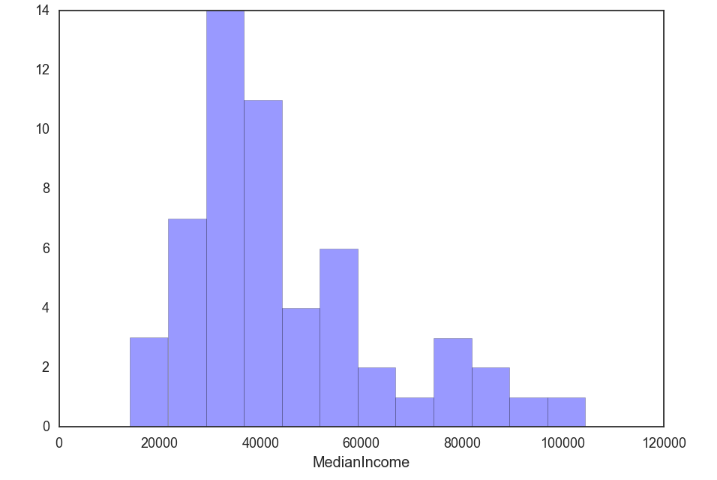
**Introduction**

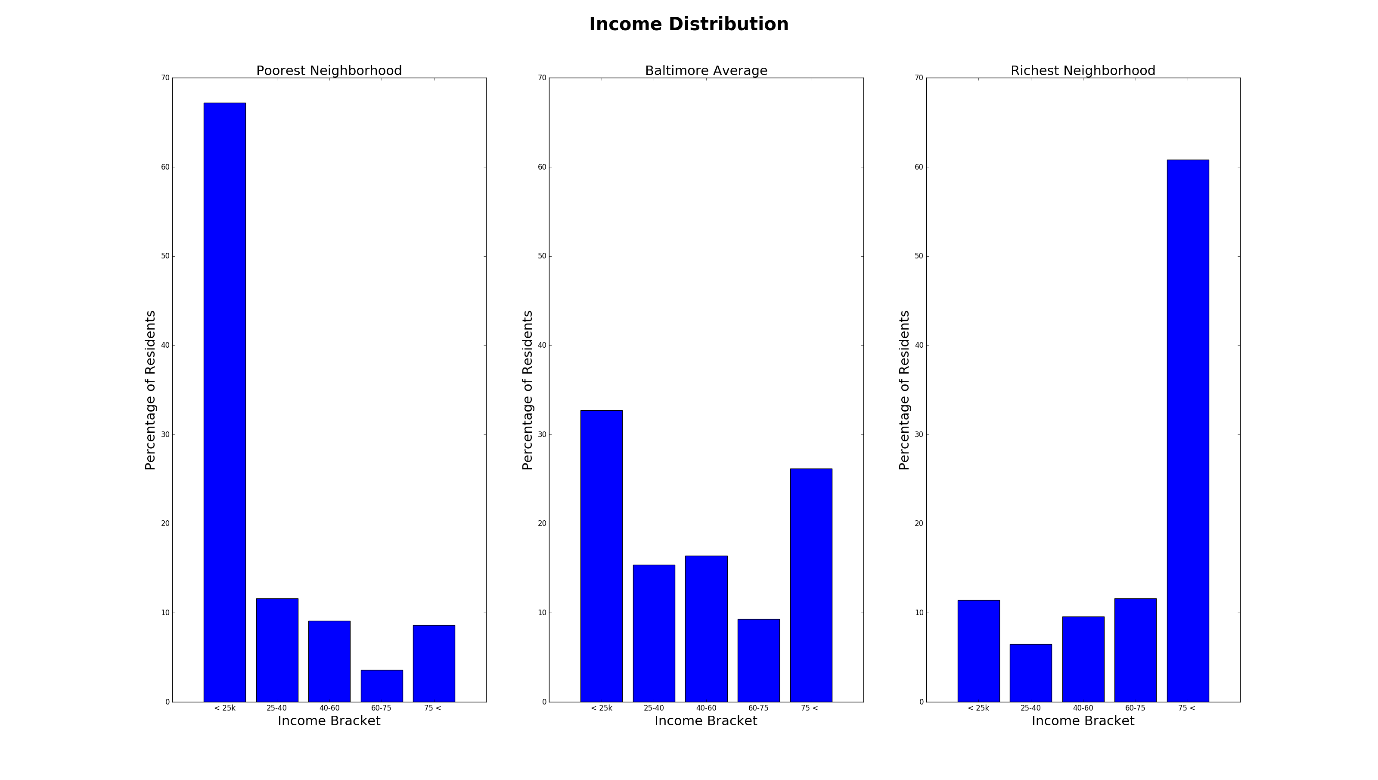
Baltimore is the largest city in Maryland, and the 29th most populous city in the country. It is also the largest independent city in the United States of America. This city has always been a city of neighbourhoods, everyone a universe unto itself. Each section has its own history, culture, and personality- an identity that can only be understood after years of experiences.

