Commercial Goals: Digital Traffic

Utilizing Synthetic Controls to Baseline Projected Performance

Insights and Analytics | Last Updated: 03 February, 2022

Table of Contents

Top Line

We built a forecast model to set a useful baseline for the business to build from. The business owner for this metric will ultimately set where were they want the business goal to land.

4 different forecasting models were applied, and are displayed below. These different models allow us to understand the potential impact of different scenarios.

- **Model 1 (Low/Flat):** Covid data included, and covid period counted as 3/1/2020 - 5/1/2021, flattened trend.
  - This model produces the smallest forecasted traffic.
- **Model 2 (Mid-point):** Covid data included, includes growth trend
  - This model producess a forecasted traffic that is close to 2019 traffic.
- **Model 3 (High):** Covid data not included (then forecast 3 years out)
  - This model essentially is built with the belief that the trends pre-2020 will continue. Ultimately the business question here is whether shopper behavior will fully return to pre-pandemic behaviors in 2022.

	2017	2018	2019	2020	2021	2022[1]	2022FC_Flat	2022FC_High	2022FC_Mid	Business Goal
Jan	1.24M	1.10M	1.26M	1.23M	0.98M	0.79M	**1.11M**	**1.52M**	**1.38M**	*TBD*
Feb	1.30M	1.19M	1.17M	1.18M	0.92M	-	**1.12M**	**1.51M**	**1.36M**	*TBD*
Mar	1.64M	1.75M	1.78M	0.95M	1.92M	-	**1.47M**	**1.87M**	**1.76M**	*TBD*
Apr	1.82M	1.48M	1.85M	0.31M	1.85M	-	**1.36M**	**1.86M**	**1.70M**	*TBD*
May	1.89M	1.68M	1.69M	1.69M	1.79M	-	**1.47M**	**1.91M**	**1.80M**	*TBD*
Jun	1.89M	1.81M	2.18M	2.12M	1.71M	-	**1.58M**	**1.96M**	**1.89M**	*TBD*
Jul	2.39M	2.38M	2.46M	1.46M	2.57M	-	**1.96M**	**2.41M**	**2.28M**	*TBD*
Aug	2.58M	2.53M	2.59M	1.22M	2.41M	-	**2.08M**	**2.59M**	**2.39M**	*TBD*
Sep	1.62M	1.54M	1.41M	0.86M	1.45M	-	**1.36M**	**1.83M**	**1.65M**	*TBD*
Oct	1.65M	1.60M	1.62M	1.39M	1.56M	-	**1.50M**	**1.95M**	**1.79M**	*TBD*
Nov	2.51M	2.53M	2.55M	2.42M	2.56M	-	**2.19M**	**2.58M**	**2.47M**	*TBD*
Dec	2.30M	2.49M	2.29M	1.86M	1.93M	-	**1.95M**	**2.31M**	**2.23M**	*TBD*
total	22.84M	22.08M	22.85M	16.70M	21.66M	—	19.15M	24.28M	22.71M	—

[1] Note that 2022 actuals represent the current traffic and is not a complete month

Reviewing Historical Data

Overview of Data

Data Generation

Web traffic data, as measured be Google web sessions.

Data Pull from DB

For forecasting purposes we have pulled data for all centers from 2009 until now (2022.01.26).

Source: Laura Simmons (server: ___)

