Recruit Restaurant Visitor Data Analysis and Forecasting Project

Running a thriving local restaurant isn't always as charming as first impressions appear. There are often all sorts of unexpected troubles popping up that could hurt business.

One common problem is that restaurants need to know how many customers to expect each day to effectively purchase ingredients and schedule staff members. This forecast isn't easy to make because many unpredictable factors affect restaurant attendance, like weather and local competition. It's even harder for newer restaurants with little historical data.

Recruit Holdings has unique access to key datasets that could make automated future customer prediction possible. The datasets comes from a restaurant review service (Hot Pepper Gourmet) and restaurant point of sales service (AirREGI, and reservation log management (Restaurant Board).

Recruit challenges people to use reservation and visitation data to predict the total number of visitors to a restaurant for future dates. This information will help restaurants be much more efficient and allow them to focus on creating an enjoyable experience for their customers.

In this paper, a key findings from a data analysis of the available datasets are presented, and a prediction algorithm is constructed with ARIMA for different types of restaurants.

The results indicate that there is a lot of interesting properties of the datasets, which could be used to improve processes. In addition, the prediction is shown to perform remarkably well in predicting the amount of visitors, in general, to different types of restaurants.

The original source of the problem and datasets can be found at:
https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting

Recruit Restaurants - Forecasting - Data Analysis - Description of datasets, distribution of visitors and reservations

- For a start, all the datasets available will be presented, for an idea of what characteristics we can analyze

Description of Datasets

Want to present the available datasets and typical values

AirREGI Reservations

- Contains information of reservations done in the Air system

Store ID: Identification of the restaurant	air_6b15edd1b4fbb96a
Visit Date: The date for the reservation	2016-01-02 17:00:00
Reservation Date: The date the reservation was done	2016-01-01 22:00:00
Visitors: The amount of spots reserved	3

Hot Pepper Gourmet Reservations

- Contains information of reservations done in the HPG system

Store ID: Identification of the restaurant	hpg_33ec1499d6b13141
Visit Date: The date for the reservation	2016-01-01 17:00:00
Reservation Date: The date the reservation was done	2016-01-01 15:00:00
Visitors: The amount of spots reserved	2

Date Information

- Contains information of dates

Calendar Date: A date	2016-01-01
Day of the Week: Which day of the week	Friday
Holiday: Whether the date is a holiday	1

AirREGI Store Information

- Contains information of the stores in the air system

Store ID: Identification of the restaurant	air_0f0cdeee6c9bf3d7
Store Genre: The type of restaurant	Italian/French
Area: The area the restaurant is located at	Hyogo-ken Kobe-shi Kumoidori
Latitude: The latitude of the restaurant's location	34.6951242
Longitude: The longitude of the restaurant's location	135.19785249999998

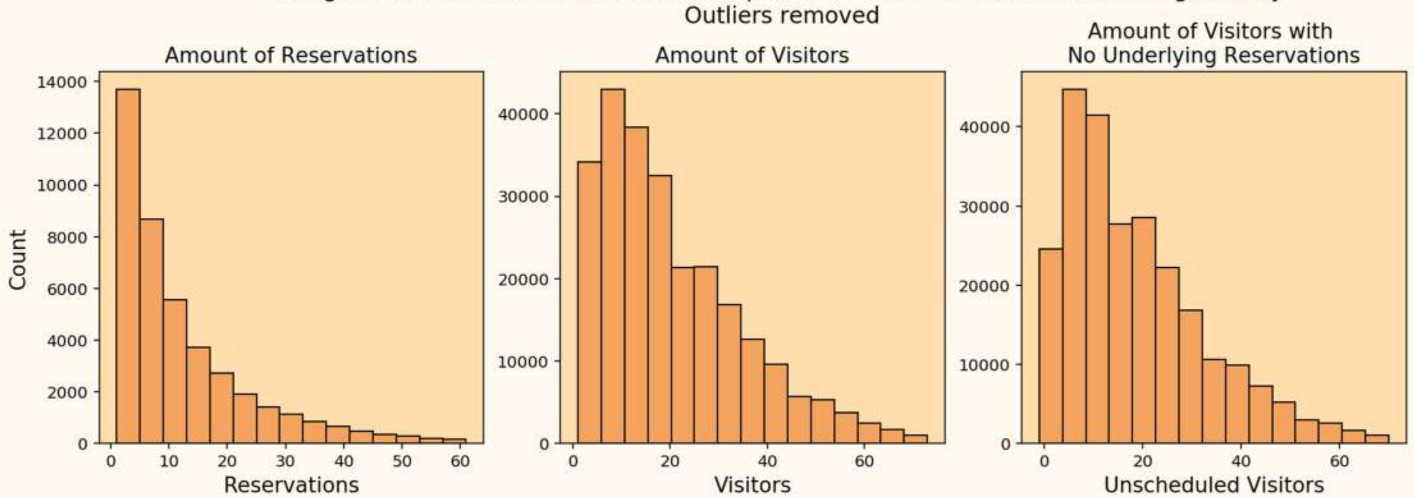
AirREGI Visitors Information

- Contains information of the visitors at restaurants for different dates

Store ID: Identification of the restaurant	air_ba937bf13d40fb24
Visit Date: A date the restaurant is open	2016-01-13
Visitors: The amount of visitors for a date	25