We all know that moving to a new city can be a daunting task, even with the help of a professional moving company. However, things start to look easy and comfortable once you know more about the area. Decision making can be made easier with a little bit of extra information up your sleeve. Is Baltimore on your list of potential lace to move? Are you planning to rent or buy a house there? If so, then keep reading because this will guide you to pick your favourite neighbourhood because everybody loves a nice neighbourhood!

**Exploratory Analysis**

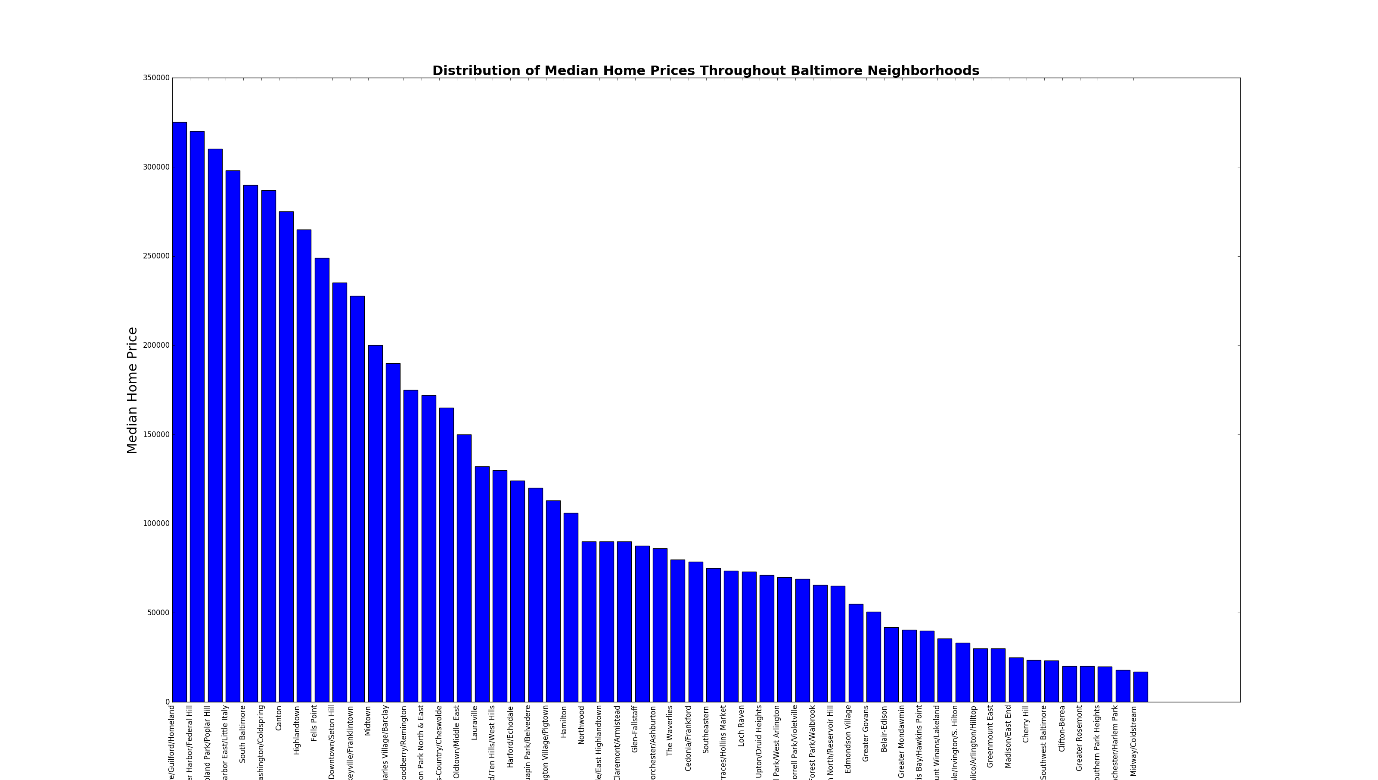
The first bit of analysis was inclined towards analysing distributions of metrics such as income, housing prices, and using the results, to look at the distributions in the richest and the poorest neighbourhoods in Baltimore.



