PROJECT REPORT (MET CS 677)

**DATA ANALYSIS**

**FOR**

**CAFETERIA VISTIOR**

* Shrutika Singodia

***Abstract:***

The aim of this project is to analyze the number of restaurant or cafeteria visitors with different variables. The program uses Python's Matplotlib and Seaborn.

***Input Data:***

The data was collected from Japanese cafeteria. the data set is small and easily accessible without requiring much memory or computing power. 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.

Input files:

1. **air\_visit\_data.csv**: It is data of the number of visitors the AirREGI cafeteria. This is essentially the main training data set.
2. **air\_reserve.csv** / **hpg\_reserve.csv**: It is the data of the reservations made by the customers through the AirREGI / Hot Pepper Gourmet systems.
3. **air\_store\_info.csv** / **hpg\_store\_info.csv**: This data set contains the details (like genre and location) about the AirREGI / Hot Pepper Gourmet restaurant.
4. **store\_id\_relation.csv**: It is for connecting the AirREGI and Hot Pepper Gourmet restaurants ids.
5. **date\_info.csv**: This data helps in flagging the Japanese holidays.

***Data Analysis:***

*AIRREGI VISITS:* In this, the total number of visitors per day over the time range together with the median visitors per day of the week and month of the year are plotted:

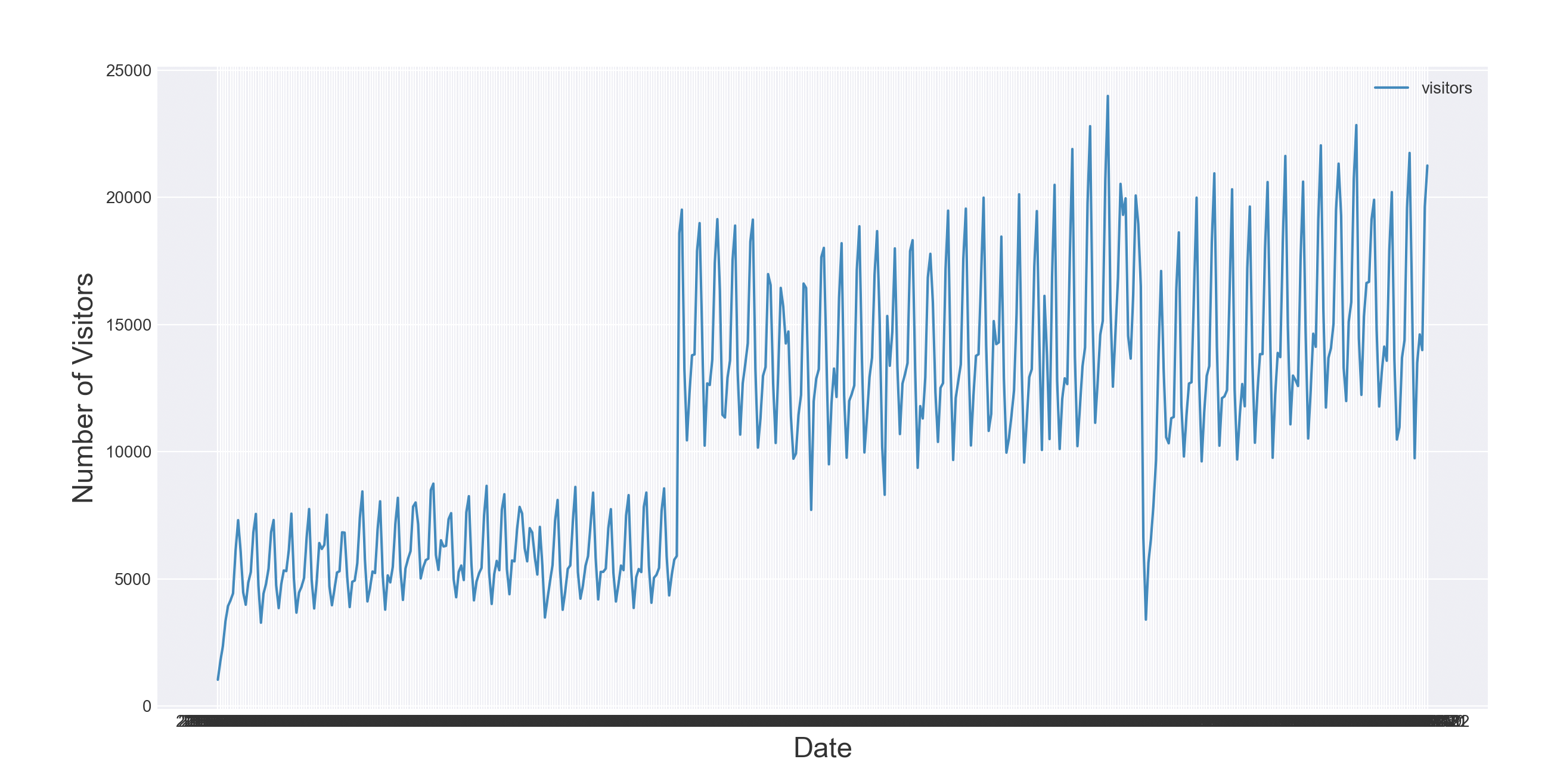


Fig 1

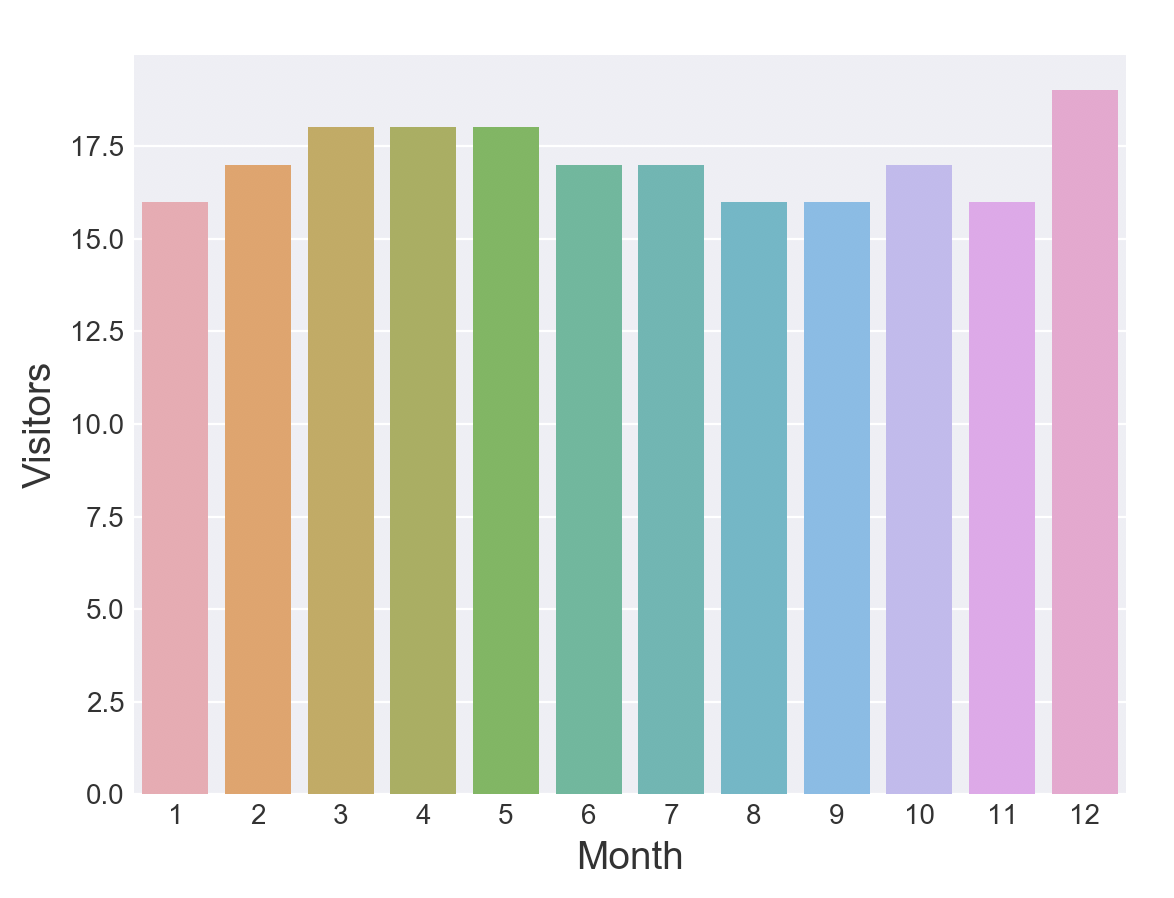
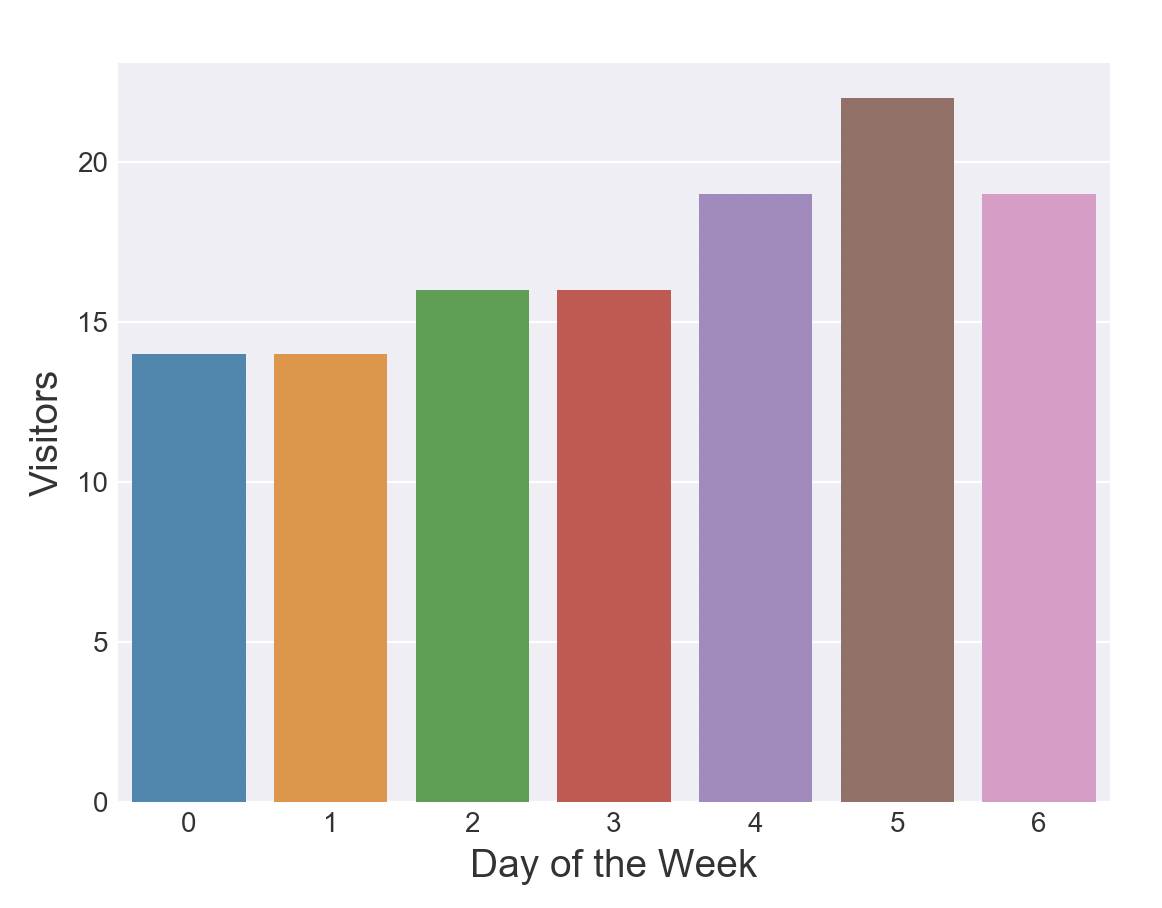


