Capstone Project - Coffee Point in Kyiv, Ukraine (Week 2)

Table of contents

Introduction: Business Problem

In this project we will try to find an optimal location for a coffee point (small coffee shop). Specifically, this report will be targeted to stakeholders interested in opening an **Coffee Point** in **Kyiv**, Ukraine.

Since there are lots of coffee points in Kyiv we will try to detect **locations that are not already crowded with coffee points**. We are also particularly interested in **locations where may be some traffic makers: metro station, shopping center, cinema**. We would also prefer locations **as close to city center as possible**, assuming that first two conditions are met.

We will use our data science powers to generate a few most promissing neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data

Based on definition of our problem, factors that will influence our decision are:

- number of existing coffee points in the neighborhood
- number of and distance to coffee points in the neighborhood, if any
- distance of neighborhood from city center
- distance to traffic makers (metro station, shopping center, cinema)

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Geocoder Python package**
- number of coffee points and their type and location and availability of traffic makers in every neighborhood will be obtained using **Foursquare API**
- coordinate of Kyiv center will be obtained using **Geocoder Python package** of well known Kyiv location (Independence Square)

Neighborhood Candidates

Let's create latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is aprox. 12x12 killometers centered around Berlin city center.

Let's first find the latitude & longitude of Berlin city center, using specific, well known address and Google Maps geocoding API.

The geograpical coordinate of Independence Square, Kyiv are 50.45016285, 30.5241869112747.
Now let's create a grid of area candidates, equaly spaced, centered around city center and within ~6km from Alexanderplatz. Our neighborhoods will be defined as circular areas with a radius of 300 meters, so our neighborhood centers will be 600 meters apart.

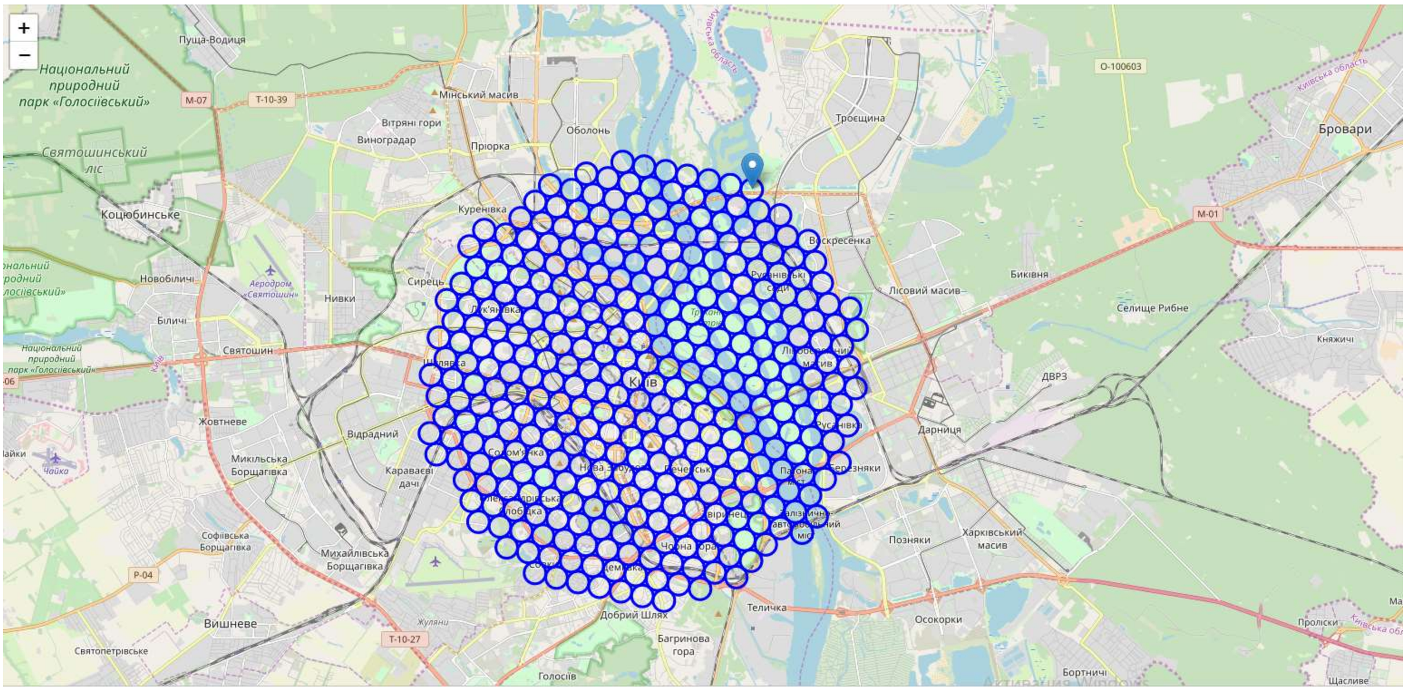
To accurately calculate distances we need to create our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on Folium map. So let's create functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

Coordinate transformation check

Kyiv center longitude=30.5241869112747, latitude=50.45016285
Kyiv center UTM X=1599455.1677679059, Y=5704824.015280547
Kyiv center longitude=30.5241869112747, latitude=50.45016284999999

Let's create a **hexagonal grid of cells**: we offset every other row, and adjust vertical row spacing so that **every cell center is equally distant from all it's neighbors**.

