

UNIVERSIDAD TÉCNICA DE MANABÍ



**FACULTAD CIENCIAS INFORMATICAS
CARRERA SISTEMA DE LA INFORMACIÓN**



Estudiantes:



Bravo Navarrete Gema Ruby
Bravo Párraga Adonys Ricardo
Chávez Reyes Antony Joseph

Docente:



Ing. Ermenson Ordoñez

Asignatura:



Inteligencia de Negocio

Paralelo:



"B"

Periodo:



Mayo - Septiembre

Tabla de contenido

<u>1. Introducción</u>	3
<u>2. Datos del negocio</u>	4
<u>3. Evaluación del caso de negocio</u>	4
<u>4. Planificación del proyecto</u>	4
<u>4.1. Alcance</u>	4
<u>4.2. Cronograma (tareas, recursos, tiempo)</u>	6
<u>4.3. Levantamiento de información</u>	8
<u>4.3.1. Entrevista</u>	8
<u>4.3.2. Infraestructura técnica</u>	8
<u>4.3.3. Áreas y procesos del negocio</u>	10
<u>4.3.4. Requerimientos de negocio</u>	10
<u>4.4. Análisis de la herramienta (Power BI)</u>	10
<u>5. Diseño y análisis de los datos</u>	11
<u>5.1. Modelo dimensional</u>	11
<u>5.1.1. Descripción de los datos (Metadatos)</u>	11
<u>5.1.2. Diseño de Data Warehouse</u>	13
<u>5.1.3. Tablas de hechos, dimensiones y medidas</u>	13
<u>5.2. Procesos ETL</u>	14
<u>5.3. Cubos OLAP</u>	14
<u>5.4. Indicadores</u>	15
<u>5.5. Reportes y Visualizaciones</u>	16
<u>6. Resultados</u>	1017
<u>6.1. Análisis del aporte a la toma de decisión</u>	17

Análisis de Datos de Redimiendo de dos Escuelas Portuguesas

1. Introducción

En el ámbito educativo, la capacidad de acceder y analizar datos es esencial para informar las decisiones pedagógicas y estratégicas. La gestión eficiente del rendimiento estudiantil es un objetivo fundamental para las instituciones educativas.

Este informe se adentra en un análisis de una base de datos que abarca el rendimiento estudiantil en dos escuelas portuguesas. Se exploran elementos que se unen en la evaluación, planificación y aprovechamiento de esta información valiosa para la mejora continua.

En primer lugar, la sección de datos del negocio presenta los procesos que impulsan el funcionamiento de las dos escuelas y aborda los desafíos educativos y administrativos.

La fase de levantamiento de información establece las bases para realizar un análisis de rendimiento estudiantil.

El diseño y análisis de los datos se revelan a través de la construcción de un modelo dimensional, la descripción completa de datos a través de metadatos y la estructura de un data warehouse sólido, incorporando tablas de hechos, dimensiones y medidas.

Los procesos ETL, la creación de cubos OLAP, la definición de indicadores y la selección de objetos visuales constituyen etapas cruciales para potenciar el análisis.

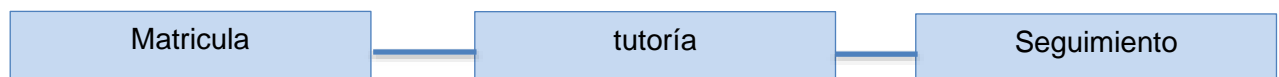
Finalmente, esta investigación culmina con la evaluación de resultados, analizando cómo esta base de datos enriquece la toma de decisiones y promueve la mejora continua del rendimiento estudiantil en el contexto educativo portugués.

2. Datos del negocio

Escuelas portuguesas ubicadas en la central de Quito, cuentan con más de 3000 estudiantes, de las cuales su objetivo es busca una educación de calidad proporcionándoles a los estudiantes las herramientas, habilidades y conocimientos necesarios para que puedan enfrentar los desafíos de la vida con confianza, ética y competencia.

Los procesos de negocios de estas instituciones educativas abarcan una amplia gama de actividades que van desde la administración y gestión hasta la entrega de contenido educativo.

Unos de esos procesos son:



3. Evaluación del caso de negocio

Las escuelas portuguesas usan hojas de Excel, un programa informático que permite organizar y analizar datos, para mostrar el rendimiento de los estudiantes en la educación secundaria (bachillerato). Los datos contienen información sobre las calificaciones, las características demográficas, sociales y escolares de los alumnos. Así, se han proporcionado dos conjuntos de datos sobre el desempeño en dos materias distintas: Matemáticas y lengua portuguesa. Esto nos permitirá encontrar los indicadores adecuados para evaluar el rendimiento de los alumnos según la escuela o la materia, entre otras opciones que los directivos podrán conocer y tomar mejores decisiones. Para ello, usaremos una herramienta de inteligencia de negocio llamada Power BI de Microsoft, donde se podrán visualizar de la mejor manera estos reportes.

