

Restaurants in Tallinn

An IBM Data Science Specialization Capstone Project

The Task

Client

A friend wants to invest his savings into opening a restaurant in Tallinn

He has some experience with hospitality industry, but he still wants a good glimpse of the market.

Context

There are quite a few established restaurants in Tallinn, so opening a good one would require understanding of local tastes.

In addition to Machine Learning solutions, he will use interviews.

Problem statement

Answering 3 distinct questions:

- Where should restaurant be located?
- What cuisine should be served?
- What other features should it have?



The Data

To acquire the data we parsed an Estonian restaurant review service: www.vabalaud.ee

Additionally, we queried Foursquare API to confirm information that we have.



Restaurant Allee

Tallinn » old Town

cuisine: international



Restaurant with beautiful views in the old town

Atmosphere ★★★★★ 4.6
Food ★★★★★ 4.1
Service ★★★★★ 4.3
Total rated: 10



This restaurant offers you the possibility to make online reservations; you will receive confirmation immediately after making the reservation.

You can check the availability of the time when you choose the date and number of people.

RESERVE A TABLE



Merike Trubetski



46. summer terrace 40



11 – 13



Kanuti 2, Tallinn, Harju County, 10111



Mon 12:00 - 22:00
Tue 12:00 - 22:00
Wed 12:00 - 22:00
Thu 12:00 - 22:00
Fri 12:00 - 22:00

Data

We have collected data on 50 restaurants in Tallin.

It's not a huge city, so that's still quite a lot!

Data has been cleared and has no empty cells.

This data will be used to train all ML algorithms in this project

Column name	Description
Restaurant	Restaurant name is displayed here
Details URL	This is url from which data is acquired
Address	Address of the Restaurant
Cuisine	Comma separated types of cuisine
Avg_Bill	Average bill (float), estimated
Additional	List of properties that the venue has (comma separated)
Atmosphere	Atmosphere rating (1-5), float
Food	Food rating (1-5), float
Service	Service rating (1-5), float
Latitude	Latitude of venue
Longitude	Longitude of venue

