

How to predict the future numbers of restaurant visitors?

Yonghua Su

Institute of Statistics and Big Data
Renmin University of China

03.17 2021

Introduction

The aim is to predict the future numbers of restaurant visitors. This makes it a Time Series Forecasting problem. The data comes in the shape of 8 relational files which are derived from two separate Japanese websites that collect user information: “Hot Pepper Gourmet (hpg): similar to Yelp” (search and reserve) and “AirREGI / Restaurant Board (air): similar to Square” (reservation control and cash register). The training data is based on the time range of Jan 2016 - most of Apr 2017, while the test set includes the last week of Apr plus May 2017.

Overview: File structure and content

```
> air_visits
# A tibble: 252,108 × 5
  air_store_id     visit_date visitors wday month
  <chr>           <date>      <dbl>   <ord> <ord>
1 air_ba937bf13d40fb24 2016-01-13     25   三    " 1"
2 air_ba937bf13d40fb24 2016-01-14     32   四    " 1"
3 air_ba937bf13d40fb24 2016-01-15     29   五    " 1"
4 air_ba937bf13d40fb24 2016-01-16     22   六    " 1"
5 air_ba937bf13d40fb24 2016-01-18      6   —    " 1"
6 air_ba937bf13d40fb24 2016-01-19      9   二    " 1"
7 air_ba937bf13d40fb24 2016-01-20     31   三    " 1"
8 air_ba937bf13d40fb24 2016-01-21     21   四    " 1"
9 air_ba937bf13d40fb24 2016-01-22     18   五    " 1"
10 air_ba937bf13d40fb24 2016-01-23     26   六   " 1"
# ... with 252,098 more rows
```

Overview: File structure and content

```
> air_reserve
# A tibble: 92,378 × 12
  air_store_id visit_datetime     reserve_datetime
  <chr>          <dttm>           <dttm>
1 air_877f797... 2016-01-01 19:00:00 2016-01-01 16:00:00
2 air_db4b38e... 2016-01-01 19:00:00 2016-01-01 19:00:00
3 air_db4b38e... 2016-01-01 19:00:00 2016-01-01 19:00:00
4 air_877f797... 2016-01-01 20:00:00 2016-01-01 16:00:00
5 air_db80363... 2016-01-01 20:00:00 2016-01-01 01:00:00
6 air_db80363... 2016-01-02 01:00:00 2016-01-01 16:00:00
7 air_db80363... 2016-01-02 01:00:00 2016-01-01 15:00:00
8 air_3bb99a1... 2016-01-02 16:00:00 2016-01-02 14:00:00
9 air_3bb99a1... 2016-01-02 16:00:00 2016-01-01 20:00:00
10 air_2b8b29d... 2016-01-02 17:00:00 2016-01-02 17:00:00
# ... with 92,368 more rows, and 9 more variables:
```

Overview: File structure and content

```
> air_store
```

```
# A tibble: 829 x 9
```

	air_store_id	air_genre_name	air_area_name	prefecture	latitude
	<chr>	<fct>	<fct>	<chr>	<dbl>
1	air_0f0cdee...	Italian/French	Hyōgo-ken	Kō... Hyōgo-ken	34.7
2	air_7cc17a3...	Italian/French	Hyōgo-ken	Kō... Hyōgo-ken	34.7
3	air_fee8dcf...	Italian/French	Hyōgo-ken	Kō... Hyōgo-ken	34.7
4	air_a17f077...	Italian/French	Hyōgo-ken	Kō... Hyōgo-ken	34.7
5	air_83db5af...	Italian/French	Tōkyō-to	Min... Tōkyō-to	35.7
6	air_99c3eae...	Italian/French	Tōkyō-to	Min... Tōkyō-to	35.7
7	air_f183a51...	Italian/French	Tōkyō-to	Min... Tōkyō-to	35.7
8	air_6b9fa44...	Italian/French	Tōkyō-to	Min... Tōkyō-to	35.7
9	air_0919d54...	Italian/French	Tōkyō-to	Min... Tōkyō-to	35.7
10	air_2c6c79d...	Italian/French	Tōkyō-to	Min... Tōkyō-to	35.7
# ... with 819 more rows, and 4 more variables: longitude <dbl>,					

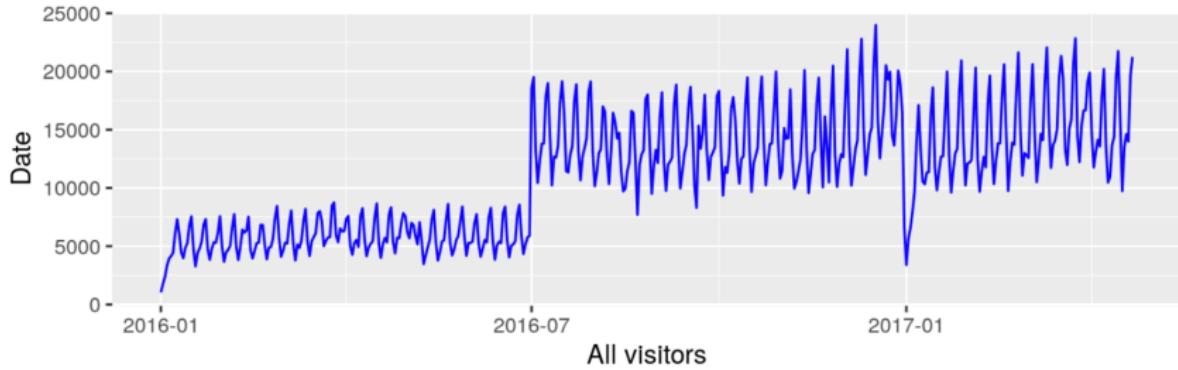
Overview: File structure and content

```
> air_store
```

