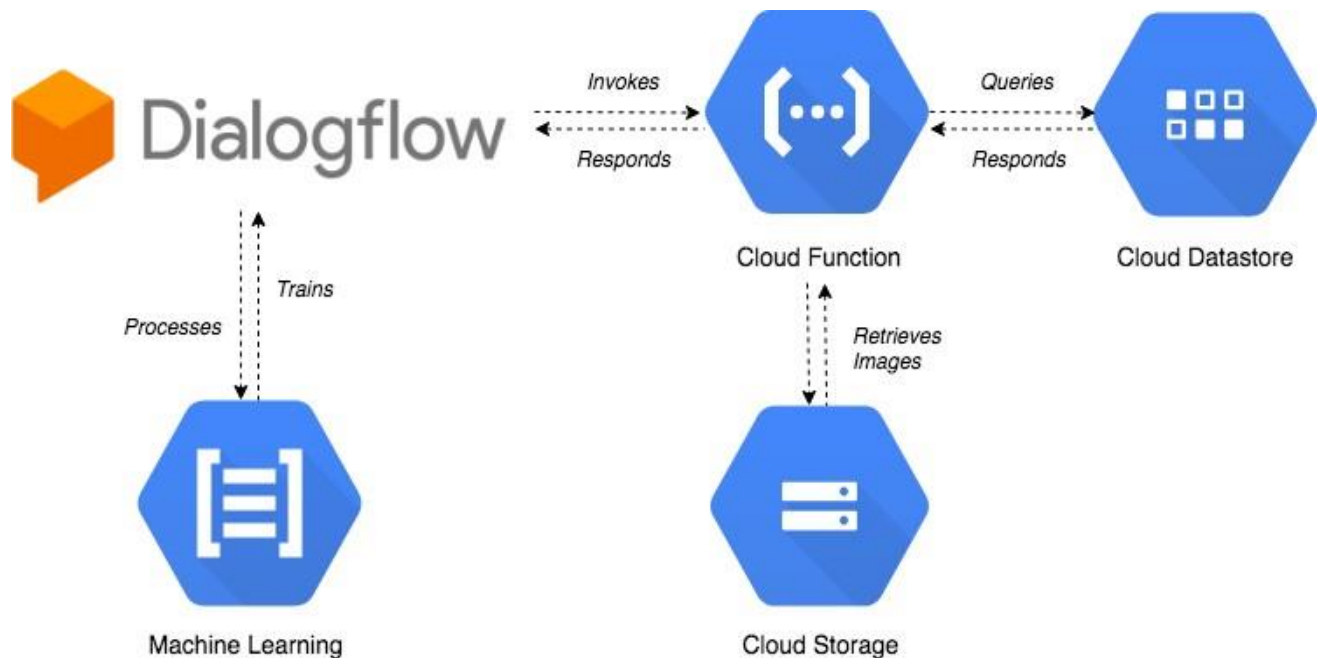


# BIG QUERY + ML+ BOT

TIEMPO DE RESPUESTA.



