



**Universidad Autónoma del Estado de México**

**Material didáctico multimedia**

**Sólo visión**

**Título del material: La metodología CRISP-DM**

**Unidad de Aprendizaje Minería de Datos  
Licenciatura de Ingeniería en Computación  
Facultad de Ingeniería**

**Elaborado por M en I Sara Vera Noguez  
Durante el período intensivo verano 2016**

- **Minería de datos**
- **Unidad 5. Metodología CRISP-DM y minería de datos complejos**

M en I Sara Vera Noguez

# Contenido

- Metodología CRIPS-DM
- Ciclo en espiral para proyectos de minería

# Modelos de proceso para desarrollo de SI

- ¿Qué nos proponen los modelos de procesos par el desarrollo de software?
- ¿Qué flujos de trabajo tiene la metodología del Proceso Unificado?

# ¿Para qué un proceso?



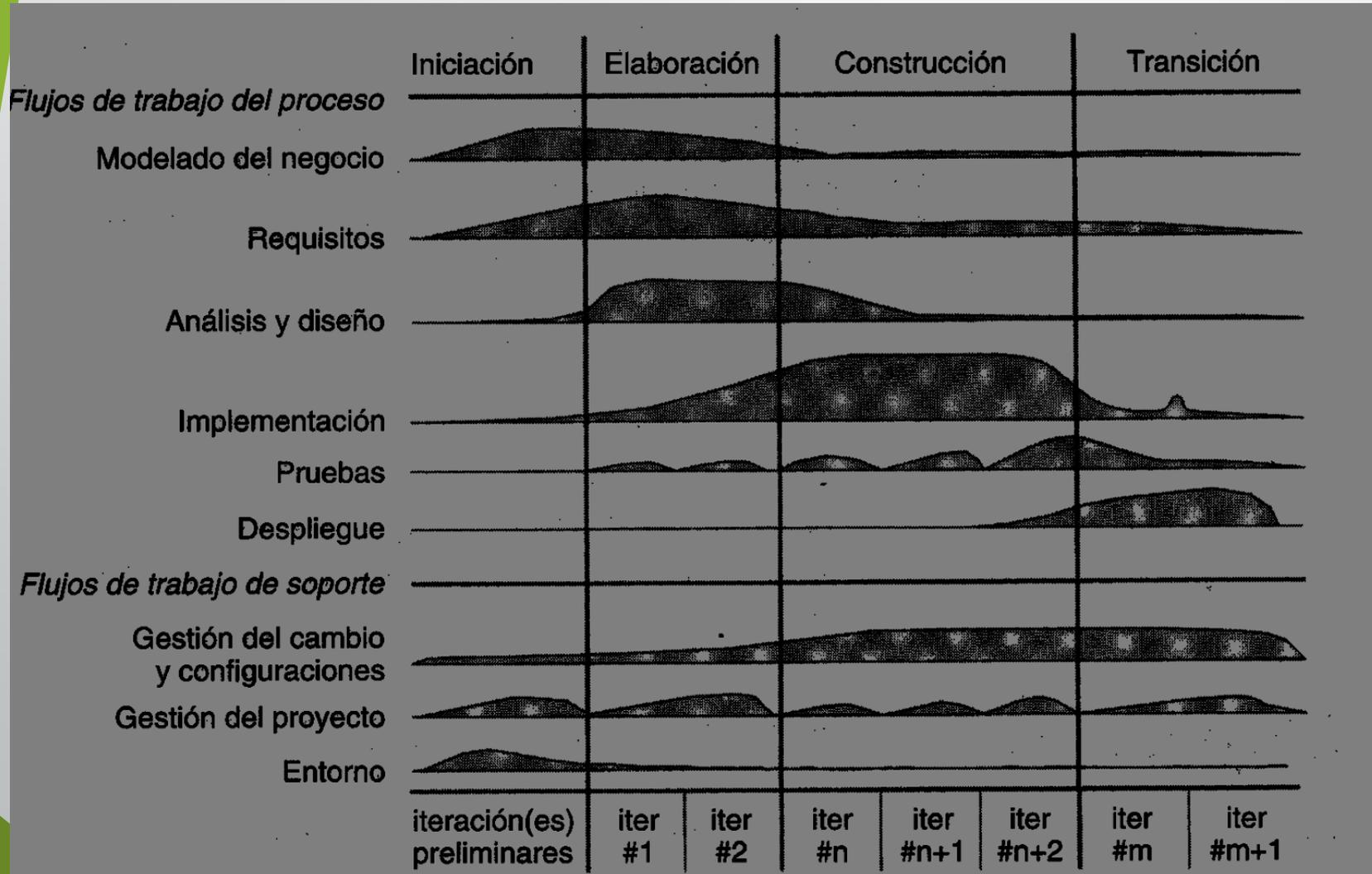
# ¿Para qué un proceso?

