Analysis Submission

Kasey Zapatka

Step 1: Load data

My code below uses a function to load `channel-spend.rds` into my R session and then calls that function and assigns the output to the `channels` variable.

```
# ============================================================================
# STEP 1: LOAD DATA
# ============================================================================

# STEP 1 --------------------------------------------
# function to load data
load_in_channel_data <- function(data_path){
  data <- read_rds(paste0(data_path, "channel-spend.rds"))
  glimpse(data)
}

# Call your function below
channels <- load_in_channel_data(data_path)
```

```
## Rows: 100
## Columns: 8
## $ influencers         <dbl> 0.1466605, 0.0000000, 0.0000000, 0.0000000, 0.00000~
## $ linear_tv           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ mailers             <dbl> 0.000000, 0.000000, 0.000000, 0.000000, 0.000000, 0~
## $ meta_prospecting    <dbl> 2.111382, 2.258977, 2.626174, 2.780627, 2.925106, 2~
## $ meta_retargeting    <dbl> 0.3804446, 0.4042905, 0.4024203, 0.3972770, 0.37873~
## $ podcast             <dbl> 0.0000000, 0.0000000, 0.0000000, 0.1617786, 0.00000~
## $ search_non_branded <dbl> 0.2805409, 0.3011138, 0.2983084, 0.2813201, 0.29534~
## $ depvar              <dbl> 4.94647, 4.98352, 4.20996, 5.55403, 6.00975, 4.6784~
```

Step 2: Sample from the predict prior

Do you think the priors we have set are reasonable? Yes, or No?

No, I do not think the set priors are reasonable. Summary Table 1 shows that the prior predictive covers a range (7 and 28) that is greater than what we might observe in our data (4 to 9). While I don't expect the priors to perfectly capture the observed data, this is a red flag that suggests the priors do not adequately capture the range of possible, observed values.

Table 1: Summary statistics comparing observed and predicted values.

Statistic	Observed	Predicted
Min	4.210	7.281
1st Quartile	6.890	13.070
Median	7.390	16.326
Mean	7.266	17.913
3rd Quartile	7.892	23.068
Max	9.364	28.278

What is your justification for your answer above?

Panel A in Figure 1 visualizes the observed distribution, with two peaks around `4.9` and `7.5`. The prior predictive in Panel B stands in stark contrast: it shows a large bimodal distribution with peaks around `~13` and `~23`, shifted approximately `~7` units to the right of the observed data, and with both peaks of similar height. Although the prior captures the bimodal shape of the observed data, it is shifted too far to the right to reflect the true parameter values, and the left-hand peak is disproportionately large. Panel C, which plots daily revenue over time, suggests the observed bimodality may stem from a slower uptick in revenue at the start of the study period——perhaps due to seasonality, early-stage operations, or recent market entry.

