



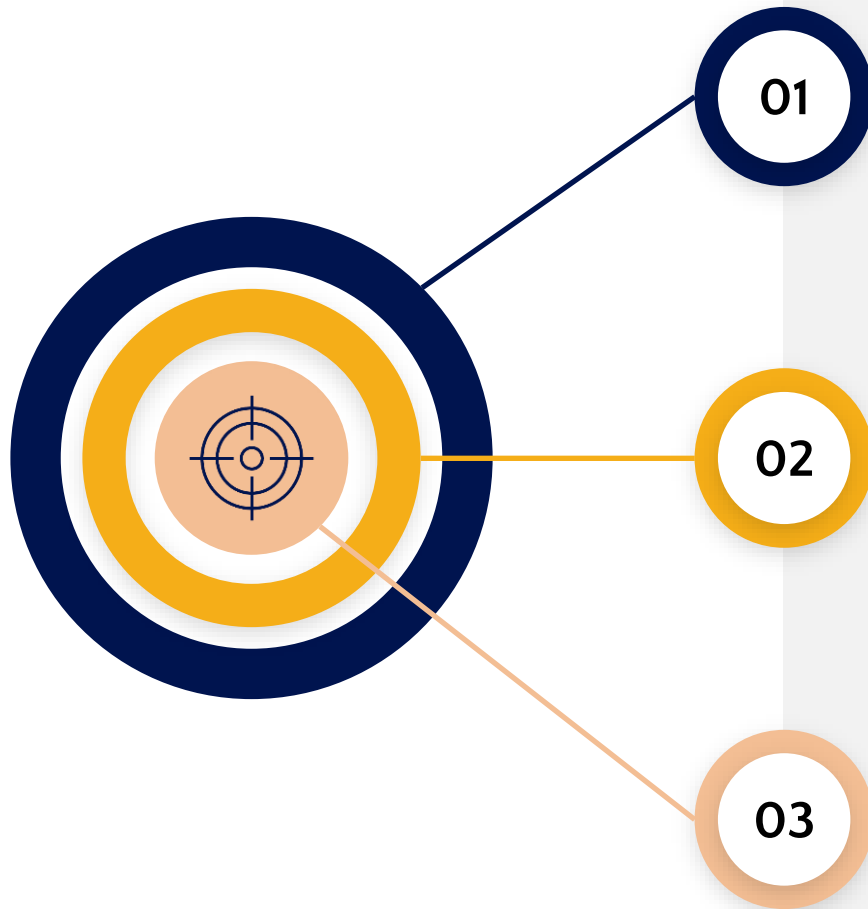
# Banking Analytics

# Introducción

Los bancos necesitan una arquitectura big data porque les permite almacenar y analizar grandes cantidades de datos de diversas fuentes, como transacciones financieras, registros de clientes y datos de redes sociales. Esto les ayuda a mejorar su toma de decisiones, detectar fraudes y riesgos, y proporcionar mejores servicios a sus clientes.



