**Integrantes:   
  
  
Juan Diego Castellanos Bonilla-201816067**

**Carlos perez- 201718653  
  
  
Ana Sofia Medina- 202416921  
  
  
CASO A TRABAJAR: Análisis sobre la encuesta multipropósito**

El proyecto busca aprovechar los datos de la encuesta multipropósito de los años 2017 y 2021 para guiar a secretarios de salud, gerentes de EPS o IPS, entre otros, a entender y caracterizar comportamientos de enfermedades en la población de Bogotá, con el fin de definir las intervenciones o acciones que se requieren en beneficio de la calidad de vida de esas personas y en el futuro extenderlo a otras ciudades y municipios de Colombia.

El proyecto incluye la identificación e implementación de requerimientos de analítica descriptiva basada en tableros de control, dirigidos a actores específicos de una organización para apoyar sus funciones. Esto debe estar acompañado de la automatización del proceso de consumo de fuentes de datos, transformación y almacenamiento en un data mart, de donde se toman los datos para presentarlos en los tableros de control.

Este desarrollo debe tener en cuenta las recomendaciones dadas por clientes similares, a los entregables en semestres anteriores para no repetir los mismos errores. En particular, visualizaciones solo-gráficas de barras no son suficientes.

El proyecto se realiza para un grupo de médicos y estudiantes de medicina, interesado en conocer factores como los demográficos, localización, vivienda y salud que caracterizan las personas que participan en la encuesta multipropósito. De igual manera, se tendrán en el grupo estudiantes de administración que apoyan en los análisis a realizar. El grupo de profesores y estudiantes de medicina será considerado el cliente de este proyecto. Cada grupo del curso de IN, contará con un grupo de estudiantes de medicina quienes tendrán el rol de cliente, al igual que estudiantes de administración que vienen a apoyar la selección de los temas y análisis a realizar, los análisis con la información presentada en los tableros de control y la gestión del proyecto. Cada grupo debe trabajar de forma conjunta para lograr los diferentes entregables del proyecto. A este último lo llamaremos, grupo extendido.

Este alcance se detalla a continuación.

**ENUNCIADO SEGUNDO PROYECTO**

**OBJETIVOS**

**Este trabajo tiene como objetivo principal**

• Desarrollar un proyecto de inteligencia de negocios centrado en el uso de datos que permitan monitorear métricas e indicadores de interés para un grupo de expertos.

**A nivel de objetivos específicos**

• Identificar requerimientos analíticos que puedan ser resueltos con análisis descriptivos a desplegar en tableros de control.

• Proponer una arquitectura de solución de IN para los análisis a construir.

● Desarrollar *data marts* que permitan resolver requerimientos analíticos de interés para el cliente.

● Diseñar e implementar procesos de ETL.

● Implementar requerimientos analíticos basados en tableros de control que son de interés para los expertos.

Tenga presente que las entregas parciales, descritas en la infografía tendrán un aporte del 20% a la nota total del proyecto, el 70% restante corresponde a los entregables y el 10% a la sustentación. Las entregas parciales también se hacen por BloqueNeón en una actividad específica.

**ENTREGABLES**

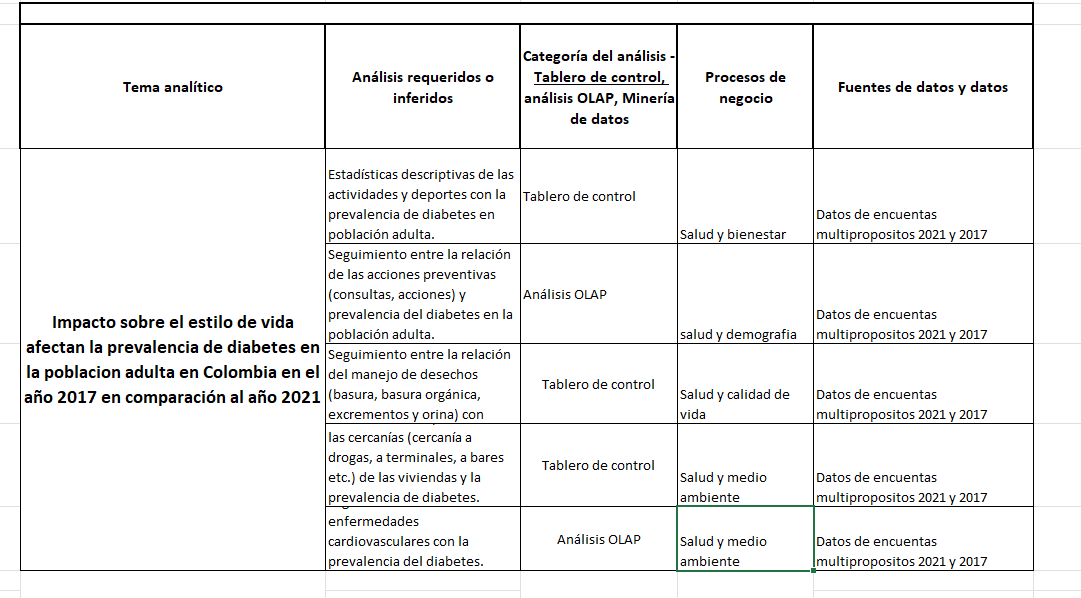
La descripción detallada de las actividades a realizar y de los elementos mínimos a incorporar en el documento final a entregar se presenta a continuación:

**1. [10%] Identificar necesidades analíticas** siguiendo la metodología *Business Life Cycle* [4],[6]–

Grupo Kimball (las referencias corresponden a las descritas en el cronograma del curso).

