
SDML ML CASE STUDY

FEBRUARY 2025



PROBLEM DESCRIPTION

You work for ABC Social. The VP of Finance meets with you and says, “You know that almost all of our revenue comes from selling ads on our website, right?”

He continues, “We’re losing too many advertisers. I need you to tell the sales team who we’re about to lose so that we can keep them.”

“Can you do that?”

QUESTIONS

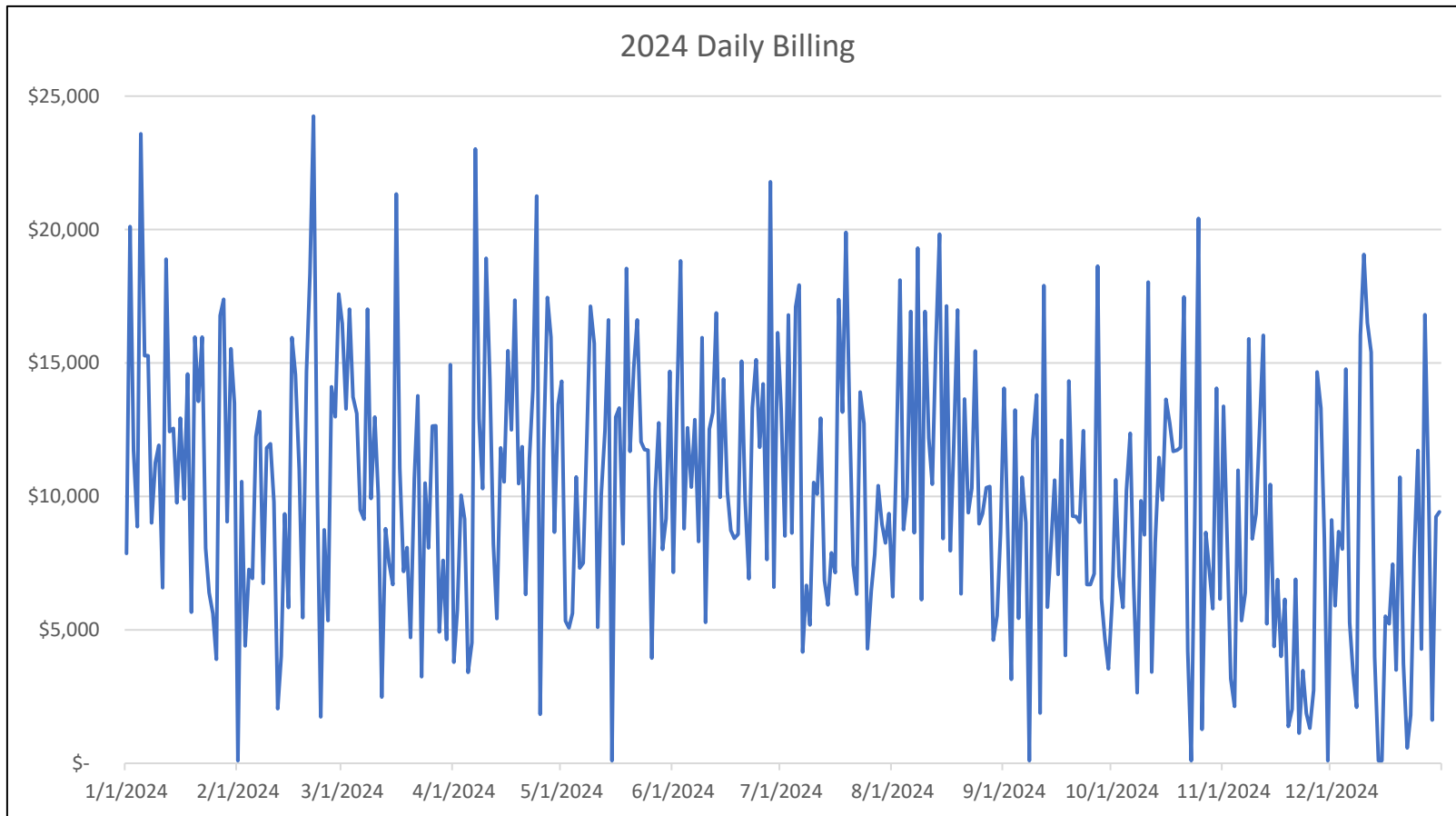
What questions are you going to ask (someone else or yourself)?

What decisions are you going to make?

