Finding the best locations for new pharmacies

IBM DATA SCIENCE CAPSTONE PROJECT

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION / BUSINESS PROBLEM

PROBLEM

The city of Vienna (capital city of Austria, Europe) has been experiencing a lot of growth in the recent years, which makes it attractive for old and new businesses alike. But this growth also comes with a number of challenges, one of them being the development of the healthcare system to accomodate an increasing number of inhabitants. Judging by the fact that the predictions about population increase had to be adjusted upwards every year since 1989 (see image below, in German), it becomes clear that planning has to start now in order to avoid shortages.

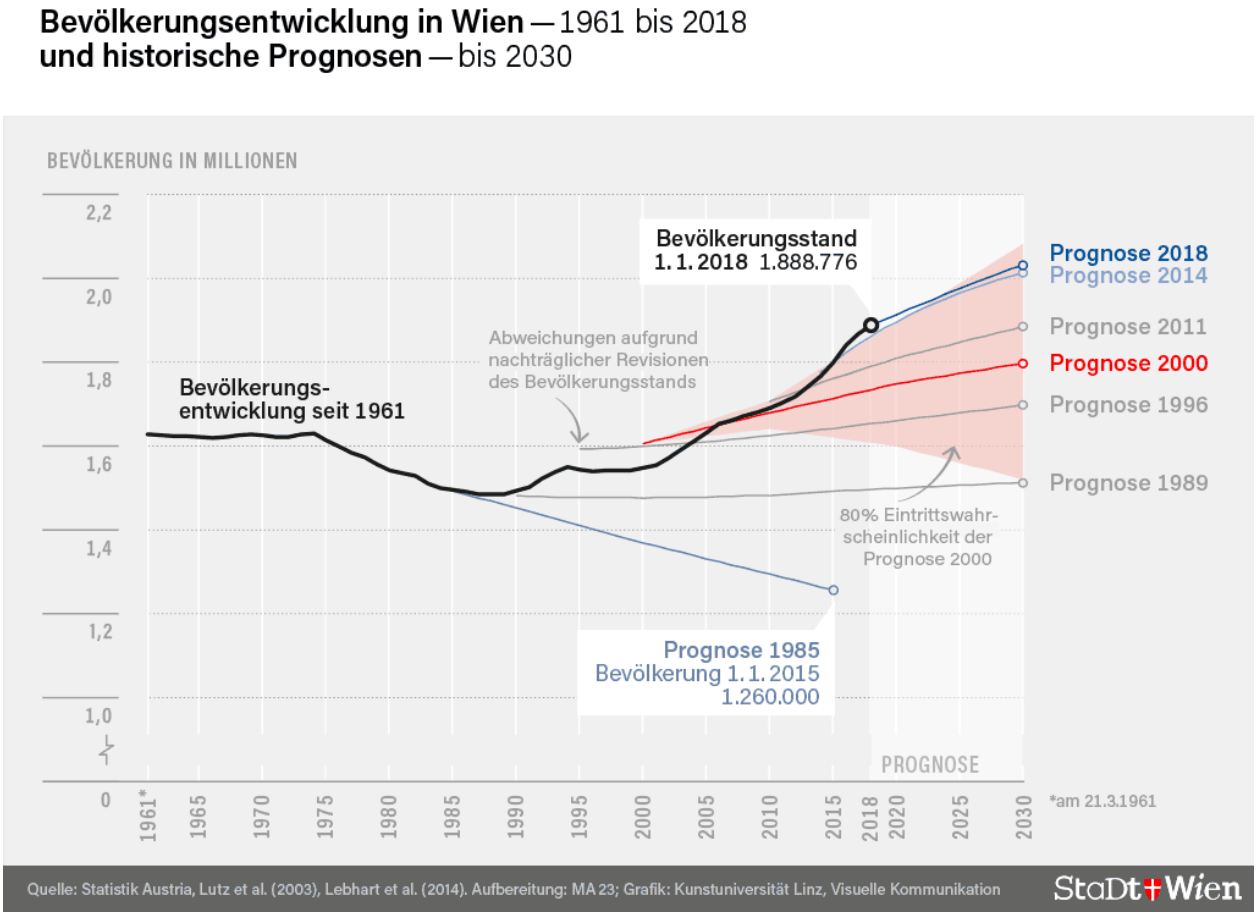

Figure 1: Population development in Vienna incl. predictions

Since rental space for businesses is not unlimited (especially in the center of the city), prices are ever increasing, making it essential to pick the right location for setting up new business. In addition, setting up businesses like pharmacies is regulated by local legislation, applying for a permit to set up a new pharmacy is a costly and time consuming process in itself, but even more so if a bad spot was chosen. While this is an optimization problem in itself, it all starts with knowing where demand will be coming from and evaluating locations from there.

Lastly, there has been a lot of political debate whether or not doctors should be allowed to sell drugs directly to their patients, breaking the current legal monopoly currently held by pharmacies. As with most things, there are arguments for either side, but we hope to contribute to the discussion with some data-driven analysis.

AUDIENCE

As with any serious analysis, it is important to think about the audience that it addresses. For this analysis we have identified the following:

- Investors in the pharmaceutical industry, interested in making an informed decision about whether and where to invest in Vienna by setting up new pharmacies.
- Public administration and city developers, wanting to decide where space should allocated to maximize public service in the coming decade.
- Universities in and around Vienna, trying to plan how many students of pharmacology will best suit the needs of Vienna.
- Politicians in Vienna, interested in some facts about how many pharmacies will be needed in the near future.

DATA

To perform our analysis, we will need some data. Luckily, we live in the information age and there is a wealth of data available online. For this project, we will be using two data sources, Foursquare and Open Data Austria.

FOURSQUARE

Foursquare specializes in location based data that is enriched with user based imagery and reviews. For our project, we will pull data from Foursquare that allows us to find the current location of pharmacies and match them up with the various districts of Vienna. Using this data, we'll be able to calculate densities, like persons per pharmacy, today and 15 years in the future, which allows us to recommend locations where additional pharmacies might be needed.

Note to non-German speaking readers: "Apotheke"/"Wien" are the German words for "Pharmacy"/"Vienna", should you encounter them.

The most common way to use Foursqure for regular users is through their website, using a URL like this: https://de.foursquare.com/explore?mode=url&near=Wien%2C%20%C3%96sterreich&nearGeoId=72057594040689305&q=apotheke

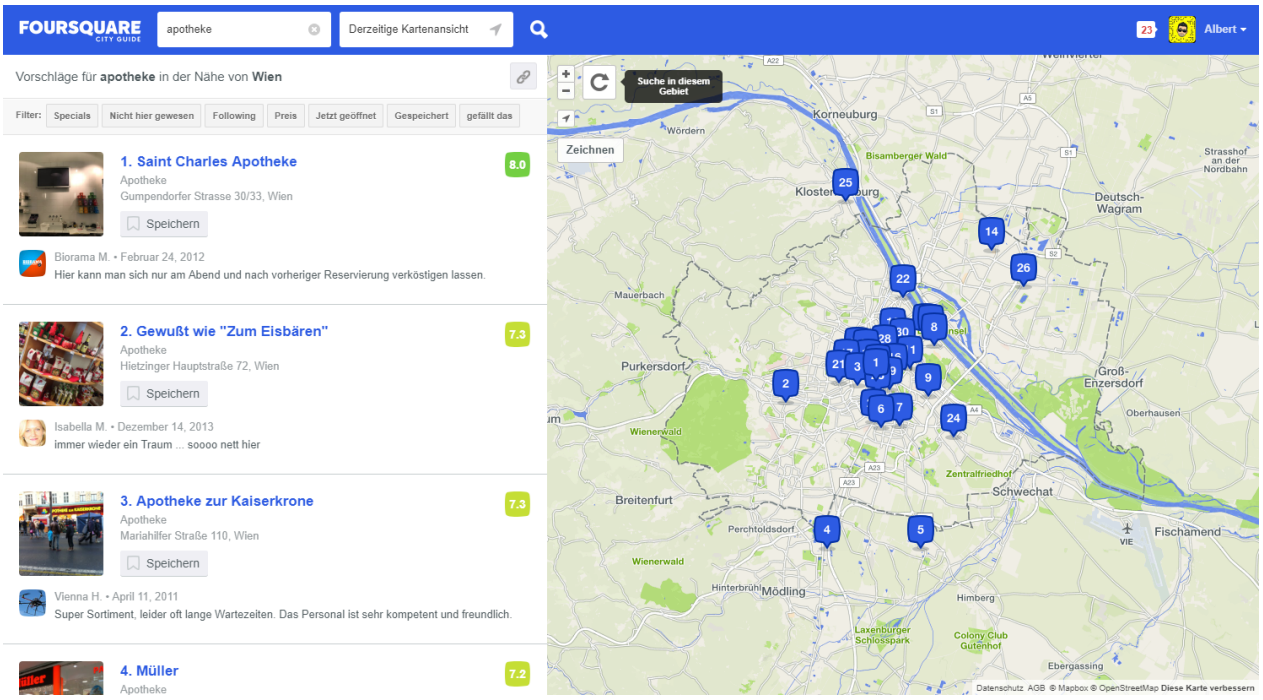

Figure 2: Venue search on Foursquare Website

For Data Scientists however, using their developer API (https://developer.foursquare.com/docs/api) might be more advantageous, like this for example: https://api.foursquare.com/v2/venues/search?client_id=XXX&client_secret=XXX&v=20180604&query=Apotheke&limit=500&near=Vienna,AT&intent=browse