Fig 2

Fig 3

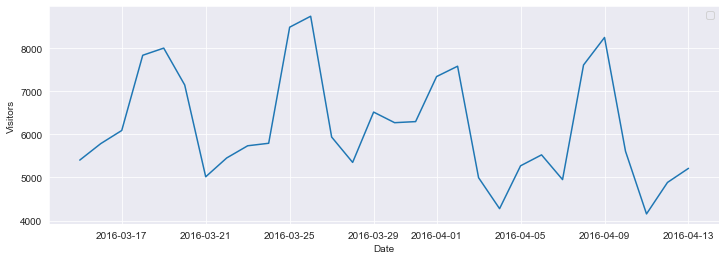


Fig 4

Observation:

1. There is an interesting long-term step structure in the overall time series. Which might be due to the fact that new restaurants were being added to the data base. In addition, we already see a periodic pattern that corresponds to a weekly cycle. (using fig 1 and fig 4)
2. Friday, Saturday and Sunday (weekend) appear to be the most popular days, while Monday and Tuesday have the lowest numbers of average visitors over the training time span of the data. (using fig 2)
3. December appears to be the most popular month for restaurant visits. While, March, April and May appear to be consistently busy months. (using fig 3)

*AIRREGI RESERVATIONS:*  It shows how reservations data compares to the actual visitor numbers. This is done by using the AirREGI restaurants and visualizing their visitor volume through reservations for each day, along with the time duration of these visits. In addition to this, hours of these visits and the time between making a reservation and visiting the restaurant was also visualized.