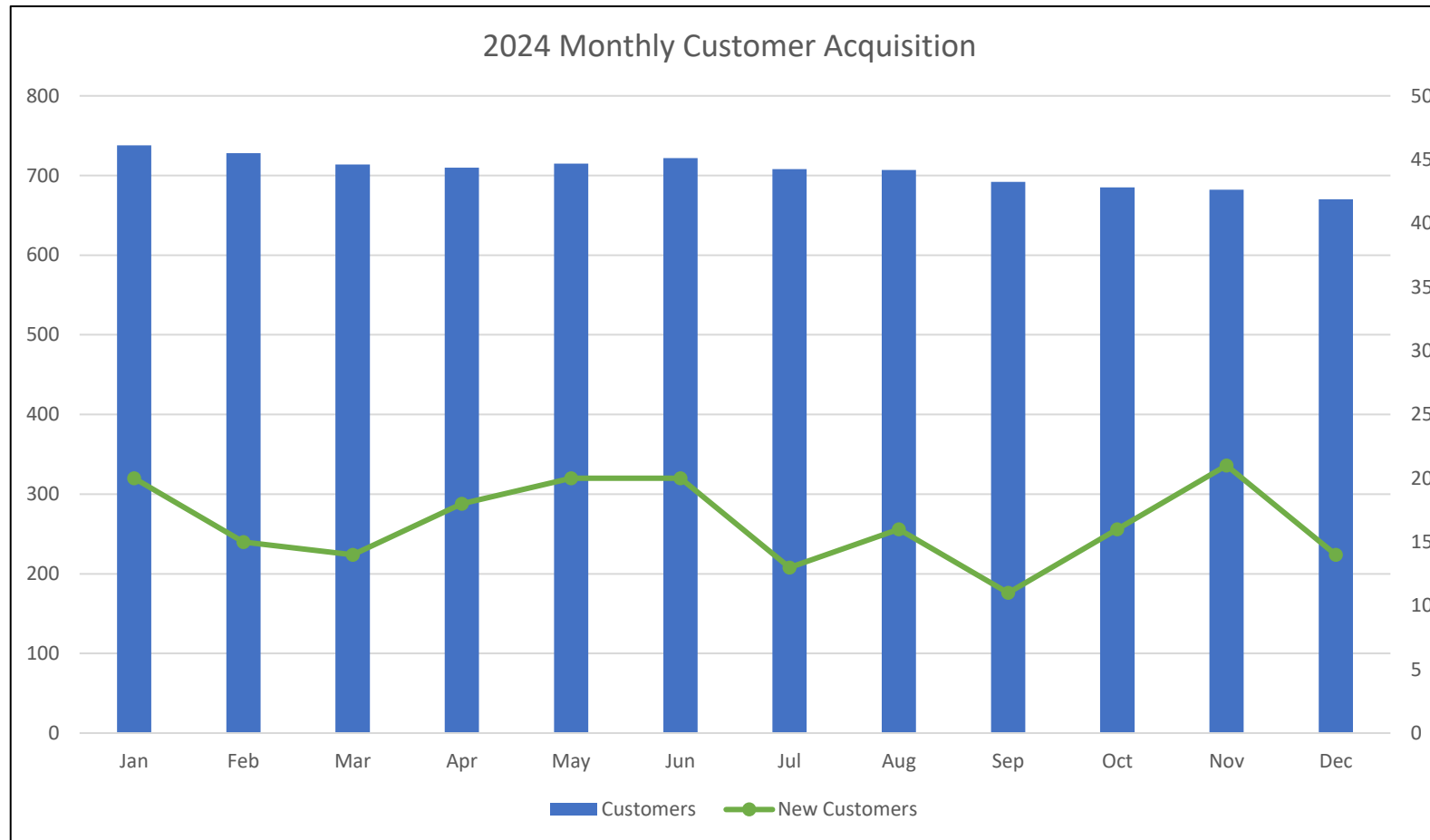
QUESTIONS

How big of a problem is this?