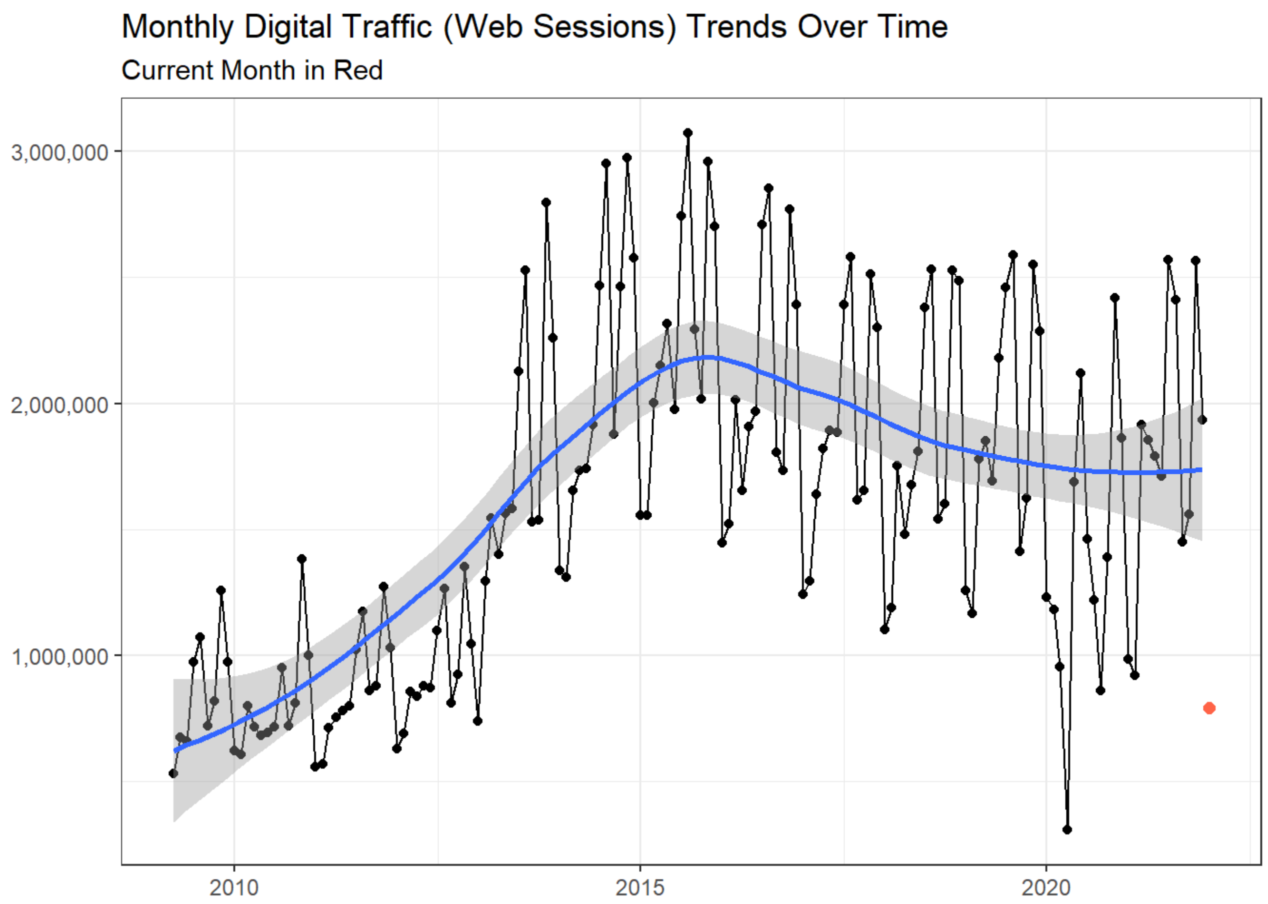
Tables: N/A

Not sure if there was any additional filtering logic done on either LS side or in the data that is available in Google Analytics.

View of Trends

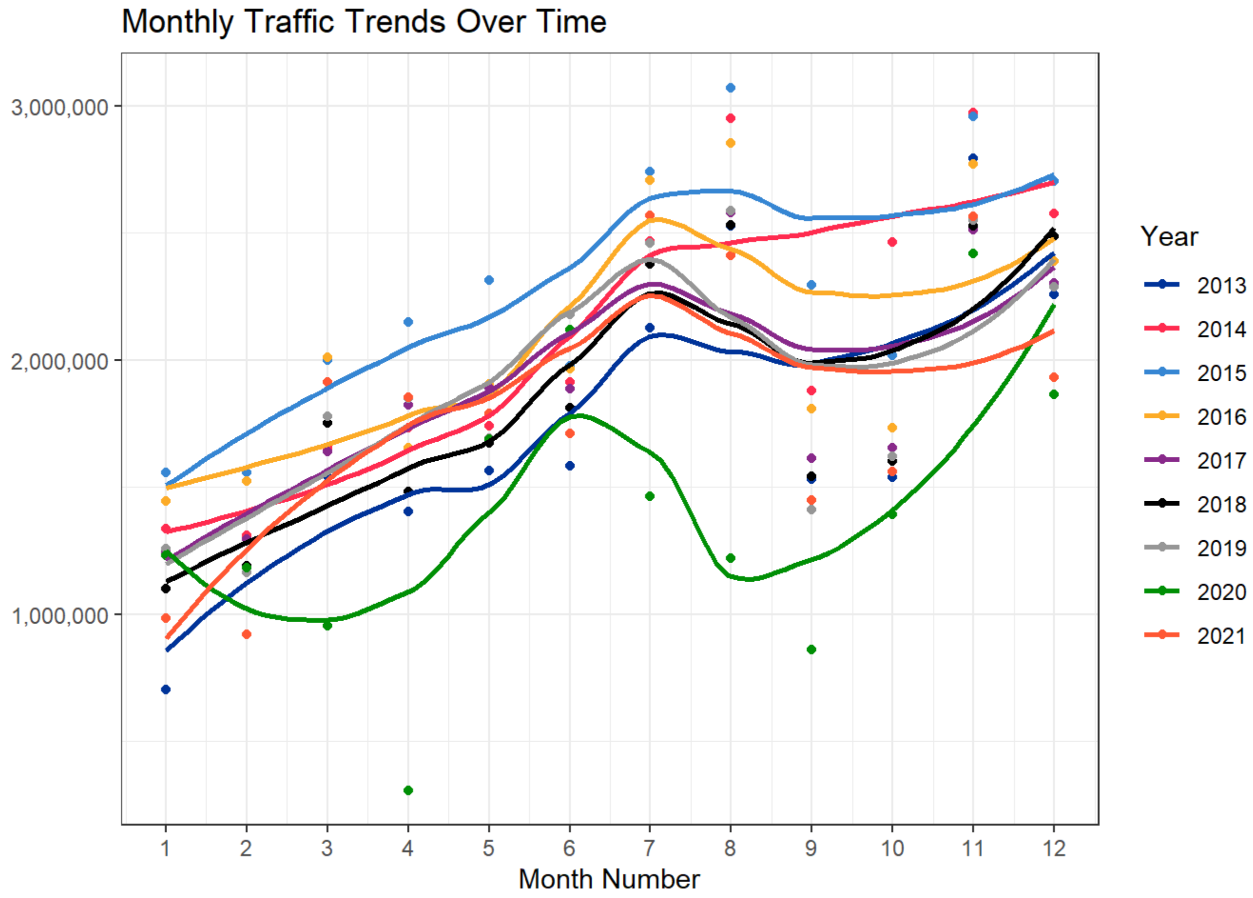
Before we begin to model this work, we want to visualize the general trends. This allows us to to flag for large outliers or anomalies within our data.