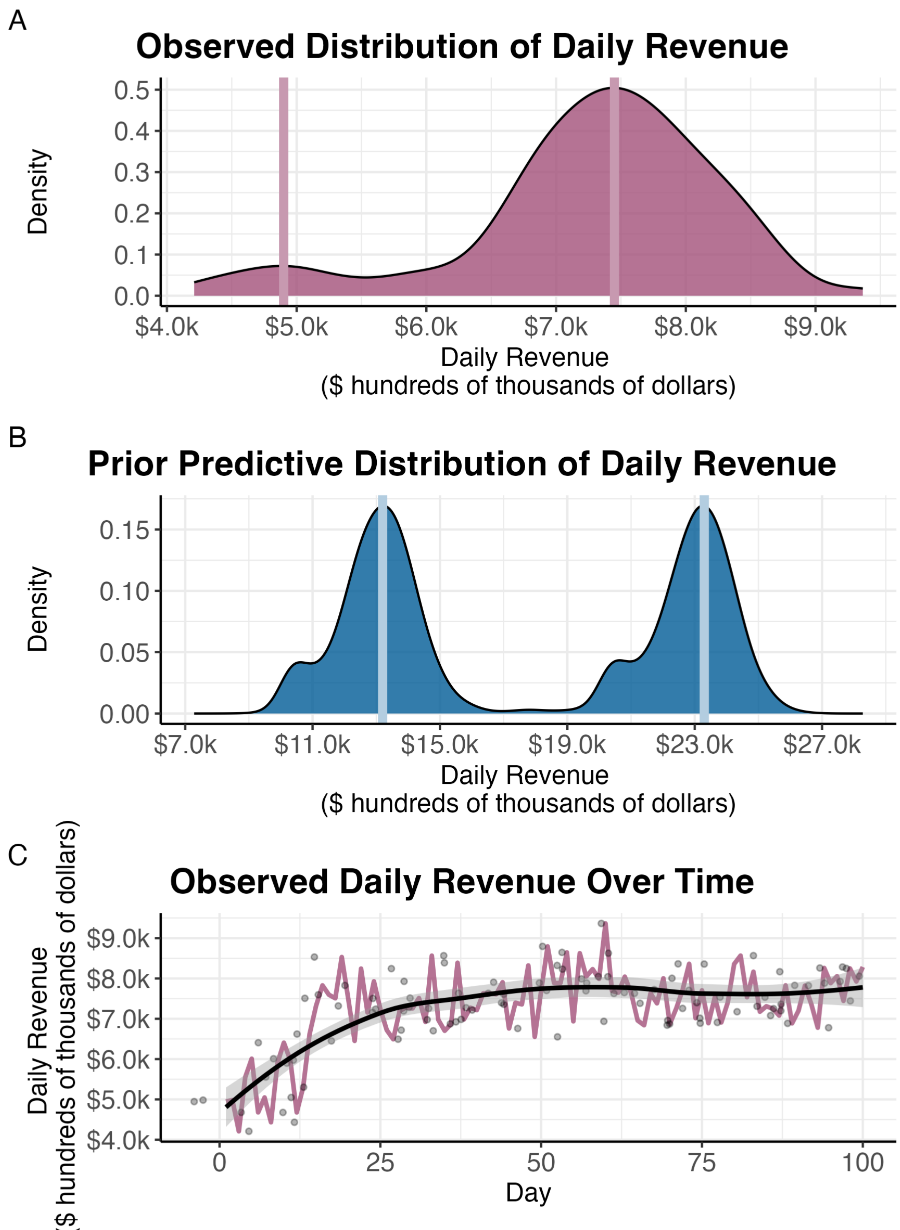

Figure 1: Comparing observed and prior distributions

Step 3: Sample from the posterior prior

What is the output of fit$cmdstan_diagnose()? What does this information tell you about the model you just sampled?

Figure 2 is the output from `fit$cmdstan_diagnose()`. Most of the diagnostics are fine: the model was efficient in exploring the sample space (Treedepth), there was a good mix of samples (E-BFMI), there was enough sample independence (ESS), and R-hat was close to 1. However, there were several divergent transitions, which indicate that the Hamiltonian Monte Carlo sampler failed to adequately explore the posterior. As a result, the model estimates are likely biased and unreliable. I should revise the priors to help the model better explore the full range of plausible parameter values.

```
Checking sampler transitions treedepth.
Treedepth satisfactory for all transitions.

Checking sampler transitions for divergences.
1348 of 1000 (134.80%) transitions ended with a divergence.
These divergent transitions indicate that HMC is not fully able to explore the posterior distribution.
Try increasing adapt delta closer to 1.
If this doesn't remove all divergences, try to reparameterize the model.

Checking E-BFMI - sampler transitions HMC potential energy.
E-BFMI satisfactory.

Rank-normalized split effective sample size satisfactory for all parameters.

Rank-normalized split R-hat values satisfactory for all parameters.

Processing complete.
```

Figure 2: Diagnostic output

If this was a real model for a real client, would you feel comfortable proceeding with this model? Why or why not?

No, I would not feel comfortable proceeding with this model because this diagnostic is telling me the range of possible estimates was not fully explored and that I have biased and unreliable estimates. I would prefer to give clients a model that I am confident fully explored the posterior. This confirms what I suspected in Figure 1 and Table 1—that the priors would not cover the range of reasonable values and the posterior would not match the observed data well. Figure 3 visually confirms this by overlaying the prior predictive on the posterior, illustrating that a large part of the posterior (red) is not covered by the prior (blue), and that both the prior and posterior do not align well with the parameters of the observed data (4.9 and 7.5) from Figure 1.

