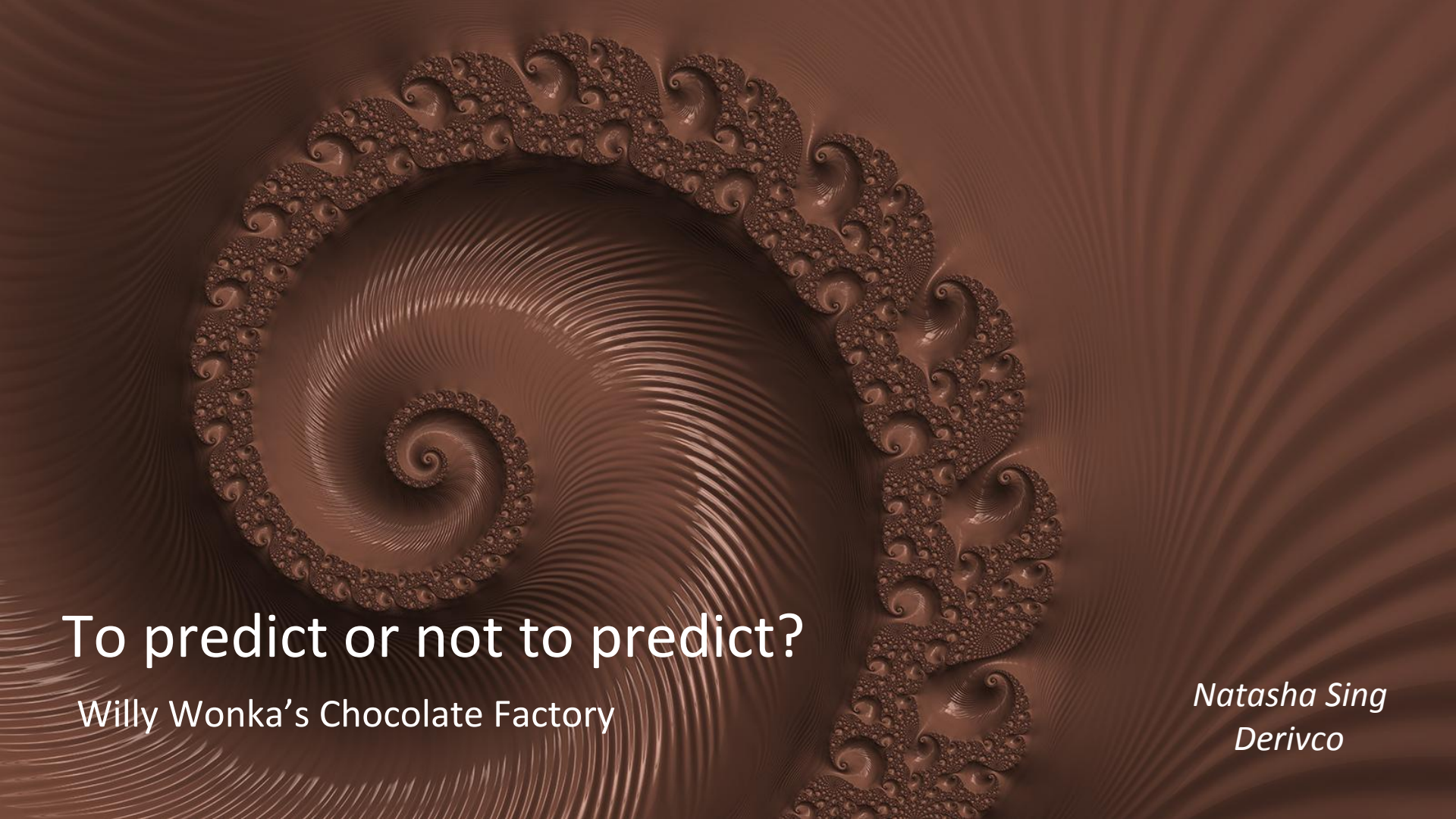


# To predict or not to predict?

## Willy Wonka's Chocolate Factory

*Natasha Sing*
*Derivco*

# WHAT WILLY WONKA'S CHOCOLATE FACTORY WOULD COST TODAY

Roald Dahl's iconic story *Charlie and the Chocolate Factory* has been retold and recreated for decades. It's about time to figure out what it would take to bring his world of pure imagination to life. After digging through tons of data, we calculated the cost of making Mr. Dahl's vision a reality. Here's what we found.