- With this data we can see how significant Covid shutdowns had on traffic throughout 2020.
- We also see a fairly consistent month to month seasonal variation.



Monthly View

Each color line represents a different year's worth of traffic aggregated on a monthly basis.


Monthly Traffic Trends Over Time
3,000,000
2,000,000
1,000,000
1
2
3
4
5
6
7
8
9
10
11
12
Month Number
Year
2013
2014
2015
2016
2017
2018
2019
2020
2021

Forcasting model

Why Synthetic Controls?

This method can account for the effects of confounders changing over time, by weighting the control group to better match the treatment group before the intervention. Another advantage of the synthetic control method is that it allows researchers to systematically select comparison groups. It has been applied to the fields of political science, health policy, criminology, and economics.

In our case, we are using this methodology to help set a baseline of where the business would be if we continue with the same historical processes and procedures as before.

Step 1: Build Forecast Model

Building model for all entire portfolio.

- In this case, we are already into 2022, but we want to build the forecast model excluding the new year.
- In the future data frame, we go forwards 365 days (1 year) to be able to predict against.

```
Traffc <-
  CleanTraffc %>%
  filter(Date < as.Date("2022-01-01")) %>%
  mutate(ds = Date) %>%
  group_by(ds) %>%
  summarise(y = sum(Sessions, na.rm=T)) %>%
  ungroup()

# m <- prophet(Traffc, holidays = covid)
m <- prophet(holidays = covid)
m <- add_country_holidays(m, country_name = 'US')
m <- fit.prophet(m, Traffc)

future <- make_future_dataframe(m, periods = 365)
# tail(future)


forecast <- predict(m, future)
# tail(forecast[c('ds', 'yhat', 'yhat_lower', 'yhat_upper')])
```

