STUDENT INTERNSHIP PROGRAM (SIP) REPORT

Submitted By,

Samruddhi Yadav
T184103
0120160126

SCHOOLOFCOMPUTER ENGINEERING AND TECHNOLOGY

MIT Academy of Engineering
(An Autonomous Institute affiliated to Savitribai Phule Pune University)

CERTIFICATE

This is to certify that the "Student Internship Program (SIP)" report submitted by Samruddhi Yadav PRN 0120160126 is work done by her and is submitted during the academic year 2018 – 2019.

FacultyMentor
(Mrs. Mayura Kulkarni)

School - InternshipCoordinator
(Mr. Roudragouda Patil)

Coordinator–SIP
(Mr. Senthil Kumar)

SchoolDean
(Mrs. Ranjana Badre)

MIT Academy of Engineering
(An Autonomous Institute affiliated to Savitribai Phule Pune University)

ACKNOWLEDGEMENT

The internship opportunity I had with Fyra Insights was a great chance for learning and professional development. Therefore, I consider myself as a very lucky individual as I was provided with an opportunity to be a part of it. I am also grateful for having a chance to meet so many wonderful people and professionals who led me though this internship period.

Bearing in mind previous I am using this opportunity to express my deepest gratitude and special thanks to the Director of Fyra Insights **Mr.Sopan Shewale** , who instead being extraordinarily busy with his duties, took time out to hear, guide and keep me on the correct path and allowing me to carry out my project at their esteemed organization and extending during the training.

I perceive as this opportunity as a big milestone in my career development. I will strive to use gained skills and knowledge in the best possible way, and I will continue to work on their improvement, in order to attain desired career objectives. Hope to continue cooperation with all of you in the future.

Sincerely,
Samruddhi Yadav.

Place: Pune.
Date: 20th August, 2019

TABLE OF CONTENT

1. INTRODUCTION:

- In service industry, an increasing number of business owners improve the quality of service and reduce cost using big data techniques. Based on historical data of business, predictive models of machine learning can estimate future moves of customers.
- Especially, forecasting the number of future visitors can help restaurant owners make the best operations to maximize the revenue.
- With an accurate visitor forecasting model, restaurant owners can prepare suitable amount of ingredients that exactly satisfy future visitors. In addition, restaurant owners can also schedule a suitable number of staff that can serve these visitors.
- Despite various visitor forecasting techniques for other purposes such as national tourism and hotel demand in the literature, little is known for restaurant owners to estimate the number of future visitors. New customers may outnumber old customers in some restaurants of tourist destinations. Hence, it is necessary to develop a new method that can predict the total number of future visitors to a restaurant on a specific day.
- To bridge this gap, in this project an approach to forecast how many future visitors will go to a restaurant using big data and supervised learning. The approach collects big data involving restaurant information, historical visits and historical reservations. Regular restaurants can easily collect such big data on their own without any complex computing infrastructure (e.g., third-party cloud computing services). From such big data together with time information, we constructed four groups of features correspondingly.

2.INTERNSHIP DISCUSSION:

2.1 PROJECT DETAILS

For efficient and economical operation, restaurant owners need to accurately estimate the number of future customers. In this project, an approach to predict how many future visitors will go to a restaurant using supervised learning. The included data involves restaurant information, historical visits and historical reservations. With features constructed from the data, our approach generates predictions by performing machine learning operations. Evaluation of approach is done using large-scale real world datasets from two restaurant booking websites. The evaluation results show the effectiveness of the approach, as well as useful insights for future work.

2.2 KNOWLEDGE ACQUIRED

Machine learning concepts used:

Keras: Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result with the least possible delay is key to doing good research.Use Keras if you need a deep learning library that:Allows for easy and fast prototyping (through user friendliness, modularity, and extensibility).Supports both convolutional networks and recurrent networks, as well as combinations of the two.Runs seamlessly on CPU and GPU.

