
Consulta SQL

Actividad 3 – Análisis del Impacto de los Días Festivos em las Ventas Semanales
(CONSULTAS_SQL_ACT3)

Erika Samara Alvares Angelim
Guillermo Andrés Fernández Olivares

Ingeniería para el Procesado Masivo de Datos
27 de junio de 2025

Asignatura	Datos del alumno	Fecha
Ingeniería para el Procesado Masivo de Datos	Apellidos: Alvares Angelim	27/06/2025
	Nombre: Erika Samara	

Índice

Documentación Técnica.....	2
1. Subida de datos a GCP	2
Captura 1.1: Subida de archivos CSV (features y sales) a Cloud Storage (Bucket).....	2
2. Creación de tablas en BigQuery	2
Captura 2.1: Creación tabla nativa (stores)	2
Captura 2.2: Importación de las tablas nativas (features y sales) a BigQuery.....	2
3. Exploración inicial	3
Estadística Descriptiva de cada tabla.....	3
4. Join y preparación de tabla integrada	5
Captura 4.1: Muestra de la tabla join_ventas_completa (tablas: features, sales y stores).....	5
Captura 4.2: Resumen Estadístico: (JOIN entre las tablas: features, sales y stores)7	
6. Vista final limpia para Looker Studio.....	10
Desarrollo del dashboard	10
Justificación técnica de las visualizaciones elegidas	11

Asignatura	Datos del alumno	Fecha
Ingeniería para el Procesado Masivo de Datos	Apellidos: Alvares Angelim	27/06/2025
	Nombre: Erika Samara	

Documentación Técnica

1. Subida de datos a GCP

Captura 1.1: Subida de archivos CSV (features y sales) a Cloud Storage (Bucket)

Al subir los archivos en formato .csv, se ha creado de forma automática las dos tablas nativas.

The screenshot shows the 'Crear tabla' (Create table) interface in BigQuery. The 'Fuente' (Source) section is set to 'Google Cloud Storage' with the path 'ea-unir-dataproc-bucket/features.csv'. The 'Destino' (Destination) section shows the project 'fiery-cabinet-461611-e8' and the dataset 'grupo01'. The 'Tipo de tabla' (Table type) is set to 'Tabla nativa' (Native table). The 'Esquema' (Schema) section shows 'Detección automática' (Automatic detection) selected.

2. Creación de tablas en BigQuery

Captura 2.1: Creación tabla nativa (stores)

Presenta las primeras 5 líneas.

The screenshot shows the 'Resultados de la consulta' (Query results) interface. The 'Tabla de destino' (Destination table) is set to 'Tabla temporal' (Temporary table). The 'Tabla nativa' (Native table) is highlighted. The 'Resultados de la consulta' (Query results) section shows the first 5 rows of the 'stores' table.

Fila	Store	Size	Size
1	33	A	
2	36	A	
3	1	A	
4	26	A	
5	8	A	

Captura 2.2: Importación de las tablas nativas (features y sales) a BigQuery

Asignatura	Datos del alumno	Fecha
Ingeniería para el Procesado Masivo de Datos	Apellidos: Alvares Angelim	27/06/2025
	Nombre: Erika Samara	

Tabla: features

Presenta las primeras 5 líneas.

The screenshot shows the Google Cloud BigQuery console. The left sidebar displays the project structure with 'features' selected under 'grupo01'. The main editor shows a SQL query to select the first 5 rows of the 'features' table. The results pane shows the query execution details and a table of results.

Query:

```

1 -- paso_2_0_preview_features
2
3 SELECT *
4 FROM `fjery-cabinet-461611-e8.grupo01.features`
5 LIMIT 5

```

Results:

Fila	Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
1	1	2010-02-05	42.31	2.572	NA	NA	NA	NA	NA	211.0963582	8.106	false
2	1	2010-02-12	38.51	2.548	NA	NA	NA	NA	NA	211.2421698	8.106	true
3	1	2010-02-19	39.93	2.514	NA	NA	NA	NA	NA	211.2891429	8.106	false
4	1	2010-02-26	46.63	2.561	NA	NA	NA	NA	NA	211.3196429	8.106	false
5	1	2010-03-05	46.5	2.625	NA	NA	NA	NA	NA	211.3501429	8.106	false

Tabla: sales

Presenta las primeras 5 líneas.

The screenshot shows the Google Cloud BigQuery console. The left sidebar displays the project structure with 'sales' selected under 'grupo01'. The main editor shows a SQL query to select the first 5 rows of the 'sales' table. The results pane shows the query execution details and a table of results.

Query:

```

1 -- paso_2_0_preview_sales
2
3 SELECT *
4 FROM `fjery-cabinet-461611-e8.grupo01.sales`
5 LIMIT 5

```

Results:

Fila	Store	Dept	Date	Weekly_Sales	IsHoliday
1	1	1	2010-02-05	24924.5	false
2	1	1	2010-02-12	46039.49	true
3	1	1	2010-02-19	41595.55	false
4	1	1	2010-02-26	19403.54	false
5	1	1	2010-03-05	21827.9	false

3. Exploración inicial

Estadística Descriptiva de cada tabla

Tabla: features

Asignatura	Datos del alumno	Fecha
Ingeniería para el Procesado Masivo de Datos	Apellidos: Alvares Angelim	27/06/2025
	Nombre: Erika Samara	

Resultados de la consulta

[Guardar los resultados](#) [Abrir en](#)