# Objetivos



01

**Canales:** Analizar los canales por los que se realizan la mayor cantidad de transacciones en tiempo real

02

**Segmentar:** Analizar patrones de comportamiento para una comprensión de los hábitos financieros de los clientes.

03

**Análisis de riesgo crediticio:** Analizar el riesgo crediticio de diferentes empresas

# Datos



Estado de  
transacciones



Tipo de transacciones



Transacciones durante  
periodos de tiempo



Medio de  
transacciones

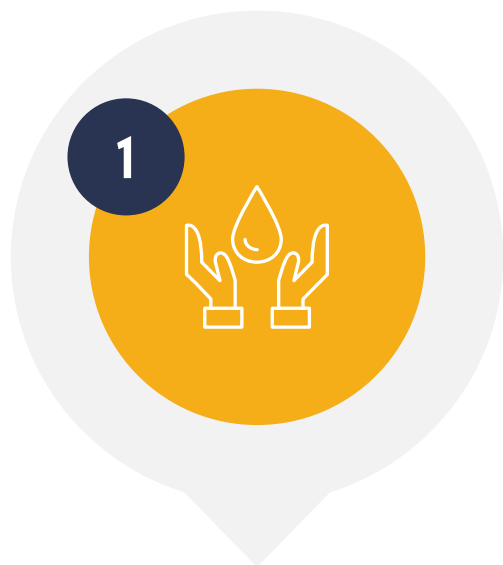
01

02

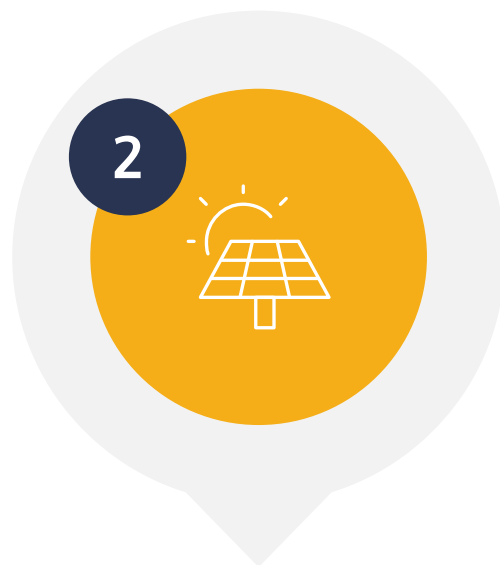
03

04

# Información



¿Cuál es el porcentaje de transacciones correctas y fallidas?



¿Cuál es el número total de transacciones diferidas e inmediatas y cuánto representan del total?