Tensorflow: Created by the Google Brain team, TensorFlow is an open source library for numerical computation and large-scale machine learning. TensorFlow bundles together a slew of machine learning and deep learning (aka neural networking) models and algorithms and makes them useful by way of a common metaphor. It uses Python to provide a convenient front-end API for building applications with the framework

Rmsle: Root Mean Squared Logarithmic Error (RMSLE) is a technique to find out the difference between the values predicted by your machine learning model and the actual values. MSE incorporates both the variance and the bias of the predictor. RMSE is the square root of MSE. In case of unbiased estimator, RMSE is just the square root of variance, which is actually Standard Deviation.In case of RMSLE, you take the log of the predictions and actual values. So basically, what changes is the variance that you are measuring. I believe RMSLE is usually used

when you don't want to penalize huge differences in the predicted and the actual values when both predicted and true values are huge numbers.

It is shown that it is feasible to predict how many future visitors will go to a restaurant using big data and supervised learning. To drive our approach, users only need to collect big data, including restaurant information, historical visits and historical reservations. Considering that restaurant owners may not have access to high-performance computers, we show that the prediction can be achieved by a mix of low-computation regressors, rather than relying on computation-intensive methods such as Deep Learning (requiring long time of training and mostly requiring highperformance GPU) and SVM (requiring huge amount of memory). In evaluation, we found that a low training error cannot ensure a low test error for a regressor. Hence, users should not minimize training error by all means during training. A relatively high training error often reflects the robustness of a model. Regarding the strongest indicators of future visitors, we found that time-related features (such as week of year, month of year and day of week) and historical visitor records (such as mean, median, maximal and minimal visitors on a day) are most useful for the prediction. In addition, the unique ID andlocation of restaurant are also important features. This indicates that each individual restaurant follows a unique pattern,even if other factors are similar. By contrast, historical reservations can hardly help the prediction.We quantified the computation time on our computer in evaluation.

2.3 SKILLS LEARNED

There was plenty of knowledge/skills that I gained while working on this project. As this was all the latest technologies and frameworks, I learned a lot about it, reading the documentation, learning it functionalities, getting hands on with simpler examples so as to have a better experience while implementing these technologies on live projects. Talking about the specific points, here are a few things I learned:

- Anaconda
- Jyupiter Notebook
- Tensorflow
- Keras
- RMSLE
- Python3
- Data science concepts

2.4 OBSERVED ATTITUDE & GAINED VALUES

Clearly, there’s no spoon-feeding when it comes to working under someone. Everyone’s busy with their work. But the good part is, if you’re willing to learn, people there are more than happy to help us with our queries and to guide us through.

While development of a project, one has to be open and positive about learning other technologies as well. If you have signed up for something, you cannot expect to be working on the same throughout.

2.5 MOST CHALLENGING TASK

Feasibility and risk analysis was the most challenging task.They are related in many ways .If project risk is great,the feasibility of producing quality software is required.It is one of the system analysis usually involves a thorough assessment of the operational, financial and technical aspects of a proposal.Feasibility study is the test of the system proposal made to identify whether the user needs may be satisfied using the current software and hardware technologies, whether the system will be cost effective from a business point of view and whether it can be developed with the given budgetary constraints.A feasibility study should be relatively cheap and done at the earliest possible time. Depending on the study,the decision is made whether to go ahead with a more detailed analysis.

Challenge on a personal (non-technical) level could be keeping up with others. As time is very precious and others being domain experts, you are expected to keep up to their expectations. There are no chance for any delay. If you're assigned a certain task, you're expected to finish it before the given time. If you delay the process, whole production line gets delayed. Therefore, you'll have to give more than you give normally to keep up with those experts.