Información del trabajo		Resultados		Gráfico	JSON	Detalles de la ejecución		Gráfico de ejecución		
Fila	variable	count	mean	std	min	25%	50%	75%	max	
1	MarkDown2	2921	3384.1765936323168	8793.5830155076419	-265.76	69.05	358.76	2131.62	104519.54	
2	MarkDown4	3464	3292.9358862586641	6792.3298613476481	0.22	305.89	1169.95	3292.07	67474.85	
3	MarkDown3	3613	1760.1001799058965	11276.46220802296	-179.26	6.68	35.76	163.96	149483.31	
4	MarkDown1	4032	7032.3717857142783	9262.7474475236741	-2781.45	1601.72	4718.7	8891.08	103184.98	
5	MarkDown5	4050	4132.216422222118	13086.690278200189	-185.17	1440.44	2722.29	4857.19	771448.1	
6	Unemployment	7605	7.8268210387902624	1.8772585939174304	3.684	6.635	7.806	8.567	14.313	
7	CPI	7605	172.46080918276158	39.7383460986084	126.064	132.3200714	182.6696737	213.9324122	228.9764563	
8	Fuel_Price	8190	3.4059918192918115	0.4313365711007135	2.472	3.045	3.521	3.742	4.468	
9	Temperature	8190	59.356197802197791	18.678606848907144	-7.29	45.88	60.59	73.87	101.95	

Resultados por página: 50 1 - 9 de 9

Tabla: sales

Resultados de la consulta

[Guardar los resultados](#) [Abrir en](#)

Información del trabajo		Resultados	Gráfico	JSON	Detalles de la ejecución	Gráfico de ejecución			
Fila	variable	count	mean	std	min	25%	50%	75%	max
1	Weekly_Sales	421570	15981.25812346748	22711.183519163009	-4988.94	2086.56	7622.31	20161.59	693099.36

Resultados por página: 50 1 - 1 de 1

Tabla: stores

Resultados de la consulta

[Guardar los resultados](#) [Abrir en](#)

Información del trabajo			Resultados	Gráfico	JSON	Detalles de la ejecución			Gráfico de ejecución		
Fila	variable	count	mean	std	min	25%	50%	75%	max		
1	Size	45	130287.59999999998	63825.271991159214	34875	70713	126512	202307	219622		

Resultados por página: 50 1 - 1 de 1

Estudio de las categorías: ¿Cuántas categorías hay en cada tabla?

Asignatura	Datos del alumno	Fecha
Ingeniería para el Procesado Masivo de Datos	Apellidos: Alvares Angelim	27/06/2025
	Nombre: Erika Samara	

Google Cloud My First Project

Buscar (/) recursos, documentos, productos y más

Explorador + Agregar datos

Act3 (Capturas de pantalla)

```
101
102 /* Estudiar las diferentes categorías de las principales variables
103 categóricas y el número de filas correspondientes a cada categoría.
104 Indicar las conclusiones en el informe y dar una interpretación de
105 qué están indicando */
106
107 -- SALES (número de filas por cada categoría perteneciente a la variable)
108 SELECT store, IsHoliday, COUNT(*) AS total_weeks_on_sale
109 FROM `vernal-landing-455321-v0.IPMD_Actividad_3.sales`
110 GROUP BY store, IsHoliday
111 ORDER BY 1,3 desc;
112
113 -- STORES (número de categorías por cada variable)
114 SELECT
115   COUNT(DISTINCT store) AS total_stores,
116   COUNT(DISTINCT Type) AS total_types
117 FROM `vernal-landing-455321-v0.IPMD_Actividad_3.stores`;
118
119 -- STORES (número de filas por cada categoría perteneciente a la variable)
120 SELECT
121   store, COUNT(*) AS total
122 FROM `vernal-landing-455321-v0.IPMD_Actividad_3.stores`
123 GROUP BY store ORDER BY store;
124
125 SELECT
126   Type, COUNT(*) AS total
127 FROM `vernal-landing-455321-v0.IPMD_Actividad_3.stores`
128 GROUP BY Type ORDER BY Type;
129
```

Esta secuencia de comandos procesará 34.72 MB cuando se ejecute.

Resultados de la consulta

Repositorio Versión preliminar Historial de trabajos

4. Join y preparación de tabla integrada

Captura 4.1: Muestra de la tabla join_ventas_completa (tablas: features, sales y stores)

paso_3_0_join_ventas																
Se completó la consulta																
Resultados de la consulta																
Información del trabajo Resultados Gráfico JSON Detalles de la ejecución Gráfico de ejecución																
Fila	Store	Dept	Date	Weekly_Sales	IsHoliday	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPX	Unemployment	Store_Type	
1	1	1	2010-02-05	24924.5	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
2	1	2	2010-02-05	50605.27	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
3	1	3	2010-02-05	13740.12	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
4	1	4	2010-02-05	39954.04	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
5	1	5	2010-02-05	32229.38	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
6	1	6	2010-02-05	5749.03	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
7	1	7	2010-02-05	21084.08	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
8	1	8	2010-02-05	40129.01	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
9	1	9	2010-02-05	16930.99	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
10	1	10	2010-02-05	30723.5	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
11	1	11	2010-02-05	26215.18	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
12	1	12	2010-02-05	5449.54	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
13	1	13	2010-02-05	41869.29	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
14	1	14	2010-02-05	15466.91	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
15	1	15	2010-02-05	10217.55	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
16	1	16	2010-02-05	13223.76	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
17	1	17	2010-02-05	4729.5	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
18	1	18	2010-02-05	1947.05	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
19	1	19	2010-02-05	5034.1	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
20	1	20	2010-02-05	8907.63	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
21	1	21	2010-02-05	13623.98	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
22	1	22	2010-02-05	24146.49	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
23	1	23	2010-02-05	8272.9	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
24	1	24	2010-02-05	11609.5	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
25	1	25	2010-02-05	11737.12	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A
26	1	26	2010-02-05	2293.0	false	42.31	2.572	null	null	null	null	null	null	211.0963582	8.106	A