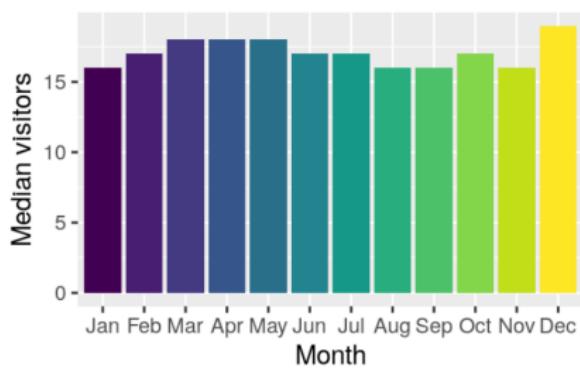
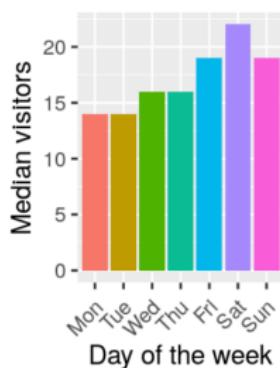
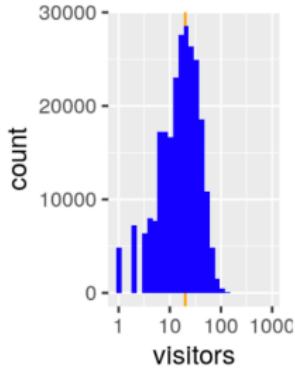
```
# A tibble: 829 x 9
```

	air_store_id	air_genre_name	air_area_name	prefecture	latitude
	<chr>	<fct>	<fct>	<chr>	<dbl>
1	air_0f0cdee...	Italian/French	Hyōgo-ken	Kō... Hyōgo-ken	34.7
2	air_7cc17a3...	Italian/French	Hyōgo-ken	Kō... Hyōgo-ken	34.7
3	air_fee8dcf...	Italian/French	Hyōgo-ken	Kō... Hyōgo-ken	34.7
4	air_a17f077...	Italian/French	Hyōgo-ken	Kō... Hyōgo-ken	34.7
5	air_83db5af...	Italian/French	Tōkyō-to	Min... Tōkyō-to	35.7
6	air_99c3eae...	Italian/French	Tōkyō-to	Min... Tōkyō-to	35.7
7	air_f183a51...	Italian/French	Tōkyō-to	Min... Tōkyō-to	35.7
8	air_6b9fa44...	Italian/French	Tōkyō-to	Min... Tōkyō-to	35.7
9	air_0919d54...	Italian/French	Tōkyō-to	Min... Tōkyō-to	35.7
10	air_2c6c79d...	Italian/French	Tōkyō-to	Min... Tōkyō-to	35.7
# ... with 819 more rows, and 4 more variables: longitude <dbl>,					