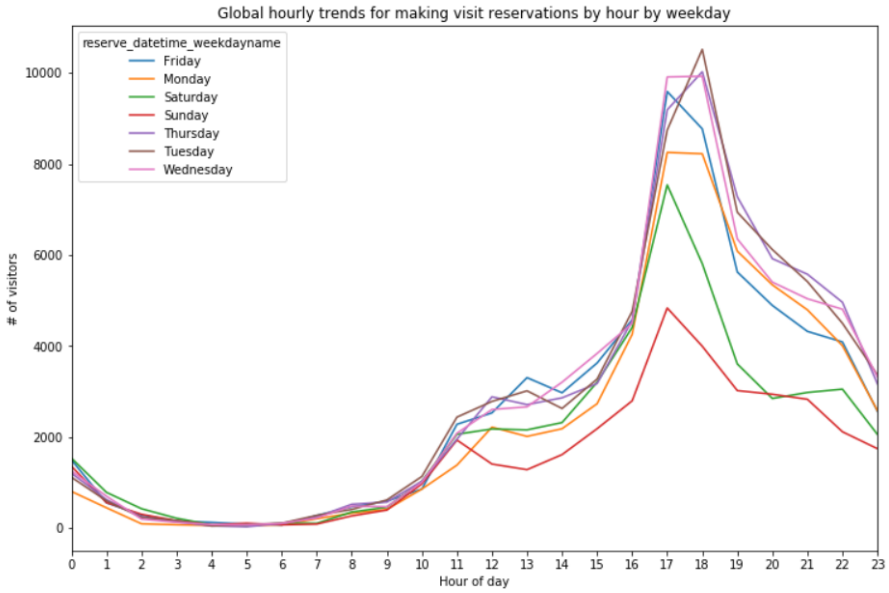
Implementation

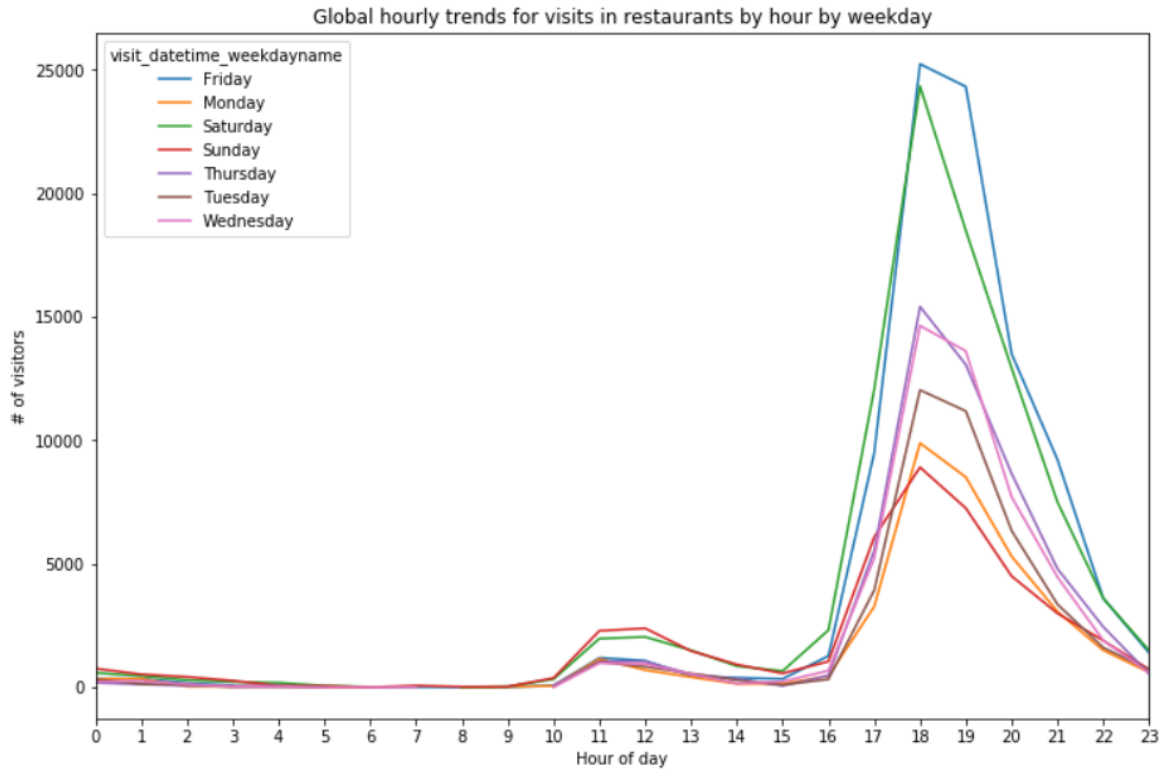
We implemented our approach using Python 3.6. The computer to run the program is a regular HP laptop with Intel Core i5-5300 CPU. Thecomputer has 8GB physical memory. The program does notrequire GPU for computation.

