# Documentación de los reportes de Looker Studio

Actualmente en BigQuery se dispone de un volumen de datos clínicos considerables pertenecientes al Sistema Nacional de Salud de República dominicana. En este proyecto se usó Looker Studio para la elaboración de tableros que permitieran visualizar estos datos. En este documento se explicarán los aspectos fundamentales a tener en cuenta para interpretar la información mostrada y crear otros tableros para mostrar otro tipo de información sobre los mismos datos. También se expondrán las particularidades de dichos datos y los defectos que estos presentan e impiden resumir su información de forma adecuada para ciertos propósitos.

1. Relación de reportes y fuentes de datos disponibles

En este proyecto se crearon varios reportes que usan diferentes bases de datos disponibles en BigQuery. La **Tabla 1** muestra la relación de los reportes realizados en Looker Studio (LS) y sus fuentes de datos. Así como una descripción general de los mismos.

**Tabla 1. Relación entre reportes y fuentes de datos**

|  |  |  |
| --- | --- | --- |
| **Nombre del reporte** | **Descripción del reporte** | **Fuentes de datos usadas** |
| PAI-Covid-19 | Muestra información relativa a la campaña de vacunación contra la COVID-19. Tiene cuatro páginas donde se muestran las estadísticas de sobre las generalidades del proceso de vacunación, los centros de vacunación, las dosis administradas a cada paciente y los datos relativos al género y los grupos etarios. | * registro\_nominal\_unificado\_v3\_prod * center\_vaccines\_agg * vaccinated\_agg * vaccines\_agg * center\_vaccinated\_agg |
| HCD-Pacientes | Muestra información relativa a la campaña de vacunación contra la COVID-19. Tiene tres páginas donde se muestra un resumen de la información disponible respecto al género y los grupos etarios, una segregación por mapas con filtros por la localización de determinados grupos de pacientes y una página dedicada a filtrar y mostrar toda la información de los grupos etarios y el género por periodo. | * patient\_v2 |
| HCD-Alergias | Muestra información relativa a las alergias/intolerancias frecuentes de los pacientes de los pacientes. También brinda información sobre la cobertura de inmunización. | * patient\_history\_alergies\_v2 * patient\_history\_inmuzations\_v2 * patient\_v2 |
| HCD-Diagnósticos | Contiene información sobre el perfilado de diagnósticos de diferentes padecimientos como ciertos tipos de hipertensión y diabetes. | * patient\_history\_diagnostics |
| HCD-Embarazos | Contiene información acerca de la relación de las prescripciones de ácido fólico con los casos de embarazos confirmados. | * patient\_history\_mediactions * patient\_history\_diagnostics |
| HCD-Laboratorios | Muestra información sobre las pruebas de laboratorios y procedimientos frecuentes llevados a cabo en las instituciones de salud. | * patient\_history\_diagnostics * patient\_history\_labs * patient\_history\_\_procedures |
| HCD-Hábitos tóxicos | Contiene información sobre los hábitos tóxicos de los pacientes como el consumo de café y de sustancias alcohólicas. | * patient\_history\_toxic\_habits\_v2 * patient\_v2 |
| HCD-Consultas | Muestra información sobre sobre los datos generales de los servicios de atención médica y también muestra información específica de los servicios médicos por provincia. | * patient\_history\_atentions\_v2 * patient\_v2 |

1. Defectos de los datos

Muchos de las tablas presentan diferentes tipos de errores, los cuales se enumerarán a continuación:

1. Campos vacíos: Las tablas presentan campos que no siempre están llenos y en algunos casos, es la mayoría de las veces. Ejemplo de esto es la tabla *patient*, donde los campos *province* (especifica la provincia a la que pertenece un paciente) y *municipality* (especifica el municipio al que pertenece un paciente), solo están disponibles para 154 pacientes, siendo el tamaño de la muestra de 1 millón de pacientes.
2. Falta de estandarización de campos: Existen campos que representan la misma información, pero están escritos de forma diferente. Esto se puede observar en las tablas *patient\_history\_alergy K* y *patient\_history\_attentions\_toxic\_habits*. En estas tablas, categorías como *Ninguno, No, Nada, 0, Null, NA …etc*, representan la misma información, pero al no estar escritas de la misma forma se dificulta su contabilización por parte de Looker Studio (LS).
3. Mala asignación del tipo de dato a diferentes campos: Datos como la presión arterial, la estatura o el peso, deberían guardarse en algún tipo de dato numérico, mientras que en la base de datos están guardados en formato de texto. Esto limita considerablemente el tratamiento de estos datos a través de LS al no poder realizar operaciones matemáticas sobre ellos. La operación CAST (que puede transformar un dato de texto en uno numérico) puede parecer una opción, pero tiene asociadas otras limitaciones. Si el dato tiene un punto decimal, esta función simplemente lo renueve y el número final será simplemente la unión de sus dígitos. Por ejemplo, si tenemos 36.66 como tipo de dato TEXT, al llevarlo a NUM usando CAST obtendremos 36666, los cual implica que se deben realizar otras operaciones condicionales sobre este dato que a gran escala pueden ralentizar el tablero.
4. Apuntes sobre los tableros de LS

En este proyecto se usan datos combinados entre la tabla *patient* y otras tablas, para realizar la georreferenciación de los datos y poder visualizar diferentes métricas en un mapa geográfico dividido por provincias. Producto de que el campo *province* no está disponible para la gran mayoría de pacientes, en los tableros que incluyen un mapa geográfico solo tienen en cuenta los datos de los pacientes con este campo rellenado. Esto se debe tener en cuenta a la hora de analizar dichos datos. Una posible solución a este problema es el completamiento de estos datos mediante un proceso de geocoding a través de APIs open source, usando el campo address como input. De esta manera se podría obtener la provincia o el municipio de la mayoría de los pacientes.

Otro problema es que, por ahora, los datos usados en los tableros provienen de Elastic Search, y se migran a BigQuery de forma manual y versionando las tablas. Esto implica que cada vez que se realice este proceso, se deberán actualizar los gráficos de LS manualmente. Esta situación es temporal y eventualmente este proceso será automatizado.

1. Potencialidades de los datos de HCD\_SNS

Los datos disponibles actualmente también pueden ser usados para aplicaciones diferentes a los requerimientos que fueron solicitados.

Una de las posibles aplicaciones puede ser la creación de reportes de datos que permitan extraer información sobre la correlación entre una zona geográfica, la edad y la incidencia de determinados hábitos tóxicos. Además de la importancia desde el punto de vista social que podrían tener la información extraída de aquí, esta podría usarse para estimar a la aparición de determinados padecimientos crónicos en la población y gestionar mejor los servicios de salud del país. Otra posible aplicación es el uso de técnicas de aprendizaje automático para estimar la demanda de medicamentos. La gestión del inventario de medicamentos es una de las tareas más complejas a realizar en una farmacia. Una buena estimación de las compras favorece el compromiso entre satisfacer la demanda de los usuarios y minimizar los costos de mantenimiento de inventario y de almacenamiento. Por ello, conocer a priori la demanda de un determinado medicamento ayuda a decidir qué cantidad se debe comprar de producto. Las aplicaciones inteligentes, como los sistemas de recomendaciones o los sistemas predictivos, son altamente demandados por la industria farmacéutica dado su potencial para optimizar la compra y/o tener un mayor control de los inventarios, entre otros beneficios [1, 2].

1. Documentación de referencia.
2. <https://support.google.com/looker-tudio/answer/6268208?hl=ES#zippy=%2Csecciones-de-este-art%C3%ADculo>
3. <https://support.google.com/looker-studio/answer/6303415?hl=ES>
4. <https://support.google.com/looker-> tudio/answer/9851950?hl=ES#zippy=%2Csecciones-de-este-art%C3%ADculo
5. https://support.google.com/looker-studio/answer/9209811?hl=ES
6. Referencias

[1] D. Rivero Albarrán, S. M. Arciniegas Aguirre, and M. F. Fernández Badillo, "Un modelo para predecir la demanda en farmacias," *Redmarka. Revista de Marketing Aplicado,* vol. 26, no. 1, pp. 1-14, 2022.

[2] A. Luque Sendra, F. Aguayo González, J. R. Lama Ruiz, E. González-Regalado Montero, and A. M. Martín Gómez, "TÉCNICAS AVANZADAS DE GESTIÓN DE STOCKS. APLICACIÓN A LA GESTIÓN DE ALMACENES EN FARMACIA HOSPITALARIA," 2016.