- To give a rough idea of how many people people, on average, visit and book reservations for restaurants, we may consider the distribution of visitor and reservation values

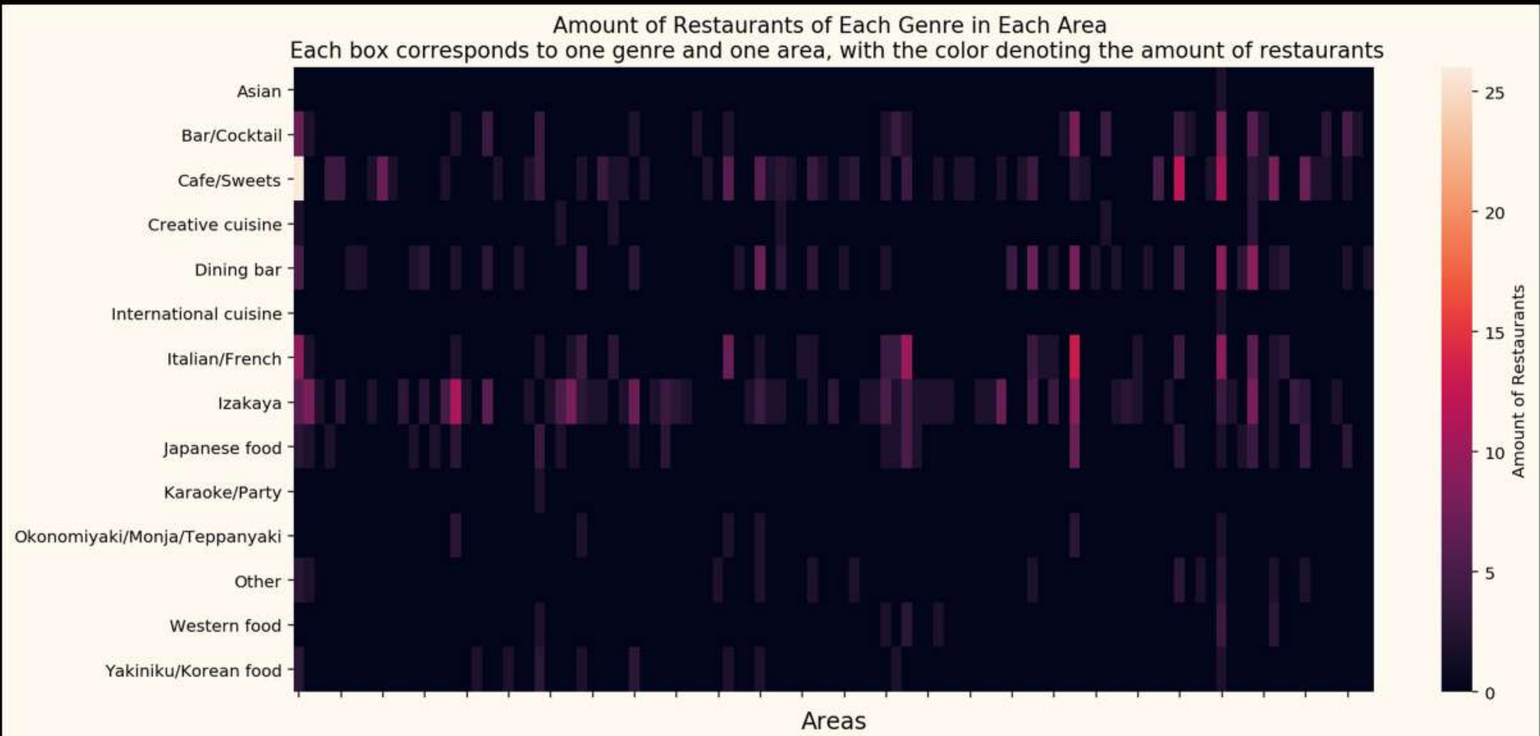
Histogram of reservations, visitors, and unplanned visitors for restaurants, for a given day
Outliers removed