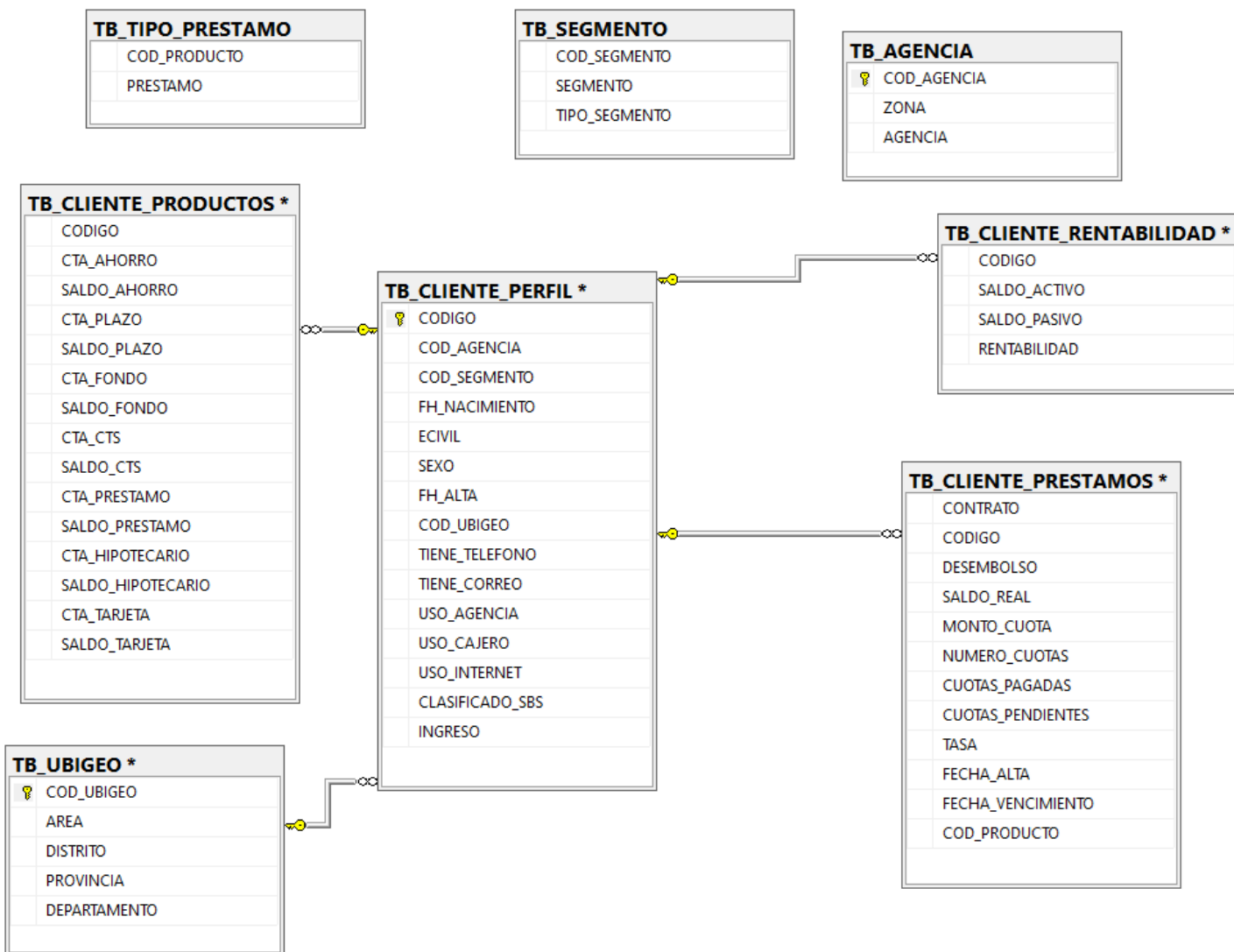


¿Cuál es el horario donde se realizan la mayor cantidad de transacciones?

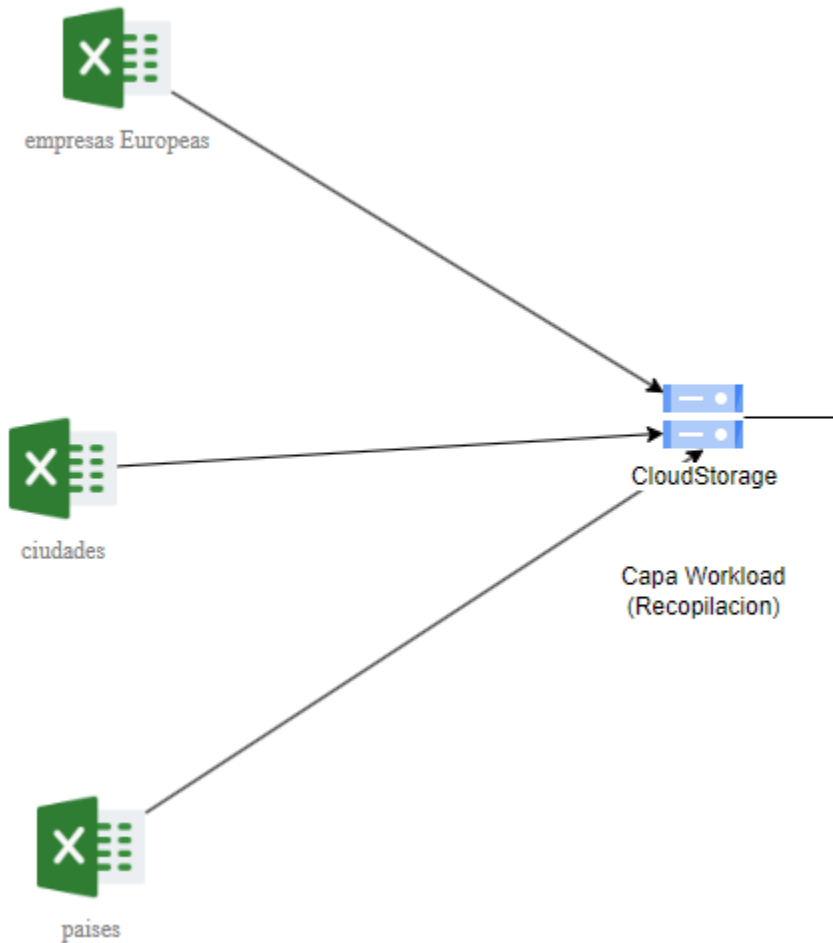


¿Cuál es el total de transacciones por app y ventanilla?

