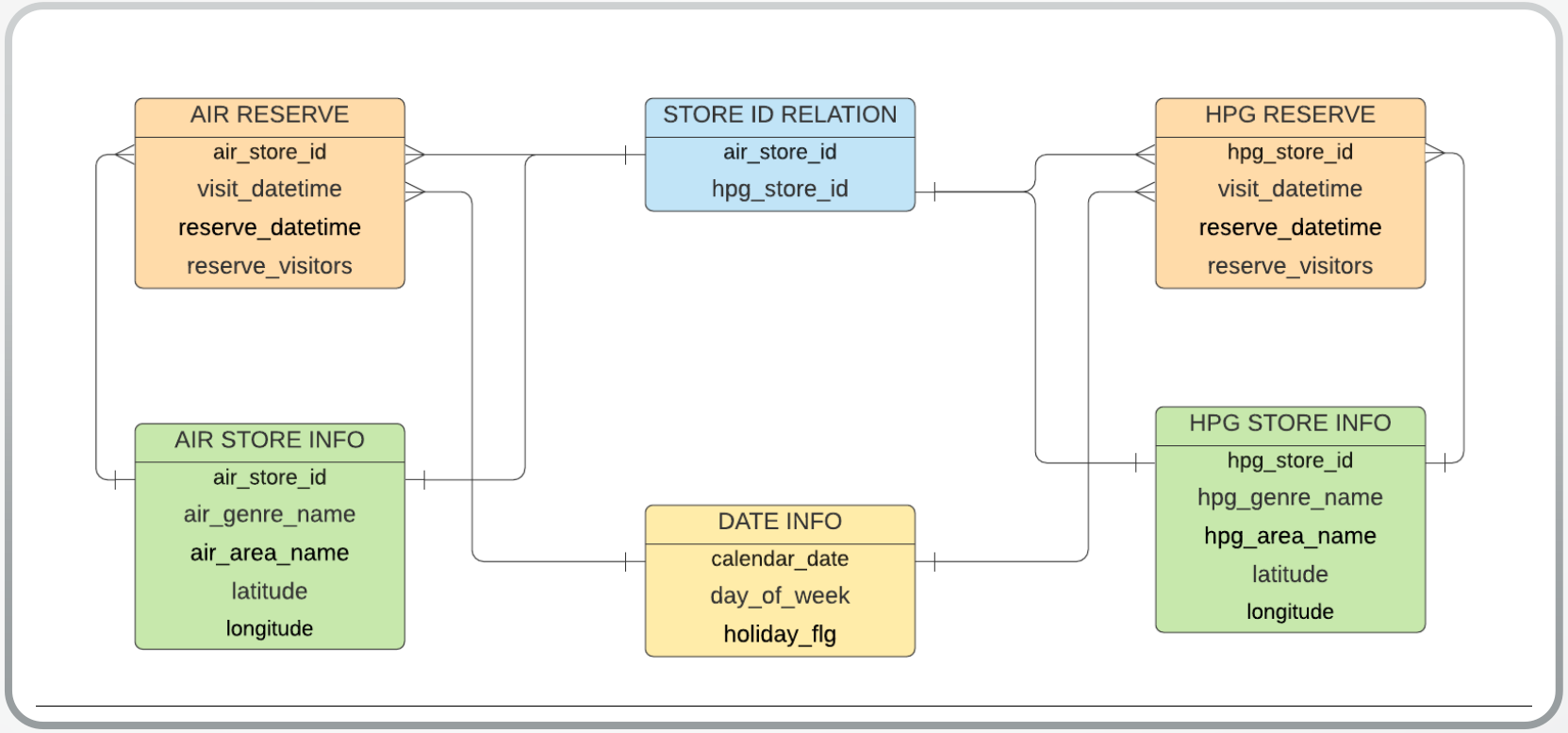
Recruit Restaurant Visitor Forecasting

Exploratory Data Analysis

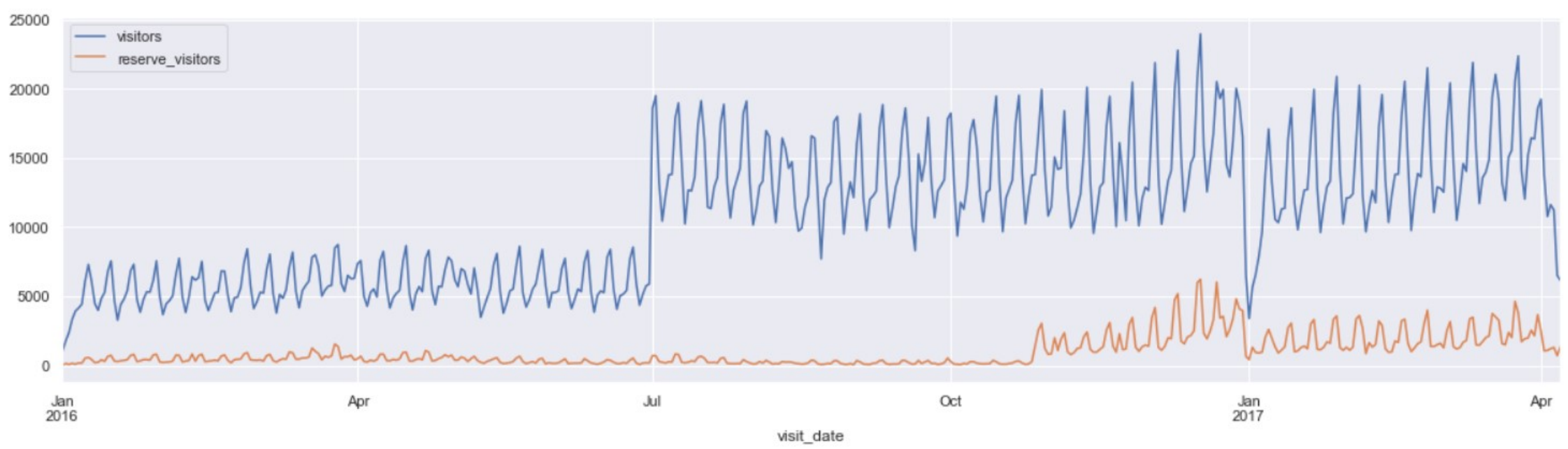
Team Name
Invincible Predictors

Team Members
Apoorv Panse
Deepti Chawda
Shahbaz Khan



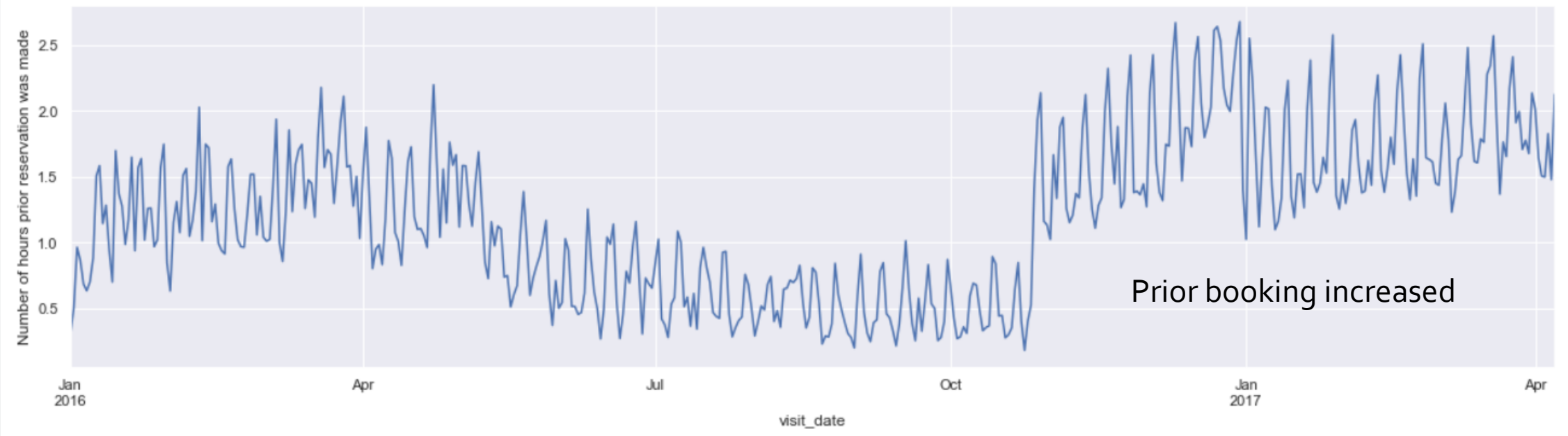
How Metadata is linked?

