

Most desirable Districts of Tallinn for Families

1.Introduction

1.1 Background

Tallinn is the capital of a rather small country Estonia, occupying only 160 square KM with 434,500 population. It has been growing, developing and shaping into modern Tallinn for around 1000 years, which makes Tallinn's districts rather diverse with its function and living conditions. While Tallinn was being built on a medieval city known as Reval consisting mostly of churches and short interconnected apartment buildings, in Soviet era certain districts were added to the city as residential areas others were industrial or commercial. Over time as some of the buildings lost their function they were replaced by residential buildings or added to the districts. Tallinn is a growing city and many residential areas have emerged all over the city without following the pattern of the city that the Soviets set. Districts have lost their intentional function but as most Soviet era buildings remain it still plays a big role to shape districts characteristics. Most offices are situated in the city center as well as most restaurants, bars and get together places located in the old town, while industrial areas are surrounding cities and recreational places located along the coastline.

1.2 Problem, interested party

As new families are being made or migrated to Tallinn, choosing the most suited location to settle down and buy a home becomes challenging. Perhaps living close to the center or industrial area to be close to the workplace is preferable to some, while others would prefer to live in more suburban areas with more parks and playgrounds for children to play. Some families of two, families without children might prefer to live closer to the center not minding noise and having recreational places close by, following a more active lifestyle. As different families have different priorities and needs, their decision to choose where their new home should be varies, but is there some pattern emerging? Are certain neighbourhoods more desirable for them?

This project aims to analyse which district is preferred by families to live and what role certain venues play in decision making. This should help new families in decision making where to choose their new home, if they have long term plans to consider neighbourhoods with certain venues like playgrounds or with more recreational venues like bars. Also such information will help families that are planning or have recently moved to Tallinn and yet are not familiar with Tallinn's neighbourhoods. Lastly it will help development companies in planning for residential areas, how many apartments with how many rooms should be planned and so on.

Below I will be using words district and neighbourhood interchangeably.

2. Data acquisition and cleaning

2.1 Data description

To achieve the goal of this project information where, in which neighbourhood do families live will be needed (Number of families per neighbourhood). In addition for better understanding motives why families chose certain neighbourhoods I will need to analyse 2 types of families separately, ones with underage (<18) children and ones without, as their priorities and needs can be significantly different (+ Number of families with children per neighbourhood). As I want to know if certain types of venues play a role in decision making for families where to reside, I will need information for venues and in which category they fall as well (+Number of venues within venue category per neighbourhood).

2.2 Data Sources

Luckily in Estonia “State Statistical Central Bureau” enables citizens to use their statistical database. [Here](#) I was able to retrieve data about families by place of residence from the “population and housing census” tab. The latest available housing census data was from 2011, but as this project carries educational purpose and housing statistics shouldn’t have changed significantly since therefore I decided to use it. From Wikipedia I was able to gather data about Areas of districts and their coordinates. As there are only 8 districts in Tallinn web scraping seemed unnecessary , therefore data was filled manually. A map geojson file for Tallinn districts was compiled using QGIS software. Shape file used for compiling geojson was retrieved from [Geoportal](#). Finally Foursquare was used to get data about venues for each district.

Downloaded data from mentioned multiple sources were uploaded to git repository and loaded into dataframes in the jupyter notebook from there.

2.3 Data cleaning

Data after loading it to dataframes where checked for missing values, unnecessary data has been removed and dataframes have been shaped for better readability.

Data from the State Statistical Central Bureau contained certain unnecessary flag columns that had been removed. Also district names were containing extra characters and unnecessary text which also had been removed. For convenience I changed the column name containing spaces to a

more lacon version from 'Place of residence' to 'District'. In geojson district names were containing extra word 'linnaosa' (district in Estonian) which I'm also removing.

3. Methodology

The goal is to understand which are desirable districts in Tallinn that families choose to live in and whether there are any certain types of venues that might be a reason for such choice, therefore analysis to identify which are mostly preferred districts has been done. On the first step which districts have high population and density(population/area) has been shown and what is the share of families out of them.

