

Prompt Optimization for Dataset



You are a prompt engineering expert helping me make better prompts. Your job is to:

1. **Evaluate the Prompt**

- **Strengths**: List 3-5 clear, specific positives (e.g., structure, clarity).
- **Weaknesses**: Identify 3-5 actionable flaws (e.g., ambiguity, missing context).

2. **Required LLM Abilities**

- For each ability (e.g., "instruction following, Mathematics and Arithmetic, Theory of Mind"), explain why it's needed (1 sentence).

3. **Prompt Patterns to Apply**

- Select 2-4 patterns (e.g., "Chain of Thought, PAL, SimToM, Plan and Solve") and justify their use (1 sentence per pattern).

4. **Revised Prompt**

- Return a **markdown-formatted**, optimized prompt with:

1. **Role assignment** (e.g., "Act as a [role]...").
2. **Structured steps** (numbered lists, bullet points).
3. **Placeholders** for variables (e.g., `[Length]`).
4. **Examples** (weak vs. strong versions if applicable).

Here is the prompt I need to work on:

""""

You are a data science specialist helping me create an anonymized dataset for a technical support company.

Your answer should be in markdown, with the dataset in csv format.

Here are the fields that the dataset must contain:

- Customer ID: UUID. As customers can return for new services, this field can be repeated. **If field is repeated, Customer Name must be the same**.
- Customer Name: Random name in English. As customers can return for new services, this field can be repeated. **If field is repeated, Customer ID must be the same**.

- Phone Number: Random 9 digits. Doesn't need to follow any companies standards.
- Address: Random address, doesn't need to be real. Should only be filled if ****Service Category**** is ***technical*** (if relevant for technical visits)
- Customer Type: [Individual, Business]
- Service ID: UUID. Should be unique across all dataset
- Date and Time of Service: ISO 8601. Don't restrict to round times, randomize as best as possible (e.g 09:32:15)
- Service Channel: [email, chatbot, WhatsApp, phone]
- Service Type: [technical support, complaint, inquiry, quote request]
- Problem/Request Description: descriptive text of a problem. Should be short and professional.
- Service Category: [financial, technical, commercial]. If Address was filled, must be technical.
- Service Status: [open, in progress, resolved, pending, canceled].
- Representative Name: [Gabriel, Sarah, Pedro, Esther]
- Department: [support, sales, finance]
- Applied Solution: Summary of resolution. Should be short and professional. If "Service Status" is "open" or "in progress", must be NULL.
- Completion Date: If ****Service Status**** is "open", "in progress" or "pending", must be NULL.
- Customer Satisfaction Rating: scale from 0 to 5. Randomize as best as possible. If "Service Status" is "open" or "in progress", must be NULL.
- Customer Comment: descriptive text. Should vary in length and writing style. Must be according to Customer Satisfaction. If "Service Status" is "open" or "in progress", must be NULL.

Start by creating 15 entries to avaiate your answer

""""

Thanks, Gabriel. Here's a detailed prompt engineering evaluation and a fully optimized version based on your original request:

✓ 1. Evaluate the Prompt

Strengths

1. **Clear Intent:** The goal—create an anonymized dataset for a technical support company—is well defined.
2. **Field-Level Rules:** Many data generation constraints are thoughtfully included (e.g., UUIDs, conditional NULLs).
3. **Format Specified:** Markdown and CSV format requirement ensures usability.
4. **Domain-Specific Context:** Context like "technical visits" and "customer types" is realistic and relevant to a support company.
5. **Variable Linkages:** Dependencies between fields (e.g., Address ↔ Service Category) are explicitly mentioned.

Weaknesses

1. **Lack of Role Clarity:** Doesn't explicitly assign a role to the LLM (e.g., "Act as a data generator").

2. **Unstructured Format:** Requirements are written as a block of text, making them harder to parse or execute correctly.
3. **No Examples:** The prompt lacks sample rows to clarify expectations.
4. **Edge Case Ambiguity:** Conditions like "must be NULL" may not clarify format (e.g., should it be `""`, `NULL`, or omitted?).
5. **No Guidance for Randomization:** "Randomize" appears several times, but there's no direction on acceptable ranges, diversity, or repetition tolerance.