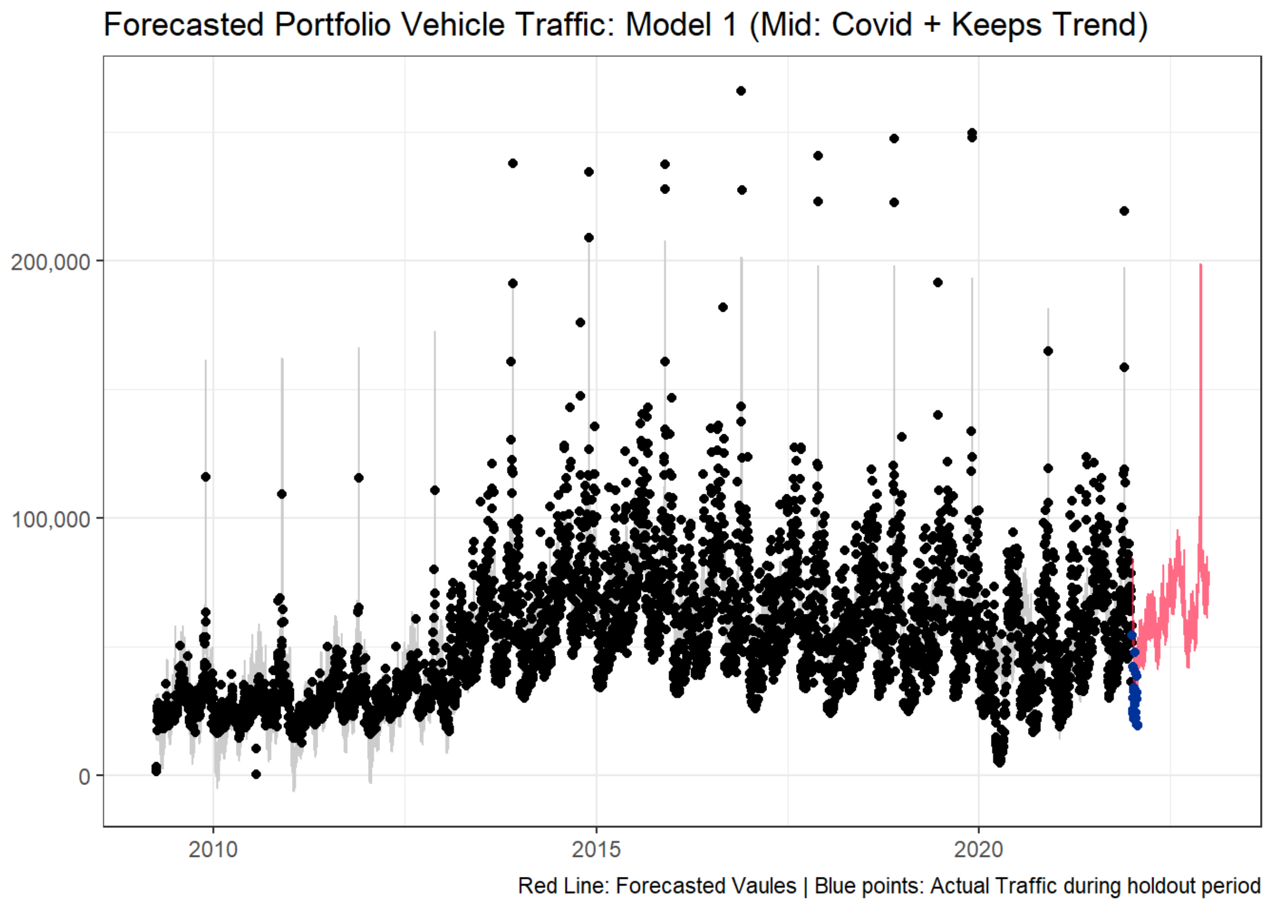
Forecast Results

Below is the forecast model.

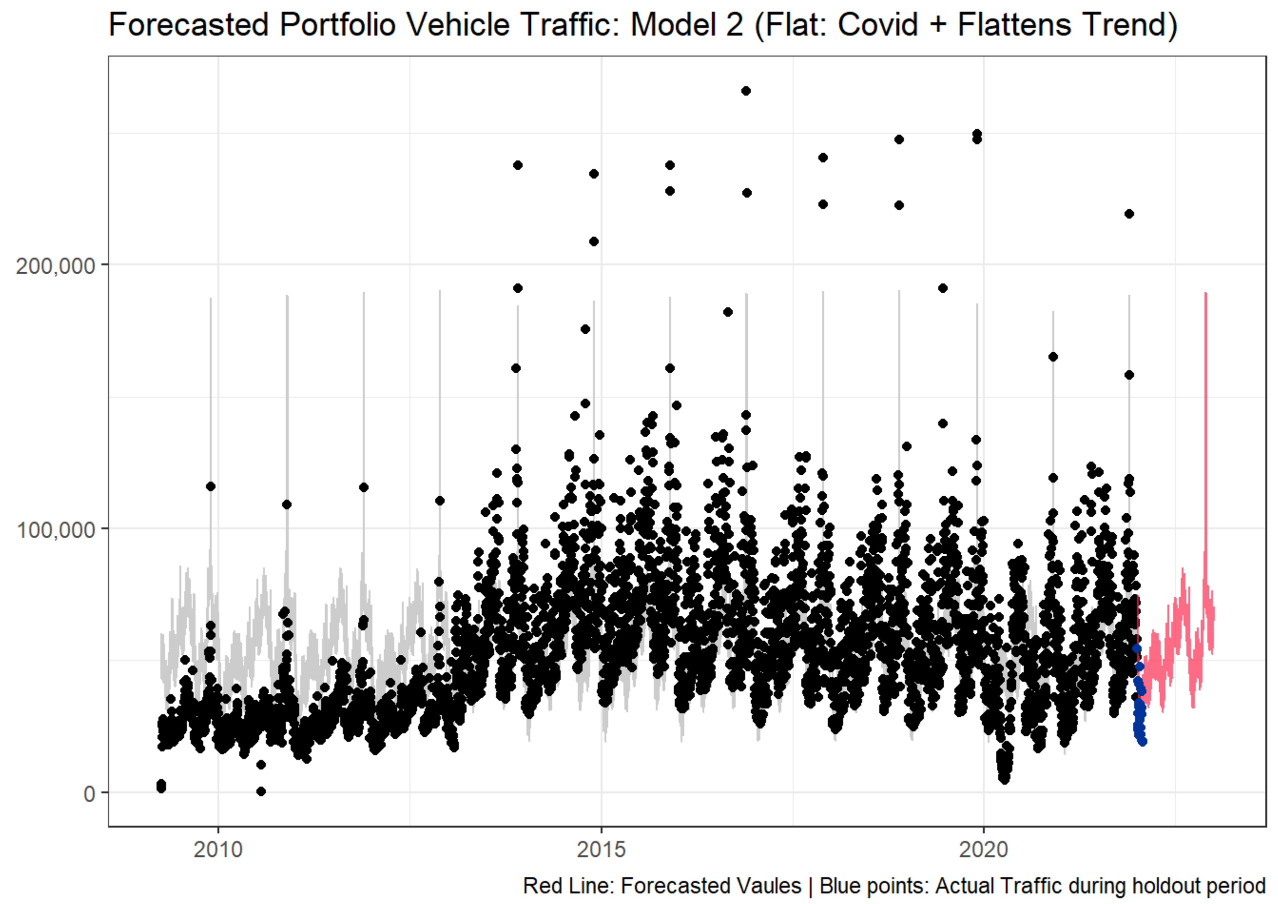
- The black dots represent to the total portfolio traffic by day.
- The blue line represents the forecasted values.