As not all families are with children, the motive for choosing residential areas with certain types of neighbouring venues might differ. For example families with children might prefer to live near school or soccer fields, while families without children might prefer to live near shopping centers or gyms. Due to that families with children were separately analysed. It has been determined whether families with children have correlation with different venue categories compared to the entire share of families overall.

Next for understanding which type of venues were more common within each district, venue categories and subcategories has been analysed.

Finally it was examined what was the correlation between share of families(with children or without) and certain venue categories.

To shortly sum up used methodology, first population distribution within Tallinn districts and share of families out of them has been analyzed and after venue categories were looked at, finally correlations between them have been examined.

4. Results

4.1 Popular Neighbourhoods for families

Tallinn is the biggest city in Estonia with the largest population. From done analysis it is evident that Tallinn residents are not distributed in accordance with district areas, the district's population has no evident correlation with it's area (Figure 1).

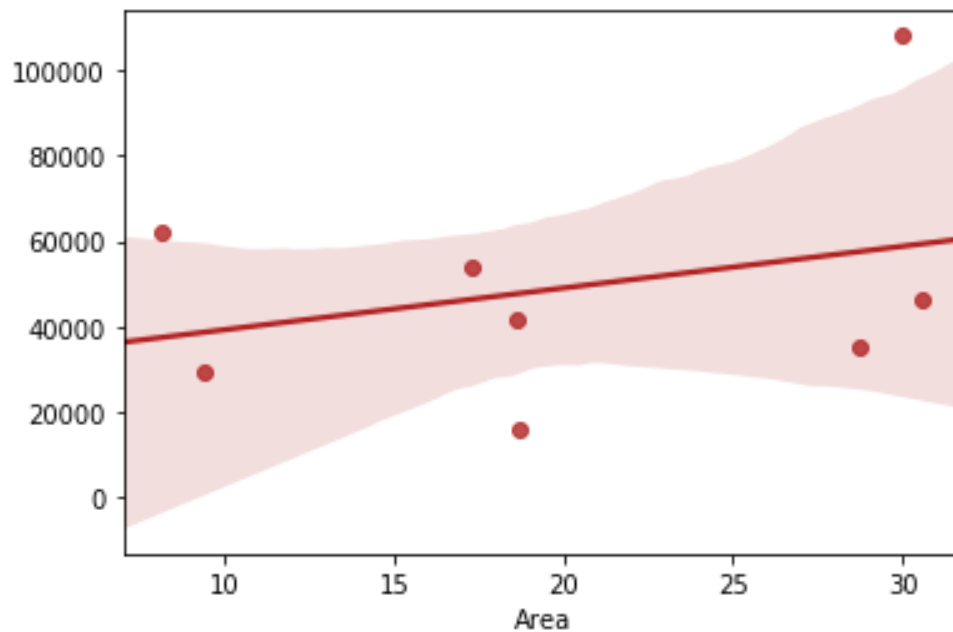


Figure 1: Correlation between Population and Area. We can see from the graph that there is no correlation due to the examples being scattered.

That means, certain districts are more densely populated than others. Mustamäe is the most densely populated district in Tallinn, but Lasnamäe has the highest population.