Asignatura	Datos del alumno	Fecha
Ingeniería para el Procesado Masivo de Datos	Apellidos: Alvares Angelim	27/06/2025
	Nombre: Erika Samara	

27	1	28	2010-02-05	1085.29	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
28	1	29	2010-02-05	7024.95	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
29	1	30	2010-02-05	5491.0	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
30	1	31	2010-02-05	3455.92	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
31	1	32	2010-02-05	11875.84	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
32	1	33	2010-02-05	6589.77	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
33	1	34	2010-02-05	14350.83	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
34	1	35	2010-02-05	2770.0	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
35	1	36	2010-02-05	120.0	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
36	1	37	2010-02-05	3438.16	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
37	1	38	2010-02-05	115564.35	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
38	1	40	2010-02-05	66780.63	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
39	1	41	2010-02-05	1011.83	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
40	1	42	2010-02-05	8366.71	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
41	1	44	2010-02-05	4674.75	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
42	1	45	2010-02-05	37.44	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
43	1	46	2010-02-05	20837.77	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
44	1	47	2010-02-05	70.0	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
45	1	48	2010-02-05	814.0	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
46	1	49	2010-02-05	11501.46	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
47	1	51	2010-02-05	1.16	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
48	1	52	2010-02-05	3508.04	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
49	1	54	2010-02-05	459.0	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A
50	1	55	2010-02-05	21249.31	false	42.31	2.572	null	null	null	null	211.0963582	8.106	A

Código SQL: script JOIN (features, sales, stores)

Ejemplo 1 de JOIN entre las tablas

```

2
3
4
5 /*
6 Este bloque SQL realiza la unión de las tres tablas principales del proyecto (sales, features, stores)
7 en una única tabla combinada, aplicando condiciones de relación adecuadas.
8 */
9
10 SELECT
11 -- Campos base de ventas
12 s.Store,
13 s.Dept,
14 s.Date,
15 s.Weekly_Sales,
16 s.IsHoliday,
17 -- Variables externas (features), sin duplicar Store y Date
18 f.Temperature,
19 f.Fuel_Price,
20 SAFE_CAST(f.MarkDown1 AS FLOAT64) AS MarkDown1,
21 SAFE_CAST(f.MarkDown2 AS FLOAT64) AS MarkDown2,
22 SAFE_CAST(f.MarkDown3 AS FLOAT64) AS MarkDown3,
23 SAFE_CAST(f.MarkDown4 AS FLOAT64) AS MarkDown4,
24 SAFE_CAST(f.MarkDown5 AS FLOAT64) AS MarkDown5,
25 SAFE_CAST(f.CPI AS FLOAT64) AS CPI,
26 SAFE_CAST(f.Unemployment AS FLOAT64) AS Unemployment,
27 -- Información de tienda
28 st.Type AS Store_Type,
29 st.Size AS Store_Size
30 FROM `fiery-cabinet-461611-e8.grupo01.sales` s
31 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.features` f
32 ON s.Store = f.Store AND s.Date = f.Date
33 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.stores` st
34 ON s.Store = st.Store
35 ORDER BY s.Store, s.Date;
36
37

```

Esta consulta procesará 13.89 MB cuando se ejecute.

Ejemplo 2 de JOIN entre las tablas (Agregación)

```

150
151
152 /* Obtener cualquier agregación (medias, sumas, desviaciones típicas...) que el
153 equipo considere relevante desde una perspectiva de negocio, guardarla en
154 vistas y explicar en el informe el motivo y las conclusiones que se deducen. */
155 CREATE OR REPLACE VIEW
156 IPMD_Actividad_3.historico_ventas_por_store AS
157 SELECT DISTINCT
158 Store, EXTRACT(YEAR FROM Date) as year, EXTRACT(MONTH FROM Date) as month, ROUND(SUM(Weekly_Sales),2) AS total_weekly_sales
159 FROM
160 (SELECT
161 s.Store, s.Dept, s.Date, s.Weekly_Sales,
162 s.IsHoliday, f.Temperature, f.Fuel_Price,
163 SAFE_CAST(f.MarkDown1 AS FLOAT64) AS MarkDown1,
164 SAFE_CAST(f.CPI AS FLOAT64) AS CPI,
165 SAFE_CAST(f.Unemployment AS FLOAT64) AS Unemployment,
166 st.Type AS Store_Type,
167 st.Size AS Store_Size
168 FROM `vernal-landing-455321-v0.IPMD_Actividad_3.sales` s
169 LEFT JOIN `vernal-landing-455321-v0.IPMD_Actividad_3.features` f ON s.Store = f.Store AND s.Date = f.Date
170 LEFT JOIN `vernal-landing-455321-v0.IPMD_Actividad_3.stores` st ON s.Store = st.Store)
171 GROUP BY Store, EXTRACT(YEAR FROM Date), EXTRACT(MONTH FROM Date);
172
173
174

```

Esta secuencia de comandos procesará 34.72 MB cuando se ejecute.