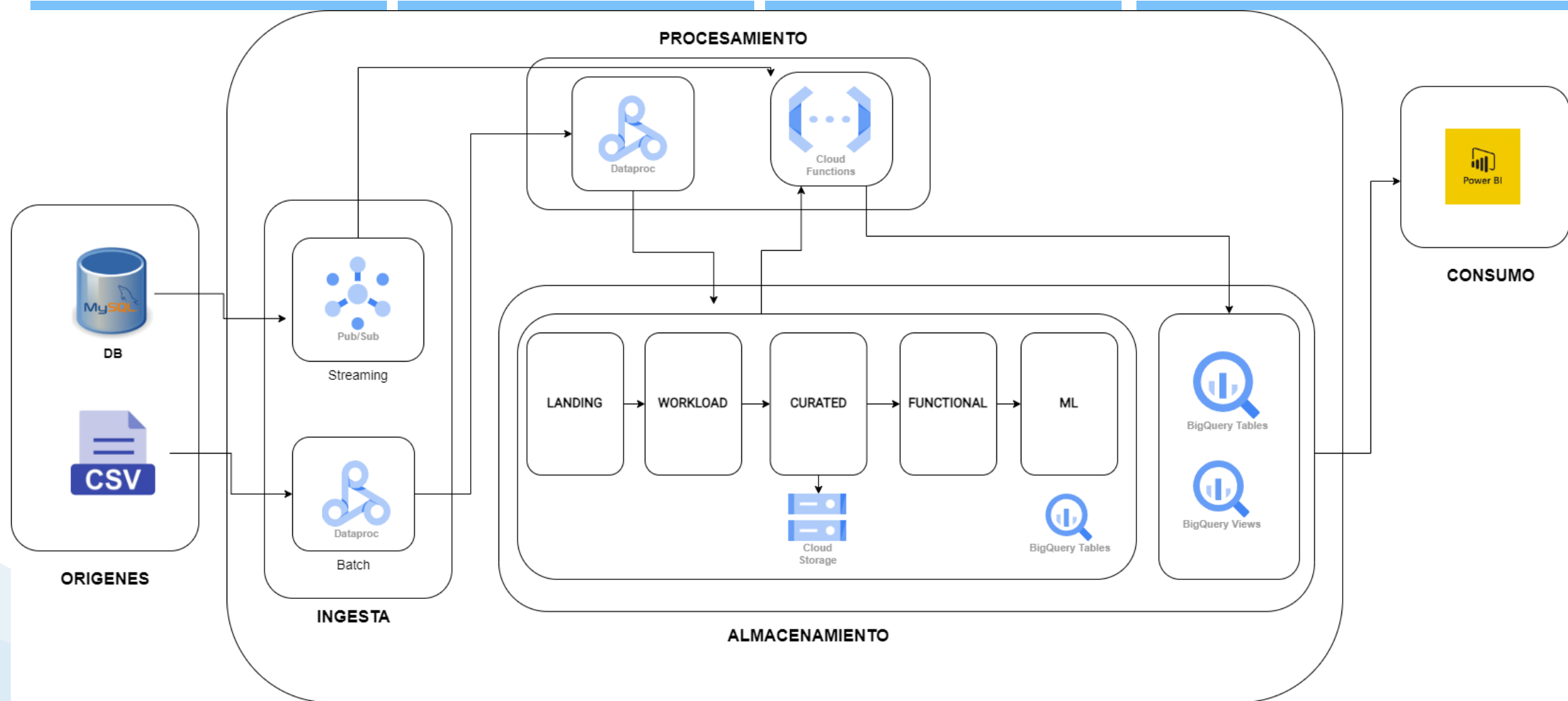
# Origen de datos



# Origen de datos Proceso Batch



# Arquitectura





# Tecnologías

## Google Storage

### Diferentes capas que se procesan en el Dataproc

#### pea-project-gr

Ubicación	Clase de almacenamiento	Acceso público	Protección
us (varias regiones en Estados Unidos)	Standard	Sujeto a LCA de objeto	Ninguno

[OBJETOS](#) [CONFIGURACIÓN](#) [PERMISOS](#) [PROTECCIÓN](#) [CICLO DE VIDA](#) [OBSERVABILIDAD](#)

Depósitos > pea-project-gr > datalake

[SUBIR ARCHIVOS](#) [SUBIR CARPETA](#) [CREAR CARPETA](#) [TRANSFERIR LOS DATOS](#) [ADMINISTRAR CONSERVACIÓN](#)