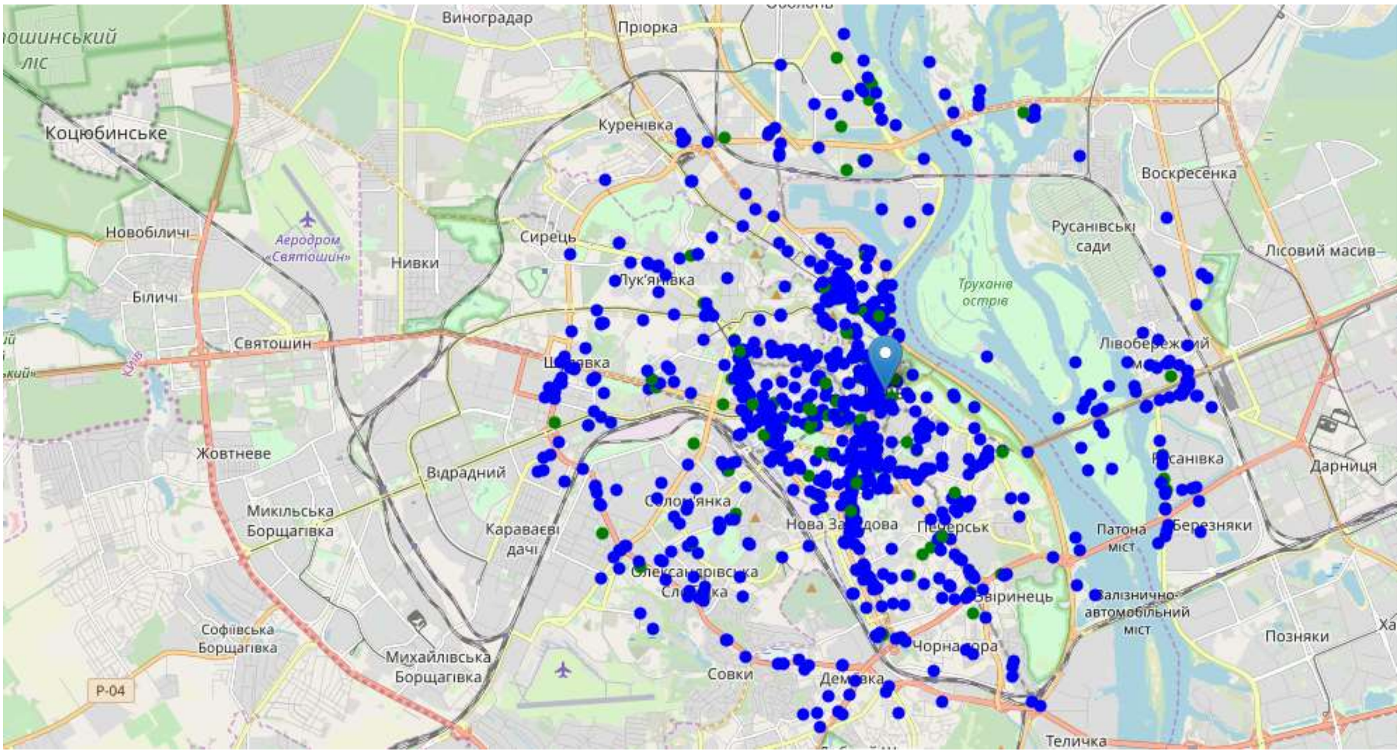
364 candidate neighborhood centers generated.
Let's visualize the data we have so far: city center location and candidate neighborhood centers:



Looking good. Let's now place all this into a Pandas dataframe.

	Address	Latitude	Longitude	X	Y	Distance from center
0	12, Яблунева вулиця, Совки, Солом'янський райо...	50.403967	30.483188	1.597655e+06	5.699108e+06	5992.495307
1	10а, Краматорська вулиця, Ширма, Голосіївський...	50.402856	30.491324	1.598255e+06	5.699108e+06	5840.376700
2	17, Гориста вулиця, Ширма, Голосіївський район...	50.401746	30.499460	1.598855e+06	5.699108e+06	5747.173218
3	43, Деміївська вулиця, Деміївка, Голосіївський...	50.400634	30.507595	1.599455e+06	5.699108e+06	5715.767665
4	7, Віктора Забіли вулиця, Деміївка, Голосіївсь...	50.399522	30.515729	1.600055e+06	5.699108e+06	5747.173218
5	Добрий Шлях, Голосіївський район, Київ, 03208	50.398410	30.523863	1.600655e+06	5.699108e+06	5840.376700
6	27, Науки проспект, Деміївка, Голосіївський ра...	50.397297	30.531997	1.601255e+06	5.699108e+06	5992.495307
7	Кадетський Гай вулиця, Олександрівська Слобідк...	50.410135	30.472486	1.596755e+06	5.699628e+06	5855.766389
8	20, Червоний провулок, Совки, Солом'янський ра...	50.409025	30.480624	1.597355e+06	5.699628e+06	5604.462508
9	20, Ясна вулиця, Совки, Солом'янський район, К...	50.407915	30.488761	1.597955e+06	5.699628e+06	5408.326913

Let's now see all the collected restaurants in our area of interest on map, and let's also show Italian restaurants in different color.