We see a strong degree of seasonality

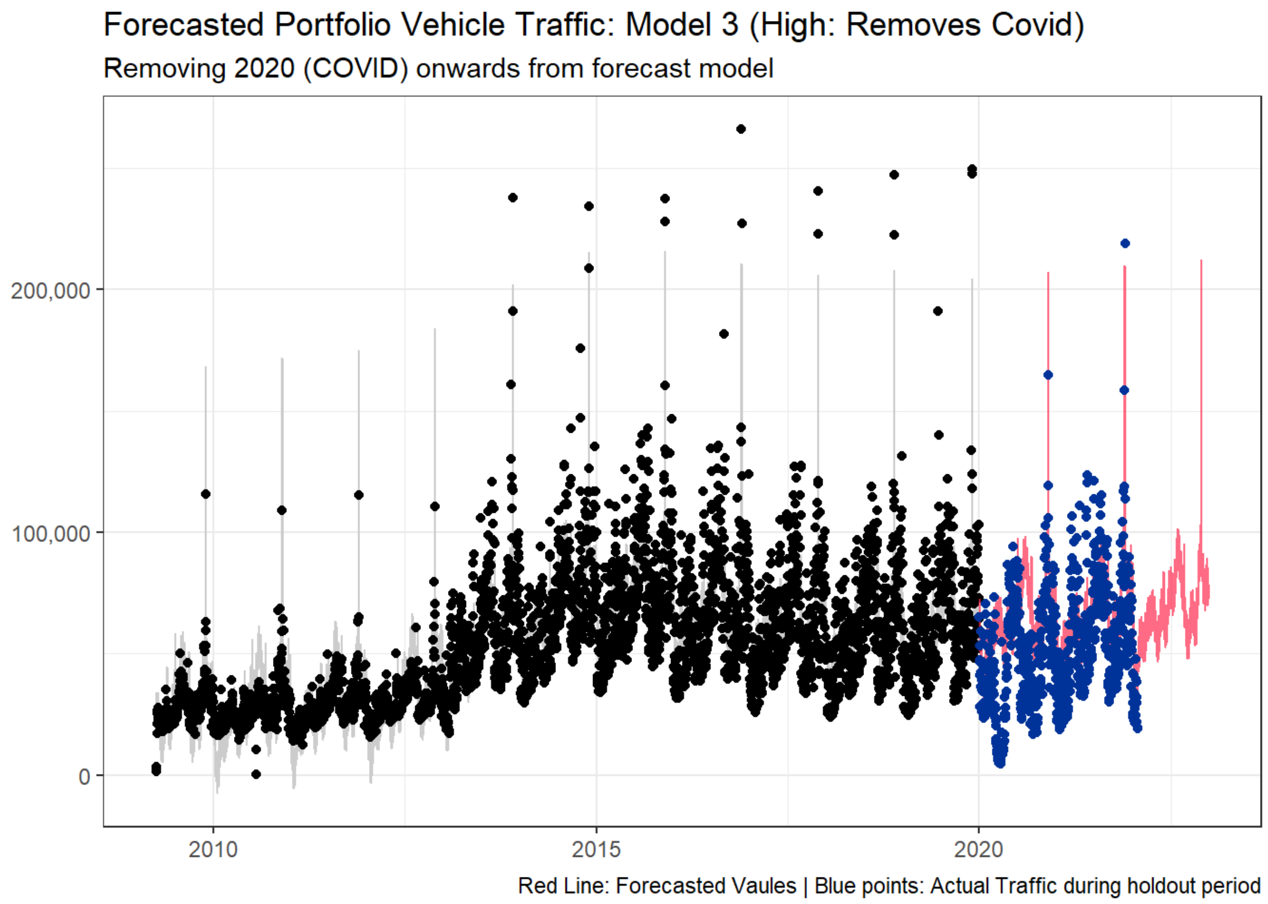
Forecast Model 1



Forecast Model 2



Forecast Model 3



Forecast Components

A forecast model can be broken up into the different elements that add (or multiply) together. By splitting them out and visualizing them, we then can view the index values and assess the impact of different items on the results