- Los modelos de proceso nos brindan una guía.
- Se pretende:
  - Reducir el riesgo del proyecto
  - Aumentar la productividad del equipo de trabajo
  - Obtener experiencia

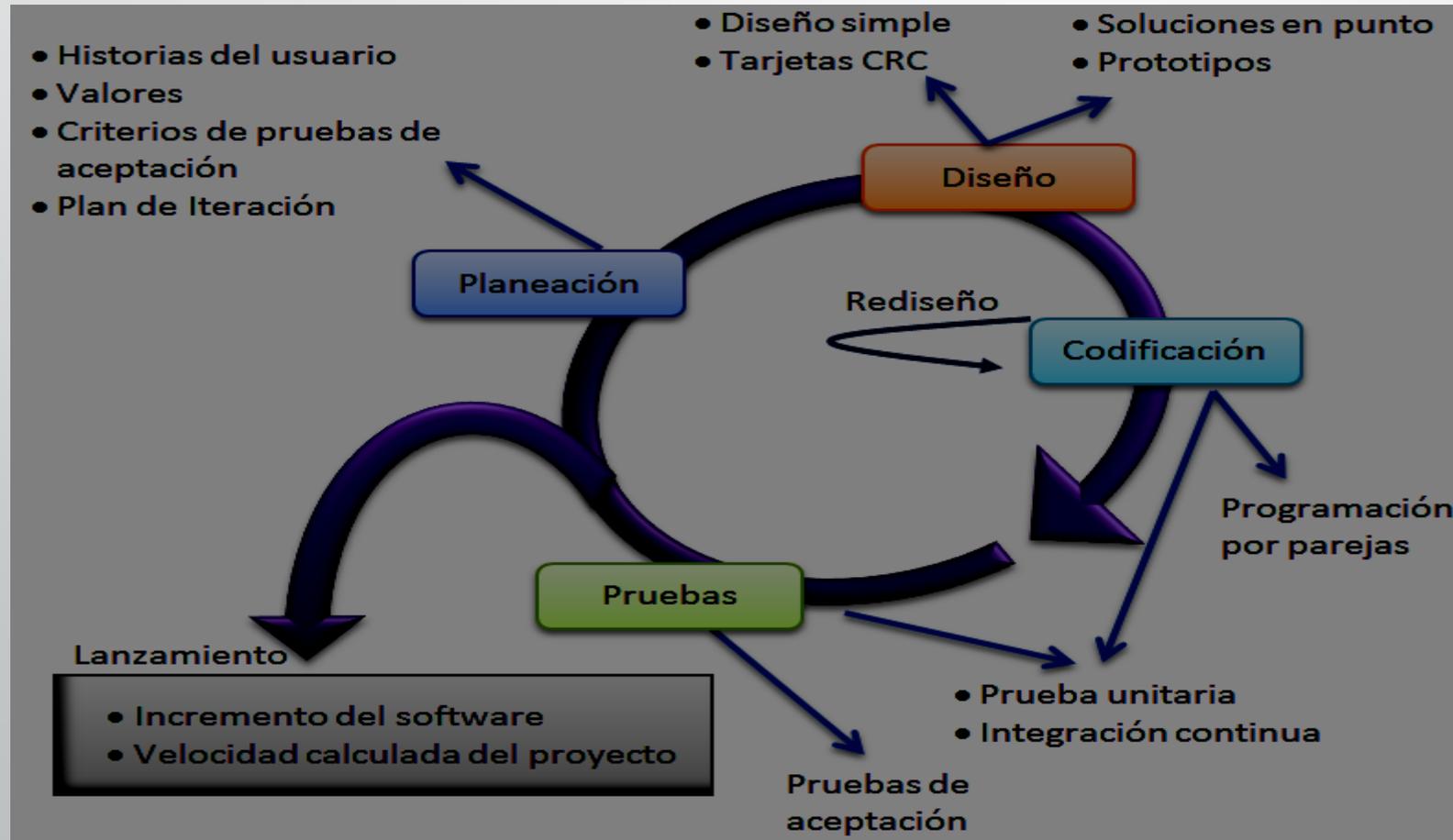
# Qué nos dice el proceso

- Nos brinda un panorama general
- Nos dice que hacer
- **No** explica cómo hacer las actividades
- Se complementa con las metodologías

