
 <p>Universidad de los Andes Colombia</p> <p>Accreditación institucional de alta calidad <b>10 años</b> Ministerio de Educación Resolución 1853 del 9 de enero de 2015</p>	<p>Ingeniería de Sistemas y Computación <b>Pregrado</b> ISIS-3301 – Inteligencia de Negocios Semestre: 2021-20</p>	
---	--	---

### CASO A TRABAJAR: Infraestructura visible

El proyecto se realiza en la temática de Infraestructura visible<sup>1</sup>, una **iniciativa de la Universidad de los Andes que permite el libre acceso a la información sobre la infraestructura colombiana y su relación con indicadores de desarrollo socioeconómico**. El proyecto busca aprovechar los datos recolectados en esta iniciativa para generar conocimiento útil que aporte a diferentes tipos de actores, desde el ciudadano normal hasta el experto en temáticas allí trabajadas, pasando por actores como periodistas, entre otros. De igual manera, quieren ampliar la gama de análisis que se pueden observar en la página web de la iniciativa.

Al final del proyecto, el cliente espera contar con tableros de control que le permitan analizar la información relacionada especialmente con las muertes en Colombia.



La empresa está muy interesada en conocer los estilos de análisis que pueden hacer y por supuesto, en ver las nuevas propuestas que pueden ayudarles a definir un mapa de ruta para futuras consultorías. En este orden de ideas, junto con los expertos en inteligencia de negocio de la organización y algunos asesores, han planteado el siguiente plan de trabajo.

1. **Etapa 1:** Desarrollo de tableros de control acompañado de *data marts* que incluyen el modelo dimensional, la carga de datos en una base de datos y el perfilamiento y preparación de los datos. Este desarrollo debe tener en cuenta las recomendaciones dadas por el cliente a los entregables del semestre anterior para no repetir los mismos errores.
2. **Etapa 2:** Identificación e implementación de nuevos requerimientos analíticos de tipo tableros de control y aprendizaje de máquina, acompañados de *data marts* que incluyen el modelo dimensional, la carga de datos en una base de datos y el perfilamiento y preparación de los datos.

Estas etapas se describen en detalle a continuación.

---

<sup>1</sup> <https://www.infraestructuravisible.org/>

 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2021-20</p>	
--	---	---

## ENUNCIADO SEGUNDO PROYECTO: ENTREGA 1

### OBJETIVOS

Este trabajo tiene como objetivos

- Desarrollar un proyecto de inteligencia de negocios centrado en el uso de datos para monitorear la estrategia de una organización.
- Comprender un requerimiento analítico que pueda ser resuelto con tableros de control.
- Implementar requerimientos analíticos basados en tableros de control que son de interés para la organización.
- Desarrollar *data marts* que permitan resolver los requerimientos analíticos seleccionados por el cliente.
- Proponer una arquitectura de solución de BI para los análisis a construir
- Crear procesos de ETL



### DESCRIPCIÓN

Implementar un requerimiento analítico que consideren es de interés para el cliente del proyecto, en la categoría de tablero de control. En esta etapa del proyecto se debe: (1) crear *data marts* con la o las tablas de hecho que considere necesarias, (2) cargar las tablas de hechos seleccionadas con sus respectivas dimensiones, y (3) diseñar e implementar los tableros de control. La descripción detallada de las actividades a realizar y de los elementos mínimos a incorporar en el documento se presenta a continuación:

1. **(20%) Modelar Data Marts:** A partir de las fuentes de datos anexas a esta entrega y de los ejemplos de requerimientos analíticos basados en tableros de control provistos por el cliente, proponga el modelo dimensional que representa mejor los requerimientos que seleccionó para realizar en esta entrega.

Para realizar esta actividad siga los siguientes pasos.



- a. Para cada tabla de hechos
  - i. (2%) Especificar y justificar la granularidad
  - ii. (5%) Definir los hechos/medidas que contiene. Para cada medida indicar y justificar el tipo de medida (aditiva, semi-aditiva y no aditiva).
- b. Para cada dimensión:
  - i. (2%) Describir los atributos y justificarlos en términos de los requerimientos analíticos propuestos y las fuentes de datos utilizadas.
  - ii. (2%) **Para cada atributo**, si se requiere, especificar el tipo de manejo de historia (1,2, 3,...) de variación lenta (*Slowly Changing Dimension*), y justificar la elección.

