

Conceptos clave en la Ciencia de Datos para negocios + Introducción a SQL

Ciencia de Datos para Negocios - FCE UBA

Nicolás Sidicaro

20/05/2025

Curso de Ciencia de Datos para Negocios

20 de Mayo de 2025

Objetivos de la clase

- Comprender la importancia del **contexto de negocio** en proyectos de ciencia de datos
- Dominar los conceptos fundamentales para **traducir datos en valor empresarial**
- Introducir **SQL** como herramienta esencial para el análisis de datos
- Entender las **consultas básicas** y su aplicación en casos de negocio reales

Parte 1: Ciencia de Datos en el Contexto Empresarial

El ciclo de vida de un proyecto de datos

1. Entendimiento del negocio

- Identificar problemas y oportunidades
- Definir objetivos claros

2. Entendimiento de los datos

- Explorar fuentes disponibles
- Evaluar calidad y relevancia

3. Preparación de datos

- Limpieza y transformación
- Integración de fuentes

1. Modelado

- Técnicas estadísticas
- Algoritmos de machine learning

2. Evaluación

- Validación de resultados
- Alineación con objetivos

3. Implementación

- Despliegue de soluciones
- Comunicación de insights

KPIs (Key Performance Indicators)

¿Qué son?

- Métricas cuantificables que evalúan el desempeño hacia objetivos específicos
- Reflejan factores críticos de éxito en una organización

Características (SMART)

- Specific (Específicos): Claramente definidos y enfocados
- **M**asurable (Medibles): Cuantificables de manera objetiva
- Achievable (Alcanzables): Realistas dentro de los recursos disponibles
- Relevant (Relevantes): Alineados con los objetivos estratégicos
- Time-bound (Temporales): Con plazos definidos para su evaluación

KPIs por industria

Retail

- Ventas por metro cuadrado
- Ticket promedio
- Tasa de conversión
- Rotación de inventario

E-commerce

- Tasa de abandono del carrito
- Costo de adquisición de cliente (CAC)
- Valor de vida del cliente (LTV)
- Tasa de retorno de productos