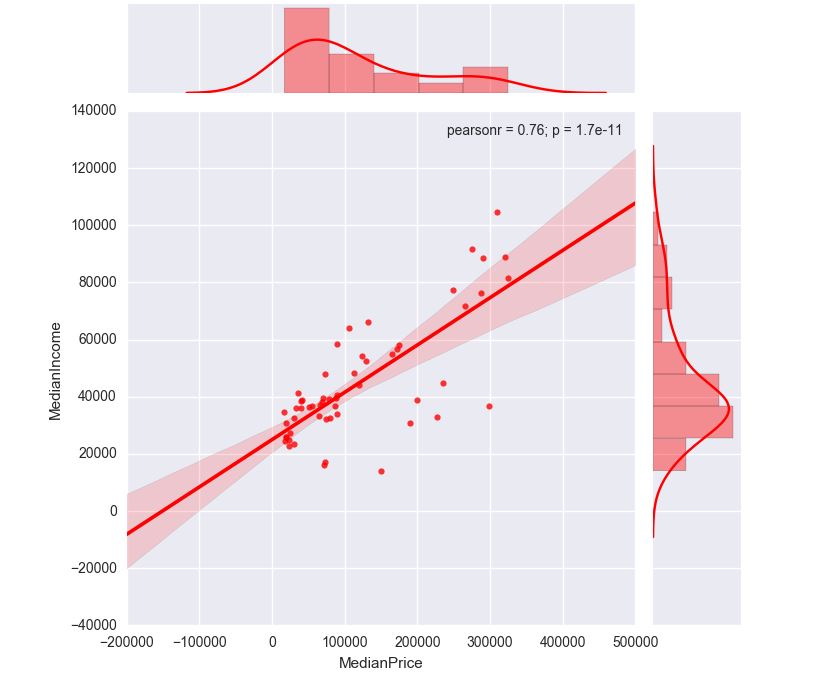


This chart confirms something that is something of a truism when we think about cities -- that neighbourhoods are extremely segregated and well defined on income boundaries. As we can see from these distributions the richest and poorest areas do not look like the Baltimore city average at all. Instead, they are, respectively, very rich, and very poor. This stark inequality represents one of the most troubling facets of city living. The inequality in wealth distribution is interesting in and of itself but it may also imply something important about the geography of the city. If wealth were more evenly spread would we see more "viable" neighbourhoods? It is hard to think that is not the case.

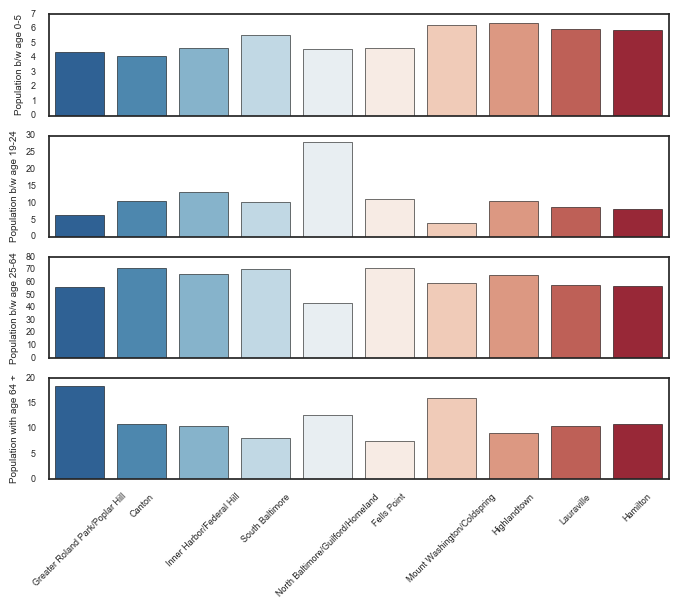
Our next move, was to analyse the distribution of housing prices across the city.