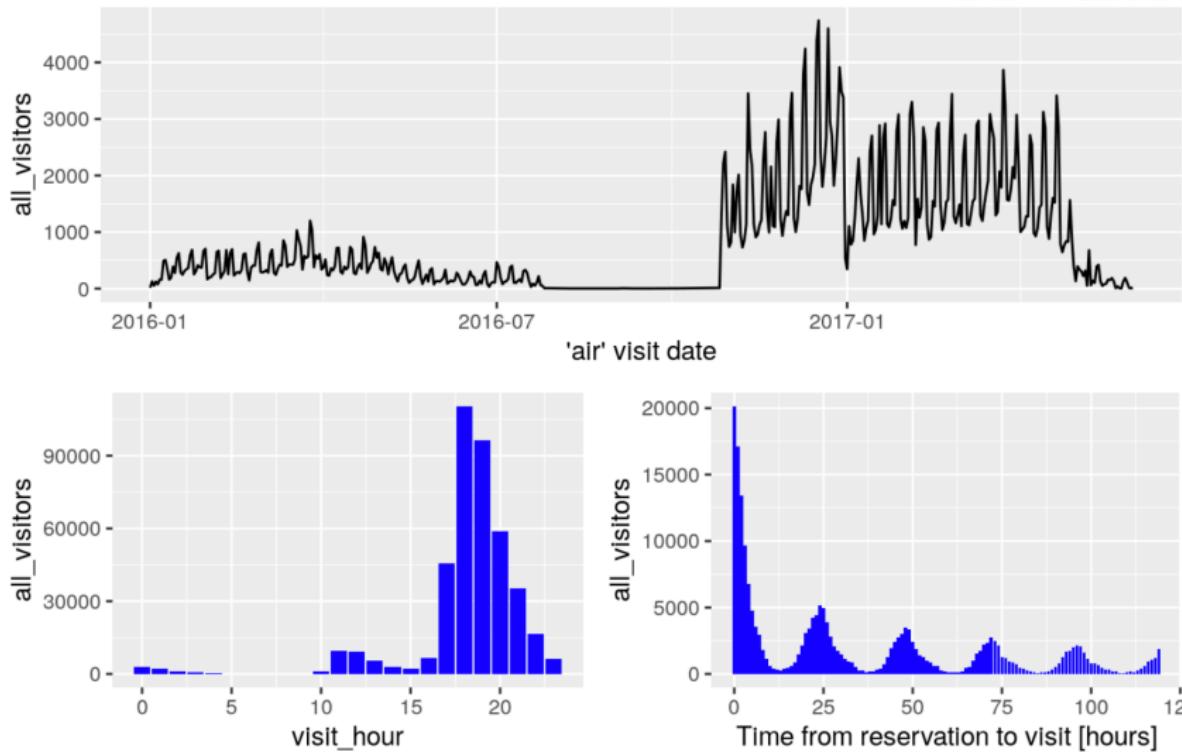
Air Visits



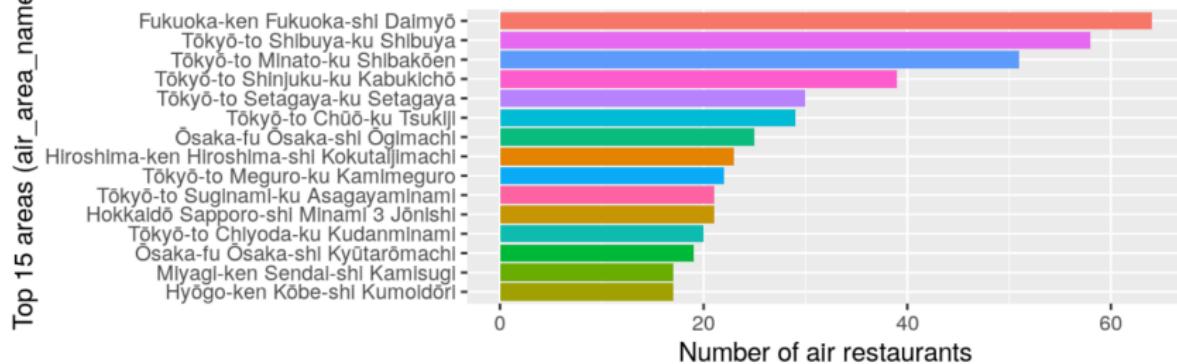
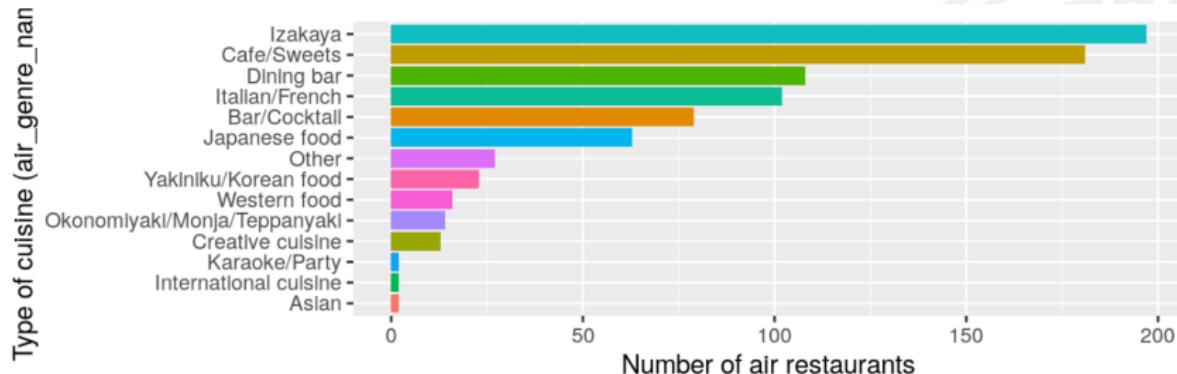
All visitors



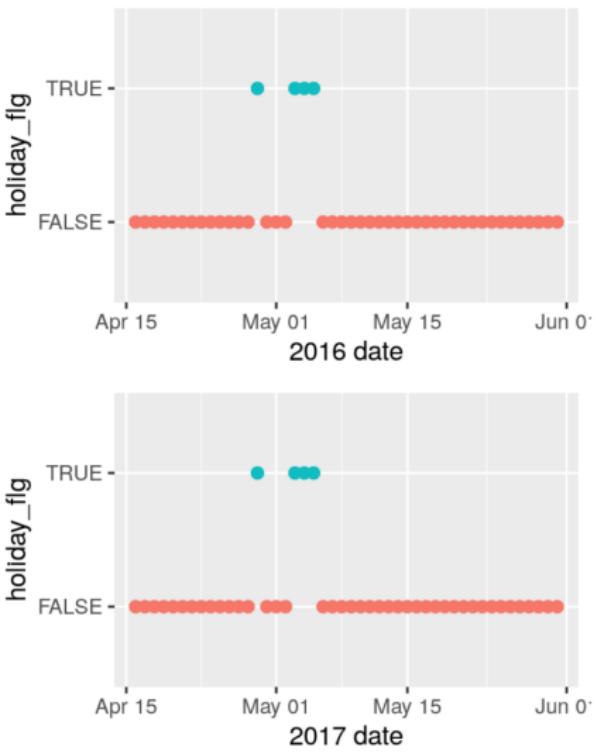
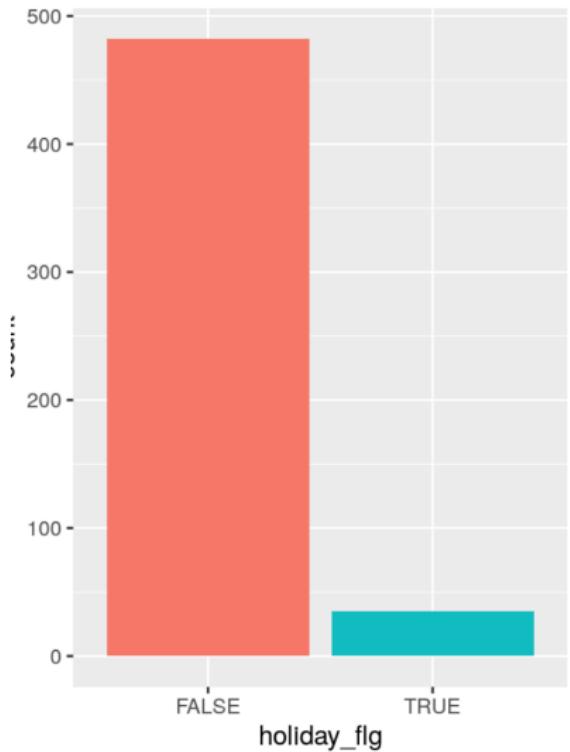
Air Reservations