 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2021-20</p>	
--	---	---

- c. (4%) Elaborar los modelos dimensionales propuestos. Debe entregar una representación gráfica de los modelos multidimensionales. En el modelo se deben representar nombres de atributos, llaves primarias, llaves foráneas y roles. Evite diagramas ilegibles.
2. (25%) Perfilamiento de datos, creación de la base de datos y procesos ETL. Tomando como base las fuentes de datos proporcionadas y los ejemplos de requerimientos analíticos de la consultoría anterior, realice los siguientes pasos.
  - a. (10%) Describir y **perfiar** las fuentes de datos recibidas y presentar el resultado del análisis. Recuerde explicar las figuras resultado del perfilamiento.
  - b. (10%) **Diseñar el proceso de ETL**. En este punto debe describir las actividades (o tareas del flujo de trabajo en la herramienta utilizada), realizadas en el proceso.
  - c. (5%) **Cargar los datos** y mostrar estadísticas de las tablas con los datos cargados.
3. (40%) **Proponer la arquitectura de solución** de BI a nivel de componentes y **diseñar e implementar** tres tableros de control. Recuerde que un tablero de control está compuesto por más de un indicador de interés para una organización.
  - a. (5%) **Diseñar los tableros de control**, que respondan análisis de interés para el cliente.
  - b. (30%) **Implementar los tableros de control**, utilizando un software especializado como PowerBI, Tableau, Microstrategy, etc.
  - c. (5%) **Proponer la arquitectura de solución** de BI para resolver los análisis realizados hasta este momento en el proyecto.
4. (10%) **Preparar** un video acompañado de una presentación para **la junta directiva de la empresa** donde venda la solución propuesta en esta etapa del proyecto de BI. Esta presentación debe estar en un video de máximo 8 minutos. Este video debe incluir en la primera parte una presentación más ejecutiva sin entrar en detalles técnicos, en la cual, muestre la interacción con los tableros de control construidos y muestre hallazgos obtenidos gracias a su uso. En la segunda parte del video, incluya detalles técnicos de los modelos dimensionales, procesos ETL entre otros elementos que considere son de interés para el área técnica de la empresa.
5. [5%] En términos de un plan de proyecto, **describir las actividades realizadas**, el tiempo asignado y el responsable o responsables de cada actividad, al igual que los entregables de cada fase. De igual manera indicar el estudiante que estuvo a cargo de cada uno de los tableros desarrollados. Adicionalmente, reparta 100 puntos entre los integrantes del grupo de acuerdo con el aporte al desarrollo del proyecto

#### ENTREGA Y EVALUACIÓN

- El proyecto se realiza en grupos de máximo 3 estudiantes.
- El documento a entregar tienen máximo 8 páginas (sin incluir portada, tabla de contenido, ni referencias), a una columna y con letra arial tamaño 12.

 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2021-20</p>	
--	---	---

- Debe entregar las fuentes de los análisis (e.g. si es tableau, los archivos twb) y todo lo necesario para poder ejecutar los tableros de control en una sustentación o en la presentación con el cliente.
- Los entregables del proyecto podrán ser utilizados por Infraestructura Visible para inspirarse en la definición y desarrollo de nuevos proyectos, sin autorización de los autores, pero dando los créditos respectivos.
- La fecha máxima de entrega es el **miércoles 24 de noviembre a las 22:00**.
- La nota del proyecto es individual a pesar de entregar un trabajo a nivel de grupo. Por eso es importante la respuesta que dan en el punto 5.

### ANEXO 1. Descripción de datos

La fuente original que quiere analizar el cliente se encuentra en esta página del DANE:

<https://www.dane.gov.co/index.php/estadisticas-por-tema/salud/nacimientos-y-defunciones/defunciones-no-fetales>

Los invitamos a revisar su descripción y contenido. Sin embargo, para efectos del proyecto, el cliente está organizando un subconjunto de los datos que desea analicen en el proyecto. Esta fuente de datos estará disponible en la sección unificada, dentro del espacio del segundo proyecto del curso.