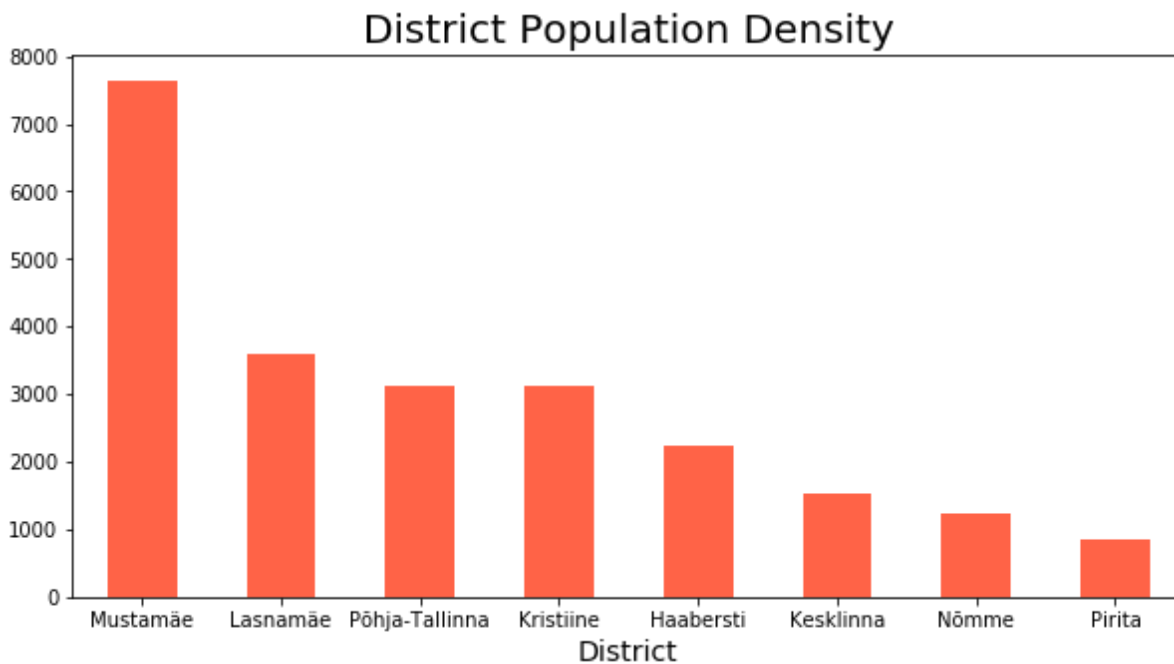


Figure 2: District population density, showing Mustamäe as the district with the highest number of residents per square meter.

However, the biggest share of the population living as families is not neither in Mustamäe nor in Lasnamäe. Pirita is the district where most family members in Tallinn leave, Nõmme and Haabersti are close second.

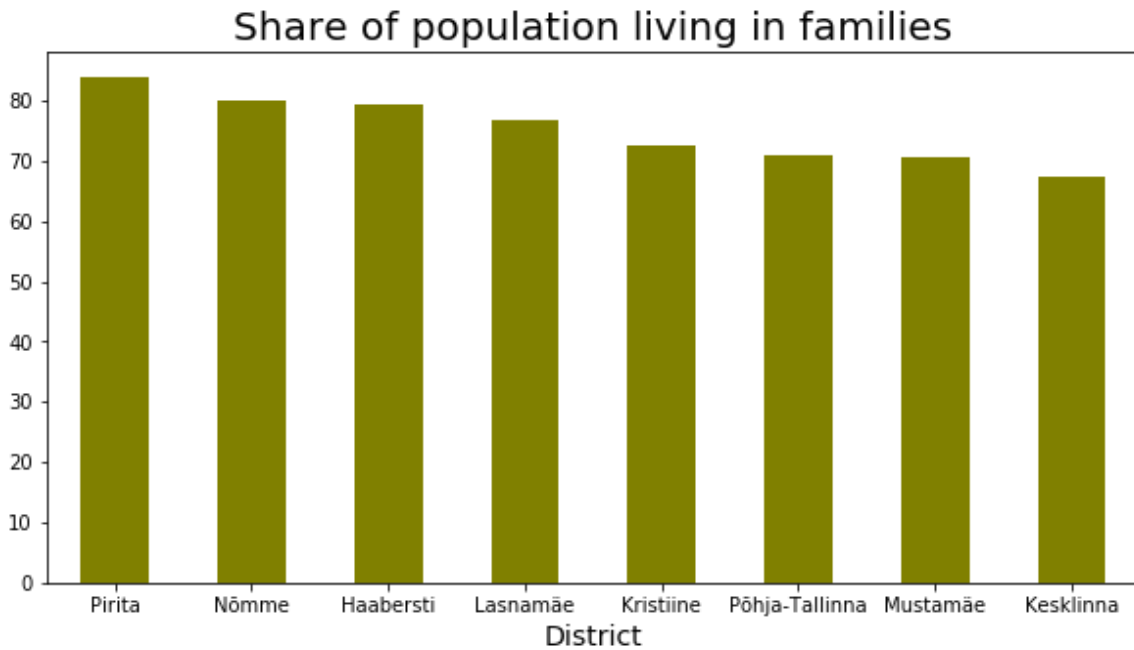


Figure 2: Share of population living in families, showing Pirita as most

Out of these families some of them include underaged(<18) children. If we look at the highest number of such families they will not be living in Pirita, rather in Lasnamäe and Mustamäe.