A typical response is returned to the client in JSON format:

```
{'meta': {'code': 200, 'requestId': '5d7770ffcf72a0002c90ed08'},
 'response': {'venues': [{'id': '4f4e057ce4b051dc48ef4e3f',
    'name': 'Apotheke zum weissen Storch',
    'location': {'address': 'Tuchlauben',
     'crossStreet': 'Steindlgasse',
     'lat': 48.209947,
     'lng': 16.370146,
     'labeledLatLngs': [{'label': 'display',
       'lat': 48.209947,
       'lng': 16.370146}],
     'postalCode': '1010',
     'cc': 'AT',
     'city': 'Wien',
     'state': 'Wien',
     'country': 'Österreich',
     'formattedAddress': ['Tuchlauben (Steindlgasse)',
      '1010 Wien',
      'Österreich']},
    'categories': [{'id': '4bf58dd8d48988d10f951735',
      'name': 'Pharmacy',
      'pluralName': 'Pharmacies',
      'shortName': 'Pharmacy',
```

```
    'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/shops/pharmacy_',
     'suffix': '.png'},
    'primary': True}],
  'referralId': 'v-1568108799',
  'hasPerk': False},
```

While our initial intention was to use Foursquare to retrieve all the pharmacy locations in Vienna, this plan had to be abandoned due to arbitrary restrictions in their API. More specifically, the maximum number of venues per search request is limited to 50 entries:

query	tacos	A search term to be applied against venue names.
limit	10	Number of results to return, up to 50.
categoryId	4bf58dd8d488d11094,	A comma separated list of categories to limit results to. I

Figure 3: Limitations of the maximum number of results returned by the Foursquare API

(Source: https://developer.foursquare.com/docs/api/venues/search)

Other query approaches like sw-ne boxing won't work due to the districts of Vienna not fitting well into rectangular shapes and/or causing issues with overlap. As a result, we'll have to find another source to work with for our analysis in order to avoid huge development efforts.

Luckily, we were able to find another data source for this problem.

OPEN DATA AUSTRIA

Open Data Austria is a website operated by the administration of Vienna, to make their data publicly available as Open Data. We will be mostly pulling census data from there, which includes current number of inhabitants per district as well as predictions 15 years into the future.

- Population projection by sex and country of birth for districts in Vienna 2014 to 2034: https://www.wien.gv.at/statistik/ogd/vie_302.csv
- Pharmacy locations Vienna: https://data.wien.gv.at/daten/geo?service=WFS&request=GetFeature&version=1.1.0&typeName=ogdwien:APOTHEKEOGD&srsName=EPSG:4326&outputFormat=csv
- Vienna District boundaries: https://data.wien.gv.at/daten/geo?service=WFS&request=GetFeature&version=1.1.0&typeName=ogdwien:BEZIRKSGRENZEOGD&srsName=EPSG:4326&outputFormat=json

METHODOLOGY

This section represents the main component of the report where we discuss and describe the exploratory data analysis that we did, the inferential statistical testing that we performed and what machine learnings were used and why.

DATA PREPARATION

The raw data we get from Open Data Austria is quite technical in nature and contains numerous columns we won't need. In addition, some of the data needs reformatting for easier consumption in our analysis.

POPULATION DATA

First, let's get rid of the stuff we don't need and rename the remaining columns to something more legible. Then, let's extract the real district number from the District field and strip off everything after the 4th digit from the Year field. Finally, let's get rid of rows that are out of date.

	NUTS1	NUTS2	NUTS3	DISTRICT_CODE	SUB_DISTRICT_CODE	POP_TOTAL	POP_MEN	POP_WOMEN	POP_NATIVE	POP_FOREIGN	REF_Date
0	AT1	AT13	AT130	90100	.	16434	7863	8571	10982	5452	20140101
1	AT1	AT13	AT130	90200	.	100016	48796	51220	61508	38508	20140102
2	AT1	AT13	AT130	90300	.	87213	41463	45750	56868	30345	20140103
3	AT1	AT13	AT130	90400	.	31947	15172	16775	20270	11677	20140104
4	AT1	AT13	AT130	90500	.	54153	26585	27568	31927	22226	20140105

Table 1: Population data raw

	District	Population	Year
115	1	15958	2019
116	2	108651	2019
117	3	91915	2019
118	4	33190	2019
119	5	56462	2019

Table 2: Population data cleaned

PHARMACY DATA

Looking at the raw data, we can basically get rid of all columns except District ("BEZIRK") and Location ("geometry"), as well as the ID ("OBJECTID") for consistency purposes. As with the population data, we'll rename the columns to something more legible. Finally, we'll also extract Longitude and Latitude from the Location column.

	id	OBJECTID	BEZEICHNUNG	BEZIRK	STRASSE	ONR	ADRESSE	TELEFON	FAX	EMAIL	WE
0	APOTHEKEOGD.fid-59fe94b2_16d1a781fd2_786	798067	Apotheke Alte Remise	16	Maroltingergasse	53	16., Maroltingergasse 53	01 494 14 92	01 494 14 92	alte.remise@apohofer.at	
1	APOTHEKEOGD.fid-59fe94b2_16d1a781fd2_787	798068	Apotheke Mag.pharm. Roland Clemens	17	Parhamerplatz	6	17., Parhamerplatz 6	01 486 14 85	01 486 14 85	apotheke.clemens@tele2.at	
2	APOTHEKEOGD.fid-59fe94b2_16d1a781fd2_788	798069	Linden-Apotheke	17	Hernalser Hauptstraße	155	17., Hernalser Hauptstraße 155	01 486 24 04	01 486 24 04	lindenapotheke@gmx.at	http://www.linden
3	APOTHEKEOGD.fid-59fe94b2_16d1a781fd2_789	798070	Apotheke "Zum weißen Kreuz"	17	Hernalser Hauptstraße	99	17., Hernalser Hauptstraße 99	01 486 33 86	01 486 33	apoweisseskreuz@hotmail.com	

Table 3: Pharmacy data raw

	ID	District	geometry	Longitude	Latitude
0	798067	16	POINT (16.30518521580027 48.21007436224088)	16.305185	48.210074
1	798068	17	POINT (16.3278525218971 48.21553926756905)	16.327853	48.215539
2	798069	17	POINT (16.31841646434104 48.22233184648672)	16.318416	48.222332
3	798070	17	POINT (16.32613097970342 48.21935539747856)	16.326131	48.219355
4	798071	17	POINT (16.30163108476474 48.22851885937867)	16.301631	48.228519

Table 4: Pharmacy data cleaned

DISTRICT DATA

As for the district boundaries, these are also supplied by the administration of Vienna. We opted to import this data into a GeoPandas dataframe, which greatly simplifies generating maps later on, while retaining all the benefits of regular Pandas dataframes.

	id	NAMEK	BEZNR	BEZ_RZ	NAMEK_NUM	NAMEK_RZ	NAMEG	LABEL	BEZ	DISTRICT_CODE	STATAUSTRIA_BEZ
0	BEZIRKSGRENZEOGD.9290	Neubau	7	VII	7., Neubau	VII. Neubau	NEUBAU	VII.	07	1070	
1	BEZIRKSGRENZEOGD.9291	Landstraße	3	III	3., Landstraße	III. Landstraße	LANDSTRASSE	III.	03	1030	
2	BEZIRKSGRENZEOGD.9292	Josefstadt	8	VIII	8., Josefstadt	VIII. Josefstadt	JOSEFSTADT	VIII.	08	1080	
3	BEZIRKSGRENZEOGD.9293	Innere Stadt	1	I	1., Innere Stadt	I. Innere Stadt	INNERE STADT	I.	01	1010	
4	BEZIRKSGRENZEOGD.9294	Ottakring	16	XVI	16., Ottakring	XVI. Ottakring	OTTAKRING	XVI.	16	1160	

Table 5: District data

Apart from loading the data, not much data preparation was needed, since there are only 23 records and the data set only contains useful data as it is.