Total visitors and reservation timeseries



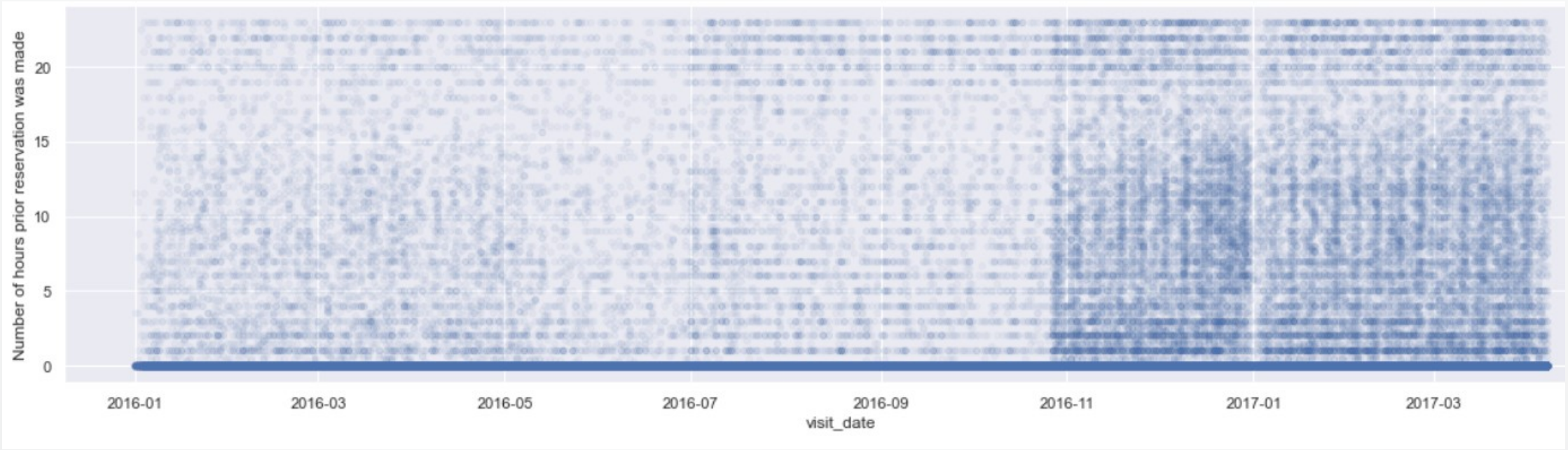
- Most of the restaurants have no prior reservations.
- Starting from November 2016, number of reservations started to grow. Which implies that new restaurants were onboarded to take reservations in advance.
- The dip during new year is caused because people prefer to spend new year with their families at home. First week of January sees less visitors.
- The reason of sudden increase in number of visitor in July 2016 is because many new restaurants were
- added in the database.

Yearly reservation activity



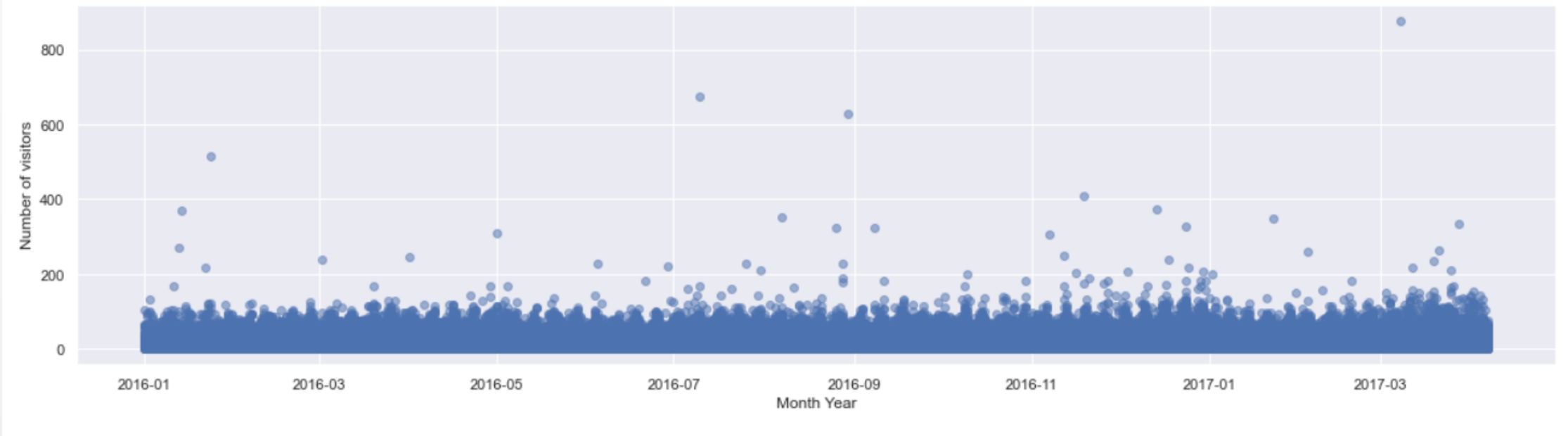
- Higher number of reservations are seen after November 2016 and hence we can clearly see that people need to book the restaurants much earlier than previously.
- People tend to reserve less during July to October. This can be a seasonal thing.

Overall Reservations Insights



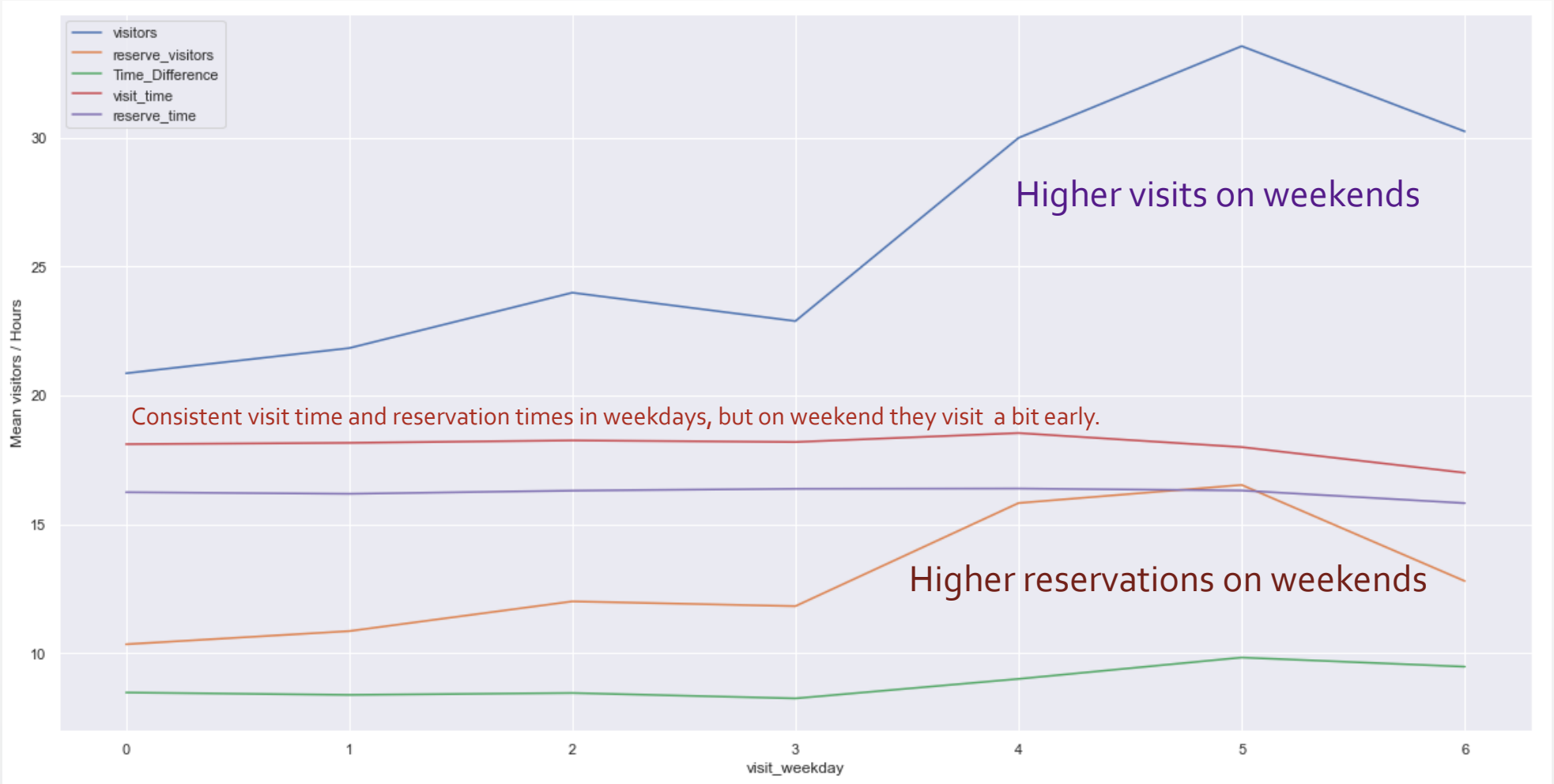
- For most of the restaurants we do not have reservations data. Hence most of the entries are 0
- Which means many restaurant do not provide reservation facility.
- After November, prior reservations were made even 24 hours before in many restaurants.

