

Nicolás Cáceres Sala
Pablo Márquez
Oriana Vivas

09-10-2021

Análisis – Atención a la Salud



Proyecto final para **CODER HOUSE**

ÍNDICE

Tabla de contenido

1.	INTRODUCCIÓN	3
2.	TABLA DE VERSIONADO	3
3.	OBJETIVO	4
4.	HERRAMIENTAS TECNOLÓGICAS IMPLEMENTADAS.....	4
5.	ALCANCE.....	5
6.	USUARIO FINAL Y NIVEL DE APLICACIÓN DE ANÁLISIS	5
7.	SEGMENTACIONES DEFINIDAS	6
8.	DIAGRAMA ENTIDAD-RELACIÓN	7
9.	TABLAS, ATRIBUTOS Y DEFINICIÓN DE TIPO DE DATOS	8
10.	IMAGEN DE MODELO RELACIONAL EN POWER BI	11
11.	TRANSFORMACIÓN DE DATOS.....	12
12.	ANÁLISIS FUNCIONAL DE TABLERO	15
13.	MEDIDAS CALCULADAS	19
14.	FUTURAS LÍNEAS	24

1. INTRODUCCIÓN

El presente trabajo es un proyecto creado para el curso de Data Analytics de Coderhouse.

Consta de la base de datos de la temática a tratar, la documentación y un tablero realizado en la herramienta Power BI.

La idea es trabajar sobre los datos de un servicio de atención a la salud del departamento de Montevideo, el cual cuenta con una red de clínicas de atención médica en diversas especialidades y que atiende a miles de pacientes a lo largo del año.

Se dispone de los datos de unas 66.500 atenciones de enero a junio del año 2021, en 21 clínicas distribuidas a lo largo de todos los municipios del departamento de Montevideo, con 187 profesionales que atienden en 10 especialidades.

2. TABLA DE VERSIONADO

Versión	Características
0.1	Se carga base de datos a Power BI, se generan transformaciones y el Modelo Relacional.
0.2	Se generan tabla calendario y tabla horas.
0.3	Se crean las solapas Global, Clínicas, Especialidad y Ausentismo. Se generan las medidas en DAX: numero_atenciones, numero_usuarios_unicos, dias_totales_consulta, porcentaje_de_inasistencias y promedio_atenciones_por_dia.
0.4	Se comienza a definir estilo, paleta de colores e iconos.
0.5	Se definen segmentaciones.
0.6	Se crea botonera para moverse entre solapas, generando marcadores para ocultar la misma.
0.7	Se crean las medidas DAX: agenda_coordinada, agenda_espontanea, agendados_coordinados_totales, agenda_domiciliaria, atenciones_sexo_femenino, atenciones_sexo_masculino, porcentaje_sexo_femenino y porcentaje_sexo_masculino.

0.8	Se definen mejoras en el aspecto visual, agregando iconos y cambios menores en paleta de colores para mejorar la experiencia del usuario.
0.9	Se realizan cambios en objetos visuales en solapas Global y Clínicas.
1	Se realizan cambios en objetos visuales en solapas Ausentismo y se definen objetos visuales a utilizar en solapa Especialidades. Primera versión funcional.
1.1	Se agregan medidas DAX comparacion_anio_2020, %variacion_2021_vs_2020 y objetos visuales para visualizar dichas medidas.
1.2	Se crea solapa Proyección Saturación con la incorporación de un parámetro.

3. OBJETIVO

El objetivo del proyecto consiste en realizar un análisis completo de los datos para conocer la situación actual del servicio y generar insumos que permitan la toma de decisiones por parte de los gestores del sistema de salud. Decisiones con respecto a, por ejemplo, que clínicas y especialidades priorizar, dónde reforzar los recursos humanos, tomar medidas que mejoren la inasistencia a consulta, generar campañas de publicidad para concientizar y prevenir enfermedades comunes en la población.

4. HERRAMIENTAS TECNOLÓGICAS IMPLEMENTADAS

- ❖ La base de datos fue definida para trabajar con el sistema de gestión de base de datos “Microsoft SQL Server 2019”.
- ❖ El diagrama entidad relación fue realizado con la aplicación online “diagrams.net” en su versión 15.3.4.
- ❖ Para la documentación se utilizó “Word Microsoft Office 2019”.
- ❖ Y finalmente para el tablero se utilizó la herramienta de análisis y visualización de datos “Microsoft Power BI” versión 2.97.921.

5. ALCANCE

Se requiere saber la situación actual de las 21 clínicas con relación a:

- ❖ Cantidad de atenciones, discriminadas a su vez por rango etario y sexo.
- ❖ Cantidad de usuarios.
- ❖ Cómo varía la atención a lo largo de los meses, por clínica.
- ❖ Comparación de atenciones con el semestre del año anterior (2020).
- ❖ Clínicas y especialidades más concurridas.
- ❖ Diagnósticos más frecuentes por cada especialidad y clínica.
- ❖ El ausentismo a consulta por clínica, por especialidad, por consultorio y por profesional.