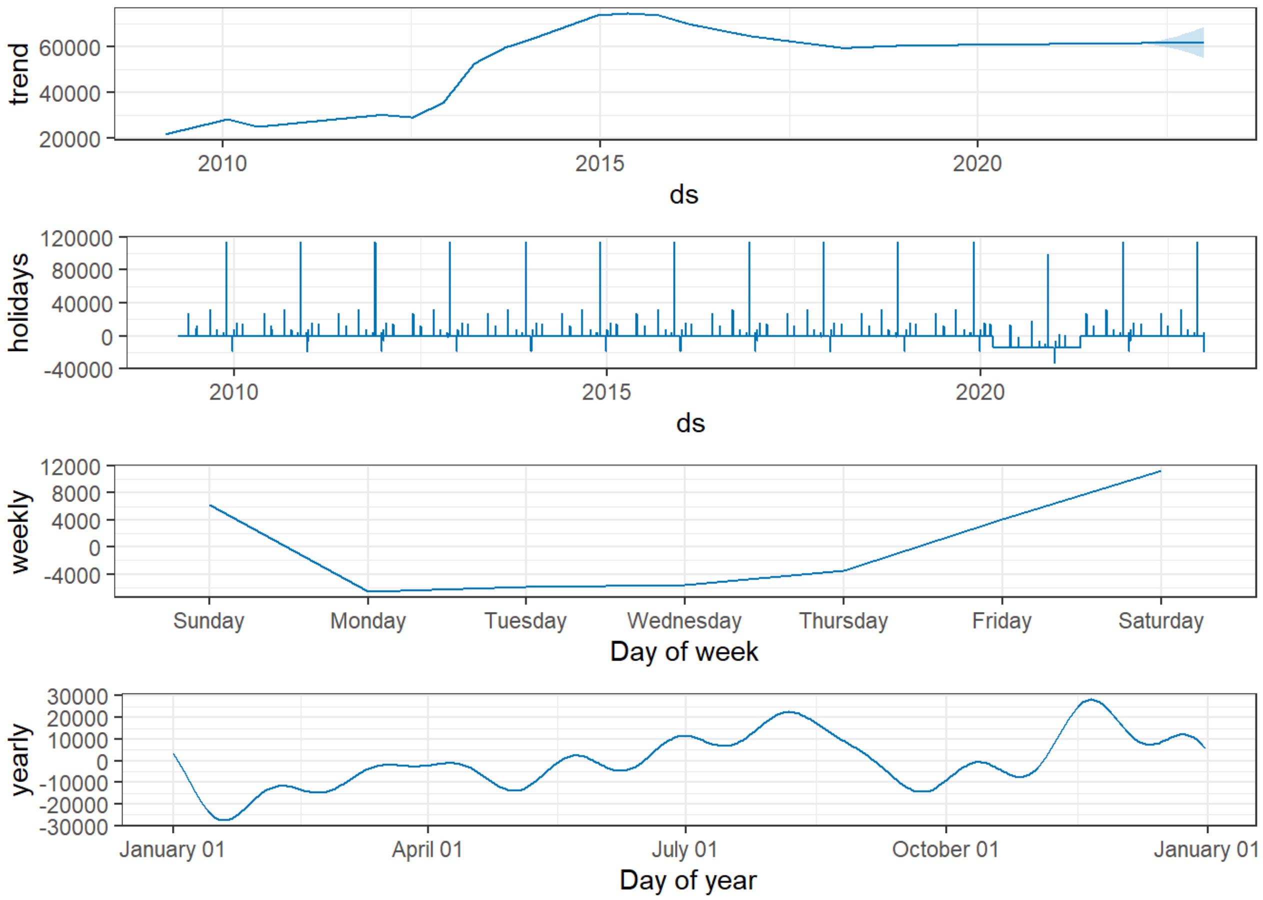
By breaking down the forecast into it's component parts we see a couple of things:

- Trend: This element shows what the general trend is year over year.
  - In this case, there is a forecasted to be a downwards trend for future years.
- Holidays: This elements flags both the US holidays as well as the date range from 3/1/2020 - 5/1/2021 as the main effect of Covid shutdowns.
- Weekly: This element shows how traffic varies within the week.
  - The outlets being a mainly weekend business really shows up in this view, with the positive indices being for Friday, Saturday, Sunday.
- Yearly: This is the element of the within year seasonality.
  - In general we see that there is a general ramp up throughout the first part of the year, a small lull in October, and then our biggest traffic volumes seen during the end of year.