Total visitors per restaurant per day



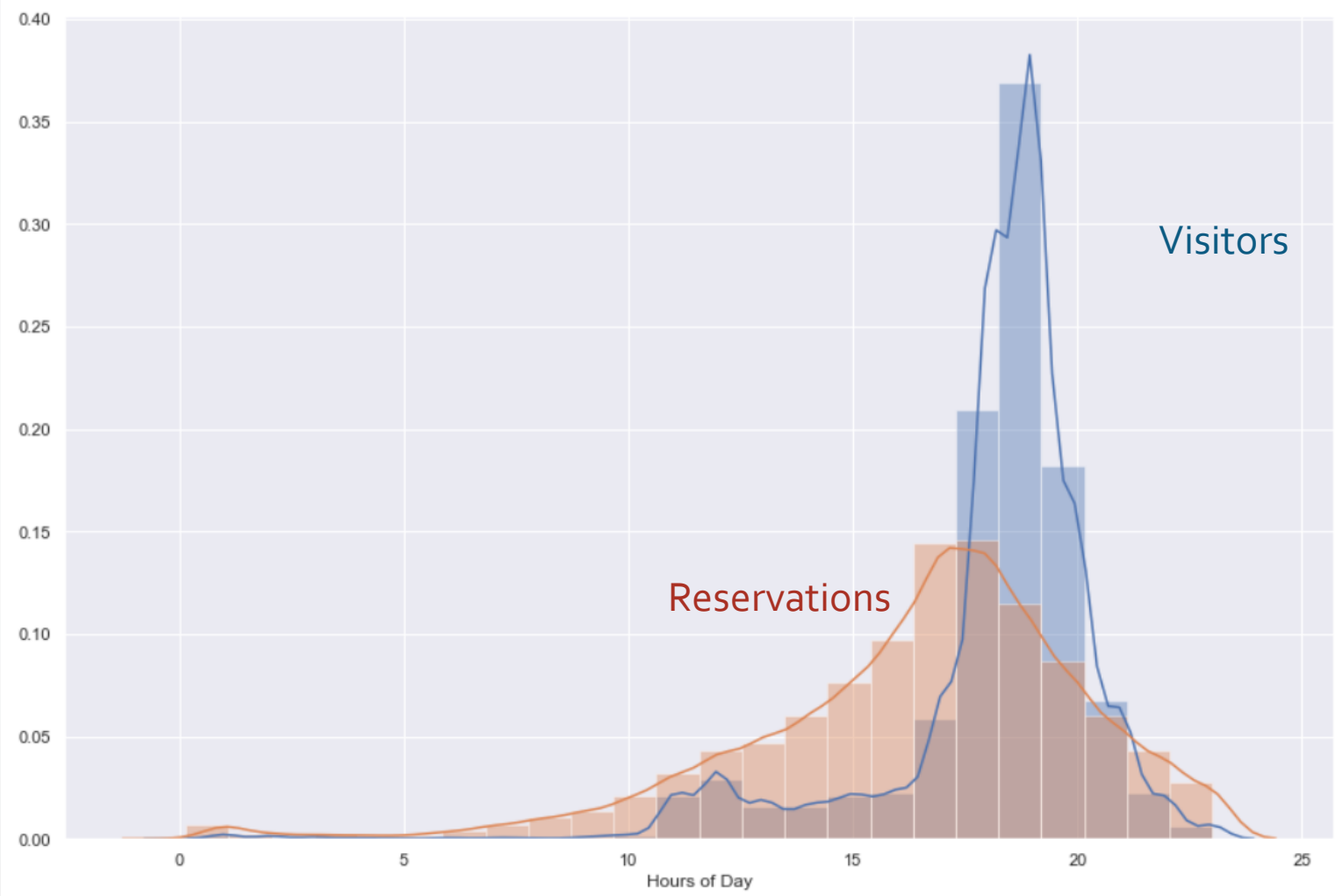
- We see the outliers here, where in a single day restaurants have more than 400 visitors.
- We can filter out outliers if needed and focus more on restaurants that have less than 200 visitors.

Activity based on day of week ?



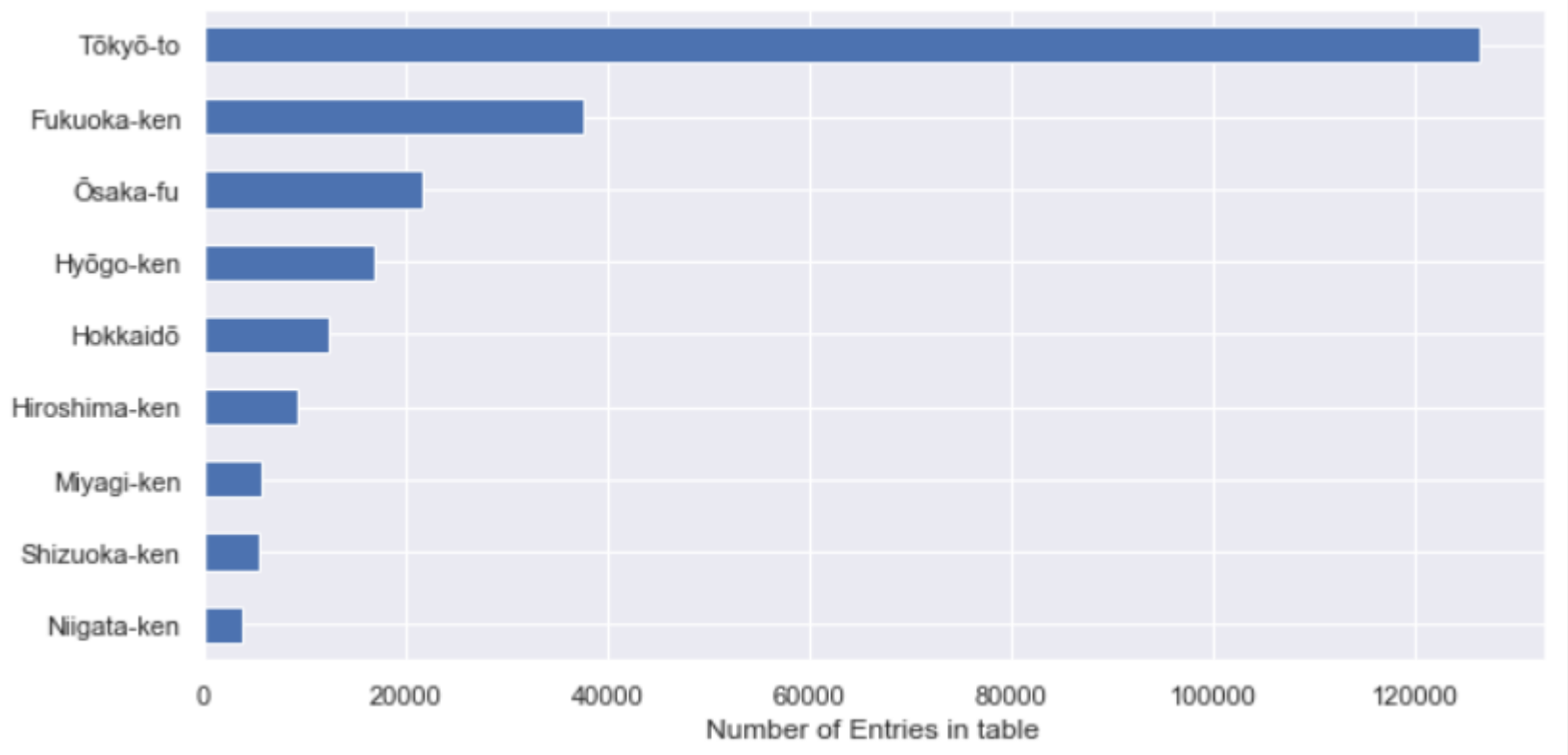
- Fridays and Saturday sees more number of visitors on average.
- Monday has lowest visitors.
- There seems to be high number of visitors when prior reservations were high.
- Day of week definitely contributes significantly to the number of visitors.
- More reservation means people have to book earlier to get the seat.