# Recordando UP



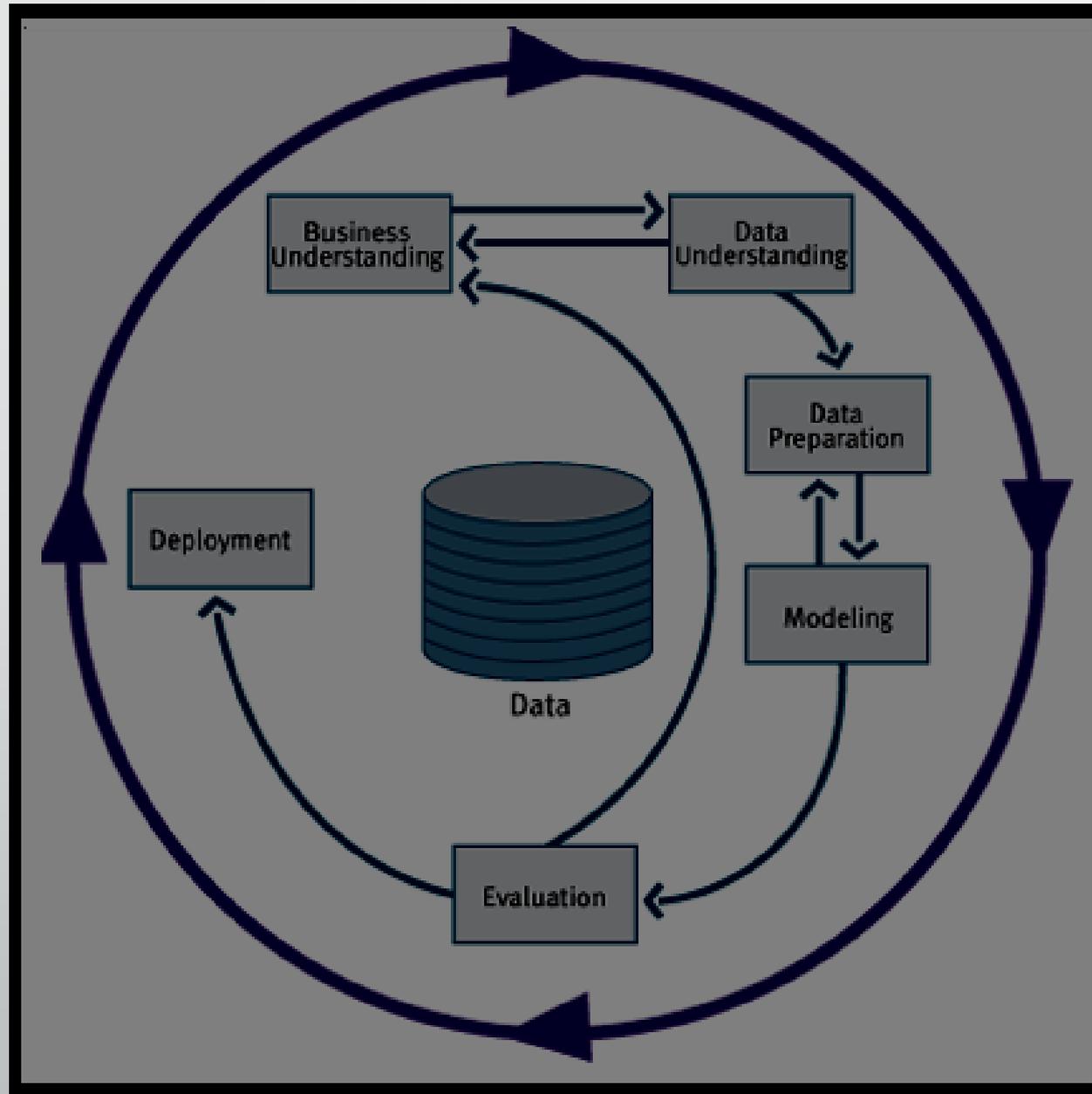
# Recordando XP



# Semejanzas y diferencias

- Se basa en el enfoque de sistemas
- El desarrollo surge de necesidades de su contexto
- Los requerimientos son la guía del desarrollo
- El flujo de trabajo (actividad) de diseño e implementación se basa en las herramientas para la implementación.
- ¿Cómo implementar minería de datos?

# La metodología CRISP-DM



# La metodología CRISP-DM

- Comprensión del negocio (Objetivos y requerimientos desde una perspectiva no técnica)
- Comprensión de los datos (Familiarizarse con los datos teniendo presente los objetivos del negocio)
- Preparación de los datos (Obtener la vista *minable* o *dataset*)
- Modelado (Aplicar las técnicas de minería de datos a los *dataset*)
- Evaluación (De los modelos de la fase anteriores para determinar si son útiles a las necesidades del negocio)
- Despliegue (Explotar utilidad de los modelos, integrándolos en las tareas de toma de decisiones de la organización)

# La metodología CRISP-DM

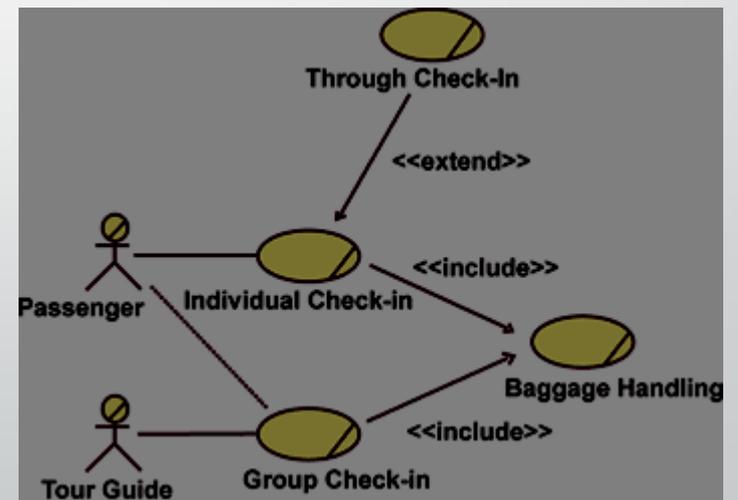
- **Comprensión del negocio** (Objetivos y requerimientos desde una perspectiva no técnica)
  - × Establecimiento de los objetivos del negocio (Contexto inicial, objetivos, criterios de éxito)
  - × Evaluación de la situación (Inventario de recursos, requerimientos, supuestos, terminologías propias del negocio,...)
