Restaurant Visitor Forecasting

Exploratory Data Analysis

29.08.2019

Vineet Joshi
Assignment - 1
MT19020

1. Introduction

This is an exploratory data analysis report on Restaurant Visitor Forecasting challenge, held on Kaggle by *Recruit Holdings*. The dataset used here is obtained from the challenge website.

Recruit Holdings owns Hot Pepper Gourmet (a restaurant review service), AirREGI (a restaurant point of sale service), and Restaurant Board (reservation log management software).The dataset from these sources provides us with information such as reservations, visits, etc to forecast future restaurant visitors on a given date.

2. About Dataset

The dataset is collected from two different sites:

- Hot Pepper Gourmet (HPG)
- AirREGI / Restaurant Board (AIR)

Both of these sites are used for reservation management and cash registration and stores data for each registered restaurant.

2.1 File Description

Data collected from AIR system.

File Name	Description	Attributes
air_visit_data.csv	This file contains visitors data for AIR restaurants.	• *air_store_id* • *visit_date* • *visitors*
air_store_info.csv	This file contains information about select AIR restaurants.	• *air_store_id* • *air_genre_name* • *air_area_name* • *latitude* • *longitude*
air_reserve.csv	This file contains data about reservations made through AIR.	• *air_store_id* • *visit_datetime* • *reserve_datetime* • *reserve_visitors*

Table 2.1

Data Collected from HPG system.

File Name	Description	Attributes
hpg_store_info.csv	This file contains information about select HPG restaurants.	• *hpg_store_id* • *hpg_genre_name* • *hpg_area_name* • *latitude* • *longitude*
hpg_reserve.csv	This file contains data about reservations made through HPG system.	• *hpg_store_id* • *visit_datetime* • *reserve_datetime* • *reserve_visitors*

Table 2.2

Other beneficial data.

File Name	Description	Attributes
store_id_relation.csv	Relation between select restaurants that uses both AIR and HPG systems.	• *hpg_store_id* • *air_store_id*
date_info.csv	Basic information about calendar dates in the dataset.	• *calendar_date* • *day_of_week* • *holiday_flg*

Table 2.3

3. Analysing AIR System's Data

3.1 Spatial Distribution Of Restaurants And Popular Locations.

Kindly consult the following workbook for detailed visualization:

Tableau Workbook: VisitorForecasting

Sheet 1 : Distribution Of Select AIR Restaurants

Sheet 2 : Visitors' Distribution Of Top AIR Areas.

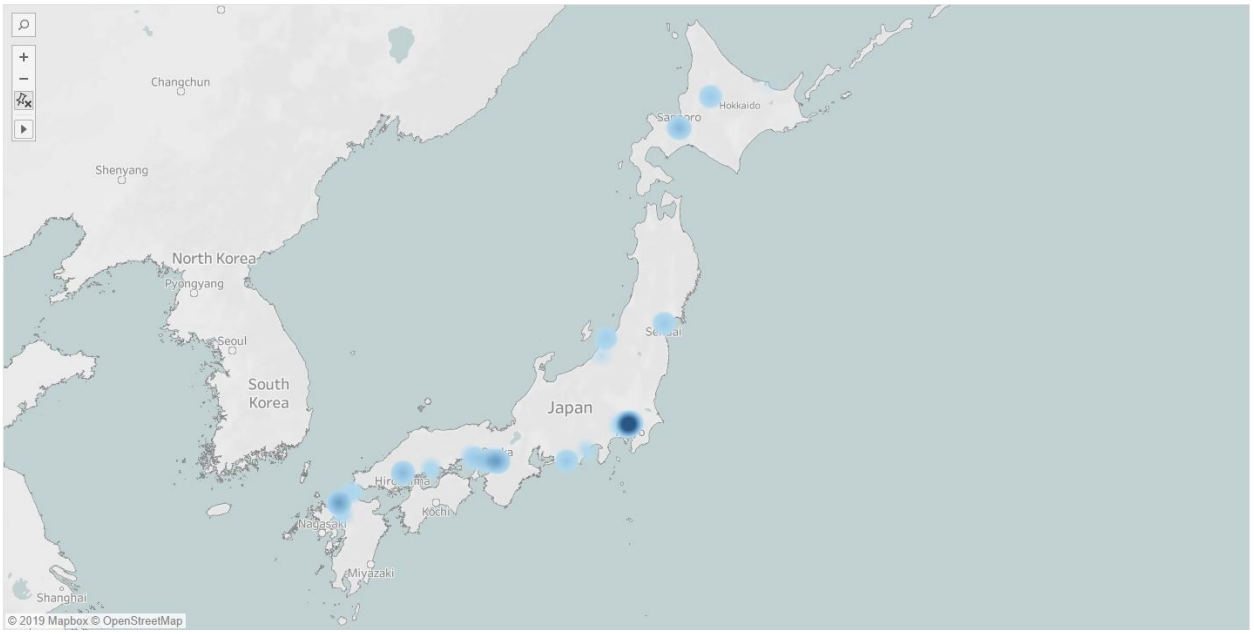


Fig 3.1.1 Spatial Distribution Of AIR Restaurants Across Japan. The darker areas represent higher density.

