IBM Capstone Presentation

DORIAN AIGNER

Your best quote that reflects your approach… “It’s one small step for man, one giant leap for mankind.”

- NEIL ARMSTRONG

Contents

1 Introduction

1 Introduction

Vienna is one of the most international cities in Europe. Being regularly ranked the city with the best living quality by the Mercer Quality of Living Survey, it attracts people from all over the world. Furthermore, the UN headquarters and the International Atomic Energy Agency are located in the city. Hosting many people from all over the world implies a diversity in ethnic cousins. Additionally, Vienna's population is one of the richest in Europe with an annual average income of more than 55,000 USD (2019). A highly international population with high disposable income for eating out regularly, provides great opportunities for flourishing restaurants and food stalls of all kinds. From restaurants with a proud and long history deeply rooted in the country's traditions to new, flashy fusion cuisine, a vast selection is available.

2 Business Problem

2 Business Problem

Although this is true for the inner districts of the city, several districts may not have access to diverse food choices. Considering the overall high wages and the vast size of middle and high-income households, this fact provides many opportunities to start up an ethnic restaurant in a district less exposed to international cuisine yet.

Thus, the main objective of this project is to cluster Vienna's districts to find out where there is least competition and highest demand for such a restaurant. This data can be highly useful in combination with financial data over the purchasing power of the districts' households and the accessibility to the public transportation system to open a successful business and serve the international community in these needs.

3 Data

3 Data

The data for this project has been retrieved and processed through multiple sources, giving careful considerations to the accuracy of the methods used:

3.1 Districts (Neighbourhoods)

3.2 Geocoding

3.3 Venue Data

4 Methodology

4 Methodology

A thorough analysis of the principles of methods, rules, and postulates employed have been made in order to ensure the inferences to be made are as accurate as possible.

4.1 Accuracy of the Geocoding API

4.2 Folium

4.3 One hot encoding

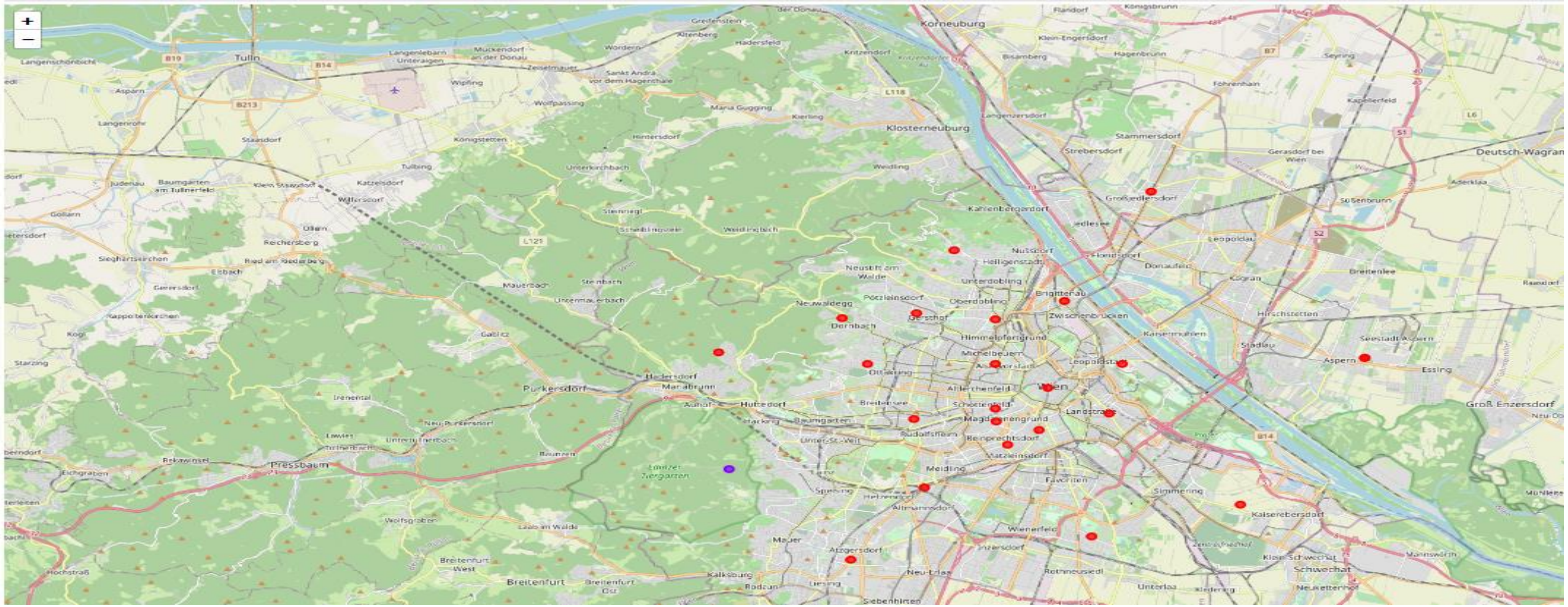
4.4 Top 10 most common venues

4.5 Optimal number of clusters

4.6 K-means clustering

5 Results

[41]:



Korneuburg
Tulln
Klosterneuburg
Stammersdorf
Strebersdorf
Deutsch-Wagram
Leopoldau
Süßenbrunn
Breitenlee
Hirschstetten
Aspern
Essling
Groß Enzersdorf
Kaiserebersdorf
Schwechat
Simmering
Favoriten
Meidling
Mauer
Atzgersdorf
Liesing
Kalksburg
Purkersdorf
Pressbaum
Breitenfurt
Hadersdorf
Mariabrunn
Hütteldorf
Ottakring
Dornbach
Neuwaldegg
Pötzleinsdorf
Gersthof
Heiligenstadt
Nussdorf
Brigittenau
Leopoldstadt
Wien
Landstraße
Kagran
Stadlau
Lainzer Tiergarten
Gablitz
Mauerbach
Königstetten
Langenrohr
Judenau
Sieghartskirchen

5 Results

The neighbourhoods are divided into n clusters where n is the number of clusters found using the optimal approach. The clustered neighbourhoods are visualized using different colours so as to make them distinguishable.

6 Discussion

6 Discussion

After analyzing the various clusters produced by the Machine learning algorithm, it becomes apparent that the Forequare API provides too limited information for several of the districts to run successfully. Therefore, it is recommended to repeat the described process with data from a different API, if applicable, a more local one.

However, even though the outcomes are only of partial use, the map and location data can be easily used to rebuild the proposed mechanism for other business cases. All in all, it shows that there is vast space for application.

7 Conclusion

7 Conclusion

As the trends described in the introduction have no end in sight, there is definitely demand. Using geodata and API information may allow entrepreneurs finding the best restaurant ideas for the best suitable area and therefore have a share of this rise in consumption. However, the process has shown that API information vary vastly between countries and local solutions of data usage has to be considered when looking deeper into this business idea.