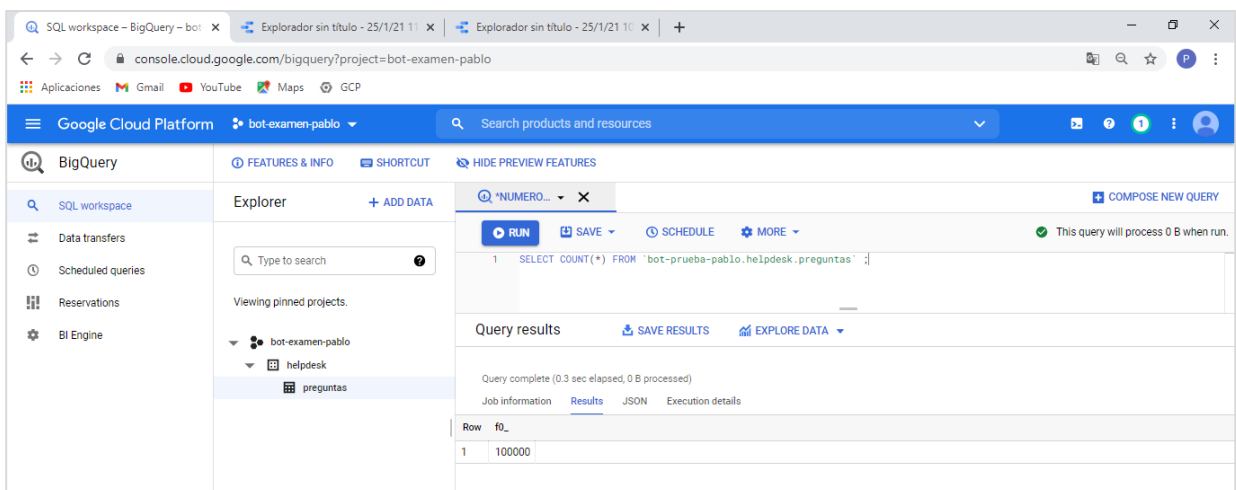
# BIG QUERY + ML+ BOT

## El problema

El problema que debe resolver el estudiante es construir de forma correcta un proyecto en GCP (con las credenciales asignadas ), deberá conectar el archivo /datos (dataset), realizar un análisis exploratorio que le ayudará a construir las query's necesarias empleando para ello Big Query y Big QueryML; a partir de su modelo ya entrenado deberá conectarlo a un chat Bot para que pueda responder en tiempo real a casos no vistos antes.

## Se pide ?

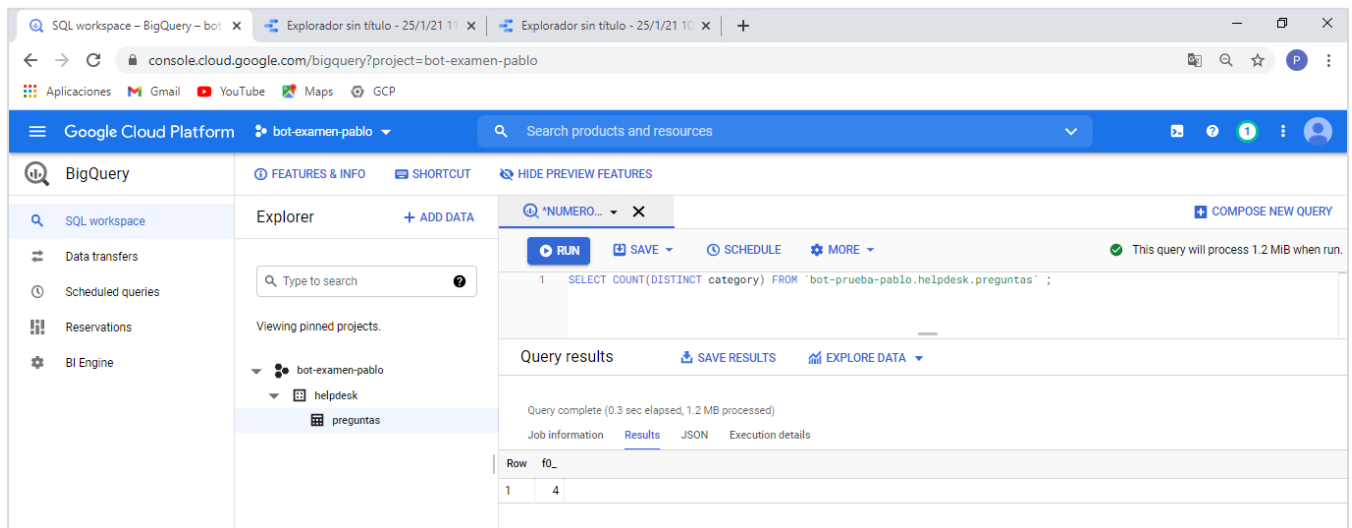
1. Configurar GCP.
2. Conectar el archivo.( DATASET )
3. Leer y hacer EDA, usando DataStudio, responder:
  - a Cuántas observaciones tiene el dataset(rows) ?



El dataset tiene 100.000 observaciones.