4. Planificación del proyecto

4.1. Alcance

El alcance del proyecto abarca la implementación de un sistema de informes de rendimiento estudiantil en escuelas portuguesas utilizando la herramienta de inteligencia de negocio Power BI de Microsoft. Estos informes se basarán en los conjuntos de datos que contienen información sobre el desempeño de los estudiantes en las materias de Matemáticas y Lengua Portuguesa durante la educación secundaria (bachillerato). El objetivo es proporcionar a los directivos y administradores escolares indicadores adecuados

para evaluar y tomar decisiones informadas sobre el rendimiento estudiantil, tanto a nivel de escuela como de materia, esto se presentará por medio de un dashboard. El sistema tendrá las siguientes funcionalidades:

1. Generar informes dinámicos y personalizables que muestren el rendimiento estudiantil por escuela, materia, género, edad u otros criterios.
2. Visualizar los informes en formatos gráficos, tabulares o textuales según las preferencias del usuario.
3. Aplicar filtros o comparaciones para analizar los datos desde diferentes perspectivas o niveles de detalle.
4. Utilización de marcadores y botones para un acceso rápido de información.

4.2. Cronograma (tareas, recursos, tiempo)

Utilizaremos la metodología Kimball donde se redactará una secuencia de fases con sus respectivas tareas para el detalle de este proyecto:

Nombre de Tareas	Recursos	Tiempo	
		Inicia	Finaliza
Planificación y Preparación			
Identificar los objetivos empresariales y las necesidades de análisis.	Personal especializado, fondo financiero para la planificación y el desarrollo del proyecto	01/05/2023	01/05/2023
Definir el alcance del proyecto, identificar las fuentes de datos y establecer prioridades.		02/05/2023	04/05/2023
Formar un equipo de proyecto y asignar responsabilidades.		04/05/2023	04/05/2023
Cronograma (tareas, recursos, tiempo)	Software de gestión de proyectos	05/05/2023	08/05/2023
Levantamiento de información			
Entrevistas	Documentación	9/05/2023	16/05/2023
Infraestructura técnica	Equipo y software necesario para analizar y comprender los requerimientos del negocio, junto a personal especializado en esas áreas	16/05/2023	19/05/2023
Áreas y procesos de negocios		22/05/2023	25/05/2023
Requerimiento de negocios		26/05/2023	31/05/2023
Diseño de tablas Hechos y dimensiones			
Identificar las dimensiones clave del negocio	Personal con experiencia en diseño de bases de datos y modelado dimensional. Recursos de Documentación. Software de modelado de bases de datos y herramientas de diseño dimensional.	01/06/2023	05/06/2023
Diseñar tablas de dimensiones que contengan atributos		05/06/2023	06/06/2023
Definir claves, atributos, jerarquías y metadatos.		07/06/2023	09/06/2023
Identificar los procesos de negocio que generan hechos medibles		12/06/2023	15/06/2023
Diseñar tablas de hechos que contengan medidas cuantificables y claves de dimensiones.		16/06/2023	19/06/2023
Definir las medidas y relaciones de hechos.		20/06/2023	22/06/2023
Diseño de ETL (Extracción, Transformación y Carga)			

Diseñar procesos de extracción, transformación y carga para mover los datos desde las fuentes a las tablas de dimensiones y hechos.	Personal con experiencia, software de ETL para diseñar, desarrollar y programar procesos de extracción, transformación y carga. Acceso a las fuentes de datos y a la infraestructura necesaria para realizar las transformaciones.	23/06/2023	28/06/2023
Transformar y limpiar los datos para garantizar la calidad y consistencia.		29/06/2023	04/07/2023
Diseño de Cubos OLAP			
Diseñar cubos OLAP (Online Analytical Processing) que permitan un análisis multidimensional.	Personal con experiencia. Software de diseño y construcción de cubos multidimensionales. Datos limpios y transformados listos para ser utilizados en los cubos OLAP.	05/07/2023	07/07/2023
Crear consultas y vistas que faciliten el acceso y análisis de datos.		10/07/2023	10/07/2023
Desarrollo e Implementación:			
Implementar los diseños de dimensiones, hechos, ETL y agregados en la infraestructura de almacén de datos.	Personal capacitado desarrolladores y administradores en base de datos, Software para implementar y gestionar la base de datos y los cubos.	11/07/2023	14/07/2023
Probar y validar los procesos ETL, la carga de datos y las consultas, por medio de indicadores, reportes y visualizaciones		17/07/2023	21/07/2023
Pruebas y Validación:			
Realizar pruebas exhaustivas para asegurar que los datos sean precisos y las consultas sean efectivas.	Personal para llevar a cabo pruebas exhaustivas en los procesos, datos y consultas.	24/07/2023	04/08/2023
Validar el rendimiento del sistema y ajustar los agregados según sea necesario.	Software de pruebas para verificar la precisión de los datos y la eficacia de las consultas.	07/08/2023	18/08/2023
Formación y Entrega:			
Capacitar a los usuarios finales para que puedan aprovechar al máximo el almacén de datos y las herramientas de análisis.	Personal de Capacitación. Documentación, tutoriales y recursos de capacitación para los usuarios. Software de presentación para entregar informes, paneles y consultas.	21/08/2023	23/08/2023
Entregar informes, paneles y consultas diseñadas para las necesidades del negocio.		24/08/2023	27/08/2023