Algunas limitaciones:

- ❖ Con relación a las inasistencias, se cuenta con la información del parte diario del profesional, dónde aparece el total de usuarios que faltaron a su consulta, pero no se sabe que usuarios en particular faltaron a la misma. Por tanto, no se puede cruzar inasistencias con el rango etario ni con el sexo del usuario.
- ❖ Con respecto año 2020, solo se cuenta con la información de la fecha de atención y la clínica, por tanto, solo se puede hacer comparaciones del semestre actual contra esas dos categorías.

6. USUARIO FINAL Y NIVEL DE APLICACIÓN DE ANÁLISIS

El siguiente análisis tiene como fin a quien brindarle la información, a la directora del servicio de salud, a los coordinadores de las clínicas y a investigadores médicos del servicio.

El nivel de aplicación del dashboard será tanto estratégico como táctico.

7. SEGMENTACIONES DEFINIDAS

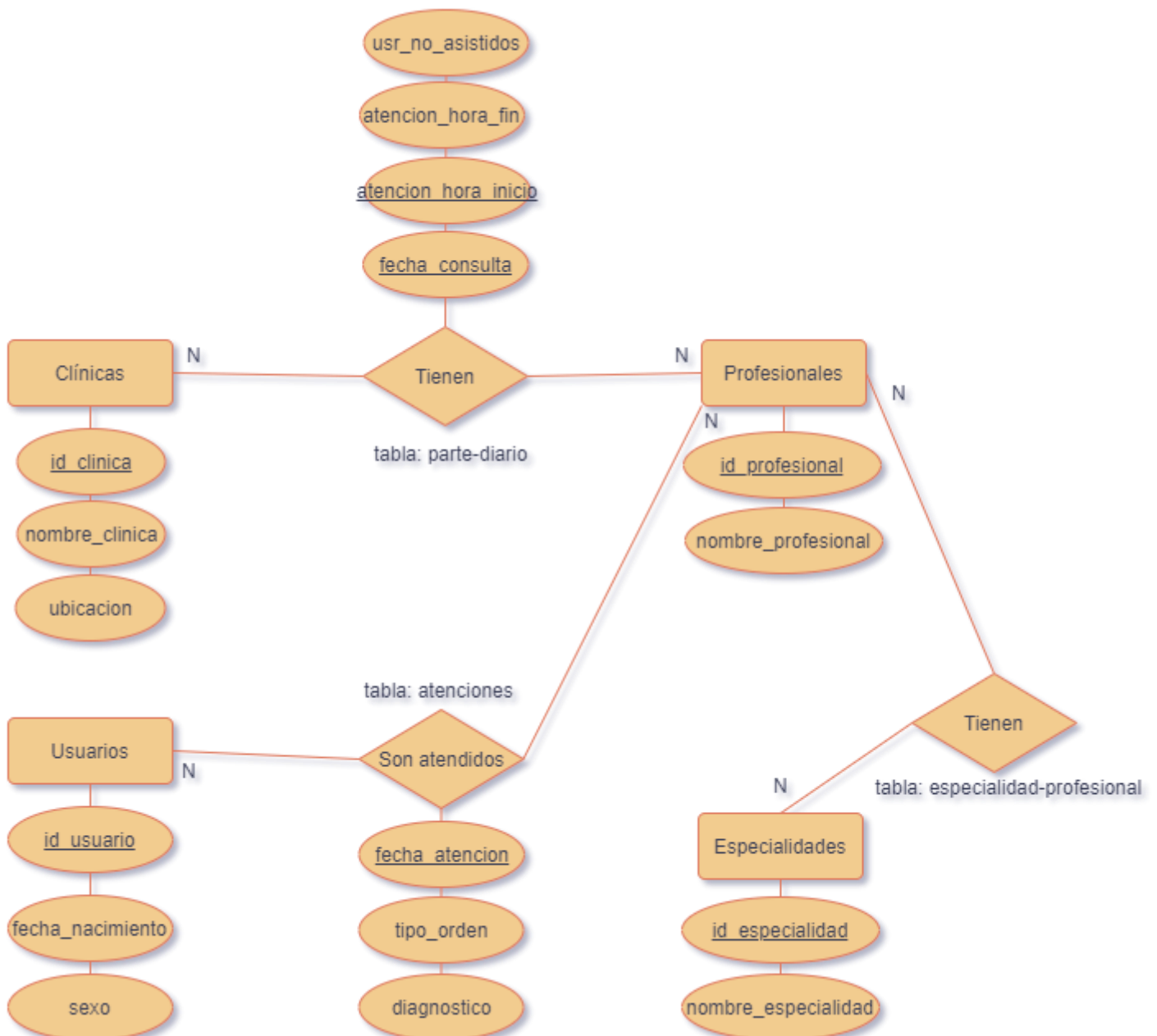
Las segmentaciones que se generaron son:

- ❖ Mes – Permite seleccionar entre los primeros 6 meses del año 2021.
- ❖ Clínicas – Permite seleccionar cada una de las clínicas.
- ❖ Especialidades – Permite seleccionar cada una de las especialidades.
- ❖ Profesionales – Permite seleccionar cada uno de los profesionales.
- ❖ Rango Etario – De acuerdo a la edad se generaron los siguientes rangos etarios: (Menores a 1), (1 a 4), (5 a 14), (15 a 19), (20 a 44), (45 a 64), (65 a 74), (Mayores a 74). Se eligieron estos rangos porque son los utilizados por el Ministerio de Salud Pública del Uruguay.
- ❖ Sexo – Permite seleccionar entre sexo Femenino o Masculino de los usuarios, pero no de los profesionales.

8. DIAGRAMA ENTIDAD-RELACIÓN

Enlace al diagrama online: <https://diagra-entidad-relacion>

Enlace a la base de datos del proyecto: <https://dataset-atencion-a-la-salud>



9. TABLAS, ATRIBUTOS Y DEFINICIÓN DE TIPO DE DATOS

A continuación, la definición de tablas del proyecto, con sus respectivas claves primarias (**PK**), claves foráneas (**FK**) y restantes atributos. Los tipos de datos fueron definidos para trabajar con Microsoft SQL Server.

Tabla: Clínicas

Tipo de Clave	Campo	Tipo de Campo
PK	id_clinica	tinyint
	nombre_clinica	nvarchar(25)
	ubicacion	geography

Contiene un listado de las clínicas.

Tabla: Parte Diario

Tipo de Clave	Campo	Tipo de Campo
PK, FK	id_profesional	smallint
PK	fecha_consulta	date
PK	atencion_hora_inicio	time
	atencion_hora_fin	time
	usr_no_asistidos	tinyint
FK	id_clinica	tinyint
FK	id_especialidad	tinyint

Cada fila de la tabla corresponde al parte diario generado por cada profesional en una fecha y horario determinados.

El campo “usr_no_asistidos” se refiere a la suma de usuarios que se ausentaron a la referida consulta del profesional.

Tabla: Profesionales

Tipo de Clave	Campo	Tipo de Campo
PK	id_profesional	smallint
	nombre_profesional	nvarchar(40)

Contiene un listado de los profesionales.

Tabla: Especialidad-Profesional

Tipo de Clave	Campo	Tipo de Campo
PK, FK	id_profesional	smallint
PK, FK	id_especialidad	tinyint

Cada fila corresponde a la asociación de un profesional a una especialidad. Hay profesionales que tienen asociada más de una especialidad.

Tabla: Especialidades

Tipo de Clave	Campo	Tipo de Campo
PK	id_especialidad	tinyint
	nombre_especialidad	nvarchar(25)

Contiene un listado de las especialidades.

Tabla: Atenciones

Tipo de Clave	Campo	Tipo de Campo
PK	fecha_atencion	datetime
PK, FK	id_usuario	int
PK, FK	id_profesional	smallint
FK	id_clinica	tinyint
FK	id_especialidad	tinyint
	tipo_orden	nvarchar(25)
	diagnostico	nvarchar(200)

Cada fila de la tabla se corresponde a la atención de un usuario en particular en una fecha y hora específicos. Se repiten los usuarios en las filas ya que los mismos pueden haberse atendido en varias oportunidades. El campo “tipo_orden” toma tres valores diferentes con la siguiente descripción:

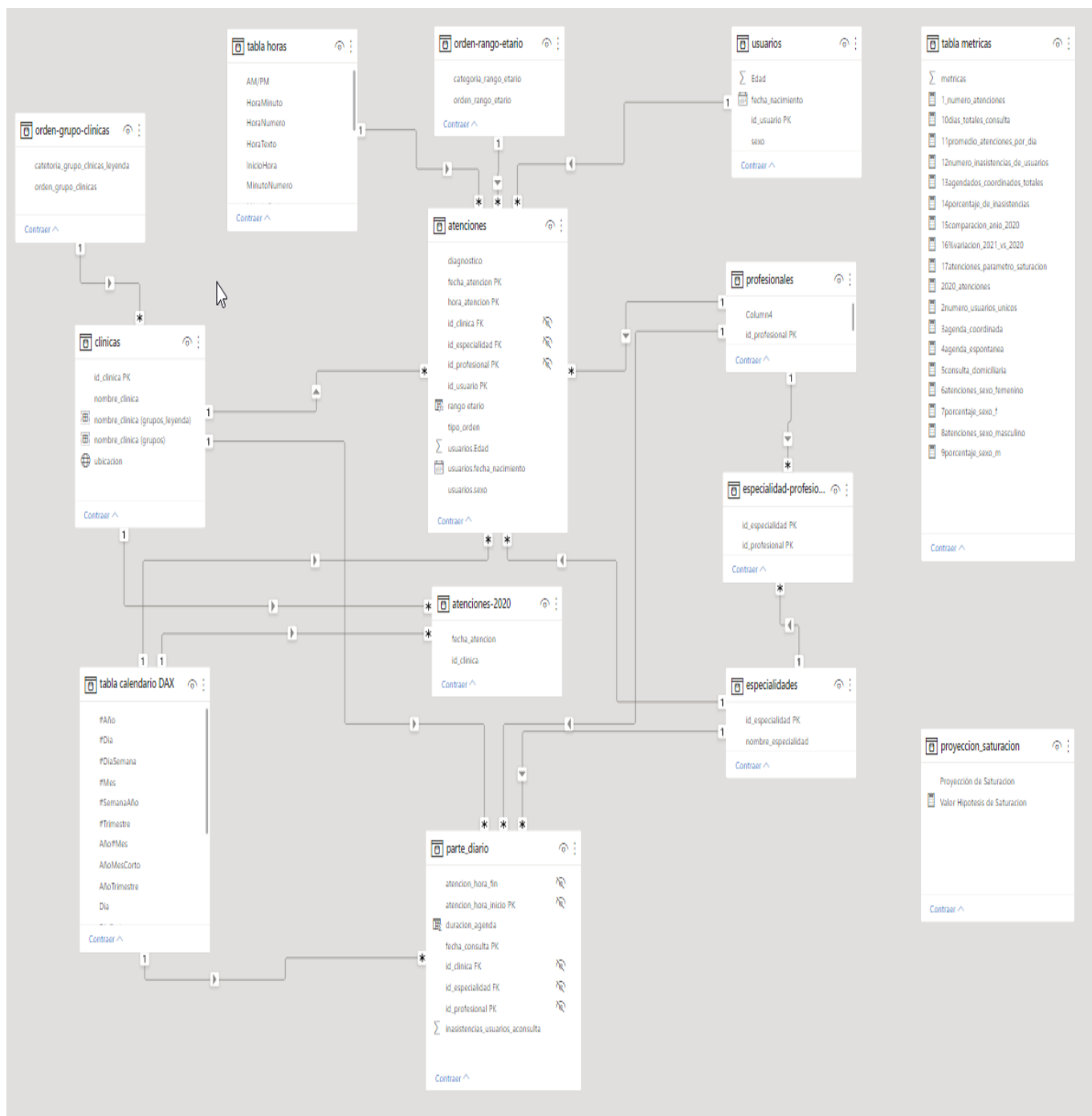
- i. **“Coordinación”**, que son las atenciones que estaban agendadas en el sistema con anterioridad,
- ii. **“Agenda Espontanea”**, que son las atenciones que se dan en el momento sin agenda previa y
- iii. **“Consulta Domiciliaria”**, que son las atenciones que se realizan a domicilio.

Tabla: Usuarios

Tipo de Clave	Campo	Tipo de Campo
PK	id_usuario	int
	fecha_nacimiento	date
	sexo	varchar(1)

Contiene un listado de usuarios únicos.

10. IMAGEN DEL MODELO RELACIONAL EN POWER BI



Nota: las tablas detalladas en el punto anterior no se vuelven a describir. Para ver el detalle de dichas tablas diríjase al punto [9.TABLAS, ATRIBUTOS Y DEFINICIÓN DE TIPO DE DATOS.](#)

- 1) Tabla **“atenciones-2020”**. Se realizó el agregado de la tabla con las atenciones del año 2020 para comparar el semestre actual con el semestre del año anterior. Esta tabla solo cuenta con la fecha y clínica de la atención.
- 2) **“tabla_metricas”**. Tabla con todas las medidas realizadas en el proyecto.
- 3) **“tabla_calendario_DAX”**. Se crea la tabla calendario en lenguaje DAX tomando como fecha de inicio la tabla de “atenciones-2020” y como fecha final la tabla “parte_diario”.
 - i. `CALENDAR (min('atenciones-2020'[fecha_atencion]),
max(parte_diario[fecha_consulta PK])`
- 4) **“tabla_horas”**. Se crea la tabla horas en lenguaje M que será utilizada en versiones futuras del proyecto.
- 5) Tablas **“orden-rango-etario”** y **“orden-grupo-clinicas”**. Estas tablas son creadas para cambiar el orden de visualización predeterminado que tienen los objetos visuales en Power BI.
- 6) **“proyeccion_saturacion”**. Tabla generada automáticamente por Power BI con la creación de un parámetro.