Hourly Visit and Reservation distribution



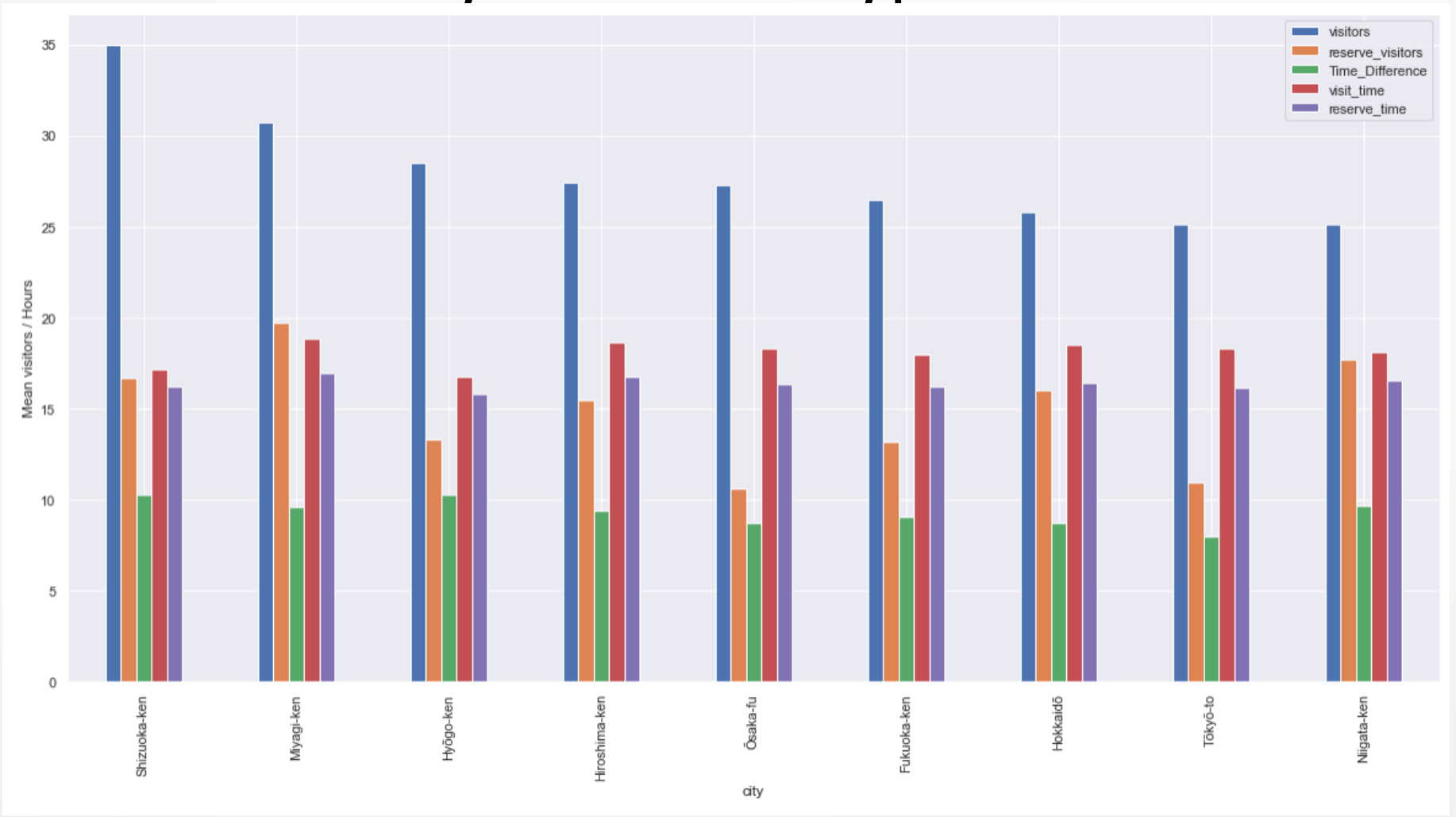
- People tend to visit the restaurants in the evening between 6 PM to 8 PM
- Most reservations are made between 3 PM and 7 PM
- Hence most people tend to have dinner rather than lunch.

Most Popular Cities as per number of entries



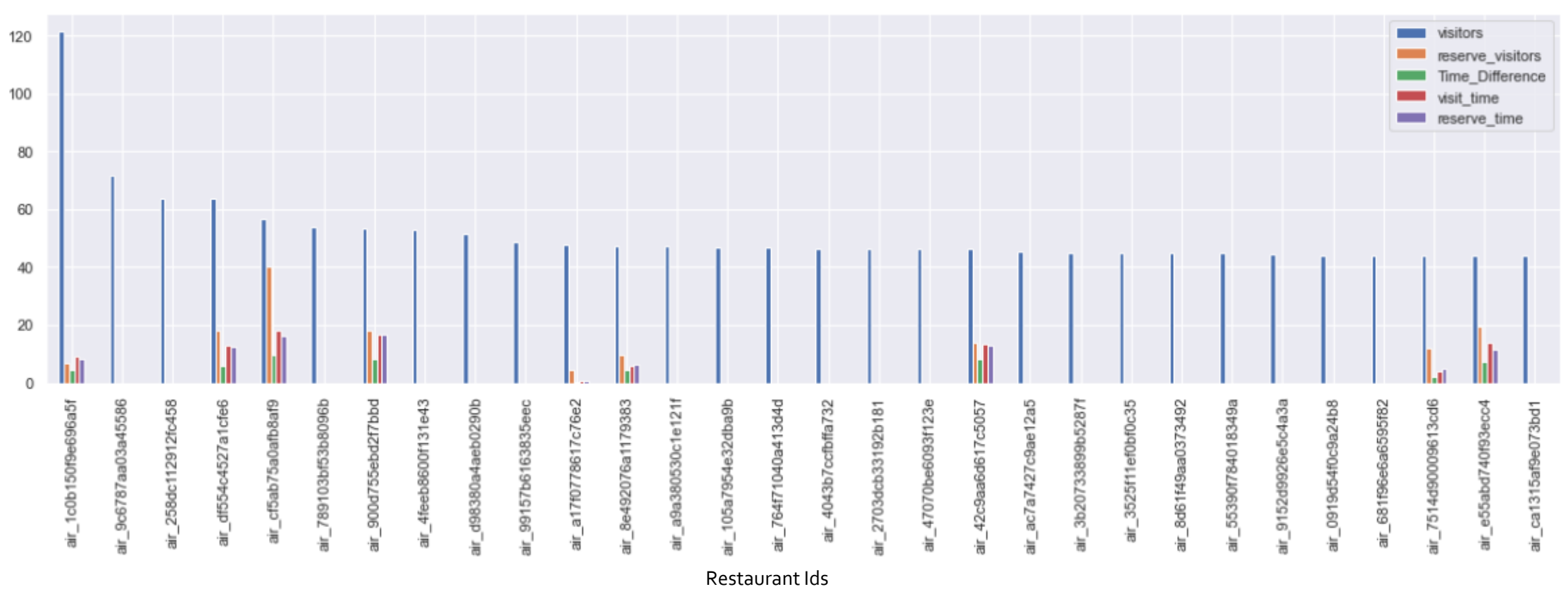
- Tokyo is the most popular city followed by Fukaoka and Osaka.
- These areas have highest number of restaurant and factually these area are populous.