- Query:

```
SELECT COUNT(*) FROM `bot-prueba-pablo.helpdesk.preguntas` ;
```

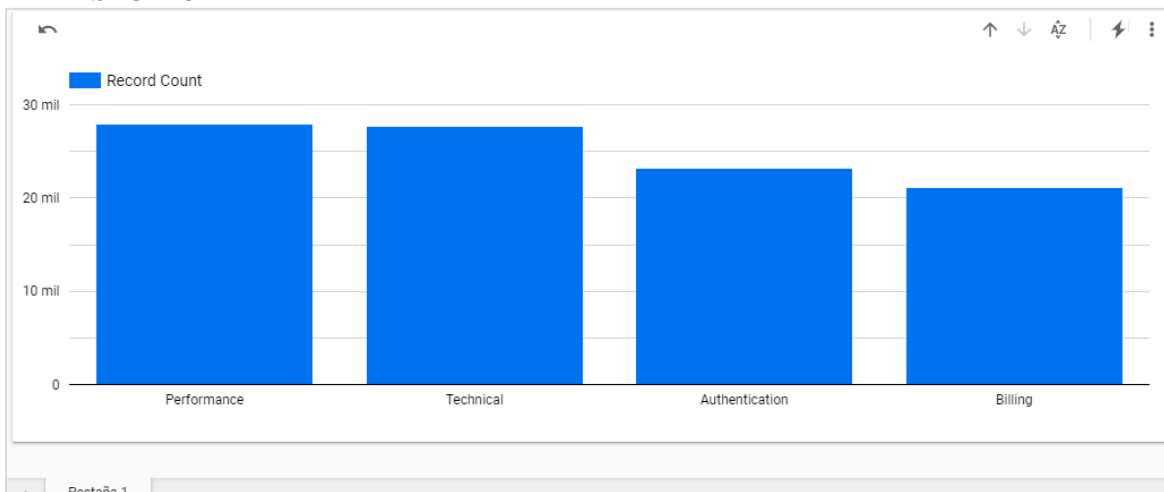
**b Cuántas categorías tiene el atributo ‘category’?**

The screenshot shows the Google Cloud Platform BigQuery console. The query editor displays the following SQL query:

```
SELECT COUNT(DISTINCT category) FROM `bot-prueba-pablo.helpdesk.preguntas`;
```

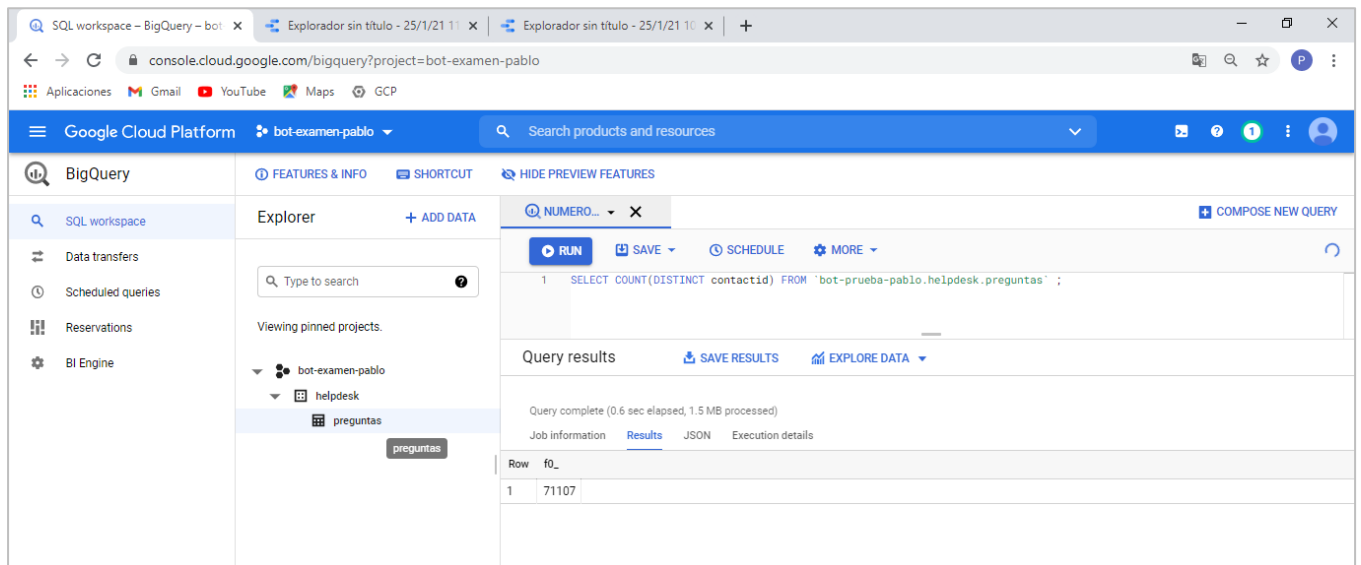
The query results are shown in a table with one row and one column:

Row	fo_
1	4

**DATA STUDIO**

4 categorías: Performance, Technical, Autentication, Billing.

- Query:  
`SELECT COUNT(DISTINCT category) FROM `bot-prueba-pablo.helpdesk.preguntas` ;`

**c Cuantos empleados diferentes hay en el dataset?**

The screenshot shows the Google Cloud Platform BigQuery console. The left sidebar contains navigation options: BigQuery, SQL workspace, Data transfers, Scheduled queries, Reservations, and BI Engine. The main area is divided into three sections: Explorer, Query editor, and Query results. The Explorer section shows a project named 'bot-examen-pablo' with a dataset named 'preguntas'. The Query editor section contains the following SQL query:

```
1 SELECT COUNT(DISTINCT contactid) FROM `bot-prueba-pablo.helpdesk.preguntas` ;
```

The Query results section shows the query is complete (0.6 sec elapsed, 1.5 MB processed). The results are displayed in a table with one row and one column:

Row	fo_
1	71107

El dataset tiene 71107 empleados distintos.

- **Query:**

```
SELECT COUNT(DISTINCT contactid) FROM `bot-prueba-pablo.helpdesk.preguntas` ;
```

#### 4. Realizar todo el proceso de entrenamiento, evaluación, predicción para los siguientes datos:

##### MODELO

The screenshot shows the Google Cloud Platform BigQuery interface. The left sidebar contains navigation options like SQL workspace, Data transfers, Scheduled queries, Reservations, and BI Engine. The main area is divided into an Explorer on the left and a query editor on the right. The Explorer shows a project named 'bot-examen-pablo' with a dataset 'helpdesk' containing tables 'pred' and 'preguntas'. The query editor shows a SQL query to create or replace a model named 'helpdesk.pred' using a linear regression model on the 'preguntas' dataset.

```

1 CREATE OR REPLACE MODEL `bot-examen-pablo.helpdesk.pred`
2   OPTIONS (model_type='linear_reg') AS
3   SELECT seniority, type, experience, category, resolutiontime as label
4   FROM `bot-examen-pablo.helpdesk.preguntas`
5

```

##### EVALUACIÓN

The screenshot shows the Google Cloud Platform BigQuery interface with the evaluation query executed. The query editor shows a SQL query that creates a table 'eval\_table' from the 'preguntas' dataset and then evaluates the model 'helpdesk.pred' against it. The results are displayed in a table below the query editor.

```

1 WITH eval_table AS (SELECT seniority, type, experience, category, resolutiontime as label
2   FROM `bot-examen-pablo.helpdesk.preguntas`)
3 SELECT * FROM ML.EVALUATE (MODEL helpdesk.pred, TABLE eval_table)
4

```

Row	mean_absolute_error	mean_squared_error	mean_squared_log_error	median_absolute_error	r2_score	explained_variance
1	1.2312840677943597	3.158530707897275	0.1162556353581454	0.9175356517186337	0.7047509930700806	0.7047510634888328

a Para un ticket : nivel-> 10 seniority, '3-Advanced' en experience, 'Technical' en category, 'Request' como type. **CUANTO DIAS ?**

- Query:

```

WITH pred_table AS (SELECT 10 AS seniority, '3-
Advance' AS experience, 'Technical' AS category, 'request' as type)
SELECT * FROM ML.PREDICT (MODEL helpdesk.pred, TABLE pred_table)