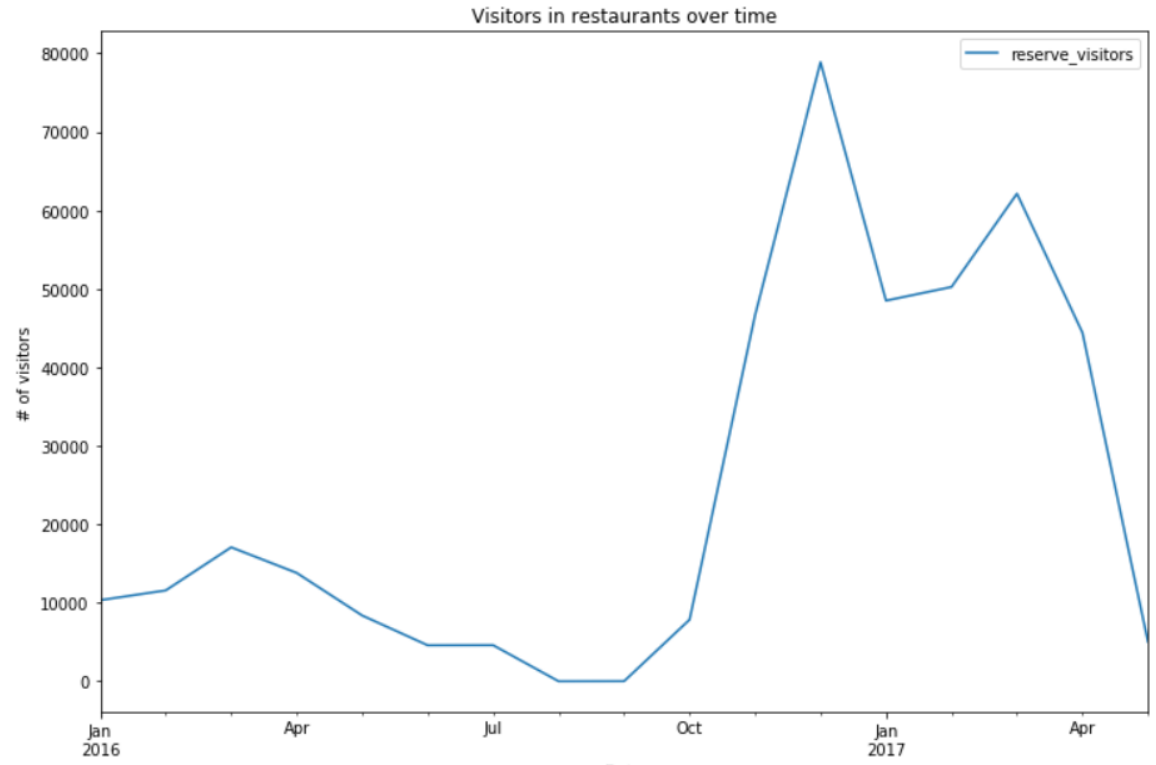
Dataset

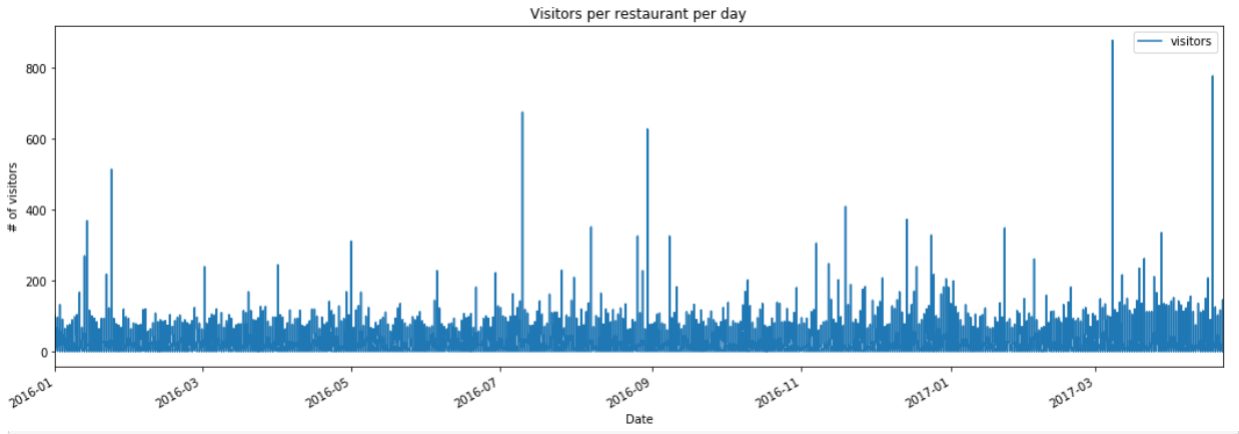
We selected two open large-scale real-world datasets fromtwo restaurant booking websites: HOTPEPPER GOURMET and AirREGI. The two datasets contain the visitorand reservation history of 150 restaurants from year 2016 tothe first season of 2017. We combined the two datasets to forma training dataset for our