Implementation

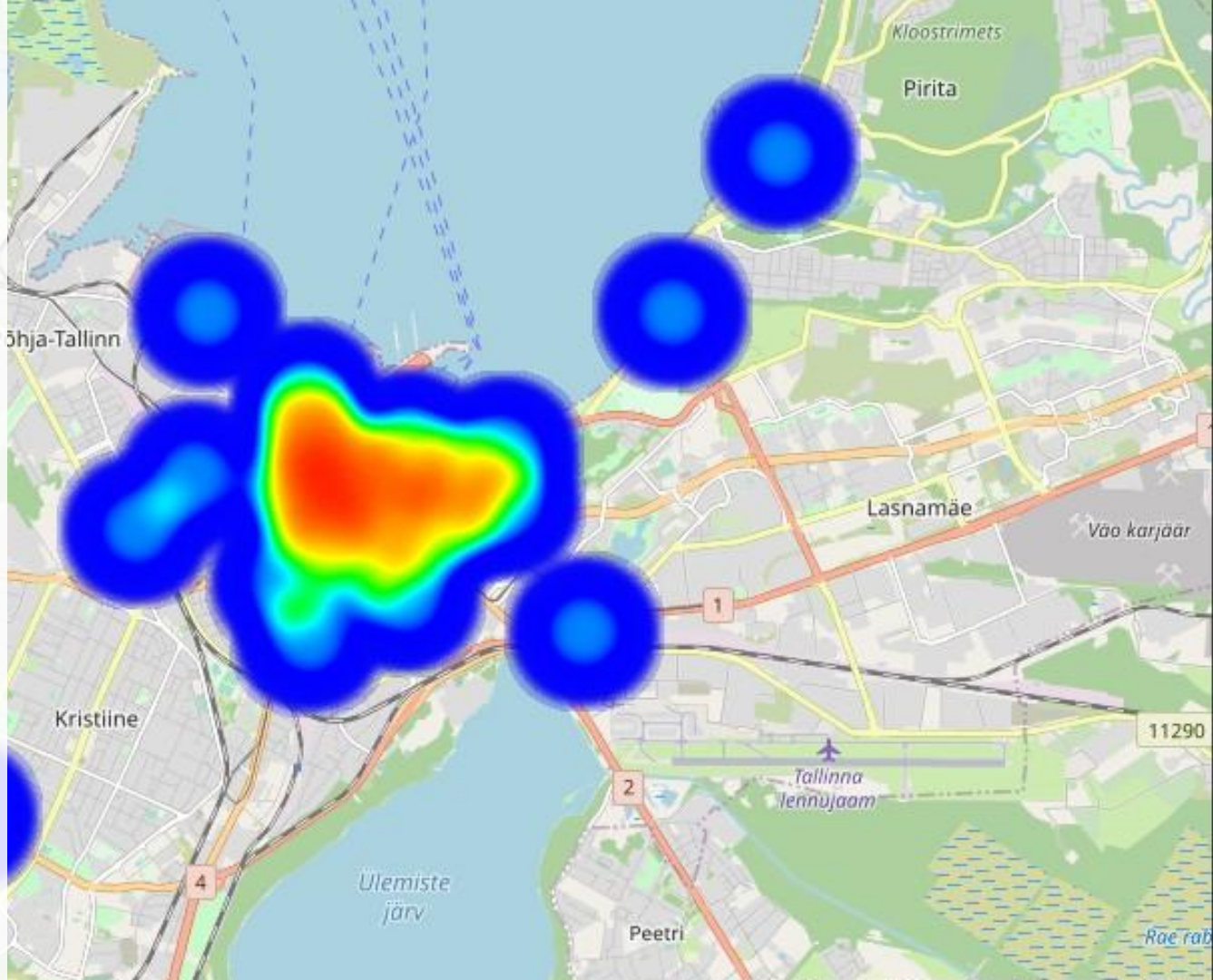


Heatmap

A certain area in the city centre was significantly denser than others.

So we used the heatmap and indeed determined that there is an area in the city center, packed much denser than the rest of the town.

Question was, why?

