

*Heather Gemmell*  
*January 2021*

# Bread for György

IBM Data Science Capstone





*Heather Gemmell*  
*January 2021*

# **BREAD FOR GYÖRGY**

---

IBM Data Science Capstone





# INTRODUCTION

---

Why is this project needed? Who is it for?

*Heather Gemmell*

*January 2021*

# INTRODUCTION

---

Throughout history, bread has been a staple food. At its most basic, bread is simply flour and water, mixed and heated, but it serves as a vital item of sustenance.

For many people, bread is considered a fundamental aspect of culinary, social, and religious culture.

*Heather Gemmell*

*January 2021*





*Heather Gemmell*

*January 2021*

## PROBLEM

---

- György has just moved to Budapest from Székelyföld (more commonly known as Transylvania), and he is keen to bring his culinary talents to the people of Budapest.
- He would like to open his bakery in one of the less crowded markets.
- György is willing to open his bakery in any part of Budapest – but would like to find the one with the least competition.





# DATA

---

What data we used and where we obtained it from

*Heather Gemmell*

*January 2021*

# DATA USED

---

- List of districts within Budapest
- Nearby venue data
- Geographical co-ordinates

*Heather Gemmell*

*January 2021*



# DATA SOURCES

---

Several tools were used to create insight



## Wikipedia

Category:Neighbourhoods of Budapest was scraped with beautifulsoup to provide initial starting data



## Foursquare

After using geocoder to obtain geographical data, venue information for neighbourhoods helped build database



## Folium

Applying data to maps in folium allowed for visualisation of data



## K-means Clustering

Lorem ipsum dolor sit amet, consectetur adipiscing elit.





# RESULTS

---

What we discovered

*Heather Gemmell*

*January 2021*

# BAKERY FREQUENCY BY DISTRICT

	District	Bakery	Cluster Labels	Latitude	Longitude
0	▶ 13th District of Budapest (2 P)	0.07	0	47.510140	19.015030
1	▶ 15th District of Budapest (5 P)	0.08	0	47.498960	19.051270
20	▶ Óbuda-Békásmegyer (1 C, 3 P)	0.07	0	47.541580	19.045010
3	▶ 2nd District of Budapest (6 P)	0.08	0	47.507450	19.066640
4	▶ Belváros-Lipótváros (14 P)	0.08	0	47.503120	19.050660
17	▶ Terézváros (1 P)	0.08	0	47.504880	19.062820
7	▶ Erzsébetváros (5 P)	0.07	0	47.500910	19.069360
9	▶ Hegyvidék (5 P)	0.08	0	47.499720	19.055080
12	▶ Kőbánya (8 P)	0.07	0	47.597900	19.043890
16	▶ Soroksár (1 P)	0.05	1	47.365480	19.091260
15	▶ Rákosmente (5 P)	0.05	1	47.371177	19.139722
14	▶ Pestszentlőrinc-Pestszentimre (3 P)	0.05	1	47.475930	19.160870

	District	Bakery	Cluster Labels	Latitude	Longitude
14	▶ Pestszentlőrinc-Pestszentimre (3 P)	0.05	1	47.475930	19.160870
13	▶ Pesterzsébet (1 P)	0.05	1	47.433330	19.116670
11	▶ Kispest (3 P)	0.05	1	47.412600	19.173110
6	▶ Csepel (2 P)	0.05	1	47.436000	19.090220
21	▶ Újbuda (5 P)	0.05	1	47.371177	19.139722
10	▶ Józsefváros (6 P)	0.06	2	47.553530	18.727300
8	▶ Ferencváros (1 C, 2 P)	0.06	2	47.476080	19.077100
5	▶ Budafok-Tétény (1 P)	0.06	2	47.472060	19.032520
18	▶ Várkerület (1 C, 15 P)	0.06	2	47.399371	18.971337
19	▶ Zugló (1 C, 3 P)	0.06	2	47.608530	19.193990
2	▶ 16th District of Budapest (3 P)	0.06	2	47.536700	19.039910
22	▶ Újpest (3 C, 4 P)	0.06	2	47.261480	19.081290





Heather Gemmell  
January 2021



01  
1



# RECOMMENDATIONS

---

Based on the findings, what is the next step?

