



DEPARTAMENTO DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

Inteligencia de negocios 202220 – Laboratorio 4

PROFESORA: Haydemar Núñez

Nombres	Apellidos	Código	Login
María Sofía	Álvarez López	201729031	ms.alvarezl
Brenda Catalina	Barahona Pinilla	201812721	bc.barahona
Álvaro Daniel	Plata Márquez	201820098	ad.plata

Proyecto 1 – Etapa 2: Automatización analítica de textos

El objetivo de este proyecto es analizar indicadores socioeconómicos y dimensiones municipales de Colombia. El proyecto busca aprovechar los datos abiertos 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

1. Identificación de necesidades analíticas

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 semestres anteriores, se identificaron y documentaron algunos requerimientos analíticos, utilizando la matriz de requerimientos de negocio, los temas analíticos, análisis requeridos, procesos de negocio, fuentes de datos y datos requeridos siguiendo la metodología Kimball.

A continuación, se presentan los temas analíticos propuestos, con sus respectivos análisis requeridos/inferidos y las justificaciones respectivas, que se pensaron para este proyecto. Para encontrar la tabla completa, remítase al archivo EntregaAnálisisRequeridos.xlsx

Tema analítico: Conflicto armado.

Tema analítico	Análisis requeridos o inferidos	Categoría del análisis - Tablero de control, análisis OLAP, Minería de datos	Procesos de negocio	Fuentes de datos y datos
Conflicto armado	Análisis de homicidios * año * departamento * causa * municipio * sexo	Análisis OLAP	Registro de afectaciones en el conflicto armado	Excel ConflictoArmado
	Análisis de número de personas secuestradas * año * departamento * municipio	Análisis OLAP	Registro de demografía y población	Excel DemografíaYpoblacion
	Visualización de las afectaciones del conflicto con base en el análisis 1 y 2	Tablero de Control	Registro de afectaciones en el conflicto armado y Registro de demografía y población	Datamart sobre el conflicto armado

Tema analítico: Educación.

Educación	Análisis de cobertura neta * año * mes * departamento * municipio	Análisis OLAP	Registro de datos sobre educación	Excel Educación
	Visualización del análisis indicadores educativos con base en el análisis 1	Tablero de Control	Registro de datos sobre educación	Datamart sobre la educacion

Tema analítico: Desempleo.

Desempleo	Análisis del indicador de desempleo * año * municipio	Análisis OLAP	Registro de mediciones de desempleo	Excel MedicionedeDesempleoDepartamental
	Análisis del indicador de desempleo * año * departamento	Análisis OLAP	Registro de mediciones de desempleo	Excel MedicionedeDesempleoMunicipal
	Visualización de análisis indicadores de desempleo municipal	Tablero de Control	Registro de mediciones de desempleo	Datamart sobre el conflicto armado
	Visualización de análisis indicadores de desempleo para departamentos	Tablero de Control	Registro de mediciones de desempleo	Datamart sobre desempleo

Tema analítico: Salud.

Salud	Análisis del nivel de satisfacción del sistema de salud * año * departamento * municipio	Análisis OLAP	Registro de datos sobre la salud	Excel Salud
	Análisis de población afiliada a regímenes * año * departamento * municipio * tipo de régimen	Análisis OLAP	Registro de demografía y población	Excel DemografiaYPoblacion
	Visualización de análisis indicadores de salud con base en el análisis 1 y 2	Tablero de Control	Registro de datos sobre la salud y Registro de demografía y población	Datamart sobre salud