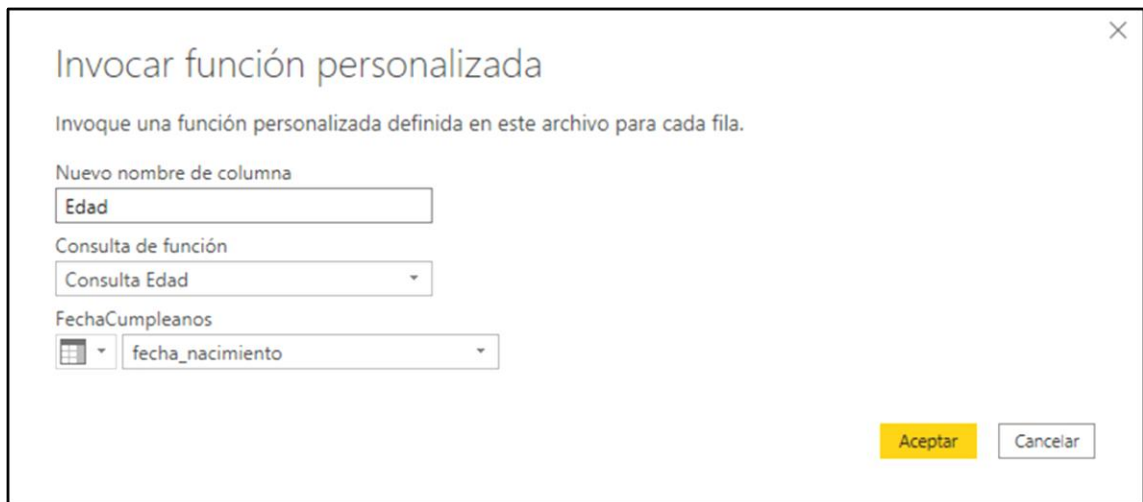
11. TRANSFORMACIÓN DE DATOS

Nota: solo aparecen las tablas a las cuales se le realizaron modificaciones.

- 1) Tabla **especialidad-profesional**:
 - a) Se quitó la columna3 ya que Power BI tomó una columna que no tenía información.
- 2) Tabla **parte-diario**:
 - a) Se quitó la columna8 ya que Power BI tomó una columna que no tenía información.
 - b) Se quitaron posibles duplicados.
 - c) Se cambió el nombre de columna usr_no_asistidos.
- 3) Tabla **usuarios**:
 - a) Se creó una nueva función llamada “Consulta Edad” con un script que permite calcular la edad a partir **fecha_nacimiento**. La misma

toma en cuenta los nacimientos en año bisiesto para que no haya diferencias de un año en las edades calculadas. También devuelve el valor -1 en caso de tener fechas de nacimiento mayores a la fecha actual, ya que se detectó en la base de datos fechas de nacimiento tales como 20/07/2028.

- b) Luego se agregó la columna **Edad** con “Invocar función personalizada” y se utilizó la función previamente creada.



- c) Se reemplazó los valores (-1) resultantes en la columna **edad** por null para que no afecte luego los cálculos con la edad. Los null totales en ese campo fueron menores a 1%.
- 4) Tabla **atenciones**:
- a) En el campo **fecha_atencion** se aplicó dividir columna por delimitador para separar la fecha del horario, la nueva columna se renombró como **hora_atencion**.
- b) Se quitaron posibles duplicados.

- c) Se combinó consultas de la tabla **atenciones** con la tabla **usuarios** a través del campo **id_usuario** ya que luego se precisó trabajar esos campos conjuntamente en una sola tabla.

Combinar

Selecione una tabla y las columnas coincidentes para crear una tabla combinada.

atenciones

fecha_atencion PK	hora_atencion PK	id_usuario PK	id_profesional PK	id_clinica FK	id_especialidad FK
2/1/2021	07:10:30	78130208	46	4	7
2/1/2021	07:18:03	68246176	46	4	7
2/1/2021	07:27:49	160309720	46	4	7
2/1/2021	07:34:41	141789632	46	4	7

usuarios

id_usuario PK	fecha_nacimiento	sexo	Edad
8	27/1/1999	F	22
16	6/10/1942	F	78
24	23/5/2009	M	12
32	17/10/1989	F	31
40	15/10/1993	F	27

Tipo de combinación

Externa completa (todas las filas de ambas)

☐ Use las coincidencias aproximadas para comparar la combinación.

Opciones de coincidencia aproximada

Aceptar

Cancelar

- 5) En todos los casos los encabezados los aplicó bien automáticamente y se corroboró que los tipos de datos estuvieran correctos, cuando no, se aplicó “Detectar tipo de datos” y se volvió a verificar.