Resultados de la consulta

Historial de trabajos

Asignatura	Datos del alumno	Fecha
Ingeniería para el Procesado Masivo de Datos	Apellidos: Alvares Angelim	27/06/2025
	Nombre: Erika Samara	

Ejemplo 3 de JOIN entre las tablas

Google Cloud

My First Project

Buscar (/) recursos, documentos, productos y más

Explorador

Mostrar solo los destacados

Repositorios

Consultas

Consultas compartidas

Act3 (Capturas de pantalla)

Notebooks

Lienzos de datos

Preparaciones de datos

Canalizaciones

Conexiones externas

IPMD_Actividad_3

features

historico_ventas_por_st...

historico_ventas_por_ti...

sales

Act3 (Capturas de pantalla)

Ejecutar

Abrir en

Más

Guardar consulta

Descargar

```
178
179
180 /* Obtener cualquier agrupación y agregación (cualquier agregación aplicada
181 en diferentes grupos) que el equipo considere relevante desde una perspectiva
182 de negocio, guardarlas en vistas y explicar en el informe el motivo y las
183 conclusiones que se deducen. */
184 CREATE OR REPLACE VIEW
185 IPMD_Actividad_3.historico_ventas_por_tipo_store AS
186 SELECT DISTINCT
187 Store_Type,IsHoliday,EXTRACT(YEAR FROM Date) as year,
188 ROUND(SUM(Weekly_Sales),2) AS total_weekly_sales
189 FROM
190 (SELECT
191 s.Store,s.Dept,s.Date,s.Weekly_Sales,
192 s.IsHoliday,f.Temperature,f.Fuel_Price,
193 SAFE_CAST(f.MarkDown1 AS FLOAT64) AS MarkDown1,
194 SAFE_CAST(f.CPI AS FLOAT64) AS CPI,
195 SAFE_CAST(f.Unemployment AS FLOAT64) AS Unemployment,
196 st.Type AS Store_Type,
197 st.Size AS Store_Size
198 FROM `vernal-landing-455321-v0.IPMD_Actividad_3.sales` s
199 LEFT JOIN `vernal-landing-455321-v0.IPMD_Actividad_3.features` f ON s.Store = f.Store AND s.Date = f.Date
200 LEFT JOIN `vernal-landing-455321-v0.IPMD_Actividad_3.stores` st ON s.Store = st.Store)
201 GROUP BY Store_Type, IsHoliday, EXTRACT(YEAR FROM Date);
202
```

Resultados de la consulta

Historial de trabajos

Repositorio

Versión preliminar

Captura 4.2: Resumen Estadístico: (JOIN entre las tablas: features, sales y stores)

Resultados de la consulta

Guardar los resultados

Abrir en

Información del trabajo

Resultados

Gráfico

JSON

Detalles de la ejecución

Gráfico de ejecución

Fila	variable	count	nulos	mean	std	min	25%	50%	75%	max
1	CPI	421570	0	171.20194682296682	39.1592756230134	126.064	132.0644333	182.598178	212.0769346	227.2328068
2	Dept	421570	0	44.260317385012577	30.492054015785879	1.0	18.0	38.0	72.0	99.0
3	Fuel_Price	421570	0	3.361026527029896	0.45851453712775336	2.472	2.931	3.459	3.74	4.468
4	MarkDown1	150681	270889	7246.4201959105076	8291.2213454748	0.27	2228.83	5307.81	9198.08	88646.76
5	MarkDown2	111248	310322	3334.6286210988164	9475.3573254581243	-265.76	42.04	190.49	1891.53	104519.54
6	MarkDown3	137091	284479	1439.4213841900437	9623.07829031279	-29.1	5.17	24.94	104.01	141630.61
7	MarkDown4	134967	286603	3383.1682560922263	6292.3840308699891	0.22	499.0	1482.03	3554.45	67474.85
8	MarkDown5	151432	270138	4628.9750791773713	5962.8874552546986	135.16	1872.37	3358.94	5539.68	108519.28
9	Store_Size	421570	0	136727.91573878669	60980.583328101231	34875.0	93638.0	140167.0	202505.0	219622.0
10	Temperature	421570	0	60.090058732831146	18.447931147616043	-2.06	46.54	61.97	74.26	100.14
11	Unemployment	421570	0	7.9602886946413705	1.8632960384302153	3.879	6.891	7.866	8.572	14.313
12	Weekly_Sales	421570	0	15981.25812346748	22711.183519163009	-4988.94	2086.56	7622.31	20161.59	693099.36

