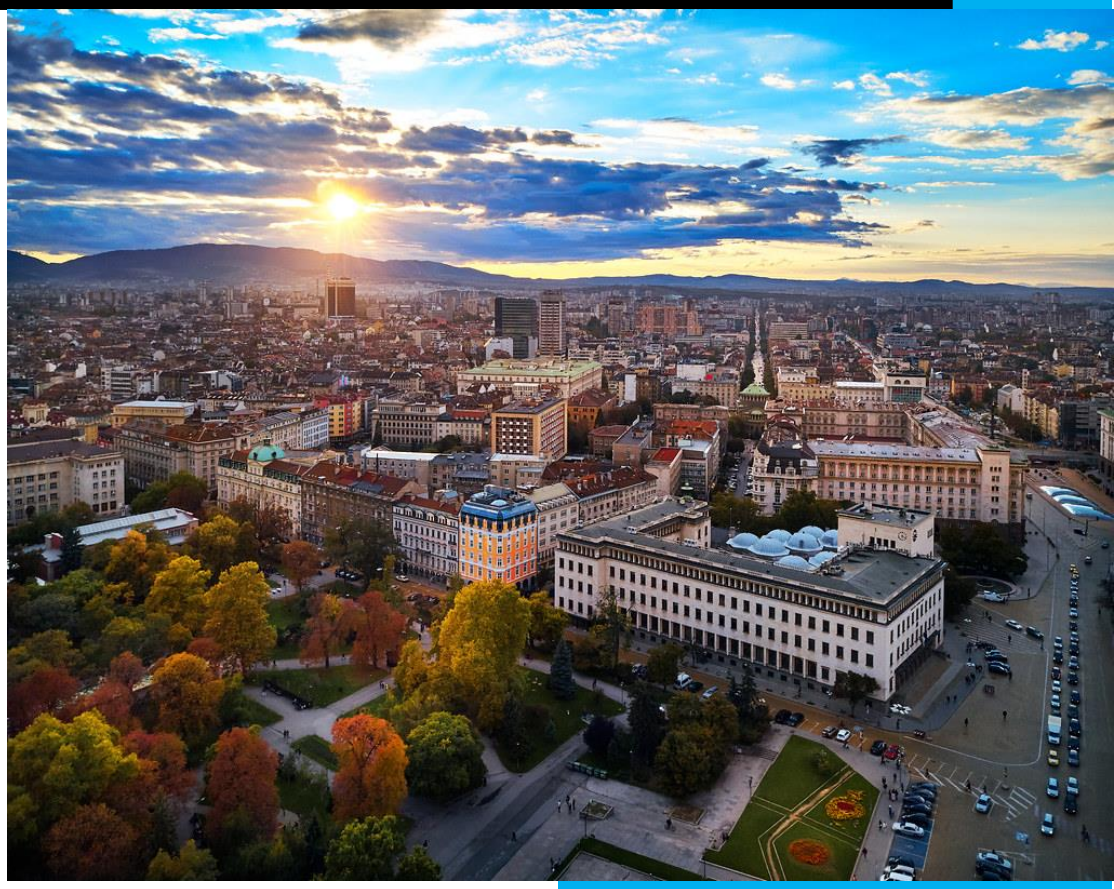
2020

Hipster Neighborhoods of Sofia



Anton Gerunov

3/7/2020

Contents

Introduction

The increasing standard of living and the generational shifts globally have produced a plethora of varied lifestyles that are reflected in the architecture, environmental planning, consumption habits, and social expression. One of the most distinct lifestyles is that of a so-called "hipster" that is often associated with refined consumption, vintage outlook, and a general ironic attitude towards established social norms and institutions. This has served as a magnet for many across different generations, and particularly Millennials, and has been widely adopted.

The interest in a hipster lifestyle also leads to the formation of specific spatially-delimited communities that are conducive for this lifestyle and feature its outlook, feel, and places and modes of consumption. Those are the so-called "hipster neighborhoods" that attract disproportionate interest from buyers and renters that identify or are sympathetic with this subculture. The rise in interest, and thus market demand, is also reflected in the rise of prices and change of community composition that partly fuels the process of gentrification.

This is observed in many cities globally but it particularly fascinating in the case of emerging economies as this proceeds both rapidly and chaotically. For this project we take the case of Bulgaria's capital city Sofia and try to identify its hipster neighborhoods and see whether they remain invariant or shift with the passage of time.