evaluation (2350992 instances in total). The test set, containing 2019 instances, records how manyvisitors go to these restaurants in the two following months inyear 2017.

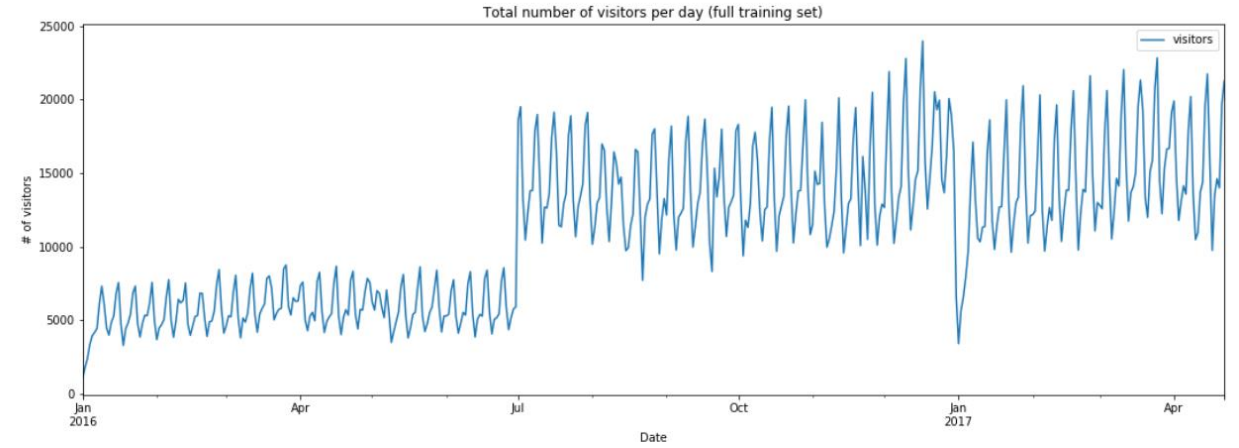
Model


Global hourly trends for making visit reservations by hour by weekday
reserve_datetime_weekdayname
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
# of visitors
Hour of day