Mean visitors, reserved visitors, prior reservations



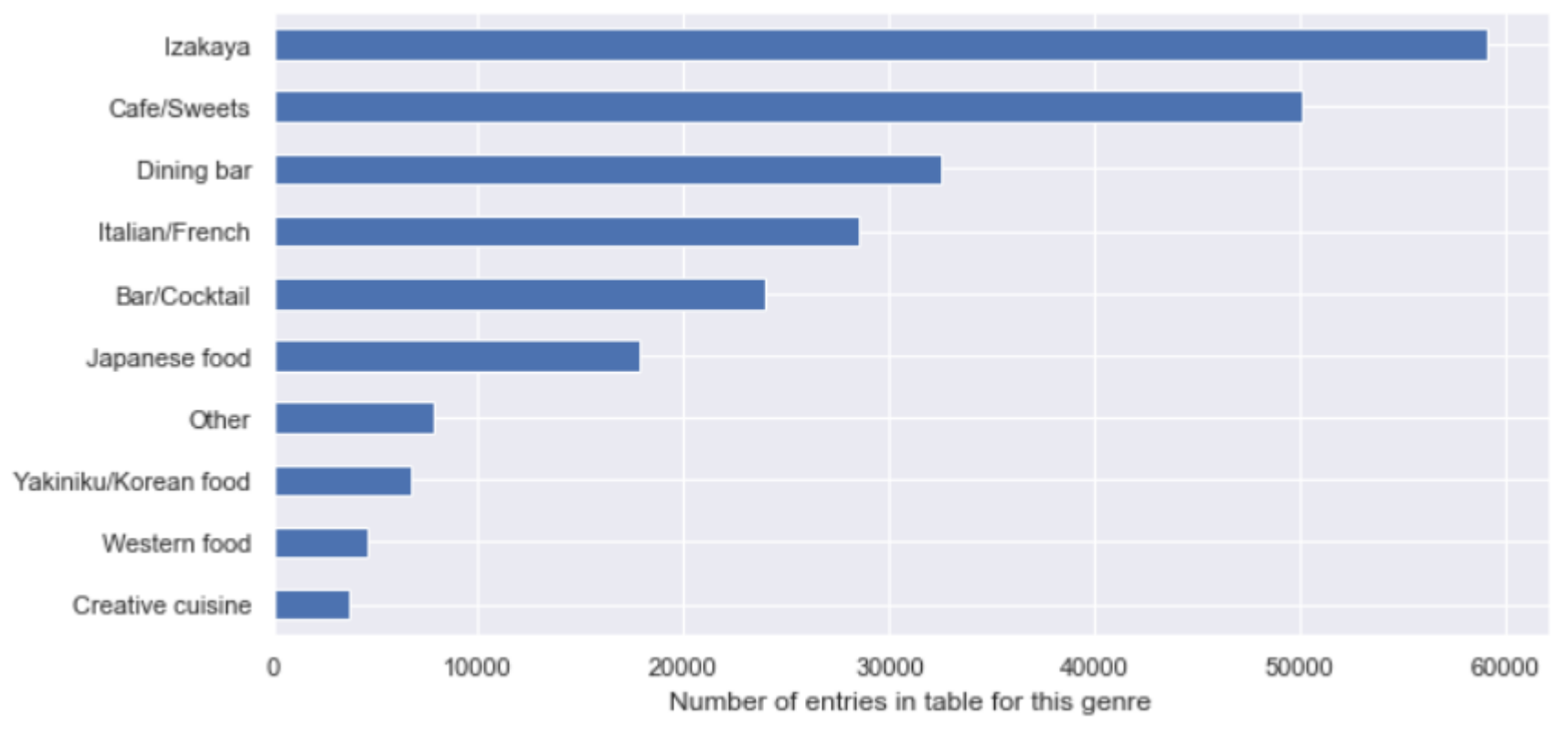
- Earlier we saw Tokyo, Fukuoka and Osaka were most populous cities.
- But Shizuoka, Myagi, Hyogo have higher number of mean visitors. This implies that Shizuoka have less number of restaurants but more visitor capacity.
- People of Myagi seems to reserve more before the visit.
- Tokyo must be having more number of small restaurants as mean visitor count is less.
- Visit time is almost consistent in all the cities, that is in the evening.

Top 30 Restaurants



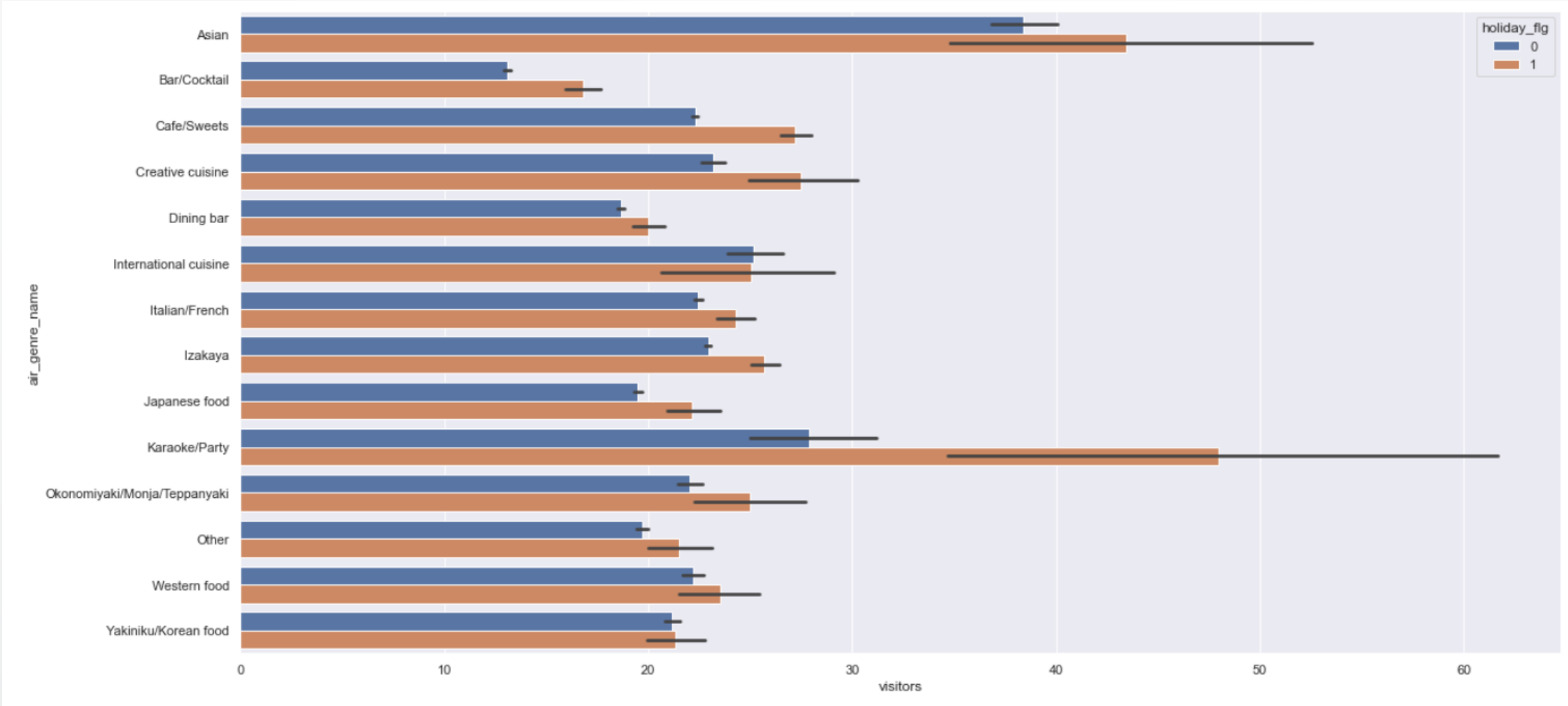
- Most of the top restaurants do not have reservation data.
- Most popular restaurants have visitors as high as 40-120 on average.

Most Popular Genre / Cuisine



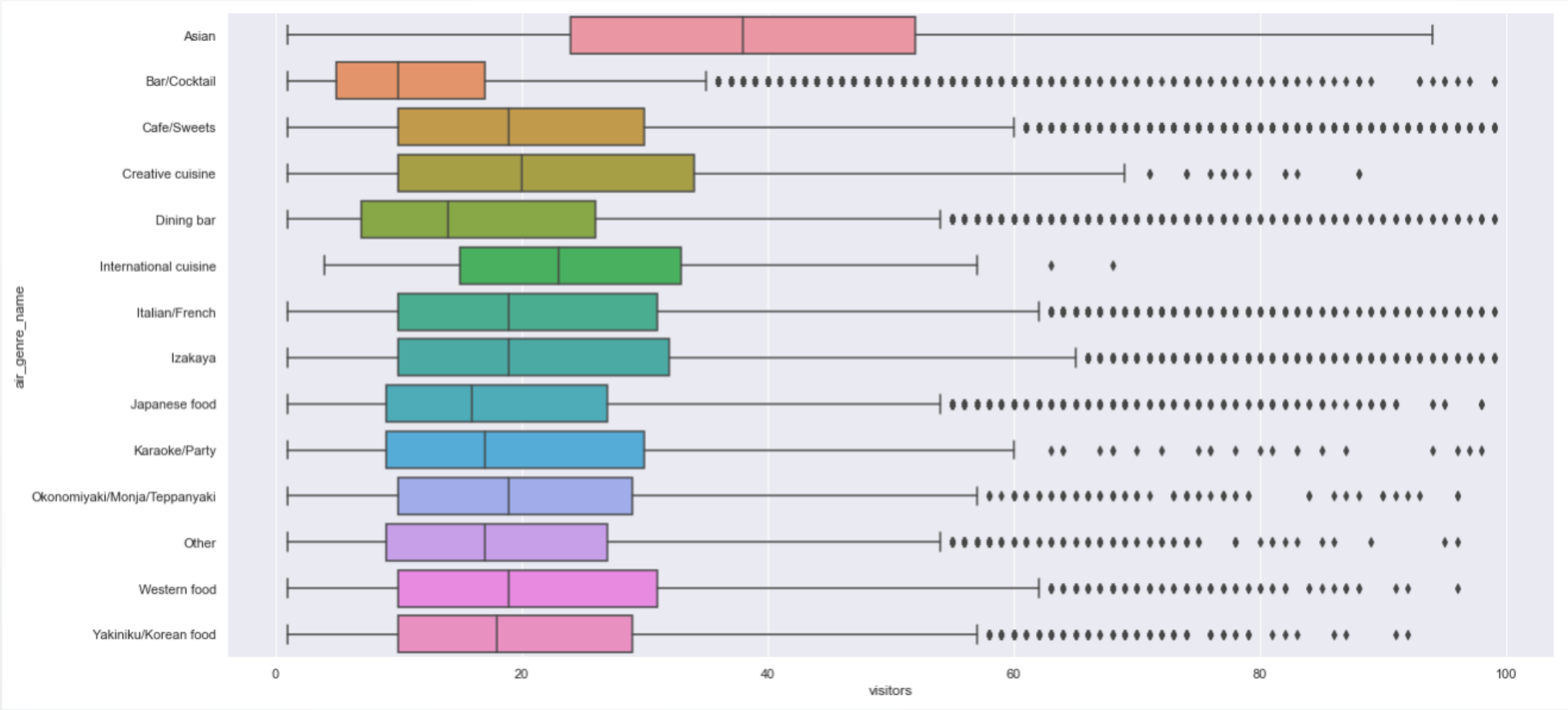
- Izakaya, café and Dining bars are the most popular genre and number of restaurants serving these genre is high.