EXPLORATORY ANALYSIS

POPULATION DEVELOPMENT

Let's look at the population development in the various districts of Vienna from 2014 to 2034. To do this, we first pivot the original relational data into something that can be plotted more easily.

District	1	2	3	4	5	6	7	8	9	10	...	14	15	16	17	18	19	20	21	22	23
Year																					
2019	15958	108651	91915	33190	56462	32902	33129	26422	43160	201388	...	88668	77061	104909	57694	50668	70652	92949	158075	184859	103770
2020	15838	109814	92634	33305	56729	33235	33492	26617	43385	203359	...	88800	77198	105474	57968	50752	70509	94944	160316	189533	103970
2021	15712	111093	94264	33374	56902	33506	33650	26763	43538	205393	...	89536	77035	105631	58390	50762	70272	96427	162366	190821	106149
2022	15595	112060	95668	33448	57079	33773	33804	26905	43690	207152	...	90268	76901	105800	58815	50781	70059	98014	163373	193105	108349
2023	15486	113400	96864	33529	57263	34036	33955	27046	44204	209578	...	90319	76798	105986	58981	50811	69873	99311	163399	197517	109320
2024	15396	114116	97364	33654	57527	34339	34142	27220	44792	211516	...	90542	76819	106319	59228	50913	69778	100437	162808	203383	109835
2025	15278	115339	97633	33670	57581	34515	34227	27289	45314	214112	...	90550	76635	106318	59264	50848	69580	100835	164623	207568	110997
2026	15170	115846	98571	33704	57855	34703	34322	27368	45835	216725	...	90606	77125	106377	59331	50814	69426	101275	166391	209780	111396
2027	15066	116318	99464	33729	58106	34875	34403	27436	45873	219210	...	90369	77581	106410	59383	50774	69277	101680	168091	212616	111780
2028	14990	116828	100570	33841	58320	35135	34566	27576	46029	219903		90359	78242	106736	59600	50872	69288	102353	169805	214252	112227

Table 6: Pivot table of predicted population development in Vienna

Using this data, we can plot the population development using MatPlotLib.

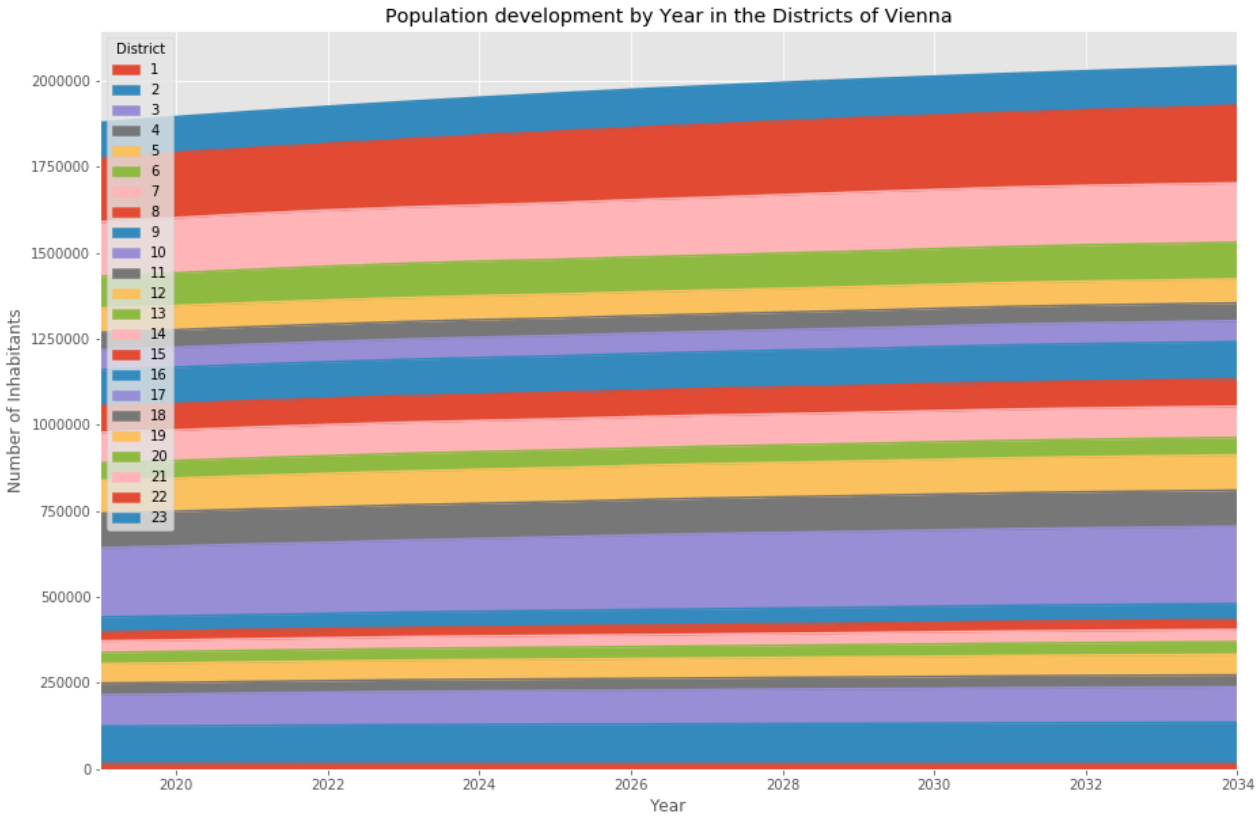

Figure 4: Plot of predicted population development in Vienna

While looking at the population development per district over all the years is nice, it might be more helpful to see the change between now (2019) and 15 years in the future (2034). For this purpose, we'll create a new dataframe containing only those years.

	District	Population2019	Population2034	PopulationChangePrc
0	1	15958	14628	-8.33
1	2	108651	120833	11.21
2	3	91915	103546	12.65
3	4	33190	34567	4.15
4	5	56462	59692	5.72
5	6	32902	36659	11.42
6	7	33129	35503	7.17
7	8	26422	28373	7.38
8	9	43160	46961	8.81
9	10	201388	224691	11.57
10	11	100061	104993	4.93
11	12	95550	102710	7.49
12	13	51298	50524	-1.51
13	14	88668	90724	2.32
14	15	77061	78704	2.13
15	16	104909	108909	3.81
16	17	57694	61025	5.77
17	18	50668	51662	1.96
18	19	70652	69795	-1.21
19	20	92949	106436	14.51
20	21	158075	172580	9.18
21	22	184859	225819	22.16
22	23	103770	114078	9.93

Table 7: Population development in Vienna 2019/2034

Looking at this data it already becomes apparent that the population development is different in each district of Vienna, with some districts shrinking in population by over 8 percent, while others are predicted to grow as much as 22 percent.

PHARMACY LOCATIONS

Let's first take a look at how many pharmacies there are today in the respective districts of Vienna: 329! The overall number will not help us in our analysis, so we are going to break it down by district:

	District	Pharmacies2019
0	1	21
1	2	19
2	3	16
3	4	8
4	5	9
5	6	7
6	7	9
7	8	5
8	9	13
9	10	28
10	11	12
11	12	16
12	13	8
13	14	15
14	15	14
15	16	16
16	17	8
17	18	12
18	19	12
19	20	12
20	21	27
21	22	25
22	23	17