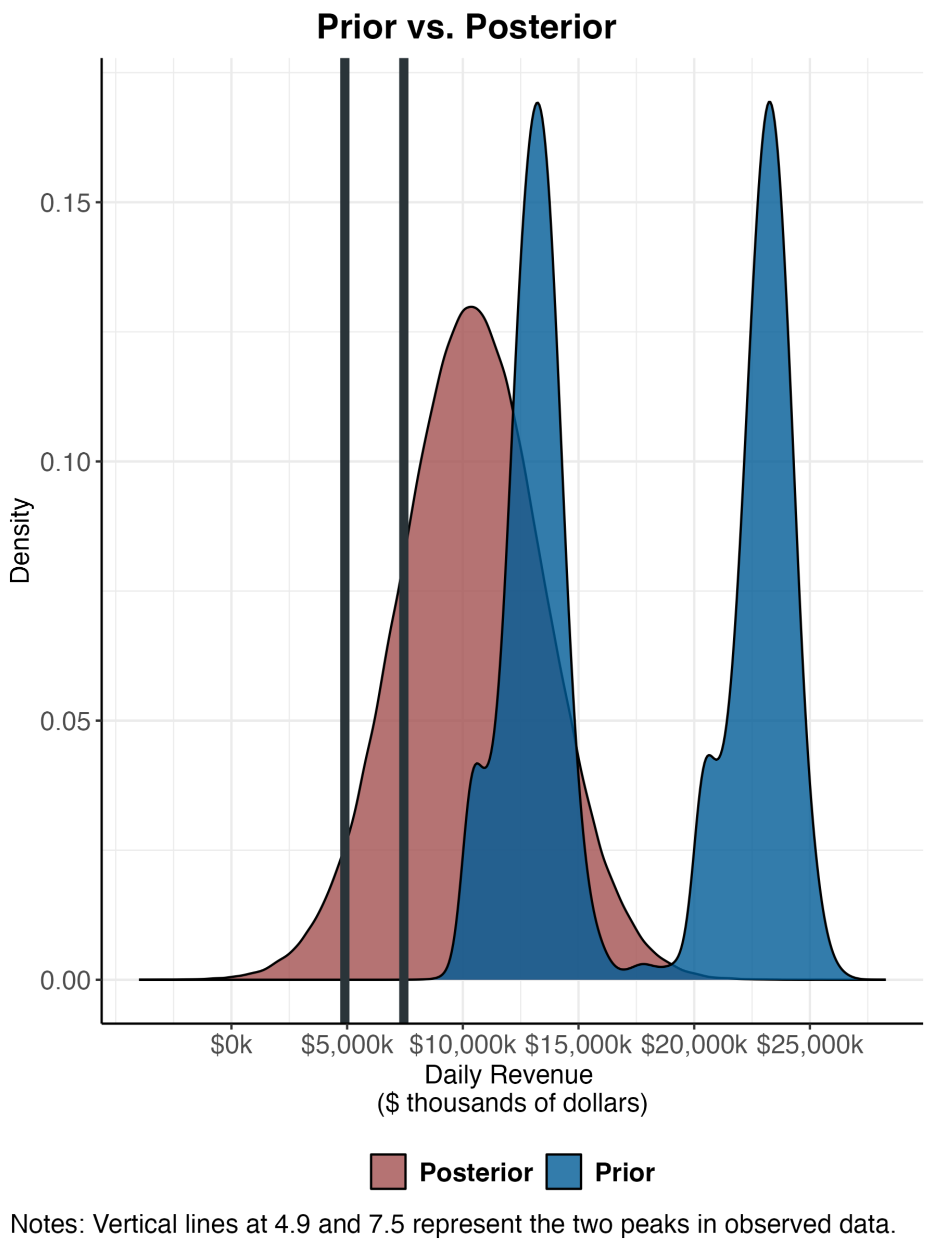

Figure 3: Prior vs Posterior

Step 4: Update priors

If you think this model is deficient in some manner, try editing the priors in order to improve the model. What did you change?

I think the model is deficient and edited the priors to improve the model. Table 2 lists the changes I made:

Table 2: Comparison of original and updated priors for the model intercept.

Parameter	Original	Updated
intercept_lb	10	2
intercept_ub	20	6
intercept_eta_mean	0	0
intercept_eta_scale	100	1

Why did you change the priors you selected?

My main concern was making sure my priors and posteriors covered plausible values observed in the data. So, I narrowed the intercept bounds from 10–20 to 4–10, but prior predictive checks showed the mean was still too far to the right, so I adjusted them further to 2–6, resulting in a mean closer to the empirical baseline of 7.5. I also tightened `intercept_eta_scale` to 1 to concentrate prior mass around the prior mean; updated prior and posterior predictive plots aligned well with the observed data, as shown in Figure 4. The priors for `beta`, `kappa`, `conc`, and `shift` seemed reasonable and didn't need changes after updating the intercept priors. Given the model's sensitivity to temporal structure, I chose not to alter the shift prior at this stage but may revisit it if future checks suggest the need.