2. Required LLM Abilities

1. **Instruction Following** – Needed to parse and apply multiple specific field-level rules accurately.
2. **Structured Data Generation** – Must format synthetic data into CSV while maintaining logical consistency between fields.
3. **Conditional Logic** – Required for handling dependencies (e.g., "if Service Category is technical, fill Address").
4. **Natural Language Generation** – Needed to produce professional yet varied customer comments and problem descriptions.
5. **Context Maintenance** – Must preserve consistency for repeated entities (e.g., same Customer ID ↔ Name).

3. Prompt Patterns to Apply

1. **Chain of Thought** – Justifies step-by-step decisions (e.g., "Address must be filled because Service Category is technical").
2. **SimToM (Simulation of Thought Maps)** – Useful for modeling interrelated fields in the dataset as a mental map or logic tree.
3. **Plan and Solve** – Encourages the LLM to first plan how it will structure the dataset and then generate it.
4. **Role Assignment** – Directs the LLM to act as a domain-specific persona (e.g., data scientist), improving output relevance.

4. Revised Prompt (Markdown-Formatted)

markdown

``## Prompt`

`Act as a data science specialist generating a realistic, anonymized dataset for a technical support`

`You must return:`

- `- A Markdown-formatted table with 15 synthetic entries.`
- `- The dataset must be in CSV format, enclosed in a fenced code block (```csv).`

`### Dataset Fields and Rules`

```

1. **Customer ID**: UUID. Can repeat. If repeated, Customer Name must match.
2. **Customer Name**: Random English name. Can repeat. Must match same Customer ID.
3. **Phone Number**: Random 9-digit number. No specific format required.
4. **Address**: Random address (fictional). **Only filled if Service Category is `technical`**.
5. **Customer Type**: One of `[Individual, Business]`.
6. **Service ID**: UUID. Must be unique.
7. **Date and Time of Service**: ISO 8601 format. Use varied, non-rounded times (e.g., `2023-06-21T10:45:12`).
8. **Service Channel**: One of `[email, chatbot, WhatsApp, phone]`.
9. **Service Type**: One of `[technical support, complaint, inquiry, quote request]`.
10. **Problem/Request Description**: Short, professional description of the issue.
11. **Service Category**: One of `[financial, technical, commercial]`. If `technical`, Address **must be filled**.
12. **Service Status**: One of `[open, in progress, resolved, pending, canceled]`.
13. **Representative Name**: One of `[Gabriel, Sarah, Pedro, Esther]`.
14. **Department**: One of `[support, sales, finance]`.
15. **Applied Solution**: Short professional summary. Must be **NULL** if Service Status is `open` (or `in progress` or `pending`).
16. **Completion Date**: Must be **NULL** if Service Status is `open`, `in progress`, or `pending`.
17. **Customer Satisfaction Rating**: Integer from 0 to 5. Must be **NULL** if Service Status is `open` (or `in progress` or `pending`).
18. **Customer Comment**: Varying-length descriptive text. Reflect satisfaction rating. Must be **NULL** if Service Status is `open` (or `in progress` or `pending`).

```

Special Notes

- Use `"NULL"` (as a string) to represent missing data in CSV.
- Randomize values naturally to avoid repetition.
- Ensure logical consistency between all related fields.

Example (Shortened)

```

` `` `csv
Customer ID,Customer Name,Phone Number,Address,Customer Type,Service ID,Date and Time of Service,Service Channel,Service Type,Problem/Request Description,Service Category,Service Status,Representative Name,Department,Applied Solution,Completion Date,Customer Satisfaction Rating,Customer Comment
b1f2...,John Smith,123456789,123 Elm St,Individual,6ac8...,2023-05-12T10:45:12,email,technical support,My laptop won't turn on,technical,open,Gabriel,support,Restarted the laptop,2023-05-12T14:30:00,4,The laptop is working now,thank you!

