una imagen procesada que es perfectamente comparable con una copia fotoestática real.

Spotify, Netflix, instagram  
Les dejo una imagen que explica muy visualmente la diferencia entre Machine Learning y Deep Learning:  


# Flujo de trabajo en ciencia de datos: fases, roles y oportunidades laborales

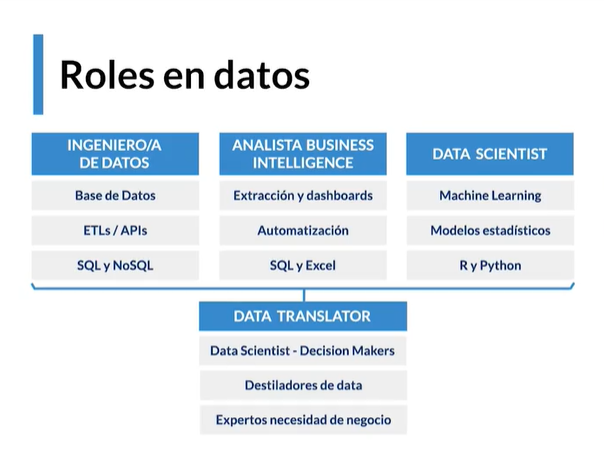
**Flujo de trabajo en ciencia de datos: fases, roles y oportunidades laborales.**

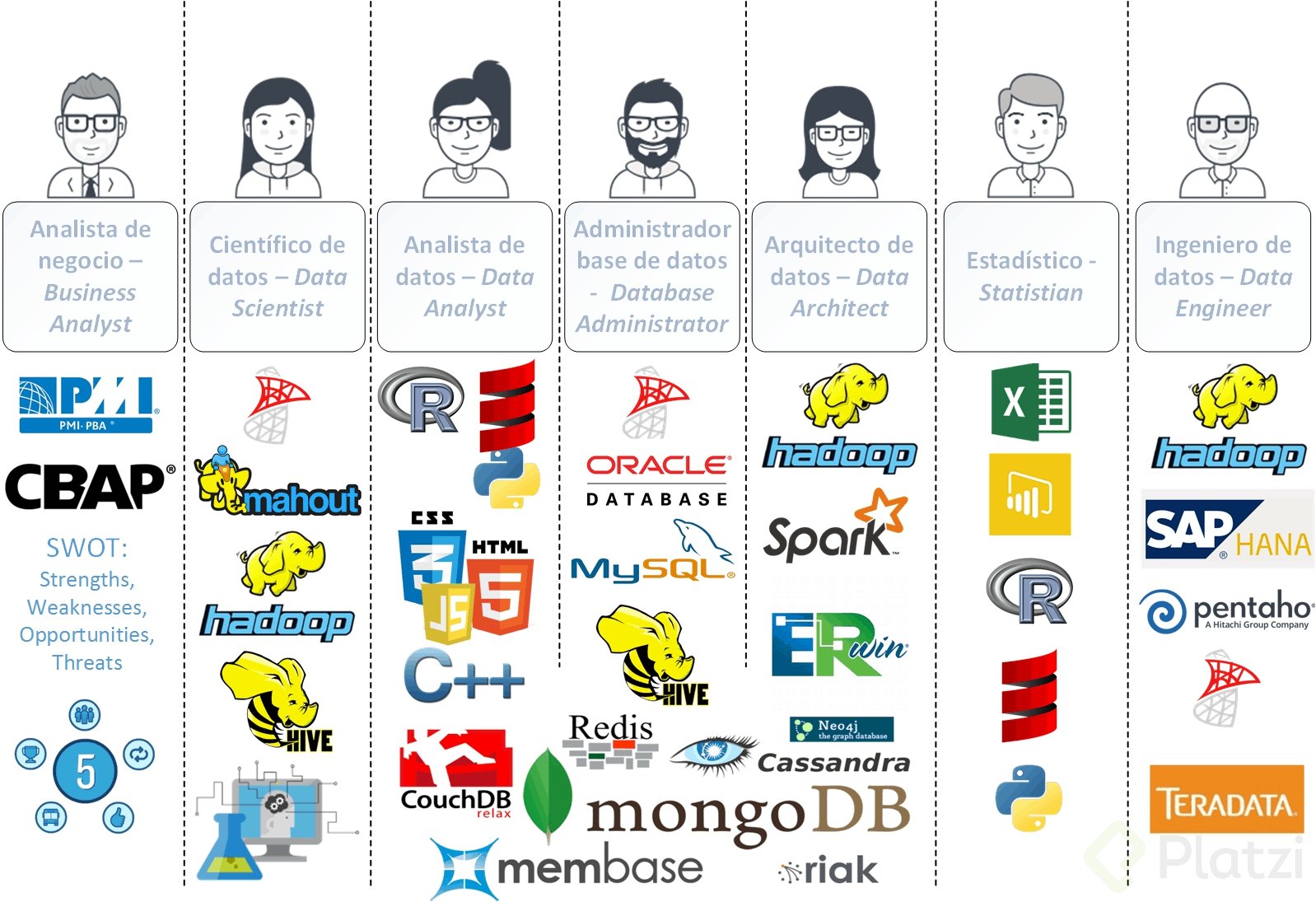
Dentro de un equipo encargado de analizar los datos podemos encontrar diferentes roles:

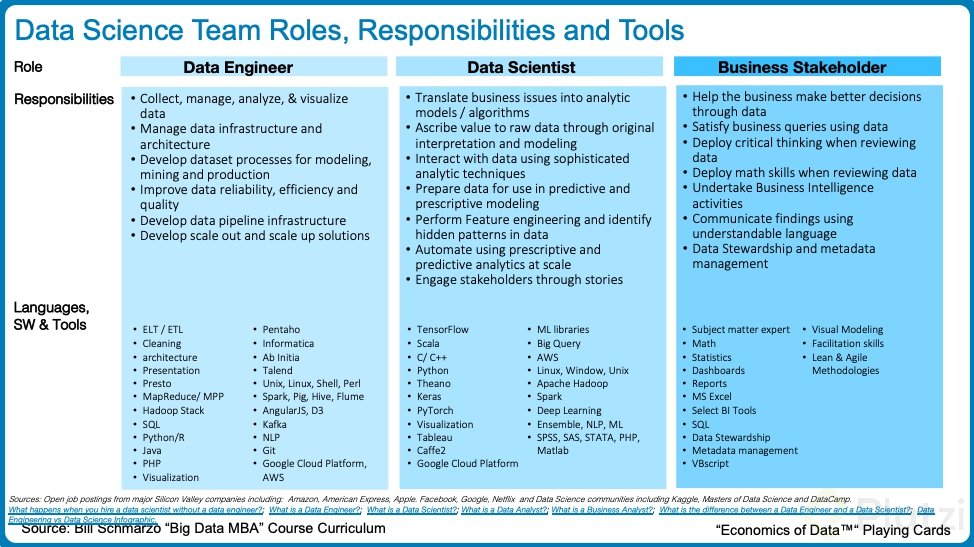
* **Ing de dato** : Es una persona que se encarga de conectar los dispositivos, estableces la estructura encargado de **captar los datos y administrarlos**. Tambien cabe destacar que es el principal responsable de que los datos de un negocio existan. Las herramientas o skills que necesita son los siguientes:
  + Hacer APIs y eTLs
  + Conectores
  + Herramientas de SQL y NoSQL
* **Analista o Business Intelligence**: Este rol le pertenece a quien, a partir de los datos generados u obtendios por el Ing de datos, extrae la **información** importante para que nos pueda dar el **status actual** de la empresa. Para realizar su trabaho es necesario que:
  + Cree Dashboards o cuadros de control
  + Automatice la generación de los mismo.
  + Utilice SQL y Excel.
* **Data Scientist**: Este rol le pertenece a la persona que, al igual que el *Analista* extrae la información, pero además su enfoque esta en **predecir** el comportamiento de los datos valiendose de su background en matematicas, estadistica y probabilidad. Sus principales herramientas son:
  + Python y R

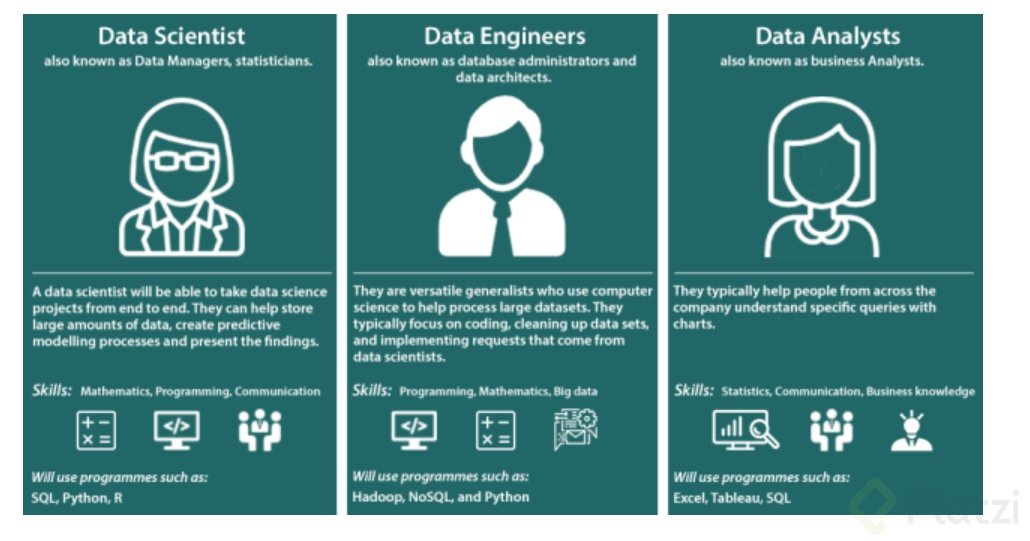
Finalmente estos roles dentro de un equipo de analisis de datos, no puede desarrollarse sólo debido a sus habilidades tecnicas y especificas. Para poder alinear los objetivos de este equipo con los de la organización es necesario que exista un Traductor:

* **Data Translator**: Esta persona se encarga de liderar al equipo de trabajo, para que su trabajo se haga de manera efectiva, y oportuna. Además debe ser capaz de interpretar los datos pero sobre todo:
  + Tener una comunicación efectiva con otros equipos de la organización
  + Ser Experto en las necesidades del negocio.





Cada uno de los roles conlleva manejar herramientas y actividades que me llaman mucho la atención, sin embargo, por mi gusto por los datos podría estar en la parte del ingeniero de datos obteniendo información y guardándola. Otra parte que me gusta es la manera como se presenta el análisis de datos. Les comparto otra imagen con más herramientas: 



# Herramientas para cada etapa del análisis de datos

**¿Qué clientes compraron ayer y hoy? Sin duda que SQL.** Estamos con la mirada en el pasado y el presente, no nos están pidiendo que predigamos el futuro. Con SQL podremos saber quienes compraron ayer y hoy.

**Herramientas en Data Science**

**Extracción de la información con SQL**  
Aquí extraemos la información, sintetizar una base de datos y crear un cuadro de control de la operación.  