A partir de entrevistas realizadas al grupo de estudiantes de medicina que trabajan con cada grupo del curso, los datos compartidos y la retroalimentación a consultorías realizadas en semestres anteriores, identificar y documentar 2 requerimientos analíticos, utilizando la matriz de requerimientos de negocio, que contiene 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. Estos análisis son de analítica basados en tableros de control. Adicionalmente debe describir y justificar los análisis propuestos. La justificación debe incluir elementos de la entrevista realizada a los expertos de

medicina 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 p royectos similares relacionados con analítica sobre las temáticas de interés identificadas en las entrevistas iniciales.

**Respuesta: Las necesidades analíticas se Acordaron con los estudiantes de medicina (clientes) y estos están documentados en el xlsx llamado Plantilla-EntregaAnalisisRequeridos YETL.  
**

**2. (14%) Modelar *Data Marts***: A partir de las fuentes de datos anexas a esta entrega y de requerimientos analíticos identificados por ustedes como resultado de la entrevista a realizar al cliente, proponga el modelo multidimensional que representa mejor los tipos de requerimientos que tiene en mente el cliente. En particular usted debe:

a. (7%) Elaborar los modelos dimensionales para realizar los análisis seleccionados y que representen las fuentes compartidas, **al igual que los indicadores necesarios para saber si el análisis es relevante o no en la práctica clínica**. 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.  
  
**(se generará un modelo para cada uno de los análisis y al final se tendrá el modelo completo)**  
**Modelo1:   
Análisis:** **Estadísticas descriptivas de las viviendas, correlación entre material y prevalencia de diabetes en población adulta.**

**Dimensión 1: Vivienda**

**Atributos: vivienda\_id (llave primaria).**

**Dimensión 2: Persona\_Encuestada**

**Atributos:**

**persona\_id (llave primaria)**

**edad**

**genero**

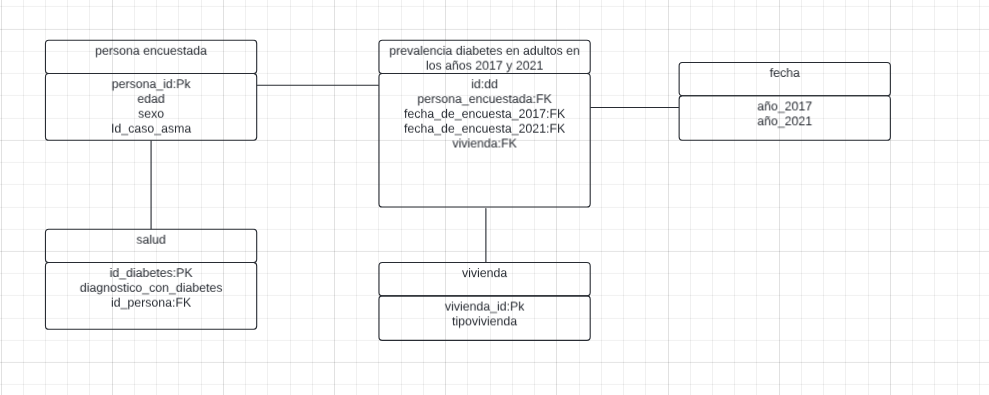
**Dimensión 3: salud   
diagnosticado\_con\_diabetes(Si, No)  
persona\_id (llave foránea a Dimensión Personas)**

**Indicador:**

**1. Porcentaje de prevalencia de diabetes por tipo citas de salud**

**2. Porcentaje de prevalencia de diabetes por estilo de vida.**

**3. Porcentaje de prevalencia de diabetes por tipo vivienda**

**  
  
  
Al mirar el Tablero de control y ver estos porcentajes, se puede saber si la prevalencia del diabetes tiene que ver con el tipo de vivienda, si este tiene goteras y/o por el material predominante de la vivienda.**

**Modelo2:  
Análisis: Seguimiento entre la relación de la ubicación (localidad) y prevalencia del diabetes en la población adulta.  
  