Filtrar solo por prefijo de nombre

Filtro Filtrar objetos y carpetas

<input type="checkbox"/>	Nombre	Tamaño	Tipo	Fecha de creación	Clase de almacenamiento
<input type="checkbox"/>	curated/	—	Carpeta	—	—
<input type="checkbox"/>	functional/	—	Carpeta	—	—
<input type="checkbox"/>	landing/	—	Carpeta	—	—
<input type="checkbox"/>	workload/	—	Carpeta	—	—

### Código que utiliza Cloud Functions

#### bk\_code\_message\_db

Ubicación	Clase de almacenamiento	Acceso público	Protección
us (varias regiones en Estados Unidos)	Standard	Sujeto a LCA de objeto	Ninguno

[OBJETOS](#) [CONFIGURACIÓN](#) [PERMISOS](#) [PROTECCIÓN](#) [CICLO DE VIDA](#) [OBSERVABILIDAD](#)

Depósitos > bk\_code\_message\_db

[SUBIR ARCHIVOS](#) [SUBIR CARPETA](#) [CREAR CARPETA](#) [TRANSFERIR LOS DATOS](#) [ADMINISTRAR CONSERVACIÓN](#)

Filtrar solo por prefijo de nombre

Filtro Filtrar objetos y carpetas

<input type="checkbox"/>	Nombre	Tamaño	Tipo	Fecha de creación	Clase de almacenamiento
<input type="checkbox"/>	process_pubsub_message_clientes.zip	1.5 KB	application/zip	16 sept 2023 23:12:32	Standard
<input type="checkbox"/>	process_pubsub_message_transaccione...	1.1 KB	application/zip	16 sept 2023 23:12:32	Standard

# Tecnologías

## PubSub

### Temas

Google Cloud	bank-project	pubsub
Pub/Sub	Temas	<a href="#">+ CREAR TEMA</a> <a href="#">BORRAR</a>
Pub/Sub	LISTA	MÉTRICAS
Temas	Filtro	Filtrar temas
Suscripciones	ID del tema ↑	Clave de encriptación
Instantáneas	clientes-topic	Google-managed
Esquemas	transactions-topic	Google-managed

### Suscripciones

Suscripciones	<a href="#">+ CREAR SUSCRIPCIÓN</a>	<a href="#">BORRAR</a>
LISTA	MÉTRICAS	
Filtro	Filtrar las suscripciones	
Estado	ID de la suscripción ↑	Tipo de envío
✓	gcf-process-pubsub-message-clientes-us-central1-clientes-topic	Envío
✓	gcf-process-pubsub-message-transacciones-us-central1-transactions-topic	Envío
✓	transactions-topic-subs	Extracción

# Tecnologías

## Cloud Functions

Las funciones se disparan cuando el PUB/SUB recibe mensajes y realiza las transformaciones correspondientes

The screenshot shows the Google Cloud Functions console for the 'bank-project'. It includes a navigation bar with the Google Cloud logo and a dropdown menu for the project. Below the navigation bar, there's a section for 'Cloud Functions' with a 'Funciones' tab, a '+ CREAR FUNCIÓN' button, and an 'ACTUALIZAR' button. A 'Filtro' section allows filtering functions. The main content area displays a table of functions with columns for checkboxes, environment, name, and last implementation time.

<input type="checkbox"/>	Entorno	Nombre ↑	Última implementación
<input type="checkbox"/>	✓ 1st gen	<a href="#">process_pubsub_message_clientes</a>	16 sept 2023 23:14:11
<input type="checkbox"/>	✓ 1st gen	<a href="#">process_pubsub_message_transacciones</a>	16 sept 2023 23:14:13

# Tecnologías

## Notebook/ PySpark – DataProc

bank-project-398102 > cluster-3eec [Sign out](#)

File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

/ GCS / ml-process /

Name	Last Modified
schemas	5 days ago
FillCurated.ipynb	8 hours ago
FillLanding.ipynb	7 hours ago
landing.py	7 hours ago
FillFunctional.ipynb	7 hours ago
functional.py	7 hours ago
curated.py	6 hours ago
mlprocess.py	5 hours ago
Modelo.ipynb	5 hours ago