**Análisis y visualización con R y Python**  
Se analiza y visualizar la información extraída a través de gráficos, y con los cuales podremos crear modelos predictivos. Los 2 lenguajes de programación que usan aquí son:

* **R:** Sirve para realizar análisis de datos con un enfoque estadístico y es un más complejo aprenderlo al inicio. Algunas de las herramientas (packpages) que se complementan con R son ggplot2 y dplyr .
* **Python:** Sirve también para realizar análisis de datos, pero con un enfoque hacia la ingeniería. Python es un lenguaje más sencillo y es similar a otros lenguajes de programación. Alguna herramientas (librerías) que complementan con Python son Pandas y Numpy.  
  

# ¿Qué es y cómo usar una base de datos relacional con SQL?

Si alguien quiere profundizar en SQL, estos cursos tienen un excelente contenido:

[<https://platzi.com/cursos/bd/>] -> Fundamentos Bases de Datos  
[<https://platzi.com/cursos/postgresql/>]-> Curso PostgreSQL  
[<https://platzi.com/cursos/practico-sql/(url)> -> Curso Practico SQL

**¿Qué tipos de comandos hay dentro del SQL?**  
Los comandos del lenguaje SQL se dividen según su función en estos 5 tipos:

**DDL (Data Definition Language)**: definen el esquema o estructura de la base de datos. Ejemplos: CREATE (crear); ALTER (alterar); DROP (eliminar objetos); RENAME (renombrar); TRUNCATE (quita todos los registros de una tabla, incluidos los espacios de los registros eliminados); COMMENT (comentar); entre otros.  
**DQL (Data Query Language):** sirven para hacer consultas sobre los datos en el esquema de objetos. Un objeto puede ser desde un resultado de búsqueda a una tabla. El propósito del comando es el de establecer una relación, basada en la consulta, dentro de la estructura de la base de datos, como la función de búsqueda. Ejemplo: SELECT (recuperar registros de la base de datos).  
**DML (Data Manipulation Language):** tratan la manipulación de los datos presentes en la base de datos. La mayoría de los comandos pertenecen a este tipo. Ejemplo: INSERT (insertar un objeto); DELETE (eliminar registros); UPDATE (actualizar); CALL; MERGE (3 en 1, inserta, elimina y actualiza); LOCK TABLE (bloquear tabla); EXPLAIN PLAN (determina el plan de acceso); …  
**DCL (Data Control Language)**: se encargan de los derechos, los permisos y otros controles del sistema de la base de datos. Ejemplos: GRANT (proporcionar privilegios acceso a un usuario); REVOKE (revocar el derecho de accceso dado a un usuario); etc.  
**TCL (Transaction Control Language)**: sirve para las transacciones con la base de datos. Es decir, con estos comandos se puede llevar un control sobre otros comandos y cómo afectan a la base de datos. Ejemplos: COMMIT (llevar a cabo una transacción); ROLLBACK (revertir una transacción en caso de que ocurra algún error); SAVEPOINT (establecer un punto de rescate dentro de una transacción.).

**10 preguntas frecuentes de SQL en entrevistas de trabajo**  
Según: [java67](https://bit.ly/3kOupic)

1: SQL para encontrar el segundo salario más alto de los empleados:

**select** **MAX**(Salary) **from** Employee **WHERE** Salary **NOT** **IN** (**select** **MAX**(Salary) **from** Employee ); ```

2: SQL para encontrar el salario máximo de cada departamento.

SELECT DeptID, MAX(Salary) FROM Employee GROUP BY DeptID.

3: Escriba una consulta SQL para mostrar la fecha actual.

SELECT GetDate(); ```

4: Escriba una consulta SQL para comprobar si **la** fecha pasada a **la** consulta es **la** fecha del formato dado o **no**.

SELECT ISDATE(‘1/08/13’) AS “MM/DD/YY”;```

5: Escriba una consulta SQL para imprimir el nombre del empleado distinto cuya fecha de nacimiento es entre el 01/01/1960 al 31/12/1975.

**SELECT** **DISTINCT** EmpName **FROM** Employees **WHERE** DOB **BETWEEN** ‘01/01/1960’ **AND** ‘31/12/1975’;```

6: Escriba una consulta SQL para encontrar el número de empleados según el género cuya fecha de nacimiento sea entre el 01/01/1960 y el 31/12/1975.

SELECT COUNT(\*), sex from Employees WHERE DOB BETWEEN ‘01/01/1960’ AND ‘31/12/1975’ GROUP BY sex```

7: Escriba una consulta SQL para encontrar un empleado cuyo salario sea igual o superior a 10000.

SELECT EmpName FROM Employees WHERE Salary>=10000;```

8: Escriba una consulta SQL para encontrar el nombre del empleado cuyo nombre comience con "M"

SELECT \* FROM Employees WHERE EmpName like ‘M%’;

9: busque todos los registros de empleados que contengan la palabra “Joe”, independientemente de si se almacenó como JOE, Joe o joe.