Tema analítico: Vivienda y servicios públicos

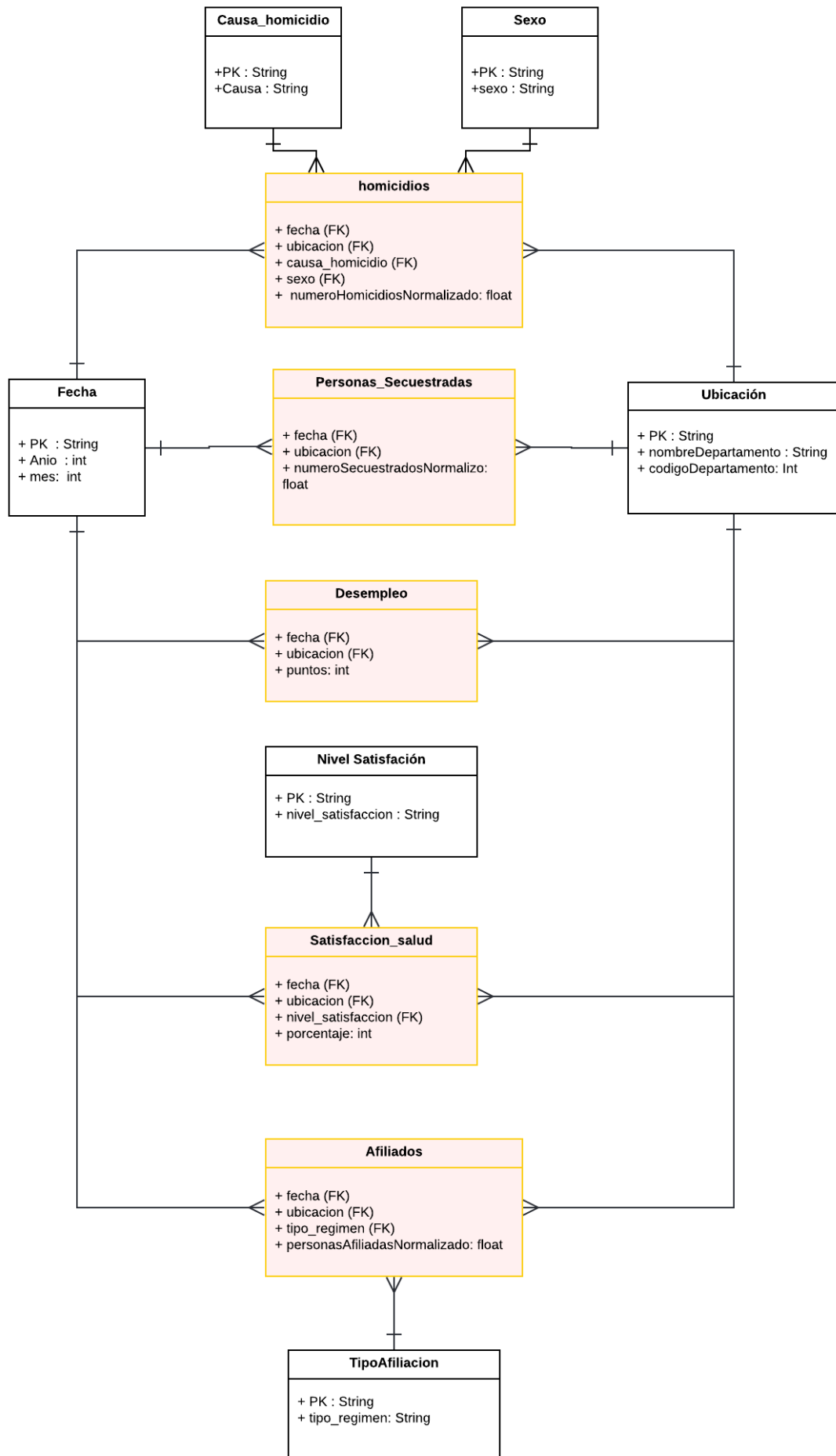
Vivienda y servicios públicos	Análisis del indicador de cobertura * año * departamento * municipio * tipo de servicio	Análisis OLAP	Registro de vivienda y servicios públicos	Excel VivienDayServiciosPublicos
	Análisis de penetración de banda ancha * año * departamento * municipio	Análisis OLAP	Registro de demografía y población	Excel DemografiaYPoblacion
	Visualización de análisis indicadores de salud con base en el análisis 1 y 2	Tablero de Control	Registro de vivienda y servicios públicos y Registro de demografía y población	Datamart sobre Vivienda y servicios públicos

2. Modelado de *DataMarts*

A partir de las fuentes de datos anexas a esta entrega y los requerimientos analíticos identificados en el literal anterior, se realizó un modelo dimensional que representa los tipos de requerimientos del cliente.

Con esto en mente, lo primero que se hizo fue elaborar el modelo multidimensional, con los nombres de atributos, llaves primarias, llaves

foráneas y roles claramente definidos. El modelo obtenido fue el siguiente:



Para la creación de este modelo seguimos el siguiente proceso para la creación del modelo dimensional:

1. Seleccionar el proceso de negocio: para los requerimientos que deseamos realizar, deseamos colocar gran enfoque en los siguientes procesos de negocio: registro de Homicidios, registro de personas secuestradas, registro de la proporción de satisfacción de salud, registro de afiliados al sistema de salud y registro de la medición de desempleo departamental.

2. Seleccionar granularidad:

En general, el nivel de granularidad de la mayoría de procesos es parecido. Escogimos este nivel de granularidad para satisfacer los requerimientos de análisis que identificamos anteriormente. Al analizar los datos nos damos cuenta que en estos podemos realizar una distinción de algunos indicadores según la ubicación y la temporalidad.