  - × Establecimiento de los objetivos de la minería de datos (objetivos y criterios de éxito)
  - × Generación del plan del proyecto (plan, herramientas, equipo y técnicas)

# La metodología CRISP-DM

- **Comprensión del negocio** (Esto se puede hacer por ejemplo como en UP, mediante el modelado organizacional, o a la manera de XP, mediante charlas con el personal de la organización)



Unidad 5.



# La metodología CRISP-DM

- **Comprensión de los datos** (Familiarizarse con los datos teniendo presente los objetivos del negocio)
  - × Recopilación inicial de datos
  - × Descripción de los datos
  - × Exploración de los datos
  - × Verificación de calidad de datos

# La metodología CRISP-DM

- **Preparación de los datos** (Obtener la vista *minable* o *dataset*)
  - × Selección de los datos
  - × Limpieza de datos
  - × Construcción de datos
  - × Integración de datos
  - × Formateo de datos



# La metodología CRISP-DM

- **Modelado** (Aplicar las técnicas de minería de datos a los *dataset*)
  - × Selección de la técnica de modelado
  - × Diseño de la evaluación
  - × Construcción del modelo
  - × Evaluación del modelo

# La metodología CRISP-DM

- **Evaluación** (De los modelos de la fase anteriores para determinar si son útiles a las necesidades del negocio)
  - × Evaluación de resultados
  - × Revisar el proceso
  - × Establecimiento de los siguientes pasos o acciones