Finanzas

- Tasa de morosidad
- Índice de eficiencia
- Rentabilidad del capital

Marketing

- Retorno de inversión (ROI)
- Tasa de engagement
- Costo por lead/conversión

Recursos Humanos

- Tasa de rotación
- Tiempo de contratación
- Índice de satisfacción laboral

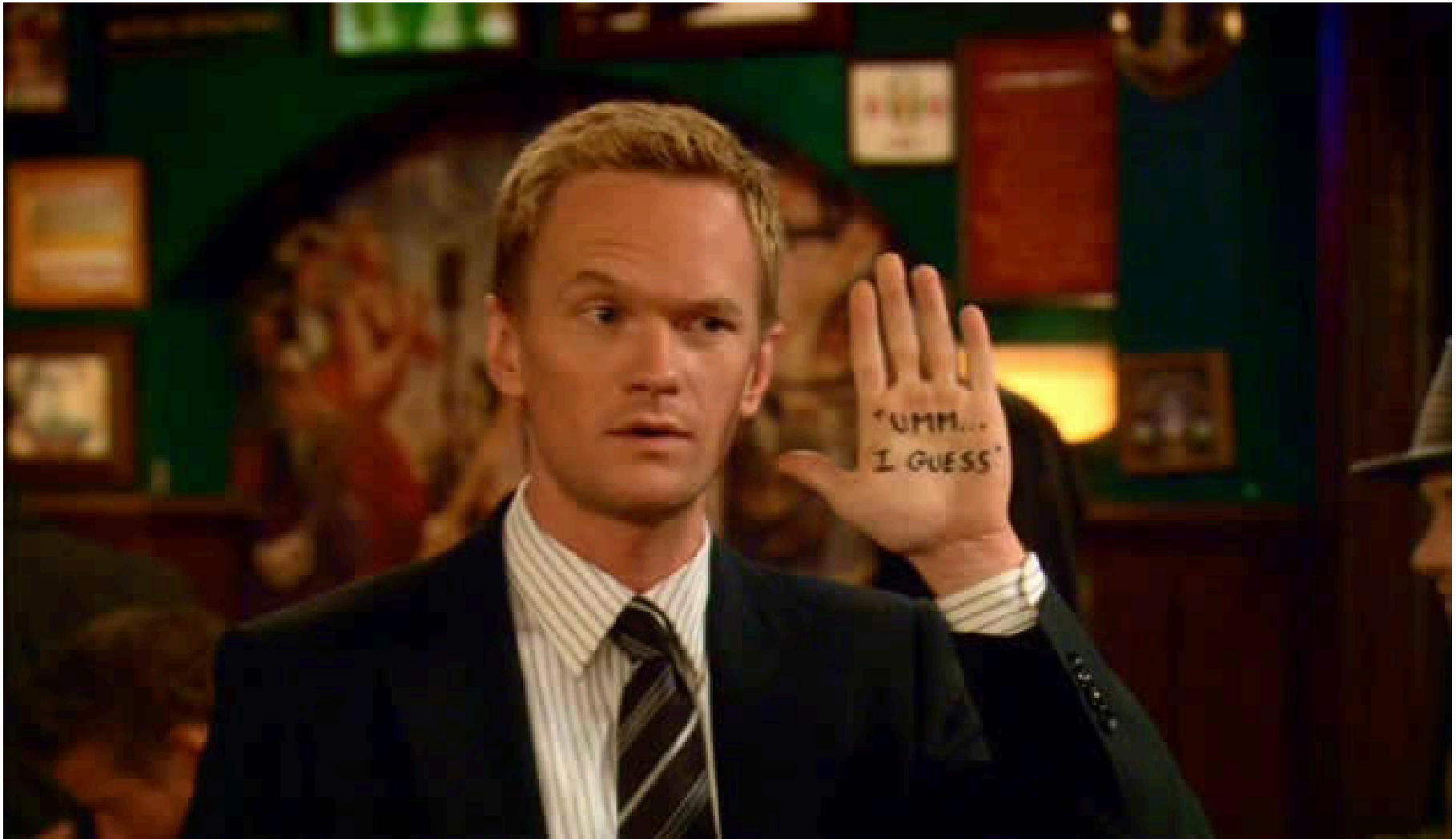
Del dato al KPI



Del dato al KPI

1. Identificación de objetivos estratégicos
2. Selección de métricas que indican progreso
3. Establecimiento de valores base y metas
4. Creación de sistemas de monitoreo
5. Evaluación periódica y ajustes

Del dato al KPI



Entendimiento del negocio

La importancia del contexto empresarial

- El valor de los datos solo existe en el contexto de las decisiones de negocio
- Cada industria y organización tiene sus propias dinámicas y desafíos
- El analista debe familiarizarse con la terminología específica del sector

Preguntas fundamentales para entender el negocio

- ¿Cuál es el **modelo de negocio**? ¿Cómo genera valor la empresa?
- ¿Quiénes son los **clientes** y cuáles son sus necesidades?
- ¿Quiénes son los **competidores** y cómo se diferencia la empresa?
- ¿Cuáles son los **procesos clave** y cómo se miden?
- ¿Qué **decisiones críticas** se toman regularmente?

Métodos para desarrollar el entendimiento del negocio

¿Sos o querés ser un consultor? Entonces...

- **Entrevistas** con stakeholders clave
- Inmersión en **documentación** corporativa
- **Observación** de procesos operativos
- Análisis de la **cadena de valor**
- **Benchmarking** con competidores

Data Storytelling

¿Qué es?

Combinación de **datos + visualización + narrativa** que:

- Transforma datos complejos en insights accionables
- Comunica hallazgos de manera memorable y persuasiva