Looking good. So now we have all the caffees in area within few kilometers from Independence Square, and we know some traffic makers in our locations! This concludes the data gathering phase - we're now ready to use this data for analysis to produce the report on optimal locations for a new coffee point!

Methodology

In this project we will direct our efforts on detecting areas of Kyiv that have low caffee density, particularly those near of some traffic makers. We will limit our analysis to area ~6km around city center.

In first step we have collected the required **data: location and type (category) of every caffee within 6km from Kyiv center** (Independance Square). We have also **identified traffic makers** (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of '**caffees density**' across different areas of Kyiv - we will use **heatmaps** to identify a few promising areas close to center with low number of restaurants in general (*and* traffic makers in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations with **no more than two caffees in radius of 250 meters**, and we want locations **near traffic makers in radius of 400 meters**. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Analysis

Let's perform some basic explanatory data analysis and derive some additional info from our raw data. First let's count the **number of caffees in every area candidate**:

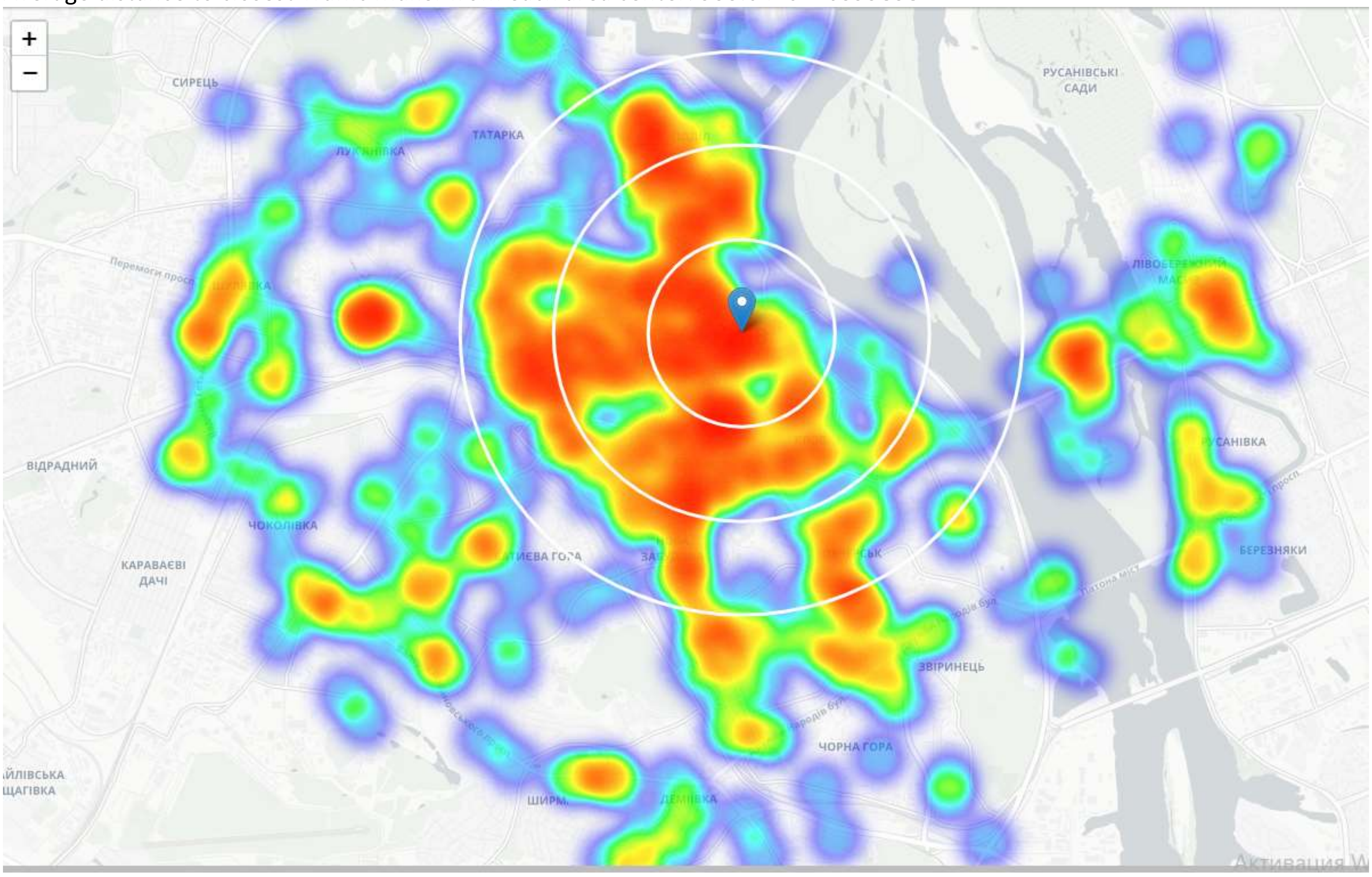
Average number of caffees in every area with radius=300m: 2.217032967032967