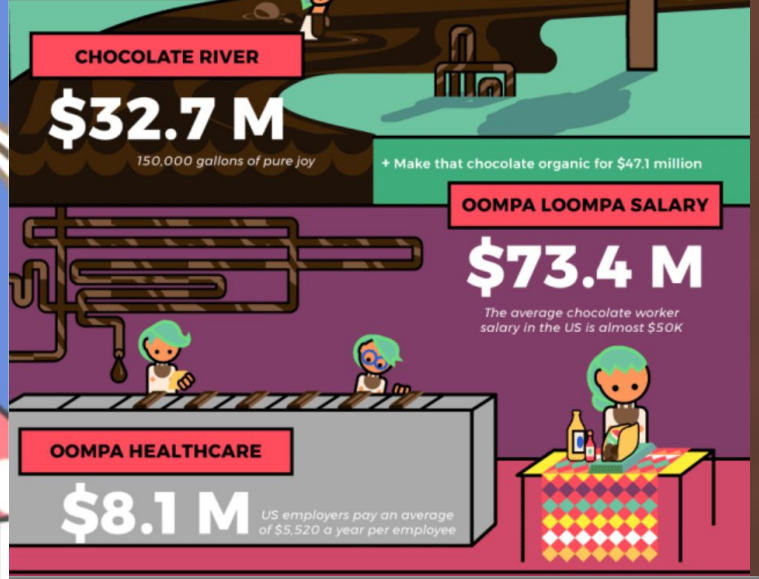
*All calculations below.*

## YEARLY ENERGY COSTS

**$2 M**

*Gets you one year with the lights on*

+ Going solar would require 3,221 solar panels, costing $2.37 billion



## CHOCOLATE RIVER

**$32.7 M**

*150,000 gallons of pure joy*

+ Make that chocolate organic for $47.1 million

## OOMPA LOOMPA SALARY

**$73.4 M**

*The average chocolate worker salary in the US is almost $50K*

## OOMPA HEALTHCARE

**$8.1 M**

*US employers pay an average of $5,520 a year per employee*

| | |
|---|---|
| GRAND TOTAL | $224,674,600 |
| ENERGY COSTS | $2,000,000 |
| CHOCOLATE RIVER | $32,600,000 |
| OOMPA LOOMPA SALARY | $73,000,000 |
| OOMPA LOOMPA HEALTHCARE | $8,000,000 |
| COST OF EDIBLE WALLPAPER | $2,600 |
| 12 GOLDEN GOOSE EGGS | $72,000 |
| FLYING GLASS ELEVATOR | $109,000,000 |