Average visitors by Genre and Holidays



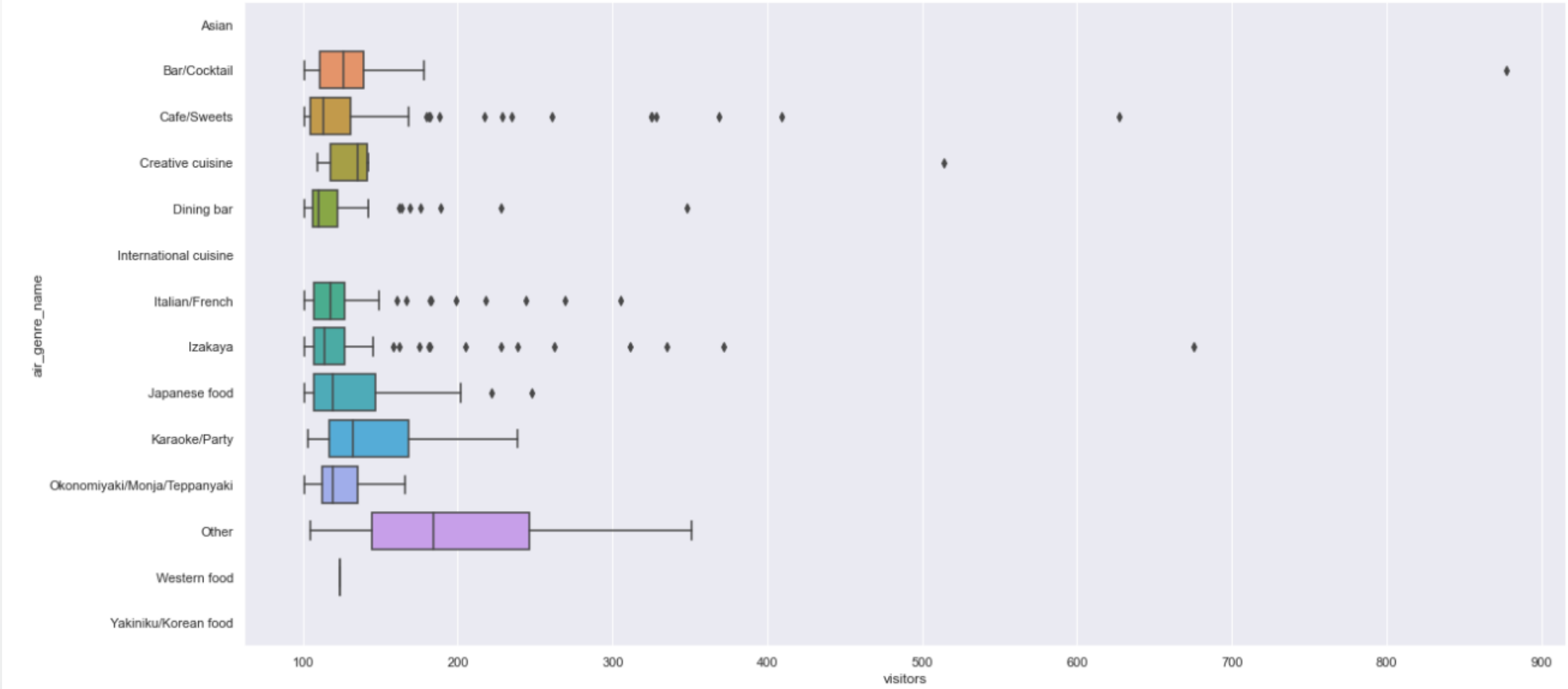
- There are more visits on holidays compared to normal days.
- Asian and Karaoke Party genre seems to have more average visitors, especially on holidays.
- Our earlier question about restaurants being closed or open on holidays, is resolved. Restaurants are open on holidays.

Genres people seem to prefer in smaller restaurants (Visitors < 100)



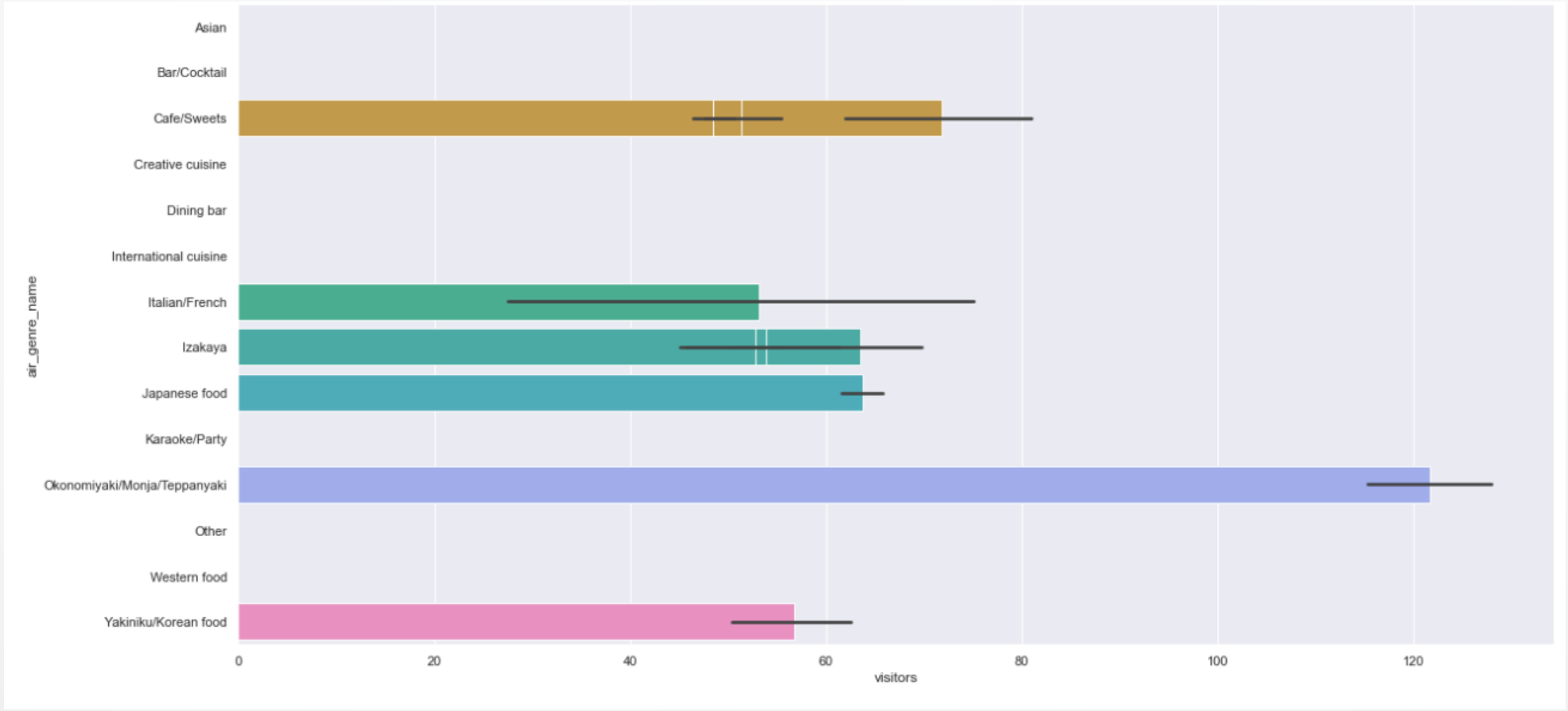
- Asian genre seems to have more average visitors, especially on holidays.
- Even though Izakaya has most number of entries in table, we have comparatively less visitors which implies that restaurants serving Izakaya must have small visitor capacity.

