**A tribute to my wrongness: Airbnb rents and distance to health tourism centers in Istanbul.**

*Are the prices of Airbnb rentals in Istanbul correlated with distance to institutions related to health tourism? I really hoped that the answer was yes. However, it was not. Here is a tribute to my wrongness.*

*A picture containing map

Description automatically generatedA sneak peek into the results of the analysis conducted in this project.*

Influenced by my previous analysis, I wanted to know more about the health institutions in Istanbul. I had two particular types of health institutions in my mind: **aesthetic clinics and hair transplant centers.**

Why? Well, before the Coronavirus pandemic, if you were to come to Istanbul, there was one thing that you would surely notice: the amount of people walking around with bloody scalps.

Graphical user interface, text, application, email

Description automatically generated

*Health tourism (especially for hair follicle transplants) is big business in Turkey.*

A quick google search using the keywords “Istanbul” and “hair transplant” will reveal to you the reasons of this odd sighting. Istanbul is a hotspot for health tourism. There are many aesthetics clinics and hair transplant centers throughout the city. Their customers are mostly foreigners and tourists that are looking to sneak in some kind of a cosmetics-related medical procedure into their vacation time.

Motivated by this very niche type of tourism, I started wondering if I can see the effects of health tourism related institutions on the overall tourism economy of the city.

As in other tourism hotspots, Airbnb rentals are a very popular choice of accommodation for tourists in Istanbul. **Perhaps I could find some kind of a linear or monotonic relationship in between the price of an Airbnb rental and its closeness to a health tourism institution?**

**The distribution of health tourism related institutions across Istanbul**

The question that I set out to answer had a spatial dimension to it. I figured that looking at the distribution of health tourism institutions and Airbnb rentals across Istanbul could be a great first step in understanding the relationship in between these two. Drawing data from IMM’s (Istanbul Metropolitan Municipality) dataset on health institutions and from web scraping scripts that I have wrote, I plotted out all health tourism related institutions (aesthetic clinics and hair transplant centers) both on the individual level and on the district level.

Diagram, schematic

Description automatically generated

*An overview of the distribution of health tourism institutions across Istanbul featuring two maps and a bar chart. Aesthetic clinics and hair transplant centers seem to be focused around centralized and well-connected districts.*

As it can be seen on the district level choropleth map above, health tourism related institutions in Istanbul seem to be congregated around central and southern districts such as Şişli, Beşiktaş and Kadıköy. These three, with 33, 19 and 17 health tourism related institutions respectively, are very well connected to other districts. Along with Ataşehir and Üsküdar, the 4th and 5th districts on the list, these districts serve as major economic centers that draw in people from other districts on a daily basis. The districts that lie on the southern shoreline of the European side of the city are also major population centers.

Looking at the distribution of health tourism related institutions on an individual level seems to confirm our previous statement related to their distribution. Even within southern districts, most health tourism related institutions congregate around the southern shoreline. There are very few institutions that are outside the central southern cluster.

These results are hardly shocking when we consider that the gravity center of Istanbul, at least tourism wise, lies on the southern tip of the city. These are the districts that are most frequented by tourists and operating a health tourism related clinic around tourist hotspots makes sense.

Perhaps delving deeper into the specifics of this spatial distribution can help us gain a deeper insight about the topic before moving on. Considering this, we can look at how the number of institutions related to health tourism within a district relates to some staple statistics such as population and yearly average household income. We can use a scatterplot of these variables to investigate them.

Chart, scatter chart

Description automatically generated

*Two scatterplots that show the relationship in between the number of institutions related to health tourism (dependent variable) in a district and two other independent variables. The dependent variable has no linear relationship whatsoever with population and somewhat of a mild positive linear relationship with yearly average household income.*

What we discover is that there is practically no linear relationship in between the population of a district and the number of health tourism related institutions that same district has. Considering this, the fact that the southern districts are also population hotspots does not seem to be relevant. However, there appears to be a mild linear relationship in between the yearly average household income of a district and the number of health tourism related institutions that the district houses. However, it is impossible to make any comments about the causality and the direction of this mild relationship with only the data at hand.

Moving on, we can now look at the second element of the question that we are seeking to answer: Airbnb rentals.