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area
0	12, Яблунева вулиця, Совки, Солом'янський райо...	50.403967	30.483188	1.597655e+06	5.699108e+06	5992.495307	0
1	10a, Краматорська вулиця, Ширма, Голосіївський...	50.402856	30.491324	1.598255e+06	5.699108e+06	5840.376700	0
2	17, Гориста вулиця, Ширма, Голосіївський район...	50.401746	30.499460	1.598855e+06	5.699108e+06	5747.173218	0
3	43, Деміївська вулиця, Деміївка, Голосіївський...	50.400634	30.507595	1.599455e+06	5.699108e+06	5715.767665	3
4	7, Віктора Забіли вулиця, Деміївка, Голосіївсь...	50.399522	30.515729	1.600055e+06	5.699108e+06	5747.173218	2
5	Добрий Шлях, Голосіївський район, Київ, 03208	50.398410	30.523863	1.600655e+06	5.699108e+06	5840.376700	0
6	27, Науки проспект, Деміївка, Голосіївський ра...	50.397297	30.531997	1.601255e+06	5.699108e+06	5992.495307	0
7	Кадетський Гай вулиця, Олександрівська Слобідк...	50.410135	30.472486	1.596755e+06	5.699628e+06	5855.766389	0
8	20, Червоний провулок, Совки, Солом'янський ра...	50.409025	30.480624	1.597355e+06	5.699628e+06	5604.462508	0
9	20, Ясна вулиця, Совки, Солом'янський район, К...	50.407915	30.488761	1.597955e+06	5.699628e+06	5408.326913	1

OK, now let's calculate the **distance to nearest traffic maker from every area candidate center** (not only those within 300m - we want distance to closest one, regardless of how distant it is)

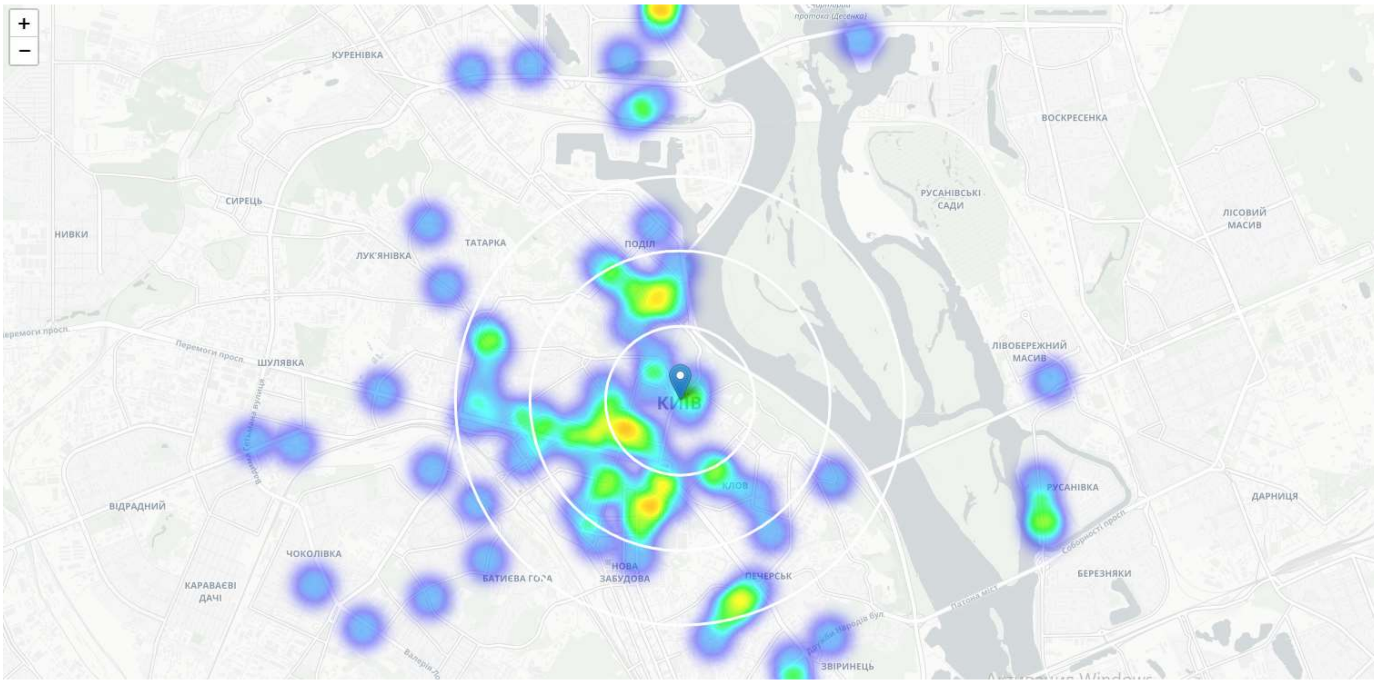
	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area	Distance to Italian restaurant
0	12, Яблунева вулиця, Совки, Солом'янський райо...	50.403967	30.483188	1.597655e+06	5.699108e+06	5992.495307	0	2520.099113
1	10a, Краматорська вулиця, Ширма, Голосіївський...	50.402856	30.491324	1.598255e+06	5.699108e+06	5840.376700	0	2562.828972
2	17, Гориста вулиця, Ширма, Голосіївський район...	50.401746	30.499460	1.598855e+06	5.699108e+06	5747.173218	0	2109.001775
3	43, Деміївська вулиця, Деміївка, Голосіївський...	50.400634	30.507595	1.599455e+06	5.699108e+06	5715.767665	3	1745.761905
4	7, Віктора Забіли вулиця, Деміївка, Голосіївсь...	50.399522	30.515729	1.600055e+06	5.699108e+06	5747.173218	2	1538.662008
5	Добрий Шлях, Голосіївський район, Київ, 03208	50.398410	30.523863	1.600655e+06	5.699108e+06	5840.376700	0	1551.540177
6	27, Науки проспект, Деміївка, Голосіївський ра...	50.397297	30.531997	1.601255e+06	5.699108e+06	5992.495307	0	1779.627227
7	Кадетський Гай вулиця, Олександрівська Слобідк...	50.410135	30.472486	1.596755e+06	5.699628e+06	5855.766389	0	1543.900931
8	20, Червоний провулок, Совки, Солом'янський ра...	50.409025	30.480624	1.597355e+06	5.699628e+06	5604.462508	0	1944.998962
9	20, Ясна вулиця, Совки, Солом'янський район, К...	50.407915	30.488761	1.597955e+06	5.699628e+06	5408.326913	1	2262.900420