# La metodología CRISP-DM

- **Despliegue** (Explotar utilidad de los modelos, integrándolos en las tareas de toma de decisiones de la organización)
  - × Planificación de despliegue
  - × Planificación de la monitorización y del mantenimiento
  - × Generación de informe final
  - × Revisión del proyecto

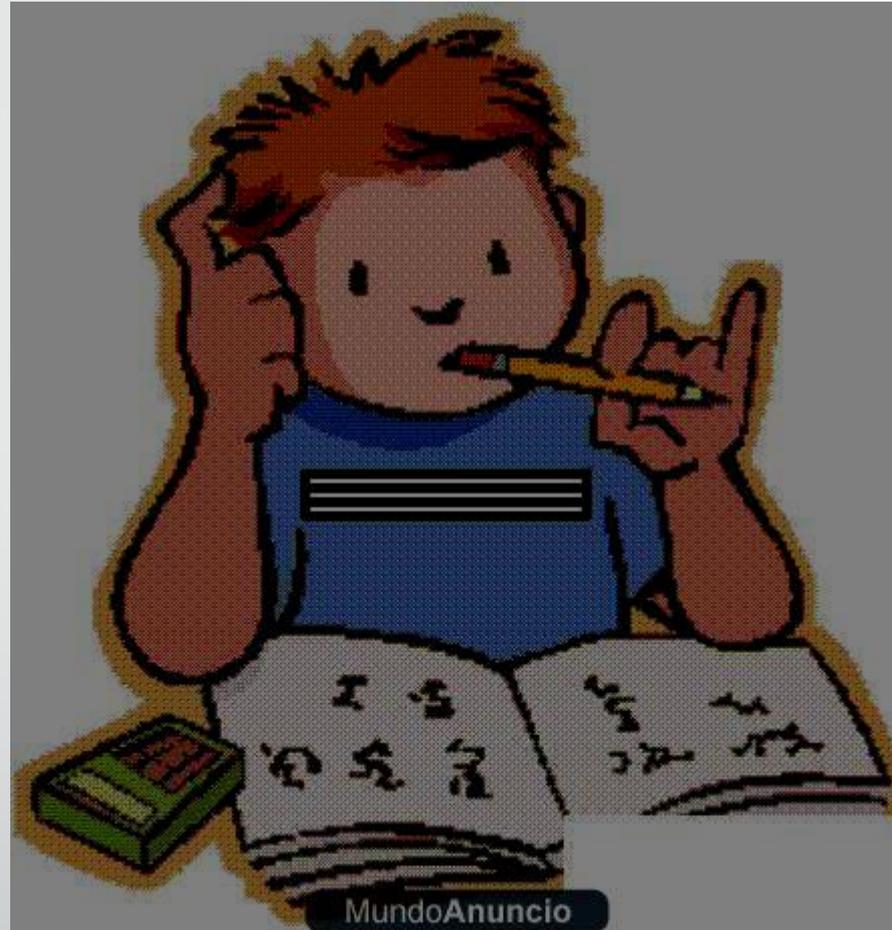
# Actividad

- De forma individual realiza un resumen de las tareas que se debe hacer de acuerdo a la metodología CRISP-DM