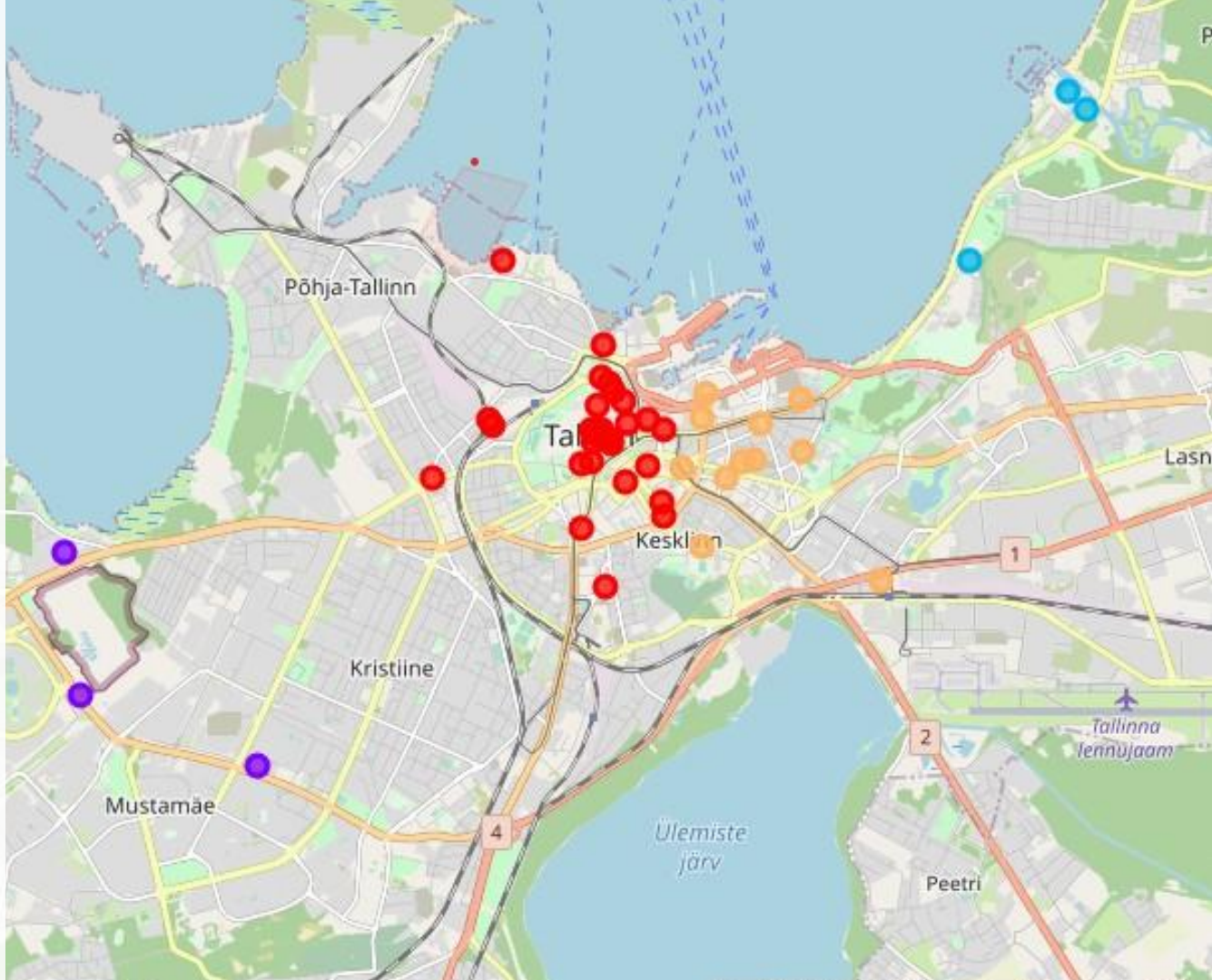


Clustering

K-Means Clustering was used to determine if there are any significant groups.

After looking at the clusters, grouping became obvious:

- Red - Old City and Castle
- Orange - Business Centre
- Blue - Expensive district with a yacht club
- Purple - Decent district of Mustamäe