As we can see from the home price chart home prices are high in a relatively small number of neighbourhoods before declining rather drastically. This seems to reinforce the above: a few neighbourhoods contain much of the wealth of Baltimore. This can be further substantiated by the following graph, which demonstrates the relationship between median price and income. There exists a reasonably high correlation coefficient of 0.76 at a p-value less than 0.05.



Further, we looked at age distributions amongst the neighbourhoods the neighbourhoods with highest incomes.

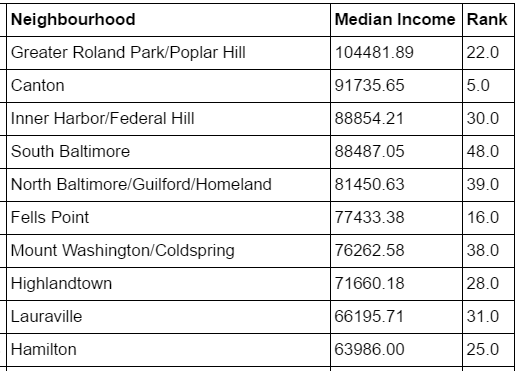
s

Comparing this chart to the income distributions across the neighbourhoods, we observed that the neighbourhoods with generally high incomes, are ones with a high population of people between the ages 19-64, that is, a working-class population. (North and South Baltimore). This gives us a fair idea of where the working population on the city is clustered. Also, we observed, that areas like Cold Spring have high populations of elderly people, and a relatively large population of children under the age of 5, and between the age of 25-64. This gives us an insight into the where families, with people involved in profession can be found.

We also considered racial diversity indices, percentage of people living below poverty line and the gender ratio for each neighbourhood. The above plots, were, in our opinion, the most relevant to our eventual goal.

**Rankings**

The goal of this project is to output rankings based on the wealth of data provided through the Baltimore Neighbourhood Indicators Alliance. But what sort of rankings? Well, one way to do it would be to output ordered lists that reflect key neighbourhood vitals. A perfectly valid ranking under this single variable ordering would be the neighbourhoods ordered by median income. You'd end up with something like this:



This is interesting -- we know where the rich prefer to live. But it doesn't get us much else. There are a multitude of interesting ways to rank

CSA2010 adult\_arrest\_rate

Downtown/Seton Hill 241.5

Madison/East End 154.6

Southwest Baltimore 147.3

Clifton-Berea 131.9

Washington Village/Pigtown 129.5

.........

Interesting -- If we valued our safety perhaps we'd best avoid the Downtown/Seton Hill area. Looks like trouble.

But these rankings don't tell the whole story. They don't even tell half the story. You might say they're rather...one dimensional. We should them out. But how? If you're rich, well, perhaps you can simply trust the taste of other rich folks and choose the area with the highest median income. If you did that, you'd be using median income as a proxy for quality. But do people of the same economic cohort really want the same things as you? I’d venture a guess that they don't.

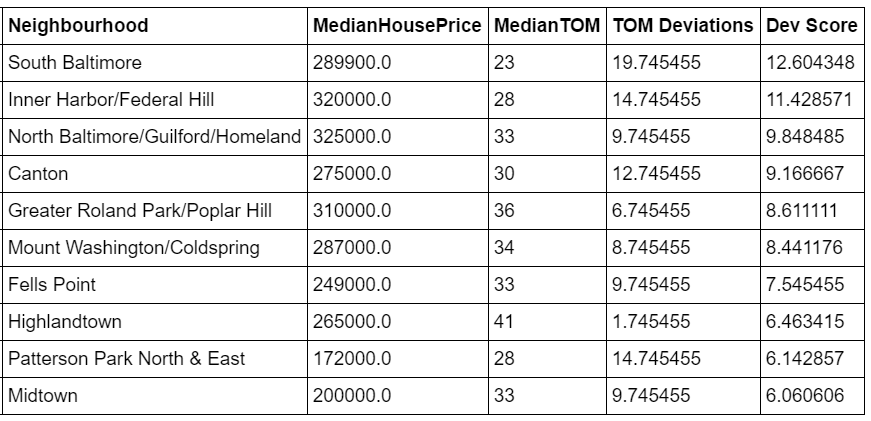
And what if you're not rich? Maybe you should find the safest neighbourhood that you can afford and stick to it. That might be a sensible approach. But what if you're a tough young guy and would tolerate a little less safety for better proximity to the music venues you like. Then you'd need some sort of intersection of the rankings for safety, entertainment, and price.