Dimensión 1: Localidad**

**Atributos:**

**localidad\_id (llave primaria)**

**nombre\_localidad  
  
Dimensión 2: Persona\_Encuestada  
  
Atributos:**

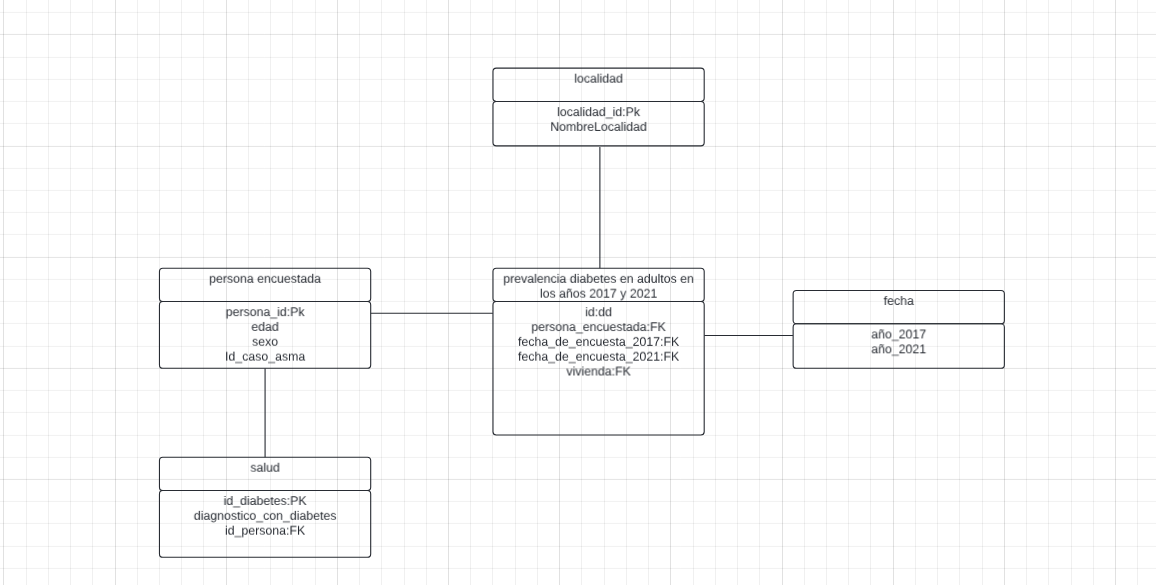
**persona\_id (llave primaria)**

**edad**

**genero**

**localidad\_id (llave foránea a Dimensión Localidad)  
  
Dimensión 3: salud   
diagnosticado\_con\_diabetes (Si, No)**

**persona\_id (llave foránea a Dimensión Personas)  
  
Indicadores:   
1. Porcentaje de Prevalencia de diabetes por Localidad.**

**2.** **Número Total de Casos de diabetes por Localidad.  
**

**Modelo3:   
Análisis: Seguimiento entre la relación del manejo de desechos (basura) con prevalencia de la diabetes en la población adulta.**

**Dimensión 1: Persona\_Encuestada  
Atributos:**

**usuario\_id (llave primaria)**

**Dimensión 2: Manejo de basuras**

**Atributos:**

**Manejo\_de\_Basuras\_id (llave primaria)**

**Manejo\_bueno (si, no)**

**Dimensión 3: Manejo de basuras orgánicas**

**Atributos:**

**Manejo\_de\_Basuras\_organicas\_id (llave primaria)**

**Manejo\_bueno (si, no)   
  
Dimensión 4: presencia animales**

**Atributos:**

**Cercanía\_excrementos\_o\_orina\_id (llave primaria)**

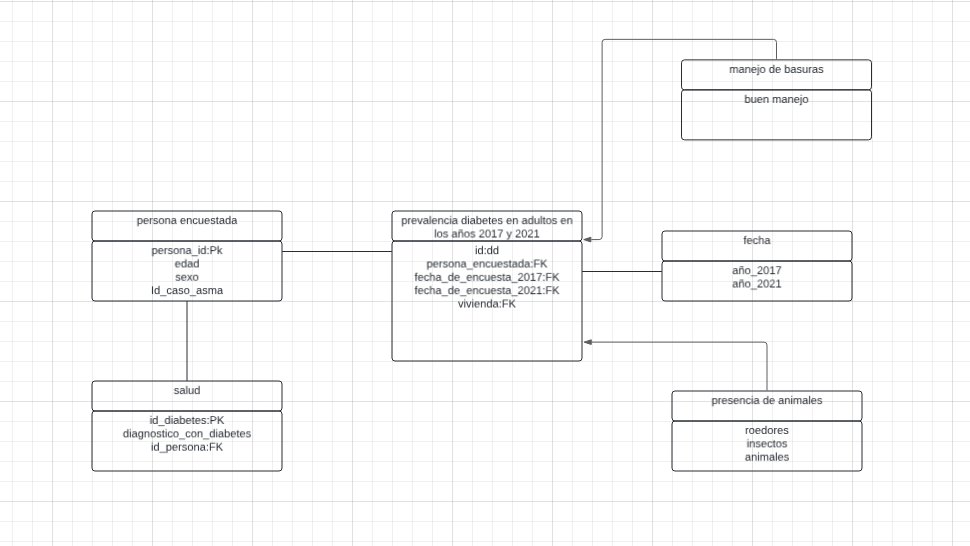
**Esta\_cerca (si, no)   
  
