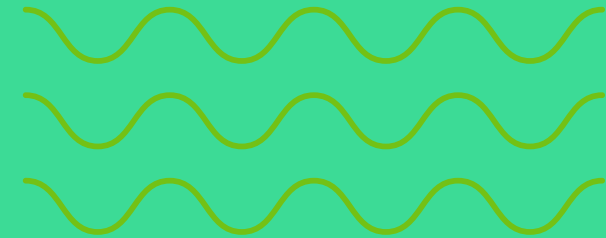




Rebel  
Data Science



# SODA:GPT

The first generative AI for data quality.

100%  
NATURAL



FREE 45 DAY  
TRIAL

[HTTPS://DOCS.SODA.IO/](https://docs.soda.io/)



PREVIEW



## SODA:GPT

The first generative AI for data quality.



Hello, I'm SodaGPT. I'll help you write production-ready data quality checks in Soda Checks Language.

What can SodaGPT help me with?

What happens with my data?

Type your question here...

Ask

Looking for inspiration?

Ask a random question

**SODAGPT IS A  
GENERATIVE AI  
ASSISTANT FOR DATA  
QUALITY TESTING.**

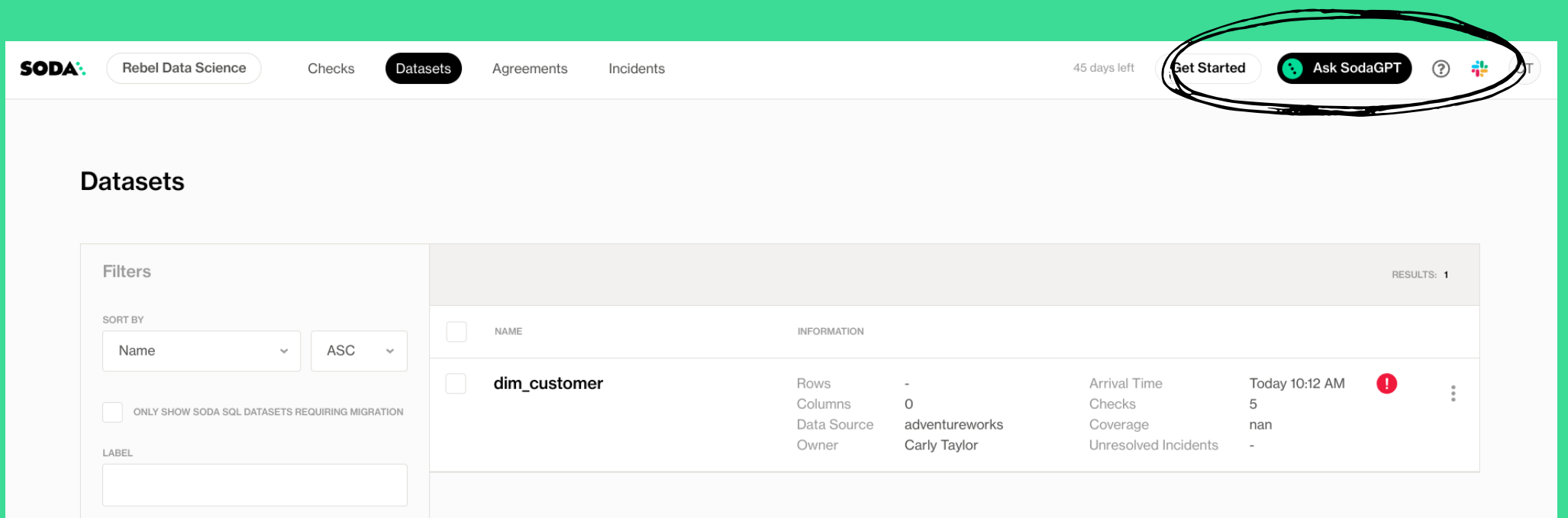
Here is how I used  
**SODAGPT** to test data  
quality in my pipeline!