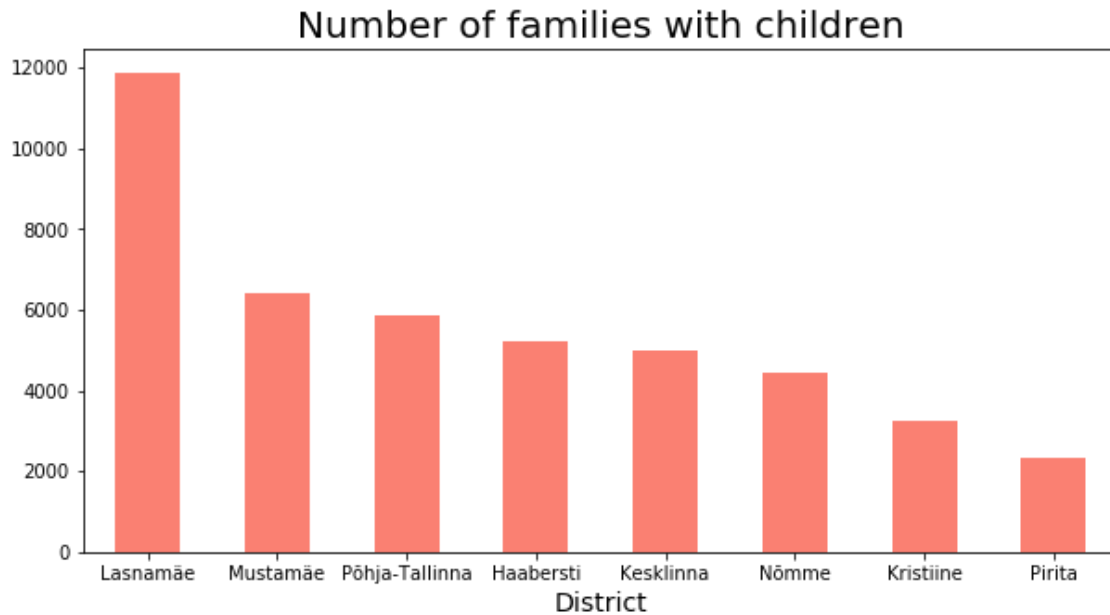


Figure 3: Families with children

Determined popular neighbourhoods for families turned out to be **Pirita** with the biggest share of families(also with high share Nõmme and Haabersti) and **Lasnamäe** with residing most number of families with children.

4.2 Venues and correlation

As for venue categories in all of the districts most types of venues were 'Shop & Services' and 'Travel & Transport'. Needs of the population, living in families or not, will fall in these venue categories, for everyone it is important to have shopping centers or bus stops located conveniently near home. But was there any special motive why families preferred Pirita? and families with children Lasnamäe? Therefore the third most frequent venue category 'Outdoors & Recreation' was explored as this could have been more relevant for families.