· Amount of reservations done for a restaurant, for a given day, is often smaller than 10 but there exists cases where reservations reach larger values like 10, 20, 30 - perhaps they corresponds to holiday/weekend days
· Amount of visitors, generally, exceeds the number of reservations, implying that restaurants often receive more customers than the amount of reservations - indicates consumers seldom have to worry about a restaurant running out of places
· Further illustrating the relation between reservations and visitors, the amount of visitors with no underlying reservations are mostly non-negative, implying again that restaurants tends to receive more customers than the amount of reservations they receive
· Also note that from the amount of reservations and visitors, we can see that there exists a lot of cases where a restaurant may receive visitors without a single reservation, for a particular day - which might corresponds to weekdays or other low-traffic inducing days
· Days with high amount of visitors most likely correspond to weekends or holidays, which might explain why these cases occur much less than days with low amount of visitors (which probably, in turn, corresponds to weekdays)

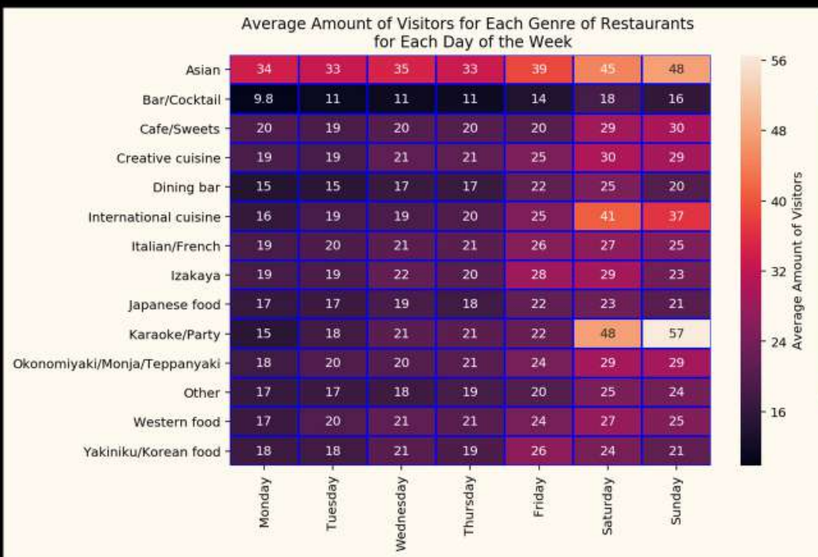
Recruit Restaurants - Forecasting - Data Analysis - Type of restaurants in areas, amount of visitors to every type of restaurant over the week

- As all areas are different, there is an interest in considering what type, and how many, restaurants exists in each area - perhaps there is some area with a lot of Dining Bars



· There is one area with a lot of Cafe/Sweets - might correspond to a place where spending moderate amount of money on food/drinks is really popular, like a tourist place, shopping center, or close to perhaps schools
· Areas that have a large amount of restaurants of one type usually have a large/moderate amount of restaurants of other types - might indicate that the underlying area is a popular place to eat/drink, for example inner cities, or areas with a large nightlife
· As expected from the analysis of the amount of restaurants of each type, Cafe/Sweets, Izakaya, and Italian/French restaurants are spread, more or less, over all different areas. The large spread of Cafe/Sweets and Izakaya over different areas indicates that informal restaurants are attractive among Japanese people

- Intuitively, the traffic to restaurants should change over the week, with perhaps an uptick of visitors during the weekend. This can be considered by the average amount of visitors for each type of restaurant over the week