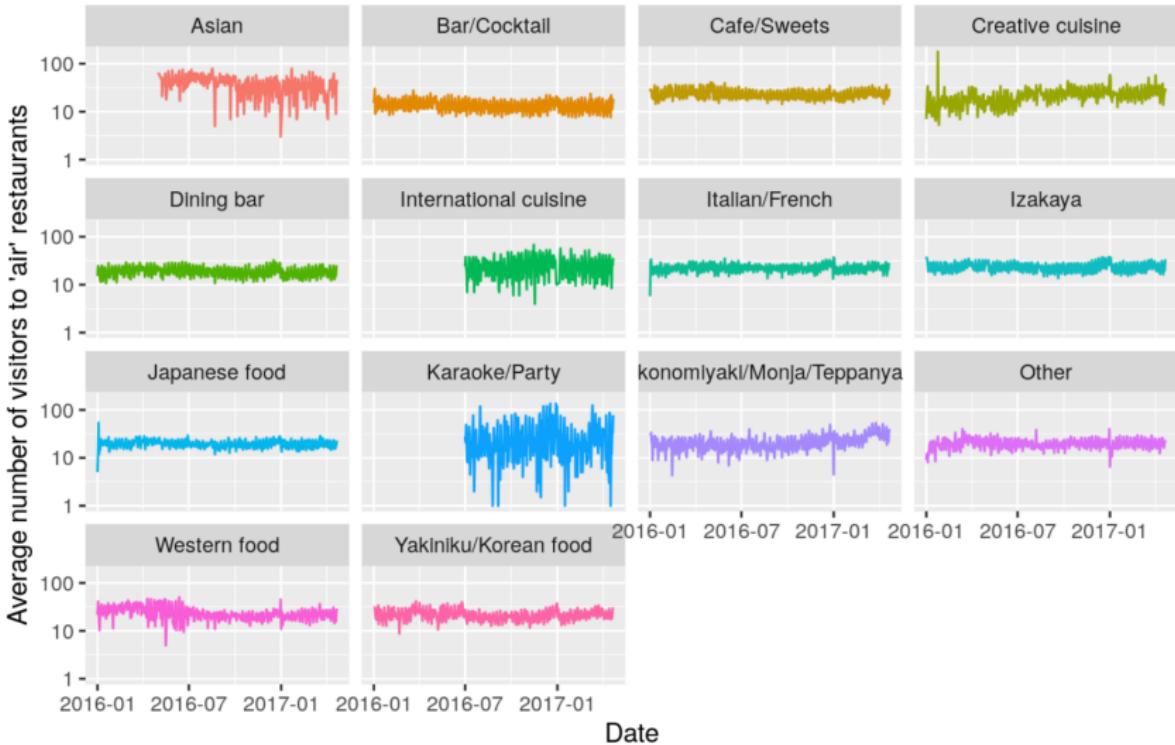
Air Store



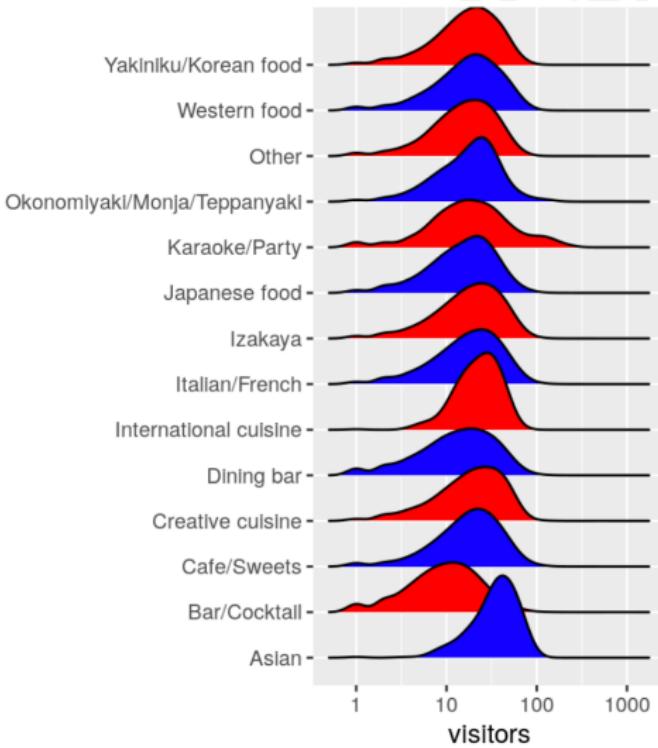
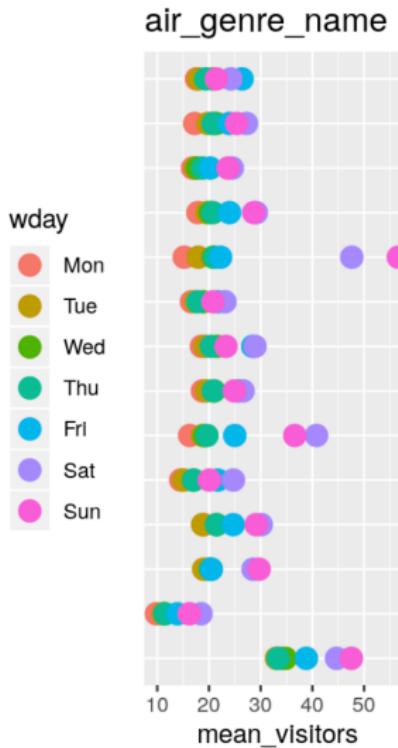
Holidays