4.3. Levantamiento de información

4.3.1. Entrevista

Para conocer el rendimiento de los estudiantes se han recopilado datos sobre dos materias en dos escuelas, una de portugués y de matemáticas. Se obtuvieron los datos de estudiantes que incluyen calificaciones, edad y actividades extracurriculares e información de los padres y la asesoría que reciben.

En cuanto a la relación con los padres se indica que en la mayoría de los casos viven juntos y tienen una buena relación familiar, así mismo, ambos padres trabajan en áreas como servicios, salud y docencia.

Pocos estudiantes han tenido fracasos en materias en un rango de 1 a 3 y aproximadamente la mitad ha tenido ausencias considerablemente altas en la escuela. Dedicar horas de estudio que oscila entre 1 a 4 horas y mantienen en promedio buena salud. Y de los más de mil estudiantes en general solamente setenta han mantenido una calificación final considerablemente aceptable sobre veinte.

Dedicar entre 1 y 4 horas de viaje para trasladarse de la escuela al domicilio y de una a cinco horas en salidas con amigos lo que en su mayoría superan las tres horas lo cual supone una buena relación social.

Menos de la mitad de la lista de estudiantes frecuenta beber alcohol entre los días de la semana y aproximadamente la mitad frecuenta beber alcohol los fines de semana.

En general se observa que el rendimiento de los estudiantes ha comenzado con calificaciones regulares y pasado un tiempo ha subido el rendimiento académico y finalmente decayó el promedio de sus calificaciones.

4.3.2. Infraestructura técnica

Se requiere de un equipo para realizar las tareas y de software de procesamiento de datos como Excel para visualizar los datos obtenidos, Power BI para realizar transformaciones, cálculos de medidas y campos calculados, análisis de datos, predicciones y visualizaciones de rendimiento.

4.3.3. Áreas y procesos del negocio

Las áreas aplicadas por las escuelas portuguesas para medir el rendimiento de los estudiantes son: por medios de pruebas, calificaciones, tutorías, tareas en la aula y asesoramientos. En este análisis se ha considerado la educación secundaria (bachillerato), cuyo objetivo es buscar una educación de calidad que proporcione a los estudiantes las herramientas, habilidades y conocimientos necesarios para que puedan enfrentar los desafíos de la vida con confianza, ética y competencia.

4.3.4. Requerimientos de negocio

Para cumplir con los requisitos comerciales establecidos, llevamos a cabo una reunión con los líderes de estas instituciones. Durante este encuentro, se nos brindó una descripción detallada de los problemas que enfrentan, así como los datos necesarios para realizar un análisis exhaustivo. Estos datos son esenciales para comprender las especificaciones detalladas que desean ver reflejadas en la ejecución de este proyecto. Al interpretar esta información, estamos capacitados para tomar decisiones estratégicas que no solo impulsarán la implementación exitosa del proyecto, sino que también contribuirán a lograr resultados superiores en el ámbito de sus instituciones educativas.

El proyecto debe cumplir con los siguientes requerimientos pedidos por el usuario:

1. ¿Cuál es el rendimiento general de los estudiantes por escuela?
2. ¿Cuál es el rendimiento por asignatura?
3. ¿Cuál es la cantidad y porcentaje de los estudiantes que obtuvieron más de 17 o mas en su calificación final?
4. ¿Cuál es la asignatura más demanda?
5. ¿Cuál es el índice de fracasos por edades?
6. ¿Cuál es el índice de fracaso por tiempo de estudio?
7. ¿Cuál es el recuento de fracaso por género?
8. ¿Cuál es el grupo predominante de representantes legales de los estudiantes en términos de género?

4.4. Análisis de la herramienta (Power BI)

Power Bi de Microsoft es una plataforma unificada y escalable, es una colección de servicios de software, aplicaciones y conectores que trabajan juntos para convertir el origen de los datos sin relaciones con información coherente, interactiva y visualmente atractiva.