12. ANÁLISIS FUNCIONAL DE TABLERO

7) Solapa **Global**:

Se puede visualizar,

- ❖ Como fue variando el número de atenciones durante los seis meses.
- ❖ La discriminación de las atenciones por: Atenciones Coordinadas, Agenda Espontánea y Consulta Domiciliaria.
- ❖ El porcentaje y cantidad de atenciones según sexo masculino y femenino.
- ❖ Distribución de atenciones según rangos etarios dividido por categorías femenino y masculino.
- ❖ Las atenciones de las 21 clínicas pudiendo a su vez discriminar a grandes rasgos las especialidades más demandadas en cada una de ellas.
- ❖ En gráfico de columnas agrupadas una comparación de atenciones con igual período del año anterior (los primeros 6 meses del año 2020). Se puede comparar solo por fecha y clínicas.



Ilustración Solapa Global

8) Solapa **Clínicas**:

Se puede visualizar,

- ❖ Ubicación de todas las clínicas en el mapa con el tamaño de la burbuja según cantidad de atenciones.

- ❖ El Top 5 especialidades con mayor cantidad de atenciones, la idea es que al seleccionar en los filtros las clínicas de una en una se pueda identificar cuáles son las primeras 5 especialidades más demandadas en cada una de ellas.
- ❖ Formando grupos de 5 clínicas, ordenadas de acuerdo a las atenciones en forma descendente, se busca identificar cuanto es el peso sobre el total de atenciones de cada grupo.

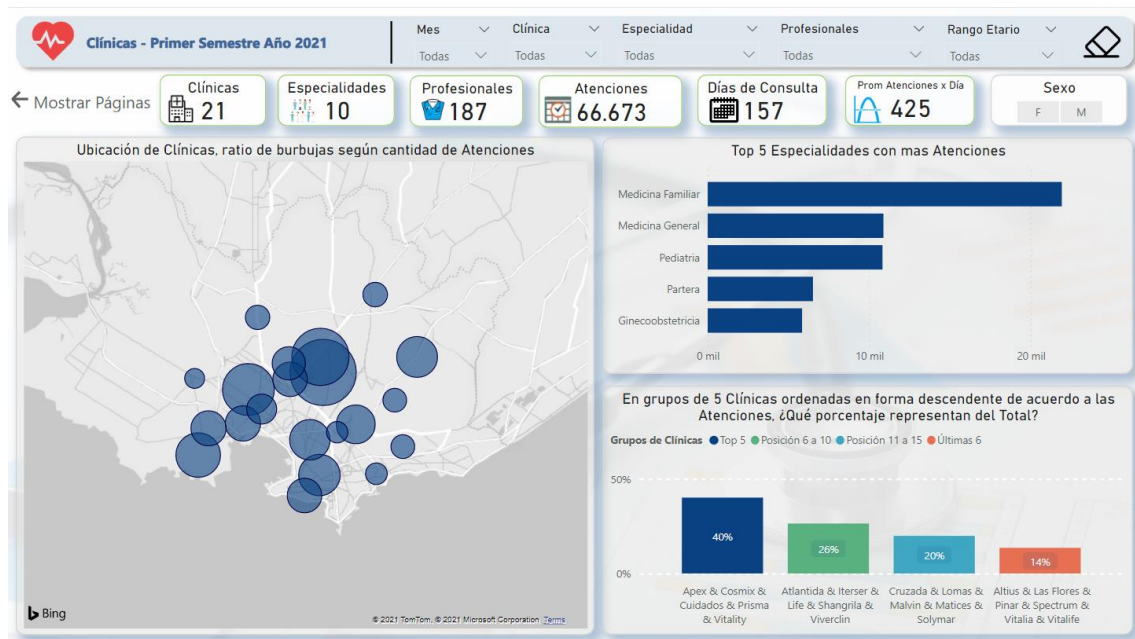


Ilustración Solapa Clínicas

9) Solapa **Especialidades**:

Se puede visualizar,

- ❖ En un Treemap, como se distribuyen las atenciones por especialidad. Se ve claramente desde las especialidades más demandadas a las menos concurridas.
- ❖ En un WorldCloud a simple vista se pueden ver los diagnósticos más frecuentes de una forma visualmente llamativa.

- ❖ Una tabla con el top 15 diagnósticos para que se pueda seleccionar por especialidad y ver con más detalle, en orden descendente, los 15 diagnósticos más importantes según cada una.