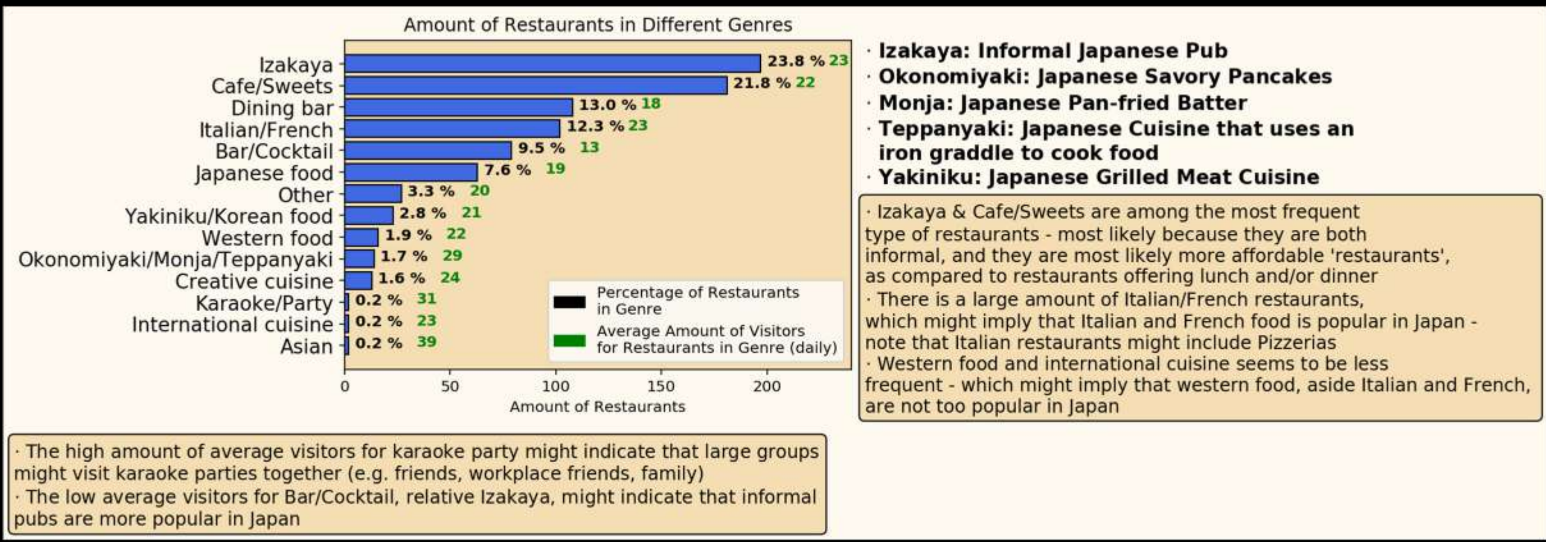


	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Asian	34	33	35	33	39	45	48
Bar/Cocktail	9.8	11	11	11	14	18	16
Cafe/Sweets	20	19	20	20	20	29	30
Creative cuisine	19	19	21	21	25	30	29
Dining bar	15	15	17	17	22	25	20
International cuisine	16	19	19	20	25	41	37
Italian/French	19	20	21	21	26	27	25
Izakaya	19	19	22	20	28	29	23
Japanese food	17	17	19	18	22	23	21
Karaoke/Party	15	18	21	21	22	48	57
Okonomiyaki/Monja/Teppanyaki	18	20	20	21	24	29	29
Other	17	17	18	19	20	25	24
Western food	17	20	21	21	24	27	25
Yakiniku/Korean food	18	18	21	19	26	24	21

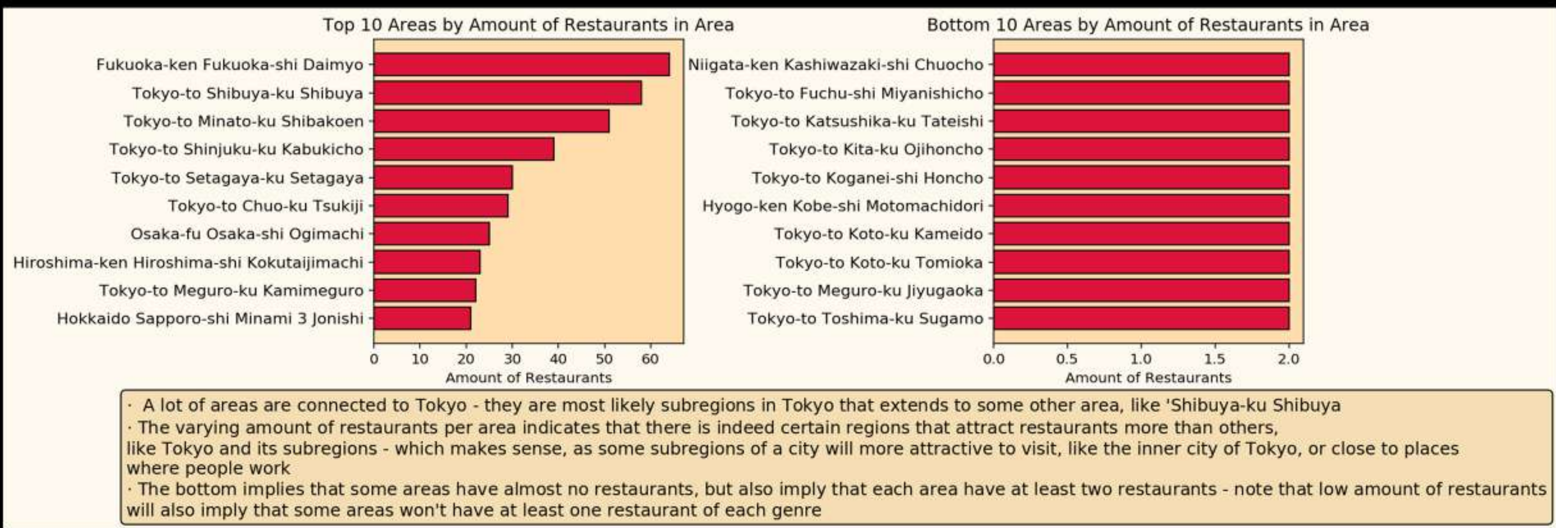
· Asian has the highest amount of visitors in general all weekdays, by a wide margin. Implies that Asian-oriented restaurants attract a lot of Japanese people. Might be because Japan is in Asia, and a generic Asian restaurant might offer a diverse set of meals for most people's preference, as opposed to e.g. Italian/French which attract people who like European food.
· International cuisine has a big jump in visitors during the weekend. A possible reason is that people want to eat something different toward the weekend, and since Japan is an Asian country, international food might be an attractive choice. For the same reason that Westerner might want to eat Asian food during the weekend
· Karaoke/Party, likewise, has a big jump in visitors during the weekend. Most likely, the weekend is an appropriate day to go out with people to either a a karaoke or party
· The weekday with the most visitors for all genres is friday (except weekend). No surprise as this is the last working day before the weekend, and people feel like eating something special to celebrate the weekend