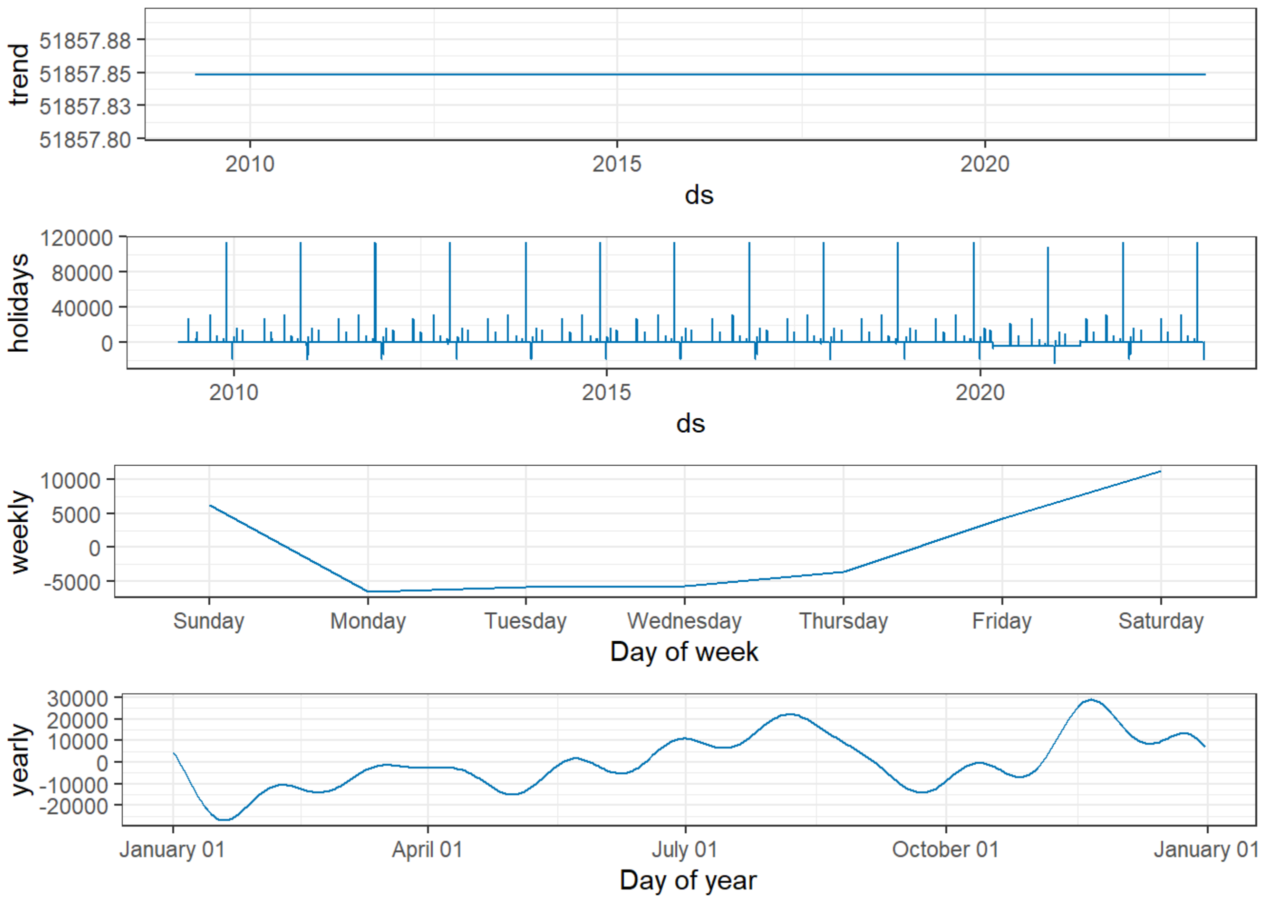
Note about negative values: These 4 items below are added together to create the forecast model. This is why there are some items that are negative and some that are positive. A negative value does not mean that we would forecast negative vehicles, but instead that at those instances the traffic would decrease from the trend.

- An example here would be the impact of Christmas day on digital traffic, where with the centers closed it's less likely people are logging into the website.
- We do see a yearly spike around Thanksgiving/Black Friday time period.