Utiliza alineación con Microsoft 365, Teams y Azure Synapse, combinación precio/valor, cartera de potencia y ambición de producto.

El producto Power BI se compone de una serie de aplicaciones, cada una con sus propias características y usos. Éstos incluyen:

- **Power Query:** una herramienta de conexión de datos que le permite transformar, combinar y mejorar datos de varias fuentes
- **Power Pivot:** una herramienta de modelado de datos para crear modelos de datos.
- **Power View:** una herramienta de visualización de datos que genera gráficos interactivos, gráficos, mapas y otros elementos visuales
- **Power Map:** otra herramienta de visualización para crear imágenes 3D envolventes
- **Power Q&A:** un motor de preguntas y respuestas que le permite hacer preguntas sobre sus datos en un lenguaje sencillo

En nuestro arsenal de recursos materiales, contamos con dos computadoras portátiles completamente equipadas, cada una de ellas dotada con la potencia necesaria para llevar a cabo las tareas requeridas. Además, disponemos de una valiosa licencia completa de Power BI, una herramienta líder en el análisis y visualización de datos, que amplía nuestras capacidades y posibilidades en el manejo eficiente de la información.

El proyecto tiene las siguientes limitaciones:

1. El presupuesto asignado es de \$5.000 dólares americanos.
2. El tiempo estimado para el desarrollo y análisis de los datos es de 4 meses.
3. Los recursos humanos disponibles son un analista de datos, un desarrollador de Power BI y un coordinador del proyecto.
4. Los recursos materiales disponibles son dos computadoras portátiles y una licencia de Power BI.

5. Diseño y análisis de los datos

5.1. Modelo dimensional

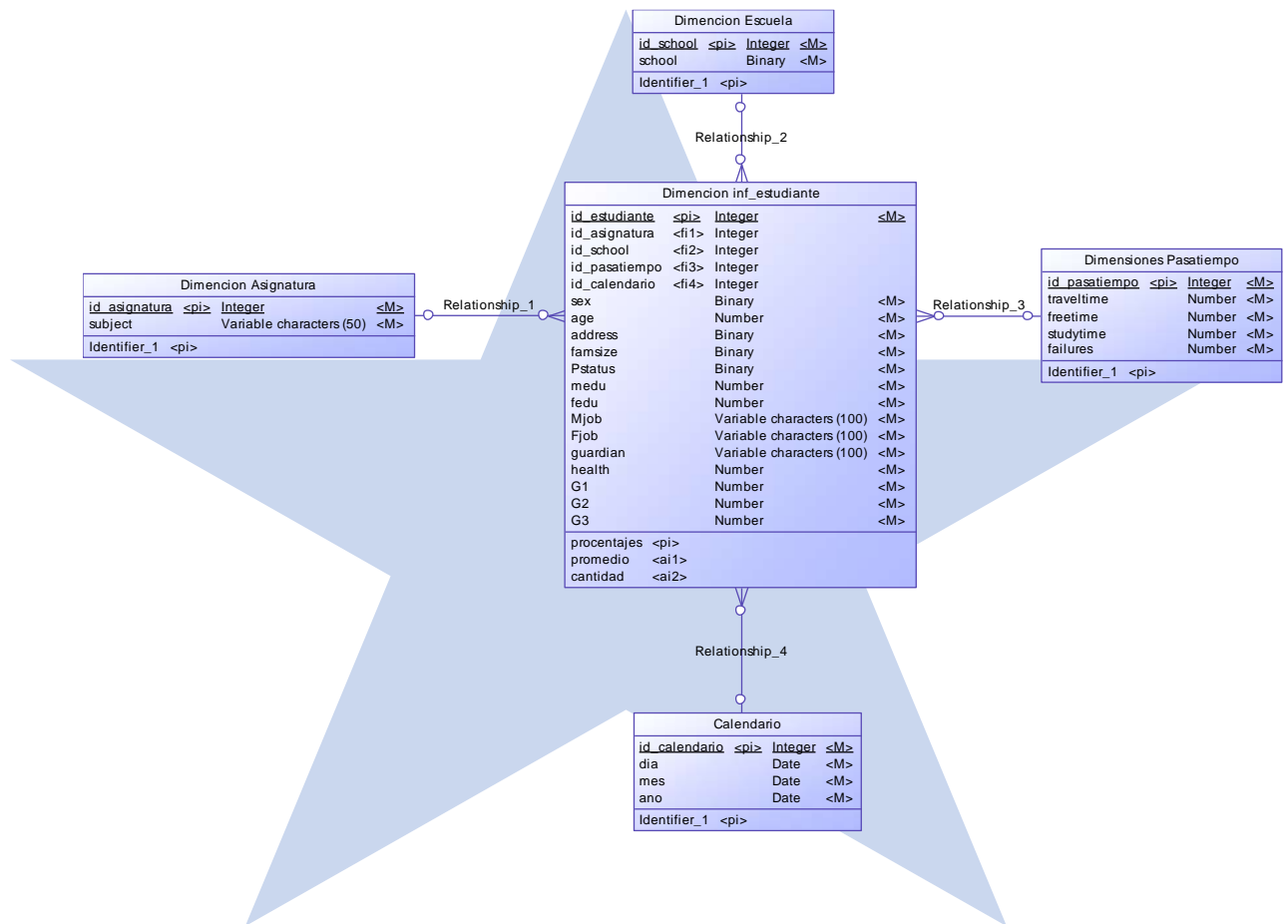
5.1.1. Descripción de los datos (Metadatos)