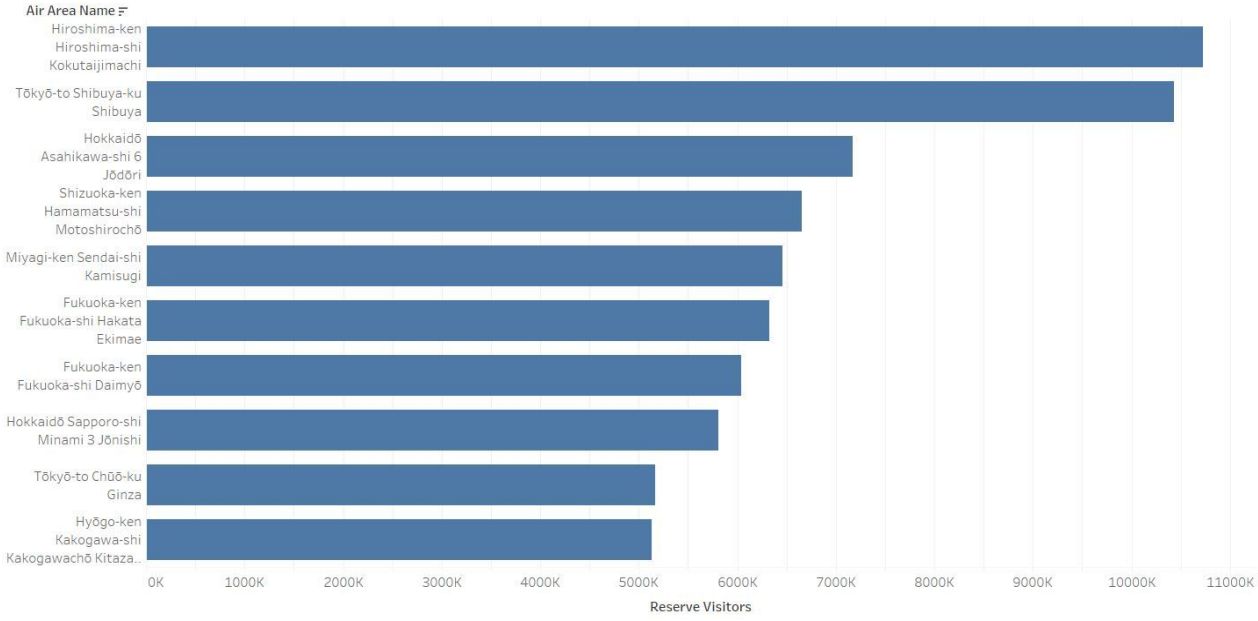


Fig 3.1.2 Top Localities In Japan With Highest Number Of Visitors Via AIR.

Inferences:

Based on the above figure (Fig 3.1.2),,it is quite clear that Hiroshima has the highest number of people visiting AIR restaurants, followed closely by an area in Tokyo. This can be due to the fact that Hiroshima has one of the highest densities of AIR restaurants (Fig 3.1.1).

3.2 Temporal Distribution Of Visitors.

Kindly consult the following workbook for detailed visualization:

Tableau Workbook: VisitorForecasting

Sheet 3 : Monthly Visit Data For AIR Restaurants

Sheet 4 : Visit Data For AIR Restaurants - Distributed Over Weekdays

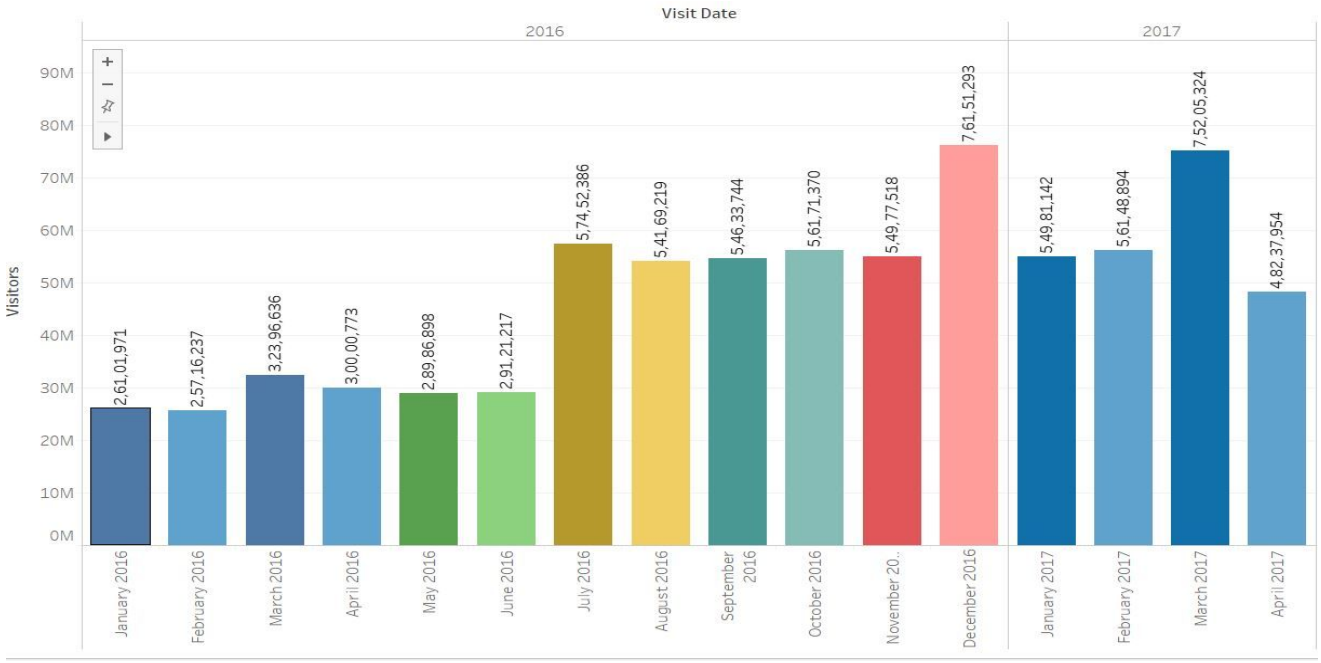


Fig 3.2.1 Numbers Of Visitors In AIR Restaurants Each Month For The Year 2016 & 2017 (Quarter 1).

Inferences:

- During the entire year (2016), we can see slight variations in the number of visitors each month. This is quite understandable.
- December 2016 appears to be the most popular months for restaurant visit.
- The months of March 2016, July 2016 and December 2016 showed a sudden hike in the number of visitors.
- Number of visitors in the first quarter of 2017 is far more than that of 2016. Hence people are shifting towards AIR restaurants.

Fig 3.2.2 Numbers Of Visitors In AIR Restaurants During Each Weekday, For The Year 2016 & 2017 (Quarter 1). Top Line is for 2016.

Inferences:

- As clear from the graph the number of visitors is maximum for Fridays and Saturdays. The count is minimum for Mondays and Tuesdays. Both of these are quite understandable.

3.3 Distribution Of Visitors Across Various Genres

Kindly consult the following workbook for detailed visualization:

Tableau Workbook: VisitorForecasting

Sheet 5 : Distribution Of Visitors Across Various Genres (AIR)

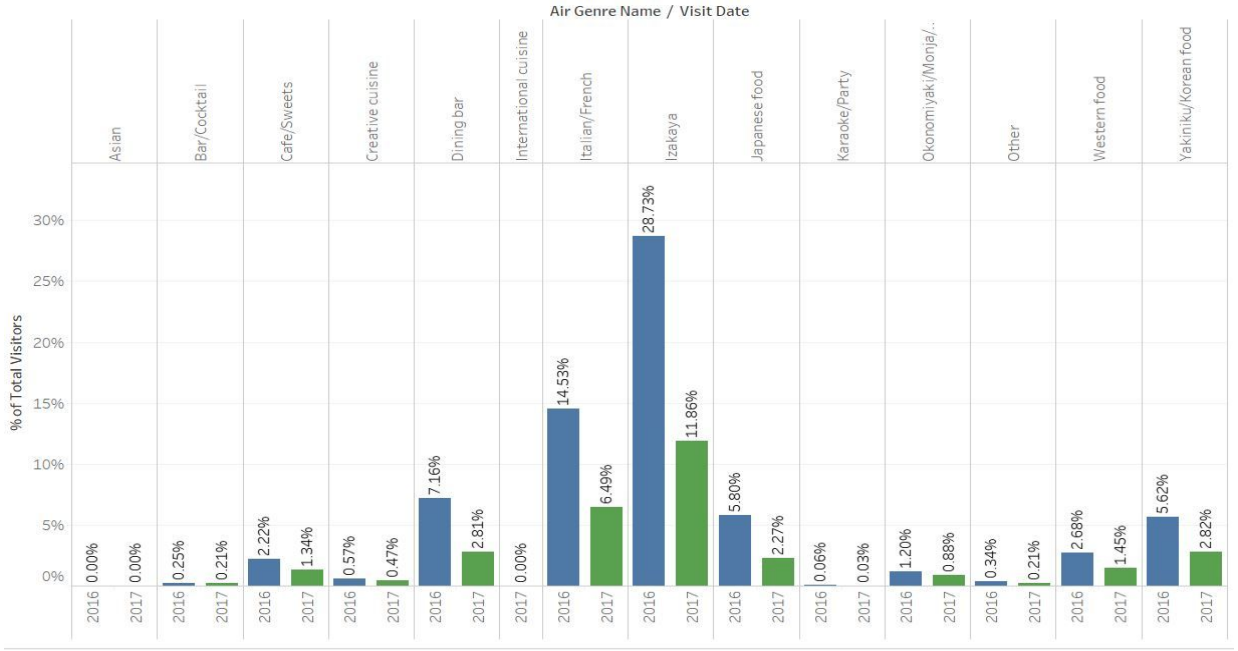


Fig 3.3.1 Numbers Of Visitors In AIR Restaurants Across Various Genres, For The Year 2016 & 2017 (Quarter 1).

Inferences:

- As clear from the graph above, across all the genres available with the AIR restaurants, most of the visitors are visiting Izakaya, a type of informal Japanese pub.