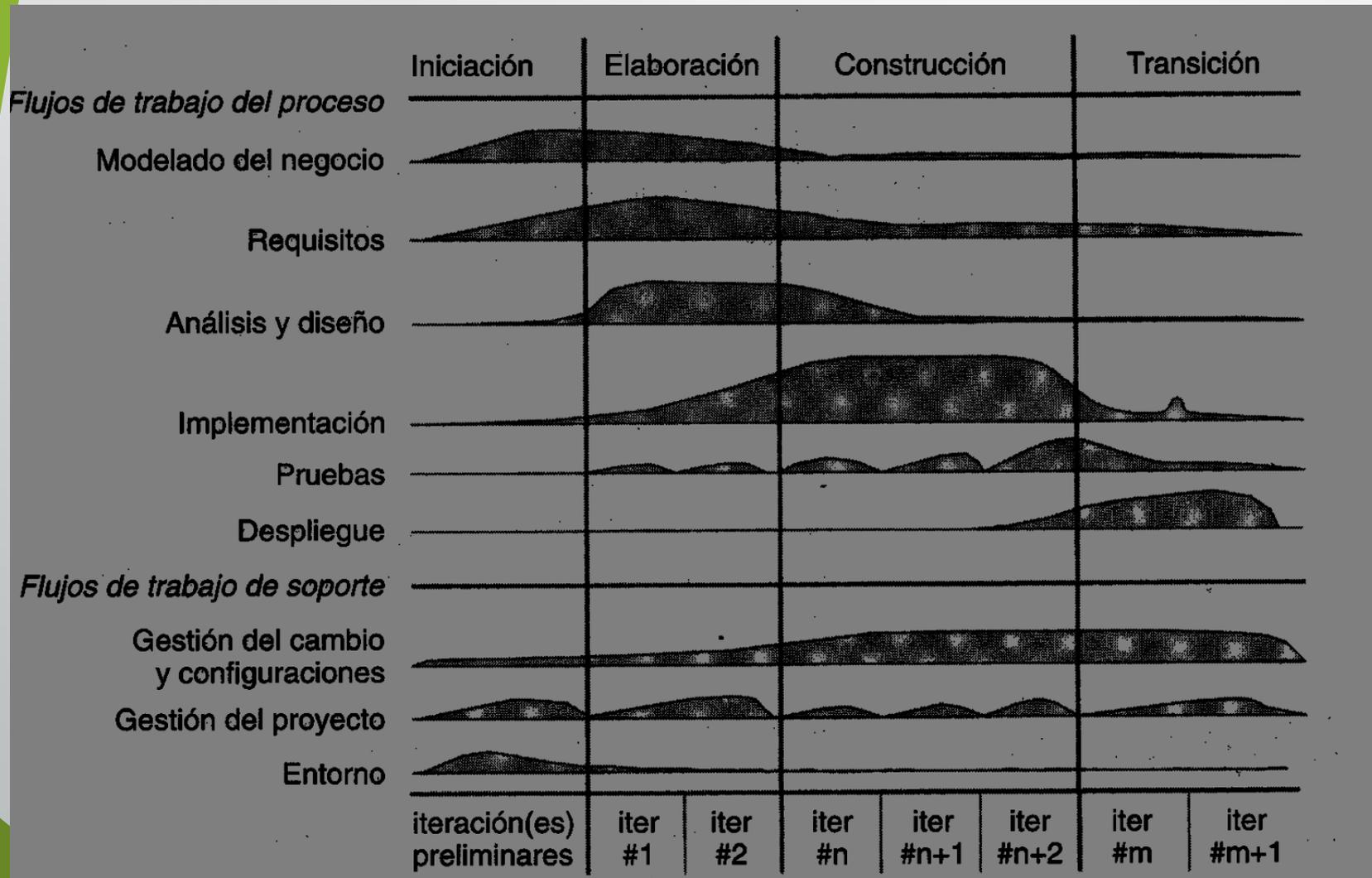
# Actividad

- En equipos de tres integrantes propone mecanismo concretos tomados de UP, XP, bases de datos, UML, u otra notación o metodología para plasmar los resultados de las tareas propuestas por CRISP-DM

# Repaso



# Como lo plantea UP



# El proceso de desarrollo

- En pocas palabras es:
  - Identificar requerimientos
  - Diseñar una solución
  - Implementar
  - Probar

# Productos generados en UP

- Del modelado del negocio se obtiene un modelo que incluye diversas vistas de la organización como una vista externa, una estructural y una de sus procesos (dinámica); es el insumo para la identificación de los requisitos
- Del flujo de requisitos se obtiene los requerimientos funcionales (¿qué?) plasmados en los casos de uso y los no funcionales (¿con qué?)