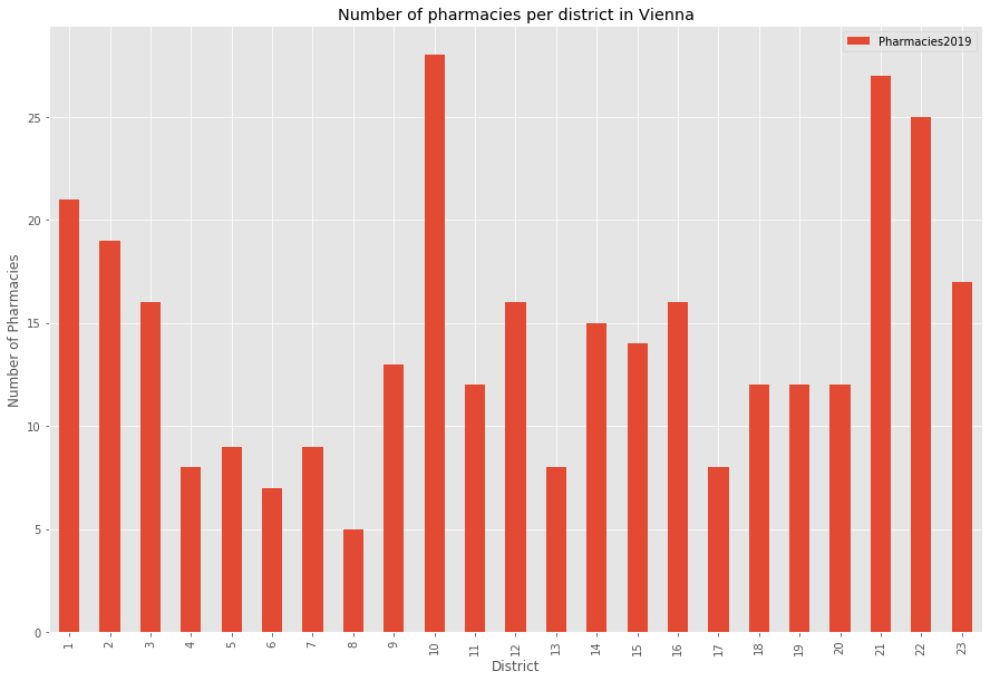

Figure 5: Number of pharmacies per district in 2019

Now that we know how many pharmacies there are per district, let's plot them on a map to get a more visual impression.

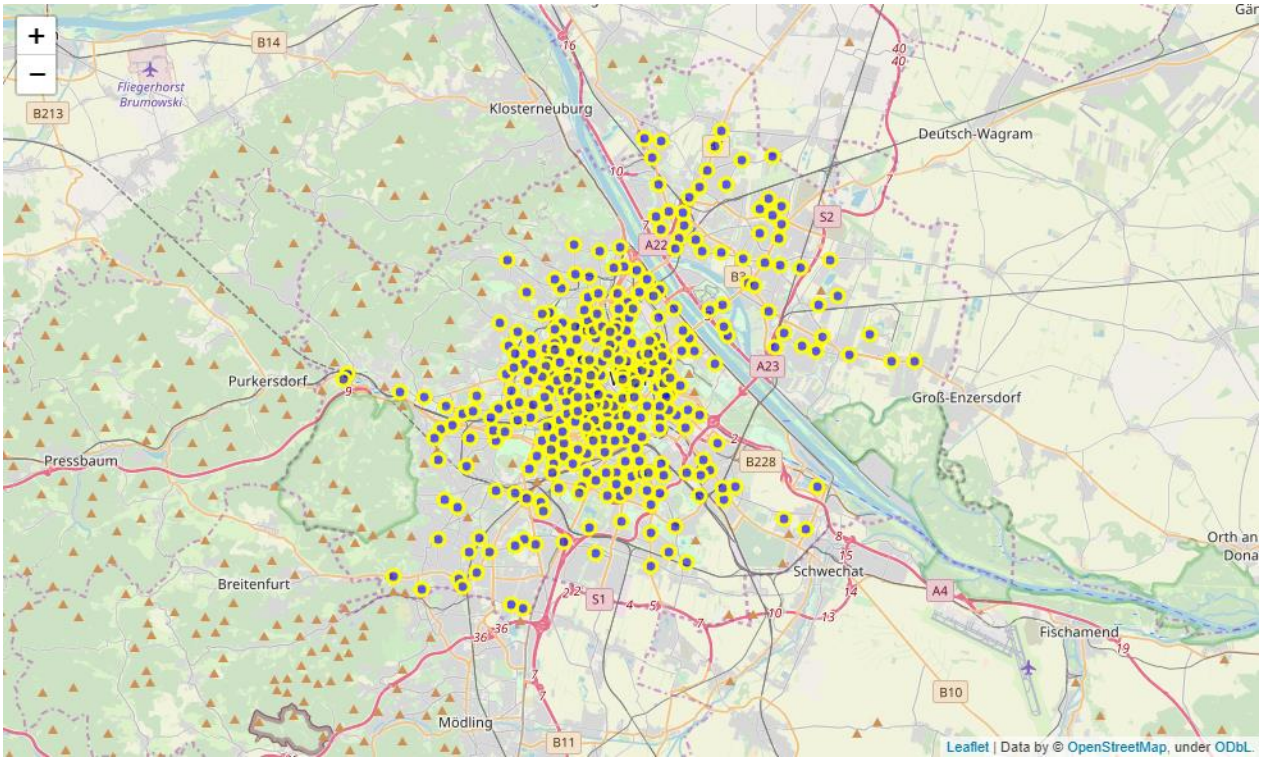

Figure 6: Map plot of pharmacies in Vienna

Again, having these many circlemarkers on a map isn't useful, so we'll at the districts and create a choropleth map. One thing to note already is that most pharmacies seem to be placed in areas of dense population and/or main transportation routes.

Note: Due to a bug in Folium and Chrome, it was not possible to proceed with Folium for mapping and we switched to using the plotting function included in GeoPandas.

Let's therefore try a simpler plot that highlights the different districts and plot the pharmacies in there. To do this, we merge the pharmacy count per district with the district boundary data prepared earlier.

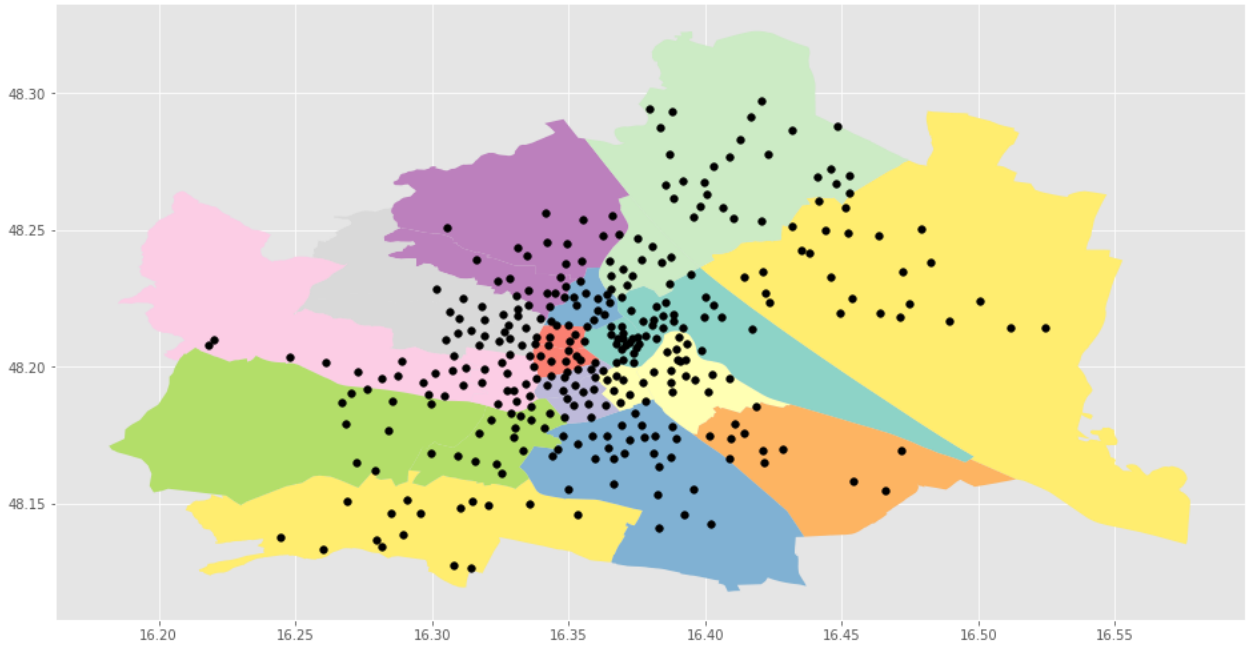

Figure 7: Map plot of pharmacies by district

The figure above shows us the pharmacy locations (black dots) and the 23 districts of Vienna all in one map. Looking at this plot, we can see that there are quite densly populated areas which also seem to have lots of pharmacies, while there also districts on the outskirts of Vienna that have fewer pharmacies. Let's inspect this further in the next section and calculate some metrics to highlight the situation.

FEATURE ENGINEERING

Since our main objective is to understand where more/less pharmacies are going to be needed, let's combine the population data with pharmacy data to see how many people a pharmacy has to serve in each of the districts.