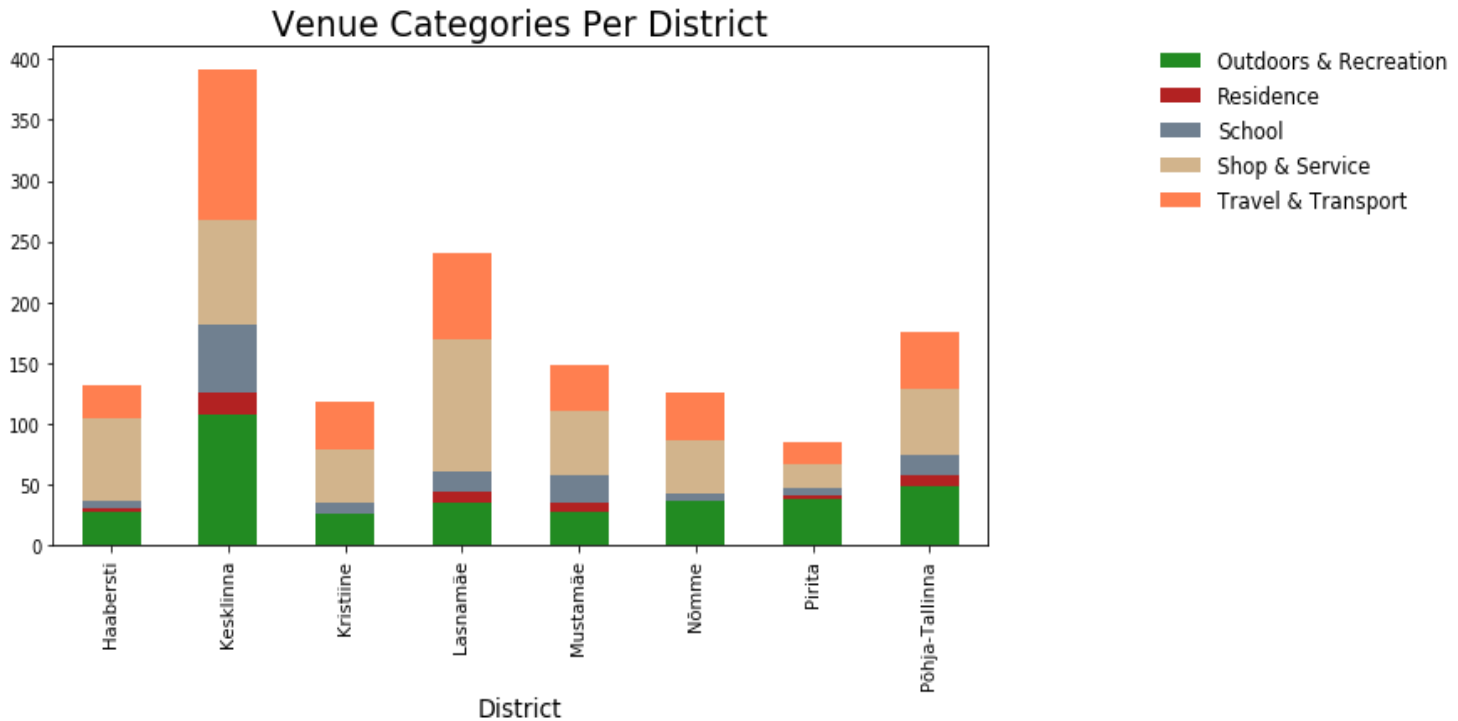


Figure4: Venue categories per district

Indeed in Pirita the number of venues belonging to the 'Outdoors & Recreation' category appeared to be larger than any other category. It is not quite the same for Nõmme and Haabersti or overall for the entire Tallinn. As it turns out there is no correlation between the share of population living in families and 'Outdoors & Recreation' venues for the entire city.

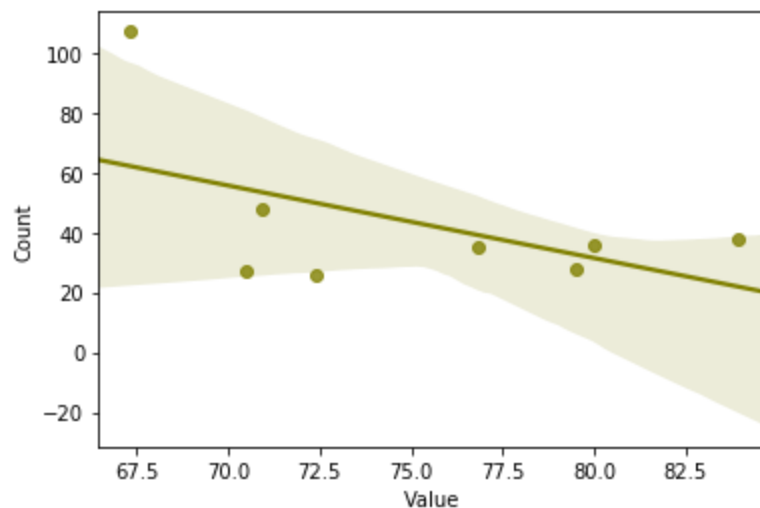


Figure 5: No correlation between the share of population living in families and 'Outdoors & Recreation' venues in Tallinn

But considering only districts with a high share of population living in families and high number of families with children below age 18 has positive correlation with 'Outdoors & Recreation' category.