*Heather Gemmell*

*January 2021*

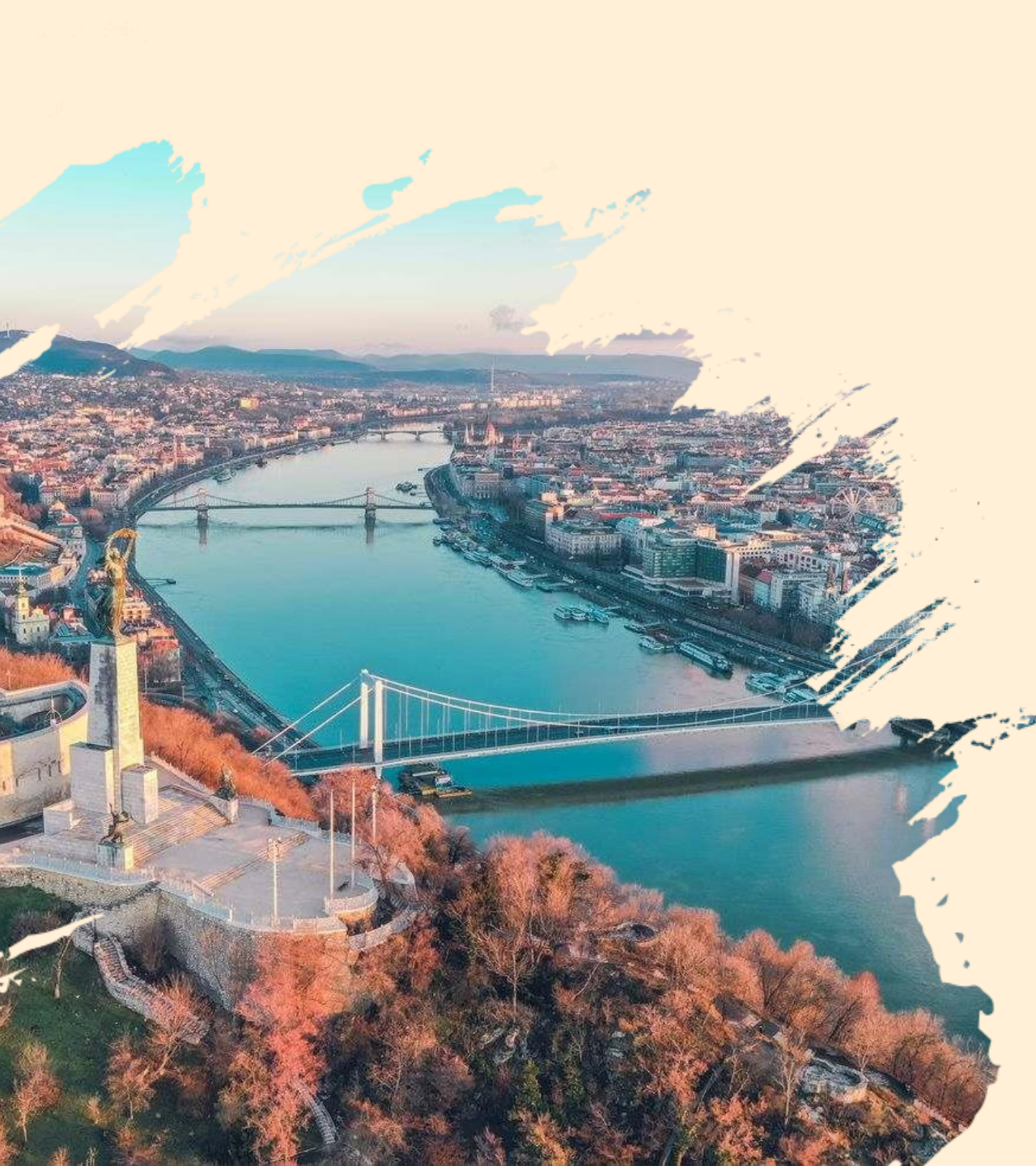




## CONCLUSIONS

---

- Budapest displays a fairly even spread of bakeries
- The marginally higher frequency in the red cluster is not sufficient to eliminate those districts as options
- Focus should be on Székely as UVP
- More research must be done into potential rental costs, as well as district-based tax breaks



# THANK YOU!

---

**Heather Gemmell**

Email:  
[magyar.insights@wedata.com](mailto:magyar.insights@wedata.com)

Phone:  
+36 305 3910-737