For this purpose, a new dataframe is created based upon the population dataframe which contains the following calculated metrics and engineered features about each district:

- Density2019: Population in 2019 divided by Pharmacies in 2019
- Density2034: Population in 2034 divided by Pharmacies in 2019
- DensityChangePrc: Change in percent of the two above
- Pharmacies2034Pop: Number of pharmacies needed in 2034 to accommodate population in 2034
- Pharmacies2034Den: Number of pharmacies needed in 2034 to approach the average density of 2019

- PharmaciesChangePop: Change in number of pharmacies needed due to population change 2019->2034
- PharmaciesChangeDen: Change in number of pharmacies needed to smooth out density across Vienna
- PharmaciesChangeTot: Total change needed, sum of the two above

District	Population2019	Population2034	PopulationChangePrc	Pharmacies2019	Density2019	Density2034	DensityChangePrc	Pharmacies2034Pop	Pha
1	15958	14628	-8.33	21	759	696	-9.05	19	
2	108651	120833	11.21	19	5718	6359	10.08	21	
3	91915	103546	12.65	16	5744	6471	11.23	18	
4	33190	34567	4.15	8	4148	4320	3.98	8	
5	56462	59692	5.72	9	6273	6632	5.41	10	
6	32902	36659	11.42	7	4700	5237	10.25	8	
7	33129	35503	7.17	9	3681	3944	6.67	10	
8	26422	28373	7.38	5	5284	5674	6.87	5	
9	43160	46961	8.81	13	3320	3612	8.08	14	
10	201388	224691	11.57	28	7192	8024	10.37	31	
11	100061	104993	4.93	12	8338	8749	4.70	13	
12	95550	102710	7.49	16	5971	6419	6.98	17	
13	51298	50524	-1.51	8	6412	6315	-1.54	8	
14	88668	90724	2.32	15	5911	6048	2.27	15	
15	77061	78704	2.13	14	5504	5621	2.08	14	
16	104909	108909	3.81	16	6556	6806	3.67	17	
17	57694	61025	5.77	8	7211	7628	5.47	8	
18	50668	51662	1.96	12	4222	4305	1.93	12	
19	70652	69795	-1.21	12	5887	5816	-1.22	12	
20	92949	106436	14.51	12	7745	8869	12.67	14	
21	158075	172580	9.18	27	5854	6391	8.40	29	
22	184859	225819	22.16	25	7394	9032	18.14	31	
23	103770	114078	9.93	17	6104	6710	9.03	19	

Table 8: Engineered Features per district

Using these features, we will be able to conduct machine learning later in the analysis in the form of Clustering, which will help us in grouping the districts in order to formulate a strategy for each of them.

DISTRICT ANALYSIS

With all the relevant metrics calculated, let's now join these with the district boundary data in order to make some visualization plots. Let's start off by looking at the density of pharmacies in 2019 and 2034.

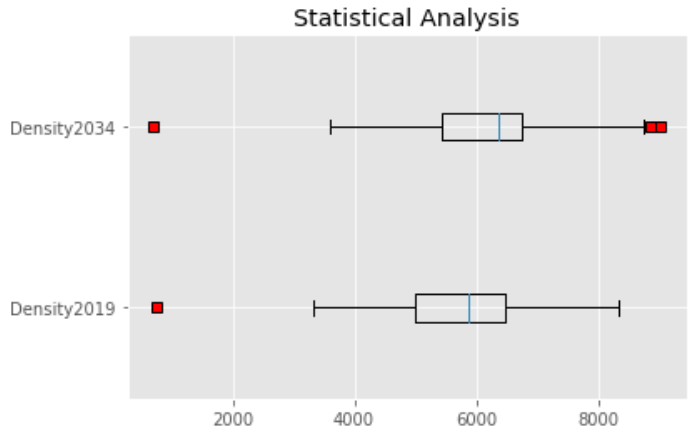

Figure 8: Statistical analysis of the pharmacy density development 2019/2034

In the plot above we can see that in 2019 we have one outlier on the lower end of the density spectrum, while the average number of people serviced by a pharmacy seems to be just below 6000. Fast forward to the year 2034 and we see that with the predicted population development, this average will increase to well above 6000 and we also are going to see outliers on the upper side of the density spectrum. This should be enough for further analysis.

So let's first take a look at the population in the year 2019 to possibly identify hotspots visible today.

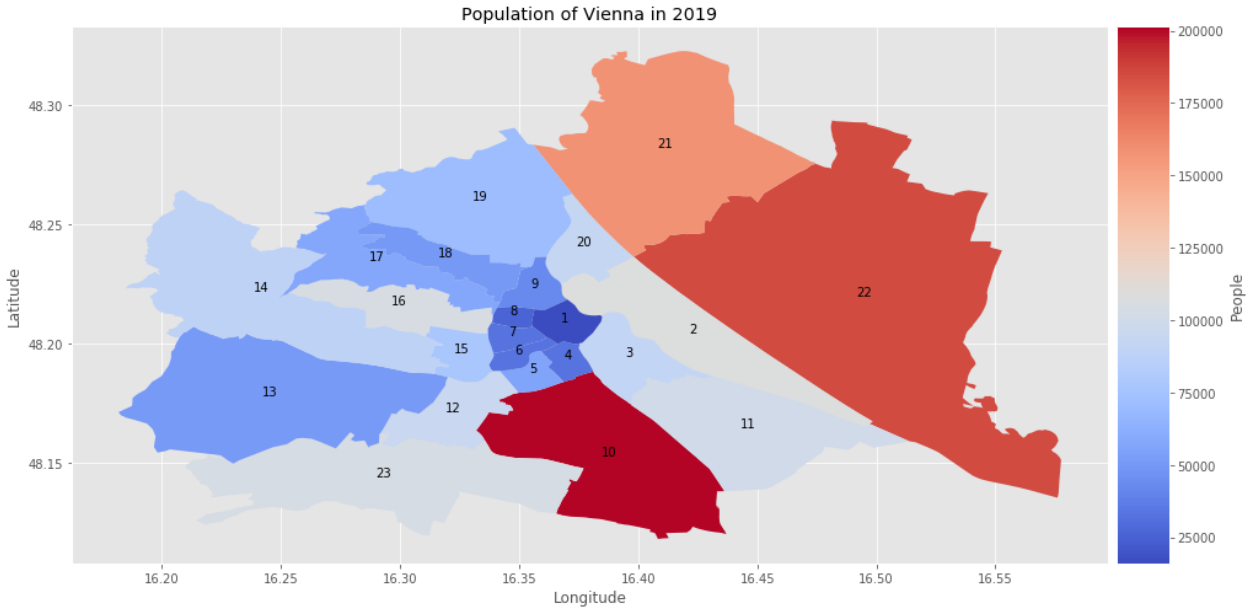

Figure 9: Population of Vienna in 2019

Looking at the plot above we can already see that the districts of Vienna are far from homogeneous. While the districts in the center have low population, districts 10, 22 and 21 are the ones with the highest population.

Now let's look at how many inhabitants each pharmacy in a given district is servicing today, in the year 2019.

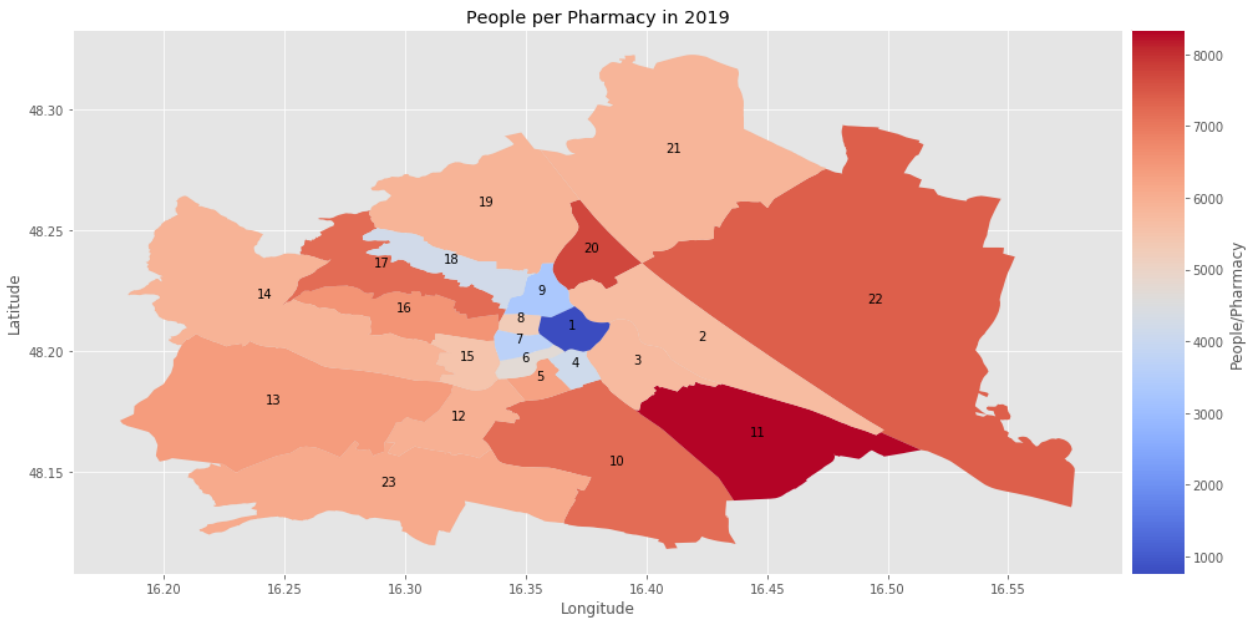

Figure 10: People per pharmacy in 2019

In the plot above we can already see that there are great differences in the number of serviced inhabitants per district. In some districts (11, 20, 22, 17 and 10) the number of people per pharmacy is very high, while in certain others (1, 9, 7 and 4) there seems to be more than enough pharmacies today.