Genres people seem to prefer in bigger restaurants (Visitors > 100)



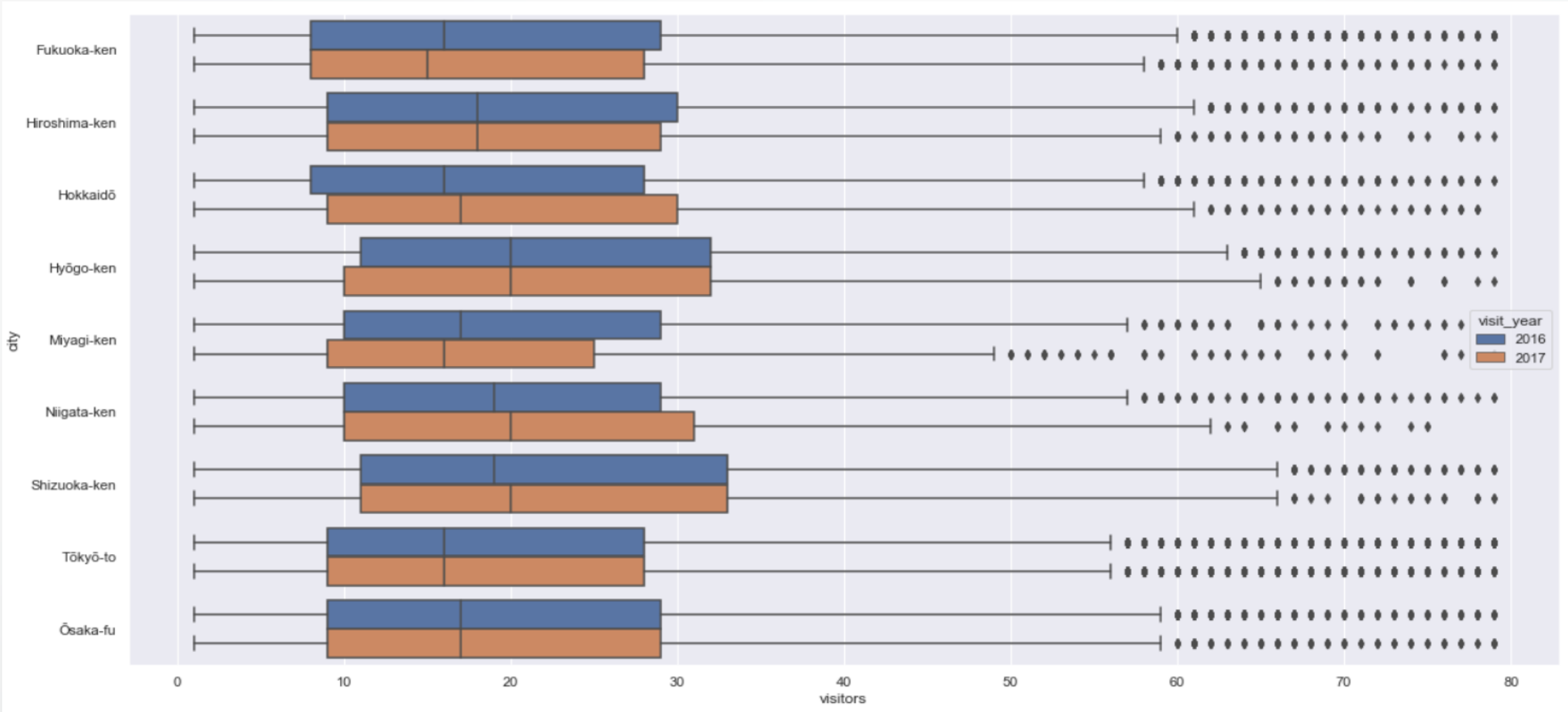
- Many big restaurants throw Karaoke Parties, Serve Japanese food.
- Even some cafes accommodate more visitors.
- Almost all of the restaurants serving Asian / Korean / Western food / International cuisine have small accommodation capacity.
- Size of restaurant is directly proportional to number of visitors, hence Genres they serve play important role.

Most Popular Genre / Cuisine in Top 10 restaurants



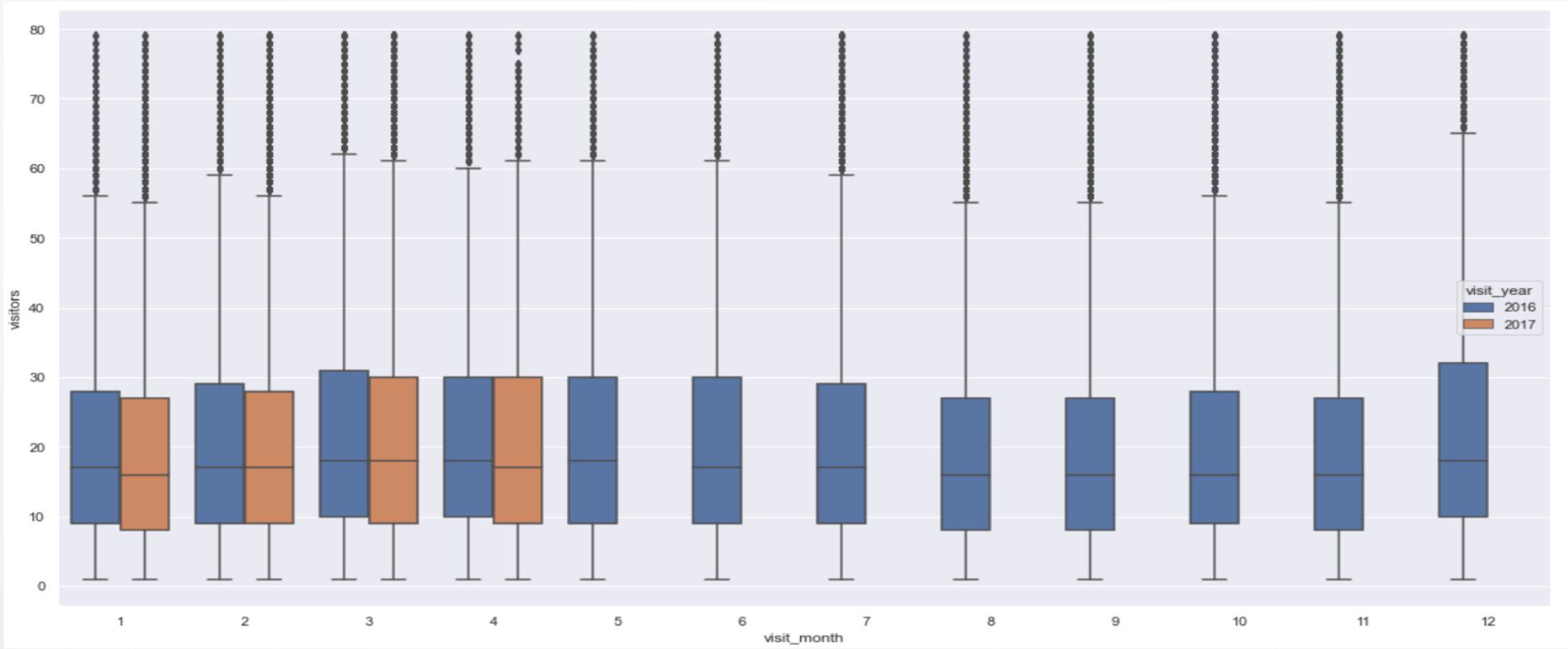
- Even though there are not many restaurants serving Monja, we see highest number of average visits. Rare genres might naturally see more visitors.
- Café, Izakaya, Italian Japanese and Korean genre seems more popular in top 10 restaurants.