Dimensión 5: Salud**

**Atributos:**

**diagnosticado\_con\_diabetes (Si, No)**

**persona\_id (llave foránea a Dimensión Personas)  
  
Indicador:**

1. **Porcentaje de Prevalencia de diabetes Manejo de basuras:**
2. **Porcentaje de prevalencia de diabetes presencia de animales**
3. **Porcentaje de prevalencia de Diabetes Cercanía de Excremento:**

****

**Modelo4:   
Análisis: Estadística descriptiva entre la cercanía bar de las viviendas y la prevalencia de diabetes.   
  
 Dimensión 1: Usuario  
Atributos:**

**usuario\_id (llave primaria)**

**Dimensión 2: Cercanías de bar**

**Atributos:**

**Cercanía\_bar\_id (llave primaria)**

**Esta\_cerca (si, no)**

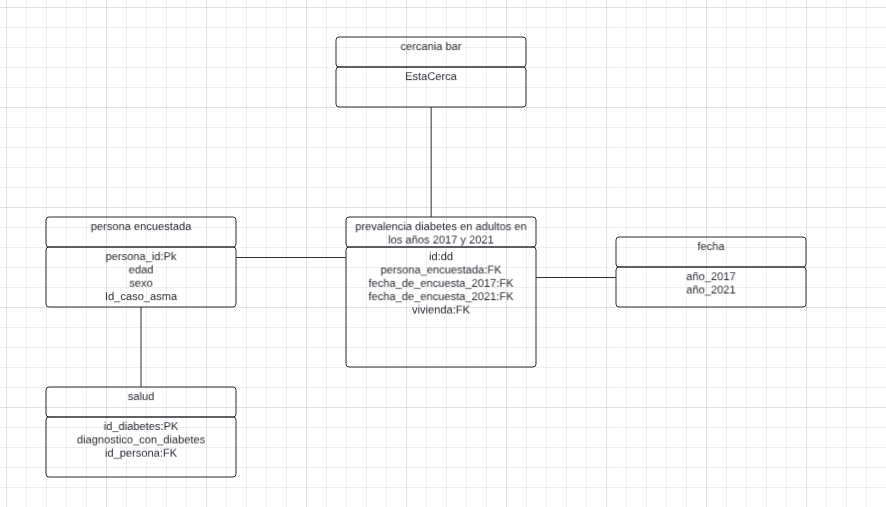
**Dimensión 3: Salud**

**Atributos:**

**diagnosticado\_con\_diabetes (Si, No)**

**persona\_id (llave foránea a Dimensión Personas)  
  
Indicador:**

1. **Porcentaje de Prevalencia de Diabetes por cercanía bar**
2. **Número Total de Casos de Diabetes por cercanía bar**