Next, let's take a look at the administration data predicting the population development for the year 2034 in each district.

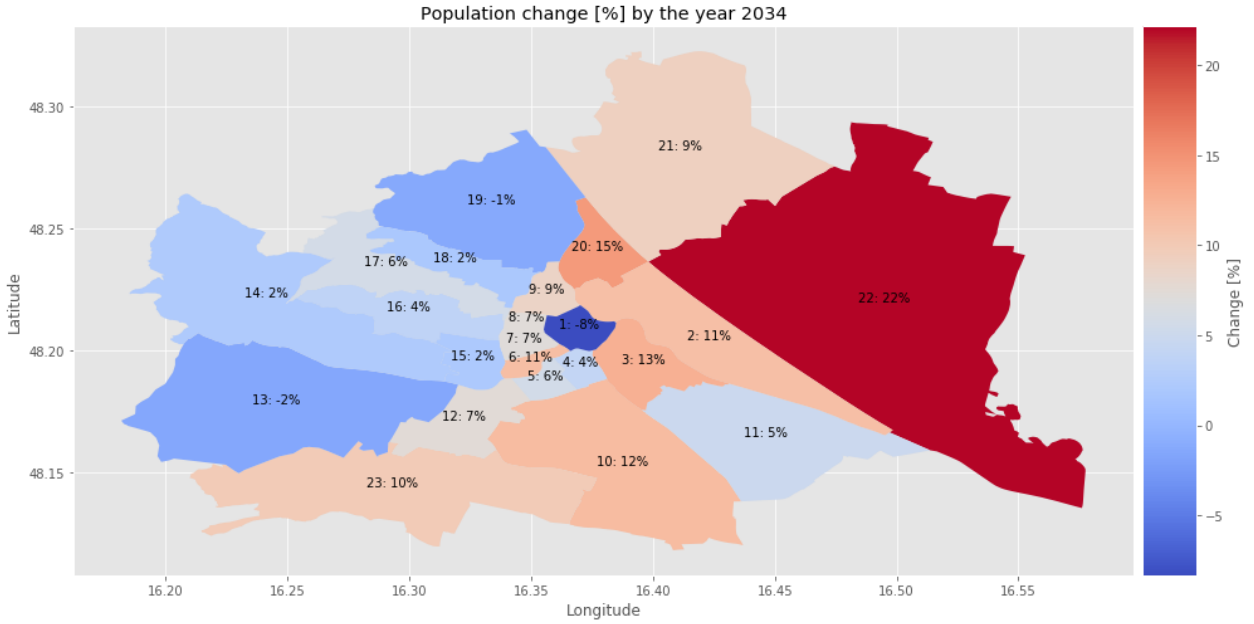

Figure 11: Population change in percent by the year 2034

In the plot of the expected population change in percent we can see that there are districts in the center and west of Vienna (1, 13 and 19) that will actually shrink in population, while some other are predicted to see substantial growth (22, 10 and 23).

Considering the density of pharmacies today and the development of the population in Vienna in the years until 2034, it is now time to make predictions how the number of pharmacies should be shifted in each district. There are two factors influencing the development of the number of pharmacies: One is the actual predicted change in population, the other is the smoothing of the density of pharmacies across the city of Vienna.

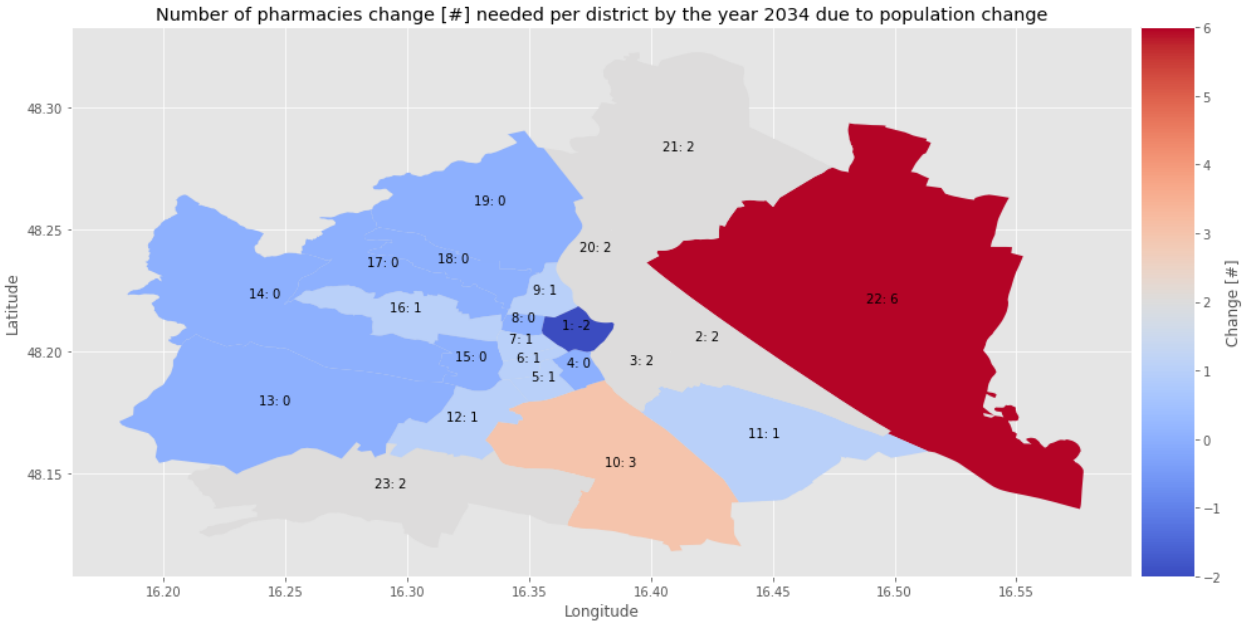

Figure 12: Number of pharmacies change in percent due to population change

Judging by the map plot above, there appear two interesting districts: 1 and 22. While district 1 seems to need a lot fewer pharamacies in the future, district 22 will need six additional ones due to increasing population alone.