Average visitors per city, year wise



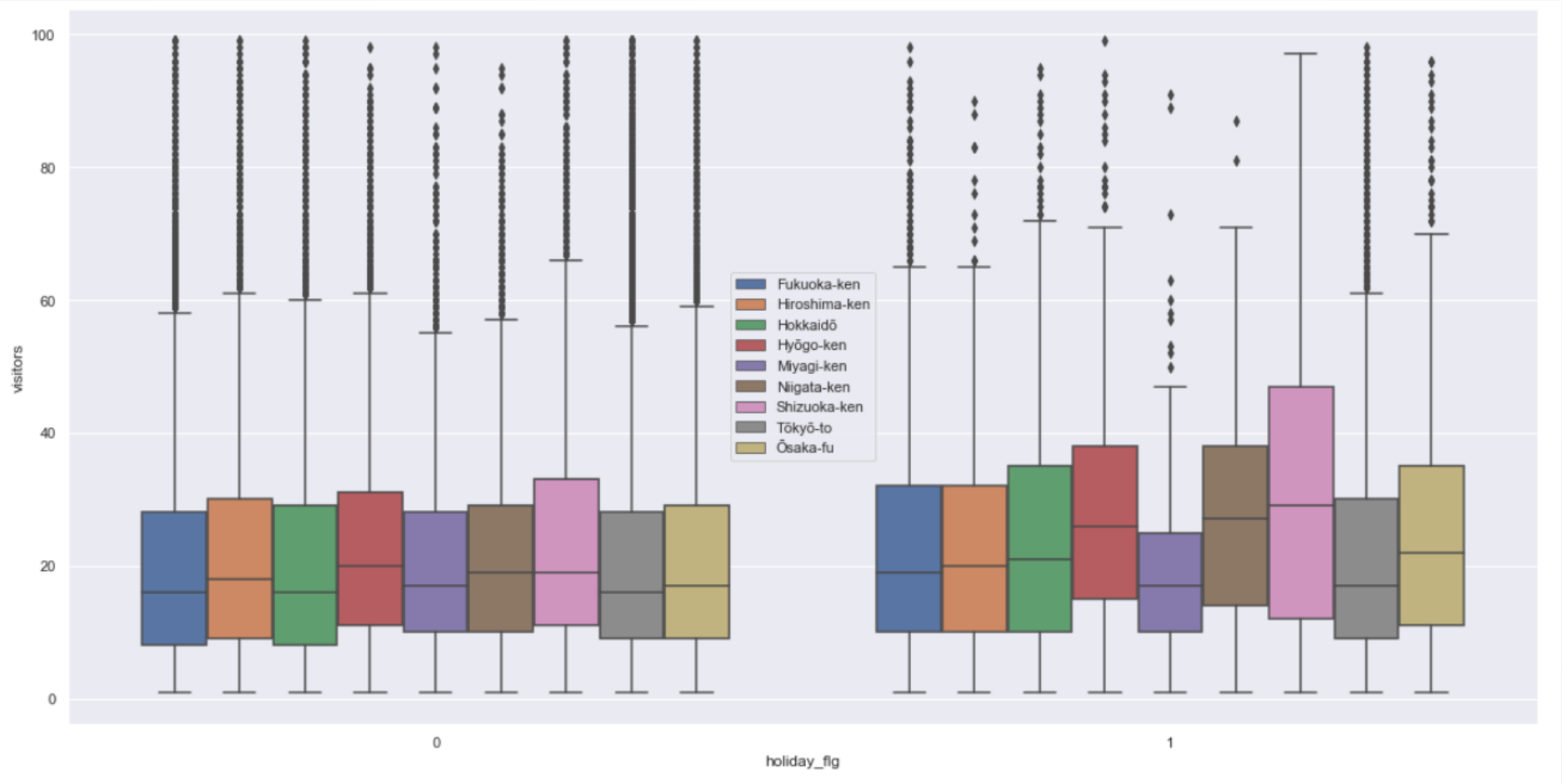
- Here we can see Fukuoka and Myagi saw less visits compared to previous year.
- Whereas Hokkaido, Shizuoka, Nigata saw slight increase in number of visitors.
- Tokyo , Osaka, Hiroshima, Hyogo saw consistent activity.

Average visitors per month, year wise



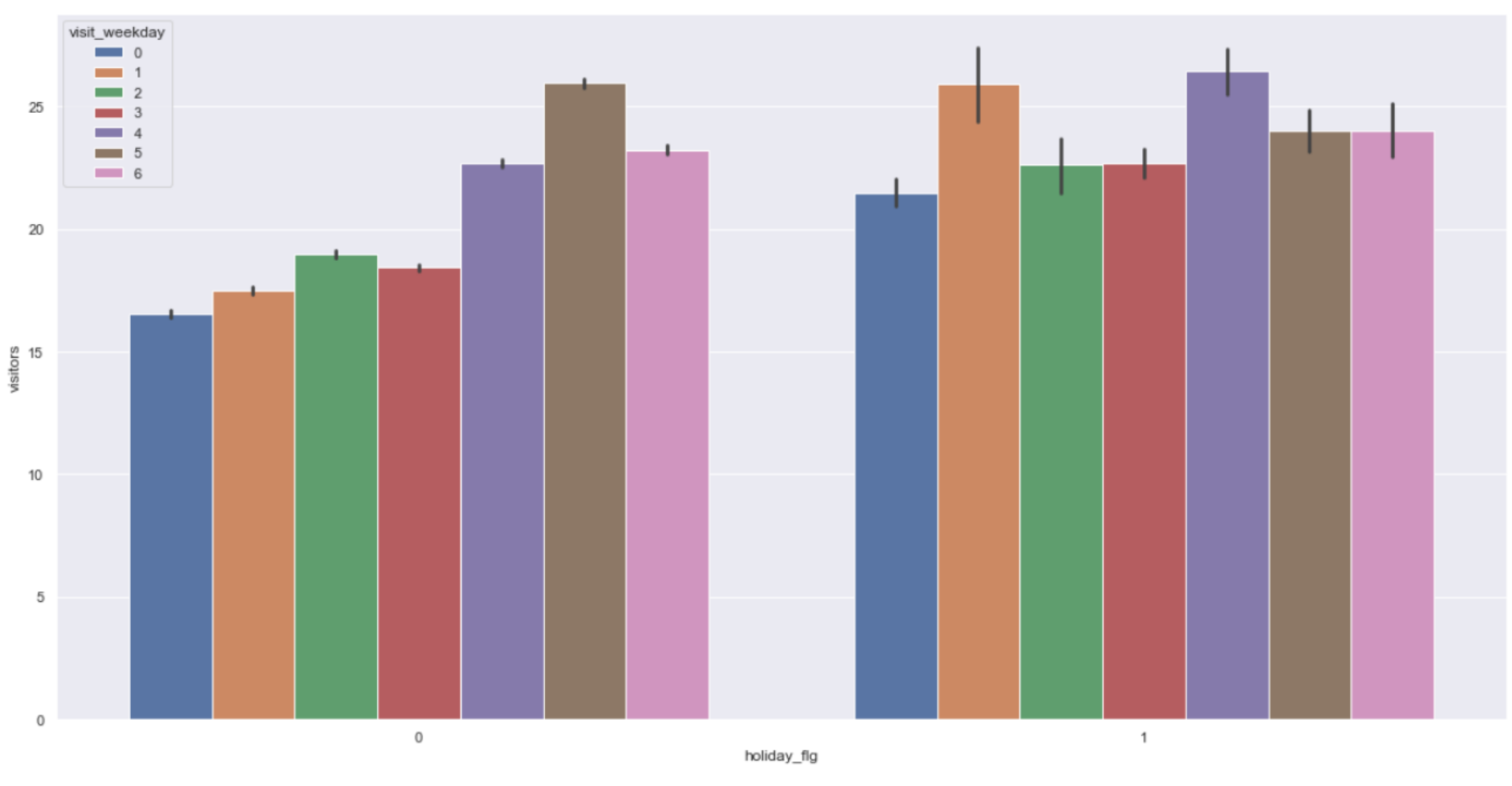
- The average visitors seems to be concentrated between the range 10 to 30
- We have excluded the outliers and considered max visitors as 80 for this chart.
- So now we can expect more predictions in this range.

Cities active on holidays



- Shizuoka is most active on holidays.
- Myagi seems to be less active on holidays.
- We can still see higher activity on holidays on average

Relation between weekdays and holidays



- We can see when there is no holiday, there are less visitors on weekday.
- When there are holidays, even Monday, Tuesday have more visitors.

Feature Selection

#	Column	Non-Null Count	Dtype
0	air_store_id	239673 non-null	category
1	visit_date	239673 non-null	datetime64[ns]
2	visitors	239673 non-null	int64 TARGET
3	air_genre_name	239673 non-null	category } Label encodings
4	air_area_name	239673 non-null	category }
5	reserve_visitors	239673 non-null	float64 } Imputations, as there are lot of missing values. Will drop if results are not as expected.
6	Time_Difference	239673 non-null	float64
7	visit_time	239673 non-null	float64
8	reserve_time	239673 non-null	float64
9	visit_year	239673 non-null	int64
10	visit_month	239673 non-null	int64
11	visit_weekday	239673 non-null	int64 } One hot encoding.
12	city	239673 non-null	category } Label encodings
13	ward	239673 non-null	category
14	neighborhood	239673 non-null	category
15	holiday_flg	239673 non-null	int64 } one hot encoding

As discussed, will add important new feature of restaurant_size (0-small restaurant, 1 – big restaurant)