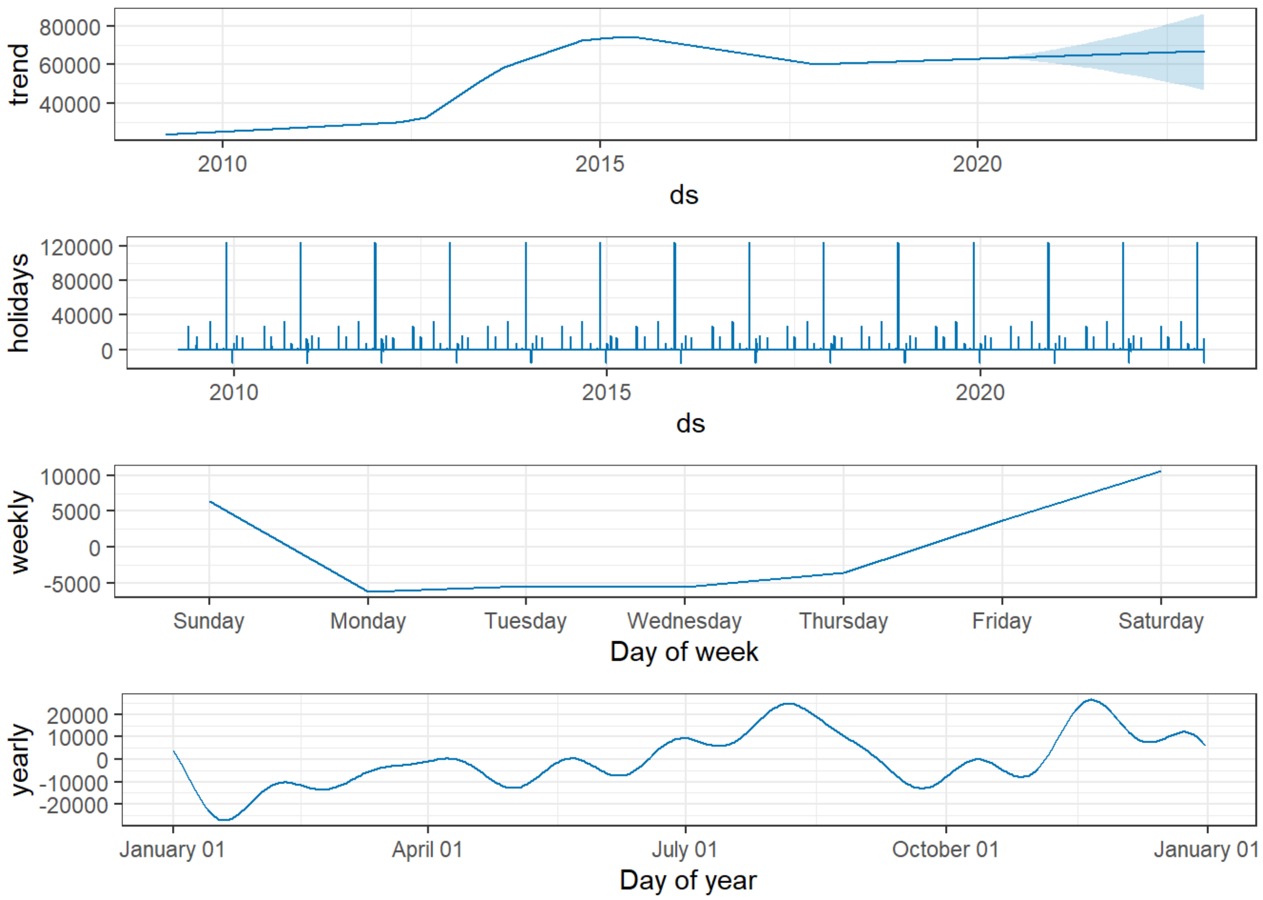
Model 1



Model 2


trend
51857.88
51857.85
51857.83
51857.80
2010
2015
2020
ds
holidays
120000
80000
40000
0
2010
2015
2020
ds
weekly
10000
5000
0
-5000
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day of week
yearly
30000
20000
10000
0
-10000
-20000
January 01
April 01
July 01
October 01
January 01
Day of year

Model 3


trend
80000
60000
40000
2010
2015
2020
ds
holidays
120000
80000
40000
0
2010
2015
2020
ds
weekly
10000
5000
0
-5000
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day of week
yearly
20000
10000
0
-10000
-20000
January 01
April 01
July 01
October 01
January 01
Day of year

Next Steps

The "Top Line" section presented both the actuals traffic by month for prior years with the forecasted values for 2022 (bolded).

The forecasted values could represent what would be considered as the baseline for goal setting by the business. Upon which the business leaders responsible for this metric would review and adjust the values according to where they believe they can made an impact. It will be up to the business to determine what strategies they will implement and adjust the goals accordingly.

This analysis body of work is set up to allow for the adjustment of goals based upon the values decided upon by the business owners.

Appendix

Export: Digital Traffic

Daily raw counts for Tangeroutlet.com since 4/2/2009.

Forecast Model

The forecast model code is available within the I&A OneDrive.

TangerOutlets