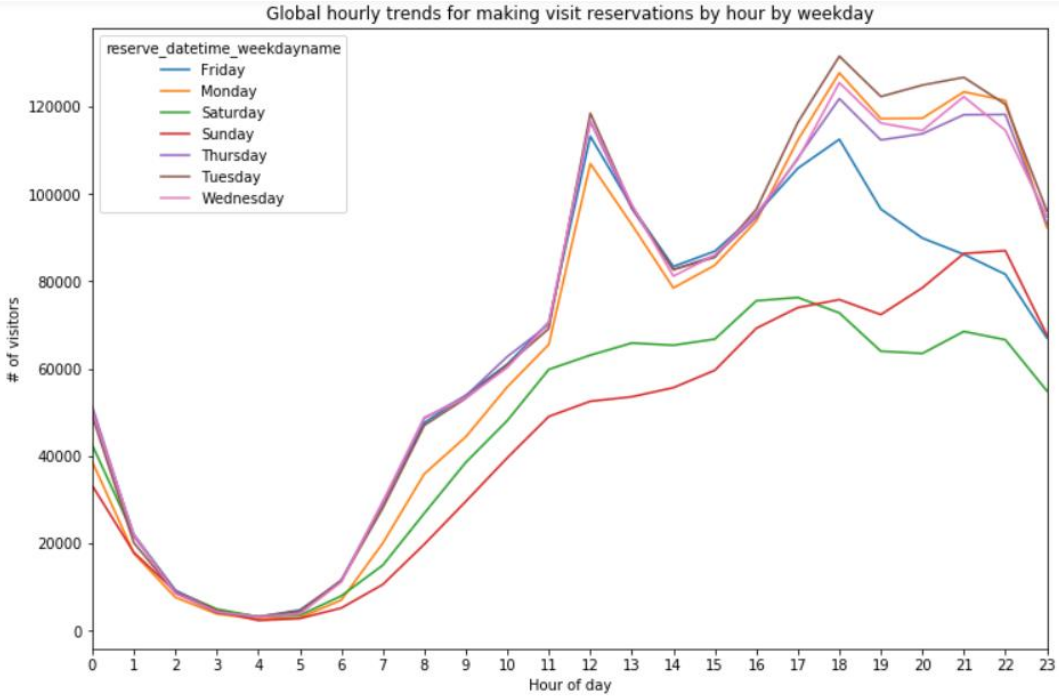

Global hourly trends for visits in restaurants by hour by weekday
visit_datetime_weekdayname
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
# of visitors
Hour of day