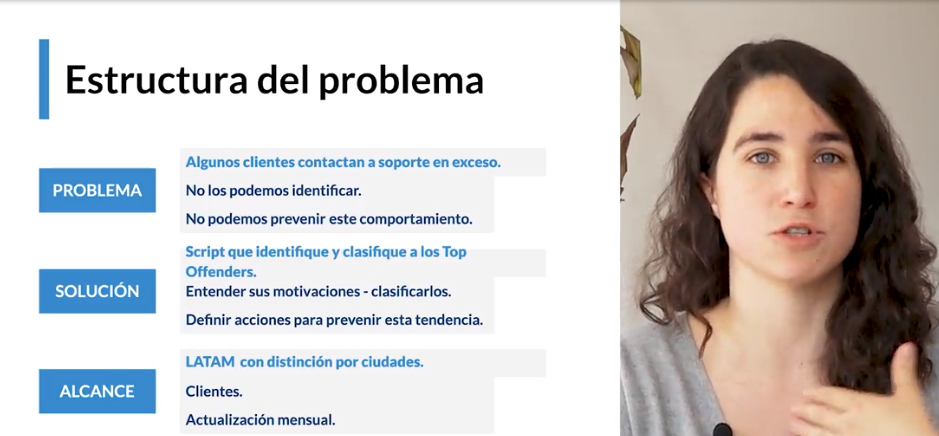
**SELECT** \* **from** Employees **WHERE** **UPPER**(EmpName) **like** '%JOE%';```

10: Escriba una consulta SQL para encontrar el año desde la fecha.

SELECT YEAR(GETDATE()) as “Year”;

# Conflictos y retos actuales sobre la ética y tratamiento de datos

# Aplica técnicas de storytelling para convertir problemas de datos en historias



<https://platzi.com/cursos/storytelling/>

<https://platzi.com/cursos/curso-de-thick-data/>

<https://platzi.com/cursos/canvas/>

# Cómo estructurar un caso de negocio



¿Cómo estructurar un caso de negocio?

Desglosar un problema de negocio en una hipótesis estructurada:  
**¿Qué?** > Problema de negocio -> Hipótesis: Partir de una pregunta específica.  
**¿Por qué?** > Clasificar todos los motivos y sintetizarlos en pocas categorías, ejemplo:

* Motivaciones económicas
* Preguntas
* Problemas tecnológicos (relacionado con problema de como comunica la empresa).
* Política de la empresa.

**¿Cómo?** > Estrategia de cómo vamos a diseñar y organizar todo el análisis:

1.- **Análisis cuantitativo**. Debe ir siempre primero ya que es toda la información. Hacer una clasificación numérica.  
2.- **Análisis cualitativo**. Buscar categorías en función de texto. Nos ayuda a identificar características, tipos, relacionados con el problema.  
3.- **Matriz cuantitativa-cualitativa**. Poner en conjunto la información, entender los números y explicarlos más a profundidad.  
4.- **Acciones de prevención**. Definir acciones con la información que tenemos. **Pasar de los insights a acciones**.  
5.- **Validación**. Verificar si nuestro análisis ha servido y si nuestras acciones de prevención están teniendo efecto.

Paso 1- Identificar  
Paso 2-encontrar las categorías  
Paso 3- poner en conjunto esta información  
Paso 4- crear acciones con esta información  
Paso 5- validar que todo nuestro estudio esté sirviendo

Problema: La empresa tiene montos de dinero sin aplicar de facturas pagadas por los clientes en toda latinoamérica

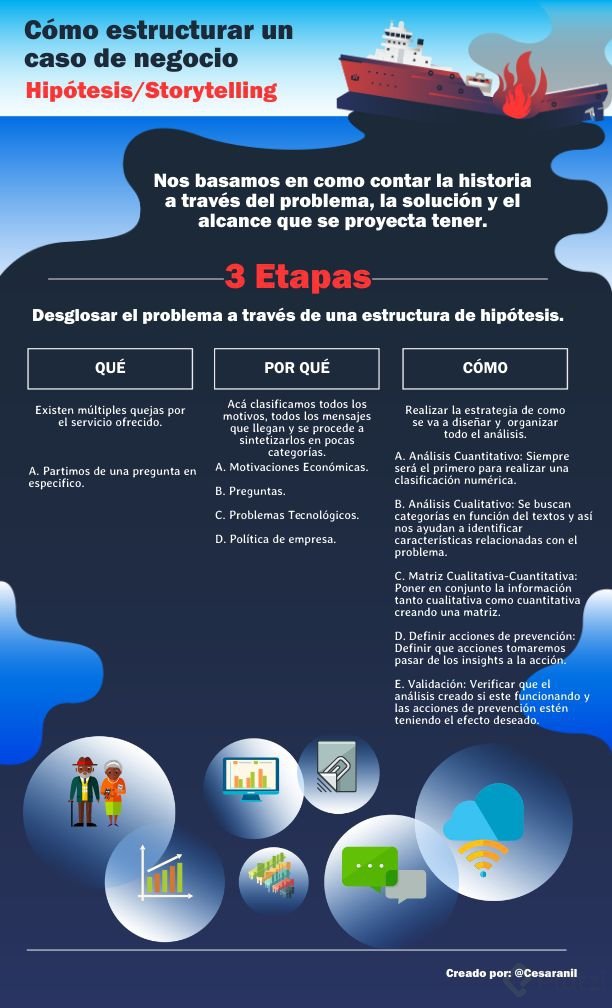
**Análisis Cuantitativo**: Ubicar la información de extractos de pagos en las diferentes locaciones y unificarlas / Revisar cuanto se aplica a las facturas el día en que llega el extracto

**Análisis Cualitativo** : Determinar categorías del por qué no se aplica (falta de tiempo, falta de información, la factura no existe, el cliente no registra datos completos)