N°	Metadato	Tipo de Dato	Descripción
1	school	binario	Escuela del estudiante
2	Sex	binario	Sexo o genero del estudiante
3	age	numérico	Edad del estudiante
4	address	binario	Tipo dirección de domicilio del estudiante
5	famsize	binario	Tamaño de la familia
6	Pstatus	binario	Estado de convivencia de los padres
7	medu	numérico	Educación de la madre
8	fedu	numérico	Educación del padre
9	Mjob	varchar	Trabajo de la madre
10	Fjob	varchar	Trabajo del padre
11	reasol	varchar	Razón para elegir esta escuela
12	guardian	varchar	Tutor del estudiante
13	traveltime	numérico	Tiempo de viaje del domicilio a la escuela
14	studytime	numérico	Tiempo de estudio semanal
15	failures	numérico	Número de fallas de clases pasadas
16	schoolsup	binario	Apoyo educativo extra
17	famsup	binario	Apoyo educativo familiar
18	paid	binario	Extra pagadas dentro de la asignatura del curso
19	activities	binario	Actividades extracurriculares
20	nursery	binario	Asistencia de los niños en la guardería

21	higher	binario	Quiere tomar educación superior
22	internet	binario	Acceso a internet en el domicilio
23	romantic	binario	Relación romántica
24	famrel	numérico	Calidad de relación familiar
25	freetime	numérico	Tiempo libre después de la escuela
26	goout	numérico	Salida con amigos
27	Dalc	numérico	Consumo de alcohol en la jornada laboral
28	Walc	numérico	Consumo de alcohol el fin de semana
29	health	numérico	Estado de salud actual
30	absentes	numérico	Número de ausencias escolares
31	G1	numérico	Calificación del primer periodo
32	G2	numérico	Calificación del segundo periodo
33	G3	numérico	Calificación final
34	subject	varchar	Asignaturas matemáticas y portugués
35	Dia	date	Dia
36	mes	date	Mes
37	ano	date	Año

5.1.2. Diseño de Data Warehouse

Presentaremos un diseño estrella, donde podrán visualizar las respectivas tablas de dimensiones y hechos:

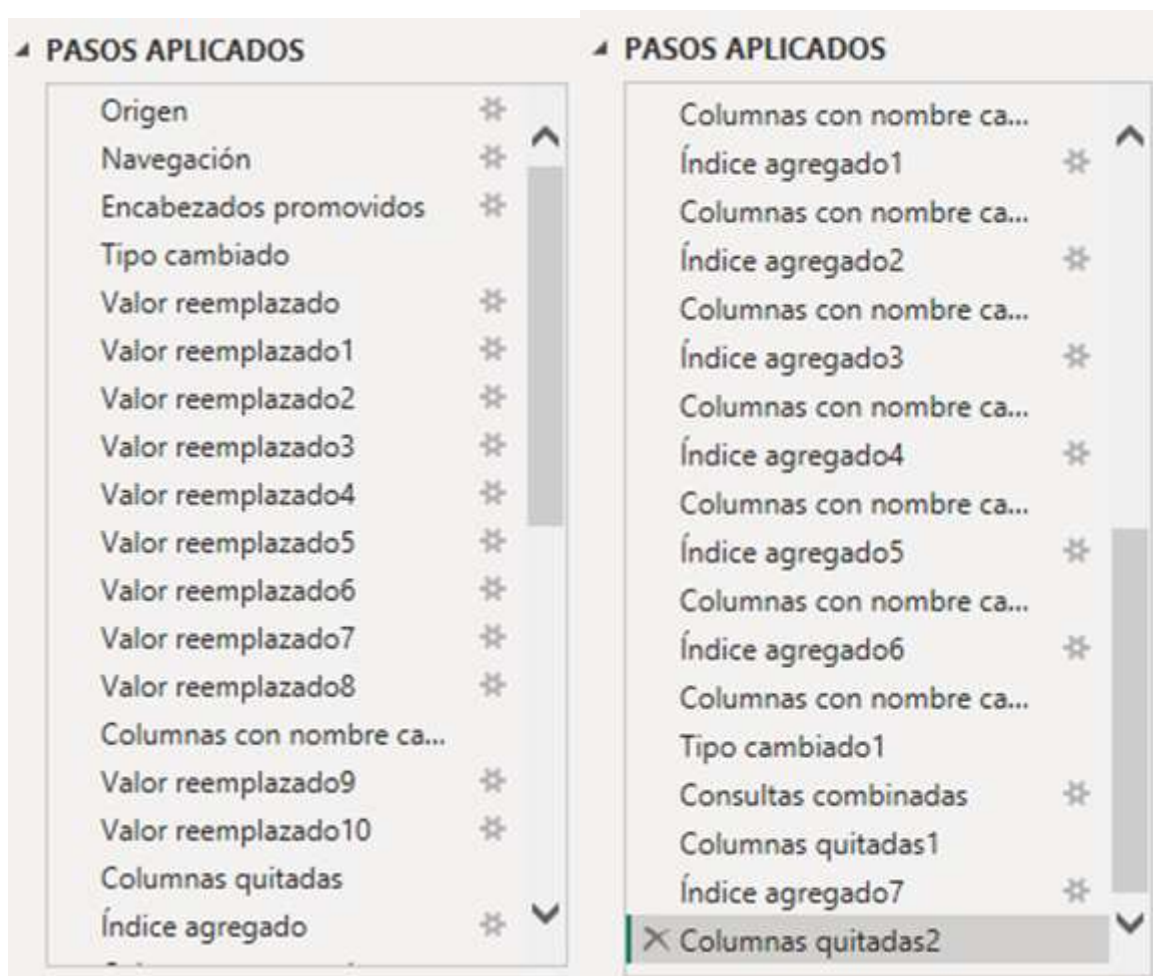


5.1.3. Tablas de hechos, dimensiones y medidas

Hechos	info_estudiante
Dimensiones	Escuela, Calemdario, Pasatiempo, Asignatura
Medidas	Promedio, porcentajes, cantidad.

5.2. Procesos ETL

En el ETL de Power Query se llevaron a cabo una serie de operaciones destinadas a modelar y optimizar una tabla de datos. Estas acciones abarcan desde la designación de la primera fila como encabezados de columna, el ajuste de los tipos de datos en las columnas, la sustitución de valores en columnas particulares, hasta la modificación de nombres de columnas y la introducción de índices en las filas. Además, hemos conseguido combinar tablas basándonos en una columna compartida y suprimir aquellas columnas que no aportan a nuestro análisis. La convergencia de estos pasos nos ha brindado la posibilidad de darle un formato adecuado a los datos y de prepararlos de manera óptima para un análisis más eficaz.