Problem Description

The desirability of hipster neighborhoods poses particular challenges for both the supply and the demand side of the housing market. Since it is difficult to define what a hipster neighborhood is there is significant information asymmetry between buyers and sellers and thus pricing cannot be set appropriately to reflect market conditions. Furthermore, the lack of sufficient data may lead to mislabeling a neighborhood as "not hipster" (i.e. undesirable) and thus impede its future growth.

The key problem therefore is:

- **There is no formal definition of hipster neighborhoods using objective criteria (i.e. data) and thus confusion reigns in the market. This results in improperly put price premia, leading to some suboptimal outcomes.** Sometimes sellers suffer as they fail to reach a fair price for a property in a hip cluster; and sometimes buyers suffer as they overpay for a supposedly hip place that is actually not so.

To this end we need to **complete a number of tasks**:

1. Study definitions of "hipster" and find convergence
2. Operationalize the term so that it can be investigated formally
3. Collect necessary data as per the term operationalization
4. Using data identify and map hipster neighborhoods
5. Investigate whether those communities shift over time (i.e. "a temporal map of coolness")
6. Outline most hipster neighborhoods in the city of Sofia to inform both buyer and sellers of the possible price premium

The problem is a classic asymmetric information one and can thus be fruitfully solved using data science approaches.

Background and Related Work

What makes a neighborhood hipster has captured the imagination of many observers and as a topic found its distinguished place in both popular and academic discourse. As far back as 2012, *The Atlantic* (Doll, 2012) identified the overwhelming use (and mis-use) of the term and the plethora of popular ways to describe it. The closest to being a data-driven one is the analysis, prepared by travel website *Travel + Leisure* (Hunt, 2013) – *America's Best Cities for Hipsters*. It assessed each area's walkability, number of coffee shops, food trucks, farmer markets, locally owned bars and restaurants, residents in artistic occupations and how often hipster-related words appeared on a given neighborhood's web pages. While an interesting first foray, this analysis includes a very narrow set of indicators, and only focuses on US-American cities. This provides ample opportunity to step and build upon this approach, significantly expanding it, and applying it to a novel location such as Sofia.

To do this, we need to better operationalize the term "hipster". While this has been amply researched such as in Jake Kinzey's book *The Sacred and the Profane* (Kinzey, 2010), complete consensus has failed to emerge. While authors such as Schiermer (2014), in his *Acta Sociologica* article *Late-modern Hipsters: New Tendencies in Popular Culture* focuses on behavioral and sociological manifestation of hipster culture at the individual level, many authors instead focus on externa and environmental signifiers of this community. Over the past decade, the academic community has observed somewhat of a convergence in views that we will use to operationalize the concept of hipster.

For instance, Cronin et al. (2012) focus very much on food consumption as one defining facet of the hipster identify. More specifically, they outline three main practices that hipsters use to distinguish themselves from the mainstream and assert their lifestyle in a narrow and bounded community, namely:

- Preference for vegetarian meals (or options);
- Choices of specific (craft) brands and avoidance of others (mainstream brands);
- Decommodification of the process of consuming food.

The authors further point out that those practice serve as points of resistance against the mainstream culture and prevent the co-optation of hipsters into the more established and even mundane lifestyles.

Michael (2015) also focuses on consumption choices that are in opposition to the mainstream - particularly the fascination with vintage goods, kitsch retro styles and artisanal production. Hubbard (2016) adds to this definition further venues that are symptomatic to a hipster identify such as authentic coffee shops, vintage stores, and bars that serve microbrews. However, while most authors and indeed the general public are rather positive of hipsters, Hubbard (2016) sees them as agents that seek to colonialize ethnic neighboorhoods, gentrify them, and displace the locals through economic means. This further validates the need of appropriately pricing hipster vs. non-hipster neighborhoods that this project undertakes. Finally, LeBesco and Naccarato (2015) also investigate how hipsters express themselves through their food practices which enables them to compete with other alternative communities and subcultures.