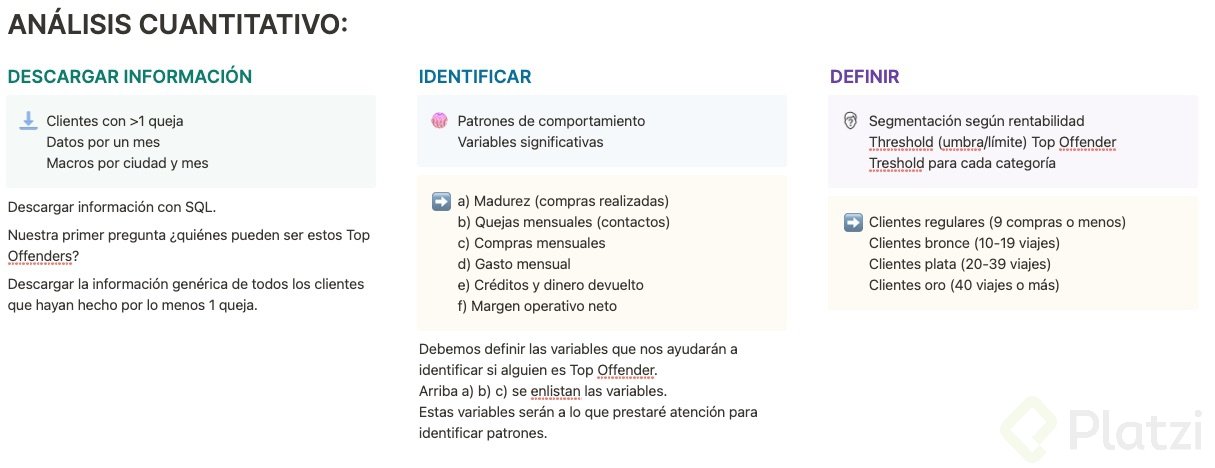
**Matriz Cuantitativa / Cualitativa**: Cruzar la base de información de pagos sin aplicar, con las razones más frecuentes de por qué no se aplica y ordenar en que razones tienen más peso para este error.

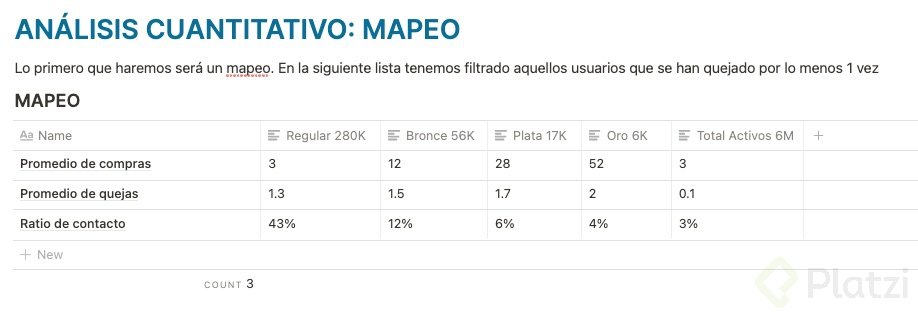
**Definir acciones** : Presentar una estrategia de trabajo con los clientes para atacar las razones más fuertes que recaen en este error.

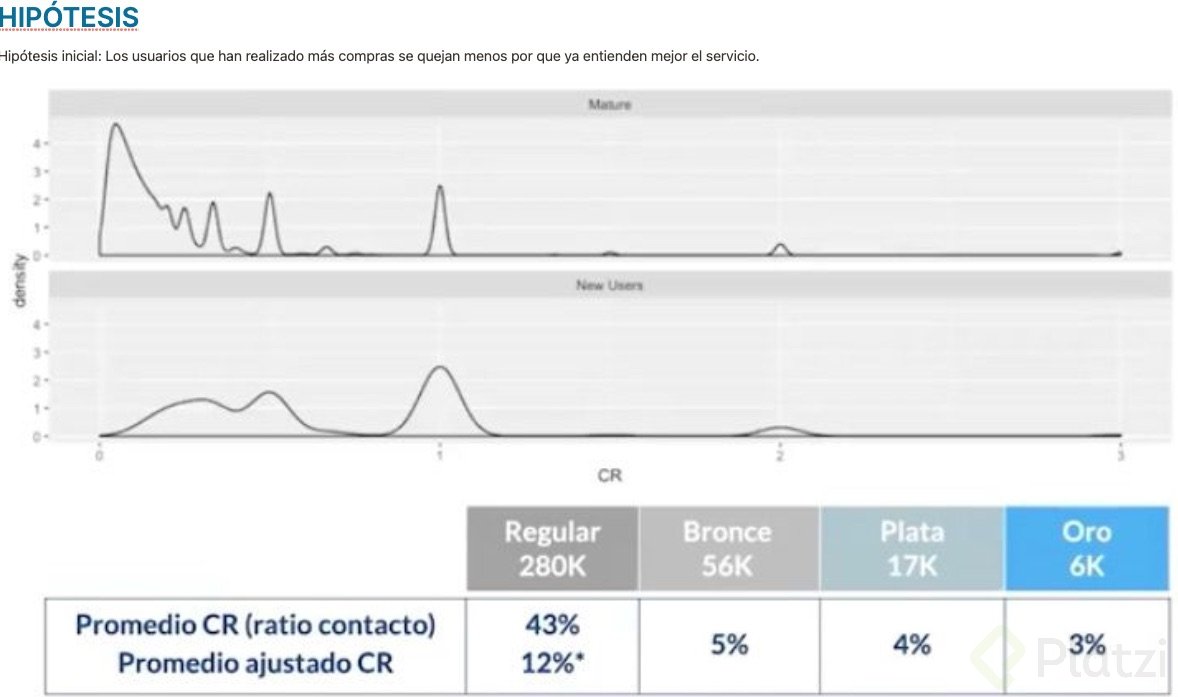
**Validación** Luego de un periodo de tiempo, evaluar nuevamente los resultados cuantitativos y cualitativos para revisar si las estrategias funcionaron.