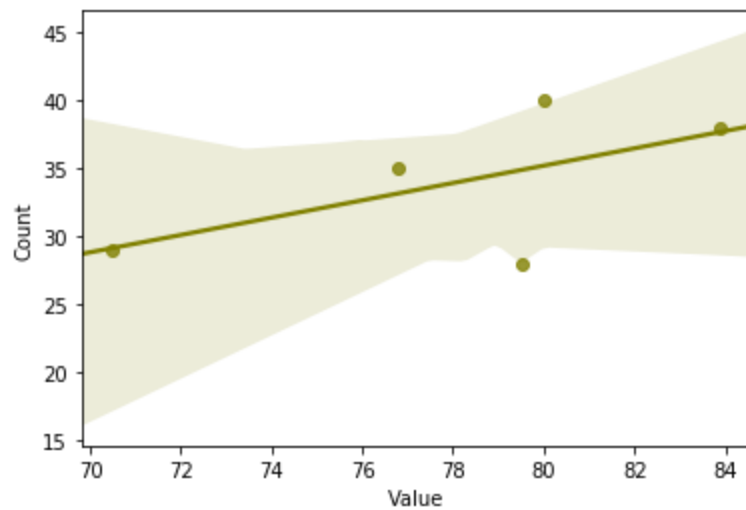


Figure 5: Positive correlation between the share of population living in families and 'Outdoors & Recreation' venues in (Pirita, Haabersti, Nõmme, Lasnamäe, Mustamäe)

Therefore subcategories of 'Outdoors & Recreation' were examined in more detail in hope of finding out patterns of preferences families might have. Most venues in Pirita consist of scenic lookouts, beaches and harbor/marina. However Nõmme has more venues from the 'Outdoors & Recreation' category (kesklinn has most, but it is an outlier as it includes an island that is not inhabited) including parks, trails and forests.