The key takeaways from the overview of the research on hipsters and their environment is **that hipster neighborhoods are distinguished by the following**:

Presence of a large variety of ethnic restaurants

- Venues with artistic functions (e.g. dance studios, galleries, etc.)
- Outlets conducive to activities associated with the subculture (e.g. yoga studios, photography studios)
- Coffee shops
- Local breweries (esp. micro-breweries)
- Shops that offer a variety of non-traditional food options (e.g. organic food stores)
- Places where vintage or retro items can be purchases, preferably at low prices (e.g. flea markets but not upscale antique shops)

We conjecture that the **larger the concentration of hipster-related venues in a given neighborhood, the more hipster** this neighborhood is.

Data

The business problem defined - the difficulty in identifying hipster neighborhoods - calls for a data-driven solution. The operationalization of the term naturally calls for a rich database of location and venue data that will help identify, group, and map the different types of neighborhoods. We now review each step of the data pipeline.

Analytic Approach

The business problem defined is largely exploratory. It builds upon previous work but additional efforts are needed to further operationalize and measure the concept under study (hipster neighborhoods). It will be useful to identify and map those neighborhoods and possibly cluster them together. Those trends need to also be studied over time and their dynamics - precisely measured. Given all this, **a descriptive approach will be most suitable** for this project. It will enable exploration and still ensure that insights can be gleaned through analysis of data.

Data Requirements

To appropriately address the problem at hand we need three distinct types of data:

- **Map data of Sofia** - we need to have an overview of the city and its geography and neighborhoods. This is available out-of-the-box in dedicated Python libraries and can be used directly for the analysis.
- **Neighborhood Coordinates** - to identify hipster neighborhoods we must first be able to identify and separate different neighborhoods in the city. There is a little caveat here that administrative borders of the neighborhoods include a number of regions each with its specific features. We thus need to find the latitude and longitude of those sub-regions, and not merely the administrative regions.
- **Data on venues** - since the concept of a hipster neighborhood is largely operationalized by the types and numbers of venues that are located in it, we need to have a rich and detailed database with venues and their coordinates.

Data Collection

The two major sources of data needed, and their possible sources are as follows:

- **Neighborhood Coordinates** - a sufficiently detailed database of neighborhood coordinates in Sofia are surprisingly hard to find. An extensive research could not find the necessary data in a structured form. The closest to that which is avaiable can be found on the website Guide Bulgaria. On the positive side this data is very detailed with 83 different neighborhood regions (or just neighborhoods from now on) and their respective coordinates. This level of granularity is much higher that admnistrative regions (only 24). On the negative side - this data is not structured and needs to be scraped from the site using a Python libary such as Beautiful Soup and then further processed into a pandas data frame.

- **Data on venues** - the Foursquare location data can be used to obtain a rich list of different venues around given coordinates. This is furthermore useful as the level of detail of the venue categories is very high with over a hundred different types of venues. Such a level of detail enables us to segment each venue as a "hipster" (e.g. flea market) or "non-hipster" (e.g. gas station or fast food) venue and thus identify neighborhoods with a higher concentration of the former. This data is easy to obtain from Foursquare's API using Python requests library after creating a Foursquare developer account to obtain the credentials. This data is then processed and put into a pandas data frame.

Data Understanding and Preparation

The data understanding phases helps answer the question of whether appropriate data has been collected. It thus includes activities to process and gain initial insight into the data. This is done particularly through visualizations and descriptive statistics. Here we will use univariate statistics but will also focus on mapping and creating . In terms of the data preparation, we will need to cleanse the data (being particularly careful about missing values), transform it so that it is suitable for the analytics phase, and then apply domain knowledge to possibly engineer new features.

More specifically, the data preparation will need to convert parsed html for Sofia coordinates into a data frame that can be used for further processing. Additionally, location data downloaded from Foursquare is obtained at the individual (venue level), whereas the unit of analysis is the neighborhood. Data will thus have to be aggregated at the neighborhood level and analyzed. Processing will also include sorting out hipster and non-hipster venues. Initial understanding will focus on Sofia's geography and its neighborhoods.

Modeling and Evaluation

The final phase of the analysis is the modeling and evaluation. This will focus on identifying hipster neighborhoods by means of highest number of hipster venues. We explore the snapshot of the distribution as well as the trends over time. Furthermore, by means of the k-clustering algorithm we group similar neighborhoods and investigate their specific characteristics by investigating the cluster centroids per hipster category. This pahse is aimed to answer the questions posed for the project by identifying hipster neighborhoods, outlining clyster of similar hipster locales and investigating stability over time.