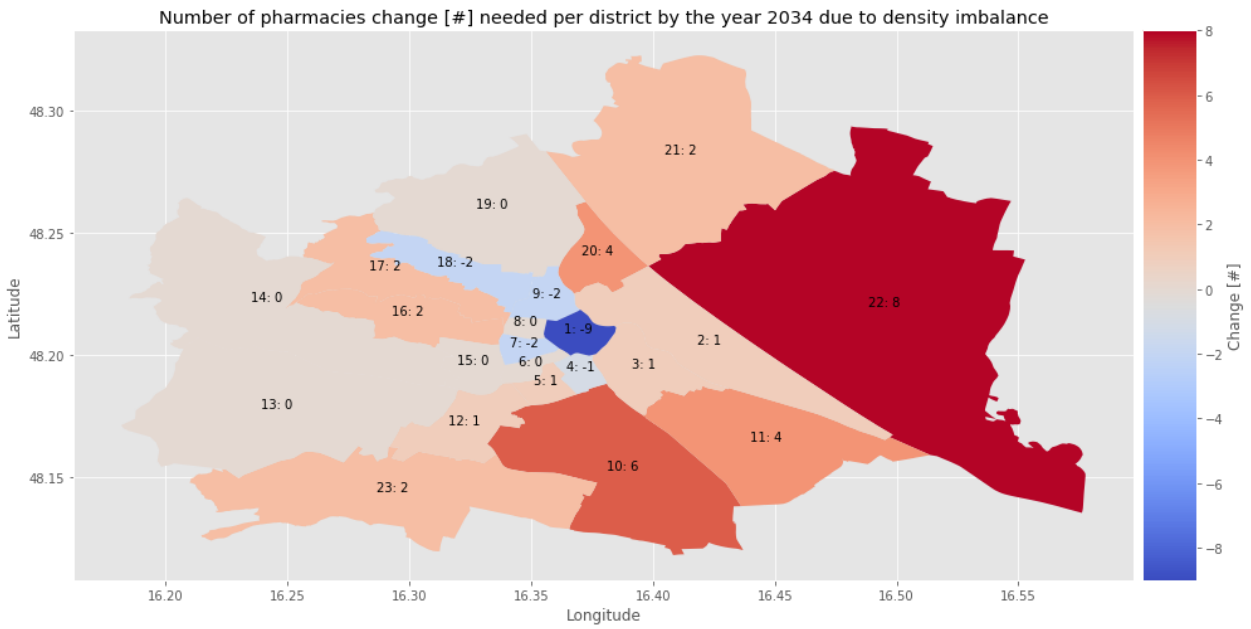

Figure 13: Number of pharmacies change in percent due to density imbalance

As mentioned earlier, there appears to be somewhat of an imbalance when it comes to the number of people served by each pharmacy between the districts. To level out 50% of this imbalance, the number of pharmacies should change by certain mounts. Again, it is districts 1 and 22 with the outlier values.

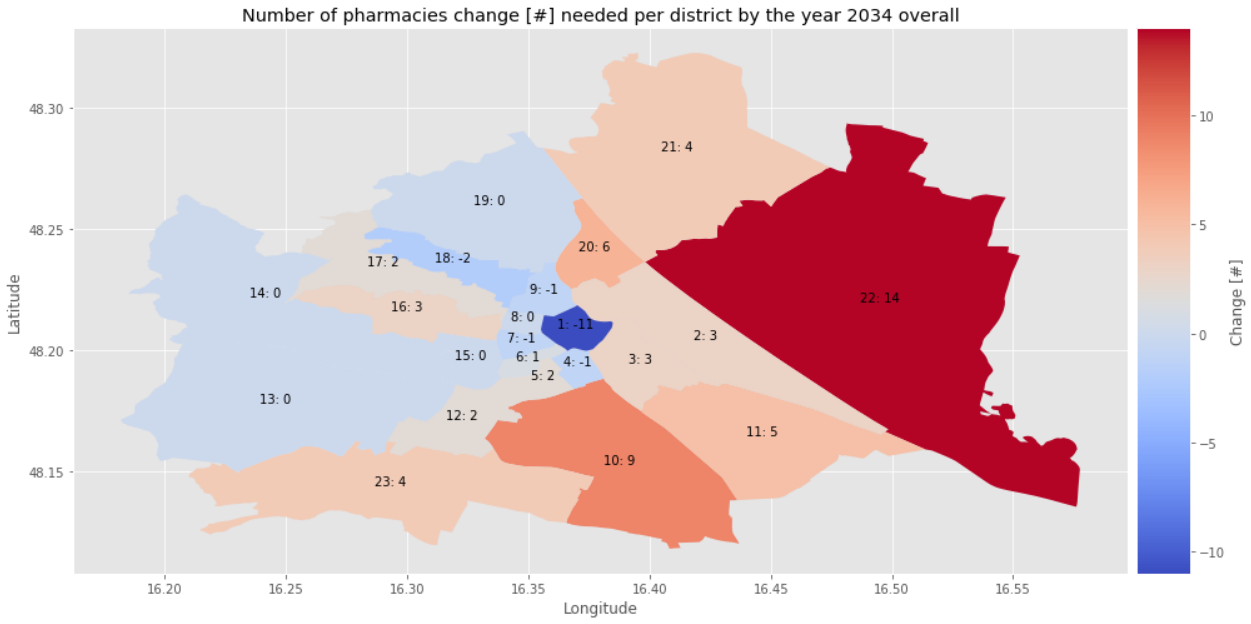

Figure 14: Number of pharmacies change by the year 2034 overall

Putting the previous two plots together, we can calculate the overall number of change in pharmacy numbers per district that is recommended for the year 2034. Looking at the map plot above, it will come as little surprise that the most extreme values are to be found in the central district 1 and the outskirts distrct 22 respectively.

This concludes our exploratory analysis and allows us to use machine learning to finalize our analysis methodology.

CLUSTERING

Finally, with the data from the last map plot, we proceed to put the 23 districts of Vienna into 3 clusters for which development strategies should be formulated later on.

CLUSTER 0: FEWER PHARMACIES NEEDED

The first cluster identified by Kmeans contains the five districts which are going to need fewer pharmacies in the future, either due to shrinking population or due to excess supply present already today.

	id	NAMEK	BEZNR	BEZ_RZ	NAMEK_NUM	NAMEK_RZ	NAMEG	LABEL	BEZ	DISTRICT_CODE	...	Pharmacies2019
5	BEZIRKSGRENZEOGD.9295	Alsergrund	9	IX	9., Alsergrund	IX. Alsergrund	ALSERGRUND	IX.	09	1090	...	13
19	BEZIRKSGRENZEOGD.9286	Wieden	4	IV	4., Wieden	IV. Wieden	WIEDEN	IV.	04	1040	...	8
0	BEZIRKSGRENZEOGD.9290	Neubau	7	VII	7., Neubau	VII. Neubau	NEUBAU	VII.	07	1070	...	9
7	BEZIRKSGRENZEOGD.9297	Währing	18	XVIII	18., Währing	XVIII. Währing	WÄHRING	XVIII.	18	1180	...	12
3	BEZIRKSGRENZEOGD.9293	Innere Stadt	1	I	1., Innere Stadt	I. Innere Stadt	INNERE STADT	I.	01	1010	...	21

5 rows × 32 columns

Table 9: Cluster 0 - Fewer pharmacies needed

CLUSTER 1: JUST RIGHT

The second cluster appears to contain those district where no change in the number of pharmacies is going to be necessary.

	id	NAMEK	BEZNR	BEZ_RZ	NAMEK_NUM	NAMEK_RZ	NAMEG	LABEL	BEZ	DISTRICT_CODE	...	Pharmacies2019
11	BEZIRKSGRENZEOGD.9301	Döbling	19	XIX	19., Döbling	XIX. Döbling	DÖBLING	XIX.	19	1190	...	12
21	BEZIRKSGRENZEOGD.9288	Hietzing	13	XIII	13., Hietzing	XIII. Hietzing	HIETZING	XIII.	13	1130	...	8
22	BEZIRKSGRENZEOGD.9289	Rudolfsheim-Fünfhaus	15	XV	15., Rudolfsheim-Fünfhaus	XV. Rudolfsheim-Fünfhaus	RUDOLFSHEIM-FÜNFHAUS	XV.	15	1150	...	14
10	BEZIRKSGRENZEOGD.9300	Penzing	14	XIV	14., Penzing	XIV. Penzing	PENZING	XIV.	14	1140	...	15
2	BEZIRKSGRENZEOGD.9292	Josefstadt	8	VIII	8., Josefstadt	VIII. Josefstadt	JOSEFSTADT	VIII.	08	1080	...	5

5 rows × 32 columns

Table 10: Cluster 1 - just right

CLUSTER 2: MORE PHARMACIES NEEDED

The third and biggest cluster contains those districts where more pharmacies will be needed until the year 2034, either due to population growth or a current lack of supply.

	id	NAMEK	BEZNR	BEZ_RZ	NAMEK_NUM	NAMEK_RZ	NAMEG	LABEL	BEZ	DISTRICT_CODE	...	Pharmacies2019
12	BEZIRKSGRENZEOGD.9302	Donaustadt	22	XXII	22., Donaustadt	XXII. Donaustadt	DONAUSTADT	XXII.	22	1220	...	25
16	BEZIRKSGRENZEOGD.9283	Favoriten	10	X	10., Favoriten	X. Favoriten	FAVORITEN	X.	10	1100	...	28
9	BEZIRKSGRENZEOGD.9299	Brigittenau	20	XX	20., Brigittenau	XX. Brigittenau	BRIGITTENAU	XX.	20	1200	...	12
15	BEZIRKSGRENZEOGD.9282	Simmering	11	XI	11., Simmering	XI. Simmering	SIMMERING	XI.	11	1110	...	12
13	BEZIRKSGRENZEOGD.9303	Floridsdorf	21	XXI	21., Floridsdorf	XXI. Floridsdorf	FLORIDSDORF	XXI.	21	1210	...	27

5 rows × 32 columns

Table 11: Cluster 2 - more pharmacies needed

Looking at the clusters found by the Kmeans algorithm, it should be faily easy to derive adequate strategies for each of them to address the challenges ahead.