3.4 Distribution Of Total Reservations

Kindly consult the following workbook for detailed visualization:

Tableau Workbook: VisitorForecasting

Sheet 6 : Total Reservations On Given Dates (AIR)

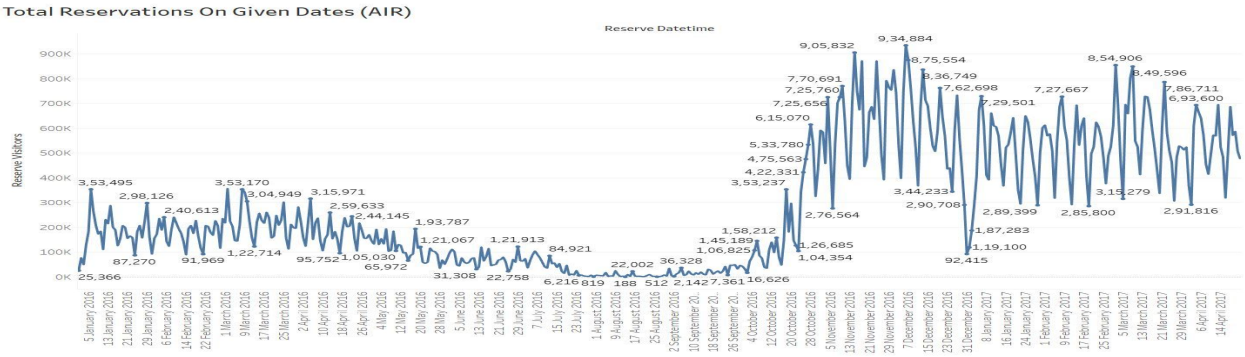


Fig 3.4.1 Total Number Of Reservations Made Each Day For 2016 and 2017, AIR Restaurants.

Inferences:

- The number of visitors reserving online has substantially increased for the year 2017.
- From the last week of July 2016, to the first week of October 2016, the number of reservations reaches minimum for the year 2016.

3.5 Distribution Of Reservation Hours

Kindly consult the following workbook for detailed visualization:

Tableau Workbook: VisitorForecasting

Sheet 7 : Hourly Distribution Of Reservations (AIR)

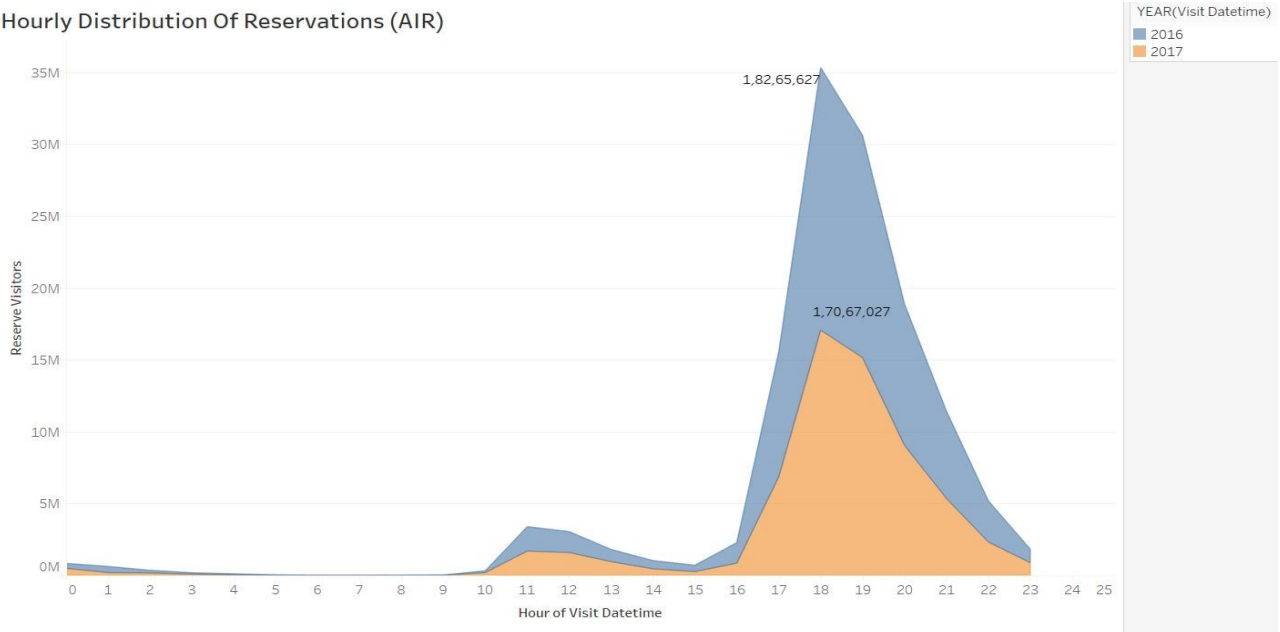


Fig 3.5.1 Number Of Reservations Made For Each Hour Slot

Inferences:

- The most popular time getting reserved is from 4.00 PM to 10.00 PM, that is, most of the people are reserving AIR restaurants for Dinner time.
- The least popular time is Breakfast, 12 NOON - 10.00 AM. Hardly a handful people are reserving tables for these hours.