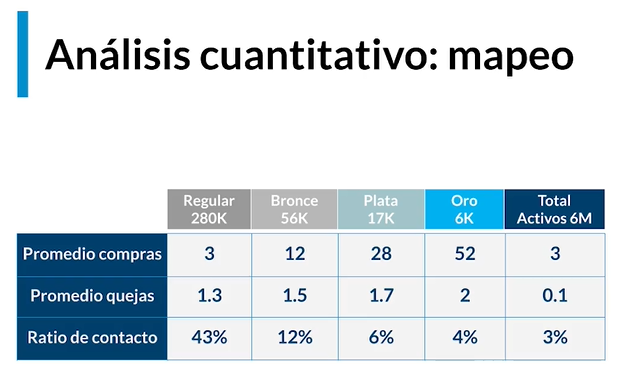
# Análisis cuantitativo en un caso de negocio (numeros)











Los datos cuantitativos nos dicen qué clientes se quejaron y cuantas veces  
Los datos cualitativos nos explican el motivo de su queja (devolución de dinero, problemas técnicos, dudas, etc)

Análisis cualitativo en un caso de negocio

**Clusterizar**: Agrupar las problemáticas, motivaciones y contactos por parte de los clientes a las que nos enfrentamos. Este ejemplo se clasificaron en 4 categorías las más de 500 tipos contactos:

* **Motivación económica**. Buscan que se les devuelva el dinero. Ésta es la más grave ya que además de ser una queja, se les devuelve el dinero.
* **Preguntas**. Los clientes tienen dudas.
* **Problemas tecnológicos**. Relevante para identificar fallas técnicas del servicio o producto.
* **Política de empresa**.  
  

## Clasificación

Definir los motivos detrás de los mensajes, cuando los hayamos clasificado dentro de una categoría profundizaremos en los motivos de contacto de una manera geolocalizada por lo que tendremos especificidad en cada contexto social.



Problema: La empresa tiene montos de dinero sin aplicar de facturas pagadas por los clientes en toda latinoamérica

Categorías:

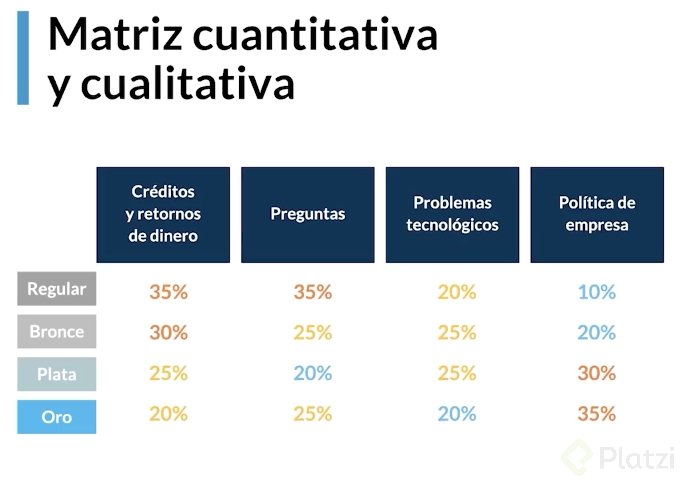
* La factura no está disponible
* El pago no indica que facturas debe aplicarse
* El cliente pago más del valor de la facturación
* El cliente pago menos del valor de la facturación

## Fusión cuanti-cualitativa en un caso de negocio

Resolver de manera conjunta la información cuantitativa y cualitativa para sacar conclusiones.

Una vez que tenemos información sobre cuáles son las razones por las que los clientes nos contactan así como una clasificación de los top offenders vamos a ver por qué se queja cada uno de los tipos de top offenders para encontrar las claves del análisis.

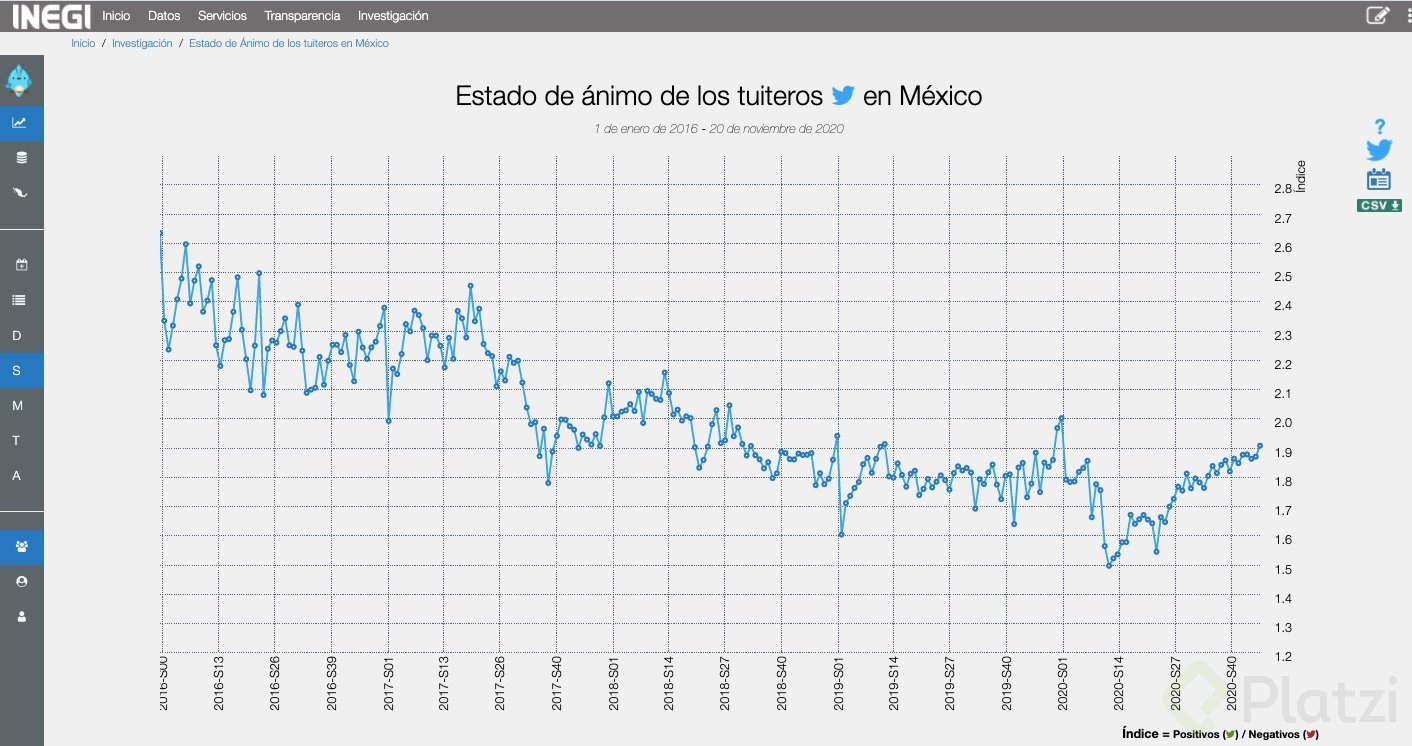
Para este análisis se utiliza una tabla con un mapa de calor para resaltar los porcentajes más elevados y poder concentrar nuestra atención en los problemas mayores más rapidamente.