Methodology

The exploratory nature of the this project means that the methodology used will largely focus on processing and visualizing data. The analytics process goes through its standard phases as follows:

- **Initial Preparation**
- **Obtaining Data**
- **Data Processing**
- **Visualization**
- **Applying Machine Learning Algorithm**
- **Results Interpretation**

Detailed steps and code follow and are structured in a way to ensure both transparency and replicability. The first step in data collection is scraping the website Guide Bulgaria to obtain granular location data for Sofia's neighborhoods. For this purpose we initially get the contents of the page itself and then use Beautiful Soup to parse it and only retain the necessary data. We then proceed to clean the database of missing observations (there is 1) and to convert the coordinates to their appropriate numeric format. We also further inspect the data frame itself to make sure the resultant data is suitable for further processing and analysis.

As a next step we proceed to visualize Sofia's neighborhoods in order to obtain a feeling of their geographical distribution and relative positions. To this end we use Nominatim to obtain the latitude and longitude of the city. We create a map of the city using the folium library. Markers, or indeed pins, for each neighborhood are added to the map and then it is displayed. This enables us to have an overview of this capital, together with a very granular idea of the neighborhoods that form it.

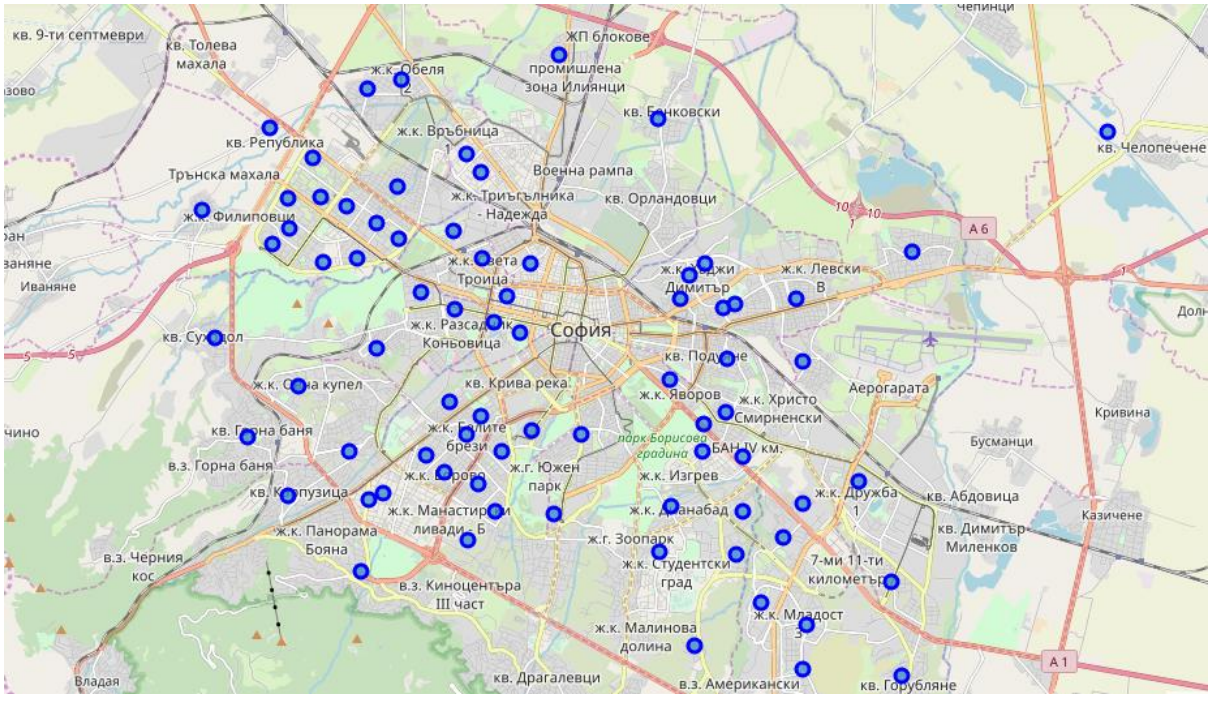


FIGURE 1: MAP OF SOFIA AND ITS NEIGHBORHOODS

The next step of the analytic process is to obtain data about the venues that can be found in each of the neighborhoods. To this end we use a Foursquare developed account and get the data through their API. We can further see all the venue categories in the dataset, displaying each unique type of venue that is found in the city. Among those some are hipster in nature, and some are not. The differentiation between the two will be one of our challenges.