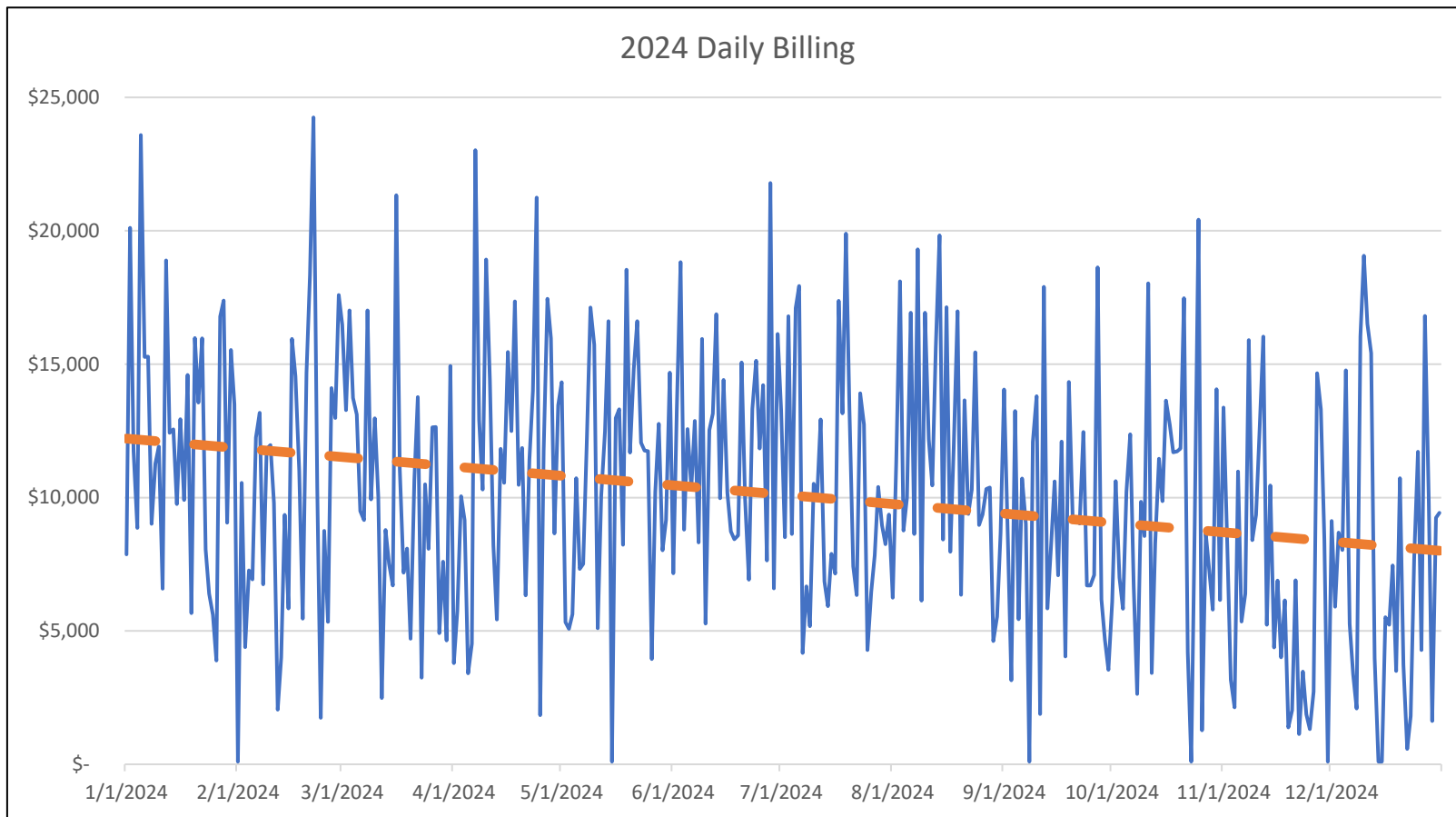
BILLING DATA



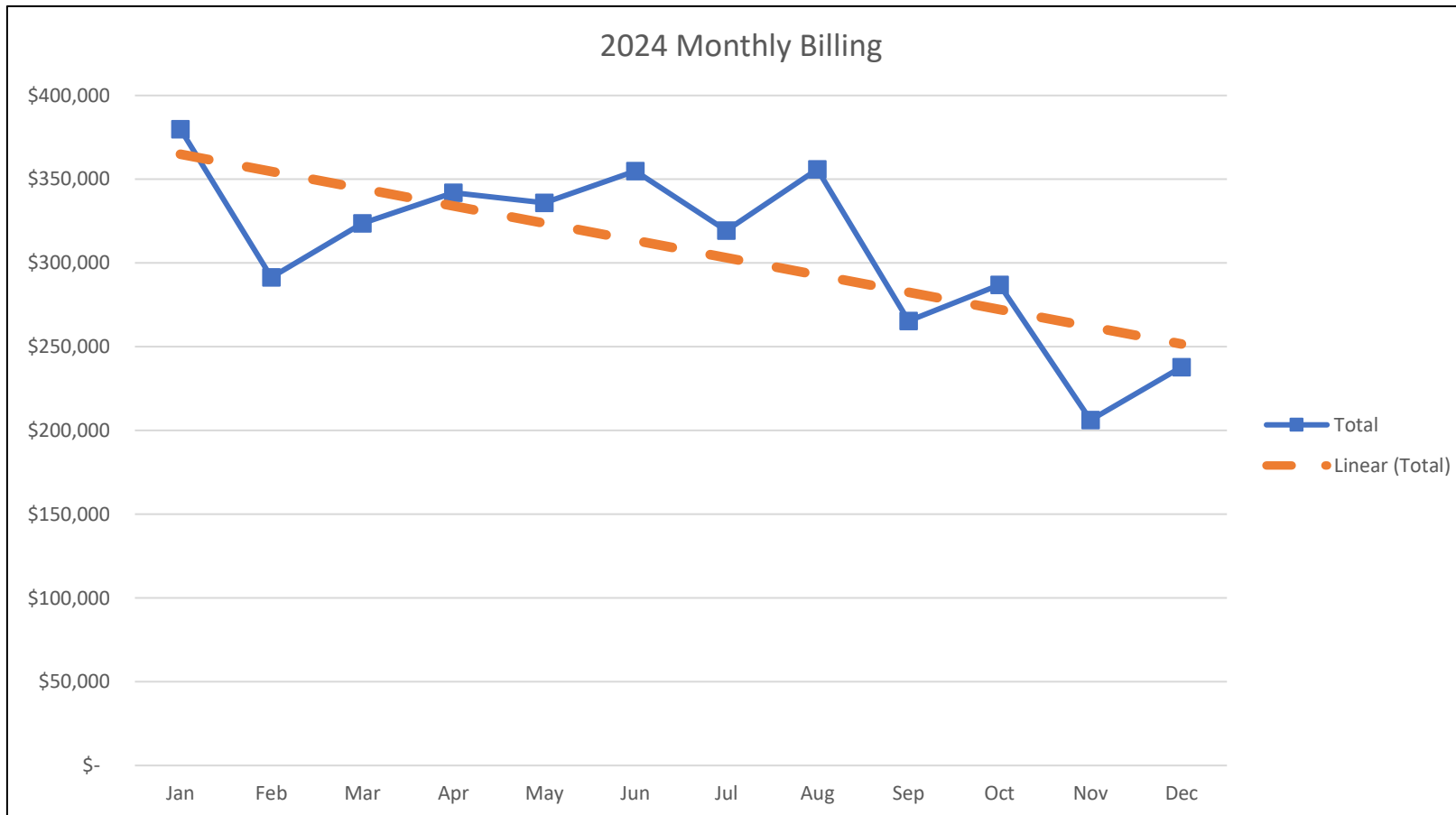
CUSTOMER DATA



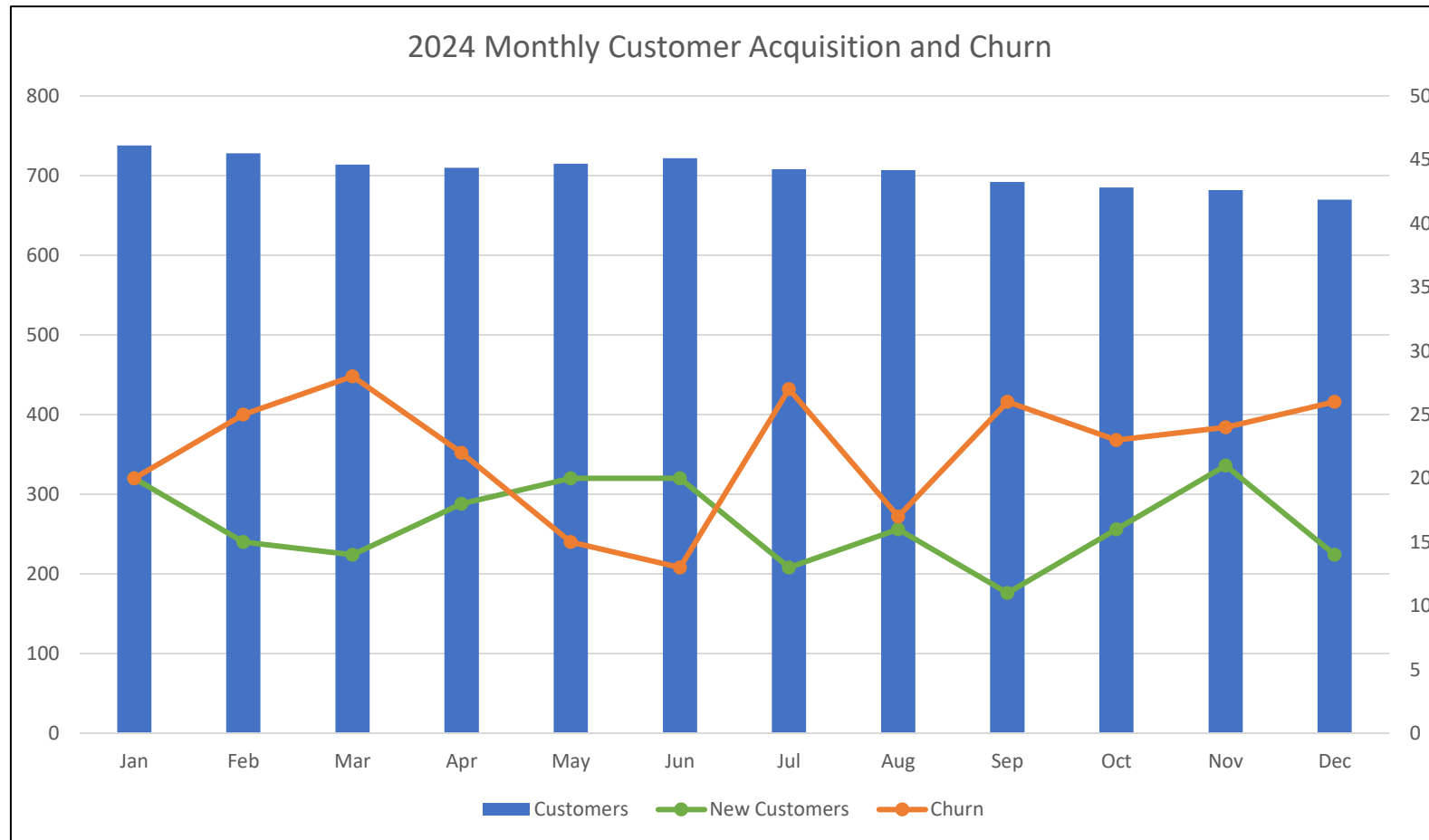
BILLING DATA



BILLING DATA



CUSTOMER DATA



SALES DATA WAREHOUSE

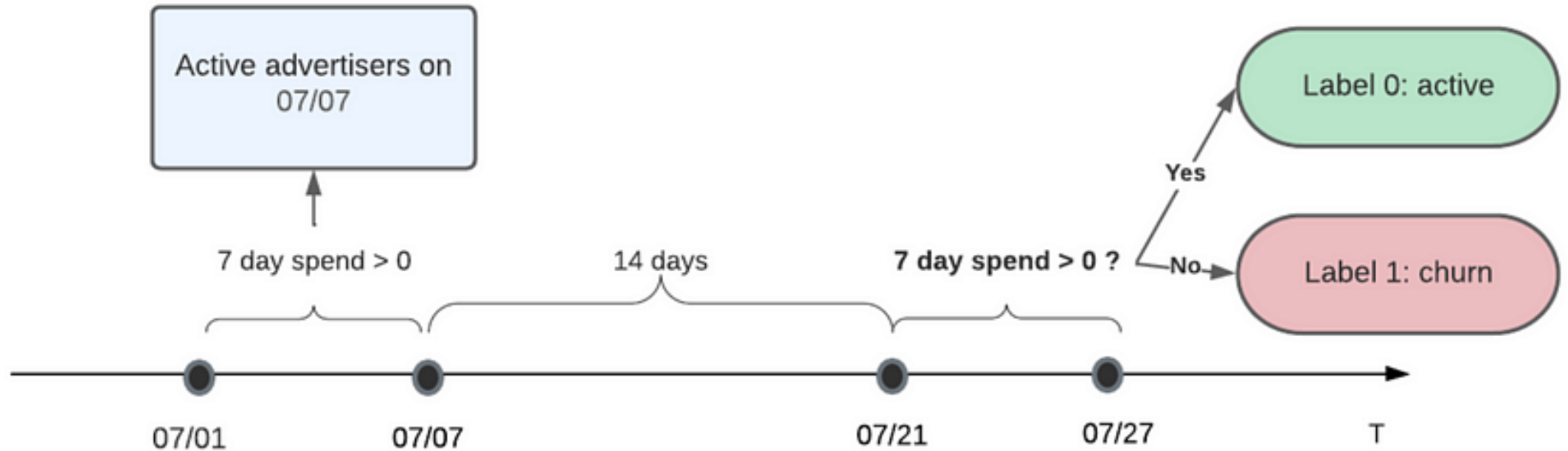
Customer ID	Sale Date	Invoice Date	Paid Date	Volume	Rate	Run Start	Run End	Run Days	Cust Founded	Cust Vertcal	Cust State	Cust Contact	Contact Level	...	Target Age	Target Gender	Target Region	Target State	Target Metro	Target Edu	Target Income	Target Marital	...
301	1/1/2024	1/5/2024	1/31/2024	700	150	1/2/2024	1/8/2024	7	1992	13	CA	Bob Tree	Director		C	F	4	NULL	NULL	2	3	5	
301	1/1/2024	1/5/2024	2/16/2024	250	150	1/2/2024	1/3/2024	2	1992	13	CA	Bob Tree	Director		A	B	1	NULL	NULL	3	4	1	