Modelo.ipynb

Las características: nivel de riesgo crediticio + Facturación + EBIT + PLTax + Sector + País + Apalancamiento + ROE + TAsset de años anteriores, para predecir.

```
[27]: X = restricted_df[['MScore.'+str(año_widget.value)+'.int', 'Turnover.'+str(año_widget.value), 'EBIT.'+str(año_widget.value), 'PLTax.'+str(año_widget.value), 'Leverage.'+str(año_widget.value)]]
X = pd.concat([X, restricted_df.loc[:, 'Automobiles and Components':'Real Estate']], axis=1)
X = pd.concat([X, restricted_df['Spain']], axis=1)
X = pd.DataFrame(preprocessing.MinMaxScaler().fit_transform(X), columns=X.columns) # MinMax Scaling
y = restricted_df[['MScore.'+str(año_widget.value + 1)+'.int']]
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0, test_size=0.2, shuffle=True, stratify=y)
X
```

[27]:

	MScore.2018.int	Turnover.2018	EBIT.2018	PLTax.2018	Leverage.2018	ROE.2018	TAsset.2018	Automobiles and Components	Capital Goods	Commercial and professional services	Energy	Food Beverage and Tobacco	Food and Staples Retailing	Health Care Equipment and Services
0	1.0	0.158884	0.777574	0.809337	0.086342	0.152948	0.007275	0.0	1.0	0.0	0.0	0.0	0.0	0.0
1	1.0	0.912978	0.772238	0.806730	0.086212	0.152937	0.017136	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.856485	0.791195	0.821109	0.086208	0.152946	0.031393	0.0	1.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.701328	0.776359	0.809904	0.086215	0.152954	0.005507	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.641379	0.790593	0.820282	0.086213	0.152966	0.014624	0.0	1.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
24315	0.0	0.032581	0.775291	0.809366	0.086229	0.152947	0.001933	0.0	1.0	0.0	0.0	0.0	0.0	0.0
24316	1.0	0.048264	0.775386	0.809299	0.086246	0.152946	0.001715	0.0	0.0	0.0	0.0	0.0	1.0	0.0
24317	0.0	0.026216	0.776107	0.810130	0.086209	0.152980	0.000322	0.0	0.0	0.0	0.0	0.0	0.0	0.0
24318	0.0	0.056912	0.782384	0.814631	0.086208	0.152950	0.006792	0.0	0.0	0.0	0.0	0.0	0.0	0.0
24319	1.0	0.011202	0.775118	0.809249	0.086268	0.152942	0.001048	0.0	1.0	0.0	0.0	0.0	0.0	0.0

# Tecnologias

## DataProc / Workflow

Dataprocc

Jobs on Clusters

Clusters

Jobs

Workflows

Autoscaling policies

Serverless

Batches

Interactive

Metastore Services

Metastore

Federation

Utilities

Component exchange

Workbench

Workflow template details

template-mlprocessing

Cluster placement

Type	Auto managed cluster
Template region	us-central1
Cluster name prefix	cluster-ml
Zone	Auto-zone
Master node	Standard (1 master, N workers)
Machine type	n1-standard-2
Primary disk type	pd-standard
Primary disk size	100 GB
Worker nodes	2
Machine type	n1-standard-2
Primary disk type	pd-standard
Primary disk size	500 GB
Local SSDs	0
Secondary worker nodes	0
Image version	2.1-debian11

Job details

Step ID	Dependencies	Job type
job-landing	No dependencies	PySpark
job-curated	job-landing	PySpark
job-functional	job-curated	PySpark

## Jobs Executions

Dataprocc

Jobs on Clusters

Clusters

Jobs

Workflows

Autoscaling policies

Serverless

Batches

Interactive

Workflow instance details

55a40f2a-e1d4-425a-928e-35b76d56366d

Template: [template-mlprocessing](#)

Job details

Filter Filter jobs

Step ID	Status	Job ID	Dependencies
job-landing	Completed	<a href="#">job-landing-vkarvzugcrrqy</a>	No dependencies
job-curated	Completed	<a href="#">job-curated-vkarvzugcrrqy</a>	job-landing
job-functional	Completed	<a href="#">job-functional-vkarvzugcrrqy</a>	job-curated