**Mi solución al reto**

**Problema: 🧑Algunos clientes no reciben sus entregas y el sistema marca que SÍ se realizaron las entregas.**

* En mi problema puedo relacionar que el sistema le da mayor prioridad de entrega a los clientes más importantes, y esto sería la relación entre sistema-cantidad de compras.
* También puedo relacionarlo al problema con si contabilizamos todas las compras que hace por día cada usuario, hacen que los repartidores no puedan entregar todo lo que correspondía al día.
* El último punto también se puede ver afectado por qué tan lejos está el domicilio al que se iba a realizar la entrega, por lo tanto habría que revisar la cantidad de entregas que se realizan por día y el domicilio de estos.
* Por último, puede existir un bug en el sistema que haga que marque que SÍ se realizaron las entregas o que marque la hora de las entregas fuera del horario laboral, y que esto se haya debido a la enrome cantidad de pedidos.

Un ejemplo de análisis cualitativo, minería de textos, uso de machine learning y crowdsourcing es la plataforma ESTADO DE ÁNIMO DE LOS TWITTEROS EN MÉXICO 🇲🇽 les recomiendo checarla es muy interesante, puedes ver el estado de ánimo en determinadas zonas y fechas:  
<https://www.inegi.org.mx/app/animotuitero/#/app/multiline>  


La **minería de textos** es una rama específica de la [minería de datos](https://es.wikipedia.org/wiki/Miner%C3%ADa_de_datos) que se refiere al proceso de analizar y derivar información nueva de textos. Por medio de la identificación de patrones o correlaciones entre los términos se logra encontrar información que no está explícita dentro del texto. Los textos que se usan como recursos pueden ser páginas web, libros, correos electrónicos, reseñas de clientes, artículos, entre otros.

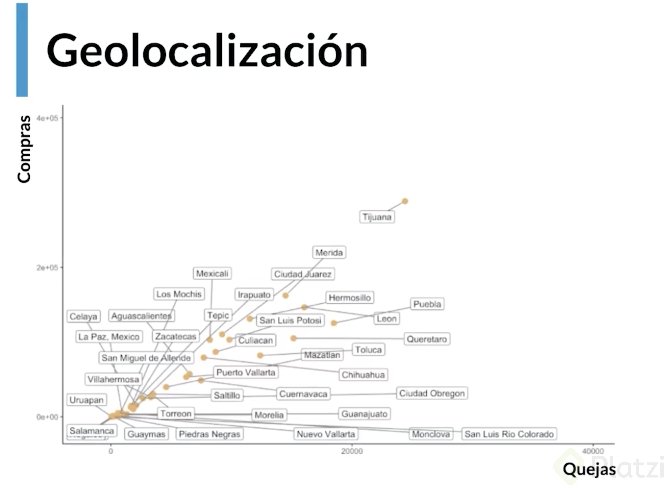
La minería de textos es un área multidisciplinaria basada en la recuperación de información, [aprendizaje automático](https://es.wikipedia.org/wiki/Aprendizaje_autom%C3%A1tico), [estadísticas](https://es.wikipedia.org/wiki/Estad%C3%ADstica) y la [lingüística computacional](https://es.wikipedia.org/wiki/Ling%C3%BC%C3%ADstica_computacional). Como la mayor parte de la información (más de un 80%) se encuentra actualmente almacenada como texto, se cree que la minería de textos tiene un gran valor comercial.

Por ejemplo: Twitter, tiene su base de negocio en la exploración de los mensajes.



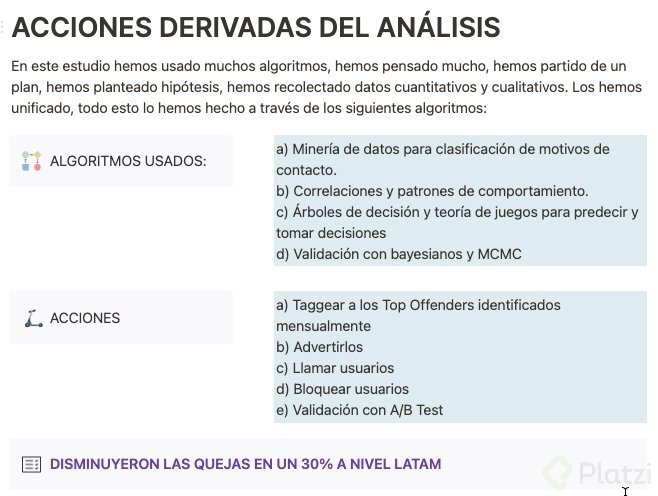
## Variación de comportamientos a partir de la geolocalización

Relevancia del comportamiento de acuerdo al origen geográfico por país, ciudad, distrito.