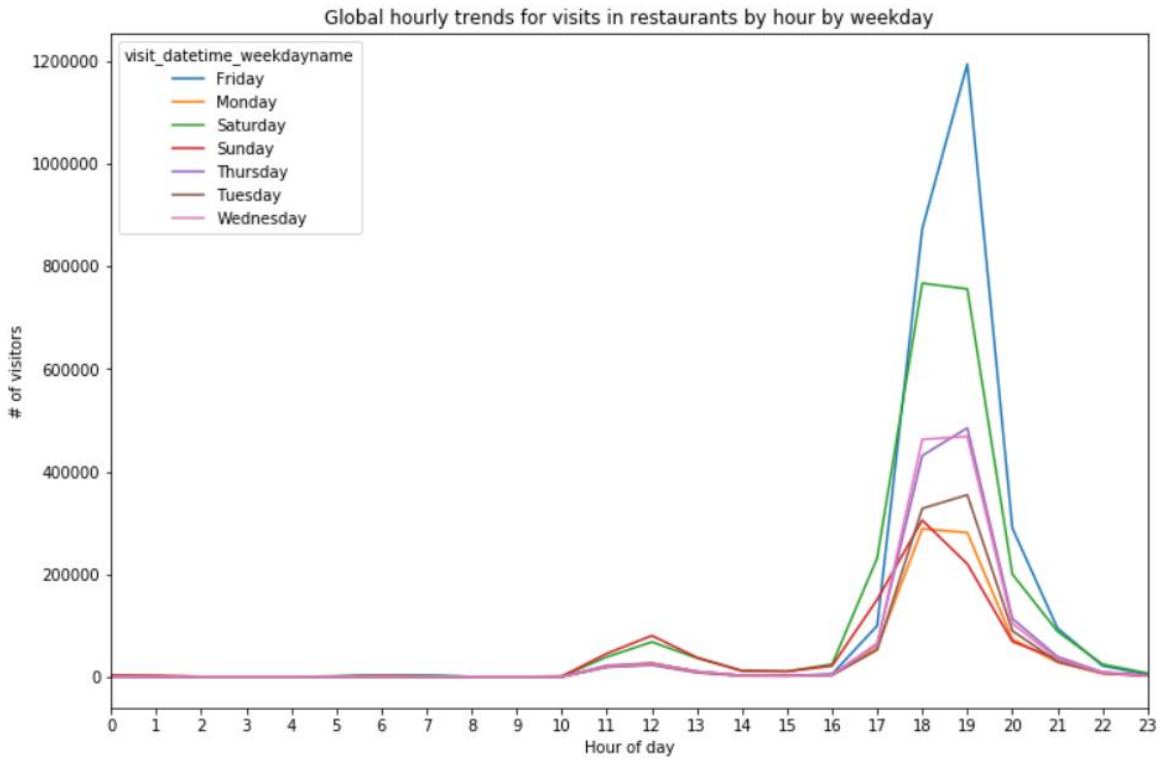

Visitors in restaurants over time
reserve_visitors
# of visitors
80000
70000
60000
50000
40000
30000
20000
10000
0
Jan
2016
Apr
Jul
Oct
Jan
2017
Apr
Date