### ENUNCIADO SEGUNDO PROYECTO: ENTREGA 2

#### OBJETIVOS

Este trabajo tiene como objetivos

- Identificar requerimientos analíticos a partir de oportunidades de negocio, de unas fuentes de datos, y de la revisión de proyectos similares, todo en el contexto de una organización.
- Priorizar los requerimientos analíticos de acuerdo con los criterios definidos para la organización.
- Implementar requerimientos analíticos basados en tableros de control y en tareas de aprendizaje de máquina, que sean de interés para la organización de estudio.
- Desarrollar *data marts* que permitan resolver los requerimientos analíticos seleccionados por el cliente.



#### DESCRIPCIÓN

 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2021-20</p>	
--	---	---

Esta entrega se enfoca en el proceso de identificación de necesidades analíticas, su relación con la estrategia del negocio y la priorización e implementación de los análisis de interés para el cliente para el cual se realiza el proyecto. Estos análisis requieren la implementación de *data marts* que contienen la información que debe ser accedida desde ese repositorio de datos.

La entrega debe incluir como mínimo los siguientes elementos:

1. [15%] **Identificar necesidades analíticas** siguiendo la metodología *Business Life Cycle* [3],[6]– Grupo Kimball (las referencias corresponden a las descritas en el cronograma del curso).
  - a. [10%] A partir de entrevistas realizadas a personas relacionadas con el proyecto que están trabajando, los datos compartidos y la retroalimentación a la consultoría realizada en el semestre 202120, identificar y documentar requerimientos analíticos, utilizando la matriz de requerimientos de negocio, los temas analíticos, análisis requeridos, procesos de negocio, fuentes de datos y datos requeridos (en algunas referencias asociadas al curso, en la matriz de documentación, en cambio de “fuentes de datos y datos” aparece el título “comentario” o solo “datos”). Para este punto descargue el documento de Excel llamado *Plantilla-Entrega1.xlsx*, del enlace asociado a esta entrega. En ese documento de Excel debe llenar el formato de temas analíticos y análisis requeridos/inferidos que se encuentra en la primera hoja. Usted debe categorizar los análisis requeridos en Tableros de control o basados en aprendizaje de máquina. Adicionalmente debe describir y justificar los análisis propuestos. La justificación debe incluir elementos de la entrevista realizada al cliente o de la revisión de proyectos similares a nivel nacional o mundial. La invitación en este punto es a hacer una revisión del estado del arte que en este caso corresponde a revisar proyectos similares relacionados con [analítica sobre muertes](#).
  - b. [5%] Completar la matriz de bus de la bodega de datos a construir con los procesos y dimensiones que identificó para la empresa en la que desarrolla el proyecto. (Hoja 2 del documento de *Excel*).
2. [5%] **Realizar la priorización de los procesos de negocio.**
  - a. [2%] Para revisar la factibilidad de los análisis descritos en el punto anterior, identifique las fuentes de datos que debería consultar (incluya un resumen de lo presentado en la hoja Excel), describa su estructura y revise la disponibilidad de las mismas. Con la información previa defina “factibilidad” y proponga la forma de medirla. Utilice la hoja 3 del documento de Excel.
  - b. [2%] Para analizar el impacto de los análisis descritos, relaciónelos con el plan estratégico de la organización (información disponible en las páginas web de la empresa – en este caso páginas web: <https://www.infraestructuravisible.org/>, o con los casos revisados del estado de arte). Esto es utilizar la hoja 4 del documento



 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2021-20</p>	
--	---	---

Excel para indicar las metas u objetivos del negocio que apoya el análisis. Adicionalmente, proponga un mecanismo para cuantificar el impacto que tienen los análisis para el negocio. Por ejemplo, cuantas metas están relacionadas con el análisis identificado.

- c. [1%] Concluya, justificando, cuál debería ser el primer proceso de negocio que se debe implantar en el proyecto de inteligencia de negocios. Recuerde justificar la decisión.
3. **(10%) Modelar Data Marts:** A partir de los requerimientos analíticos seleccionados por el cliente, proponga el modelo dimensional que representa mejor el proceso. El modelo propuesto puede ser una extensión de los modelos entregados en la etapa anterior del proyecto.

Para realizar esta actividad siga los siguientes pasos.