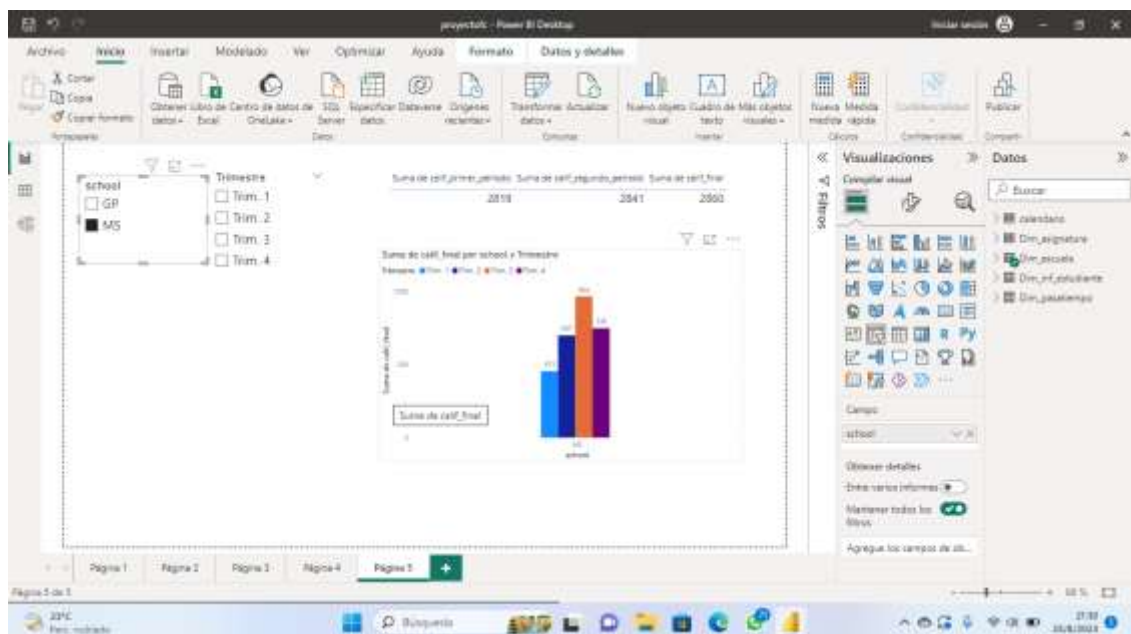


5.3. Cubos OLAP

Presentaremos el cubo OLAP que visualizamos en la plataforma Power Bi para su análisis de datos. Por medio de la construcción del modelo estrella involucramos la conexión de tres tablas principales: "dim_escuela", "calendario" y "Dim_inf_estudiante". A través de una vista OLAP en el informe, incorporaremos dos segmentaciones de datos: una

basada en la categoría de escuela y otra en los trimestres del calendario. Utilizaremos una matriz y un gráfico de columnas agrupadas para presentar de manera clara y visualmente atractiva las calificaciones finales de los estudiantes.

Gracias al uso del cubo OLAP, tenemos la capacidad de realizar búsquedas eficientes en las tres tablas. En nuestro escenario particular, comenzamos buscando en la categoría de "escuela" y luego nos dirigimos a la categoría de "trimestre". Esto nos permite analizar de manera ágil el rendimiento de las dos escuelas en cada trimestre y conocer sus respectivas calificaciones.



5.4. Indicadores

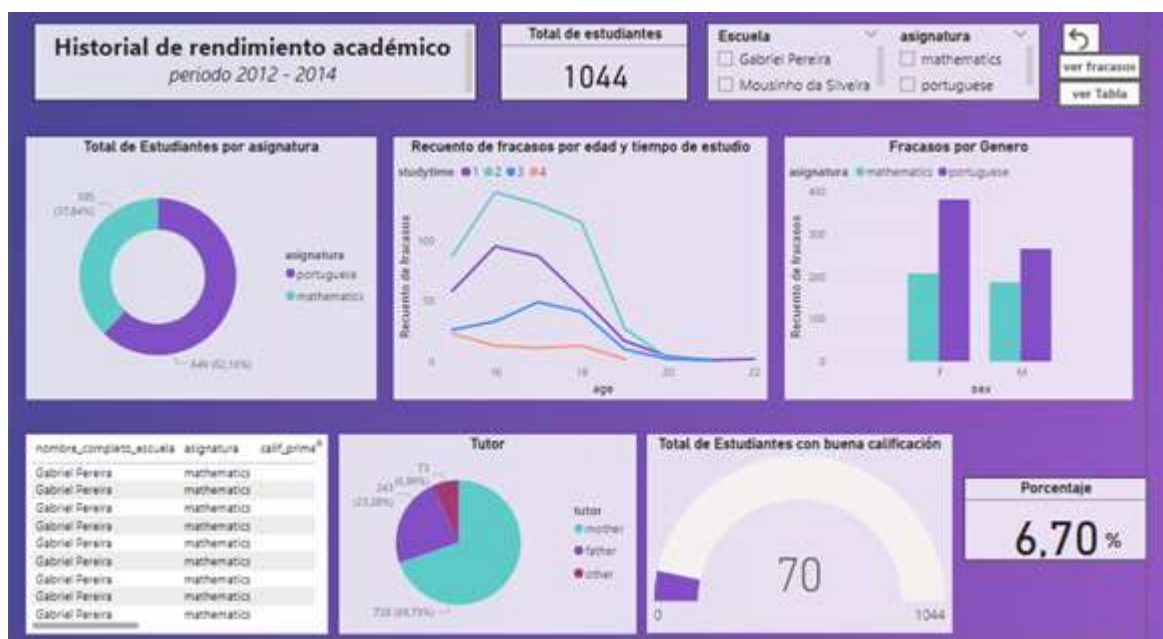
En esta sección, se presentan procesos que cuentan con indicadores de medidas y la fórmula correspondiente para poder calcularlo y los atributos de cada dimensión.

Procesos Escuelas Potuguesas		
Etapa	Nombre	Formula
Rendimiento	1. Total, de estudiantes	1. TotalEstudiantes = COUNTROWS(Dim_inf_estudiante)
	2. Total, de Estudiantes con Calificaciones	2. TotalEstudiantesEnRango = COUNTROWS(FILTER(Dim_inf_estudiante,

	<p>Sobresalientes en el Rango Específico.</p> <p>3. Porcentaje de Estudiantes con Calificaciones Sobresalientes en el Rango Específico</p> <p>4. Calculo trimestral para el Período Académico de la Institución</p>	<p>Dim_inf_estudiante[calif_final] >= 17 && Dim_inf_estudiante[calif_final] <= 20))</p> <p>3. PorcentajeEnRango = $\text{DIVIDE}([\text{TotalEstudiantesEnRango}], [\text{TotalEstudiantes}]) * 100$</p> <p>4. Trimestre = $\text{calendario}[\text{date}].[\text{Trimestre}]$</p>
--	---	--

5.5. Reportes y Visualizaciones

Se presenta un Dashboard donde visualizamos los resultados obtenidos con los datos proporcionados en las dos escuelas.