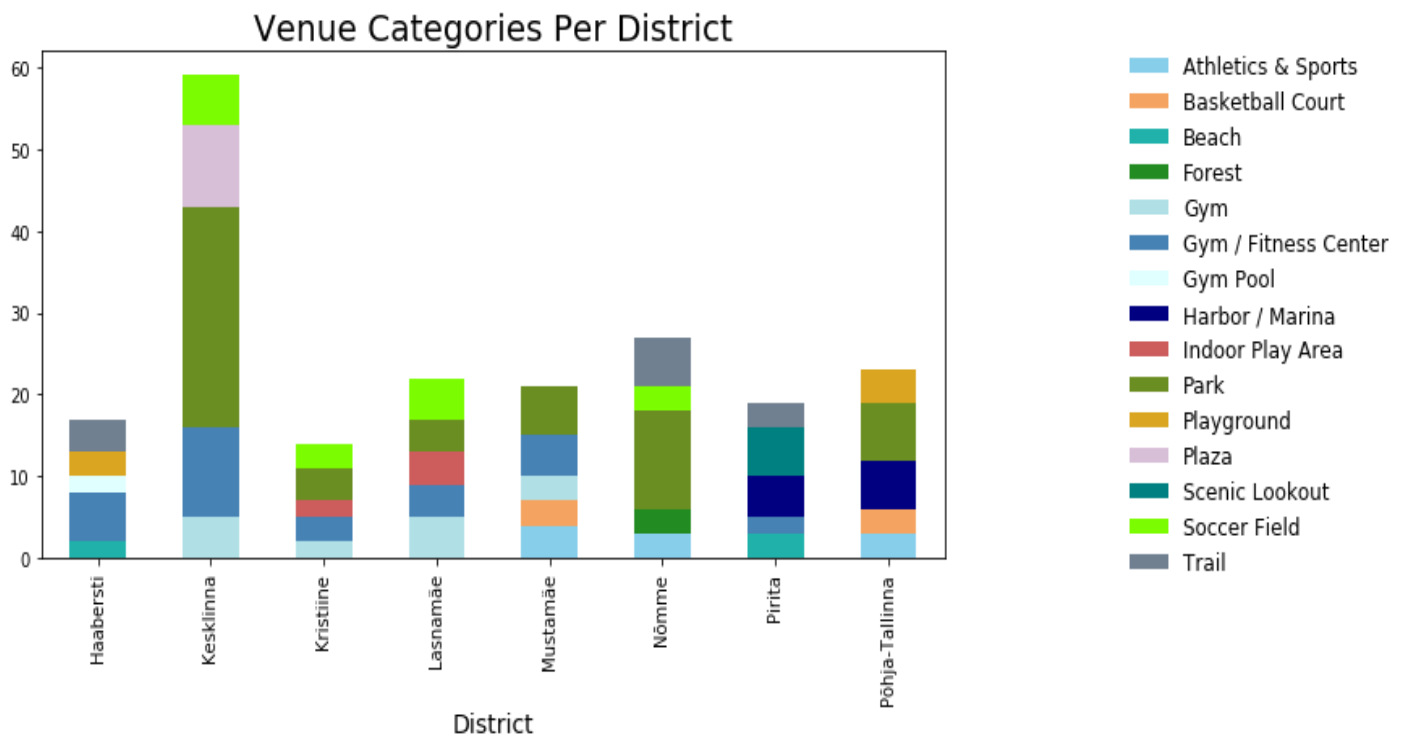


Figure 6: Venue types distribution per district

Overall in Tallinn parks and gyms are the most popular Outdoors & Recreation type of venues. For districts where most families with underaged children reside also have basketball, soccer fields and other athletic & sports venues compared to other districts.

Overall there is no direct relation between preferred neighbourhoods by families and venue categories. 'Shop & Services' and 'Travel & Transport' are dominating any other type of venue categories in all Tallinn districts. But if we examine only Pirita (district with most families) 'Outdoors & Recreation' venue category has the most number of venues, with high number of scenic lookout and Marina venue types. Lasnamäe (a district with most families with children) has a more diverse distribution of 'Outdoors & Recreation' venue types, more than half of them belong to play areas such as parks, soccer fields and indoor play areas. Also Lasnamäe has a lot more number of schools and residential venues compared to Pirita.

5. Discussion

In the venue category of 'Outdoors & Recreation' falls many similar sub-types of venues that could have been categorised together. For example Gym, Gym/Fitness Center and Gym pool are different subcategories but logically they can be grouped together. In report I leave these subcategories as they are in foursquare.

In analysis I use count of venues, but quantitative evaluation is not sensible for all types of venues. For example It makes sense if families with children will choose neighbourhoods with many playgrounds, but many parks in neighbourhoods are not necessarily an advantage. Perhaps there is only one park in the district but the area of this park is much larger than the area of many small parks accumulated in the other district. In case it was evident to me that the need of analyzing parks separately would have been beneficial, park area data should have been obtained. Foursquare doesn't offer all data for Tallinn unfortunately, as it is not a very popular API here. For example postal index columns were returned empty to me and data was missing in many other columns. Same analyses done using Google API might have given different results, however in this project usage of Foursquare API was mandatory.

6. Conclusion

In this project I analysed which Tallinn districts are mostly preferred by families to reside and whether or not there are any certain types of venues that play a role in such distribution. I identified Tallinn districts with highest population density, highest share of families residing and highest number of families with children under 18. I explored different venue categories in those districts and their relation with family data. For determining an association between these variables an inferential

statistical method of correlation has been used. As there was no correlation for the entire city's share of population living in families and venue categories, there is a positive trend of correlation for only those neighbourhoods that have a high share of population living in families. Two Tallinn districts were identified as the most popular for families and families with children and venue types have been explored in more detail.

Example suggestions for families based on the research:

Target group	Suggested neighborhood for home
Family with children, valuing outdoors	Pirita, Nõmme, Haabersti
Family on a budget	Lasnamäe, Mustamäe
Working family without children	Kesklinn