- a. Registro de homicidios:
Nivel de granularidad: a nivel de especificación de temporalidad, demografía, ubicación y causa
- b. Registro de personas secuestradas:
Nivel de granularidad: A nivel de especificación de temporalidad y de ubicación
- c. Registro de desempleo:
Nivel de granularidad: A nivel de especificación de temporalidad y de ubicación
- d. Registro de Satisfacción de salud:
Nivel de granularidad: A nivel de especificación de temporalidad, de ubicación y nivel de satisfacción
- e. Registro de afiliado:
Nivel de granularidad: A nivel de especificación de temporalidad , de ubicación y tipo de afiliación.

3. Identificar dimensiones: Para todos los procesos mencionados anteriormente decidimos realizar las vistas por fecha (mes y/o año) y lugar (departamento). Para los procesos de registro de homicidios, registro de la proporción de satisfacción de salud y registro de afiliados al sistema de salud se colocaron mas dimensiones:

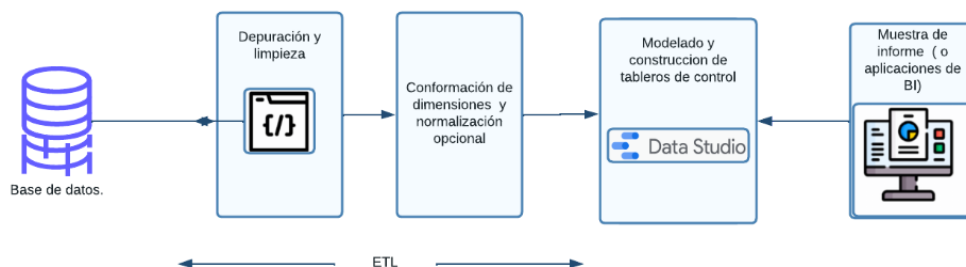
- a. Proceso de registro de homicidios: Dimensión de Causa de homicidios y Dimensión Sexo
- b. Proceso de registro de la proporción de satisfacción de salud: Dimensión nivel de satisfacción (Muy satisfecho, Muy insatisfecho, satisfecho, etc)
- c. Proceso de registro de afiliados al sistema de salud: Dimensión tipo régimen, el cual especifica a que sistema de salud se hace referencia, por ejemplo Afiliado al régimen contributivo, afiliado al régimen subsidiado, afiliado a regímenes especiales y afiliados al SGSSS

4. Identificar los hechos:

- a. Registro de homicidios:

- Porcentaje de homicidios (medida no aditiva)
- b. Registro de personas secuestradas:
Porcentaje de personas secuestradas (medida no aditiva)
- c. Registro de desempleo:
Porcentaje de personas que estan desempleadas (medida no aditiva)
- d. Registro de Satisfaccion de salud:
Porcentaje de personas que tienen un determinado nivel de satisfacción con el sistema de salud: (medida no aditiva)
- e. Registro de afiliado:
Porcentaje de personas afiliadas al regimen de salud (medida no aditiva)

Proponer arquitectura de solución:



Implementar los tableros de control :

Puede encontrar los tableros de control en el siguiente link <https://datastudio.google.com/reporting/c775cee7-acb5-427e-98e3-51ef65a4c058>

El tablero de control se realizó en GoogleDataStudio. En este puede ver en la parte izquierda que estan las opciones para realizar los filtros correspondientes. En sí,aca hay dos tableros de control, en el primero se muestran los hechos del proceso de registro del conflicto armado y en el segundo se muestran los hechos del proceso de registro de Salud. A continuación, se muestra una imagen del tablero de control. Para pasar al segundo tablero de control, debe darle clic en el numero 2 de la barra vertical izquierda.



Descripción de las actividades realizadas:

- Identificación de las necesidades analíticas : Brenda, Alvaro, Sofía
- Modelado del datamarts: Brenda, Alvaro, Sofía
- Entender las fuentes de datos: Brenda, Alvaro, Sofía (5 horas)
- Diseñar e implementar el proceso de ETL: Brenda, Alvaro, Sofía (2 horas)
- Proponer arquitectura de decisión: Brenda, Alvaro (1 hora)
- Implementar los tableros de control : Brenda, Alvaro (6 horas)
- Preparacion del video : Brenda, Sofía, Alvaro (1:15 horas)
- Realización del video : Sofía (1 hora)
- Realización del informe: Brenda, Sofía, Alvaro (2 horas)

Repartición de puntos:

Si tuviéramos que repartir 100 puntos entre los integrantes del grupo, repartiríamos 33.33 a cada uno