Showing the total number of venues in a neighborhood is a useful first pass for getting an initial idea of concentrations across the different parts of the city. We count the total number of venues across each neighborhood and then visualize the top 10 neighborhoods with most places of interest. The Poligona neighborhood is most densely filled with locales - just a bit less than 60, and it is followed by Iztok, Studentski grad, Lozenets, and Beli Brezi. Those followers all have between 40 and 50 different venues.

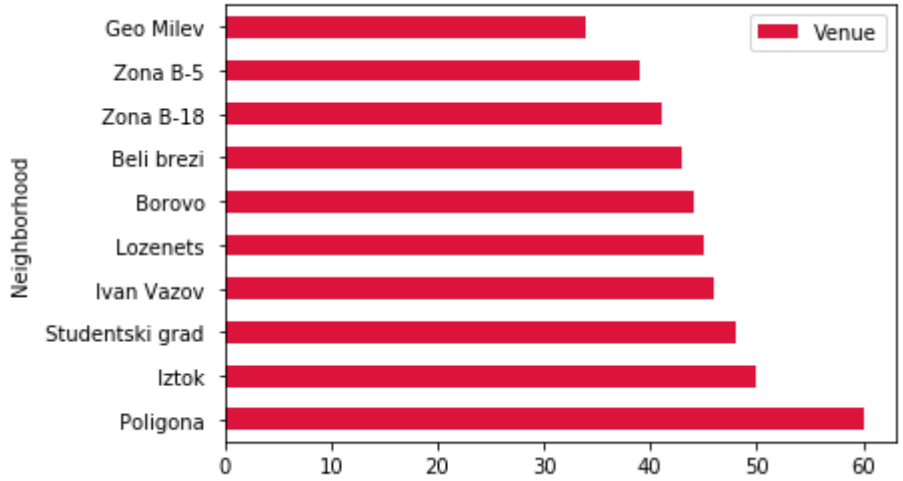


FIGURE 2: NEIGHBORHOODS WITH HIGHEST CONCENTRATION OF VENUES

Naturally, the total number of venues is hardly an indicator of how hipster a given space is - the neighborhood all needs to have the "right" venue in order to be elevated to the hipster status. Using the literature review and related work, we managed to **identify a number of distinguishing hipster-type of venues that are commonly associated with this subculture**. We mostly focused on the following:

- Presence of a large variety of ethnic restaurants
- Venues with artistic functions (e.g. dance studios, galleries, etc.)
- Outlets conducive to activites associated with the subculture (e.g. yoga studios, photography studios)
- Coffee shops
- Local breweries (esp. micro-breweries)
- Shops that offer a variety of non-traditional food options (e.g. organic food stores)
- Places where vintage or retro items can be purchases, preferably at low prices (e.g. flea markets but not upscale antique shops)

To be able to distinguish better, we can go through unique venue categories and only retain those that meet the above criteria. All those types are collected in a list called `hipster` that contains typical categories, including ethnic resturants, photgraphy and yoga studios, cafes, breweries, performing arts centers, flea markets, and many others.

We then count the number of hipster venues in each of Sofia's neighborhoods and reach a final list of hip places to live. **Top five most hipster neighborhoods in the city are**:

1. Iztok;
2. Zona B-18;
3. Ivan Vazov;
4. Poligona;
5. Lozenets.

Notably, the places with the most venues do not necessarily coincide with the place with the most hip venues, showing the importance of composition. High concentration of things to do does not always imply a high concentration of cool things to do.