6. Resultados

6.1. Análisis del aporte a la toma de decisiones

En el contexto de Power BI, al explorar los datos, identificamos que dos instituciones educativas, Gabriel Pereira y Mousinho da Silveira, se destacan. Entre ambas, Gabriel Pereira presenta una matrícula estudiantil más amplia. Sin embargo, se evidencia un bajo rendimiento académico en ambas instituciones.

Particularmente, el análisis revela que la asignatura de portugués es la más demandada, contando con una cantidad significativamente mayor de estudiantes en comparación con matemáticas. La representación gráfica de líneas muestra que los estudiantes de edades entre 15 y 18 años experimentan tasas de fracaso más elevadas. En contraste, los estudiantes mayores de 18 años presentan índices de fracaso más bajos, posiblemente indicando una mayor madurez en su enfoque académico.

Un hallazgo relevante es que aquellos estudiantes que invierten tiempo en el estudio exhiben índices de fracaso más bajos. Aunque esta población es reducida, demuestra la importancia del compromiso con el aprendizaje.

En relación al rendimiento general de los estudiantes en ambas instituciones (un total de 1044 estudiantes), apenas 70 obtienen calificaciones sobresalientes, representando un modesto 6.70% del total. En el caso específico del Colegio Gabriel Pereira, con 772 estudiantes, solo 57 logran calificaciones excelentes, equivalente al 7.38% del total. Por su parte, el Colegio Mousinho da Silveira, que cuenta con 272 estudiantes, presenta solo 13 estudiantes con calificaciones destacadas, representando un 4.78% del total.

Toma de decisiones

Basándonos en el análisis realizado, se pueden proponer varias medidas para mejorar la situación académica en los colegios Gabriel Pereira y Mousinho da Silveira.

- **Programas de Apoyo Académico:** Implementar programas de apoyo académico dirigidos a estudiantes que presentan dificultades, especialmente en la materia de matemáticas. Esto podría incluir tutorías después de clases, sesiones de estudio en grupo y recursos en línea.
- **Enfoque en la Materia de Matemáticas:** Dado que la materia de Matemáticas no es la más demandada, considerar enriquecer el plan de estudios de

esta materia para hacerla más atractiva y efectiva en términos de enseñanza y aprendizaje.

- **Intervenciones para Grupos de Edad Vulnerables:** Puesto que se ha identificado un mayor índice de fracasos entre estudiantes de 15 a 18 años, implementar intervenciones específicas para este grupo. Esto podría involucrar orientación académica, consejería y programas para abordar posibles factores desencadenantes.
- **Incentivos para el Estudio:** Diseñar incentivos para motivar a los estudiantes a dedicar más tiempo al estudio. Esto podría incluir reconocimientos por logros académicos, programas de recompensas y actividades extracurriculares relacionadas con el aprendizaje.
- **Estrategias de Enseñanza Innovadoras:** Explorar enfoques de enseñanza innovadores que fomenten la participación activa de los estudiantes y abordan diversos estilos de aprendizaje. Esto podría incluir el uso de tecnología educativa, proyectos prácticos y metodologías colaborativas.
- **Monitoreo Continuo del Rendimiento:** Establecer sistemas de seguimiento continuo del rendimiento académico para identificar problemas en etapas tempranas y tomar medidas preventivas antes de que los estudiantes enfrenten dificultades mayores.
- **Colaboración con Padres y Tutores:** Involucrar a los padres y tutores en el proceso educativo. Mantener una comunicación abierta para que estén al tanto del progreso de sus hijos y puedan brindar apoyo en casa.
- **Diversificación de Estrategias de Evaluación:** Utilizar una variedad de métodos de evaluación para medir el progreso y el conocimiento de los estudiantes de manera más completa y precisa.
- **Formación Docente y Desarrollo Profesional:** Brindar capacitación y desarrollo profesional a los docentes para que puedan aplicar técnicas pedagógicas efectivas, adaptarse a las necesidades cambiantes de los estudiantes y promover un entorno de aprendizaje positivo.
- **Análisis Continuo y Ajustes:** Realizar análisis periódicos para evaluar la efectividad de las medidas implementadas y ajustar las estrategias según los resultados y las necesidades identificadas.