Código SQL: script que ha generado el Resumen Estadístico

Asignatura	Datos del alumno	Fecha
Ingeniería para el Procesado Masivo de Datos	Apellidos: Alvares Angelim	27/06/2025
	Nombre: Erika Samara	

```
1  -- paso_3_1.resumen.estadistico_union
2
3  /*
4  En BigQuery, aunque ya tengamos una consulta guardada como paso_3_8_join_ventas,
5  no podemos referenciarla directamente como si fuera una tabla dentro de otra consulta.
6  Porque las consultas guardadas no crean objetos persistentes como tablas o vistas.
7  Son solo scripts reutilizables, no datasets almacenados.
8
9  Este código devuelve el resumen estadístico de 5 variables numéricas
10 tras el JOIN de las 3 tablas, manteniendo todo intacto.
11 */
12
13 -- paso_3_1.resumen.estadistico_union
14 -- paso_3_1.resumen.estadistico_union
15
16 -- Weekly_Sales
17 SELECT 'Weekly_Sales' AS variable,
18 COUNT(s.Weekly_Sales) AS count,
19 COUNTIF(s.Weekly_Sales IS NULL) AS nulos,
20 AVG(s.Weekly_Sales) AS mean,
21 STDDEV(s.Weekly_Sales) AS std,
22 MIN(s.Weekly_Sales) AS min,
23 APPROX_QUANTILES(s.Weekly_Sales, 4)[OFFSET(1)] AS '25%',
24 APPROX_QUANTILES(s.Weekly_Sales, 4)[OFFSET(2)] AS '50%',
25 APPROX_QUANTILES(s.Weekly_Sales, 4)[OFFSET(3)] AS '75%',
26 MAX(s.Weekly_Sales) AS max
27 FROM `fiery-cabinet-461611-e8.grupo01.sales` s
28 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.features` f ON s.Store = f.Store AND s.Date = f.Date
29 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.stores` st ON s.Store = st.Store
30
31 UNION ALL
32
```

```
33 -- Dept
34 SELECT 'Dept' AS variable,
35 COUNT(s.Dept) AS count,
36 COUNTIF(s.Dept IS NULL) AS nulos,
37 AVG(s.Dept) AS mean,
38 STDDEV(s.Dept) AS std,
39 MIN(s.Dept) AS min,
40 APPROX_QUANTILES(s.Dept, 4)[OFFSET(1)] AS '25%',
41 APPROX_QUANTILES(s.Dept, 4)[OFFSET(2)] AS '50%',
42 APPROX_QUANTILES(s.Dept, 4)[OFFSET(3)] AS '75%',
43 MAX(s.Dept) AS max
44 FROM `fiery-cabinet-461611-e8.grupo01.sales` s
45 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.features` f ON s.Store = f.Store AND s.Date = f.Date
46 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.stores` st ON s.Store = st.Store
47
48 UNION ALL
49
50 -- Temperature
51 SELECT 'Temperature' AS variable,
52 COUNT(SAFE_CAST(f.Temperature AS FLOAT64)) AS count,
53 COUNTIF(SAFE_CAST(f.Temperature AS FLOAT64) IS NULL) AS nulos,
54 AVG(SAFE_CAST(f.Temperature AS FLOAT64)) AS mean,
55 STDDEV(SAFE_CAST(f.Temperature AS FLOAT64)) AS std,
56 MIN(SAFE_CAST(f.Temperature AS FLOAT64)) AS min,
57 APPROX_QUANTILES(SAFE_CAST(f.Temperature AS FLOAT64), 4)[OFFSET(1)] AS '25%',
58 APPROX_QUANTILES(SAFE_CAST(f.Temperature AS FLOAT64), 4)[OFFSET(2)] AS '50%',
59 APPROX_QUANTILES(SAFE_CAST(f.Temperature AS FLOAT64), 4)[OFFSET(3)] AS '75%',
60 MAX(SAFE_CAST(f.Temperature AS FLOAT64)) AS max
61 FROM `fiery-cabinet-461611-e8.grupo01.sales` s
62 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.features` f ON s.Store = f.Store AND s.Date = f.Date
63 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.stores` st ON s.Store = st.Store
64
65 UNION ALL
66
```

```
66 -- Fuel_Price
67 SELECT 'Fuel_Price' AS variable,
68 COUNT(SAFE_CAST(f.Fuel_Price AS FLOAT64)) AS count,
69 COUNTIF(SAFE_CAST(f.Fuel_Price AS FLOAT64) IS NULL) AS nulos,
70 AVG(SAFE_CAST(f.Fuel_Price AS FLOAT64)) AS mean,
71 STDDEV(SAFE_CAST(f.Fuel_Price AS FLOAT64)) AS std,
72 MIN(SAFE_CAST(f.Fuel_Price AS FLOAT64)) AS min,
73 APPROX_QUANTILES(SAFE_CAST(f.Fuel_Price AS FLOAT64), 4)[OFFSET(1)] AS '25%',
74 APPROX_QUANTILES(SAFE_CAST(f.Fuel_Price AS FLOAT64), 4)[OFFSET(2)] AS '50%',
75 APPROX_QUANTILES(SAFE_CAST(f.Fuel_Price AS FLOAT64), 4)[OFFSET(3)] AS '75%',
76 MAX(SAFE_CAST(f.Fuel_Price AS FLOAT64)) AS max
77 FROM `fiery-cabinet-461611-e8.grupo01.sales` s
78 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.features` f ON s.Store = f.Store AND s.Date = f.Date
79 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.stores` st ON s.Store = st.Store
80
81 UNION ALL
82
83 -- MarkDown1 a MarkDown5
84 SELECT 'MarkDown1' AS variable,
85 COUNT(SAFE_CAST(f.MarkDown1 AS FLOAT64)) AS count,
86 COUNTIF(SAFE_CAST(f.MarkDown1 AS FLOAT64) IS NULL) AS nulos,
87 AVG(SAFE_CAST(f.MarkDown1 AS FLOAT64)) AS mean,
88 STDDEV(SAFE_CAST(f.MarkDown1 AS FLOAT64)) AS std,
89 MIN(SAFE_CAST(f.MarkDown1 AS FLOAT64)) AS min,
90 APPROX_QUANTILES(SAFE_CAST(f.MarkDown1 AS FLOAT64), 4)[OFFSET(1)] AS '25%',
91 APPROX_QUANTILES(SAFE_CAST(f.MarkDown1 AS FLOAT64), 4)[OFFSET(2)] AS '50%',
92 APPROX_QUANTILES(SAFE_CAST(f.MarkDown1 AS FLOAT64), 4)[OFFSET(3)] AS '75%',
93 MAX(SAFE_CAST(f.MarkDown1 AS FLOAT64)) AS max
94 FROM `fiery-cabinet-461611-e8.grupo01.sales` s
95 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.features` f ON s.Store = f.Store AND s.Date = f.Date
96 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.stores` st ON s.Store = st.Store
97
98 UNION ALL
99
```