123	1/1/2024	1/2/2024	1/23/2024	2000	210	1/2/2024	2/15/2024	45	2022	3	TX	Flo Real	Analyst		X	F	NULL	TX	NULL	1	1	1	
299	1/1/2024	1/31/2024	2/7/2024	300	100	1/8/2024	1/14/2024	7	2007	1	NY	Tom New	Manager		B	M	2	NULL	NULL	4	6	2	
301	1/2/2024	1/5/2024	1/31/2024	500	175	1/5/2024	1/7/2024	3	1992	13	CA	Bob Tree	Director		C	M	NULL	VA	NULL	3	2	2	
444	1/2/2024	1/12/2024	2/26/2024	5000	120	1/3/2024	3/31/2024	89	2004	21	WA	Amy Mya	Director		E	M	1	NULL	NULL	2	2	6	
123	1/2/2024	1/3/2024	1/23/2024	2000	210	1/4/2024	1/4/2024	1	2022	3	TX	Flo Real	Analyst		X	M	1	NULL	NULL	1	3	1	
505	1/2/2024	1/3/2024	2/15/2024	250	190	1/3/2024	1/9/2024	7	2018	7	MA	B Gunn	Analyst		A	C	3	NULL	NULL	4	5	2	
301	1/3/2024	1/5/2024	2/16/2024	700	150	1/4/2024	1/5/2024	2	1992	13	CA	Bob Tree	Director		A	B	5	NULL	NULL	3	3	1	
123	1/3/2024	1/4/2024	1/24/2024	100	230	1/5/2024	1/7/2024	3	2022	3	TX	Flo Real	Analyst		B	F	NULL	FL	NULL	2	2	2	

QUESTIONS

How do you define churn?

QUESTIONS

How do you define churn?



QUESTIONS

What kind of problem set up will we use?

QUESTIONS

What are we predicting?

How does the business benefit
from our predictions?

QUESTIONS

How would you calculate the ROI of this model?

QUESTIONS

What kind of model algorithm or architectures might be good to use?

QUESTIONS

What metrics will be used to measure performance?

How does each metric reflect business value?

QUESTIONS

Why might we measure performance differently, such as on only a subset of the predictions?

QUESTIONS

We have two models. One is slightly higher recall, but the other is slightly higher precision. Which model is better?

QUESTIONS

How do we set a threshold for binary classification?

QUESTIONS

If we want multiple categories of predictions, such as High and Medium risk, how do we set those thresholds?

QUESTIONS

How could we understand and communicate to the sales team why a customer was predicted to churn?

QUESTIONS

What kind of features work well or don't work well with our chosen algorithm or architecture?

REFERENCE

Inspiration for this case study
was drawn in large part from a
Pinterist blog post:

<https://medium.com/pinterest-engineering/an-ml-based-approach-to-proactive-advertiser-churn-prevention-3a7c0c335016>