Average distance to closest Traffic Maker from each area center: 933.0416226399858



Looks like a few pockets of low restaurant density closest to city center can be found **south, south-west and west from Independance Square**.

Let's create another heatmap map showing **heatmap/density of Traffic Makers** only.



This map is not so 'hot' (Traffic makers represent a subset of ~15% of all restaurants in Kyiv) but it also indicates higher density of existing Traffic makers directly north and west from Independance Square, with closest pockets of **high Traffic makers density positioned east, south-east and south from city center**.

Based on this we will now focus our analysis on areas *south-west, south, south-east and east from Kyiv center* - we will move the center of our area of interest and reduce it's size to have a radius of **2.5km**. This places our location candidates mostly in boroughs **Lipky and Shevchenkovsky** (another potentially interesting borough is **Klov** with large low caffee density north-east from city center, however this borough is less interesting to stakeholders as it's mostly residental and less popular with tourists).

Lipky and Shevchenkovsky

Analysis of popular travel guides and web sites often mention Lipky and Shevchenkovsky as beautifull, interesting, rich with culture, 'hip' and 'cool' Kyiv neighborhoods popular with tourists and loved by Kyivites.

"Bold and brazen, Lipky's creative people, places, and spaces might challenge your paradigm." Tags: Nightlife, Artsy, Dining, Trendy, Loved by Berliners, Great Transit (airbnb.com)

"Lipky has long been revered for its diverse cultural life and as a part of Kyiv where alternative lifestyles have flourished. Envisioning the glamorous yet gritty nature of Kyiv often conjures up scenes from this neighbourhood, where cultures, movements and artistic flare adorn the walls of building and fills the air. Brimming with nightclubs, street food, and art galleries, Lipky is the place to be for Kyiv's young and trendy." (theculturetrip.com)

"Imagine an art gallery turned inside out and you'll begin to envision Shevchenkovsky. Single walls aren't canvases for creative works, entire buildings are canvases. This zealously expressive east Kyiv neighborhood forgoes social norms" Tags: Artsy, Nightlife, Trendy, Dining, Touristy, Shopping, Great Transit, Loved by Kyivites (airbnb.com)

"As anyone from Lipky will tell you, this district is not just the coolest in Kyiv, but the hippest location in the entire universe. Lipky has long been famed for its diverse cultural life, its experimental alternative lifestyles and the powerful spell it exercises on young people from across Ukraine. In 2001, Lipky and Shevchenkovsky were merged to form one administrative borough. When it comes to club culture, Shevchenkovsky is now out in front – with southern Shevchenkovsky particularly ranked as home to the highest density of clubs in the city." (visitkyiv.ua)

Popular with tourists, alternative and bohemian but booming and trendy, relatively close to city center and well connected, those boroughs appear to justify further analysis.

Let's define new, more narrow region of interest, which will include low-restaurant-count parts of Lipky and Shevchenkovsky closest to Independance Square.

Not bad - this nicely covers all the pockets of low restaurant density in Lipky and Shevchenkovsky closest to Kyiv center.

Let's also create new, more dense grid of location candidates restricted to our new region of interest (let's make our location candidates 100m appart).

2261 candidate neighborhood centers generated.
OK. Now let's calculate two most important things for each location candidate: **number of restaurants in vicinity** (we'll use radius of **250 meters**) and **distance to closest Italian restaurant**.