3.6 Pre-Booking Behaviour For AIR Restaurants

Kindly consult the following workbook for detailed visualization:

Tableau Workbook: VisitorForecasting

Sheet 8 : Pre-Booking Details For AIR Restaurants

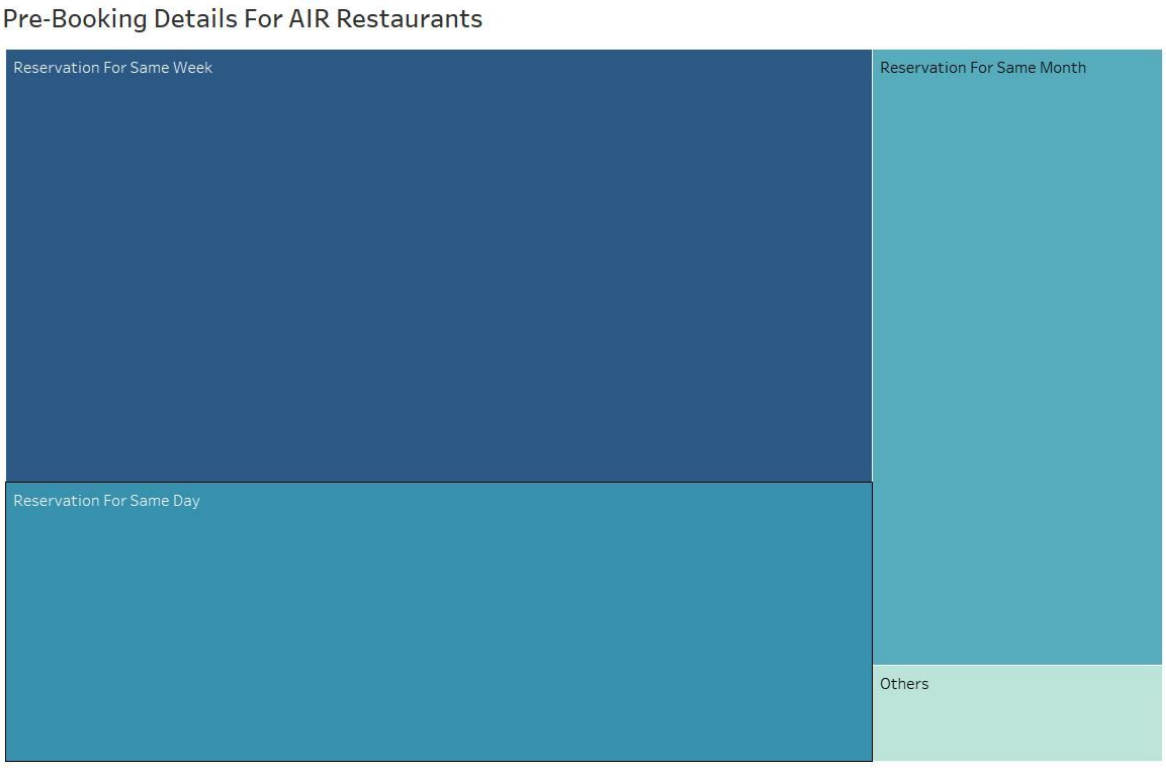


Fig 3.6.1 Number Of Days Before Which People Usually Reserve A Table

Inferences:

- Most people reserve tables for the same week or same day.
- There are few clients who reserve tables several months beforehand.

4. Analysing HPG System's Data

4.1 Spatial Distribution Of Restaurants And Popular Locations.

Kindly consult the following workbook for detailed visualization:

Tableau Workbook: VisitorForecasting

Sheet 9 : Distribution Of Select HPG Stores

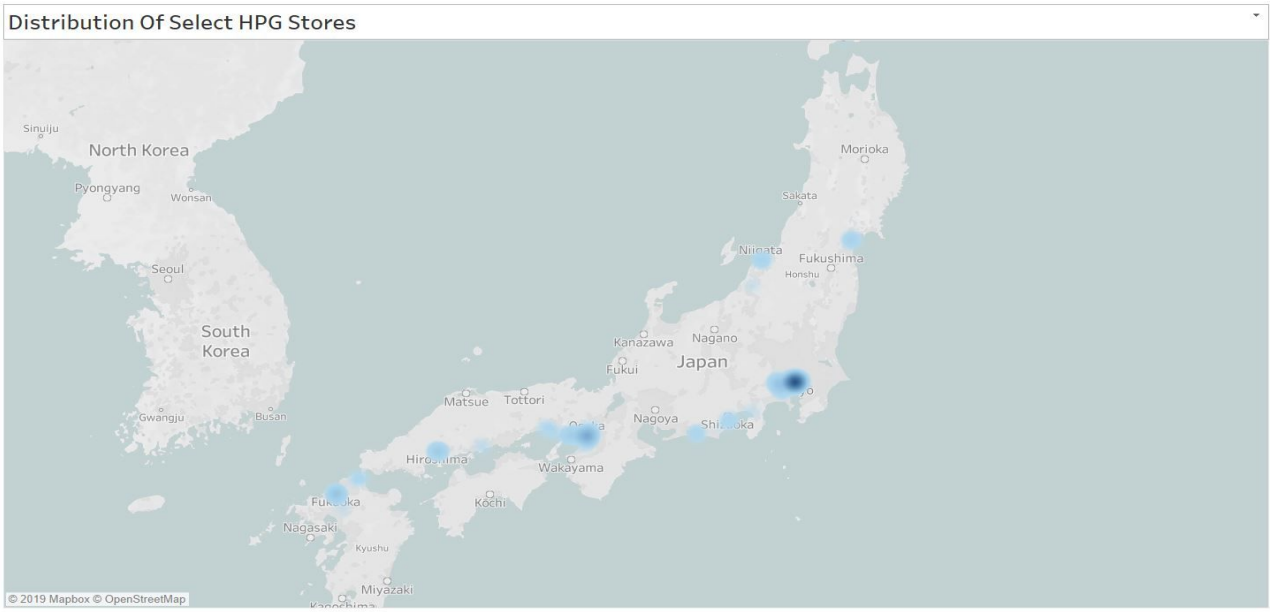


Fig 4.1.1 Spatial Distribution Of HPG Restaurants Across Japan. The darker areas represent higher density.