Componentes del storytelling efectivo con datos

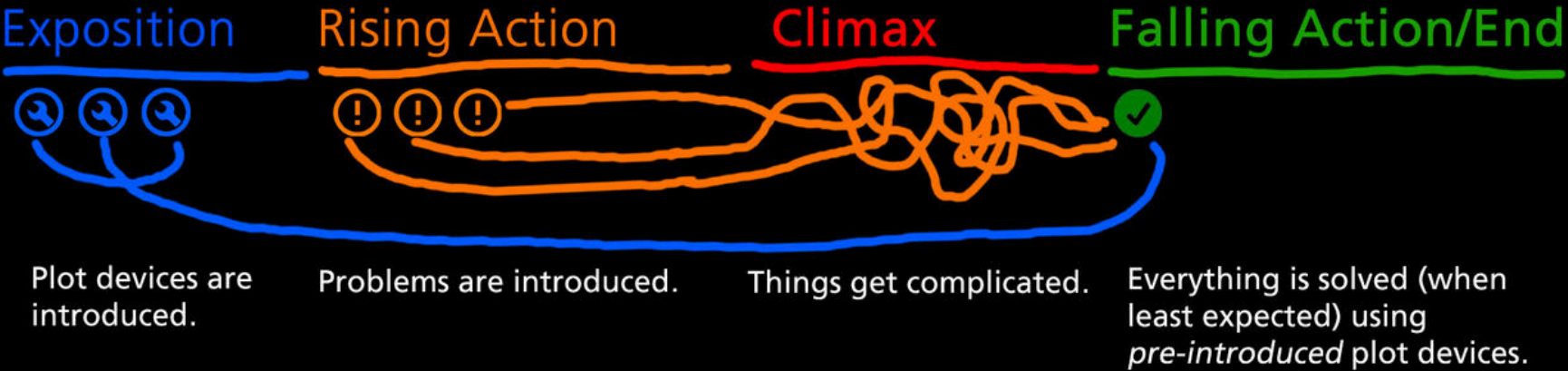
- **Contexto:** Situar los datos en el marco adecuado
- **Relevancia:** Conectar con las preocupaciones de la audiencia
- **Estructura narrativa:** Problema, complicación, resolución
- **Visualización apropiada:** Elegir el formato visual que mejor comunica el mensaje
- **Llamado a la acción:** Indicar claramente los próximos pasos

Hacer algo aburrido

Good vs. Bad Storytelling

sinandira.deviantart.com

Pixar



Anime



Errores comunes a evitar en data storytelling

- **Sobrecarga** de información
- Gráficos **engañosos** o manipuladores
- Narrativa **desconectada** de los datos
- **Ignorar** la audiencia específica
- Falta de conclusiones **accionables**

Técnicas de storytelling con datos

- "Zoom in, zoom out": Alternar entre la visión general y el detalle
- Comparaciones y contrastes para dar contexto
- Uso de analogías y metáforas para explicar conceptos complejos
- Personalización con ejemplos concretos y casos de estudio
- Presentación progresiva de la información

Parte 2: Introducción a SQL para Analistas de Datos

¿Qué es SQL y por qué es importante?

Definición y contexto

- Structured Query Language: lenguaje estándar para interactuar con bases de datos relacionales
- Creado en los años 70, se ha convertido en EL LENGUAJE para el trabajo con datos estructurados
- Permite a los analistas acceder directamente a los datos sin depender de áreas de gobierno (todo en *cloud/servidor*)

Ventajas de SQL para analistas de datos

- **Acceso directo** y específico a grandes volúmenes de datos
- Procesamiento **eficiente** en el servidor de base de datos
- Capacidad para **combinar datos** de múltiples fuentes
- Lenguaje **declarativo**: se especifica qué datos se quieren, no cómo obtenerlos
- Habilidad **transferible** entre diferentes sistemas y empresas

Sistemas de gestión de bases de datos relacionales

- PostgreSQL

- Open source
- Robusto y escalable
- Excelente para análisis

- MySQL/MariaDB

- Open source
- Muy usado en aplicaciones web
- Fácil de configurar

- Oracle Database

- Empresarial
- Alto rendimiento
- Funciones avanzadas

- Microsoft SQL Server

- Ecosistema Microsoft
- Power BI integración
- Buenas herramientas de administración

- SQLite

- Ligero, embebido
- No requiere servidor

Fundamentos de bases de datos relacionales

Conceptos clave

- **Base de datos:** colección organizada de información estructurada
- **Tabla:** estructura que organiza datos en filas y columnas
- **Fila/Registro:** instancia individual de datos
- **Columna/Campo:** atributo específico que describe los datos
- **Clave primaria:** identificador único para cada registro
- **Clave foránea:** campo que vincula tablas entre sí