****

**Modelo5: seguimiento entre la relación escasez de aire con la prevalencia del diabetes.**

**Dimensión 1: escasez de aire**

**Atributos: calidad\_aire\_id (llave primaria),**

**El\_aire\_tiene\_buena\_calidad (si, no)**

**Dimensión 2: usuario**

**Atributos:**

**persona\_id (llave primaria)**

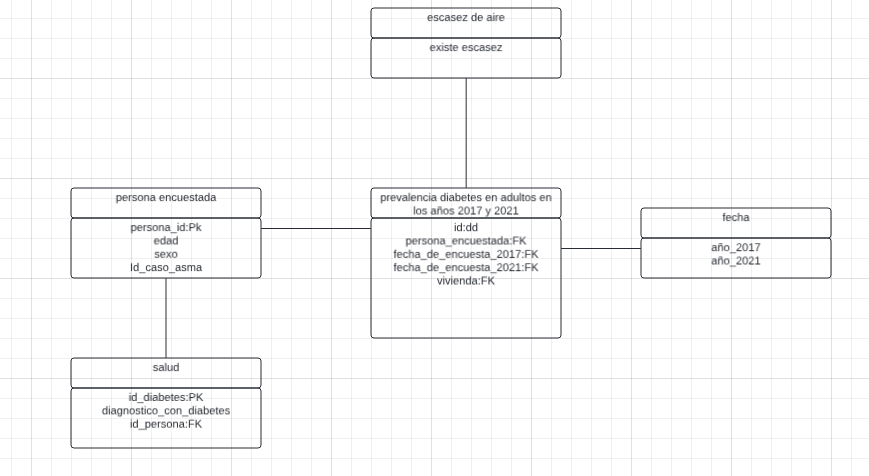
**edad**

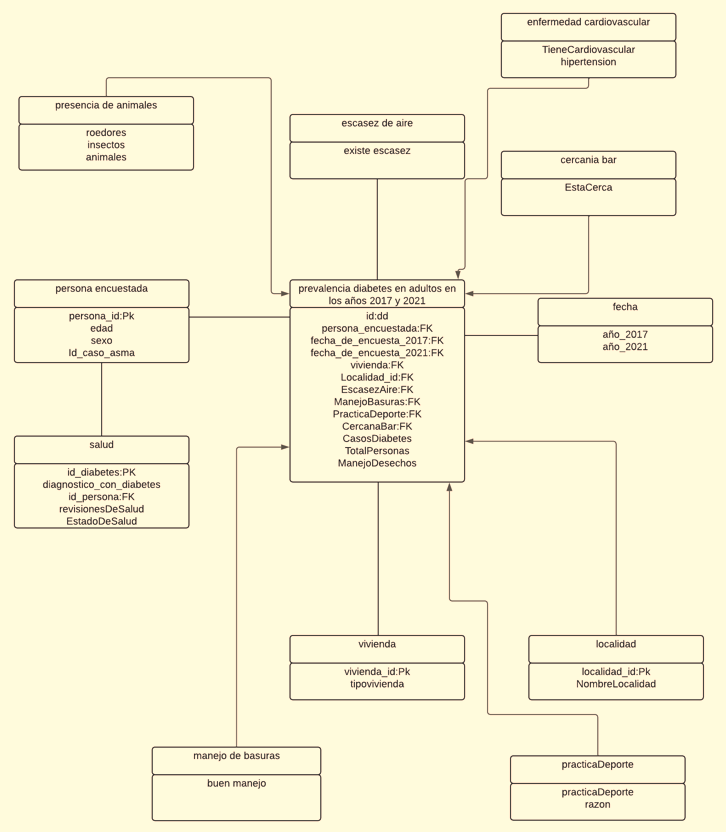
**genero**

**Dimensión 3: salud   
diagnosticado\_con\_diabetes (Si, No)  
persona\_id (llave foránea a Dimensión Personas)**

**Indicador:**

1. **Porcentaje de Prevalencia de Diabetes por calidad de aire.**
2. **Número Total de Casos de Diabetes por calidad de aire.**

  
  
**Modelo definitivo:**

  
**Roles: En cuanto a los roles principales se tiene:**

1. **Dimensión del tiempo: representa la dimensión del tiempo (año) con los casos de Diabetes**
2. **Dimensión de Persona\_Encuestada: representa la persona encuestada, la cual nos da la edad y si esta fue diagnosticada con diabetes o no.**

**Rol secundario:**

1. **Cercanías: Todas las dimensiones de cercanía son un rol secundario, ya que se quiere evidenciar si se tiene diabetes en cercanía o no.**
2. **Manejos de basuras: ya que estas nos sirven para saber si el mal manejo de basuras puede afectar a que el diabetes sea prevalente en las personas.**
3. **Vivienda y localidad: sirve para hacer un contexto analítico y saber si la vivienda y en donde esta ubicada afecta a los adultos en cuanto al diabetes.**
4. **Practica\_deporte: sirve para hacer un contexto analítico y saber si la realizar deporte afecta a los adultos en cuanto al diabetes.**

**al igual que los indicadores necesarios para saber si el análisis es relevante o no en la práctica clínica**.

**Al ver los indicadores solamente por análisis no nos permite saber si se tiene una relación, sin embargo, al juntarlos todos, estamos juntando ubicación, tipo de vivienda, relación a manejos de desechos, cercanías y calidad de aire, lo cual nos permite hacer un buen análisis, con el fin de saber si estos factores ayudan a la prevalencia del Diabetes.**

b. (7%) Justificación del modelo

Para cada tabla de hechos

i. (2%) Especificar y justificar la granularidad

|  |  |  |
| --- | --- | --- |
| Hecho | Granularidad | Justificación |
| Característica vivienda-persona. | Combinación entre una vivienda y un numero de personas que viven en ella. | Este nivel de granularidad permite un análisis detallado de las características de la vivienda y las personas asociadas, proporcionando una vista general de cómo estas características están relacionadas con otras variables como lo son: cantidad de ventilación, humedad. |
| Asociación entre realizar deporte y prevalencia de diabetes. | combinación entre deportes y la prevalencia de diabetes asociada a esa vivienda. | analizar cómo el deporte está correlacionado con la prevalencia de diabetes en un momento determinado |
| Cercanías de vivienda y prevalencia de diabetes. | combinación entre una cercanía específica y la prevalencia de diabetes asociada a esa cercanía | la proximidad a elementos específicos está relacionada con la prevalencia de diabetes en una vivienda en un momento determinado |
| Prevalencia de diabetes y calidad de aire. | Combinación entre la calidad de aire donde vive la persona encuestada y la prevalencia de diabetes asociada a esa calidad. | La calidad de aire en una zona en específico está relacionada con la prevalencia de diabetes en un momento determinado. |

ii. (3%) Definir los hechos/medidas que contiene. Para cada medida indicar y justificar el tipo de medida (aditiva, semi-aditiva y no aditiva).

Para cada dimensión:

**Para responder este punto se realizo la tabla en Excel, la cual se encuentra en la carpeta proyecto 2 y se llama: hechos medidas dimensiones.xlsx**

iii. (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.  
  