This concludes the methodology section and we now move on to reviewing the results.

RESULTS

During the exploratory analysis it became increasingly clear that the districts of Vienna are going to be experiencing different developments regarding population growth. While some districts are going to shrink, other are thriving and growing in population.

For city developers, it is vital to know where these developments are going, since the pharmacy approval process is quite bureaucratic and time consuming. To sum up the results so far, please inspect the following bar plot:

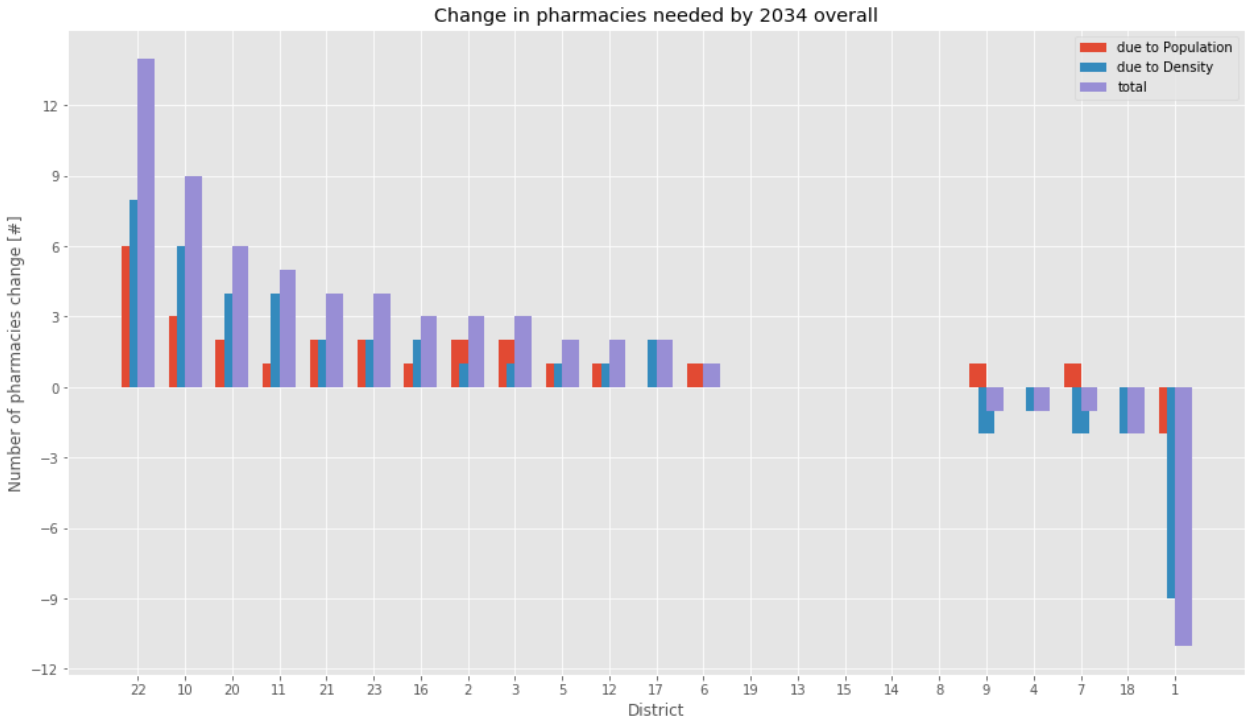

Figure 15: Change in pharmacies needed by 2034 overall

To ease analysis, we have sorted the districts by the predicted change in number of pharmacies until the year 2034. Thanks to this, it is easy to see that there are 3 types of districts:

- Those that are going to need more pharmacies (Cluster 0)
- Those that require no change (Cluster 1)
- Those that might do with fewer pharmacies (Cluster 2)

In the plot above, we have split the total increase/decrease needed into its components, change due to population and change due to density. While in most districts the sign of change is identical, there are some interesting things to note in districts requiring fewer pharmacies overall in the future. In some of them, the increased need due to population increase is mostly cancelled out by the already high density, while in central district 1 there seems to be excess supply paired with shrinking population.

In our last map plot below, we can see quite a clear picture of the 3 clusters. While the central districts of Vienna might need fewer pharmacies in the future, most districts in the south and north-east are going to need quite a number of extra pharmacies by the year 2034. Finding locations for these, however, should prove to be less of problem compared to the cramped situation in the center of Vienna.

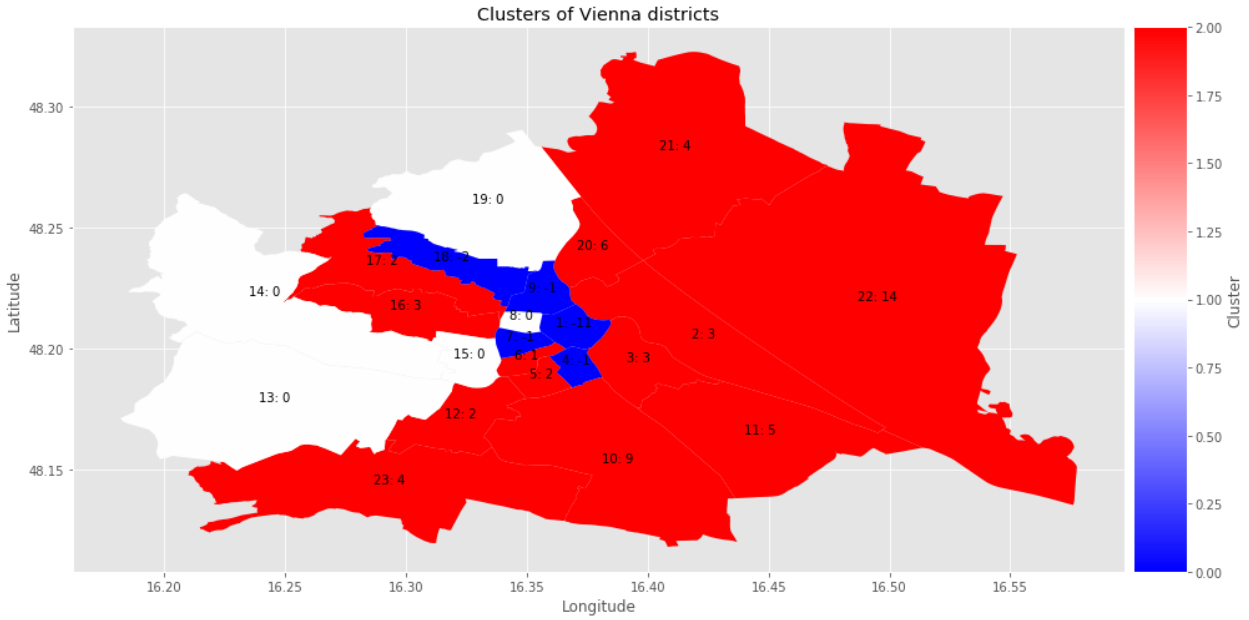

Figure 16: Strategic clusters of districts in Vienna

DISCUSSION

Vienna is the biggest and capital city of Austria, having a long history of growth and city development. It is no surprise that the distribution of pharmacies is not at all even across the city. As we have seen in the results section, there seem to be a lot more pharmacies in district 1 right in the center of Vienna. To explain this, it is helpful to know that this central district is where the majority of shopping and commerce is done, in addition to it being very well connected through public transport.

Looking at the pure numbers, one might think that the city administration should scramble to get new pharmacies to open up in the outskirts of the city. While this is certainly true and needs to be addressed, at the same time people in Vienna have no problem using public transport to get anywhere in the city, including pharmacies. Only when the number of available pharmacies is greatly below the numbers we calculated above (i. e. district 22), things might become problematic due to needless traffic and long journeys across the city just for some medicine.

While out of scope for this analysis, the current debate whether doctors should be allowed to sell medicine in their offices might become interesting for the future. After all, these extra locations where patients are already going to get treatment could remove part of the need for extra pharmacies, especially in the outskirts of the city.

CONCLUSION

Planning demand for certain services in big city like Vienna has always been challenging. Thanks to modern data sources like Foursquare and various Open Data initiatives, it has however become a bit easier. We managed to answer the question of how many pharmacies are going to be needed by the year 2034 using free data, free tools and Data Science methodology alone. As a result, stakeholders interested in knowing what to plan for in next 15 years hold a great analysis in their hands that can easily be detailed further, i. e. to find specific locations where we have determined that a future need will develop.