# Tecnologias

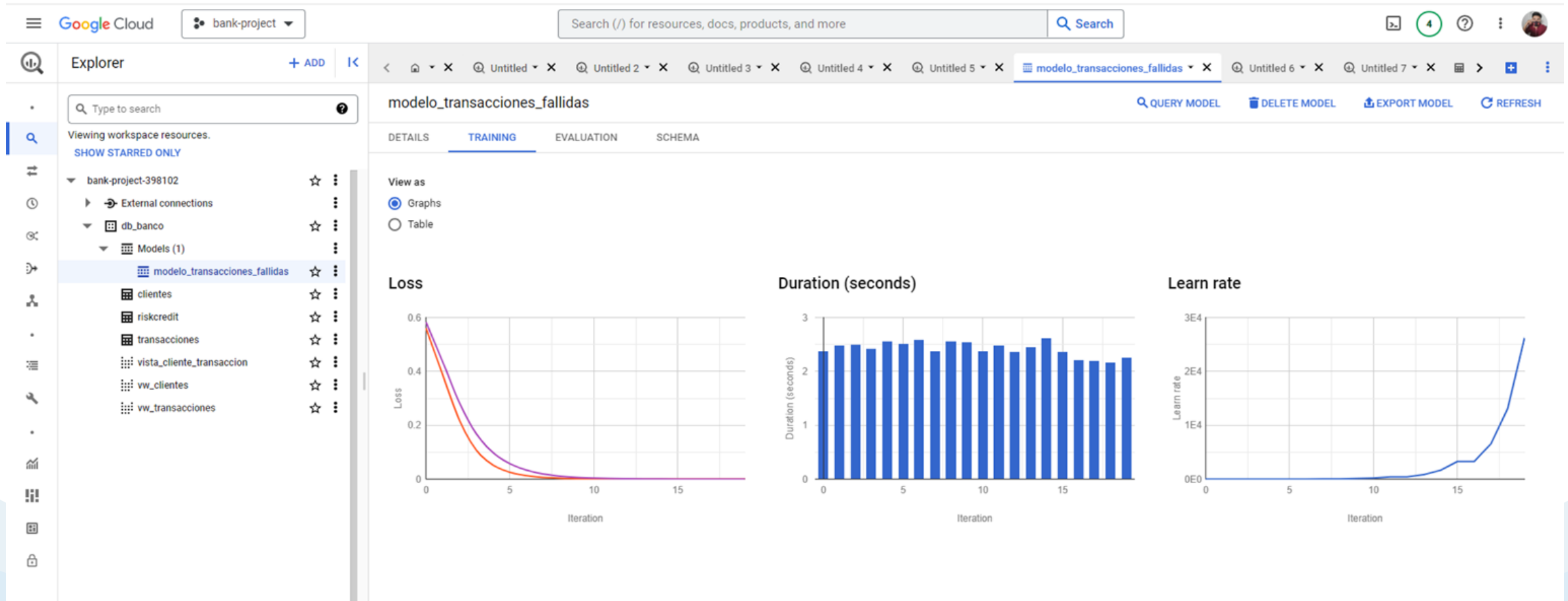
## BigQuery Tables

The screenshot displays the Google Cloud BigQuery Explorer interface. On the left, the Explorer pane shows the project hierarchy: bank-project > db\_banco > Models (1) > transacciones. The main pane shows the 'transacciones' table in the 'PREVIEW' tab. The table has 11 columns: Row, id\_deposito\_cliente, id\_recibe\_cliente, monto, tipo\_transaccion, agencia, medio\_transaccion, fallo, fecha\_transaccion, and motivo. The data shows a list of transactions with various details like amounts, transaction types, and dates.