**Primero se explicara a que se refiere con el SCD   
1. SCD1: cuando el atributo no cambia o cambia muy raramente.**

**2. SCD2: cuando es necesario rastrear cambios históricos, pero no es necesario mantener múltiples versiones de la dimensión.  
3. SCD3: Cuando el atributo cambia constantemente y es necesario mantener múltiples versiones de la dimensión.**  
 **La tabla para responder este punto se encuentra en Excel, la cual se encuentra en la carpeta proyecto 2 y se llama variación atributos.xlsx**

**3. (28%) Entendimiento de los datos, creación del *Data mart* y proceso ETL.** Perfilamiento de datos y análisis de calidad, creación de la base de datos y procesos ETL. Tomando como base las fuentes de datos proporcionadas, realice los siguientes pasos.

a. (10%) **Entender las fuentes de datos** recibidas y presentar el resultado del análisis

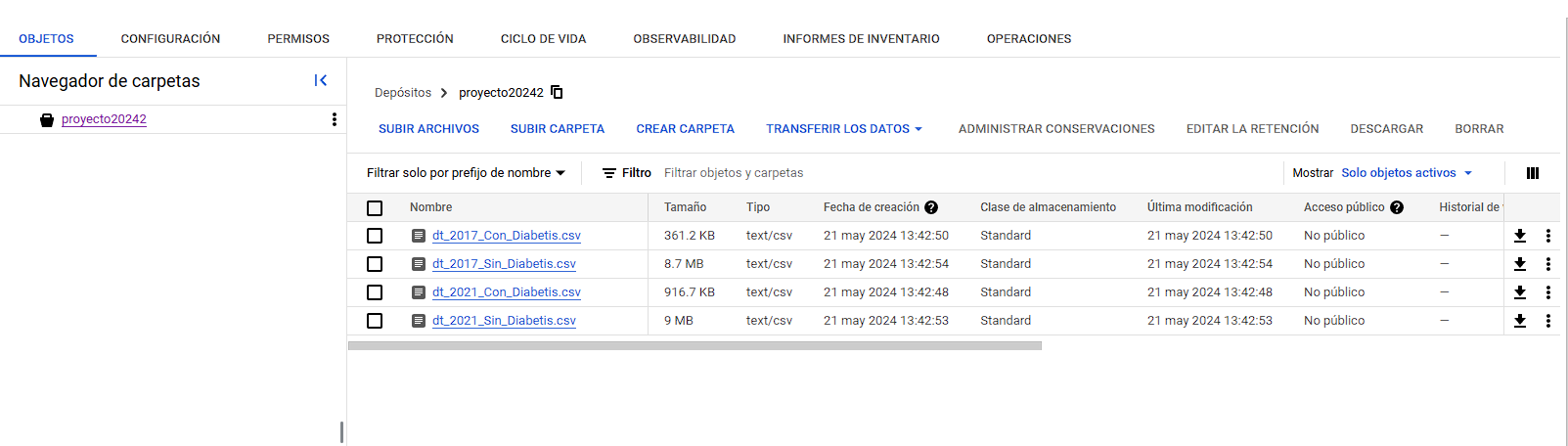
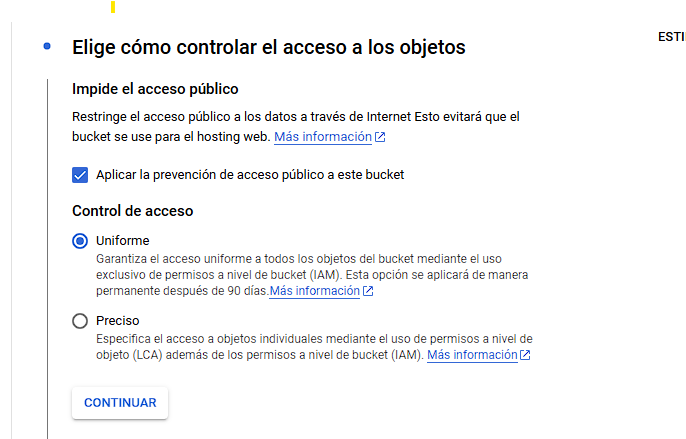
(estadísticos de los datos y análisis de calidad de datos).  
  
 **para entender las fuentes de datos y presentar el resultado del análisis, se hizo en un cuaderno de jupyter el cual se llama datos\_proyecto2 el cual se encuentra en la carpeta Proyecto 2  
  
Se utilizo profile report para ver los estadísticos de los datos, para cada uno de los dataframes.**

b. (18%) **Diseñar e implementar el proceso de ETL**. En este punto debe entregar el diseño del ETL y su implementación. A nivel del diseño utilice la plantilla compartida en excel.  
 **Se descargo el ETL el cual se guardó el json en la carpeta llamada proyecto 2 y se adjunta imagen del proceso ETL en el cual se muestra el diseño e implementación. A nivel de diseño se lleno la plantilla compartida en Excel la cual aparece en el archivo Plantilla-EntregaAnalisiRequeridoYETL.xlsx**   
  
  
  
**Debido a que se quiere hacer una comparación para saber la prevalencia, en este caso. Se realizo un ETL para guardar datos del 2017 y otro del 2021, los cuales hacen lo mismo, a excepción de guardar los datos en sus tablas correspondientes.**

**4. (28%) Proponer la arquitectura de solución** de IN a nivel de componentes e **implementar** dos tableros de control. Recuerde que un tablero de control está compuesto por más de un indicador o métrica de interés para una organización e incluye elementos para facilitar la interacción.

a. (8%) **Proponer la arquitectura de solución** de IN para resolver los análisis realizados

hasta este momento en el proyecto.  
  