## GRAND TOTAL

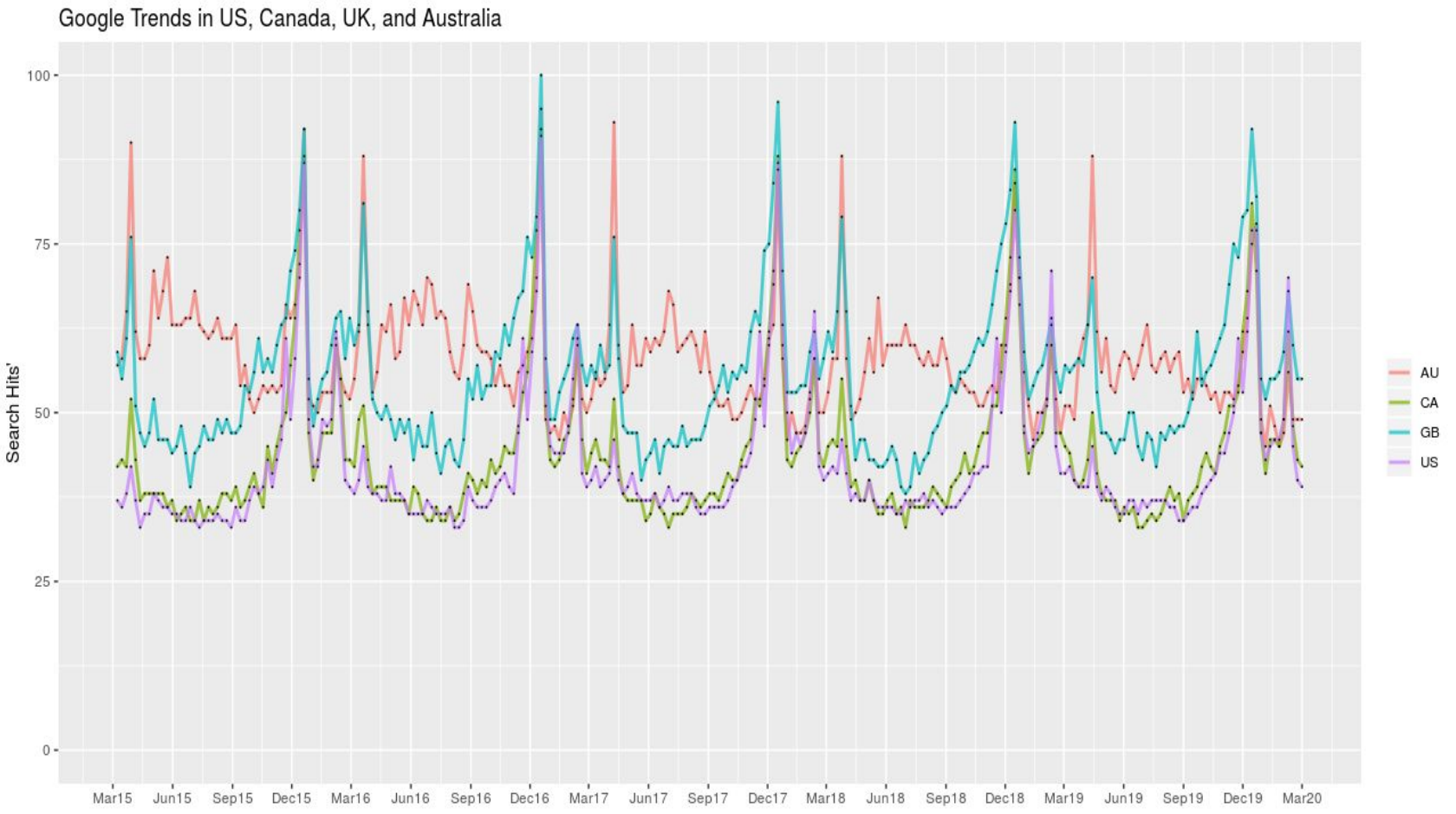
**$224.6 M**

# PROPHET

**Trend** | **Seasonality** | **Holidays**

Open Source
Accurate and fast
Analysts don’t require specialized skills in time series
Tunable parameters
Available for both Python and R