Recruit Restaurants - Forecasting - Data Analysis - Type of restaurants in Japan, amount of restaurants in areas

- To get an idea of what type of restaurants exists in Japan, and in what proportions, we consider an analysis of the air_genre_name attribute in the datasets

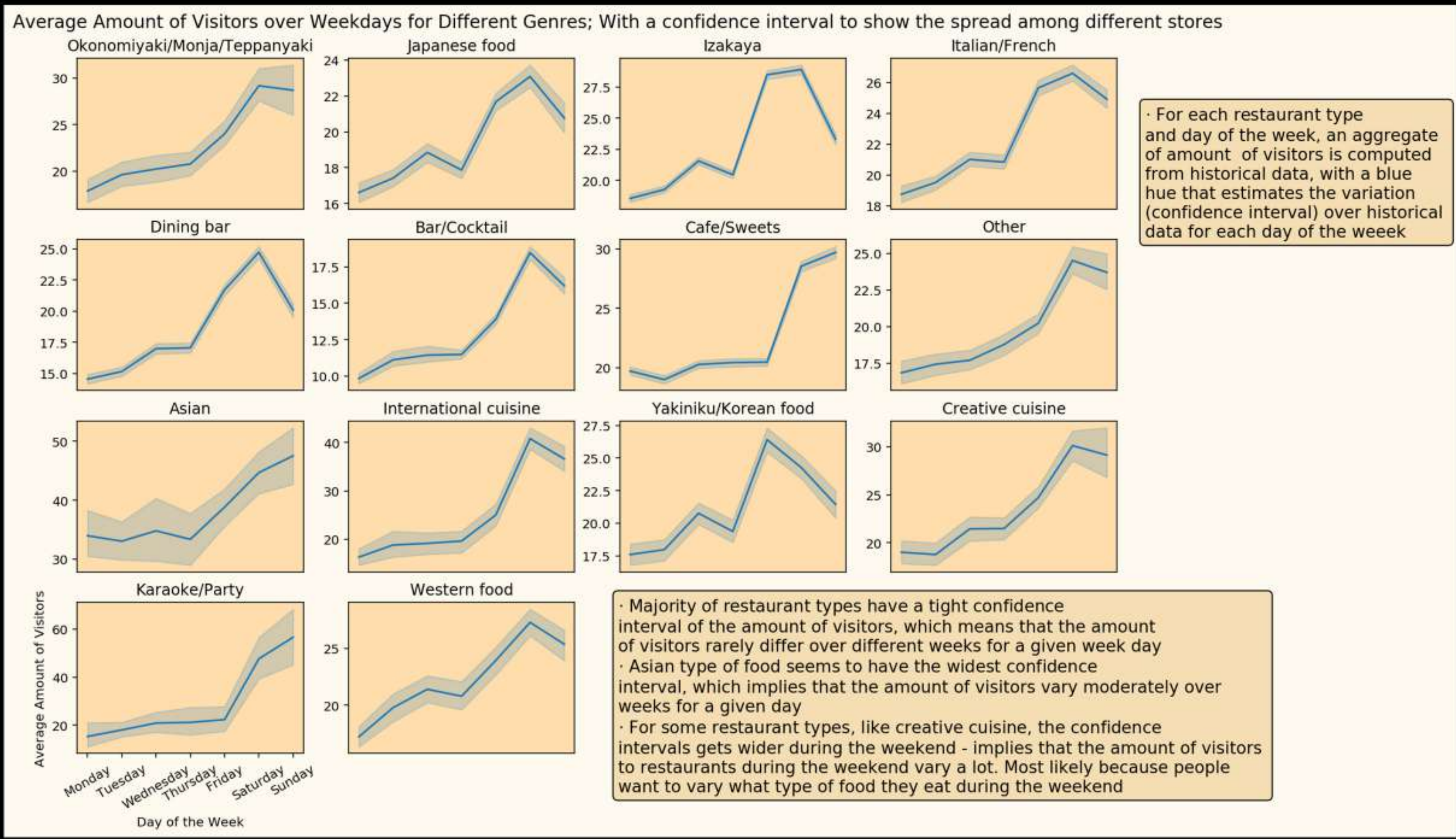


- Another interesting property is what areas of Japan exists, and how many restaurants exists in each area. Because the amount of areas exceed 100, only the top ten and top bottom areas by count are considered