Row	id_deposito_cliente	id_recibe_cliente	monto	tipo_transaccion	agencia	medio_transaccion	fallo	fecha_transaccion	motivo
1	445	48617	377.78	Inmediato	Camacho	Ventanilla	1	2023-09-05T05:13:57	Fondos
2	446	48620	641.01	Diferido	Caminos del Inca	App	0	2023-09-05T05:13:55	
3	447	48606	248.2	Diferido	Online	App	1	2023-09-05T05:13:57	Fondos
4	447	48577	0.72	Diferido	La Punta	Web	1	2023-09-05T05:13:54	Fondos
5	448	48572	185.82	Inmediato	Grau (Lima)	Web	1	2023-09-05T05:13:58	Fondos
6	448	48607	0.16	Inmediato	La Punta	App	1	2023-09-05T05:13:58	Fondos
7	448	48590	232.29	Inmediato	Online	Web	1	2023-09-05T05:13:55	Fondos
8	449	48580	38.51	Inmediato	Grau (Lima)	Ventanilla	0	2023-09-05T05:13:54	
9	449	48612	0.41	Diferido	Grau (Lima)	Ventanilla	0	2023-09-05T05:13:57	
10	450	48591	0.19	Diferido	Online	App	0	2023-09-05T05:13:56	
11	1586	191774	610.42	Inmediato	Caminos del Inca	Web	1	2023-08-11T03:50:13	Fondos
12	1588	191799	101.83	Diferido	Camacho	App	0	2023-08-11T03:50:15	
13	1588	191814	251.04	Inmediato	Asia	App	0	2023-08-11T03:50:16	
14	1588	191768	432.86	Inmediato	La Punta	Ventanilla	0	2023-08-11T03:50:03	
15	1589	191778	27.77	Inmediato	Online	Ventanilla	0	2023-08-11T03:50:12	
16	1593	191807	0.81	Diferido	Grau (Lima)	Web	1	2023-08-11T03:50:14	Fondos
17	1594	191783	265.4	Diferido	Asia	App	0	2023-08-11T03:50:11	
18	1594	191813	0.34	Diferido	Puente Piedra	App	0	2023-08-11T03:50:04	
19	1594	191768	71.94	Inmediato	Caminos del Inca	Web	0	2023-08-11T03:50:19	
20	1596	191803	0.91	Diferido	Online	Ventanilla	0	2023-08-11T03:50:08	
21	1598	191796	0.45	Inmediato	Grau (Lima)	Ventanilla	0	2023-08-11T03:50:14	
22	1598	191816	83.9	Diferido	Camacho	App	0	2023-08-11T03:50:14	
23	1601	191767	0.88	Inmediato	Camacho	Ventanilla	0	2023-08-11T03:50:12	

# Tecnologías

## BigQuery ML



# POWER BI

PEA DATA ENGINEER

---

## **Análisis Bancario**

Seguimiento Transacciones

Seguimiento Clientes

Riesgo Crediticio



# Seguimiento de Transacciones



# Seguimiento de Clientes



# Modelo Riesgo Crediticio

Empresa	Riesgo_credificio
- SOMASCHINI SPA -	1
@PACK SRL	1
1001 CRECHES	1
2 ERRE S.R.L.	0
2 L S.R.L.	0
2 M DECORI SPA	0
25 GIUGNO SOCIETA' COOPERATIVA	0
2BR MOBILITE	1
2B-SOFTEAM DATA	0
2C CASADO CONSULTING SL	0
2C DI COVERI S.R.L.	0
2FC+NET	0
2H ENERGY	0
2M S.R.L.	0
2P ASFALTI S.R.L.	0
2SB	1
3 C - LAVORAZIONE PELLI S.R.L.	0
3 C IPERCASH S.R.L.	1
3 EFFE DISTRIBUZIONE S.R.L.	0
3 EMME S.R.L.	0
3 I D	0
3 ID RENOVATION	1
3 M SOCIETA' A RESPONSABILITA' LIMITATA	0
3 M.C. - SOCIETA' PER AZIONI	1
3 P BIOPHARMACEUTICALS SL	0
3 ZINNEN SPA	1
31 Y 32 BEBICASH SA	0
3A DEI F.LLI ANTONINI S.P.A.	0
3D IMMOBILIARE S.R.L.	0
3D-MICROMAC AG	0
3F DI FERRECCHI SILVANO SOCIETA' PER AZIONI SIGLABILE 3F DI FERRECCHI SILVANO S.P.A.	0
3G IMMO-CONSULTANT	1