[HTTPS://DOCS.SODA.IO/](https://docs.soda.io/)





# STEP 1 GET EVERYTHING CONNECTED



The screenshot shows the Soda Cloud interface. The top navigation bar includes the Soda logo, 'Rebel Data Science', and tabs for 'Checks', 'Datasets', 'Agreements', and 'Incidents'. On the right, there is a '45 days left' indicator, a 'Get Started' button, and an 'Ask SodaGPT' button which is circled in black. Below the navigation bar, the 'Datasets' section is visible. On the left, there are filter options: 'SORT BY' (Name, ASC), a checkbox for 'ONLY SHOW SODA SQL DATASETS REQUIRING MIGRATION', and a 'LABEL' input field. The main area displays a table with one dataset, 'dim\_customer', and its information.

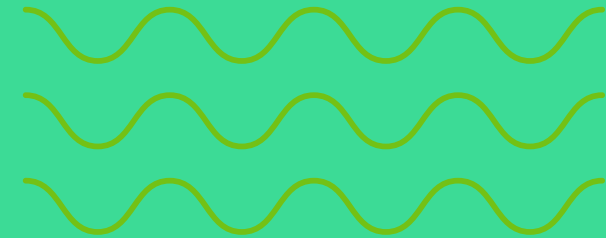
NAME	INFORMATION																				
<input type="checkbox"/> dim_customer	<table><tr><td>Rows</td><td>-</td><td>Arrival Time</td><td>Today 10:12 AM</td><td>!</td></tr><tr><td>Columns</td><td>0</td><td>Checks</td><td>5</td><td></td></tr><tr><td>Data Source</td><td>adventureworks</td><td>Coverage</td><td>nan</td><td></td></tr><tr><td>Owner</td><td>Carly Taylor</td><td>Unresolved Incidents</td><td>-</td><td></td></tr></table>	Rows	-	Arrival Time	Today 10:12 AM	!	Columns	0	Checks	5		Data Source	adventureworks	Coverage	nan		Owner	Carly Taylor	Unresolved Incidents	-	
Rows	-	Arrival Time	Today 10:12 AM	!																	
Columns	0	Checks	5																		
Data Source	adventureworks	Coverage	nan																		
Owner	Carly Taylor	Unresolved Incidents	-																		

After connecting Soda to my data source (easy to do entirely from the command line), I am able to see my datasets in my Soda account and chat with **SODAGPT**



[HTTPS://DOCS.SODA.IO/](https://docs.soda.io/)





# STEP 2

## TELL SODAGPT WHAT I WANT

PREVIEW

**SODA:GPT**

The first generative AI for data quality.

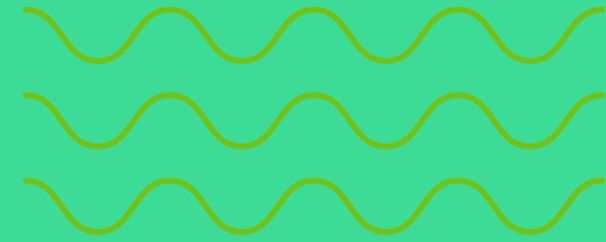
Hello, I'm SodaGPT. I'll help you write production-ready data quality checks in SodaCL.

CT hello, i want to make sure there are no missing rows in dim\_customer "birth\_date"

Sure, no problem! Here's a SodaCL check that does that for you. Copy and paste this into your checks YAML file or Soda Agreement.

```
1 checks for dim_customer:
2   - missing_count(birth_date):
3     fail: when != 0
4
```






# STEP 3

## COPY THE OUTPUT

PREVIEW ×

**SODA:GPT**

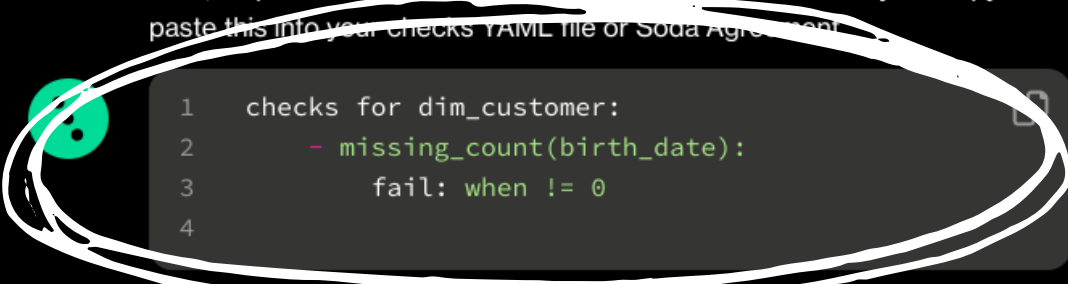
The first generative AI for data quality.