Inferences:

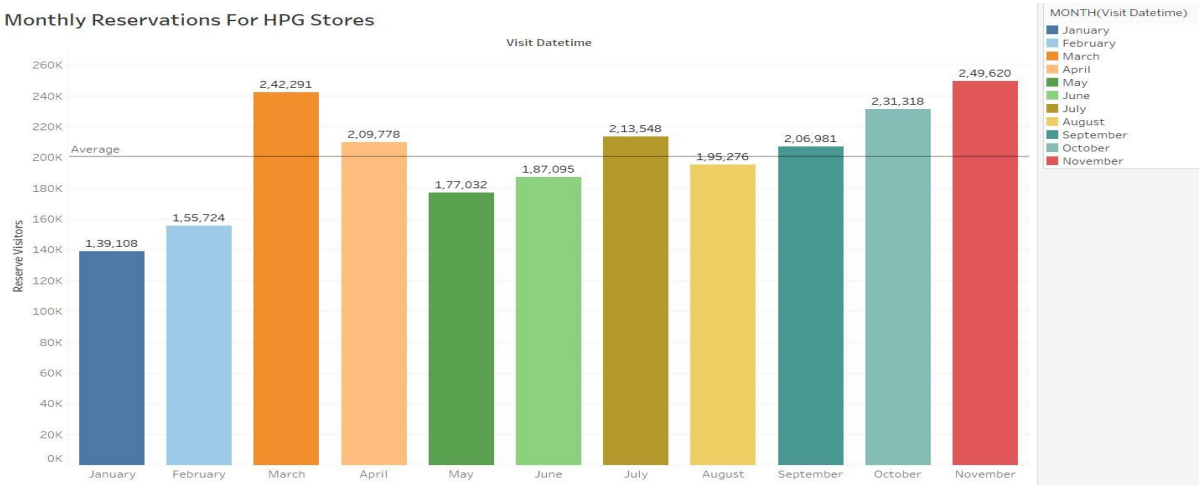
Based on the above figure ,it is quite clear that the density of HPG restaurants is far less than that of AIR restaurants. Moreover unlike AIR restaurants, HPG has higher density in Tokyo region.

4.2 Temporal Distribution Of Visitors.

Kindly consult the following workbook for detailed visualization:

Tableau Workbook: VisitorForecasting

Sheet 10: Monthly Reservations For HPG Stores



Inferences:

- March and November are two of the most profitable months for HPG system.
- January, February, May and August shows below average performance.

4.3 Distribution Of Total Reservations

Kindly consult the following workbook for detailed visualization:

Tableau Workbook: VisitorForecasting

Sheet 11 : Total Reservations On Given Dates (HPG)