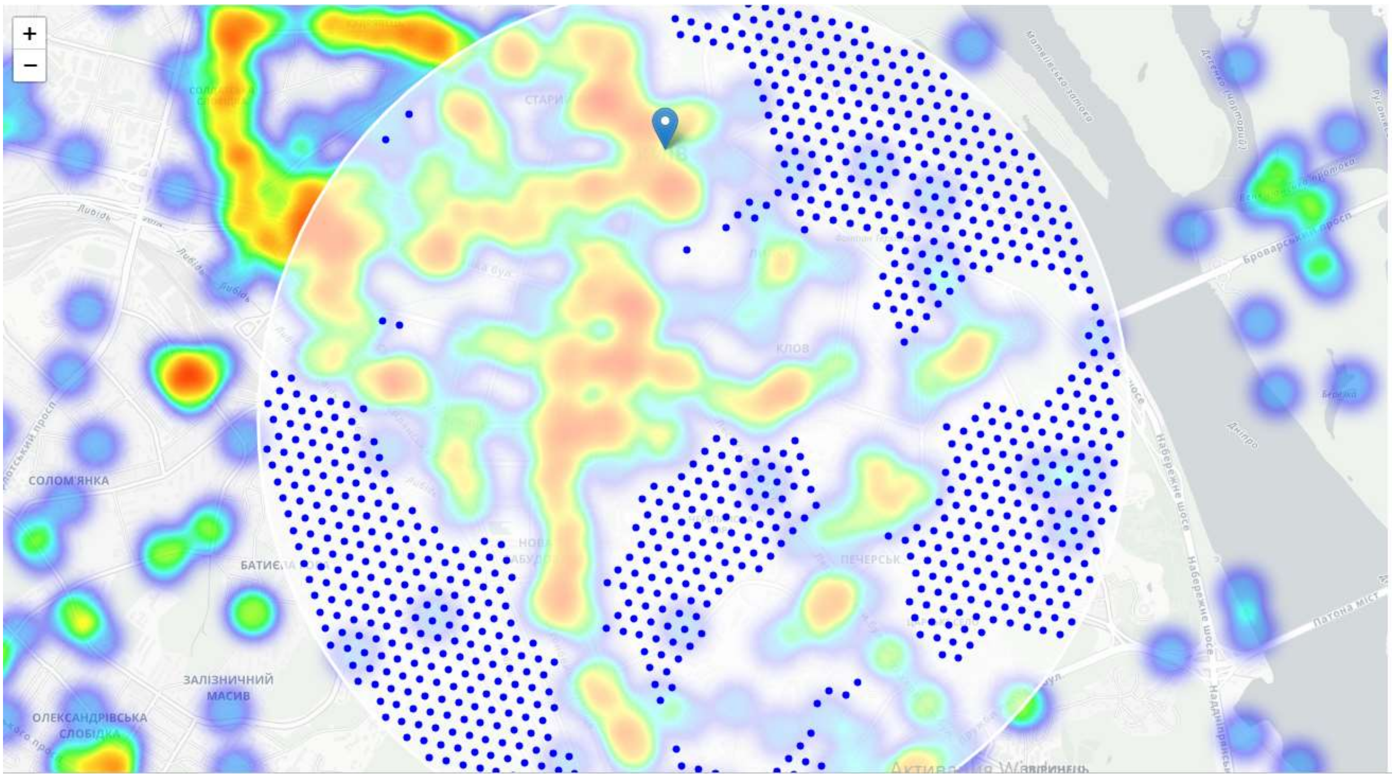
Generating data on location candidates... done.

	Latitude	Longitude	X	Y	Restaurants nearby	Distance to Italian restaurant
0	50.414668	30.518673	1.599905e+06	5.700824e+06	1	462.515749
1	50.414482	30.520029	1.600005e+06	5.700824e+06	1	374.904666
2	50.416438	30.511465	1.599355e+06	5.700911e+06	0	1008.540903
3	50.416253	30.512822	1.599455e+06	5.700911e+06	0	913.119622
4	50.416067	30.514178	1.599555e+06	5.700911e+06	0	818.791875
5	50.415882	30.515534	1.599655e+06	5.700911e+06	0	725.984039
6	50.415696	30.516890	1.599755e+06	5.700911e+06	1	635.362507
7	50.415511	30.518246	1.599855e+06	5.700911e+06	1	548.012962
8	50.415326	30.519603	1.599955e+06	5.700911e+06	1	465.779881
9	50.415140	30.520959	1.600055e+06	5.700911e+06	2	391.897421

OK. Let us now **filter** those locations: we're interested only in **locations with no more than caffees in radius of 250 meters**, and **traffic maker in radius of 400 meters**.

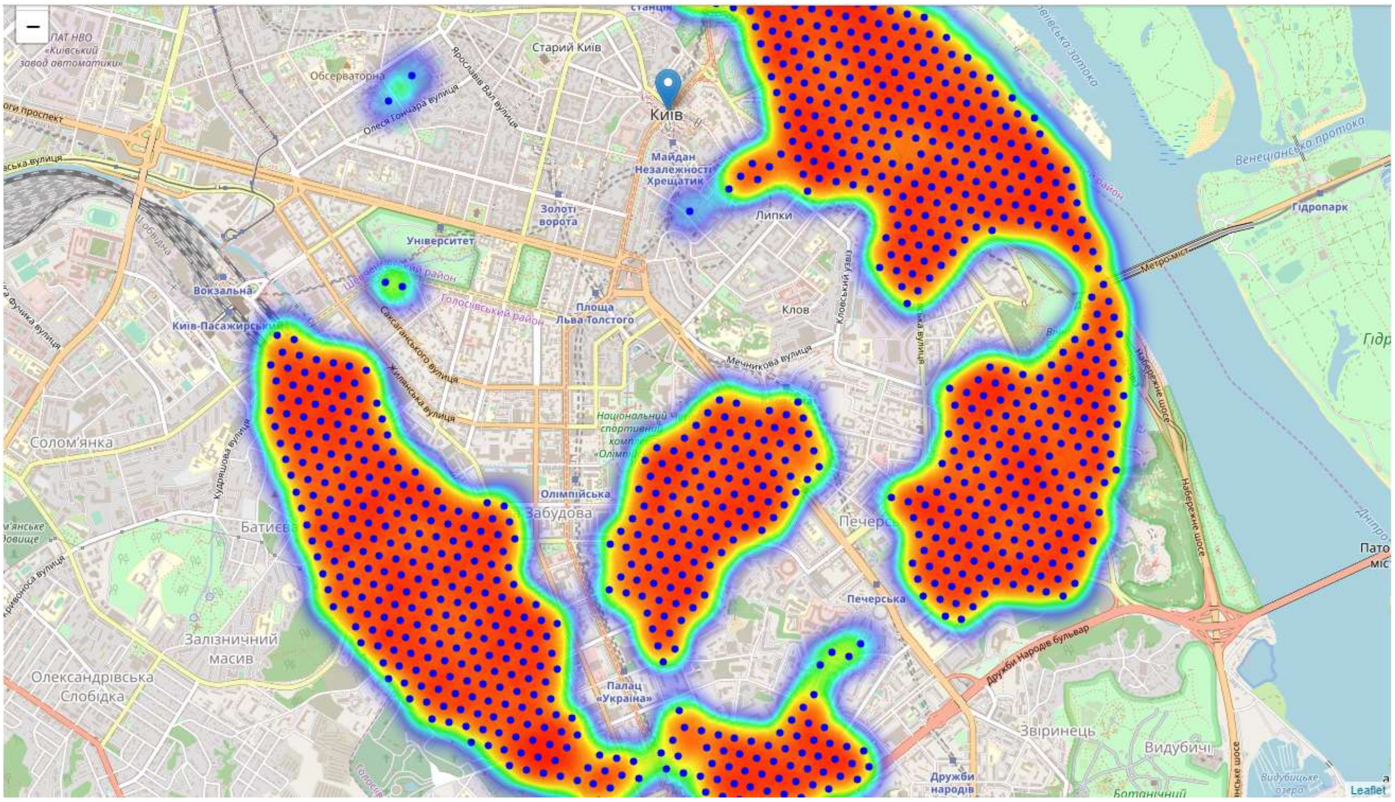
Locations with no more than two restaurants nearby: 1199
Locations with no Italian restaurants within 400m: 1158
Locations with both conditions met: 915