Asignatura	Datos del alumno	Fecha
Ingeniería para el Procesado Masivo de Datos	Apellidos: Alvares Angelim	27/06/2025
	Nombre: Erika Samara	

```

99
100 SELECT 'MarkDown2' AS variable,
101 COUNT(SAFE_CAST(f.MarkDown2 AS FLOAT64)) AS count,
102 COUNTIF(SAFE_CAST(f.MarkDown2 AS FLOAT64) IS NULL) AS nulos,
103 AVG(SAFE_CAST(f.MarkDown2 AS FLOAT64)) AS mean,
104 STDEV(SAFE_CAST(f.MarkDown2 AS FLOAT64)) AS std,
105 MIN(SAFE_CAST(f.MarkDown2 AS FLOAT64)) AS min,
106 APPROX_QUANTILES(SAFE_CAST(f.MarkDown2 AS FLOAT64), 4)[OFFSET(1)] AS '25%',
107 APPROX_QUANTILES(SAFE_CAST(f.MarkDown2 AS FLOAT64), 4)[OFFSET(2)] AS '50%',
108 APPROX_QUANTILES(SAFE_CAST(f.MarkDown2 AS FLOAT64), 4)[OFFSET(3)] AS '75%',
109 MAX(SAFE_CAST(f.MarkDown2 AS FLOAT64)) AS max
110 FROM `fiery-cabinet-461611-e8.grupo01.sales` s
111 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.features` f ON s.Store = f.Store AND s.Date = f.Date
112 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.stores` st ON s.Store = st.Store
113
114 UNION ALL
115
116 SELECT 'MarkDown3' AS variable,
117 COUNT(SAFE_CAST(f.MarkDown3 AS FLOAT64)) AS count,
118 COUNTIF(SAFE_CAST(f.MarkDown3 AS FLOAT64) IS NULL) AS nulos,
119 AVG(SAFE_CAST(f.MarkDown3 AS FLOAT64)) AS mean,
120 STDEV(SAFE_CAST(f.MarkDown3 AS FLOAT64)) AS std,
121 MIN(SAFE_CAST(f.MarkDown3 AS FLOAT64)) AS min,
122 APPROX_QUANTILES(SAFE_CAST(f.MarkDown3 AS FLOAT64), 4)[OFFSET(1)] AS '25%',
123 APPROX_QUANTILES(SAFE_CAST(f.MarkDown3 AS FLOAT64), 4)[OFFSET(2)] AS '50%',
124 APPROX_QUANTILES(SAFE_CAST(f.MarkDown3 AS FLOAT64), 4)[OFFSET(3)] AS '75%',
125 MAX(SAFE_CAST(f.MarkDown3 AS FLOAT64)) AS max
126 FROM `fiery-cabinet-461611-e8.grupo01.sales` s
127 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.features` f ON s.Store = f.Store AND s.Date = f.Date
128 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.stores` st ON s.Store = st.Store
129
130 UNION ALL
131
132

```

```

132
133 SELECT 'MarkDown4' AS variable,
134 COUNT(SAFE_CAST(f.MarkDown4 AS FLOAT64)) AS count,
135 COUNTIF(SAFE_CAST(f.MarkDown4 AS FLOAT64) IS NULL) AS nulos,
136 AVG(SAFE_CAST(f.MarkDown4 AS FLOAT64)) AS mean,
137 STDEV(SAFE_CAST(f.MarkDown4 AS FLOAT64)) AS std,
138 MIN(SAFE_CAST(f.MarkDown4 AS FLOAT64)) AS min,
139 APPROX_QUANTILES(SAFE_CAST(f.MarkDown4 AS FLOAT64), 4)[OFFSET(1)] AS '25%',
140 APPROX_QUANTILES(SAFE_CAST(f.MarkDown4 AS FLOAT64), 4)[OFFSET(2)] AS '50%',
141 APPROX_QUANTILES(SAFE_CAST(f.MarkDown4 AS FLOAT64), 4)[OFFSET(3)] AS '75%',
142 MAX(SAFE_CAST(f.MarkDown4 AS FLOAT64)) AS max
143 FROM `fiery-cabinet-461611-e8.grupo01.sales` s
144 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.features` f ON s.Store = f.Store AND s.Date = f.Date
145 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.stores` st ON s.Store = st.Store
146
147 UNION ALL
148
149 SELECT 'MarkDown5' AS variable,
150 COUNT(SAFE_CAST(f.MarkDown5 AS FLOAT64)) AS count,
151 COUNTIF(SAFE_CAST(f.MarkDown5 AS FLOAT64) IS NULL) AS nulos,
152 AVG(SAFE_CAST(f.MarkDown5 AS FLOAT64)) AS mean,
153 STDEV(SAFE_CAST(f.MarkDown5 AS FLOAT64)) AS std,
154 MIN(SAFE_CAST(f.MarkDown5 AS FLOAT64)) AS min,
155 APPROX_QUANTILES(SAFE_CAST(f.MarkDown5 AS FLOAT64), 4)[OFFSET(1)] AS '25%',
156 APPROX_QUANTILES(SAFE_CAST(f.MarkDown5 AS FLOAT64), 4)[OFFSET(2)] AS '50%',
157 APPROX_QUANTILES(SAFE_CAST(f.MarkDown5 AS FLOAT64), 4)[OFFSET(3)] AS '75%',
158 MAX(SAFE_CAST(f.MarkDown5 AS FLOAT64)) AS max
159 FROM `fiery-cabinet-461611-e8.grupo01.sales` s
160 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.features` f ON s.Store = f.Store AND s.Date = f.Date
161 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.stores` st ON s.Store = st.Store
162
163 UNION ALL
164