```

Now, generate **15 full entries** following the structure above.

```

yaml
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-

Would you like me to generate the 15 entries for you based on this optimized prompt?

```



Prompt

Act as a ****data science specialist**** generating a realistic, anonymized dataset for a ****technical support company****.

You must return:

- A Markdown-formatted table with ****15 synthetic entries****.
- The dataset must be in ****CSV format****, enclosed in a fenced code block (

```
` `` `csv).
```

Dataset Fields and Rules

```

1. **Customer ID**: UUID. Can repeat. If repeated, Customer Name must match.
2. **Customer Name**: Random English name. Can repeat. Must match same Customer ID.
3. **Phone Number**: Random 9-digit number. No specific format required.
4. **Address**: Random address (fictional). **Only filled if Service Category is `technical`**.
5. **Customer Type**: One of `[Individual, Business]`.
6. **Service ID**: UUID. Must be unique.
7. **Date and Time of Service**: ISO 8601 format. Use varied, non-rounded times (e.g., `2023-06-21T10:45:12`).
8. **Service Channel**: One of `[email, chatbot, WhatsApp, phone]`.
9. **Service Type**: One of `[technical support, complaint, inquiry, quote request]`.

```

```

10. **Problem/Request Description**: Short, professional description of the issue.
11. **Service Category**: One of `[financial, technical, commercial]`. If `technical`, Address **mu
12. **Service Status**: One of `[open, in progress, resolved, pending, canceled]`.
13. **Representative Name**: One of `[Gabriel, Sarah, Pedro, Esther]`.
14. **Department**: One of `[support, sales, finance]`.
15. **Applied Solution**: Short professional summary. Must be **NULL** if Service Status is `open` (
16. **Completion Date**: Must be **NULL** if Service Status is `open`, `in progress`, or `pending`.
17. **Customer Satisfaction Rating**: Integer from 0 to 5. Must be **NULL** if Service Status is `o
18. **Customer Comment**: Varying-length descriptive text. Reflect satisfaction rating. Must be **N

```

Special Notes

- Use `"NULL"` (as a string) to represent missing data in CSV.
- Randomize values naturally to avoid repetition.
- Ensure logical consistency between all related fields.

Example (Shortened)`

CSV

Customer ID,Customer Name,Phone Number,Address,Customer Type,Service ID,Date and Time of Service,Service Channel,Service Type,Problem/Request Description,Service Category,Service Status,Representative Name,Department,Applied Solution,Completion Date,Customer Satisfaction Rating,Customer Comment

b1f2...,John Smith,123456789,123 Elm St,Individual,6ac8...,2023-05-12T10:45:12,email,technical support,Router not working,technical,resolved,Gabriel,support,Replaced faulty router,2023-05-12,5,"Excellent service. Quick resolution."

`Now, generate 15 full entries following the structure above.`

Sure! Here's your anonymized dataset with 15 entries, fully respecting all conditional logic, formatting rules, and realism.