```

The screenshot shows the Google Cloud Platform BigQuery interface with the prediction query executed. The query editor shows a SQL query that creates a table 'pred\_table' with specific input values and then predicts the label using the model 'helpdesk.pred'. The results are displayed in a table below the query editor.

```

1 WITH pred_table AS (SELECT 10 AS seniority, '3-Advanced' AS experience, 'Technical' AS category, 'Request' as type)
2 SELECT * FROM ML.PREDICT (MODEL helpdesk.pred, TABLE pred_table)
3

```

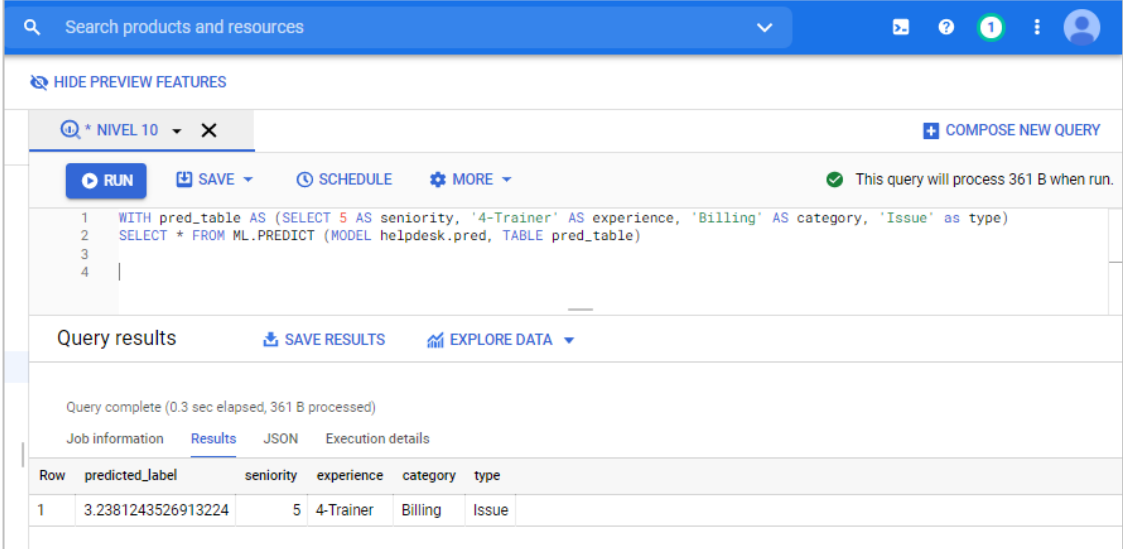
Row	predicted_label	seniority	experience	category	type
1	3.09599775257164	10	3-Advanced	Technical	Request

Tarda 3.09 días

**b Para un ticket : nivel-> 5 seniority, '4-Trainer' en experience, 'Billing' en category, 'Issue' como type. CUANTO DIAS ?**

• **Query:**

```
WITH pred_table AS (SELECT 5 AS seniority, '4-Trainer' AS experience, 'Billing' AS category, 'Issue' as type)
SELECT * FROM ML.PREDICT (MODEL helpdesk.pred, TABLE pred_table)
```



Query results

Query complete (0.3 sec elapsed, 361 B processed)

Row	predicted_label	seniority	experience	category	type
1	3.2381243526913224	5	4-Trainer	Billing	Issue

Tarda 3.2381 dias

**c Cuánto tiempo tardo el cluster de bigquery en responder a la pregunta del punto 4a ?**

Job information	Results	Execution details
Elapsed time	Slot time consumed ?	Stages ?
13.8 sec	1 min 16.519 sec	<div> <div>✓</div> Preprocess 7.230 sec           </div> <div> <div>✓</div> Train 5.127 sec           </div> <div> <div>✓</div> Evaluate 1.180 sec           </div>
		Training iterations
		Completed: 1 Planned: 20

d **Cuántos megabytes** uso el cluster de bigquery en responder a la pregunta del punto 4b?

