

Contribution title

Julian Ruppel, Johannes Weber, Tobias Rasbold, Hannes Quade und Joshua Braun

¹ Princeton University, Princeton NJ 08544, USA

² Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany lncs@springer.com
<http://www.springer.com/gp/computer-science/lncs>

³ ABC Institute, Rupert-Karls-University Heidelberg, Heidelberg, Germany
{abc,lncs}@uni-heidelberg.de

Zusammenfassung. By that time many people have the ambition to be self-employed and want to start a restaurant business. But to run a restaurant can be risky due to lots of different requirements. A well-conceived idea for a restaurant is necessary to be successful and competitive. Investors called SCHULPS want to start a restaurant business with a budget of 650,000 € somewhere in Germany. They do not have a determined idea so they are receptive to all kind of restaurant types. Nearly all criterias like the location, the size, or the menu of the restaurant can be chosen freely. The only requirement which is set is to reach monthly sales of 40,000 €. The purpose of the thesis is to find a good location for opening a restaurant in Germany based on different aspects as well as reaching the mentioned financial specifications of the investors. Therefor it was necessary to collect data from a restaurant review online platform which was yelp in this case. The data included key figures such as restaurant name, food category, rating and location. Furthermore some other key figures like the population, the buying power and the rent average for industrial buildings were collected for all cities of Germany. The aim was to fit a statistical model for the mentioned key figures to find the best cities for starting a restaurant business in Germany. For the financial aspects it was used a benchmark of Germany's catering industry to get reference values for the investors restaurant.

Schlüsselwörter: First keyword · Second keyword · Another keyword.

1 Introduction

Nowadays lots of people want to realize their dream to start an own business - mostly its associated with a own restaurant. But a own restaurant business is not for everyone because you have to deal with a lot of different requirements. First of all, a well-conceived concept is absolutely necessary. Therefor you have to ask yourself several types of questions: What kind of food should you offer? What kind of people should address the restaurant? How should the restaurant look like? Where should it be located? All these factors should be well-matched to have a unique feature as a restaurant and to be competitive. Investors called SCHULPS are holding different restaurants all over the world and now they want to start a restaurant business somewhere in Germany with a budget of 650,000 €. As the existing restaurant types of the investors do not have a determined idea due to the different countries and cultures they are open minded for a random concept in Germany. This implies, that criterias like the location, the size or the menu of the restaurant can be chosen freely. The only requirement which is specified is that the monthly sales should be around 40,000 €. The purpose of the thesis is to create a good and clear concept for a new restaurant business, especially to find a good location for opening a restaurant in Germany based on different facts as well as reaching the mentioned financial specifications of the investors. The first step for answering these questions was to collect some data from a restaurant online platform. A very popular one is obviously yelp. The yelp website offers an API which was connected with python scripts to collect the restaurant data, e.g. name, rating, location and so on. There were some other important key figures that help to find an appropriate location, especially the dataset

with all of Germany's cities and its population as well as the datasets including the buying power and the rent average for industrial buildings. All the collected data were stored in the Google Cloud Platform with the help of python scripts, more precisely in it's Datastore which is a NoSQL document database. From there it was possible to create a postgresSQL instance with the fully-managed relational database service, the so called Google Cloud SQL. For a significant analysis of the raw datasets it was necessary to clean it thoroughly. This was made with the help of python scripts and different database queries. The idea was to fit a statistical model for the mentioned key figures to find out the best cities for starting a restaurant business in Germany. For the financial aspects it was used a benchmark of Germany's catering industry which included lots of different key figures like the number of guests per day or the average consumption per guest as well as the revenues and expenses in detail. to be continued...

2 Literature Synthesis

Vergleichsliteratur zu restaurant daten, yelp

Was wurde gemacht, was ist gut / schlecht gewesen, was Machen wir anders / besser
4-5 Sätze zur Literatur

3 Material and Methods

3.1 Usecase Description

SCHULPS sind Investoren und haben bereits mehrere Restaurants weltweit geöffnet. Standorte gibt es bisher in den USA, in Indien, etc. Dabei sind die Speisekarten aber nicht identisch, sondern passen sich an die jeweiligen Kulturen des Landes an. Nun möchten Sie sich auch in Deutschland festsetzen und erst einmal ein Restaurant eröffnen. Auf einen Standort sind sie dabei nicht festgelegt. Voraussetzung ist, dass das Restaurant einen monatlichen Umsatz von 40.000€ erreicht.

3.2 Data Processing Pipeline

In the following its about the data processing pipeline.

Data sources As mentioned previously, a lot of various key figures were collected for the following data analysis relating to the restaurant business. The main data source with helpful key figures should be at least from one restaurant online review website. In this case the online platform of yelp was chosen. Yelp offers an api for delevopers to collect specific data of Germany's restaurants like name, rating or the exact location with longitudes and lattitudes. In addition to this some other datasets relating to Germany's cities were used for the data analysis:

- CSV file with the population and the area
- CSV file for the buying power
- CSV file with the rent average

Data ingestion For collecting the yelp data we used the yelp API. Therefor we wrote a script with python. In the following you can some code:

Data storage As a storage for the datasets the Google Cloud Platform was used. Before that, it is necessary to configure it first. You have to create a GCP-Project and then put the collected data in the Google Cloud Datastore which is a NoSQL document database. From there, it was possible to create a postgresSQL instance with Cloud SQL, the fully-managed relational database service of Google Cloud. For setting up a connection a proxy or current ip address is required.

Data cleaning The collected datasets in the Datastore were not messy, but there was still a lot of cleaning activities to do. The yelp data itself had some inconsistencies. Especially the columns price_range and review_count had a lot of empty values. The price_range was filled with the average value of €. For the column review_count ...

Moreover the data from yelp obviously did not match exactly with the other datasets. The buying power dataset did not include all the cities, that were listed in yelp. To solve this problem the empty values were filled with the average value of Germany. The same problem occurred with the rent average dataset which was solved the same way.

Data analysis

4 Results

4.1 Analysis

4.2 Visualizations

5 section name

Discussion

15-20 Zeilen

Vergleich zu bisherigen Algorithmen Gab es Ausreißer, die nicht untersucht wurden?

Offene Probleme?

Ausblick