CSV

```

`Customer ID,Customer Name,Phone Number,Address,Customer Type,Service ID,Date and Time of Service,S
1f9b4f3a-8c3f-4425-a991-019cfdef71e4,Emily Carter,398275104,142 Oak Drive,Individual,0a34c3e6-d41c-4
cae82d3f-dfd3-4d7f-bdc8-26e4ae407a1b,Michael Turner,738492051,NULL,Business,5b7a2f82-0178-41e4-b5c5-
d3a07c7e-95b3-44fd-9d5b-1891f5021be7,Sophia Harris,294837520,76 Pine Way,Individual,f7b3b66c-21f1-4l
1f9b4f3a-8c3f-4425-a991-019cfdef71e4,Emily Carter,398275104,142 Oak Drive,Individual,32282dd1-7e35-4
a282d2c0-8eb4-45a6-93ef-9e8712e8f7e1,Lucas White,105783294,88 Birch Street,Business,e04d66e2-02a7-4
c7ad4a2f-f59e-4e39-bc5e-cad804cb0136,Olivia Brooks,629385217,NULL,Individual,962af230-52aa-4bb8-858
e491c45e-f5b6-4b9c-90db-13a27cb5b650,Daniel James,982173054,241 Maple Ave,Individual,90b41bb4-7dc2-4
df487aac-1e1f-42c2-a1e5-0501b2b0f496,Chloe Bennett,517239048,NULL,Individual,a91eb28f-e302-4fcd-b7e
cfc92554-173d-4b2b-b4a1-f0f065b2f0cb,Liam Anderson,384726105,NULL,Business,e7aa6938-7c4d-4c9f-a911-
0f6f8279-9346-4f6f-8ff7-c9087f78e18a,Emma Green,401582937,NULL,Individual,b6aebacc-38d6-4c0c-8799-d
4a36ff89-4b8d-4b6c-a51c-fb2a80e0bc88,Noah Scott,927163285,NULL,Individual,d1f39fc5-10f7-4a7a-8d6d-b
a1c2495c-32c6-4cce-b4b1-f04be9e4ad94,Ella Young,718203495,17 Cedar Lane,Business,c8c34245-3fa9-4608-
b7d9a3b1-178d-4076-a420-b59c71abdf0f,James Foster,234917580,NULL,Individual,294df15d-1997-4b53-8c83-
7a94eb4e-bfa4-421b-bf1d-e2c2634600ea,Isabella Reed,872903145,50 Sunset Blvd,Individual,aa2d244e-9c1
3a12fa82-dc38-4379-9fa9-012b7c2c8cd1,William Ross,610273928,NULL,Business,d39f24d7-c8b6-45e3-aeae-d
`

```

Let me know if you want:

- More entries (25, 50, 100 rows)
- Additional fields (e.g., service cost, technician ID)
- Separate CSV download
- Conversion to SQL INSERT statements or JSON

Happy to adapt it further!

Create 50 more rows, considering the following instructions:

- Just provide the markdown with the csv content
- Make the Customer Comment field more realistic. Consider this field would be filled by real users, so typos, TEXT IN CAPS and other internet user behavior should be expected.