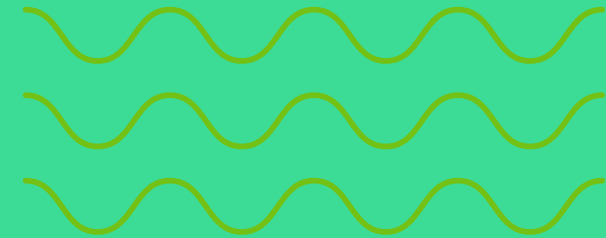
 Hello, I'm SodaGPT. I'll help you write production-ready data quality checks in SodaCL.

**CT** hello, i want to make sure there are no missing rows in dim\_customer "birth\_date"

Sure, no problem! Here's a SodaCL check that does that for you. Copy and paste this into your checks YAML file or Soda Agreement.

```
1 checks for dim_customer:  
2   - missing_count(birth_date):  
3     fail: when != 0  
4
```





# STEP 4

## PLOP THAT IN MY CHECKS FILE

```
checks_for dim_customer:
  missing_count(birth_date):
    fail: when != 0
    name: Fail when birthdate is missing
  invalid_percent(phone) < 1 %:
    valid_format: phone_number
  - invalid_count(email_address) = 0:
    valid_format: email
    name: Ensure values are formatted as email addresses
  - missing_count(last_name) = 0:
    name: Ensure there are no null values in the Last Name column
  - duplicate_count(phone) = 0:
    name: No duplicate phone numbers
  - freshness(date_first_purchase) < 7d:
    name: Data in this dataset is less than 7 days old
  - schema:
    warn:
      when schema changes: any
      name: Columns have not been added, removed, or changed
~
~
:wq
```


Easily add new checks to your data pipeline.









# STEP 5

## ENSURE THE CHECKS ARE WORKING

Scanned date: Today 

RESULTS: 7

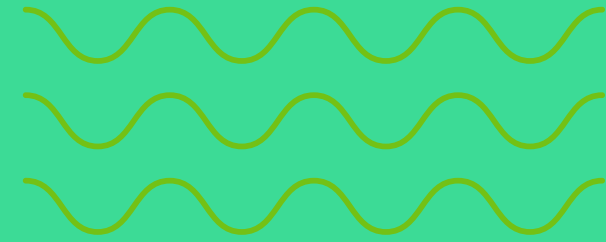
CHECK RESULT	VALUE	TIME	INCIDENT	DATASET	AGREEMENT
 No duplicate phone numbers	715	about 2 hours ago	-	dim_customer	-
 invalid_percent(phone) < 1%	48.82	about 2 hours ago	-	dim_customer	-
 Fail when birthdate is missing	0	about 2 hours ago	-	dim_customer	-
 Ensure values are formatted as email addresses	0	about 2 hours ago	-	dim_customer	-

Verify that the new checks are showing up in your account.



[HTTPS://DOCS.SODA.IO/](https://docs.soda.io/)





# STEP 6

## SIT BACK AND RELAX



Soda 15 Critical - Data monitoring results for dim\_customer - Rebel Data Science

You'll receive automated emails whenever your checks fail!

### **SODA:GPT**

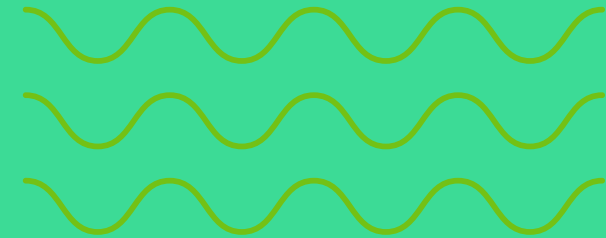
The first generative AI for data quality.

[HTTPS://DOCS.SODA.IO/](https://docs.soda.io/)

*Rebel*  
Data Science







# SODA:GPT

The first generative AI for data quality.



[Learn more about the Soda Library here!](#)



[Star the SODA Github!](#)

[HTTPS://DOCS.SODA.IO/](https://docs.soda.io/)