1	WITH pred_table AS (SELECT 10 AS seniority, '3-Advance' AS experience, 'Technical' AS category, 'request' as type)
2	SELECT * FROM ML.PREDICT (MODEL helpdesk.pred, TABLE pred_table)
3	

Job ID	bot-examen-pablo:US.bqjob_5bf3325b_177392cc0df
User	pablo@sucuzhanay.com
Location	United States (US)
Creation time	Jan 25, 2021, 11:52:50 AM
Start time	Jan 25, 2021, 11:52:51 AM
End time	Jan 25, 2021, 11:52:51 AM
Duration	0.2 sec
Bytes processed	361 B
Bytes billed	10 MB
Job priority	INTERACTIVE
Destination table	<a href="#">Temporary table</a>
Use legacy SQL	false

Vemos que ha empleado 361 MB.

e **Escriba la Query que uso para responder el punto 3C, así como el tiempo que tardo y los megabytes usados ?**

NUMERO... X

COMPOSE NEW QUERY

RUN SAVE SCHEDULE MORE

This query will process 1.2 MiB when run.

1 SELECT COUNT(DISTINCT category) FROM `bot-prueba-pablo.helpdesk.preguntas` ;

Query results SAVE RESULTS EXPLORE DATA

Query complete (0.0 sec elapsed, cached)

Job information Results JSON

Query completed in 0.021 sec

Open query in editor

1 SELECT COUNT(DISTINCT category) FROM `bot-prueba-pablo.helpdesk.preguntas` ;

La query tarda 0.021

- Query:

SELECT COUNT(DISTINCT contactid) FROM `bot-prueba-pablo.helpdesk.preguntas` ;

**5. Deberá usar dialogflow para construir el BOT, que debera llamarse (su nombre)**

Dialogflow Essentials US

**CREATE**

**+ Create Agent**

Docs

Dialogflow CX [new]

Support

Account

Logout

**DEFAULT LANGUAGE**

English - en

Primary language for your agent. Other languages can be added later.

**DEFAULT TIME ZONE**

(GMT+1:00) Europe/Madrid

Date and time requests are resolved using this timezone.

**GOOGLE PROJECT**

bot-examen-pablo

Enables Cloud functions, Actions on Google and permissions management.

**AGENT TYPE**

☒ Set as Mega Agent

Combine multiple Dialogflow agents (i.e. sub agents) into a single agent (i.e. mega agent).

**6. Haga deploy y pruebe el bot****7. Elegir y construir 3 ( tres ) intents, se deja al estudiante la elección de las mismas.**

SQL workspace - BigQuery - bot x Dialogflow x +

dialogflow.cloud.google.com/#/agent/bot-prueba-pablo/editIntent/9b413aef-dd97-4672-9c83-40d81438f44c/

Aplicaciones Gmail YouTube Maps GCP

Dialogflow Essentials US

**Submit Ticket - Issue Category** **SAVE**

Pablo en

**Intents**

Entities Knowledge [beta] Fulfillment Integrations Training Validation History Analytics

**Events**

**Training phrases**

no me arranca el ordenador

PARAMETER NAME	ENTITY	RESOLVED VALUE
category	@category	arranca

no me tula el pc

I have a problem with my laptop

My printer isn't working

My printer is slow

I need a software license



Dialogflow Essentials US

category SAVE

☒ Define synonyms ☐ Regexp entity ☐ Allow automated expansion  
☐ Fuzzy matching

Authentication	Authentication, Login, Access
Business	Business, Accounting, Purchase, Advice
Technical	Technical, Computer Issue, Computer, Software, not working, problem, isn't working, rula, arranca
Performance	Performance, Slow Computer, Slow, Need Speed Up
Billing	Billing

[Click here to edit entry](#)

**Creamos los siguientes intents:**

- Rula
- Arranca
- Activación

**8. RESPONDER : CUANTO TIEMPO TARDA UNA PREDICCION PARA UN EXPERTO NIVEL 7.( Estas respuestas debe obtenerlas del diálogo con el chat Bot )**

```

index.js  package.json
36
37
38 // The SQL Query to Run
39 const SQLQUERY = `WITH eval_table AS ( SELECT 7 as seniority, '3-Advanced' as e
40 @category as category, 'Request' as type)
41 SELECT cast(predicted_label as INT64) as predicted_label
42 FROM ML.PREDICT(MODEL helpdesk.pred, TABLE eval_table)`;
43
44 const OPTIONS = {
45   query: SQLQUERY,
46   // Location must match that of the dataset(s) referenced in the query.
47   location: "US",
48   params: {
49     category: ISSUE_CATEGORY
50

```

Try it now

Agent

USER SAYS

COPY CURL

billing

DEFAULT RESPONSE

undefined, your ticket has been created.  
Someone will you contact shortly. The  
estimated response time is 3 days.