Ilustración Solapa Especialidades

10) Solapa **Ausentismo**:

Se puede visualizar,

- ❖ El porcentaje de Inasistencias a consultas con relación a las asistencias.
- ❖ En 2 gráficos de columnas 100% apiladas ver la relación de atenciones y ausencias de todas las clínicas y especialidades.
- ❖ Una matriz con todas las clínicas listadas con la cantidad de inasistencias y el porcentaje que corresponde a esas inasistencias. De esta manera, se puede ver claramente cuáles fueron las clínicas con mayor porcentaje de ausentismo y a su vez cuanto es el peso en inasistencias absolutas. La idea es identificar fácilmente en dónde hay mayor ausentismo tanto porcentual como numéricamente.



Ilustración Solapa Ausentismo

11) Solapa **Proyección Saturación:**

Se puede visualizar,

- ❖ Aquí con un parámetro, la idea es ver como varían las atenciones con un aumento del 13% por cada deslizamiento. Tomamos el 13% ya que fue el aumento de atenciones con relación al mismo semestre del año anterior. Pusimos un tope hipotético de resistencia del sistema en 20.000 atenciones mensuales. De esta forma se puede ver, por ejemplo, con 4 deslizamientos, que sería lo mismo que decir 4 años al 13% de aumento, las atenciones superarían el límite hipotético del sistema.



Ilustración Solapa Proyección de Saturación

13. MEDIDAS CALCULADAS

2) `numero_atenciones` = `COUNTROWS(atenciones)`

- ❖ Cada fila en la tabla **atenciones** es una atención, por tanto, con `COUNTROWS` se contaron todas las filas para calcular el total de atenciones.

3) `numero_usuarios_unicos` = `DISTINCTCOUNT(atenciones[id_usuario PK])`

- ❖ A su vez se tiene el campo **id_usuario** en tabla **atenciones**, por tanto, se contaron sus apariciones distintas (`DISTINCTCOUNT`) para saber el total de usuarios únicos de esas atenciones.

4) A continuación 3 medidas diferentes:

- `agenda_coordinada` = `CALCULATE(COUNTROWS(atenciones), atenciones[tipo_orden] = "Coordinacion")`

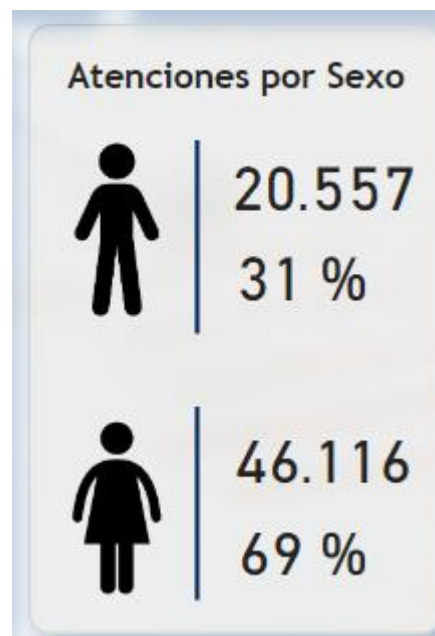
- `agenda_espontanea` = `CALCULATE(COUNTROWS(atenciones),
atenciones[tipo_orden] = "Agenda Espontanea")`
- `consulta_domiciliaria` = `CALCULATE(COUNTROWS(atenciones),
atenciones[tipo_orden] = "Consulta Domiciliaria")`

❖ Las atenciones se dividen según el campo **tipo_orden** en la tabla **atenciones** por:

- “Coordinación”, que son las atenciones que estaban agendadas en el sistema con anterioridad,
- “Agenda Espontanea”, que son las atenciones que se dan en el momento sin agenda previa y
- “Consulta Domiciliaria”, que son las atenciones que se realizan a domicilio.

Con estas 3 medidas se pretende calcular el número de atenciones correspondientes a cada **tipo_orden** filtrando según corresponda.

5) Para construir la siguiente parte del tablero:



Se utilizaron las siguientes medidas:

- `atenciones_sexo_femenino` = `CALCULATE(COUNTROWS(atenciones),
atenciones[usuarios.sexo] = "F")`
- `porcentaje_sexo_f` =
`VAR conteo_filas_masculino = CALCULATE(COUNTROWS(atenciones),
atenciones[usuarios.sexo] = "M")`
`VAR conteo_filas_femenino = CALCULATE(COUNTROWS(atenciones),
atenciones[usuarios.sexo] = "F")`

`return`
`DIVIDE(conteo_filas_femenino , (conteo_filas_femenino +
conteo_filas_masculino))`
- `atenciones_sexo_masculino` = `CALCULATE(COUNTROWS(atenciones),
atenciones[usuarios.sexo] = "M")`
- `porcentaje_sexo_m` =
`VAR conteo_filas_masculino = CALCULATE(COUNTROWS(atenciones),
atenciones[usuarios.sexo] = "M")`
`VAR conteo_filas_femenino = CALCULATE(COUNTROWS(atenciones),
atenciones[usuarios.sexo] = "F")`

`return`
`DIVIDE(conteo_filas_masculino , (conteo_filas_masculino +
conteo_filas_femenino))`

❖ Se calcularon las atenciones correspondientes al sexo Masculino y Femenino contando las filas y filtrando por el campo **usuarios.sexo** = “M” y **usuarios.sexo** = “F” según corresponda, de la tabla **atenciones**. Luego, con la función **DIVIDE**, se calculó en otras dos medidas el porcentaje del total de atenciones que corresponde a Femenino y a Masculino.