- a. Para cada tabla de hechos
    - i. (3%) Especificar y justificar la granularidad
    - ii. (3%) Definir los hechos/medidas que contiene. Para cada medida indicar y justificar el tipo de medida (aditiva, semi-aditiva y no aditiva).
  - b. Para cada dimensión:
    - i. (2%) Describir los atributos y justificarlos en términos de los requerimientos analíticos a implementar o de las fuentes de datos a utilizar.
  - c. (2%) Elaborar o ajustar los modelos dimensionales de los procesos. Debe entregar una representación gráfica de los modelos de estrella. En el modelo se deben representar nombres de atributos, llaves primarias, llaves foráneas y roles. Evite diagramas ilegibles.
4. **(32%) Perfilamiento** de datos, creación de la base de datos y procesos ETL.  
Tomando como base las fuentes de datos proporcionadas, otras fuentes externas que integran al proyecto y el modelo propuesto en el punto 3, realice los siguientes pasos.
    - a. (10%) **Describir y perfilar** las fuentes de datos a utilizar y presentar el resultado del análisis. Presentar un análisis de las diferencias encontradas en las fuentes de datos que fueron suministradas por el cliente en las dos entregas del proyecto. Recuerde explicar las figuras resultado del perfilamiento.
    - b. (10%) **Integrar las fuentes de datos suministradas** con el fin de tener la mayor cantidad de información que aporte a los análisis de la organización. En este punto es importante analizar si es posible la integración de otras fuentes de datos que buscaron y encontraron disponibles en otros repositorios (e.g., Cie10 y su estructura -<http://ais.paho.org/classifications/chapters/pdf/volume2.pdf>-, datos población del DANE). En caso de no tener nuevas fuentes para complementar los análisis documéntelo, de lo contrario, justifique su decisión e indique con cual fuente o fuentes nuevas va a trabajar.

 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2021-20</p>	
--	---	---

- c. (12%) Diseñar o ajustar el proceso de ETL. En este punto debe describir las actividades nuevas (o tareas del flujo de trabajo en la herramienta utilizada), que debió incorporar al proceso. Debe mostrar estadísticas de los datos cargados.
5. **(25%)** Diseñar e implementar dos tableros de control nuevos y una tarea de aprendizaje de máquina.
  - a. (12.5%) Diseñar e implementar dos tableros de control, utilizando un software especializado como PowerBI, Microstrategy, Tableau, Excel, etc.
  - b. (12.5%) Implementar una tarea de aprendizaje de máquina, indicando el objetivo de negocio que busca resolver, las tareas, técnicas y algoritmos utilizados, al igual que la parametrización que requieren. Recuerde justificar su decisión.
6. **(10%)** Realizar un video y una presentación en powerpoint para mostrar los resultados de la entrega. Este material será utilizado para comunicar con las personas del negocio los resultados obtenidos con el proyecto. Esta presentación debe estar en un video de máximo 8 minutos. Este video debe incluir en la primera parte una presentación más ejecutiva sin entrar en detalles técnicos e ilustrar los resultados **interactuando** con los análisis construidos ya en la herramienta utilizada para construir los tableros de control o sobre los Jupyter Notebooks. En la segunda parte del video incluir detalles técnicos de los modelos dimensionales, procesos ETL entre otros elementos que considere son de interés para el área técnica de la empresa.
7. **[3%]** Hacer la propuesta de un plan de proyecto para una nueva consultoría en temas de BI para una organización similar a infraestructura segura, para lo cual debe describir los análisis que considera son interesantes de realizar para complementar lo desarrollado en la entrega anterior del proyecto, las actividades propuestas para realizarlas, el tiempo estimado para su desarrollo, los conocimientos o competencias de las personas que deben realizar esa parte del proyecto y los entregables. Adicionalmente, indiquen el estudiante que desarrolló los tableros de control y el modelo analítico y reparta 100 puntos entre los integrantes del grupo de acuerdo con el aporte a esta entrega del proyecto.

## ENTREGA Y EVALUACIÓN

- El proyecto se realiza en grupos de máximo 3 estudiantes.
- El documento a entregar tiene máximo 10 páginas (sin incluir portada, tabla de contenido, ni referencias), a una columna y con letra arial tamaño 12.
- Se debe anexar el documento Excel que contiene el soporte de algunas de las actividades realizadas.
- Los entregables del proyecto podrán ser utilizados por Infraestructura Visible para inspirarse en la definición y desarrollo de nuevos proyectos, sin autorización de los autores, pero dando los créditos respectivos.
- La fecha de entrega es en la semana de exámenes.