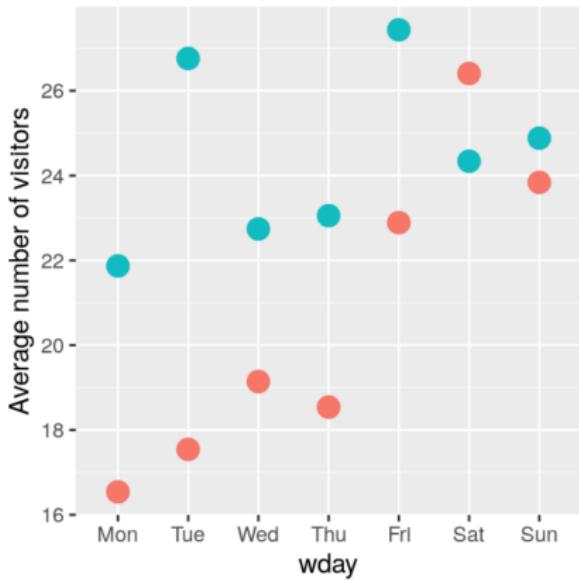
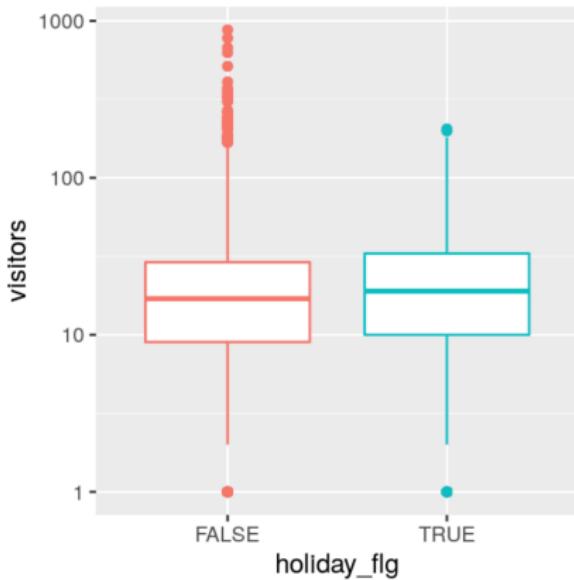
Visitors per genre