CONTEXTS

RESET CONTEXTS

Response

submitticket-collectname-followup

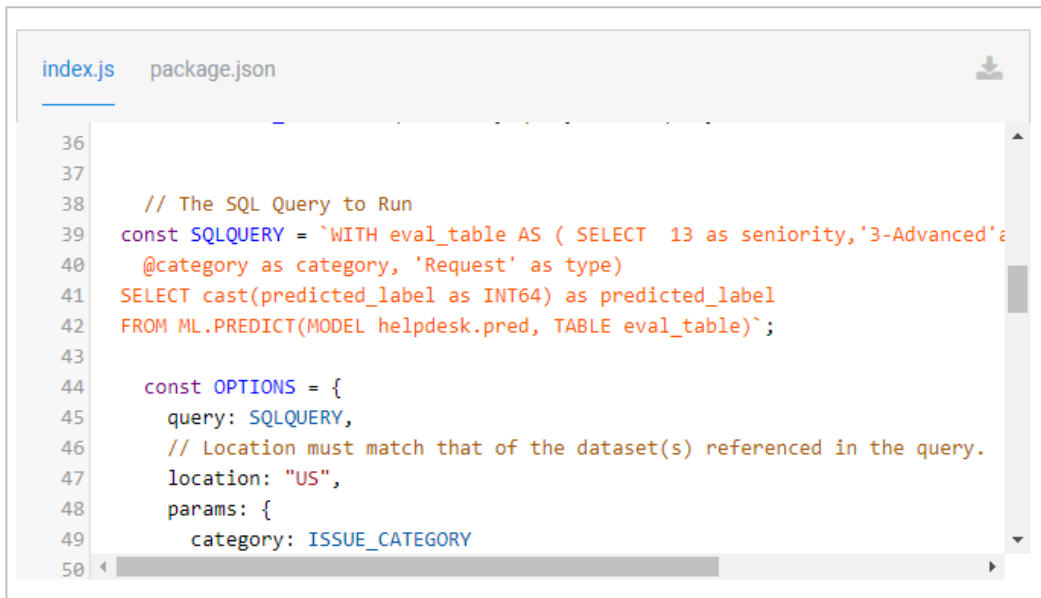
submitticket-issuecategory-followup

submitticket-custom-followup

submitticket-email-followup


Tarda 3 días

**9. RESPONDER : CUANTO TIEMPO TARDA UNA PREDICCION PARA UN EXPERTO NIVEL 13, y categoría TECHNICAL . ( Estas respuestas debe obtenerlas del diálogo con el chat Bot )**



```
index.js  package.json
36
37
38 // The SQL Query to Run
39 const SQLQUERY = `WITH eval_table AS ( SELECT 13 as seniority, '3-Advanced' as
40 @category as category, 'Request' as type)
41 SELECT cast(predicted_label as INT64) as predicted_label
42 FROM ML.PREDICT(MODEL helpdesk.pred, TABLE eval_table)`;
43
44 const OPTIONS = {
45   query: SQLQUERY,
46   // Location must match that of the dataset(s) referenced in the query.
47   location: "US",
48   params: {
49     category: ISSUE_CATEGORY
50 }
```

Try it now




Agent

USER SAYS

COPY CURL

technical

 DEFAULT RESPONSE

undefined, your ticket has been created. Someone will you contact shortly. The estimated response time is 2 days.

CONTEXTS

RESET CONTEXTS

submitticket-collectname-followup

submitticket-issuecategory-followup

INTENT

Submit Ticket - Issue Category

Tarda 2 días

## 10. Guardar la información de training, test y los modelos en GS.

## Que deberá subir.

1. Archivo .js, json
2. Añadirme como colaborador
3. Tiempo: 2h

**Cada pregunta vale 1 puntos ( 10/10 )**