How do you know the model is better than before?

I know this model is better than before because (1) the prior includes reasonable estimates I can observe in the data (`4.9` and `7.5`), (2) the updated posterior matches well to the observed data, and (3) the output of `fit_updated$cmdstan_diagnose()` reports no divergent transitions. Panel B in Figure 4 illustrates point 1, Panel C illustrates point 2, and Panel A shows both points 1 and 2 cover the data well.

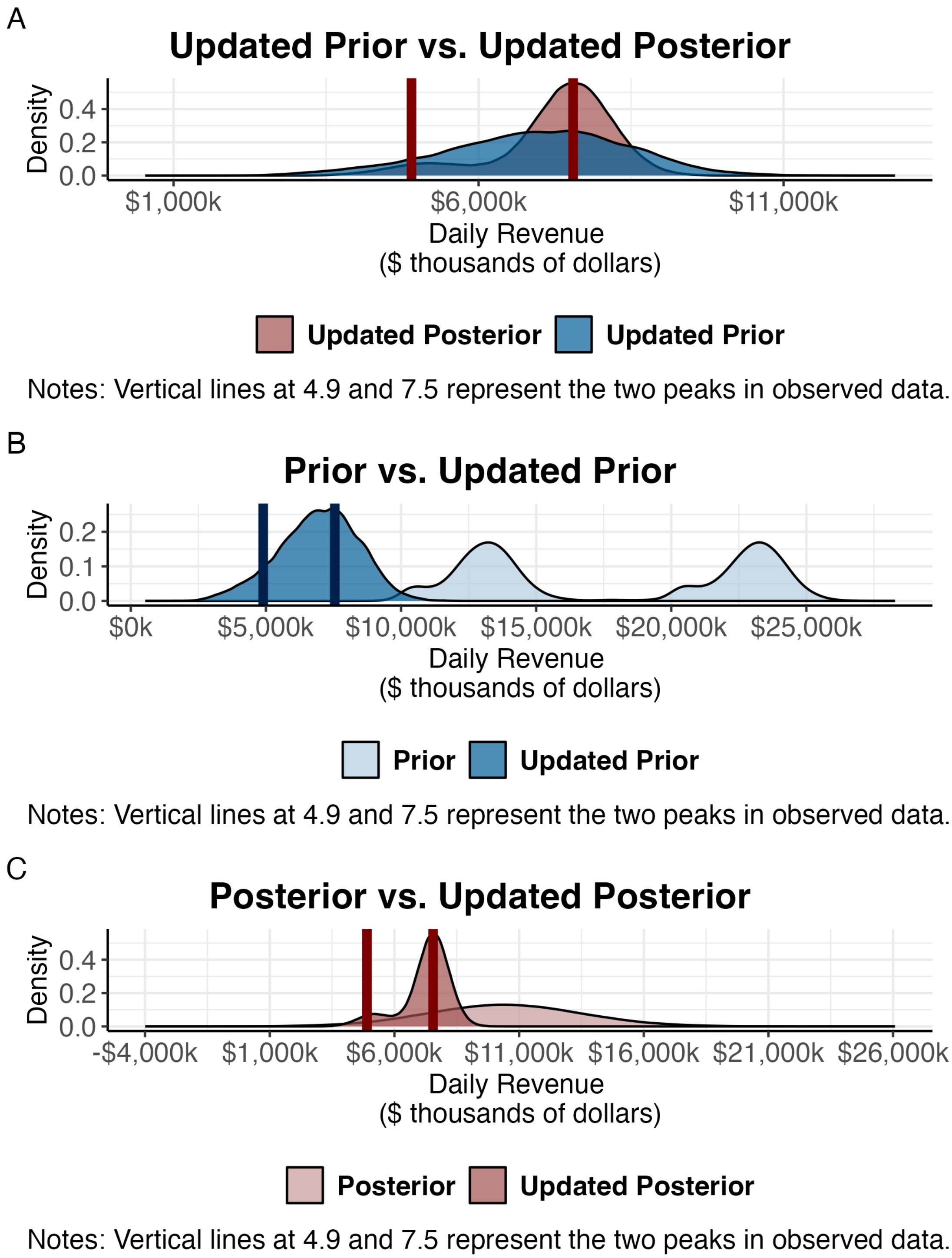

Figure 4: Comparing original and updated priors and posteriors