Let's see how this looks on a map.



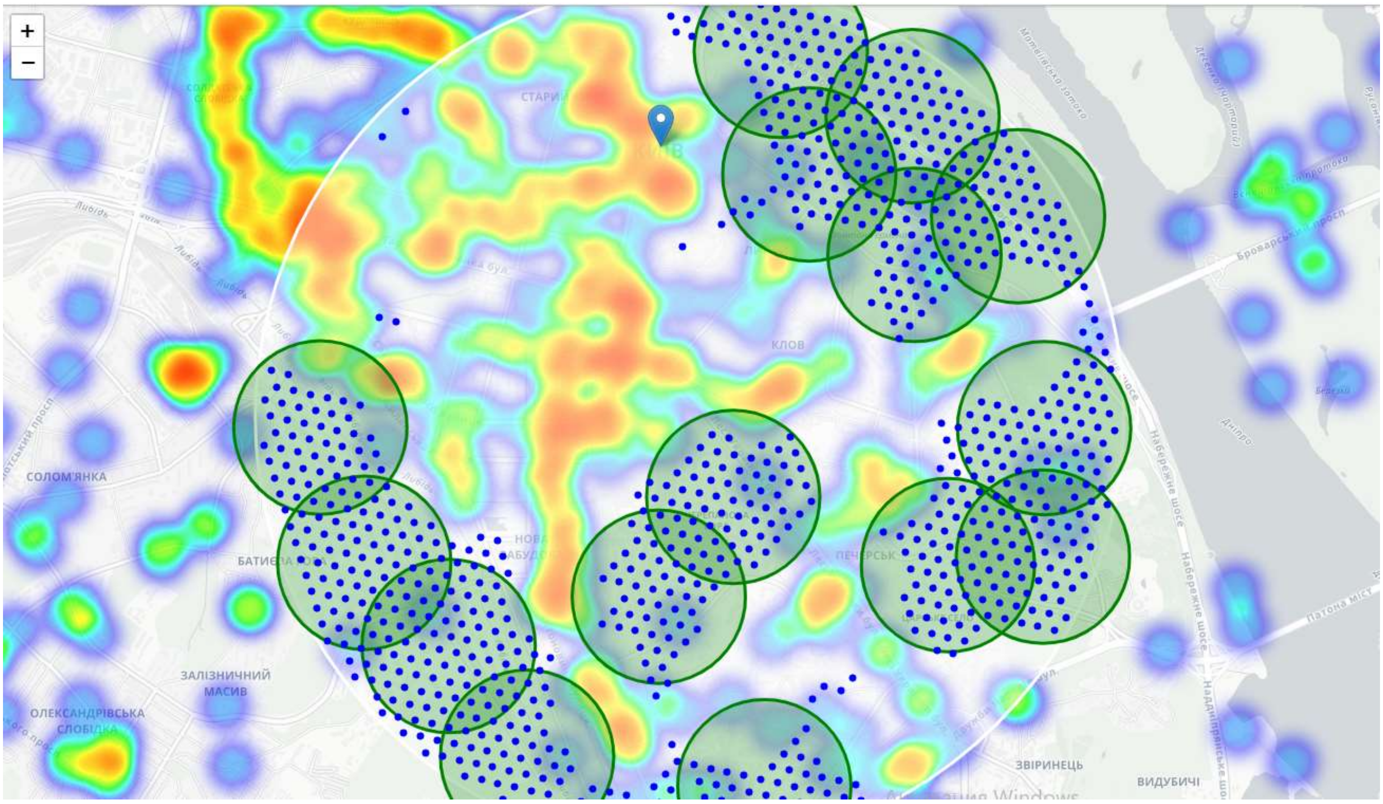
Looking good. We now have a bunch of locations fairly close to Independance Square (mostly in Lipky and Shevchenkovsky and south-east corner of Mitte boroughs), and we know that each of those locations has no more than two coffee points in radius of 250m, and traffic maker closer than 400m. Any of those locations is a potential candidate for a new coffee point, at least based on nearby competition.

