

Documentación para datamart basado en la Ejecución del Gasto de COVID-19 en el Perú (2020-presente)

Este proyecto de análisis de la ejecución del gasto de COVID-19 (2020-presente) busca identificar patrones y generar insights que permitan tomar decisiones estratégicas en cuanto al presupuesto público.

Cabe resaltar que la información es extraída de la plataforma de [datos abiertos](#) del Ministerio de Economía y Finanzas (MEF) del Gobierno Peruano.

1. Planificación del proyecto

a. Objetivos

- Evaluar la eficiencia en la ejecución del gasto destinado al COVID-19 en distintos niveles de gobierno.
- Identificar las regiones, sectores y programas que han recibido más recursos y analizar su impacto.
- Comparar el presupuesto inicial con el gasto ejecutado para determinar desviaciones y optimizar futuras asignaciones presupuestarias.

b. Alcances del proyecto

- Análisis de la ejecución del gasto público relacionado con el COVID-19 desde 2020 hasta 2025.
- Exploración de datos segmentados por nivel de gobierno, sector, pliego, entidad ejecutora y geografía.
- Creación de un Data Mart para optimizar la consulta y el análisis de datos en herramientas de BI como Power BI.
- Desarrollo de reportes que permitan visualizar la eficiencia del gasto público y la distribución de recursos.

2. Definición de requerimientos
- a. Requisitos del negocio

Identificador	R-01	Nombre	Evolución Anual del Gasto Ejecutado
Tipo	Funcional	Fecha	02-02-2025
Prioridad	Alta	Necesidad	Si
Descripción	Analiza la evolución del gasto anual desde 2020 hasta 2025, indicando la variación porcentual respecto al año anterior		
Datos dimensionales	Tiempo	Datos Hechos	Total de Monto Devengado por Año

Identificador	R-02	Nombre	Evolución Mensual del Gasto Ejecutado
Tipo	Funcional	Fecha	02-02-2025
Prioridad	Alta	Necesidad	Si
Descripción	Analiza la evolución mensual del gasto desde 2020 hasta 2025, indicando la variación porcentual respecto al mes anterior dentro de un Año seleccionado		
Datos dimensionales	Tiempo, Entidad, Ubicación	Datos Hechos	Total de Monto Devengado por Mes

Identificador	R-03	Nombre	Ejecutado vs Girado por Trimestre
Tipo	Funcional	Fecha	02-02-2025
Prioridad	Alta	Necesidad	Si
Descripción	Qué porcentaje del gasto comprometido realmente se ha girado y si hay retrasos en la ejecución presupuestaria (compromisos aún no pagados).		
Datos dimensionales	Tiempo, Entidad, Ubicación	Datos Hechos	Total de Monto Devengado y Monto Girado por Trimestre, Eficiencia Pago (Girado/Ejecutado)

Identificador	R-04	Nombre	Distribución Geográfica del Gasto Ejecutado
Tipo	Funcional	Fecha	02-02-2025
Prioridad	Alta	Necesidad	Si
Descripción	Identifica visualmente en un mapa geográfico las regiones con mayor y menor ejecución de gasto.		
Datos dimensionales	Tiempo	Datos Hechos	Total de Monto Devengado por Región

Identificador	R-05	Nombre	Distribución Geográfica del %Ejecutado
Tipo	Funcional	Fecha	02-02-2025
Prioridad	Alta	Necesidad	Si
Descripción	Identifica la distribución por región del %Ejecutado, por medio de un mapa de calor		
Datos dimensionales	Tiempo, Entidad, Ubicación	Datos Hechos	Porcentaje de Ejecución

Identificador	R-06	Nombre	Top 3 Departamentos con Mayor Gasto Ejecutado
Tipo	Funcional	Fecha	02-02-2025
Prioridad	Alta	Necesidad	Si
Descripción	Identifica los 3 departamentos con mayor gasto ejecutado		
Datos dimensionales	Tiempo, Entidad, Ubicación	Datos Hechos	Total de Monto Devengado por Región

Identificador	R-07	Nombre	Presupuesto vs Ejecución por Departamento
Tipo	Funcional	Fecha	02-02-2025
Prioridad	Alta	Necesidad	Si
Descripción	Muestra la comparación entre el Presupuesto Inicial Modificado (PIM) y el Monto Devengado por departamento		
Datos dimensionales	Tiempo, Entidad, Ubicación	Datos Hechos	Total de Presupuesto Inicial Y Monto Devengado por Región

