Projecting Natural Phenomenon Through Hotel Economics

Julian Christov

2/29/2024

Botanists often use bud development, weather forecasts, and historical data to predict when the cherry blossoms bloom. I decided to take an entirely different approach, using hotel prices as a proxy for my prediction. My goal is to use unconscious thought and instinctual behavior from tourists' understanding on when the cherry blossoms will bloom, along with market research and analysis from hotel management, to predict an exact day the cherry blossoms will bloom. Tourists will look at past precedent, but then make their travel decisions based on gut feeling. On the other hand, hotels are looking to maximize their profits by capitalizing on the tens of thousands of tourists that visit to see the cherry blossoms and list their hotel prices accordingly. An increase in tourist demand and anticipated future demand might raise hotel prices. An example of this is Ocean City Maryland, where hotel demand increases in the summer, with prices at 200%-300% over the rest of the year. Annually in September Ocean City hosts a music festival with tens of thousands of attendees, and the room prices are listed at 300% above normal, based on assumptions for when the event will take place. I will look at hotel rates in Kyoto, Washington D.C., Liesal, Vancouver, and New York City from March 15 to May 9 to determine when the cherry blossoms will bloom, since these are the earliest and latest dates the cherry blossoms have bloomed in the last hundred years at these locations.

Some of the most popular and well-regarded hotels for tourists:

1) Kyoto (Japan)

a. Four Seasons
b. Hotel Okura Kyoto Okazaki Bettei

2) Washington D.C. (USA)

a. Capital Hilton
b. State Plaza Hotel

3) Liestal (Switzerland)

a. Hotel Engel
b. Bad Schauenburg

4) Vancouver (Canada)

a. Pan Pacific
b. Fairmount

5) New York City, NY (USA)

a. The Plaza
b. Park Hyatt

All the hotel rates came from "Hotels.com" and "Booking.com". Prices are likely to increase during peak tourist seasons, such as during cherry blossom season. For Kyoto there is one major spike where prices for hotels more than tripled. This spike occurred on March 29. For Washington D.C. there are four spikes price. The first spike is on March 19, the second is on April 3, the third is on April 15, and the last spike in price is on May 5. On March 24 and April 10 there was a slight bump in Liestal's hotel prices. In Vancouver the biggest bump in price was on April 3. In NYC there were also numerous spikes in price. They were on April 6, April 13, April 20, and May 7. This may be due to NYC having a wide array of events and activities throughout Spring.