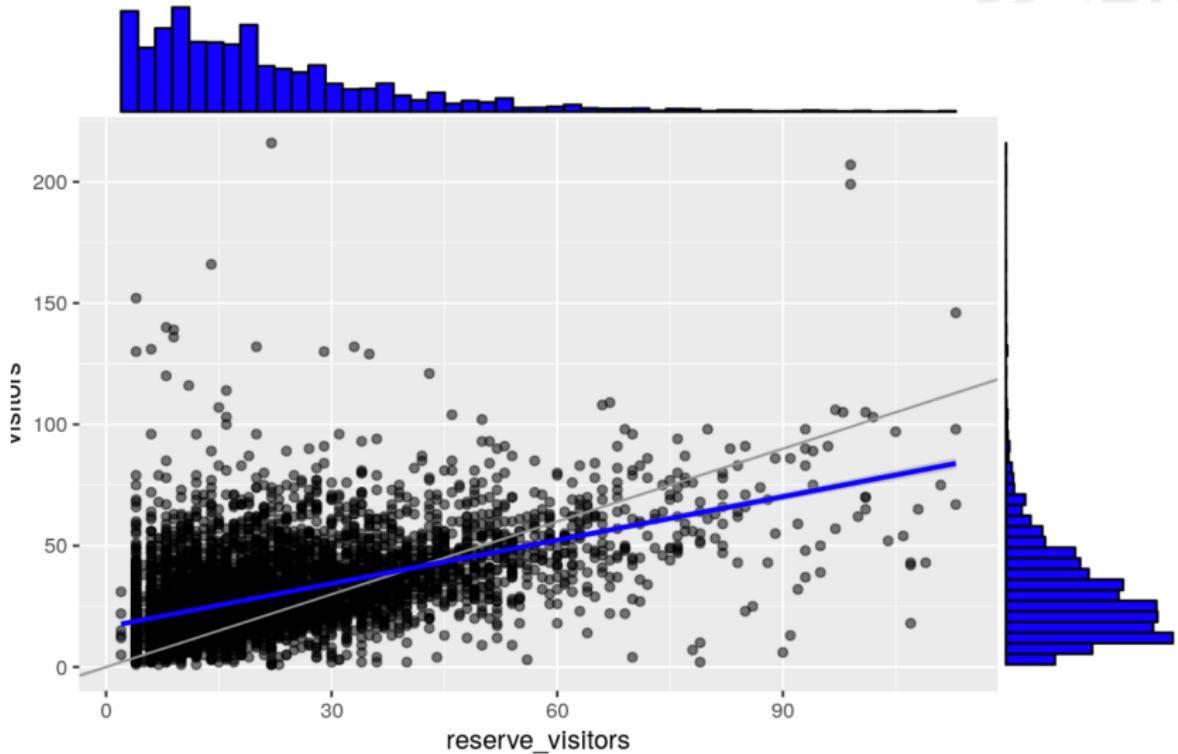
Visitors per genre



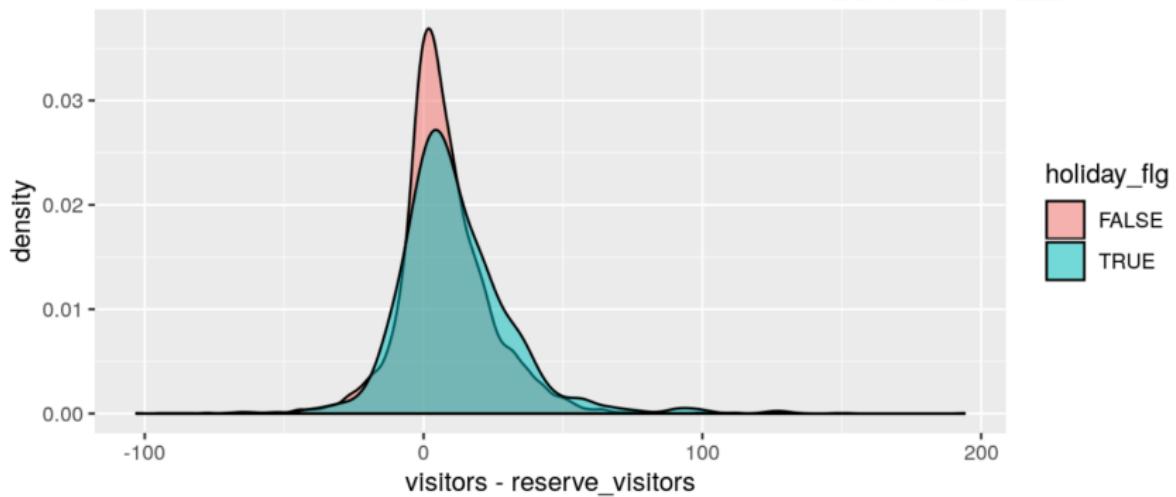
The impact of holidays



Reservations vs Visits



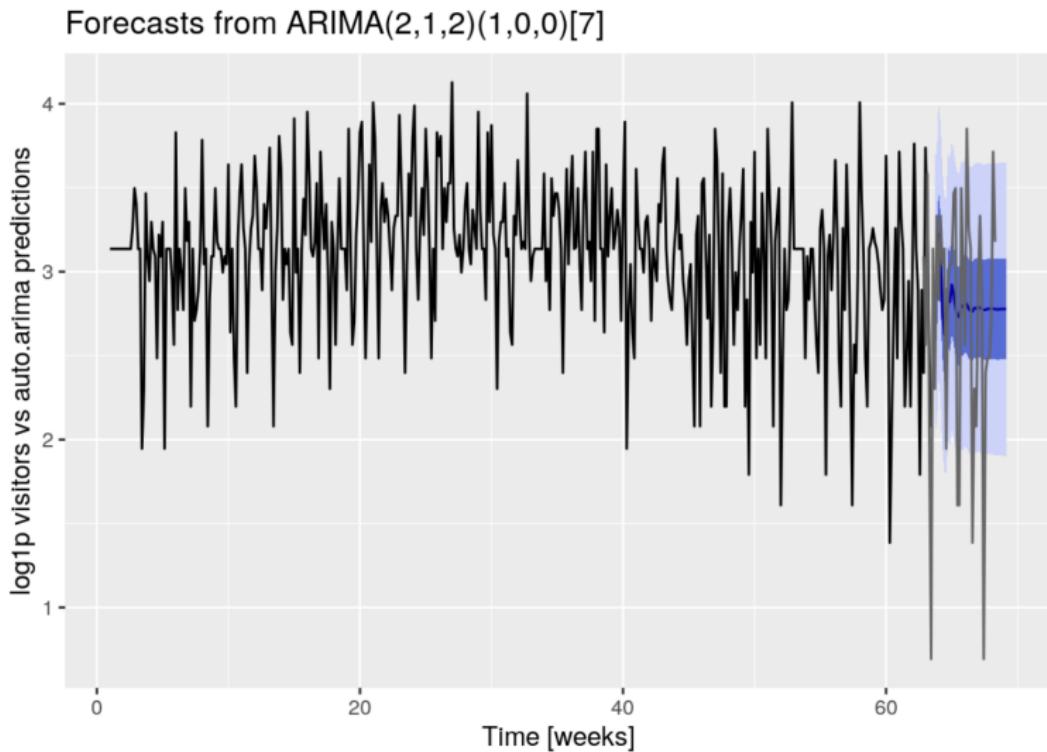
Reservations vs Visits



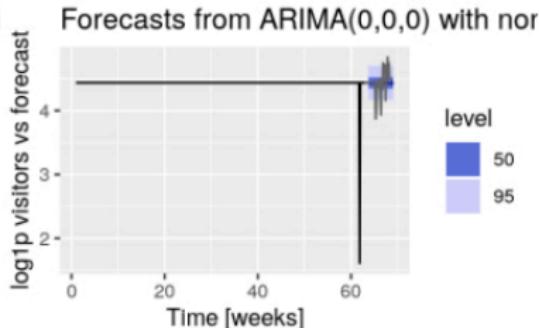
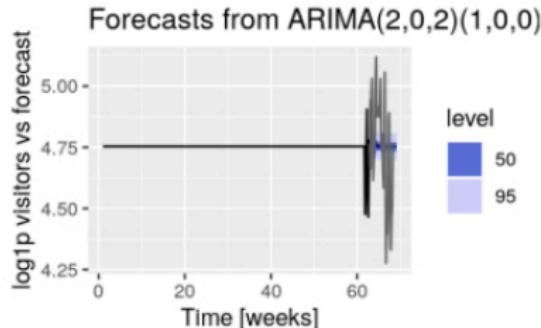
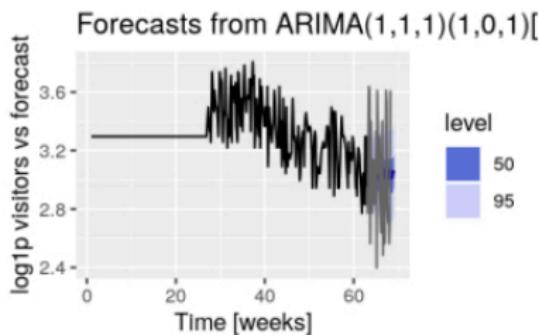
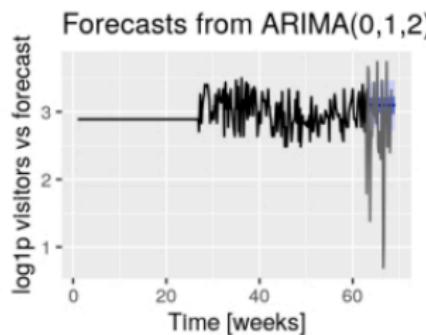
$$\left(1 - \sum_{i=1}^p \varphi_i L^i\right) (1 - L)^d X_t = \delta + \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t.$$

This defines an ARIMA(p, d, q) process with drift $\frac{\delta}{1 - \sum \varphi_i}$.