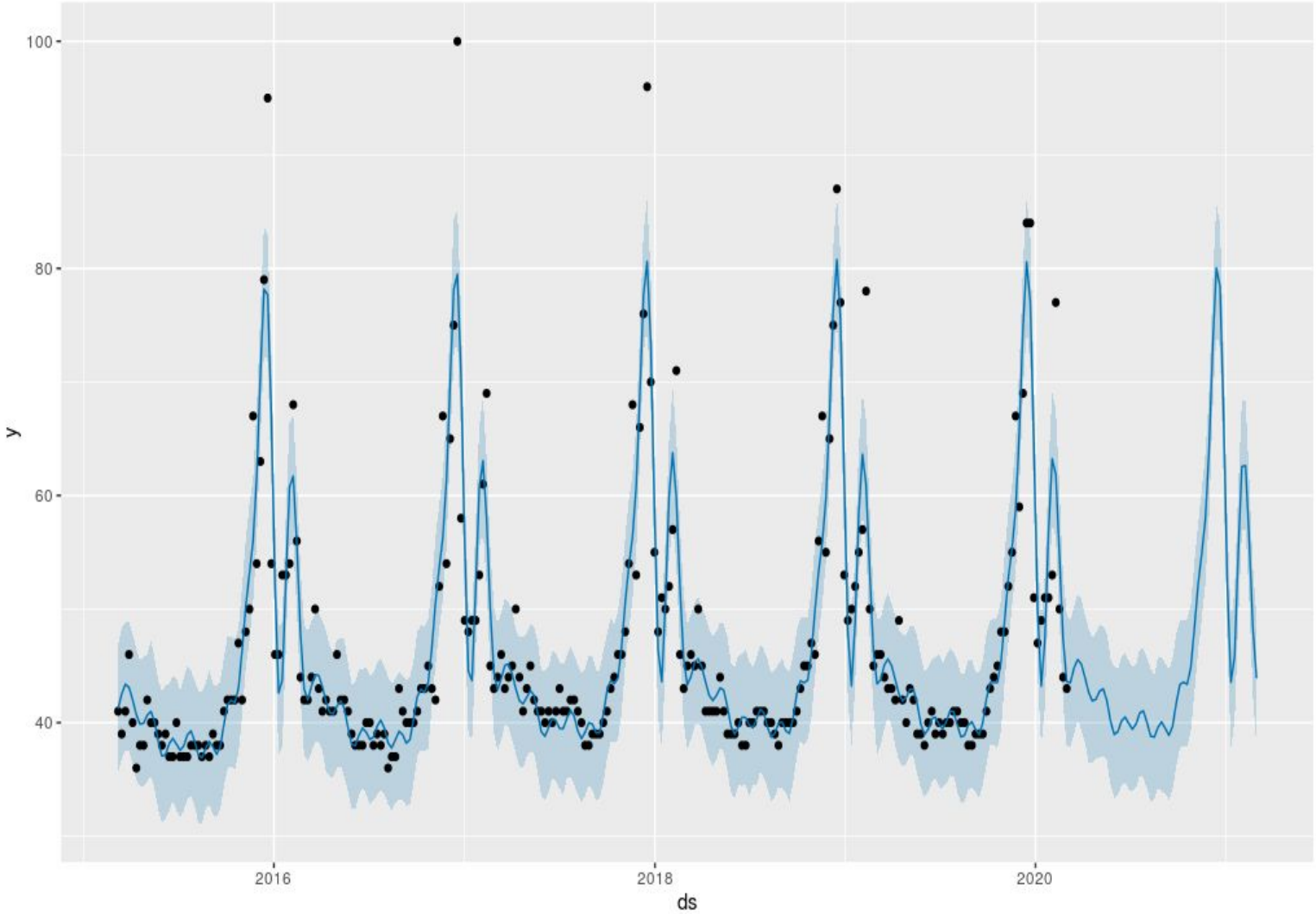
```
library(gtrendsR)
trends_chocolate = gtrends(c("chocolate"), geo = c("US"), gprop = "web", time = "today+5-y")[[1]]
```



# Prophet's Time Series Prediction Model

```
library(prophet)

#Wrangle the data to rename columns "date" and "hits"
forcast_data<-trends_chocolate %>% filter(geo=='US')%>%
  select(ds = date, y = hits)

#Build a Prophet time series model
choc_prophet<-prophet(forcast_data)

#What time periods are we going to predict?
choc_future<-make_future_dataframe(choc_prophet, periods = 52, freq='week')

#Predict future values using the prophet model
choc_forecast<-predict(choc_prophet, choc_future)

#Plot forecasted values
plot(choc_prophet, choc_forecast)
```

*


100
80
60
40
y
2016
2018
2020
ds

```r
holidays <- rbind(
  data.frame(
    holiday = "mothers_day",
    ds = as.Date(c(
      '2014-05-11',
      '2015-05-10',
      '2016-05-08',
      '2017-05-14',
      '2018-05-13',
      '2019-05-12',
      '2020-05-10'
    )),
    lower_window = -7,
    upper_window = +7
  ),
```

```r
  data.frame(
    holiday = "valentines_day",
    ds = as.Date(c(
      '2015-02-08',
      '2016-02-07',
      '2017-02-12',
      '2018-02-11',
      '2019-02-10',
      '2020-02-09'
    )),
    lower_window = -7,
    upper_window = +7 ),

  data.frame(
    holiday = "christmas_period",
    ds = as.Date(c(
      '2015-12-20',
      '2016-12-25',
      '2017-12-24',
      '2018-12-23',
      '2019-12-22',
      '2020-12-20'
    )),
```

# Prophet's Prediction Model With "Holidays"

```
#Build a model for "holiday" effect
choc_prophet_holiday<-prophet(holidays=holidays)
```