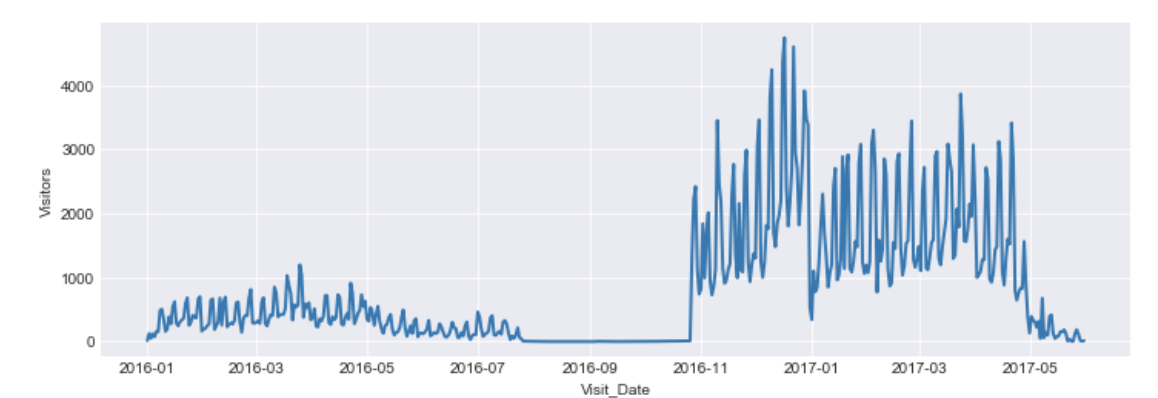


Fig 5

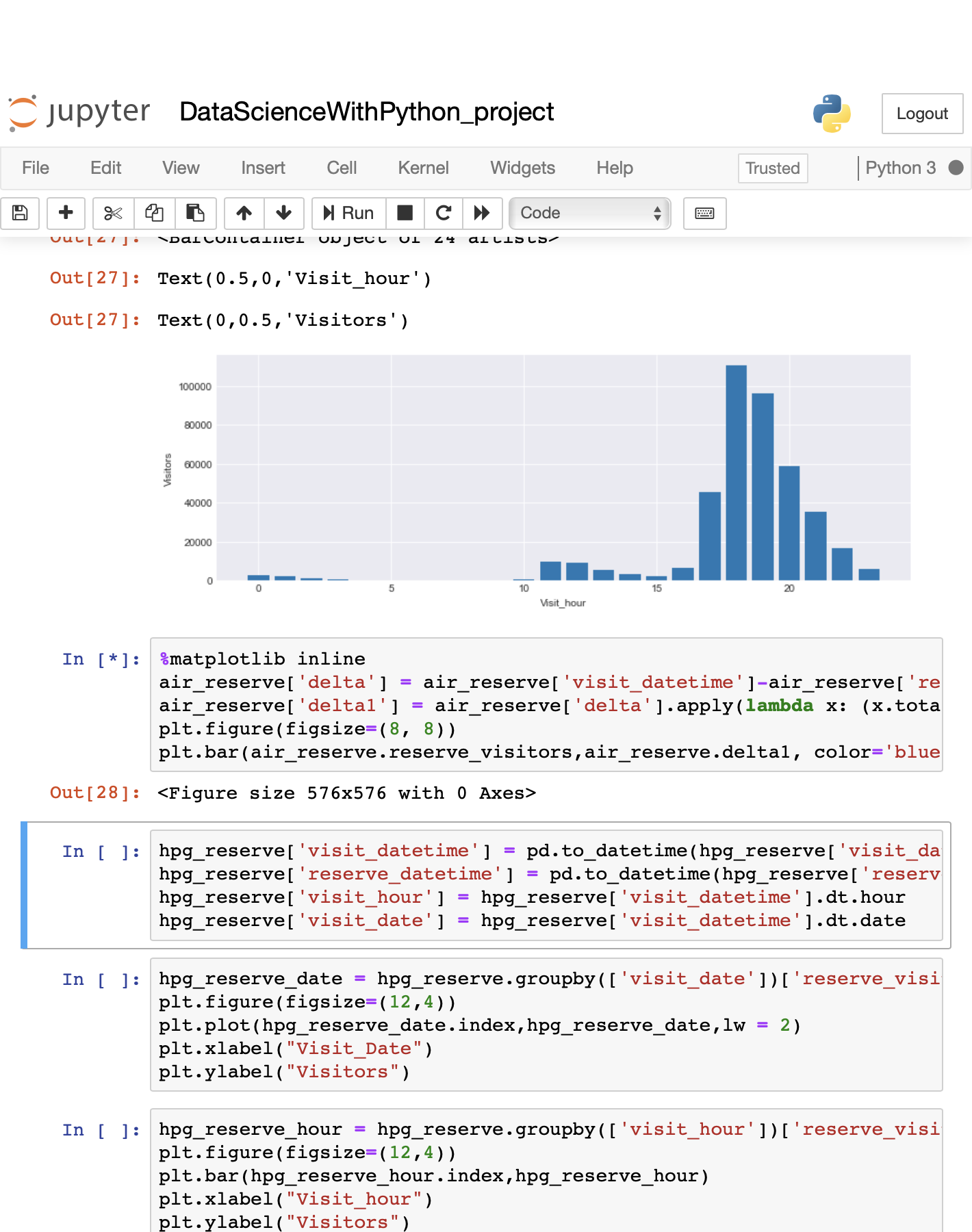


Fig 6

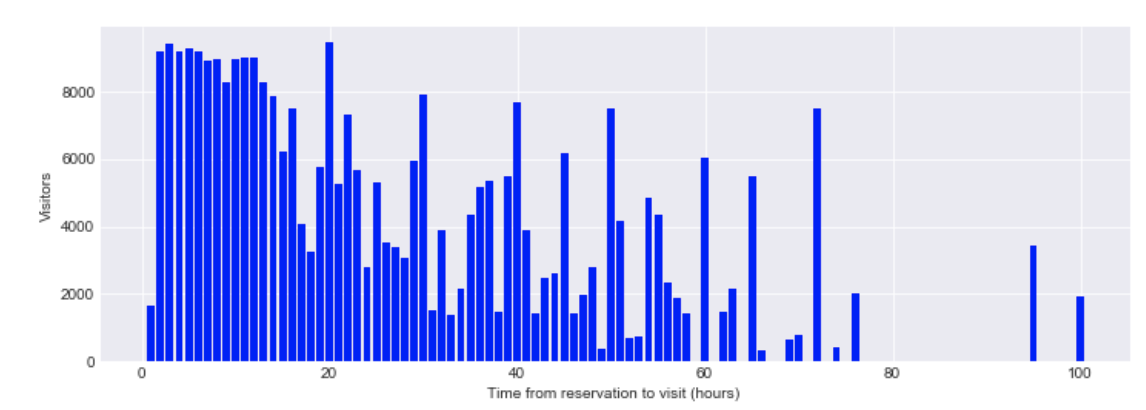


Fig 7

Observation:

* 1. There were much fewer reservations made in 2016 through the *AirREGI* system; even none at all for a long stretch of time. The volume increased during the end of that year. The decline can be seen after the first quarter which is most likely related to these reservations being at the end of the *training* time frame, which means that long-term reservations would not be part of this data set. (using fig 5)
  2. Reservations are concentrated during the evening hours (i.e., for dinner). (using fig 6)
  3. The most popular strategy is to reserve a couple of hours before the visit, but if the reservation is made more in advance then it seems to be common to book a table in the evening for one of the next evenings. Very longtime gaps between reservation and visit are not very common.  (using fig 7)

*HOT PEPPER GOURMET RESERVATIONS:* It shows how reservations data compares to the actual visitor numbers. This is done by using the *Hot Pepper Gourmet* restaurants and visualizing their visitor volume through reservations for each day, along with the time duration of these visits. In addition to this, hours of these visits and the time between making a reservation and visiting the restaurant was also visualized.

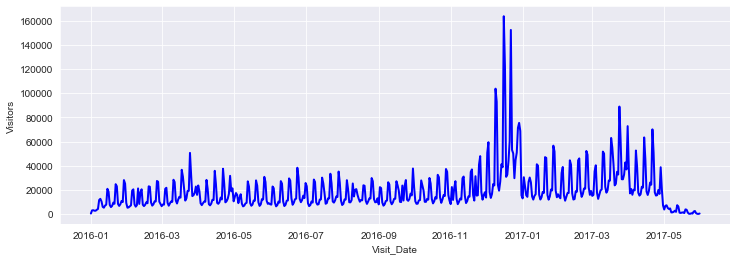


Fig 8

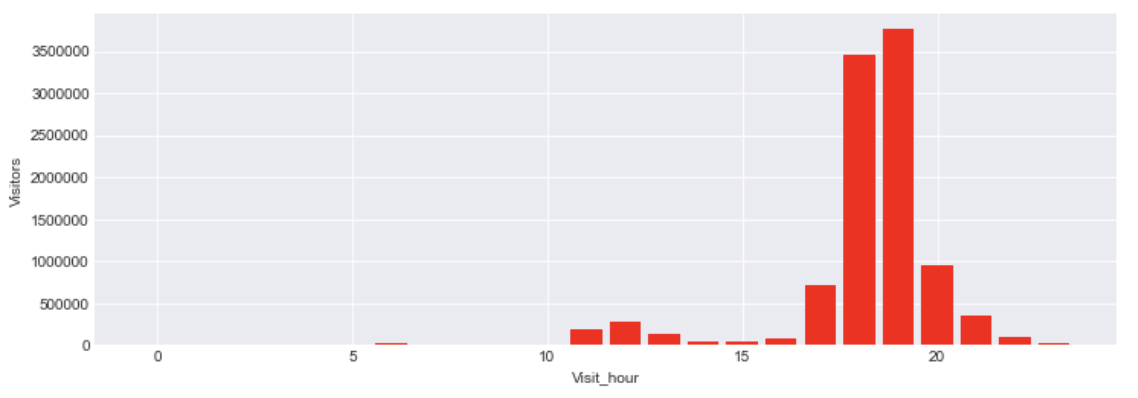
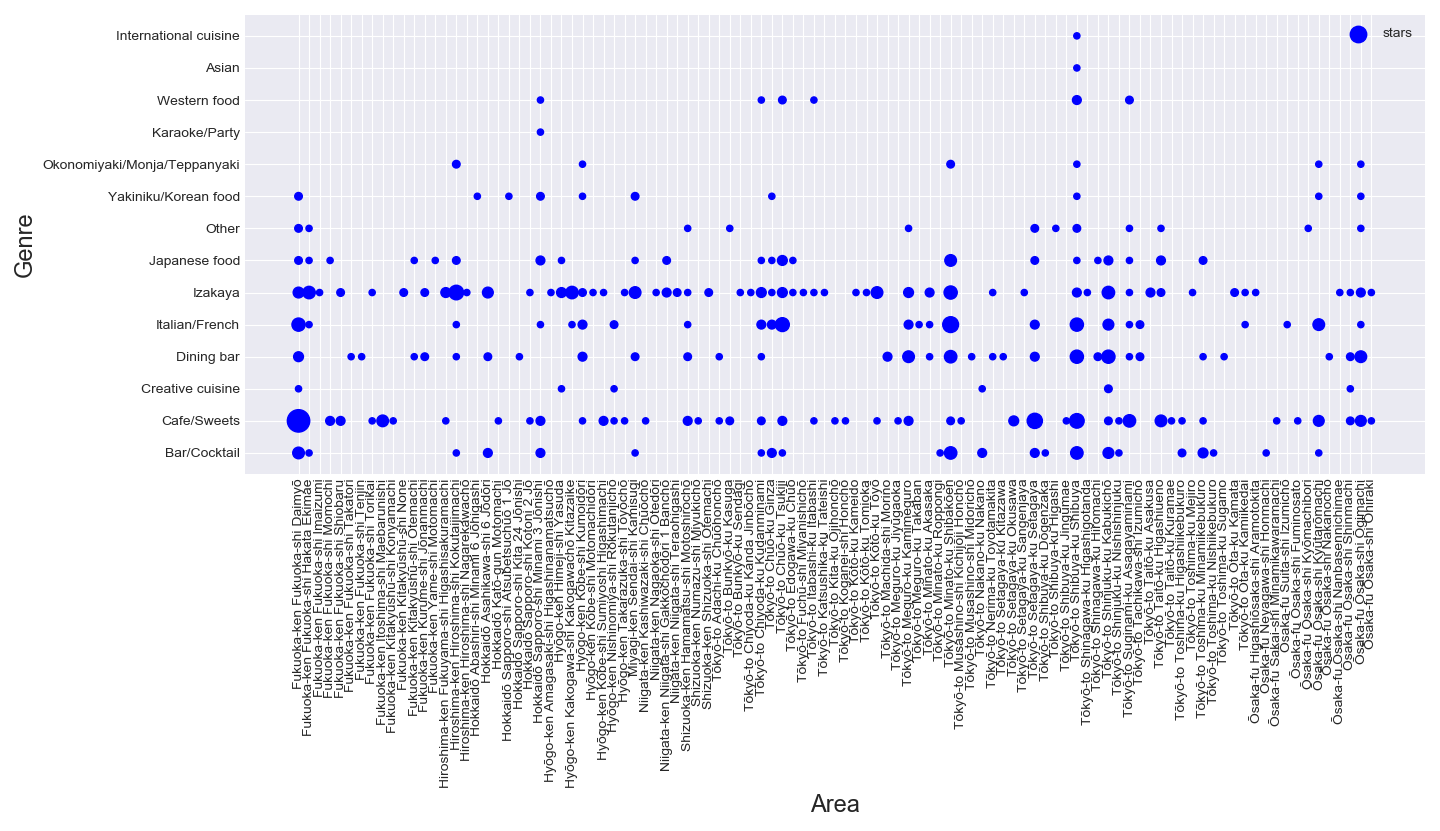


Fig 9

Observation:

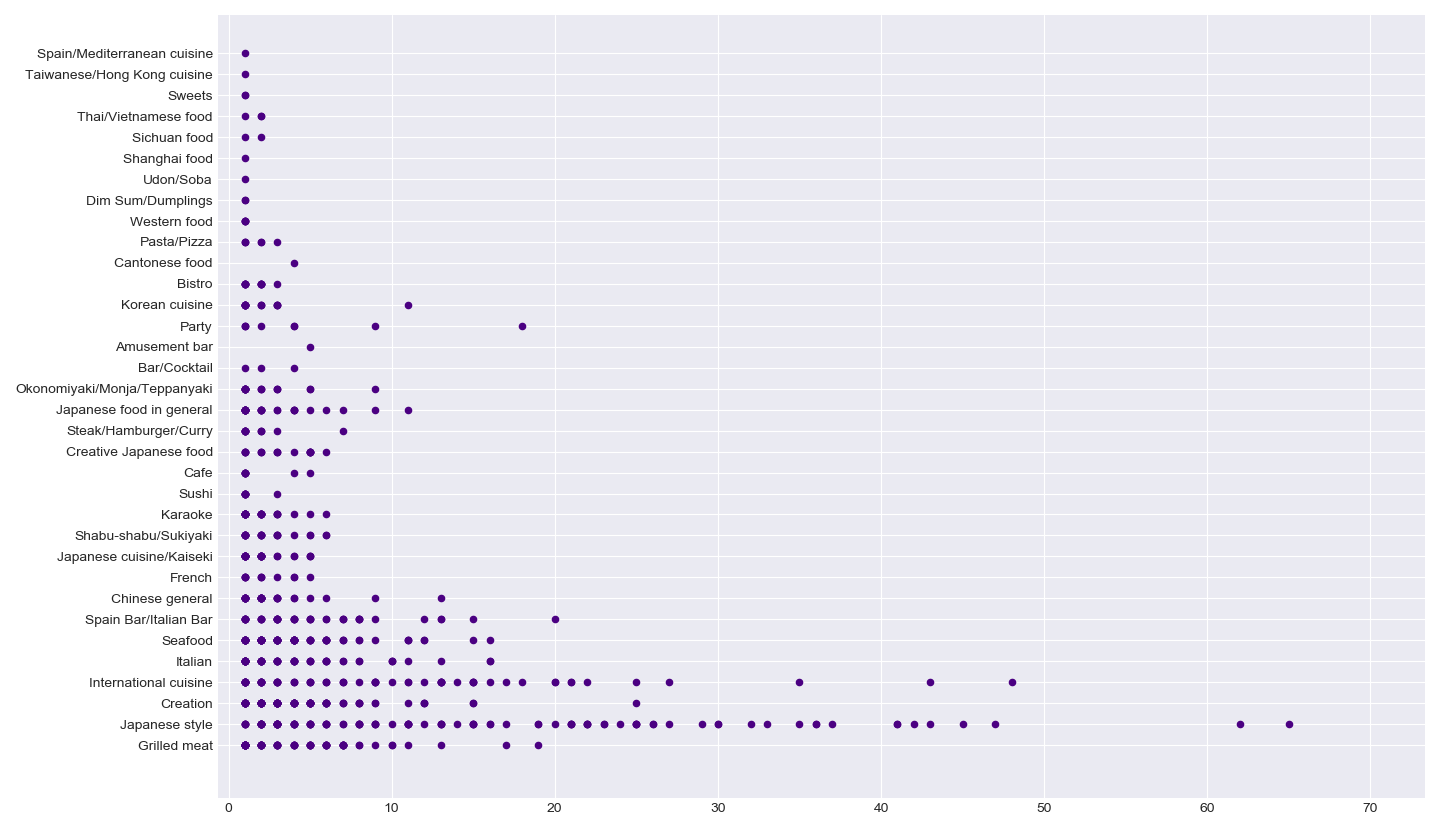
1. Here the visits after reservation follow a more orderly pattern, with a clear spike in Dec 2016. As above for the *AirREGI* data, it can also be seen that reservation visits drop off, towards the end of the time frame. (using fig 8)
2. Similar to the AirREGI reservation, here also most reservations are for dinner. (using fig 9)

*AIRREGI GENRE*: Below is the overview plot of the frequency of certain genres per area for the AirREGI data. The following *plot* show which *genres* exist in which *restaurant*, and the size of the data represent the frequency (variety) of a particular genre in a specific restaurant.

******

It can be seen that some restaurants have a lot of varieties in a particular cuisine while some only have a few varieties.

*FREQUENCY OF A CUISINE (USING HOT PEPPER GOURMET DATA):* Below is the plot of the frequency of each cuisine (genre), according to the data by Hot Pepper Gourmet . It can be observed that, the frequency of most genre is less than 10. Only a few genres like international cuisine, or Japanese style are in present in huge numbers.



**REFERENCES**

1. Kaggle: Your Home for Data Science. (2020). Retrieved from https://www.kaggle.com
2. Stack Overflow - Where Developers Learn, Share, & Build Careers. (2020). Retrieved from <https://stackoverflow.com>
3. Specifying Colors — Matplotlib 2.0.2 documentation. (2020). Retrieved from https://matplotlib.org/users/colors.html