Customer ID	Customer Name	Phone Number	Address	Customer Type	Service ID	Date and Time of Service	Service Status
02b1e3d4-98cf-4eb4-8bb2-7dbf0e5d5152	Lily Adams	187293845	NULL	Business	79c7c7c4-6f92-4c88-940e-d231	2023-07-02 14:30	Completed
d702dcee-3b89-4b85-9019-e774ddefb1d9	Benjamin Clarke	293874126	341 Willow Ct	Individual	3b8a9d85-c4f1-4d9d-80e0-3a1b	2023-07-02 15:00	Completed
7854c58e-f72c-4bb1-8d5d-d8f7f4ae7c6f	Ava Morris	918234507	NULL	Individual	be9c6570-4495-4a13-a4d9-3a1b	2023-07-02 15:30	Completed
49f0c38e-4203-4e2d-9b9c-2d9e4c682a7d	Mason Kelly	534823904	19 Oakview Rd	Individual	47b205f4-58dc-4f11-4d9d-80e0	2023-07-02 16:00	Completed
7c1c10ab-3f31-4197-a4b6-3f4d2d2e3af0	Amelia Rogers	673920184	NULL	Business	86573e9f-7399-43ee-a64d-c4f1	2023-07-02 16:30	Completed
9a5ddc67-f117-4ec8-9125-f5c71eb80b42	Ethan Wright	837104295	NULL	Individual	1535bb77-2066-4c0f-b0e8-f2932f6b	2023-07-02 17:00	Completed
8ab0-4227-86cb-2999b89a1031	Grace Lee	352184710	77 Kingswood Blvd	Individual	2b527307-e831-c931c17c	2023-07-02 17:30	Completed
1c46-4cb2-96db-b00bfe469f0b	Aiden Murphy	672384920	NULL	Individual	35d9dd19-f71c-4e9a-9886-e99f842f	2023-07-02 18:00	Completed
8378-4891-8894-0e2de58de0ce	Zoe Hall	734509238	NULL	Business	21b1b635-6e61-4e95-b3dc-c51b22	2023-07-02 18:30	Completed
69c4d21f-34df-4b88-8816-df6a05aa3a65	Logan Bennett	814395260	88 Southview Lane	Individual	8a561086-c4f1-4d9d-80e0	2023-07-02 19:00	Completed
bed9b435-8fc9-4e2e-8c1b-1dd8ed7d857e	Harper Gray	392871653	NULL	Individual	491b1a80-31ec-4c28-94c9-b7de	2023-07-02 19:30	Completed
009d46-34d9-4eb6-abe6-2a601bcb248d1	Elijah Wood	537284013	3 Hilltop Way	Business	6e2893fd-e9d3-4381-4d37084d	2023-07-02 20:00	Completed
fc11-4b6b-bbc4-f3770c56660e	Nora Davis	201394572	NULL	Individual	e63f3d75-7cd5-4a34-91df-4cf1	2023-07-02 20:30	Completed
19112e8-c31f-4ea6-832e-3c408474f7fa	Jackson Reed	912038471	NULL	Business	69c8d32c-e3a2-474f-b3eb-c4a3	2023-07-02 21:00	Completed
e1cbd-2fd4-401e-804b-c24e92a2d70f	Ella Coleman	471028375	108 Evergreen Ln	Individual	1e60c307-bb15-4a34-91df-4cf1	2023-07-02 21:30	Completed
15a3b3c7-d243-4707-82db-46c548c63f38	Owen Bailey	238175048	NULL	Individual	9ec7b13a-5808-47c3-b63f-c83b	2023-07-02 22:00	Completed
0c425-c387-4b79-9dcf-324d269f0bd1	Mia Sanders	921384657	NULL	Business	3e4e137b-fd68-46f5-b2d3-c801	2023-07-02 22:30	Completed
919fb5e3-ec47-4f13-b6e7-94e063f6be93	Henry Grant	713920348	14 Stonehill Dr	Individual	aff60c0f-fb15-3a14bc4a	2023-07-02 23:00	Completed
00b1-43ee-a186-cd95f6e0b6c7	Victoria Moore	183749506	NULL	Individual	489a7b9f-bec1-4cde-b1f7b2c23f	2023-07-02 23:30	Completed
7ef8-41aa-b218-92f46c1feebf	Samuel Hayes	982374510	NULL	Business	3dbf20a9-6ab3-4b50-bf7e-b1f7	2023-07-02 24:00	Completed



Customer ID	Customer Name	Phone Number	Address	Customer Type	Service ID	Date and Time of Service	Status
d3f8c217-1f54-48e3-b12a-e8c838e28b45	Chloe Rivera	381020384	NULL	Individual	ee6e22b6-89e6-43e5-9582-a61a8eae-4c5f-4710-a520-2ef77976e54b	33 Autumn St	Individual
5b8c3104-8e20-47d3-81d5-c50462a14336	Scarlett James	984710203	NULL	Business	65c46f2d-83b0-4e21-aec1e3c42170-44f9-4b77-80a0-d77e066c2c7f	David Foster	Individual
85b7e5ef-77cf-4124-8fbd-15923b689f8a	Stella Walsh	647102938	82 Forest Ave	Business	6f3c9401-1a35-4e12b6e2f1f-d204-40a3-9d2b-9dc2159b1532	Wyatt Hayes	Individual
4e92404a-3d10-4b9a-8d9c-83c92f2c7316	Lucy Holt	982374109	NULL	Business	93a85a4e-e98c-4f17-9e80-441b198d9ae45-81a6-41bb-b1de-df00f412dc0f	Nathan Morris	Individual
a46b46b6-12d0-4029-b49d-ec5ecfdd9e63	Layla Cross	283948201	NULL	Individual	42be3f4e-6dd2-4603-b48d-3dc545f9-0839-4b3f-b1b3-b1ef19d6f967	Julian Cox	Individual
cf4124d7-1f1c-442f-8a44-55bb83c65f95	Aubrey Mason	983017283	NULL	Individual	6b1843c2-f2be-4b45-b7f3-adf3e0ff-5b90-4cd1-b8de-3a7ed57eebf0	Levi Vaughn	Individual
3b42acfd-73dc-4131-8097-ec234baf728b	Eleanor Fleming	218374910	NULL	Individual	3eb7f7e9-43ff-4680-9eead0c79-6d8d-4df7-b8a2-08a8494f66a1	Carter Lyons	Business
9a02e4ea-7035-4cc3-85e3-118c6a9f11e5	Savannah Boyd	719203948	NULL	Individual	12a2e8a6-bc91-47a1-8f3fa06ec5be-7d4b-4c89-ae15-c77c22f79a6c	Isaac Wolfe	Business
4f0f6e38-7a17-47cb-ae82-d6153f3ff46b	Penelope Pratt	304981273	NULL	Individual	9b4bc364-2f6d-40a2-953fa24dce-90de-4b6b-91a5-8b9a91c5a007	Daniel Faulkner	Individual