**The distribution of Airbnb rentals across Istanbul.**

The dataset that I used to learn about the spatial distribution of Airbnb rentals belongs to a project called Inside Airbnb. Considering that worldwide tourism took a great hit because of the Coronavirus pandemic that we are still battling by the time I am writing about this analysis, I choose to base my analysis on data from May 2019.

Graphical user interface, map

Description automatically generated

*An overview of the distribution of Airbnb rentals across Istanbul featuring two maps and a bar chart. While it is not completely the same, the ranking on the right corner seems to mimic the previous ranking.*

Once again, the first of the maps reveal that the southern districts overlooking the Bosphorus enjoy the highest number of Airbnb rentals. Some of the districts with the highest amount of health tourism related institutions, namely Şişli, Beşiktaş and Kadıköy, also make it to the top five in the Airbnb ranking. However, the 1st place and the 3rd place in the rankings are held by Beyoğlu and Fatih respectively. These two districts were not among the top five in the health tourism related institution ranking.

On an individual level, the distribution of Airbnb rentals seems far less clustered when compared to the distribution of health tourism related institutions. The number of Airbnb rentals and the granularity level of the map presented here makes it hard to arrive at a more nuanced conclusion. However, it is clear that Airbnb rentals are present in some tourism hotspots where health tourism institutions are absent. The northeastern district of Şile, the southwestern district of Silivri and the Princes’ Islands can be considered as independent clusters. Even without making any analysis about the relationship in between distance and price, we can say that the existence of such sub-clusters will significantly lower the chances of any meaningful linear or monotonic relationship.

Before moving on, we can once again look at how the number of Airbnb rentals in a district relates to the population and the yearly average household income of that same district.

Chart, scatter chart

Description automatically generated

*Two scatterplots that show the relationship in between the number of Airbnb rentals (dependent variable) in a district and two other independent variables. The dependent variable has no linear relationship whatsoever with population and yearly average household income.*

Just like our previous correlation analysis, we see that the number of Airbnb rentals in a district has no meaningful linear relationship with population. The same applies to yearly average household income.

**From entity to attribute: Airbnb rental prices**

We have gained valuable information about the geographical location of our two elements of interest. However, what we are ultimately interested in is two specific attributes (nightly rent and the distance to a health tourism center of an Airbnb rental) that belong to one of these entities. The dependent variable of our final analysis is Airbnb rental price. Therefore, looking at how Airbnb rental prices are distributed is a great idea before moving on to the final correlation analysis. We can achieve this by utilizing a histogram:

Chart, histogram

Description automatically generated *Two histograms that show the distribution of Airbnb rental prices. The first histogram does not tell us anything except the fact that the distribution is extremely right skewed. The second distribution allows us to see the distribution at a different level of granularity.*

The first histogram is nothing but a very tall column and a few extremely small ones. This is a very bad sign for our prospects of finding a linear or monotic relationship in between the variables of interest because it signals that the distribution is heavily skewed with a skewness of 33.74. For our case, a skewness value greater than zero means that there is more weight to the right of the distribution because there are extreme outliers in the dataset. These outliers drag the distribution of rental prices away from a normal distribution shape and heavily damage the chances of finding a linear relationship.

The extreme right skew of the rental price distribution also makes it impractical to use a histogram. The first histogram looks the way it does because it has to accommodate the presence of some values that are infrequent but big in effect. We cannot see what happens at the more “normal” parts of the distribution because of this distortion. The second histogram shows the log10() transformation of the same data. Such a transformation allows us to take a peek at the smaller kinks in the distribution at the cost of losing resolution. Sadly, seeing the distribution in more detail does not give us any hope about discovering a linear relationship.

A picture containing chart

Description automatically generated

*Multiple histograms that show the distribution of Airbnb rental prices at the district level. Even with a very generous threshold, most district-level distributions are heavily skewed to the right.*

The extreme right skew of Airbnb rental prices still persists when the distributions are considered at the district level. Even with a very generous and arbitrary threshold of skewness at 5, there are districts (yellow colored) that can be considered heavily skewed. Most districts have the skewness of at least 1.

WE CAN NOW MOVE ON LOL

**Interlude: how did I conduct the nearest neighbor analysis?**

**Map

Description automatically generated**

*An abstract map of Istanbul confirming that the nearest neighbor analysis script used in this project managed to match each Airbnb rental (blue circle) with the nearest health tourism related institution (orange triangle.)*