**La arquitectura que se propone:  
  
Teniendo en cuenta que toca proponer una arquitectura para un proyecto basado en un sistema de inteligencia de negocios, el cual permita la integración, procesamiento y presentación de datos relevantes por este motivo se mencionara que se hizo en este proyecto.**

1. **Extracción y transformación de datos (ETL):  
     
    Para esta parte de la Arquitectura, lo primero que se hizo fue crear un bucket (simple storage) con el fin de poder almacenar los datos que se van a usar para el estudio y análisis de la diabetes.  
    Una vez se subieron los archivos los cuales fueron tratados previamente en Python, se revisa que el trato que se le hizo fue bueno. Este proceso sirve para guardarlos en un futuro en la base de datos. Como se mencionó se les hizo una unión a los datos del 2017 y también se unieron los datos del 2021. Una vez se hizo esto se guardaron como tablas en el Big Query.**
2. **Almacén de datos (Data Warehouse) Google storage:  
     
    Para almacenar los datos Mediante el ETL se genera una conexión con el Big Query y el Google storage en el cual se genera una conexión y se almacenan los datos en el bucket creado con anterioridad.  
   **
3. **Motor de análisis:  
     
    Gracias a GCP y que se utilizo Big Query el motor de análisis se hace en SQL mediante la creación de las tablas en el bucket y sentencias SQL se puede hacer un análisis de los datos.**
4. **Herramientas de visualización  
     
    Para poder visualizar los datos se utilizo la herramienta de Power Bi en el cual se hicieron dos tableros de control para presentar los datos.**
5. **Seguridad  
     
   En cuanto a la seguridad al bucket se selecciono al momento de la creación el control de acceso uniforme, lo cual implica que cualquier persona que use mi cuenta desde otro servicio va a poder usar los buckets.**

b. (20%) **Implementar los tableros de control**, utilizando un software especializado como PowerBI, [Tableau, Look](http://www.tableau.com/)er, etc. y conectándose a la base de datos donde tiene los datos que representan el modelo multidimensional propuesto.

**5. (10%) Preparar** un video acompañado de una presentación para el grupo de profesores y estudiantes de medicina donde presente la solución propuesta en esta etapa del proyecto de IN. 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. El video debe ser publicado en el padlet asignado para este proyecto: <https://uniandes.padlet.org/mavillam/exposici-n-proyecto-bodegas-de-datos-de-in-202410-iw8wiu8rpzmnkxte>.

**6. [10%]** Evaluación del trabajo en equipo.

a. Autoevaluación en la competencia de resolución de problemas y su capacidad para trabajar eficazmente en equipos diversos, sentido de autonomía y compromiso con sus proyectos.

b. Autoevaluación de la calidad y aporte al proyecto entregado.

c. Evaluación entre miembros del equipo sobre el aporte realizado por cada miembro al proyecto entregado.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Integrante** | **Calificación** | **Justificación** | **Nota por integrante** | **Nota por integrante** |
| **Sofia Medina** | **5** | **Casi todo el trabajo se hizo de manera individual, se presentó a la presentación de los estudiantes de medicina, se encargó de hacer todos los modelos de dimensiones.** | **Juan Castellanos:5** | **Carlos Perez:5** |
| **Carlos Perez** | **5** | **Se encargo de limpiar los datos del trabajo y se encargó de realizar los tableros de control y comunicación de verificación de datos con los estudiantes de medicina y realizo el video.** | **Juan Castellanos:5** | **Sofia Medina:5** |
| **Juan Diego Castellanos** | **5** | **Se encargo de la comunicación con los estudiantes de medicina y realizo el ETL.** | **Carlos perez:5** | **Sofia Medina:5** |

**CRITERIOS DE CALIDAD DE LOS ENTREGABLES**

● La presentación final debe incluir lo realizado en los tableros de control y evidenciar un trabajo colaborativo. En particular deben incluir:

o **Interacción** con los tableros de control para mostrar los análisis descriptivos realizados.

o Mostrar en los tableros de control los indicadore relacionados con la relevancia en la práctica clínica.

o Resultado inferencial (pruebas de hipótesis para mostrar si la evidencia es o no significativa).

o Presentación donde intervienen los estudiantes de medicina y de ingeniería.

● A nivel de tableros de control

o Identidad gráfica.

o Títulos de las gráficas.

o Nombramiento de ejes de las gráficas utilizadas.

o Lenguaje español.

o Variedad de gráficos y filtros para facilitar la interacción y apoyar los análisis a realizar.

● A nivel de proceso de ETL

o Realizarlo en ambiente de nube (utilizar las credenciales que les daremos en el laboratorio 3)

o Utilizar las herramientas sugeridas en el laboratorio de ETL

o Cargar toda la información compartida en las dos fuentes de datos.

**EQUIPO DE TRABAJO Y REUNIONES PROPUESTAS**

● El equipo de trabajo extendido del proyecto está compuesto por estudiantes del curso de inteligencia de negocio (IN) y estudiantes de medicina. Estos últimos tienen el rol de expertos en la temática del proyecto.