Let's now show those good locations in a form of heatmap:



Looking good. What we have now is a clear indication of zones with low number of coffee points in vicinity, and traffic makers at all nearby.

Let us now **cluster** those locations to create **centers of zones containing good locations**. Those zones, their centers and addresses will be the final result of our analysis.



Finaly, let's **reverse geocode those candidate area centers to get the addresses** which can be presented to stakeholders.

```
==============================================================
Addresses of centers of areas recommended for further analysis
==============================================================

БЦ «Протасів» (блок А), 2/1, Миколи Грінченка вулиця, Деміївка, Голосіївський район, Київ, 03028 => 3.2km from Independance Square
6А, Старонаводницька вулиця, Царське Село, Печерськ, Печерський район, Київ, 01196 => 3.0km from Independance Square
Набережне шосе, Печерськ, Печерський район, Київ, 01008 => 0.9km from Independance Square
Вал люнета Васильківського укріплення, Миколи Раєвського вулиця, Царське Село, Печерськ, Печерський район, Київ, 01042 => 3.8km from Independance Square
Городня вулиця, ЖК  Альпійське містечко, Батиєва гора, Солом’янський район, Київ, 03038 => 3.0km from Independance Square
Набережне шосе, Печерськ, Печерський район, Київ, 01008 => 2.1km from Independance Square
Набережне шосе, Печерськ, Печерський район, Київ, 01008 => 1.5km from Independance Square
3, Лабораторний провулок, Черепанова гора, Печерськ, Печерський район, Київ, 01901 => 2.6km from Independance Square
Києво-Печерська цитадель, Памви Беринди провулок, Царське Село, Печерськ, Печерський район, Київ, 01015 => 2.8km from Independance Square
Миколи Грінченка вулиця, Солом’янський район, Київ, 03093 => 3.6km from Independance Square
Льва Толстого, Батиєва гора, Солом’янський район, Київ, 03031 => 2.6km from Independance Square
Башта №7, Київська фортеця, 1, Івана Мазепи вулиця, Царське Село, Печерськ, Печерський район, Київ, 01010 => 1.6km from Independance Square
32, Радіальна вулиця, Царське Село, Звіринець, Печерський район, Київ, 01015 => 3.3km from Independance Square
Пам'ятник  Леся Українка, Михайла Грушевського вулиця, Черепанова гора, Печерськ, Печерський район, Київ, 01008 => 0.9km from Independance Square
Київська фортеця, Лабораторний провулок, Черепанова гора, Печерськ, Печерський район, Київ, 01901 => 2.1km from Independance Square
```

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of caffees and traffic makers nearby, all zones being fairly close to city center (all less than 4km from Independance Square, and about half of those less than 2km from Independance Square). Although zones are shown on map with a radius of ~500 meters (green circles), their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential coffee points locations. Most of the zones are located in Lipky and Shevchenkovsky boroughs, which we have identified as interesting due to being popular with tourists, fairly close to city center and well connected by public transport.

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of caffees and traffic makers nearby, all zones being fairly close to city center (all less than 4km from Independance Square, and about half of those less than 2km from Independance Square). Although zones are shown on map with a radius of ~500 meters (green circles), their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential coffee points locations. Most of the zones are located in Lipky and Shevchenkovsky boroughs, which we have identified as interesting due to being popular with tourists, fairly close to city center and well connected by public transport.

Results and Discussion

Our analysis shows that although there is a great number of coffee point in Kyiv, Ukraine (~2000 in our initial area of interest which was 12x12km around Independence Square), there are pockets of low coffee point density fairly close to city center. Highest concentration of coffee points was detected north and west from Independence Square, so we focused our attention to areas south, south-east and east, corresponding to boroughs Lipky, Shevchenkovsky and south-east corner of central Mitte borough. Another borough was identified as potentially interesting (north-east from Independence Square), but our attention was focused on Lipky and Shevchenkovsky which offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics *and* a number of pockets of low coffee point density.

After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Independence Square) we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than two coffee point in radius of 250m and those with an traffic makers closer than 400m.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 15 zones containing largest number of potential new coffee point locations based on number of and distance to existing venues - both coffee points in general. This, of course, does not imply that those zones are actually optimal locations for a new coffee point! Purpose of this analysis was to only provide info on areas close to Kyiv center but not crowded with existing coffee points - it is entirely possible that there is a very good reason for small number of coffee points in any of those areas, reasons which would make them unsuitable for a new coffee point regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

Conclusion

Purpose of this project was to identify Kyiv areas close to center with low number of coffee points in order to aid stakeholders in narrowing down the search for optimal location for a new coffee point. By calculating coffee point density distribution from Foursquare data we have first identified general boroughs that justify further analysis (Lipky, Shevchenkovsky), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby coffee points. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decission on optimal coffee point location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.