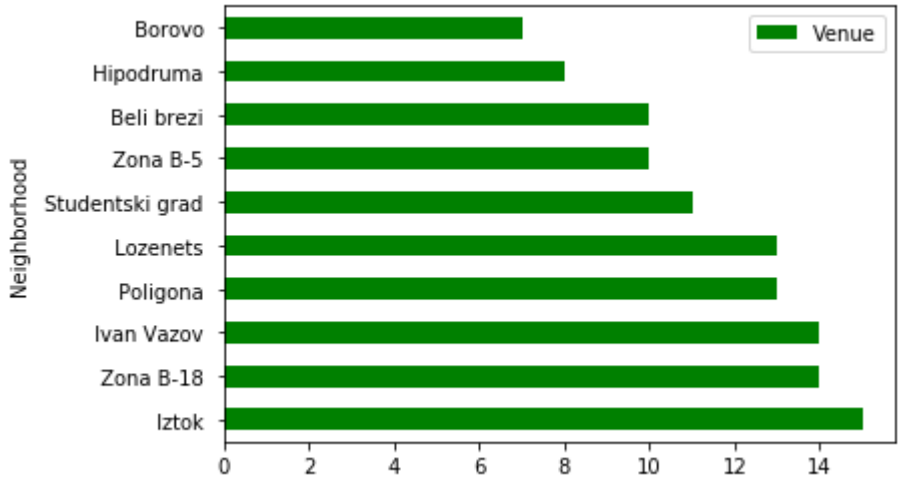


FIGURE 3: NEIGHBORHOODS WITH HIGHEST CONCENTRATION OF HIPSTER-RELATED VENUES

The current dataset contains only places where hipster-related venues are present. Note that initially we lookeds at 83 different quarters, but there are only 58 quarters which have at least one hipster-related venue. This means **that fully 25 neibghboorhoods have no tokens or venues, associated with the hipster subculture**.

Among the other 58, we note that they display significant differences among themselves. It thus makes sense to group them into clusters in such a way that each cluster is symptotatic of a different expression of hipster. **To do this we leverage the k-means algorithm since it is a useful way to from clusters based on average values of different group features**. This is particularly useful as we can work with average propensity of a given venue appearing in a given neighborhood. Furthermore, after the grouping is complete, we can investigate the cluster centroids, interpret them, and identify the type of hipster that this cluster has captured.

To this end we first prepare a data frame for the clustering taks - hip_cluster that contains only hipster-related venues. Then we do one hot encoding, thus defining a binary variable of whether a certain venue appears. Finally, we take the mean of this binary variable, giving an effective probability of this specific venue appearing in the specific neighborhood. A key questions is what is the optimum number of k that is meaningful in this case. After experimentation, it seems that k of 3 or less generates very dense and generic clusters, while k of 5 or more generates very tiny clusters (of one or two neighborhoods). This is why we set the requisite value to k = 4.