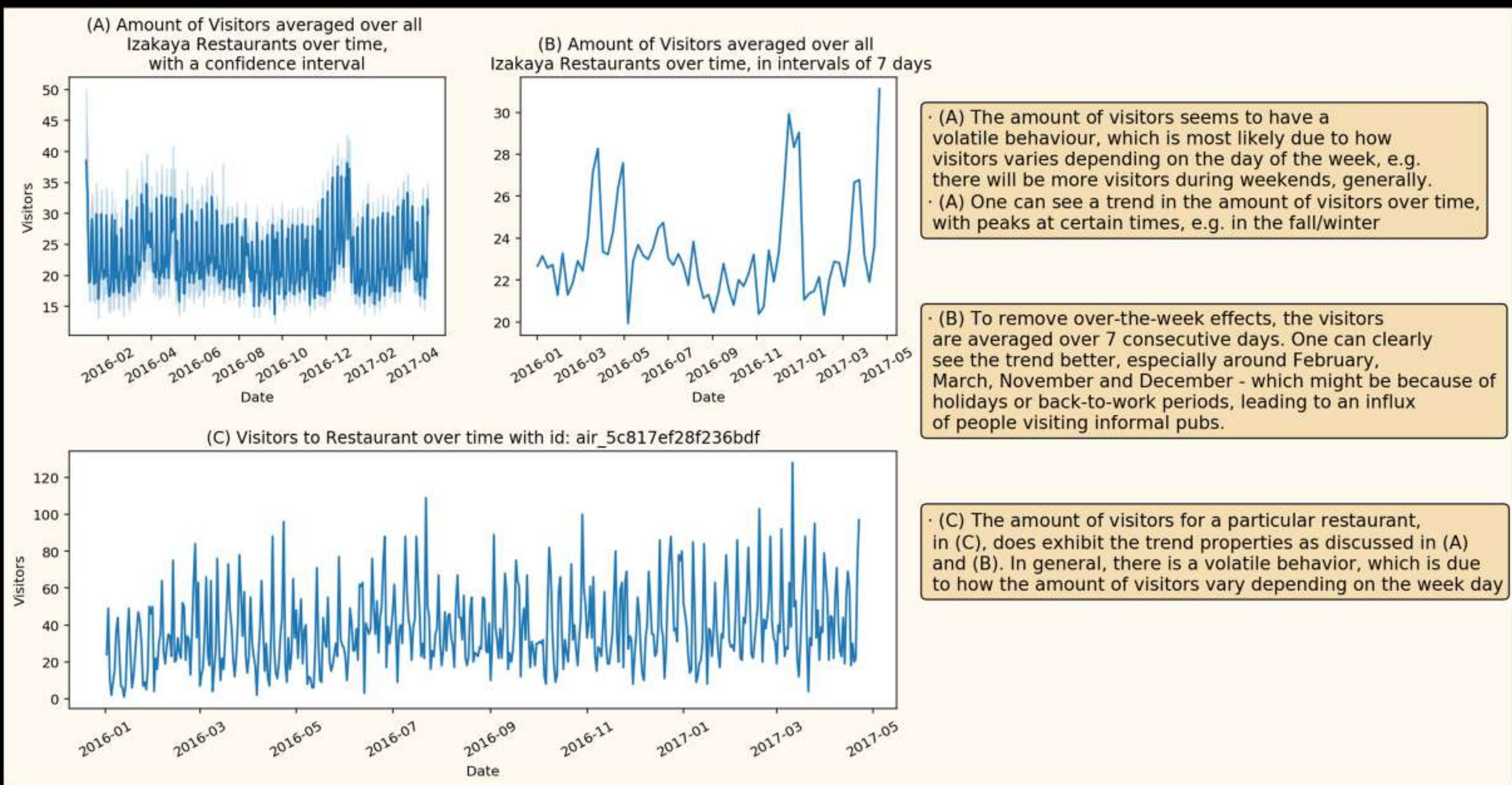


Recruit Restaurants - Forecasting - Data Analysis - Average Amount of Visitors over the week, time-series for a particular restaurant

- To analyze how the amount of visitors varies over each day of the week, the average of amount of visitors is computed for each genre over all days of the week