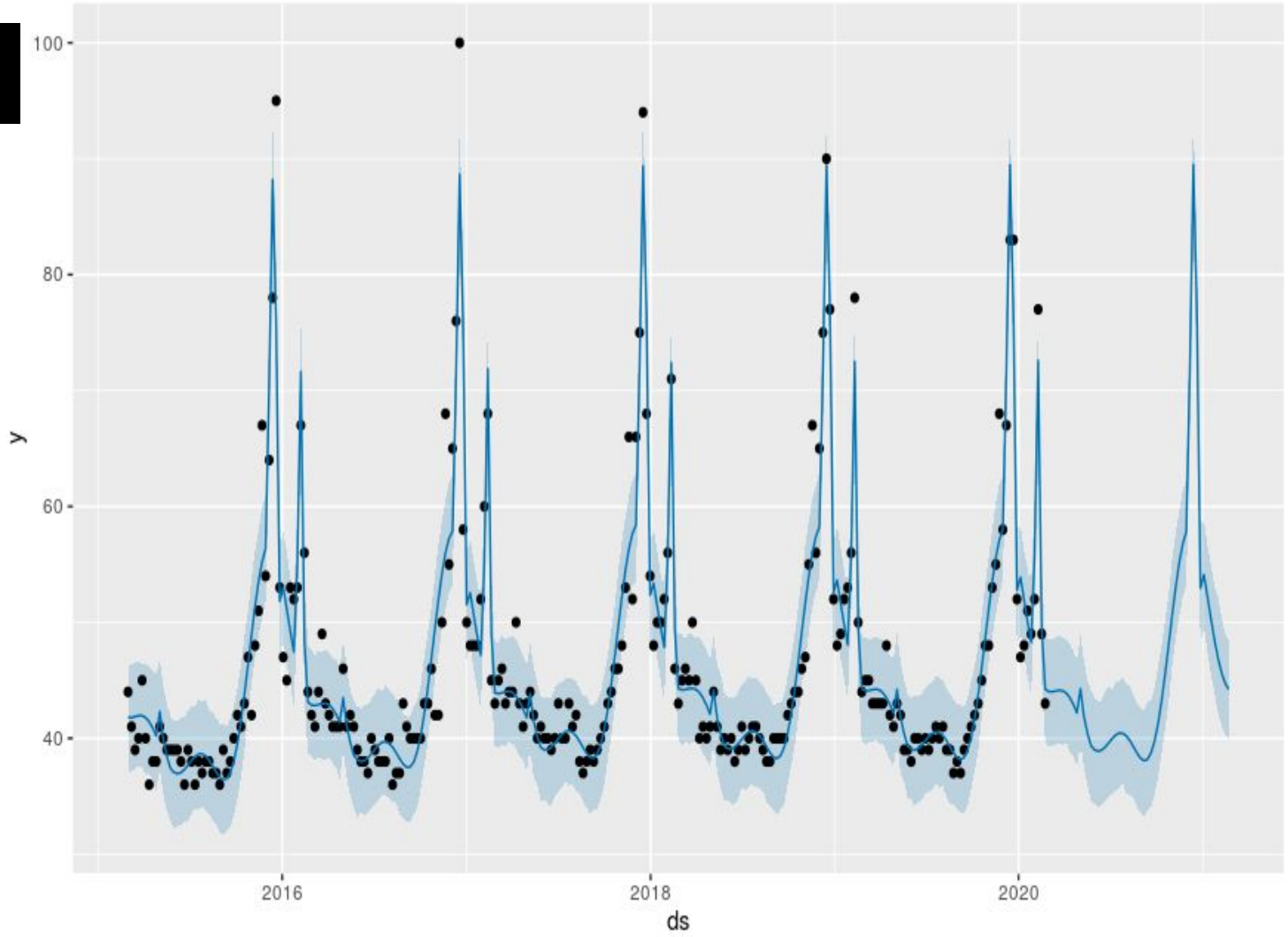
```
#Fit the model to the data
choc_prophet_fit<-fit.prophet(choc_prophet_holiday,forcast_data)
```

```
#What time periods are we going to predict?
choc_future_holiday<-make_future_dataframe(choc_prophet_fit, periods = 52, freq='week')
```