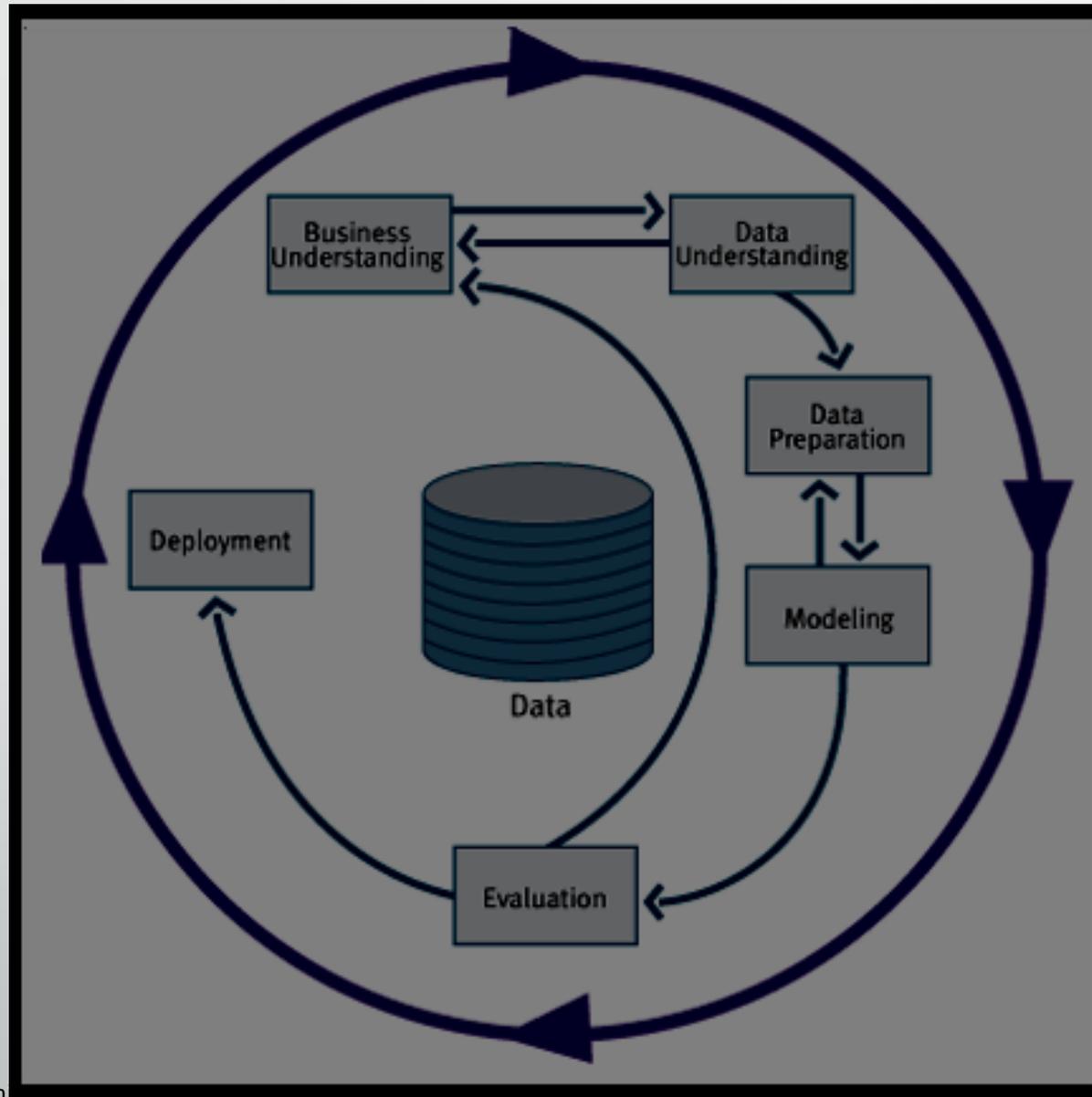
# UP

- Flujo de trabajo de análisis y diseño: se analizan los requerimientos, se especifica la arquitectura de la solución que incluye módulos, interfaces, persistencia (diseño de las BD), se planea la colaboración entre componentes (diagramas de secuencia y de colaboración), algoritmos.