ARIMA / auto.arima



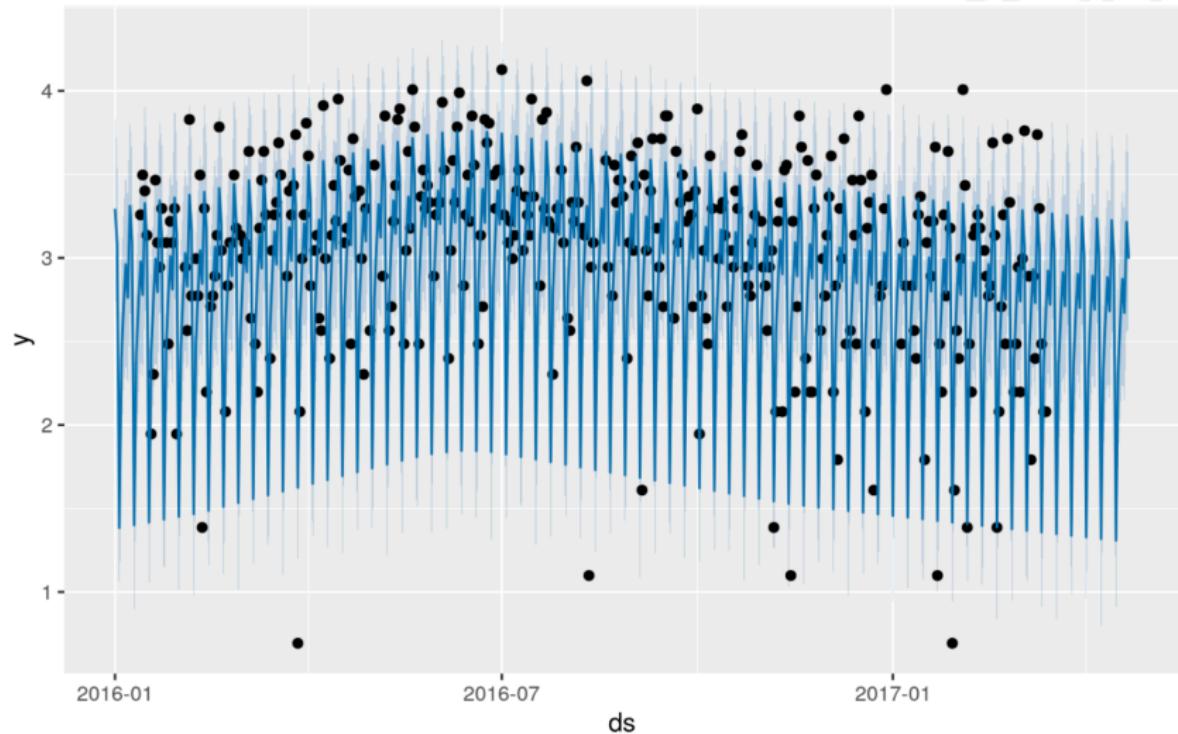
ARIMA / auto.arima



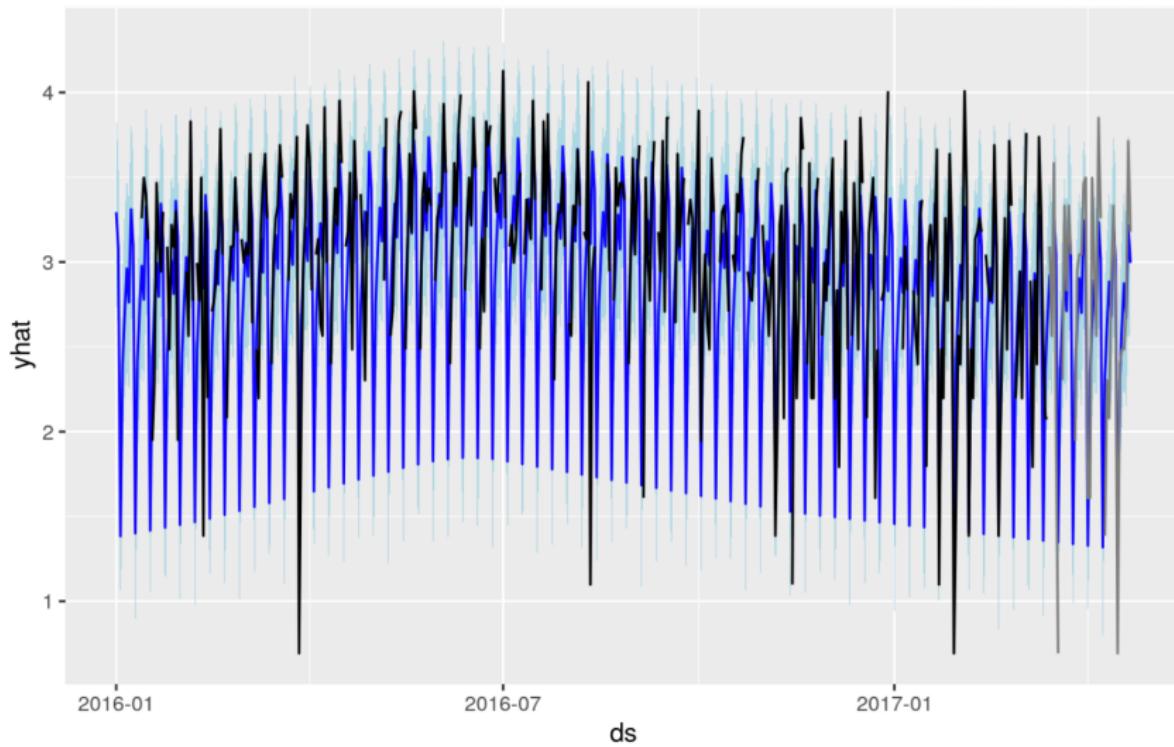
$$X(t) = g(t) + s(t) + h(t) + \epsilon_t$$

- $g(t)$: overall growth trend function, which models non-periodic changes in the value of the time series.
- $s(t)$: represents periodic changes, Seasonality - Yearly, Weekly, Daily.
- $h(t)$: represents the effects of holidays which occur on potentially irregular schedules over one or more days, X'mas, New Year.
- ϵ_t : represents any idiosyncratic changes which are not accommodated by the model.