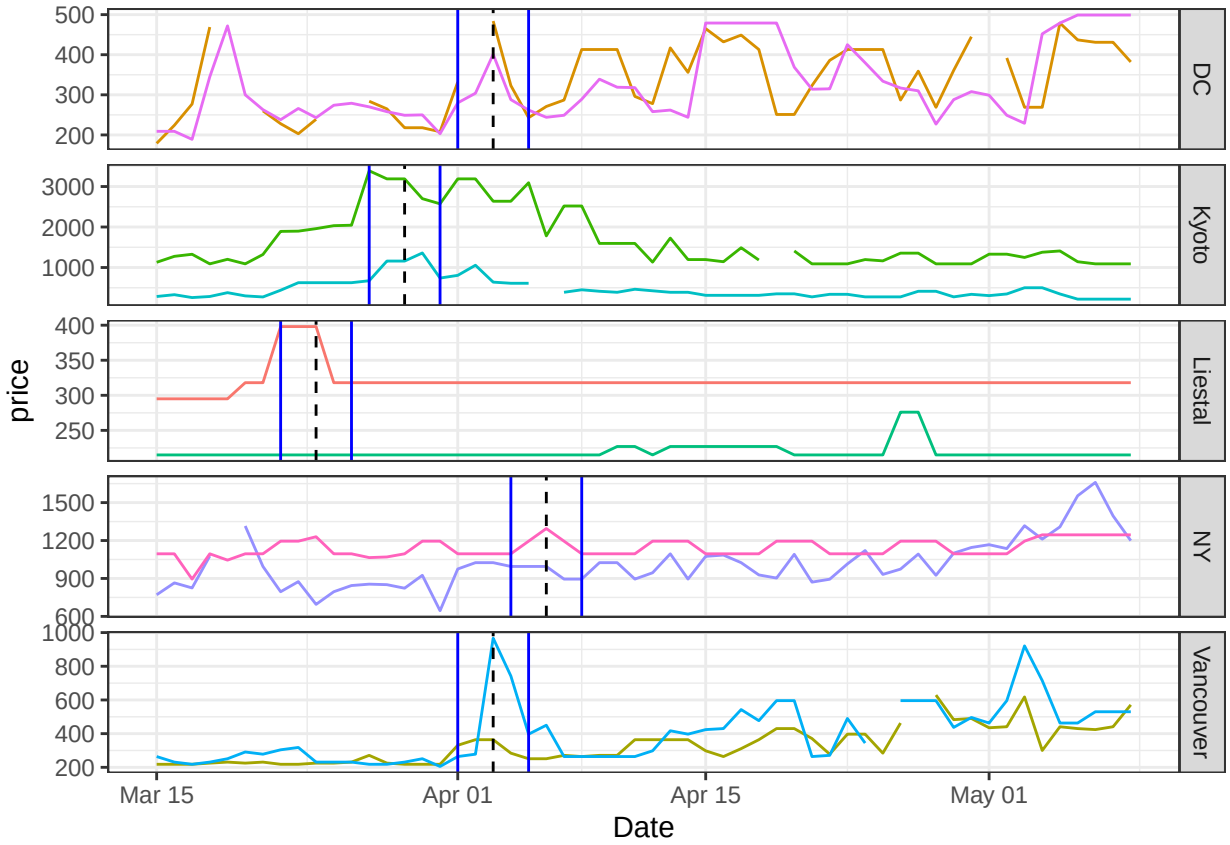
```
library(readxl)
hotels <- read_excel("~/My Kindle Content/Desktop/hotels.xlsx")

library("tidyverse")
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts
```

```
hotels %>%
  mutate(across(2:last_col(), ~ as.numeric(.))) %>%
  pivot_longer(-Date, names_to = "hotel", values_to = "price") %>%
```

```r
  mutate(location = str_split_i(hotel, ", ", 2),
         Date = as.Date(Date)) %>%
  ggplot() +
  theme_bw() +
  aes(Date, price, color = hotel) +
  geom_line() +
  facet_grid(location ~ ., scales = "free") +
  theme(legend.position = "none") +
  geom_vline(aes(xintercept = Date), linetype = 2,
             data = tibble(location = c("DC", "Kyoto", "Liestal", "NY", "Vancouver"),
                           Date = as.Date(c("2024-04-03", "2024-03-29", "2024-03-24", "2
  geom_vline(aes(xintercept = Date), color = "blue",
             data = tibble(location = c("DC", "Kyoto", "Liestal", "NY", "Vancouver"),
                           Date = as.Date(c("2024-04-01", "2024-03-27", "2024-03-22", "2
  geom_vline(aes(xintercept = Date), color = "blue",
             data = tibble(location = c("DC", "Kyoto", "Liestal", "NY", "Vancouver"),
                           Date = as.Date(c("2024-04-05", "2024-03-31", "2024-03-26", "2
```

```
## Warning: There were 6 warnings in `mutate()`.
## The first warning was:
## i In argument: `across(2:last_col(), ~as.numeric(.))`.
## Caused by warning:
## ! NAs introduced by coercion
## i Run `dplyr::last_dplyr_warnings()` to see the 5 remaining warnings.
```



Based on these spikes in hotel prices as well as my personal bias I predict the cherry blossom bloom dates in 2024 will be as follows:

1) Kyoto (Japan): March 29

a) DOY: 89
b) Interval: [87,91]

2) Washington D.C. (USA): April 3

a) DOY: 94
b) Interval: [92,96]

3) Liestal (Switzerland): March 24

a) DOY: 84
b) Interval: [82,86]

4) Vancouver (Canada): April 3

a) DOY: 94
b) Interval: [92,96]

5) New York City, NY (USA): April 6

a) DOY: 97
b) Interval: [95,99]

There are a few potential issues when using hotel prices to predict when the cherry blossoms will bloom. The first issue is there are other seasonal events such as conferences, festivals, or concerts that can drive hotel prices up. The second issue is assuming hotels and tourists have a better understanding of when the cherry blossoms will bloom than anyone else. However, it is possible that hotel prices are the most statistically significant factor in predicting the cherry blossom bloom date and people have not even considered it. As a quote from another famous historical Washington D.C. phenomenon, the Watergate Scandel, "follow the money."