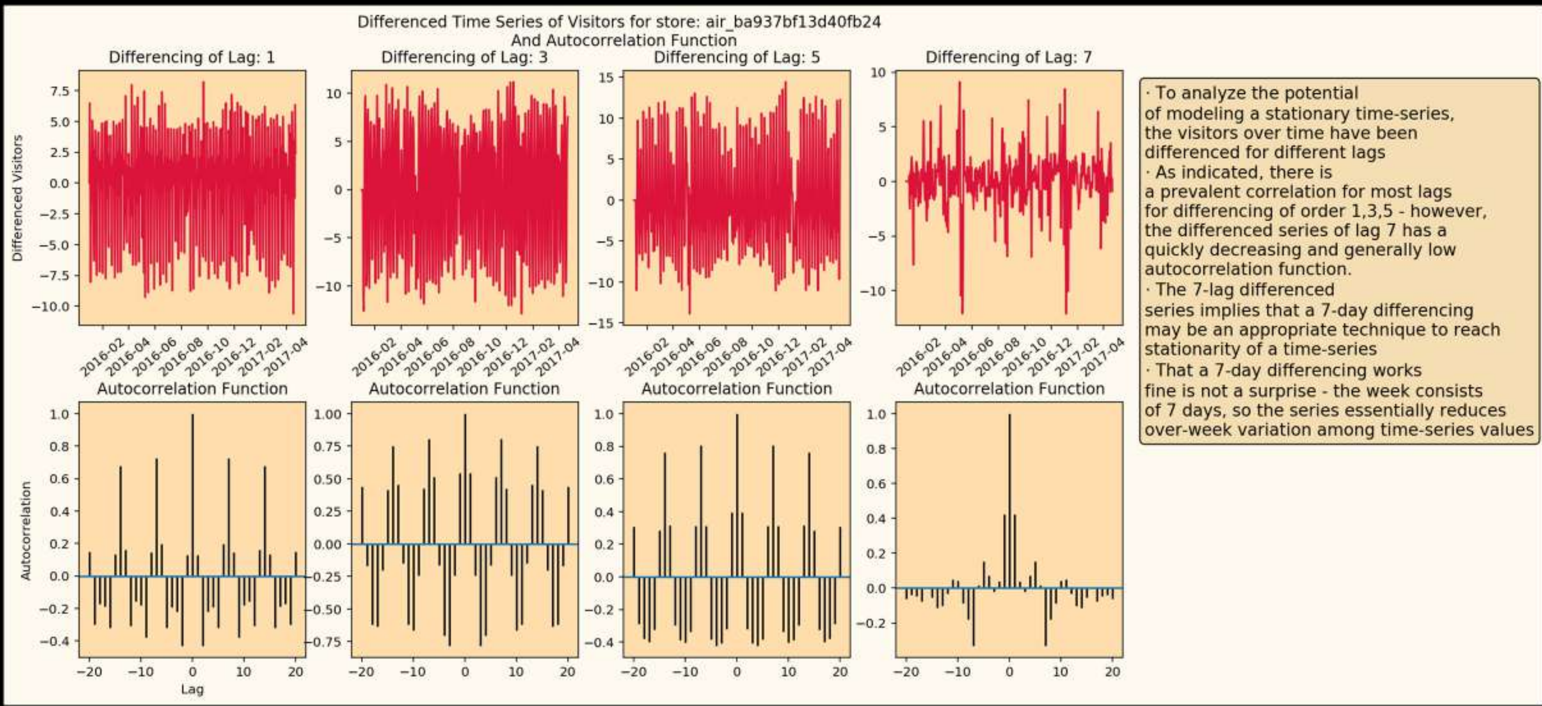


- For visitors in a particular genre, an idea is to analyze restaurants in the Izakaya genre. In particular, it is of interest to analyze the amount of visitors over time for a particular store in the Izakaya genre