Identificador	R-08	Nombre	Top 10 Sectores por Gasto Ejecutado
Tipo	Funcional	Fecha	02-02-2025
Prioridad	Alta	Necesidad	Si
Descripción	Identifica los 10 sectores con mayor Gasto Ejecutado		
Datos dimensionales	Tiempo, Entidad, Ubicación	Datos Hechos	Total de Monto Devengado por Entidad

Identificador	R-09	Nombre	Top 10 Pliegos por Gasto Ejecutado
Tipo	Funcional	Fecha	02-02-2025
Prioridad	Alta	Necesidad	Si
Descripción	Identifica los 10 pliegos con mayor Gasto Ejecutado		
Datos dimensionales	Tiempo, Entidad, Ubicación	Datos Hechos	Total de Monto Devengado por Entidad

Identificador	R-10	Nombre	Top 5 Programas Presupuestales por Gasto Ejecutado
Tipo	Funcional	Fecha	02-02-2025
Prioridad	Alta	Necesidad	Si
Descripción	Identifica los 5 programas presupuestales que han recibido más recursos a nivel de gasto ejecutado		
Datos dimensionales	Tiempo, Entidad, Actividad, Ubicación	Datos Hechos	Total de Montos por Actividad

Identificador	R-11	Nombre	Restante por Ejecutar
Tipo	Funcional	Fecha	02-02-2025
Prioridad	Alta	Necesidad	Si
Descripción	Cómo se compara el presupuesto inicial con el gasto ejecutado		
Datos dimensionales	Tiempo	Datos Hechos	Diferencia entre presupuesto inicial y ejecutado

Identificador	R-12	Nombre	Porcentaje de Ejecución
Tipo	Funcional	Fecha	02-02-2025
Prioridad	Alta	Necesidad	Si
Descripción	Qué tan eficiente ha sido la ejecución del gasto		
Datos dimensionales	Tiempo, Entidad, Ubicación	Datos Hechos	Porcentaje de ejecución del gasto (Monto Devengado / PIM)

Identificador	R-13	Nombre	Top 10 Departamentos por Cantidad de Proyectos Únicos
Tipo	Funcional	Fecha	02-02-2025
Prioridad	Alta	Necesidad	Si
Descripción	Identifica los 10 departamentos con la mayor cantidad de Proyectos/Productos únicos		
Datos dimensionales	Tiempo, Entidad, Actividad, Ubicación	Datos Hechos	Recuento único de Actividad por Región

Identificador	R-14	Nombre	Resumen de Gasto Ejecutado
Tipo	Funcional	Fecha	02-02-2025
Prioridad	Alta	Necesidad	Si
Descripción	Qué productos o proyectos han recibido más recursos		
Datos dimensionales	Tiempo, Entidad, Actividad, Ubicación	Datos Hechos	Total de Montos por Actividad

Identificador	R-15	Nombre	Gasto Ejecutado por Nivel de Gobierno
Tipo	Funcional	Fecha	02-02-2025
Prioridad	Alta	Necesidad	Si
Descripción	Analiza el total de gasto ejecutado por Nivel de Gobierno		
Datos dimensionales	Tiempo	Datos Hechos	Total de Monto Devengado por Entidad

b. Información documentada

Cada año cuenta con un archivo CSV que almacena los datos de ejecución del gasto público. Los archivos han sido unificados en un solo conjunto de datos(dataframe) para su análisis.

Reporte de ejecución del gasto de COVID 2025

ANO_EJE	MES_EJE	NIVEL_GOB	NIVEL_GOB_NOMBRE	SECTOR	SECTOR_NOMBRE	PLIEGO	PLIEGO_NOMBRE	SEC_EJEC	EJECUTORA	...
0	2025	0	E GOBIERNO NACIONAL	10	EDUCACION	10	M. DE EDUCACION	1438	117	...
1	2025	1	E GOBIERNO NACIONAL	10	EDUCACION	10	M. DE EDUCACION	1438	117	...
2	2025	0	E GOBIERNO NACIONAL	10	EDUCACION	550	U.N. AUTONOMA DE CHOTA	1375	1	...
3	2025	0	E GOBIERNO NACIONAL	10	EDUCACION	542	U.N. INTERCULTURAL DE LA AMAZONIA	1200	1	...
4	2025	0	E GOBIERNO NACIONAL	10	EDUCACION	555	U.N. INTERCULTURAL FABIOLA SALAZAR LEGUIA DE B...	1533	1	...

c. Diccionario de datos

En la misma fuente de archivos se cuenta con el Diccionario de datos en formato CSV, por lo que se usó Pandas en Python para visualizar las características de cada dato en un dataframe.