Modelo relacional

- Organización de datos en **relaciones** (tablas)
- **Normalización** para reducir redundancia
- **Integridad referencial** para mantener consistencia
- **Relaciones** entre tablas
 - Uno a uno (1:1)
 - Uno a muchos (1:N)
 - Muchos a muchos (N:M)

Ejemplo conceptual: Base de datos de ventas retail

- **Clientes:** ID_Cliente, Nombre, Email, Dirección
- **Productos:** ID_Producto, Nombre, Categoría, Precio
- **Ventas:** ID_Venta, ID_Cliente, Fecha, Total
- **Detalle_Ventas:** ID_Detalle, ID_Venta, ID_Producto, Cantidad, Precio_Unitario

Fundamentos de SQL para analistas

Las consultas SELECT: el núcleo del análisis

Estructura básica de una consulta:

- **SELECT:** campos a recuperar
- **FROM:** tablas de origen
- **WHERE:** condiciones de filtrado
- **GROUP BY:** agrupación para agregaciones
- **HAVING:** filtros sobre agregaciones
- **ORDER BY:** ordenamiento de resultados
- **LIMIT/TOP:** restricción del número de resultados

Funciones principales en SQL

Funciones de agregación

- **COUNT()**: contar registros
- **SUM()**: sumar valores
- **AVG()**: calcular promedios
- **MIN()/MAX()**: encontrar valores extremos

Funciones de texto

- **CONCAT()**: unir cadenas
- **SUBSTRING()**: extraer parte de una cadena
- **UPPER()/LOWER()**: cambiar mayúsculas/minúsculas

Funciones de fecha

- **EXTRACT()**: obtener componentes (año, mes, día)
- **DATE_TRUNC()**: truncar a nivel específico
- **DATEDIFF()**: calcular diferencia entre fechas

Funciones de ventana

- **RANK()**: asignar ranking
- **ROW_NUMBER()**: numerar filas
- **SUM() OVER()**: sumas acumulativas

JOINS: combinando datos de múltiples tablas

- **INNER JOIN:** solo registros con correspondencia en ambas tablas
- **LEFT JOIN:** todos los registros de la tabla izquierda
- **RIGHT JOIN:** todos los registros de la tabla derecha
- **FULL JOIN:** todos los registros de ambas tablas

SQL para casos de uso empresariales: Ventas

Casos de uso típicos

- Segmentación de clientes por valor y comportamiento
- Análisis de canasta de compra (productos que se compran juntos)
- Cálculo de métricas como LTV (Lifetime Value) y CAC (Customer Acquisition Cost)
- Evaluación de campañas y promociones

Preguntas que podemos responder

- ¿Cuáles son nuestros clientes más valiosos?
- ¿Qué productos se compran frecuentemente juntos?
- ¿Qué campañas generan mayor ROI?
- ¿Cómo varía el comportamiento de compra por segmento/región/temporada?

SQL para casos de uso empresariales: Finanzas

Casos de uso típicos

- Reportes de ingresos y gastos por período
- Análisis de rentabilidad por producto o cliente
- Proyecciones y comparaciones año contra año
- Detección de anomalías en transacciones

Preguntas que podemos responder

- ¿Cuáles son los productos más rentables?
- ¿Cómo se comparan los resultados con el mismo período del año anterior?
- ¿Qué unidades de negocio están superando sus objetivos?
- ¿Existen patrones inusuales en los gastos o transacciones?

SQL para casos de uso empresariales: Operaciones

Casos de uso típicos

- Monitoreo de inventarios y rotación de productos
- Evaluación de eficiencia de procesos
- Identificación de cuellos de botella
- Análisis de tiempos de servicio

Preguntas que podemos responder

- ¿Qué productos tienen riesgo de quiebre de stock?
- ¿Cuáles son los tiempos promedio de entrega por región?
- ¿Dónde se producen los mayores retrasos en el proceso?
- ¿Cómo varía la utilización de recursos a lo largo del tiempo?