● Los estudiantes de medicina serán asignados por los profesores de ellos directamente a los grupos de proyecto del curso de IN.

● El resumen de actividades sugeridas por semana para lograr los entregables es el

siguiente:

o **Semana 11: Lanzamiento del proyecto**. Se describe el alcance del proyecto y se conocen entre los integrantes de los grupos, de los dos cursos involucrados. Deben definir los canales de comunicación y las fechas de las siguientes reuniones. El lanzamiento es el viernes 27 de octubre, en el horario de 2:00 a 4:00 p.m. en el centro de prácticas, sede de la calle 116 de la Universidad de los Andes, en el salón CP 101. Se requiere, la participación de al menos un estudiante de cada grupo del curso de IN en el lanzamiento. Esa participación podrá realizarse de forma virtual, en caso de tener actividades posteriores en la sede de Germania.

* **Semana 13: Revisión y socialización de las fuentes de datos compartidas**. Los estudiantes de IN deben revisar las fuentes de datos para presentárselas de forma concreta a los estudiantes de medicina. Esta revisión corresponde al resultado de la etapa de entendimiento de los datos. Las dudas que surjan al revisar las fuentes de datos podrán registrarlas en el siguiente enlace para que los expertos en esos datos puedan darnos respuesta: https://docs.google.com/document/d/102Kpl\_SAEpPyC0d-hb058ZbNK8F65sSh/edit?pli=1

o **Semana 14: Definir los análisis requeridos.** A partir de un trabajo conjunto con los estudiantes de medicina, utilizando el análisis de las fuentes de datos compartidas e ideas de artículos o proyectos similares en Colombia u otros países, al igual que las entrevistas realizadas. De acuerdo con la sección del curso de IN a la que pertenezcan los miembros de un grupo, esta actividad podrá realizarse en los siguientes horarios, en los cuales los estudiantes de medicina podrán participar de forma presencial o virtual y se contará con el apoyo de profesores del curso de IN.

▪ S1. Viernes 10 de noviembre. 6:30-7:50 a.m. Salón RGD\_311.

▪ S2. Jueves 9 de noviembre, 8:00 – 9:20 a.m. Salón LL 107.

▪ S3. Jueves 9 de noviembre, 9:30 – 10:50 a.m. Salón SD\_801.

▪ Cada grupo es responsable de definir el canal que va a utilizar para trabajar en estas sesiones.

o **Semana 15: Primeros bocetos funcionales**. Se presentan los primeros bocetos de los tableros de control que se están desarrollando para validar con el grupo de estudiantes de medicina.

▪ S1. Miércoles 15 de noviembre. 6:30-7:50 a.m. Salón RGD\_311.

▪ S2. Lunes 13 de noviembre, 8:00 – 9:20 a.m. Salón LL 107.

▪ S3. Martes 14 de noviembre, 9:30 – 10:50 a.m. Salón SD\_801.

▪ Cada grupo es responsable de definir el canal que va a utilizar para trabajar en estas sesiones.

o **Semana 16. Cierre del proyecto** en exposición en feria de proyectos. Esta feria se realizará en la sede centro de prácticas de Uniandes, calle 116, en el horario de

12:00 a 4:00pm. Cada grupo debe organizar la presentación para ser realizada en máximo 5 minutos, esta presentación se debe repetir varias veces a los diferentes jurados. Es importante que se organicen entre ustedes para garantizar la participación de al menos un integrante del grupo de IN de forma presencial.

**ENTREGA Y EVALUACIÓN**

- El proyecto se realiza en grupos de mínimo 2 y máximo 3 estudiantes de IN.

- El documento a entregar tiene máximo 8 páginas (sin incluir portada, tabla de contenido, ni referencias), a una columna y con letra arial tamaño 12.

- 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 el grupo de profesores y estudiantes de medicina y administración para inspirarse en la definición y desarrollo de nuevos proyectos, dando los créditos respectivos, sin autorización de los autores.

- La fecha máxima de entrega completa es el jueves 23 de mayo a las 20:00.

- La fecha de sustentación se hará el viernes 24 de mayo en la feria de cierre en la franja de 12:00 a 4:00p.m.

- 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 6 “Evaluación del trabajo en equipo”.

**1. ANEXO 1. Descripción de datos**

Las fuentes originales relacionadas con los indicadores que quiere analizar el cliente se encuentran en esta página del DANE, en los siguientes enlaces:

• 2017 - <https://microdatos.dane.gov.co//catalog/565/datafile/F36>.

• 2021 - <https://microdatos.dane.gov.co/index.php/catalog/743>

Grabación de la explicación sobre la encuesta multipropósito: <https://drive.google.com/file/d/18iHUoGycLidV9jZmlnGHDM2uN9dYPWc7/view>

Los invitamos a revisar su descripción y contenido. Sin embargo, para evitar problemas por cambio en los datos, las fuentes a utilizar están disponibles en la sección unificada, dentro del espacio del segundo proyecto del curso.