6) `dias_totales_consulta` = `DISTINCTCOUNT(atenciones[fecha_atencion PK])`

❖ En esta medida contando las fechas no repetidas se calculó la cantidad de días en los que hubo atención médica.

7) `promedio_atenciones_por_dia` =
`VAR numero_atenciones = COUNTROWS(atenciones)`
`VAR dias_totales_consulta =`
`DISTINCTCOUNT(atenciones[fecha_atencion PK])`

```
return  
numero_atenciones/dias_totales_consulta
```

- ❖ Se dividió el número de atenciones por los días totales de consulta para calcular el promedio de atenciones que se hicieron por día.

```
8) numero_inasistencias_de_usuarios =  
sum(parte_diario[inasistencias_usuarios_aconsulta])
```

- ❖ Se calculó el número de inasistencias de los usuarios a la consulta sumando el campo **inasistencias_usuarios_aconsulta** en la tabla **parte_diario**.

```
9) agendados_coordinados_totales =  
VAR atenciones_agenda_coordinada =  
CALCULATE(COUNTROWS(atenciones), atenciones[tipo_orden] =  
"Coordinacion")  
VAR numero_inasistencias_de_usuarios =  
sum(parte_diario[inasistencias_usuarios_aconsulta])  
  
return  
atenciones_agenda_coordinada + numero_inasistencias_de_usuarios
```

- ❖ Se calcularon los agendados totales a consulta sumando solo las atenciones coordinadas más las inasistencias de usuarios ya que se quiere saber el porcentaje de inasistencias solo con respecto a la agenda coordinada total. No interesa saber aquí las atenciones espontáneas ni consultas domiciliarias.

```
10) porcentaje_de_inasistencias =  
VAR numero_inasistencias_de_usuarios =  
sum(parte_diario[inasistencias_usuarios_aconsulta])  
VAR atenciones_agenda_coordinada =  
CALCULATE(COUNTROWS(atenciones), atenciones[tipo_orden] =  
"Coordinacion")  
  
return  
DIVIDE(numero_inasistencias_de_usuarios,  
atenciones_agenda_coordinada + numero_inasistencias_de_usuarios)
```

- ❖ Aquí teniendo el número de inasistencias por un lado y el número de atenciones con agenda coordinada por el otro, se calculó el porcentaje que corresponde a las inasistencias.

11) `comparacion_anio_2020` = `CALCULATE(COUNTROWS('atenciones-2020'), SAMEPERIODLASTYEAR('tabla calendario DAX'[fecha]))`

- ❖ Aquí se utilizó inteligencia de tiempo (`SAMEPERIODLASTYEAR`) para luego poder hacer una comparación en una gráfica entre las atenciones del semestre de 2021 y el mismo periodo del año 2020.

```
12) %variacion_2021_vs_2020 =
VAR semestre_2020_atenciones = CALCULATE(COUNTROWS('atenciones-2020'), SAMEPERIODLASTYEAR('tabla calendario DAX'[fecha]))
VAR semestre_2021_atenciones = COUNTROWS(atenciones)

return

DIVIDE((semestre_2021_atenciones - semestre_2020_atenciones),
semestre_2020_atenciones)
```

- ❖ Con las atenciones del primer semestre de 2021 y el mismo período del año 2020 se calculó el porcentaje de variación.

```
13) atenciones_parametro_saturacion =
VAR numero_atenciones = COUNTROWS(atenciones)
VAR parametro_saturacion = 'proyeccion_saturacion'[Valor Hipotesis de Saturacion]

return
numero_atenciones * parametro_saturacion
```

- ❖ En esta medida, se multiplicó la cantidad de atenciones por el parámetro generado para poder hacer una proyección de en cuanto tiempo se saturaría el sistema.

14. FUTURAS LÍNEAS

Para las siguientes versiones se trabaja en el agregado de nuevas características:

- 1) Se agregarán nuevas visualizaciones con respecto a la productividad de los profesionales, calculando la cantidad de atenciones de acuerdo a horas trabajadas. Para ello, se cuenta con el campo calculado “duración de agenda” de los profesionales y cuantos pacientes atendieron.
- 2) Se ha realizado la “**tabla horas**”. Con ella, a futuro se agregarán visualizaciones que permitan identificar las atenciones por hora, en que horarios son más concurridas las clínicas por los usuarios y cual es la hora pico.