```

```

165
166 SELECT 'CPI' AS variable,
167 COUNT(SAFE_CAST(f.CPI AS FLOAT64)) AS count,
168 COUNTIF(SAFE_CAST(f.CPI AS FLOAT64) IS NULL) AS nulos,
169 AVG(SAFE_CAST(f.CPI AS FLOAT64)) AS mean,
170 STDEV(SAFE_CAST(f.CPI AS FLOAT64)) AS std,
171 MIN(SAFE_CAST(f.CPI AS FLOAT64)) AS min,
172 APPROX_QUANTILES(SAFE_CAST(f.CPI AS FLOAT64), 4)[OFFSET(1)] AS '25%',
173 APPROX_QUANTILES(SAFE_CAST(f.CPI AS FLOAT64), 4)[OFFSET(2)] AS '50%',
174 APPROX_QUANTILES(SAFE_CAST(f.CPI AS FLOAT64), 4)[OFFSET(3)] AS '75%',
175 MAX(SAFE_CAST(f.CPI AS FLOAT64)) AS max
176 FROM `fiery-cabinet-461611-e8.grupo01.sales` s
177 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.features` f ON s.Store = f.Store AND s.Date = f.Date
178 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.stores` st ON s.Store = st.Store
179
180 UNION ALL
181
182 SELECT 'Unemployment' AS variable,
183 COUNT(SAFE_CAST(f.Unemployment AS FLOAT64)) AS count,
184 COUNTIF(SAFE_CAST(f.Unemployment AS FLOAT64) IS NULL) AS nulos,
185 AVG(SAFE_CAST(f.Unemployment AS FLOAT64)) AS mean,
186 STDEV(SAFE_CAST(f.Unemployment AS FLOAT64)) AS std,
187 MIN(SAFE_CAST(f.Unemployment AS FLOAT64)) AS min,
188 APPROX_QUANTILES(SAFE_CAST(f.Unemployment AS FLOAT64), 4)[OFFSET(1)] AS '25%',
189 APPROX_QUANTILES(SAFE_CAST(f.Unemployment AS FLOAT64), 4)[OFFSET(2)] AS '50%',
190 APPROX_QUANTILES(SAFE_CAST(f.Unemployment AS FLOAT64), 4)[OFFSET(3)] AS '75%',
191 MAX(SAFE_CAST(f.Unemployment AS FLOAT64)) AS max
192 FROM `fiery-cabinet-461611-e8.grupo01.sales` s
193 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.features` f ON s.Store = f.Store AND s.Date = f.Date
194 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.stores` st ON s.Store = st.Store
195
196 UNION ALL
197