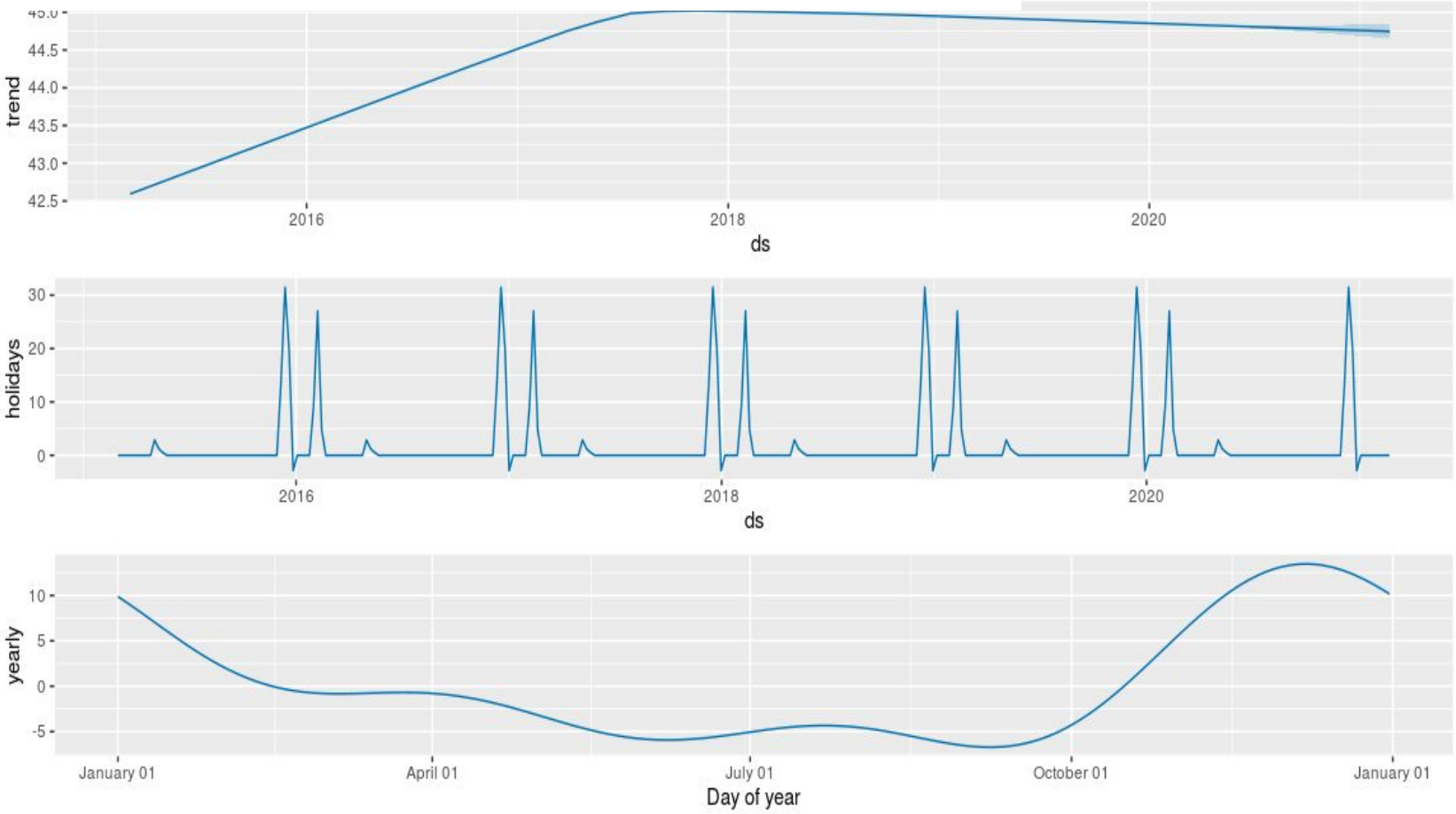
```
#Predict future values using the prophet model
choc_forecast_holiday<-predict(choc_prophet_fit,choc_future_holiday)
```

```
#Plot forecasted values
plot(choc_prophet_fit, choc_forecast_holiday)
```

Holiday Effect



```r
# Observe the Prophet's model components
prophet_plot_components(choc_prophet_fit, choc_forecast_holiday)
```




YES!!!

# Thank you!

Natasha Sing
natasha.sing@derivco.co.za

@Natasha10814

Natasha Sing

Natasha108