En el caso de top offenders se creó un gráfico que compara las quejas vs compras de acuerdo al origen geográfico de las mismas, a nivel de ciudad. Se espera obtener una línea de 45º si el comportamiento de los resultados es normal como se observa en el gráfico.  


Aquellas ciudades que se alejen de la línea son aquellas que tienen más top offenders de lo normal. Las ciudades con más top offenders son Puebla, Querétaro, Toluca; esto se define al observar que las ciudades se encuentran más alejadas del eje normal y más cercanas al eje de las quejas.

## Acciones, algoritmos y toma de decisiones según los resultados del análisis



ALGORITMOS USADOS:  
a)Minería de texto: es el analisis de las palabras  
b)Encontrar correlaciones y patrones de comportamiento: por ejemplo aprender si los usuarios que mas compran son los que mas se quejan, y porque lo hacen, etc.  
c)Arboles de decisiones: Nos ayudan a la toma de decisiones, partiendo de una hipótesis A o una hipótesis B y nos iremos acercando con teoría de juegos para predecir cual es la mejor estrategia.  
d)Validación con bayesianos: la estadística bayesiana es muy útil para identificar patrones que se comportan de maneja conjunta. Por ej. cual es la probabilidad que si yo aviso a un usuario de que es un top offender, reincida o lo entienda y deje de comportarse asi  
e)Cadenas de Montecarlo: Son probabilidades concatenadas, conjuntas. Por ej. si un mes fue top offender, el mes siguiente lo va a ser? Ya no lo sera?. Son estudios que se hacen después de nuestro análisis, son métodos de validación. Es ver como impacto el modelo.

ACCIONES:  
a)Taggear, identificar a los top offenders: saber quienes son y taggearlos, para saber si ya reincidió una vez, si darle beneficios o no.  
b)Advertirlos.  
c)Llamar usuarios: en los mejores clientes, es mejor llamarlos y se obtiene un trato personalizado.  
d)Bloquear usuarios: es en casos extremos.  
e)Validación con A/B test: donde la A puede ser una llamada, y la B puede ser mandarles un mensaje y ver el comportamiento de cada uno para ver cual funciono mejor.

1. Solo podemos analizar información numérica, ya que es la única que nos permite hacer operaciones y tener métricas. Esta afirmación es:

Falsa

2.¿Qué tipo de información podemos analizar?

Todas las respuestas son correctas.

3. Tu empresa quiere convertirse en data-driven, ¿cuál de los siguientes pasos se debe llevar a cabo primero?

Pensar en la información relevante.

4. ¿Cuál es una de las particularidades y usos más importantes del machine learning?

Se mejora y perfecciona constantemente según las nuevas realidades y situaciones de la industria.

5. ¿En qué proyectos podemos implementar deep learning?

Todas las respuestas son correctas.

6.¿Qué lenguajes o programas debes aprender inicialmente como data scientist?

Lenguajes de programación como R y Python.

7.El primer lenguaje o programa que voy a usar cuando tenga una base de datos y quiera simplificarla es:

Un lenguajes de consulta a bases de datos como SQL.

8.Pandas y NumPy pertenecen a:

Python

9.ggplot2 y dplyr pertenecen a:

R

10. ¿Cuándo usamos la expresión GROUP BY de SQL?

Siempre que una variable dentro de SELECT sí esté agrupada.

11.¿Cuál de las siguientes consultas en SQL nos permite encontrar CUÁNTAS ventas se hicieron en febrero?

**SELECT** **COUNT** (**DISTINCT** **id**) **FROM** VENTAS WHEREmonth = 2

12. ¿Cuál de las siguientes consultas en SQL nos permite encontrar QUÉ CLIENTES únicos hicieron una compra en el año 2020?

**SELECT** (**DISTINCT** client\_id) **FROM** VENTAS **WHEREyear** = 2020

13. Las técnicas de storytelling nos permiten:

Estructurar nuestros análisis de una manera más eficiente.

14. Un caso de negocio tiene que partir de una hipótesis. Esta afirmación es:

Verdadera

15. ¿Cuál es la estructura de un análisis de caída de ventas de un mes?

Qué: las ventas cayeron de manera anómala. Por qué: explorar variables que expliquen el resultado. Cómo: encontrar efectos cíclicos y cambios en nuestras variables explicativas.

16. ¿Cuál de los siguientes es un ejemplo de análisis cuantitativo?

Identificar las ventas que se hicieron por establecimiento.

17. ¿Qué tipo de dato nos permite analizar información cualitativa?

Texto

18. Fusionar información cuantitativa y cualitativa nos ayuda a interpretar características categóricas de usuarios, operaciones y conjuntos de datos. Esta afirmación es:

Verdadera

19. ¿Cuál de estas compañías crees que use más minería de texto para sus análisis?

Twitter

20. Crear análisis con un nivel geográfico de alta agregación (países, continentes) puede incurrir en conclusiones sesgadas o erróneas. Esta afirmación es:

Verdadera

21. ¿Qué debes hacer si tu análisis muestra que algunas tiendas de nuestro negocio tienen pérdidas?

Evaluar los resultados y plantearnos el cierre de algunas sedes.