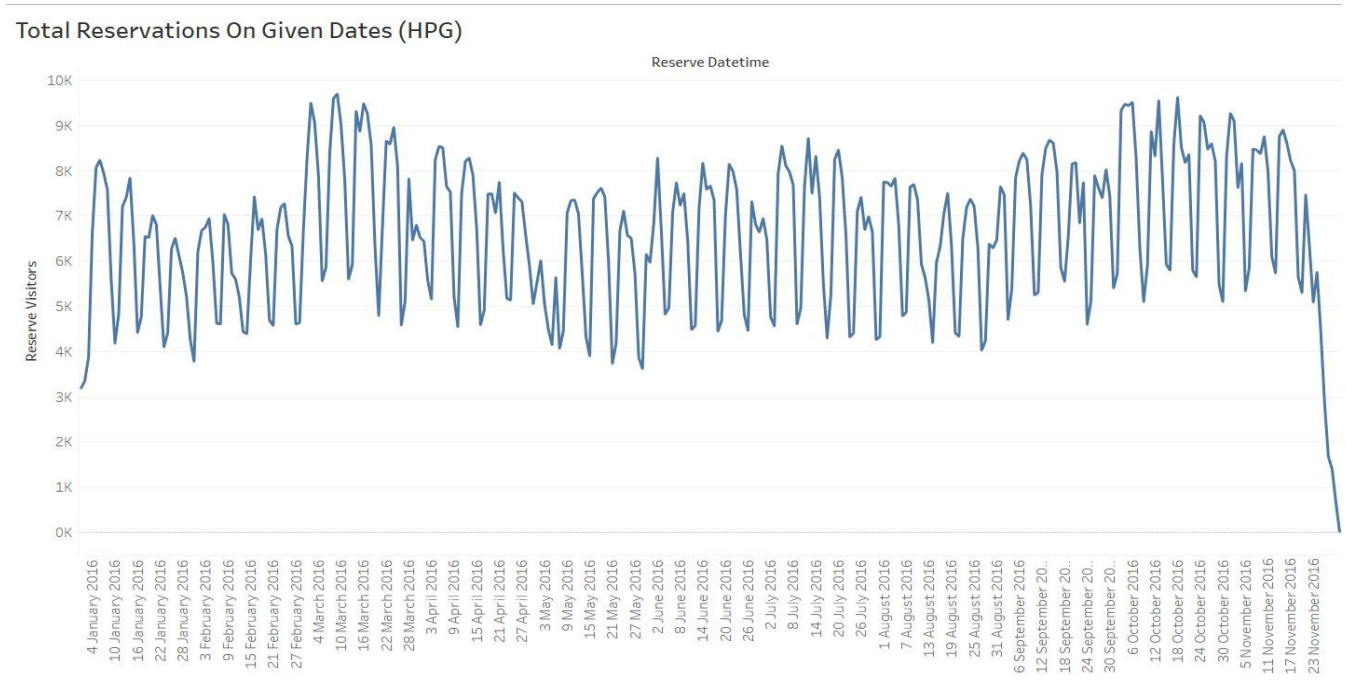


Fig 4.3.1 Total Number Of Reservations Made Each Day For 2016 and 2017, AIR Restaurants.

Inferences:

- AIR seems to be more in use as compared to HPG, for reservations. Though unlike AIR, HPG shows an almost constant pattern for the entire year.

4.4 Distribution Of Visitors Across Various Genres

Kindly consult the following workbook for detailed visualization:

Tableau Workbook: VisitorForecasting

Sheet 12 : Distribution Of Visitors Across Various Genres (AIR)

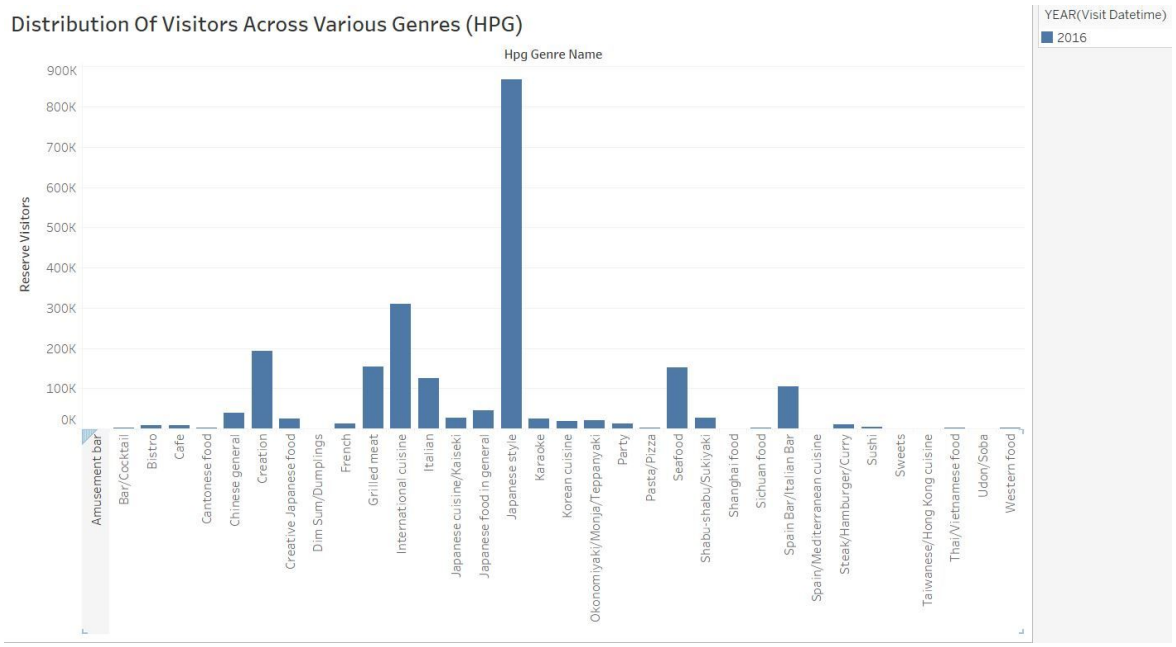


Fig 4.4.1 Numbers Of Visitors In HPG Restaurants Across Various Genres

Inferences:

- As clear from the graph above, across all the genres available with the HPG restaurants, most of the visitors are visiting Japanese Style.

4.5 Distribution Of Reservation Hours

Kindly consult the following workbook for detailed visualization:

Tableau Workbook: VisitorForecasting

Sheet 13 : Hourly Distribution Of Reservations (HPG)