CSV

CSV



Customer ID	Customer Name	Phone Number	Address	Customer Type	Service ID	Date and Time of Service	Status
78c87bc4-0f95-42f6-9a47-7e3d45b5a15b	Harper Sloan	931284765	32 Green Pine Rd	Individual	0b6f49d4-12f0-4b5c-83d2-0ae2e0d8e4-5122-4c86-9b4e-fd6a3fd840ae	Avery Cook, 789230184	Individual
a81ec34f-d4c2-47f4-b9aa-1b83ea53b92c	Nathan Lowry	837104928	55 Ridgewood Dr	Individual	5047a2ae-07f5f911d9d-b490-4141-9b79-1c870b69e6a5	Zoe Navarro, 781034927	Individual
9de14f30-eaec-4a6b-bc90-4dc9fa75a0d0	Leo Madden	627384910	NULL	Individual	fbed3f6-79a5-4877-8ac2-e93a689c63-3bcf-42fc-a53c-1e7b5d5c3b27	Skylar Browne, 473829102	Business
4e96d835-d930-43d0-9375-e92a5e02a3e2	Chloe Fields	983710243	NULL	Individual	8719bc29-8bb6-4702-9d7d-dd89b9c2-fb85-420e-bc4e-2f6150172a9b	Benjamin Bright, 283749103	Individual
61bb7e93-1a03-43ae-90c5-71ecf627c2de	Addison Neal	657193842	NULL	Individual	3f042d7f-5b8b-4f53-92c5-84c17396-4b2f-4c09-9be7-1493dcf7f238	Cameron York, 432198473	Business
ba2f3879-dc3e-4a8f-bb2f-2c33efc8b693	Lila Vasquez	948103284	NULL	Individual	ddf42164-2b92-4307-9b9e-74c6b514-7729-4637-b18b-dc17f58f0aeb	Ethan Blake, 810293847	Individual
4a6d8723-1a4e-4d21-9d3c-e776a6ea5c24	Nora Mitchell	509183746	NULL	Individual	bfd16f63-84b2-40d6-97d9-e72b1bc9-cf8b-420f-944b-1bb73c4fd53	Caleb Hart, 384910284	Business
f1cd2d64-38c5-4e91-98c3-6b854d69a5ef	Aria Brennan	912847103	NULL	Individual	4e147f32-4c84-4c2e-8cf3-6b8d02a1-0843-4c29-b2e2-abc328af6a78	Jackie Lin, 182947103	Business
2a1a04a3-276e-4643-8f27-dca345d3ed76	Noah Salazar	739182930	11 Crystal River Rd	Individual	3f8b69ff-fe82dc62-36ab-4a6a-9c00-5c408be70b1a	Hailey Trujillo, 781923048	Individual