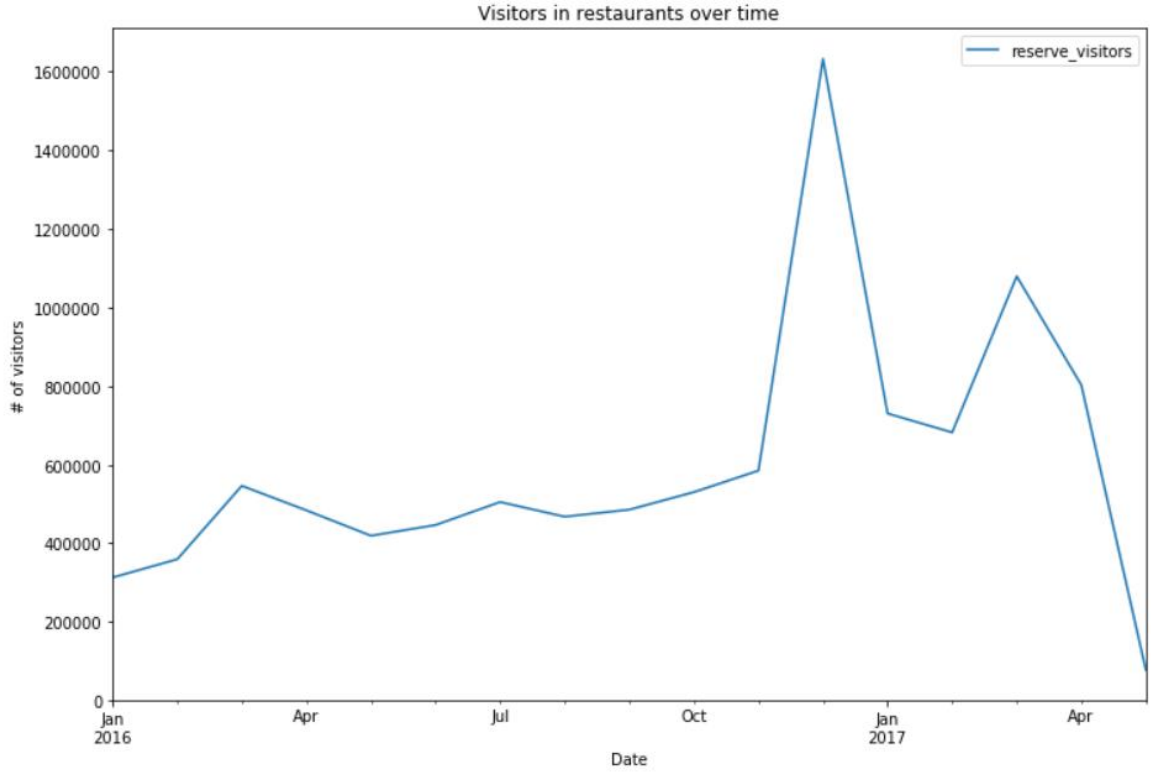

Visitors per restaurant per day
visitors
# of visitors
800
600
400
200
0
2016-01
2016-03
2016-05
2016-07
2016-09
2016-11
2017-01
2017-03
Date


Total number of visitors per day (full training set)
visitors
# of visitors
25000
20000
15000
10000
5000
0
Jan
2016
Apr
Jul
Oct
Jan
2017
Apr
Date


Sum of Visitors in all restaurants per month
visitors
# of visitors
500000
450000
400000
350000
300000
250000
200000
150000
Jan
2016
Apr
Jul
Oct
Jan
2017
Apr
Date


Global hourly trends for making visit reservations by hour by weekday
reserve_datetime_weekdayname
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
# of visitors
120000
100000
80000
60000
40000
20000
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour of day


Global hourly trends for visits in restaurants by hour by weekday
visit_datetime_weekdayname
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
# of visitors
Hour of day


Visitors in restaurants over time
reserve_visitors
# of visitors
Jan
2016
Apr
Jul
Oct
Jan
2017
Apr
Date

CONCLUSION:

I can honestly say that the time I spent at Fyra Insights was one of the best summers for me. Not only did I gain practical skills, but I also had the opportunity to meet many fantastic people. The atmosphere at the office was always welcoming which made me feel right at home. Additionally, I felt like I was able to contribute to the company by assisting and working on projects throughout the summer. Coming into this position, I felt that I had no idea where my career was going, and I lacked confidence about what I could do and what I am really good at. My internship has definitely given me a better understanding of my skill set and where my career may take me. This being my first position in an office atmosphere, I didn't know exactly what to expect. The environment here at Experience is quite relaxed, yet it taught me how to behave in the workplace. Simply working in the office and getting used to everything here has definitely prepared me for whatever my next position may be. Just observing the everyday events has taught me more about teamwork, and how people can come together to get things done. Although sometimes I have to remind myself to use my inside voice, I feel I've adapted to the office life relatively well. this internship has improved my skills a ton, both off paper and on paper. I didn't realize it all of this time, but this position served not only as a positive learning experience, but a resume builder as well. I came into this with a resume that was basically naked, now I am leaving and I have lots of updating to do. My resume doesn't need a makeover, it needs to be restarted from scratch, and that's a good thing! I underestimated how much work I did that actually translates to my resume.

And about the project, beyond features considered by our approach, many more factors can facilitate accurate prediction of future visitors to a restaurant. For instance, bad weather of the restaurant location may reduce visitors. Also, if a new restaurant is opened next to an existing restaurant, the number of future visitors going to this existing restaurant may drop. Moreover, social events can bring more visitors to restaurants of the related venues. Hence, in future work, it is necessary to include more information to the predictive model, such as weather, competitors and social events. For some kinds of restaurants, it is challenging to accurately count how many visitors have come on a day. For example, one customer may buy food for all its friends at fast food restaurants, such as McDonald's (especially when customers in a car buy food via "Drive-Thru"). Future work can also explore new customer counting methods for restaurants based on new technology (such as combining video camera and the Internet of Things, or new software/hardware.

4.BIBLIOGRAPHY:

Most of my resources constituted of online documentations of respective technologies. Links to which are provided below:

- https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting
- https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting
- http://dataanalyticsedge.com/case-studies/restaurant-visitor-forecasting/
- https://www.researchgate.net/publication/328903381_Predicting_Future_Visitors_Of_Restaurants_Using_Big_Data