SQL para casos de uso empresariales RRHH

Casos de uso típicos

- Patrones de rotación de personal
- Evaluación de desempeño por departamentos
- Análisis de compensaciones y equidad salarial
- Planificación de capacidad

Preguntas que podemos responder

- ¿Qué departamentos tienen mayor rotación de personal?
- ¿Existe correlación entre capacitación y desempeño?
- ¿Hay disparidades salariales no justificadas?
- ¿Cómo proyectar necesidades futuras de personal?

Buenas prácticas y consejos prácticos

Optimización de consultas

- Seleccionar solo las columnas necesarias (evitar SELECT *)
- Limitar el volumen de datos procesados (usar WHERE eficientemente)
- Entender y usar índices adecuadamente
- Considerar el orden de las operaciones

Organización del trabajo con SQL

- Comentar consultas complejas
- Usar CTEs (Common Table Expressions) para consultas modulares
- Crear **vistas** para consultas frecuentes
- Mantener **consistencia** en convenciones de nombres

```
744         error => $quote['sort_order'],
745     };
746     };
747     };
748     };
749     };
750     $sort_order = array();
751     foreach ($quotes as $key => $value) {
752         $sort_order[$key] = $value['sort_order'];
753     }
754     array_multisort($sort_order, SORT_ASC, $quotes);
755     $this->session->data['lpa']['shipping_methods'] = $quotes;
756     $this->session->data['lpa']['address'] = $address;
757     if (empty($quotes)) {
758         $json['error'] = $this->language->get('
759             error_no_shipping_methods');
760     } else {
761         $json['quotes'] = $quotes;
762     }
763     };
764     };
765     if (isset($this->session->data['lpa']['shipping_method']) && !
766         empty($this->session->data['lpa']['shipping_method']) &&
767         isset($this->session->data['lpa']['shipping_method']['code'])
768     ) {
769         $json['selected'] = $this->session->data['lpa']['
770             shipping_method']['code'];
771     } else {
772         $json['selected'] = '';
773     }
774     } else {
775         $json['error'] = $this->language->get('error_shipping_methods');
776     }
777     $this->response->addHeader('Content-Type: application/json');
```

```
382     if (this.paused = function () {
383         if (this.$element.find('.next, .prev').length && $.support.transition) {
384             this.$element.trigger($.support.transition.end)
385             this.cycle(true)
386         }
387         this.interval = clearInterval(this.interval)
388         return this
389     }
390     Carousel.prototype.next = function () {
391         if (this.sliding) return
392         return this.slide('next')
393     }
394     Carousel.prototype.prev = function () {
395         if (this.sliding) return
396         return this.slide('prev')
397     }
398     Carousel.prototype.slide = function (type, next) {
399         var $active = this.$element.find('.item.active')
400         var $next = next || this.getItemForDirection(type, $active)
401         var isCycling = this.interval
402         var direction = type == 'next' ? 'left' : 'right'
403         var fallback = type == 'next' ? 'first' : 'last'
404         var that = this
405         if (!$next.length) {
406             if (!this.options.wrap) return
407             $next = this.$element.find('.item')[fallback]()
408         }
409         if ($next.hasClass('active')) return (this.sliding = false)
410         var relatedTarget = $next[0]
411         var slideEvent = $.Event('slide.bs.carousel', {
412             relatedTarget: relatedTarget,
413             direction: direction
414         })
415         this.$element.trigger(slideEvent)
```

Recursos para seguir aprendiendo SQL

- Plataformas de práctica online
 - SQLZoo
 - HackerRank - Para *developers*
- Documentación específica de cada SGBD
- Comunidades como **Stack Overflow**

Conclusión

- SQL como **punto** entre datos crudos y análisis avanzados
- **Complementariedad** con herramientas como R y Python

Actividad práctica

Escenario: Retail y optimización de inventario

Contexto

Una cadena de retail quiere analizar el desempeño de sus productos por categoría y región para optimizar su inventario.

Trabajo en grupos (15 minutos)

1. ¿Qué tablas necesitaríamos para este análisis?
2. ¿Qué KPIs serían relevantes para medir el desempeño?
3. ¿Qué tipo de consultas SQL serían necesarias?