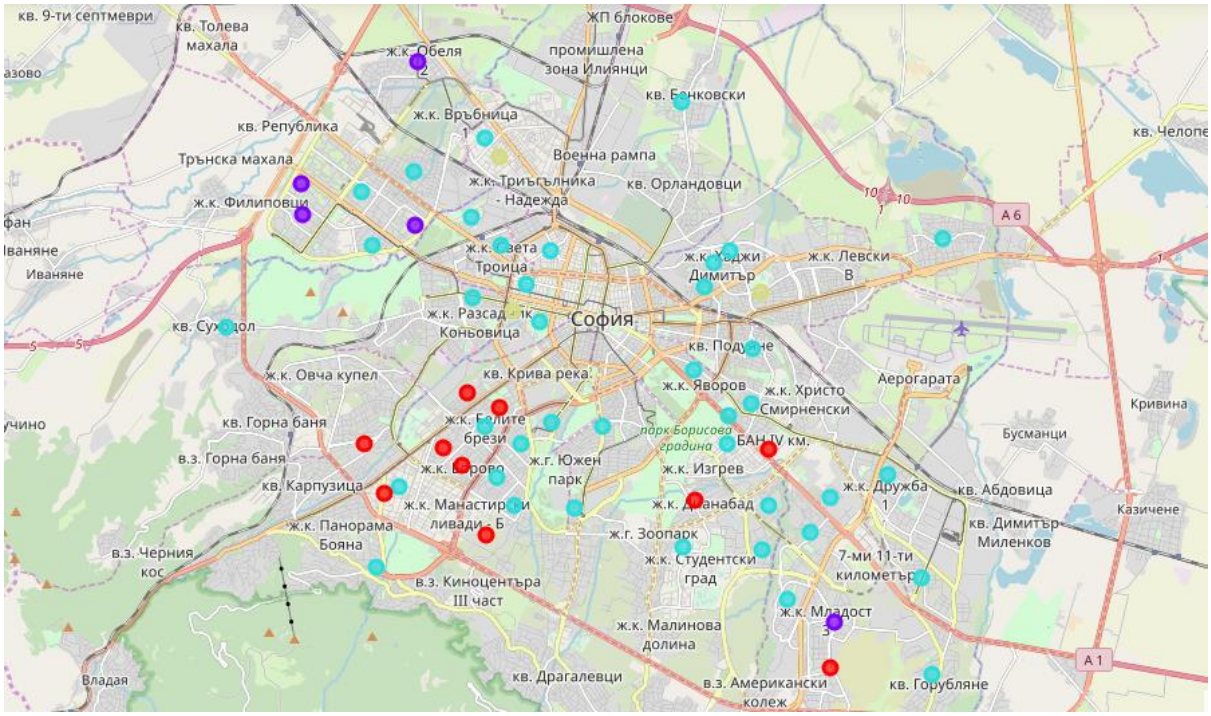


FIGURE 4: CLUSTERS OF DIFFERENT HIPSTER NEIGHBORHOODS

As a last step towards understanding the different types of neighborhoods and how they group together we can have a look at the cluster centroids and see what insight they can provide. To this end we extract the centroids from the model in order to better be able to investigate them. A better understanding of the clusters enables us to distinguish between different strands and expressions of the hipster subculture. Thus, this empowers us to better understand and, at the same time, enjoy it.

Results

The key business problem that we started with is the proper identification of hipster neighborhoods in the city of Sofia. Building upon the assumption that hipster neighborhoods are distinguished by a higher density of hipster-related venues we take the number of those within a certain radius as a proxy for the level of hipster-ness.

This research revealed that **the five most hip neighborhoods in the city of Sofia are** as follows:

1. Iztok;
2. Zona B-18;
3. Ivan Vazov;
4. Poligona;
5. Lozenets.

Going through the list of neighborhoods we note a significant concentration of venues in the top performers, i.e. top hipster neighborhoods are significantly more so that the bottom ones. We can even go further to conjecture that the bottom-listed ones are artifacts of noise in data. This begs the important question of what the minimum number and type of venues that need to be present to make a place qualify as hipster, and from the data it seems that this is around 7.

A second important insight is that we observe a **very large variance in the density of concentration** of hipster venues. There are neighborhoods with very high density, some with very low density, and a sizable group of 25 neighborhoods that do not have one single place that is associated with the hipster subculture.

The third major results from this study is that we observe **significant differences even among hipster neighborhoods in the city that we can group into four distinct clusters**. Each of those clusters is a particular expression of a different facet of the hipster lifestyle and is likely to appeal to a distinct group of individuals that are interested in the hipster subculture.

We clearly see four distinct clusters:

- **Turkish Quarters** - with a very limited selection of other hipster-related venues, this cluster prominently features Turkish restaurants. While this may not be a hipster quarter per se, it is likely of ethnic origin and non-mainstream inhabitants. We consider it the least hipster, but also the most ethnic one.

- **The Full Experience** - the second cluster features a wide range of hipster venues ranging from cafes and bookstores all the way to ethnic, vegetarian, soul food and non-traditional restaurants. It is also rich in lounges, music stores and places of artistic performance.

- **Chinese and Gourmet** - the third cluster focuses very much on Asian (Chinese) food and related experiences. It also has gourmet shops and gastropubs and a higher concentration of yoga studios. This cluster is recommended for hipster foodies looking for a more exotic experience with an Asian flavor.

- **Art and Food** - this fourth cluster goes very much under two headings - exotic food and all kinds of art. It features Indian, Chinese, European, and Vegetarian restaurant but also places to listen to music, buy records or even stock up on flowers. This is somewhat of a nightlife cluster and highly recommended for day's end rather than beginning.

Discussion

The results that stem from this small-scale project clearly show that Sofia is still a rather divided city – while there are some places with developing and vibrant hipster culture, about a third of the city shows no signs of that whatsoever. This may be due to different urban development trajectories or, more mundanely, to incomplete data. At any rate, this result is also echoed in popular opinion and media articles and is thus useful to corroborated with concrete, albeit imperfect, data.

The sharp divisions are also reflected in the divergent prices of real estate and hipster neighborhoods tend to attract more interest, and thus – differentiated price premia. The results quantify the amount of “hipster-ness” of a given neighborhood and allow sellers and buyer to compare neighborhoods based on an objective data-driven metric. The results also allow us to formulate a few recommendations:

- Buyers or renters looking for hipster quarters are well advised to consider the following hipster neighborhoods and avoid those in the non-hipster ones above.