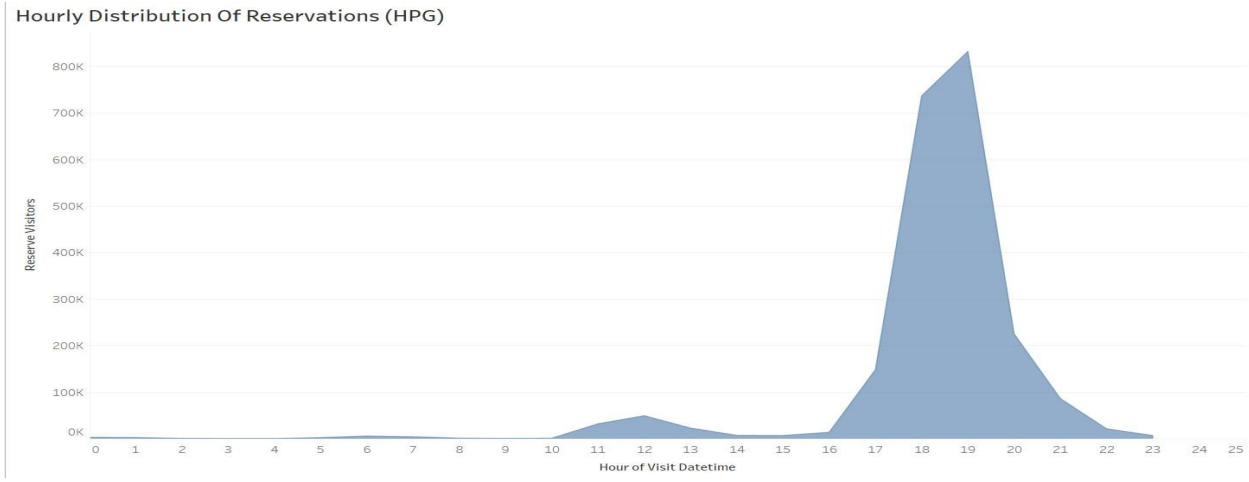


Fig 4.5.1 Number Of Reservations Made For Each Hour Slot

Inferences:

- The most popular time getting reserved is from 4.00 PM to 10.00 PM, that is, most of the people are reserving via HPG restaurants for Dinner time.
- The least popular time is Breakfast, 12 NOON - 10.00 AM. Hardly a handful people are reserving tables for these hours.

4.6 Pre-Booking Behaviour For HPG Restaurants

Kindly consult the following workbook for detailed visualization:

Tableau Workbook: VisitorForecasting

Sheet 14 : Pre-Booking Details For HPG Stores

Fig 4.6.1 Number Of Days Before Which People Usually Reserve A Table

Inferences:

- Most people reserve tables for the same week or month.
- There are few clients who reserve tables several months beforehand.

5. Distribution Of Holidays

5.1 Distribution Of Holidays Across Months

Kindly consult the following workbook for detailed visualization:

Tableau Workbook: VisitorForecasting

Sheet 15 : Distribution Of Holidays During 2016 & 2017

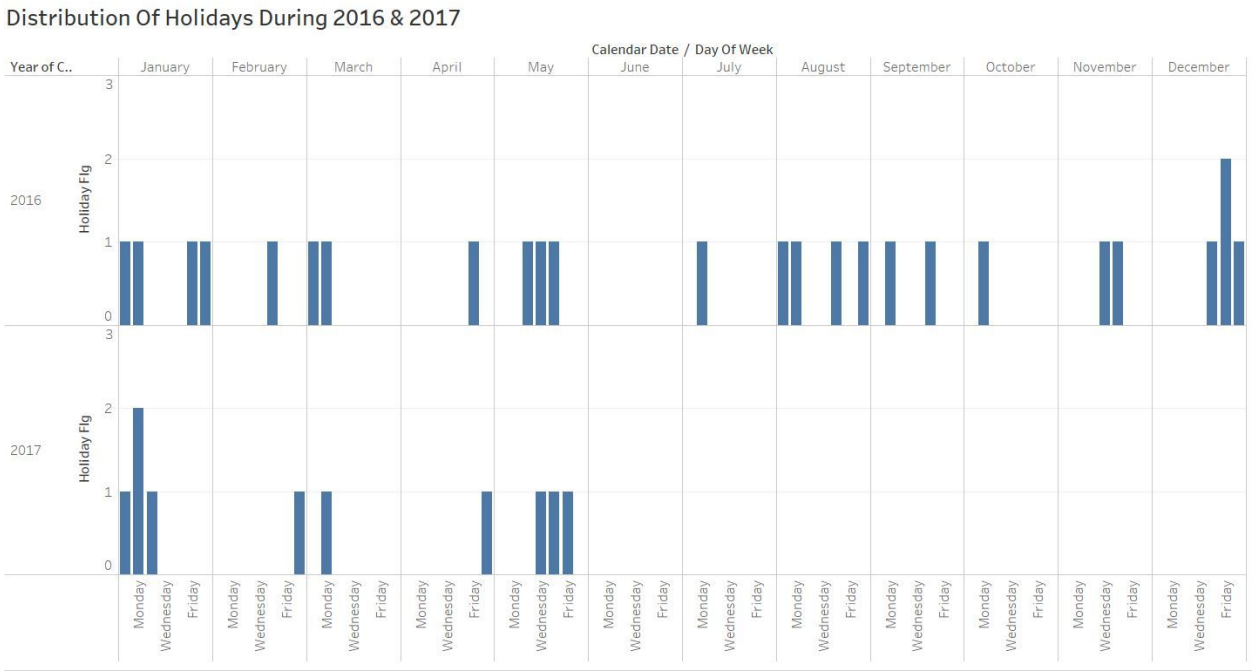


Fig 5.1.1 Distribution Of Holidays

Inferences:

- We are clearly able to see that most of the holidays are in the month of August and usually Mondays, Wednesday and Fridays are holidays.