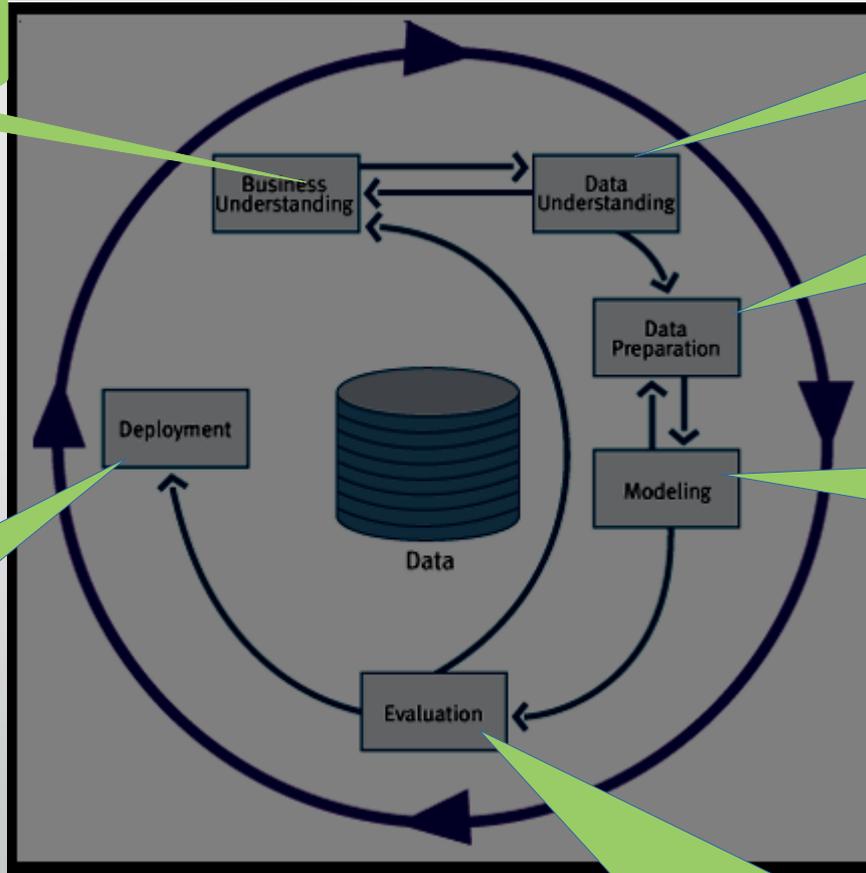
# UP

- En el flujo de implementación se genera el código que materializa el diseño.
- El flujo de trabajo de pruebas se realiza a lo largo de todo el proceso buscando probar o validar los distintos productos generados
- Finalmente en el despliegue el software es emplazado en el hardware para su operación cotidiana

# CRISP-DM y UP



# CRISP-DM y UP



Modelado de negocio

Modelado de negocio  
Y/o  
Requerimientos no funcionales

Conlleva análisis, diseño,  
implementación y pruebas

Conlleva análisis, diseño,  
implementación y pruebas

Se realiza durante el  
despliegue

Es parte de las pruebas

# Actividad

- En equipos contrastar sus propuesta de la actividad anterior con lo mostrado en la diapositiva número 27

# Conclusión

- Los flujos de trabajo del modelo no cambian
- Se debe incluir en los flujos de trabajo las tareas propias de la minería de datos

# Bibliografía básica

- Hernández Orallo, J., M. J. Ramírez Quintana, et al. (2004). Introducción a la Minería de Datos. España, Pearson Educación SA.
- Han, D. J. (2007). Principles of Data Mining, MIT Press.
- <https://anibalgoicochea.com/2009/08/11/crisp-dm-una-metodologia-para-proyectos-de-mineria-de-datos/>

# Bibliografía complementaria

- Everitt, B.S. (1994). A Handbook of Statistical Analyses using S-Plus. Chapman and Hall.
- Inmon, W.H. (1996). Building the Datawarehouse. J.Wiley & Sons.
- Han, J. and M. Kamber (2006). Data mining: concepts and techniques, Morgan Kaufmann.
- Kimball, R (1996). The Data Warehouse Toolkit. John Wiley & Sons.
- Hastie, T., R. Tibshirani, et al. (2005). The elements of statistical learning: data mining, inference and prediction, Springer
- Dunham. H. Margaret (2003). Data Mining. Introductory and Advanced Topics, Prentice Hall.
- Pyle, D. (1999), "Data Preparation for Data Mining", Morgan Kaufmann, San Francisco, CA.
- Hand, David; Mannila, Heikki; Smyth, Padhraic (2001), Principles of Data Mining, A Bradford Book. The MIT Press.
- Ian Witten and Eibe Frank (2002), Data Mining, Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann Publishers.