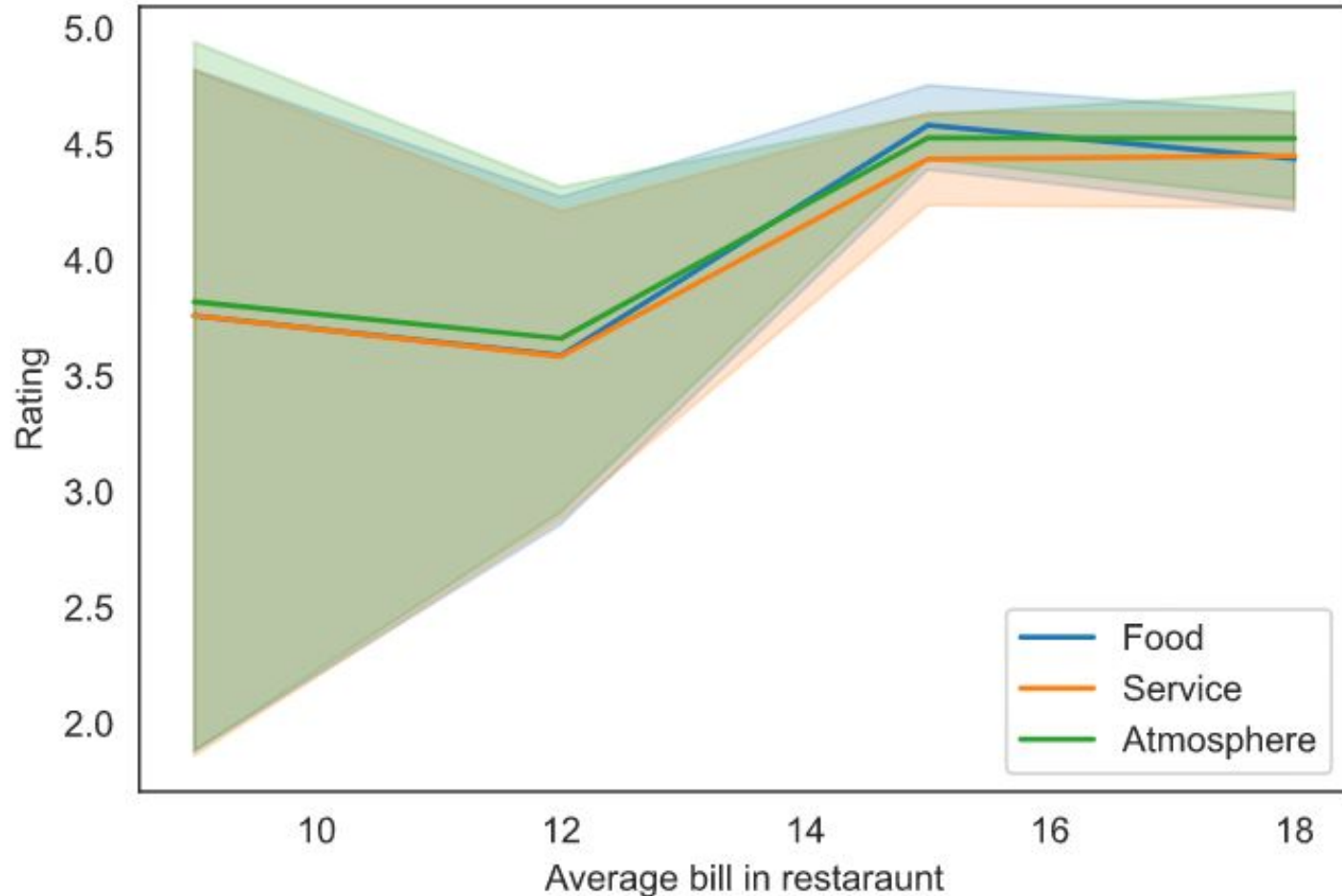


Correlation between food price and reviews

Reviews & Prices

- Expensive restaurants only get high ratings
- Cheaper restaurants have more spread
- Food is rated highest in medium and expensive restaurants
- The most expensive restaurants have better atmosphere
- Overall, people don't tend to rate restaurants less than 3.8 stars (or such restaurants close fast)

However, these ratings were so correlated, that their average was used instead.



Cuisine

Using Random Forest classification, we explored impact of different cuisine on good rating of a restaurant.

Result would imply that either the consumers' tastes are conservative based on their culture, or that the dishes in these cuisines are best known to cooks and are produced the best.

Italian	0.050941
American	0.060498
Estonian	0.065964
Russian	0.069142
Nordic	0.074943
Fusion	0.080974

Features

All restaurants had certain amount of features, that impacted their score.

Several top features were discovered, that affect rating more than others.

Wheelchair access	0.045703
Pet friendly	0.050941
Catering	0.051877
Terrace	0.058013
Business meetings	0.060498
Daily special	0.065964
Sea view	0.069142
Sauna party	0.074943
Food courses	0.080974
Wi-Fi	0.104375

Conclusion

We have provided answers for all the questions raised in the beginning of this research project.

We have dived deep into the offerings of the restaurants in the capital of Estonia

Unfortunately, our dataset was a little too small, but we managed to get some results nonetheless.

Now we know what food preferences locals and tourists prefer, as well as the kind of places that they would enjoy.

We got insight into restaurant placement strategy in Tallinn, and applied several machine learning algorithms to achieve our goals, including K-Means Clustering, SVMs, Random Forrest and Linear Regression.

Thank you for reading!