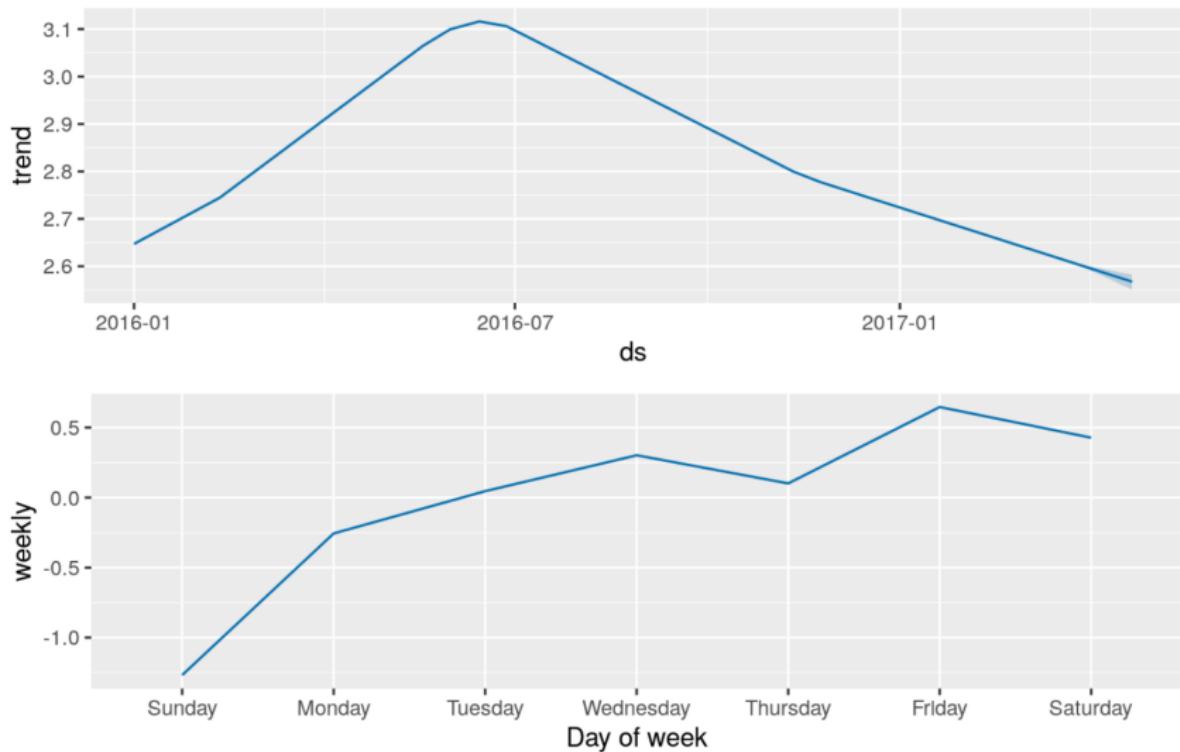
Prophet



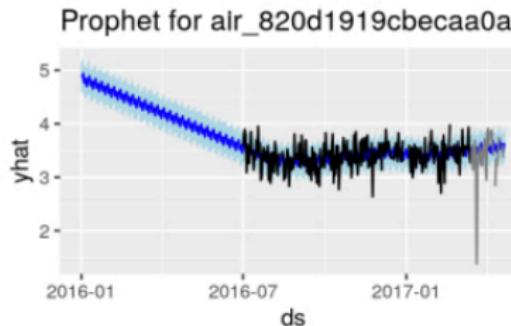
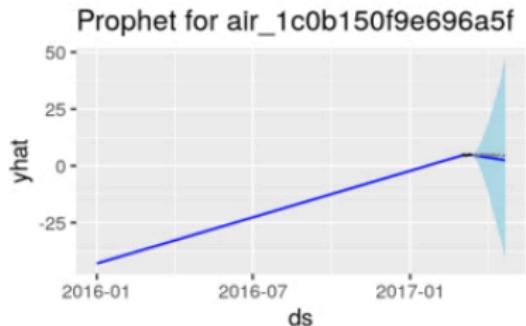
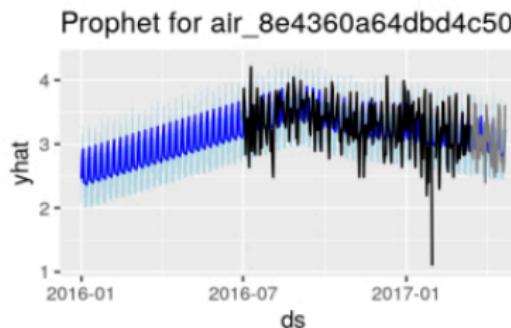
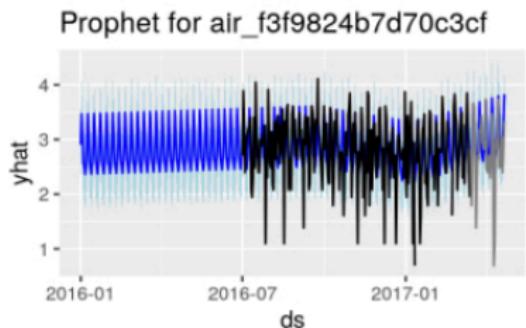
Prophet



Prophet



Prophet



Prophet