- Sellers or landlords in one of the hipster neighborhoods may find it possible to raise the prices of their real estate, thus incorporating the hipster premium in them. On the other hands sellers and landlords in non-hipster neighborhoods may have to lower the prices so that it reflects demand and thus make deals possible.

- Casual strollers and tourists that would like to sample a hipster neighborhood of the city may want to choose what sort of experience they are looking for, and thus select where they go accordingly:
    - **Turkish Quarters**: Borovo, Dianabad, Hipodruma, Krasno Selo, Lagera, Manastirski Livadi, Mladost 4, Ovcha kupel, Pavlovo, Slatina
    - **The Full Experience**: Lyulin 10, Lyulin 3, Lyulin 4, Mladost 3, Obelya 2
    - **Chinese and Gourmet**: Bakston, Banishora, Beli brezi, Benkovski, Boyana, Darvenitsa, Dragalevtsi, Drujba 1, Drujba 2, Geo Milev, Gorublyane, Gotse Delchev, Hadji Dimitar, Hladilnika, Ivan Vazov, Izgrev, Iztok, Krasna Polyana, Lozenets, Lyulin 7, Lyulin 8, Malashevtsi, Mladost 1, Mladost 2, Moderno predgradie, Motopista, Musagenitsa, Nadejda 3, Poligona, Reduta, Stefan Karadja, Strelbishte, Studentski grad, Suhodol, Sveta Troitsa, Vrajdebna, Zaharna fabrika, Zona B-18, Zona B-5
    - **Art and Food**: Nadejda, Suhata Reka, Zona B-19

The results presented here do have **some limitations** such as their accuracy and the fact that data is static and not dynamic, but they still can serve both as stand-alone and as a foundation for future work. Most notably it would be useful to combine this data with property price data per neighborhood to quantitatively measure the size of the hipster premium. Since publicly available pricing data at the neighborhood level is not available, this will have to be scraped from a number of different publications. Another venue for future research is to have snapshots of venues at given dates (results obtained from Foursquare's API on date) and these – saved in another database. While Foursquare does not provide time series, this approach will effectively do just that and enable us to follow trends of number and type of hipster venues over time.

Conclusion

This project started from the simple question of what makes a place "hipster" – something which is useful for everyone with an interest in a given neighborhood – both for people looking for goods, services and experiences; and for their suppliers. Knowing the type and particularities of a given locale enables everyone on the market to plan appropriately, set expectations, and price adequately. We combined Foursquare's rich location data with geolocation data scraped from a publicly available website and used the operationalization of what hipster is to show which are the most hipster neighborhoods in the city.

We also saw that different expressions of this subculture lead to different flavors of neighborhoods that cluster neatly together in four groups. This insight allows us to direct strollers and tourists to the exact type of hipster experience they will enjoy most. This also holds true for potential property buyers who can now consciously choose to live in exactly the type of environment they are most fascinated with. In short, the project provided useful standalone insight for the development of a subculture in an emerging economy that can inform and entertain but also serve as a foundation for future work.

References

Cronin, J. M., McCarthy, M. B., & Collins, A. M. (2014). Covert distinction: how hipsters practice food-based resistance strategies in the production of identity. *Consumption Markets & Culture*, 17(1), 2-28.

Doll, J. (2012) The Ingredients of a Hipster Neighborhood. *The Atlantic Magazine*: https://www.theatlantic.com/culture/archive/2012/09/ingredients-hipster-neighborhood/323360/ .

Hubbard, P. (2016). Hipsters on our high streets: Consuming the gentrification frontier. *Sociological Research Online*, 21(3), 1-6.

Hunt, K. (2013). America's Best Cities for Hipsters. *Travel + Leisure*: https://www.travelandleisure.com/culture-design/americas-best-cities-for-hipsters-2013 .

Kinzey, J. (2012). *The Sacred and the Profane: an investigation of hipsters*. John Hunt Publishing.

LeBesco, K., & Naccarato, P. (2015). Distinction, disdain, and gentrification: Hipsters, food people, and the ethnic other in Brooklyn, New York. In *A Place-Based Perspective of Food in Society* (pp. 121-139). Palgrave Macmillan, New York.

Schiermer, B. (2014). Late-modern hipsters: New tendencies in popular culture. *Acta Sociologica*, 57(2), 167-181.