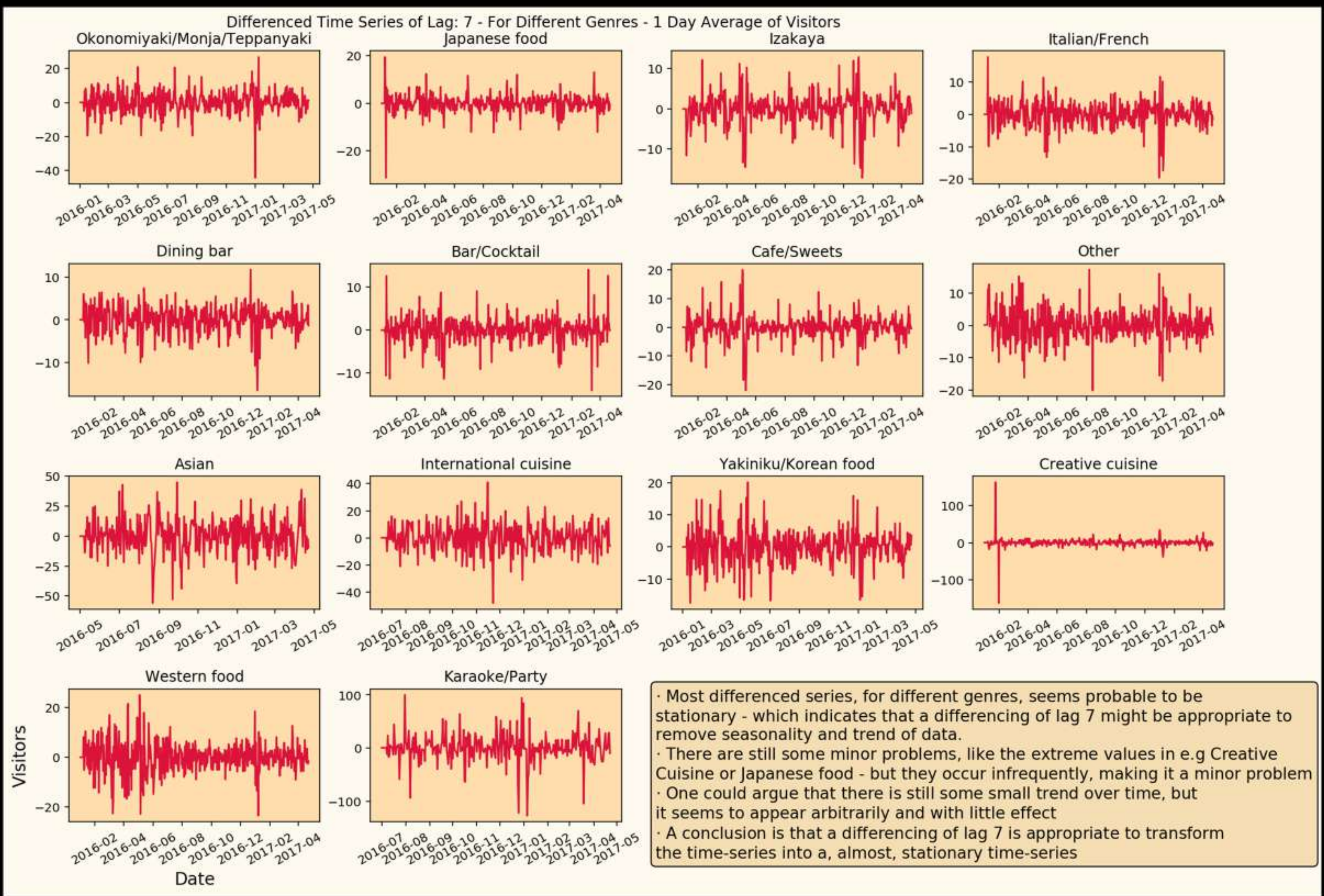


Recruit Restaurants - Forecasting - Prediction Analysis - Time series analysis

- Before applying mathematical models to our time-series data, a key idea is to analyze whether the time-series can be transformed to a stationary, or almost, time-series. One typical approach is to difference the time-series, and check for weak stationarity (constant mean over time, and time-independent autocorrelation function)



- To further illustrate the effect of differencing time-series with a lag of 7, we consider this type of transformation to the average amount of visitors (daily) for each type of restaurant, to see if the assumption of stationarity is plausible



Recruit Restaurants - Forecasting - Prediction Analysis - Optimal ARIMA models, and prediction on a week of days with optimal ARMA models

- With our 7-day differenced time-series of the daily average visitors of each genre, the next step is to fit it to optimal ARIMA models. Based on considering different AR and MA orders, and comparing the prediction performance with two metrics on a test week

Optimal Parameters for Different Genres - MAE

Genre Name	p	q	MAE
Okonomiyaki/...	5	3	2.66
Japanese food	6	3	1.37
Izakaya	5	3	1.65
Italian/French	7	3	2.58
Dining bar	6	1	0.64
Bar/Cocktail	7	1	0.96
Cafe/Sweets	4	2	1.79
Other	6	1	1.88
Asian	5	1	12.4
International cuisine	4	2	5.66
Yakiniku/Korean food	5	2	1.25
Creative cuisine	6	3	3.8
Western food	2	3	2.65
Karaoke/Party	6	3	9.86

Optimal Parameters for Different Genres - RMSE

Genre Name	p	q	RMSE
Okonomiyaki/...	7	2	3.46
Japanese food	7	3	1.91
Izakaya	6	2	2.0
Italian/French	7	3	3.0
Dining bar	6	2	0.86
Bar/Cocktail	7	3	1.2
Cafe/Sweets	7	2	2.65
Other	6	1	2.09
Asian	4	2	14.61
International cuisine	4	2	6.5
Yakiniku/Korean food	5	2	1.87
Creative cuisine	6	3	4.59
Western food	2	3	3.22
Karaoke/Party	6	3	11.45

ARIMA(p,d,q) Model - with statsmodels package (Python)
Based on utilizing d=7, i.e. a differencing of lag 7
To find optimal p and q, two metrics, MAE and RMSE, are evaluated on on a test set consisting of a week, given an ARIMA model trained on a training set.
For the search of optimal p, q, all ARIMA models for p=1,2,...,7 and q=1,2,...,7 have been considered

In above tables, the optimal choices of p and q for each genre is presented.
MAE is the mean of the residuals, with respect to true and predicted time series values.
RMSE is the square root of the mean of the residuals squared

- With our optimal models, based on MAE, the last step is to evaluate the models on a week of days, the validation set, to really measure the performance of our models