```

```

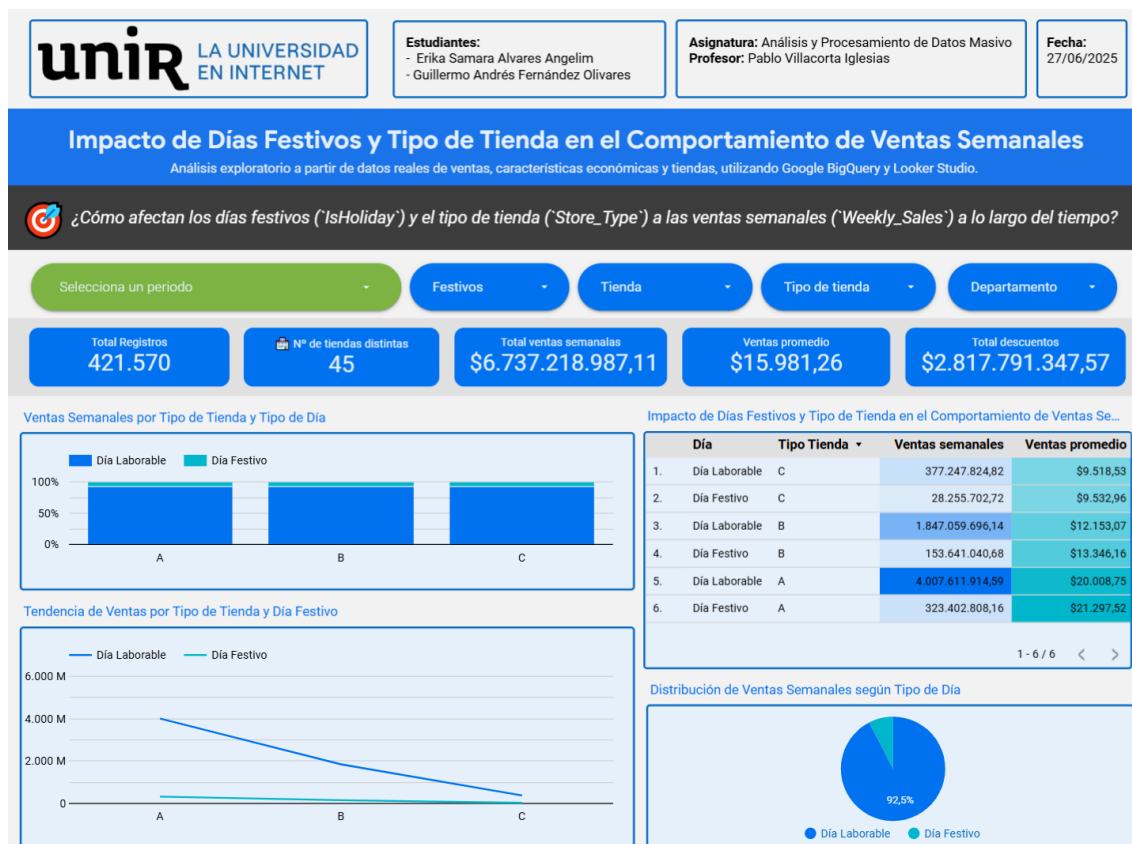
197
198 SELECT 'Store_Size' AS variable,
199 COUNT(SAFE_CAST(st.Size AS FLOAT64)) AS count,
200 COUNTIF(SAFE_CAST(st.Size AS FLOAT64) IS NULL) AS nulos,
201 AVG(SAFE_CAST(st.Size AS FLOAT64)) AS mean,
202 STDEV(SAFE_CAST(st.Size AS FLOAT64)) AS std,
203 MIN(SAFE_CAST(st.Size AS FLOAT64)) AS min,
204 APPROX_QUANTILES(SAFE_CAST(st.Size AS FLOAT64), 4)[OFFSET(1)] AS '25%',
205 APPROX_QUANTILES(SAFE_CAST(st.Size AS FLOAT64), 4)[OFFSET(2)] AS '50%',
206 APPROX_QUANTILES(SAFE_CAST(st.Size AS FLOAT64), 4)[OFFSET(3)] AS '75%',
207 MAX(SAFE_CAST(st.Size AS FLOAT64)) AS max
208 FROM `fiery-cabinet-461611-e8.grupo01.sales` s
209 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.features` f ON s.Store = f.Store AND s.Date = f.Date
210 LEFT JOIN `fiery-cabinet-461611-e8.grupo01.stores` st ON s.Store = st.Store
211
212 ORDER BY variable;

```

Se completó la consulta

Asignatura	Datos del alumno	Fecha
Ingeniería para el Procesado Masivo de Datos	Apellidos: Alvares Angelim	27/06/2025
	Nombre: Erika Samara	

6. Vista final limpia para Looker Studio



Desarrollo del dashboard

Esta sección documenta la construcción del dashboard interactivo en Google Looker Studio, como herramienta principal de visualización en esta actividad. El informe se estructura en torno a la hipótesis planteada: analizar el **impacto de los días festivos (IsHoliday) y el tipo de tienda (Store_Type) en las ventas semanales (Weekly_Sales)** a lo largo del tiempo.

Se empleó **Looker Studio** como herramienta de BI, conectada directamente con **BigQuery** mediante la tabla `join_ventas_limpia`, limitada a 20.000 registros para evitar costes en la plataforma. Este límite estuvo activo en el proceso de diseño del dashboard y al final, eliminamos este límite para sacar los pantallazos y proceder el registro de esa actividad. Looker Studio permite crear gráficos interactivos, filtros y componentes dinámicos con funciones nativas sin necesidad de programación adicional.

La fuente utilizada fue una tabla limpia derivada de la unión de los datasets **sales**, **features** y **stores**. Para optimizar el rendimiento. Se activaron las **funciones nativas de BigQuery** para evitar errores de formato. Se verificó que los filtros y tarjetas respondieran correctamente al selector de fecha.




Se diseñaron tarjetas y gráficos para representar las siguientes métricas:

Asignatura	Datos del alumno	Fecha
Ingeniería para el Procesado Masivo de Datos	Apellidos: Alvares Angelim	27/06/2025
	Nombre: Erika Samara	

Indicador	Fórmula / campo calculado	Interpretación
SUM(Weekly_Sales)	Total de ventas semanales	Evolución y volumen
AVG(Weekly_Sales)	Promedio por tienda o tipo	Comportamiento medio
COUNT(DISTINCT Date)	Total de días analizados	Alcance temporal
SUM(MarkDown1 + ... + MarkDown5)	Total de promociones activas	Relación con ventas
Segmentación por IsHoliday, Store_Type, Mes, Dept	Filtros y visualizaciones condicionales	Evaluación cruzada

Justificación técnica de las visualizaciones elegidas

Las visualizaciones fueron seleccionadas en función del tipo de variable, su rol en la hipótesis de negocio y la facilidad para detectar patrones:

Gráfica	Justificación
 Barras apiladas	Muestran claramente cómo las ventas varían según Store_Type y IsHoliday.
 Tabla con mapa de calor	Permite visualizar los montos por combinación de tienda y estado festivo.
 Línea temporal (opcional)	Útil para identificar picos, pero menos legible si hay demasiadas fechas.