Diccionario de datos del Gasto Covid 19

	VARIABLE	TIPO_DATO	DESCRIPCION
0	ANO_EJE	Numérico	Año de ejecución del presupuesto.
1	MES_EJE	Numérico	Mes de ejecución del presupuesto.
2	NIVEL_GOBIERNO	Carácter	Código (letra) que identifica el Nivel de Gobi...
3	NIVEL_GOBIERNO_NOMBRE	Carácter	Descripción de Nivel de Gobierno: Nacional, Re...
4	SECTOR	Carácter	Código de Sector al que pertenece la Entidad.
5	SECTOR_NOMBRE	Carácter	Descripción de código del Sector al que pert...
6	PLIEGO	Carácter	Código de Pliego al que pertenece la Entidad.
7	PLIEGO_NOMBRE	Carácter	Descripción de código del Pliego al que pert...
8	SEC_EJEC	Carácter	Código que identifica a una Entidad.
9	EJECUTORA	Carácter	Código de la cadena institucional que identifi...
10	EJECUTORA_NOMBRE	Carácter	Nombre de la Entidad.
11	DEPARTAMENTO_EJECUTORA	Carácter	Código de departamento donde se ubica la Entidad.
12	DEPARTAMENTO_EJECUTORA_NOMBRE	Carácter	Nombre de departamento donde se ubica la Entidad.

En este caso, el total de variables de la fuente original llegaba a la cifra de 62, variando entre tipo de dato Carácter y Numérico. Sin embargo, como el presente proyecto se enfoca en una análisis general y no específico, se seleccionaron determinadas columnas.

Asimismo, se eliminaron las variables que no son necesarias para el análisis. Por ejemplo, los códigos identificadores, ya que se dispone del nombre y el mismo script de base de datos que se ejecutará ya asigna sus propios IDs por dimensión.

A continuación, se presenta la descripción de las variables seleccionadas:

1. **ANO_EJE** (Año de ejecución)
2. **MES_EJE** (Mes de ejecución)
3. **NIVEL_GOBIERNO_NOMBRE** (Nacional, Regional, Local)
4. **SECTOR_NOMBRE** (Sector del gasto, como Salud, Transporte, etc.)
5. **PLIEGO_NOMBRE** (Entidad ejecutora del gasto, ej. Ministerio de Salud)
6. **EJECUTORA_NOMBRE** (Unidad ejecutora específica)
7. **DEPARTAMENTO_EJECUTORA_NOMBRE** (Departamento donde se ubica la entidad que ejecutó el gasto)
8. **PROVINCIA_EJECUTORA_NOMBRE** (Provincia donde se ubica la entidad que ejecutó el gasto)

9. **DISTRITO_EJECUTORA_NOMBRE** (Distrito donde se ubica la entidad que ejecutó el gasto)
10. **PROGRAMA_PPTO_NOMBRE** (Nombre del Programa presupuestal)
11. **TIPO_ACT_PROY_NOMBRE** ()
12. **ACTIVIDAD_ACCION_OBRA_NOMBRE** ()
13. **PRODUCTO_PROYECTO_NOMBRE** (Producto o proyecto específico)
14. **MONTO_PIA** (Presupuesto Inicial Anual)
15. **MONTO_PIM** (Presupuesto Inicial Modificado)
16. **MONTO_CERTIFICADO** (Monto certificado para ejecución)
17. **MONTO_COMPROMETIDO** (Monto comprometido en contratos)
18. **MONTO_DEVENGADO** (Monto realmente pagado)
19. **MONTO_GIRADO** (Monto transferido a proveedores)

d. Modelado Dimensional

Después de seleccionar las variables que permiten atender los requerimientos iniciales, se plantearon las siguientes dimensiones a fin de seguir un modelo de datos estrella:

- Dim_Tiempo: Año, Mes.
- Dim_Ubicacion: Departamento, Provincia, Distrito de la entidad ejecutora.
- Dim_Entidad: Nivel de Gobierno, Sector, Pliego, Unidad Ejecutora.
- Dim_Actividad: Programa Presupuestal, Tipo de Actividad, Producto/Proyecto, Actividad/Acción/Obra.

e. Diseño físico



f. Tablas del modelo dimensional

Tabla	Descripción
Hechos	Contiene los montos financieros del gasto, con claves foráneas que la relacionan con las dimensiones.
Dim_Tiempo	Define el período en el que se ejecuta el gasto, permitiendo análisis temporales.
Dim_Ubicacion	Contiene la información de la ubicación de la entidad ejecutora del gasto.
Dim_Entidad	Representa a las entidades responsables de ejecutar el gasto.
Dim_Actividad	Contiene información sobre los programas y actividades en los que se ha asignado el presupuesto.

Tabla Hechos: Hechos

Hechos está conformada por:

Nombre Columna	Tipo de Dato	Descripción
id_hecho	int (PK)	Clave primaria de la tabla hechos
id_tiempo	int (FK)	Clave foránea que relaciona con la dimensión de tiempo.
id_ubicacion	int (FK)	Clave foránea que relaciona con la dimensión de ubicación.
id_entidad	int (FK)	Clave foránea que relaciona con la dimensión de entidad.
id_actividad	int (FK)	Clave foránea que relaciona con la dimensión de actividad.
MONTO_PIA	float	Monto del Presupuesto Institucional de Apertura.
MONTO_PIM	float	Monto del Presupuesto Institucional Modificado.
MONTO_CERTIFICADO	float	Monto utilizado en la fase de Certificación.
MONTO_COMPROMETIDO	float	Monto utilizado en la fase de Compromiso.
MONTO_DEVENGADO	float	Monto ejecutado en la fase de Devengado.
MONTO_GIRADO	float	Monto ejecutado en la fase de Girado.

Dimensión ubicación: Dim_Ubicacion

La dimensión Dim_Ubicacion está conformado por:

Nombre Columna	Tipo de Dato	Descripción
id_ubicacion	int (PK)	Clave foránea que relaciona con la dimensión de ubicación.
DEPARTAMENTO_EJECUTORA_NOMBRE	nvarchar(255)	Nombre de departamento donde se ubica la Entidad.
PROVINCIA_EJECUTORA_NOMBRE	nvarchar(255)	Nombre de provincia del departamento donde se ubica la
DISTRITO_EJECUTORA_NOMBRE	nvarchar(255)	Nombre de distrito de la provincia del departamento donde se ubica la

Dimensión ubicación: Dim_Tiempo

La dimensión Dim_Tiempo está conformado por:

Nombre Columna	Tipo de Dato	Descripción
id_tiempo	int (PK)	Clave foránea que relaciona con la dimensión de tiempo.
ANO_EJE	int	Año de ejecución del presupuesto.
MES_EJE	int	Mes de ejecución del presupuesto.

Dimensión ubicación: Dim_Entidad

La dimensión Dim_Entidad está conformado por:

Nombre Columna	Tipo de Dato	Descripción
id_entidad	int (PK)	Clave foránea que relaciona con la dimensión de entidad.
NIVEL_GOBIERNO_NOMBRE	nvarchar(255)	Clave foránea que relaciona con la dimensión de entidad.
SECTOR_NOMBRE	nvarchar(255)	Descripción de código del Sector al que pertenece la Entidad.
PLIEGO_NOMBRE	nvarchar(255)	Descripción de código del Pliego al que pertenece la Entidad.
EJECUTORA_NOMBRE	nvarchar(255)	Nombre de la Entidad.

Dimensión ubicación: Dim_Actividad

La dimensión Dim_Actividad está conformado por:

Nombre Columna	Tipo de Dato	Descripción
id_actividad	int (PK)	Clave foránea que relaciona con la dimensión de actividad.
PROGRAMA_PPTO_NOMBRE	nvarchar(255)	Descripción del Programa Presupuestal.
TIPO_ACT_PROY_NOMBRE	nvarchar(255)	Nombre que muestra si es proyecto o producto.
PRODUCTO_PROYECTO_NOMBRE	nvarchar(MAX)	Descripción de un proyecto o producto.
ACTIVIDAD_ACCION_OBRA_NOMBRE	nvarchar(255)	Descripción de una Actividad/Acción de inversión u