**Baseline Model**

We needed a metric that could first measure for quality, and our assumption was, that using housing prices as a proxy for quality could be a good start, further upon which, we could penalize or credit a neighbourhood depending on the aforementioned metrics.

How should we parametrize this model? What sorts of data can and should it include? These are important questions whose correct answer will guarantee we get the correct answer. Our first attempt was to rank neighbourhoods depending on the median Time on Market, that is a measure of how fast a house in a neighbourhood is bought. We assumed, that a neighbourhood, which has a lower time on market, is a favourable place to stay, as opposed to another where people don’t buy real estate as fast. But what is the answer we want? Well, if we were to create a scoring function that output a ranking of neighbourhoods based on which was "best" we could do that. But defining what is best is, as we noted above, tricky. Should best be safest? Artiest? The most connected to public transit? Some arbitrary combination of everything?

We looked at the deviations of the Time on Market for each of the neighbourhoods, and decided to rank the neighbourhoods by median prices divided by these deviations. Following are the results.

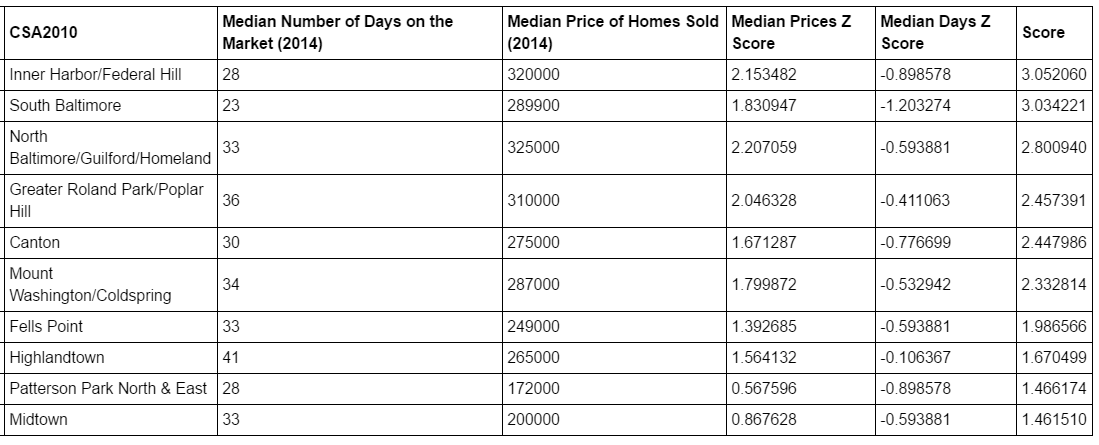


This model, we felt, over-penalized a neighbourhood with lower prices than the highest, and we also realised, that the Price Per Day of Time on Market doesn’t have any substantial meaning to our goal. We needed a better way to penalize a neighbourhood with high Time on Market.

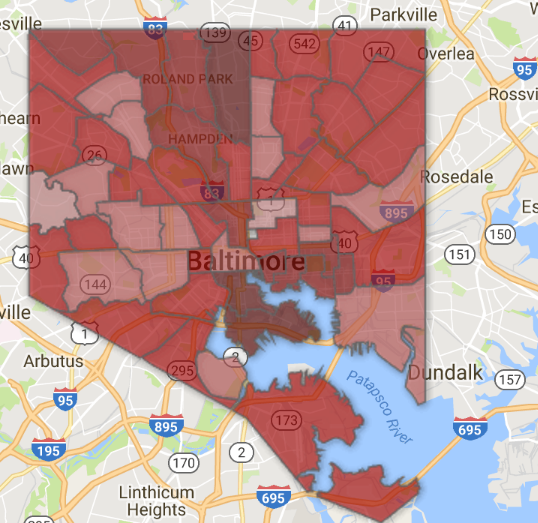
For a baseline notion of "best" we eventually settled on the following formula-

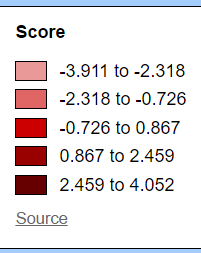
What does this formula capture? Well, we think it captures a few things very compactly. One way to compute the baseline model would be to have a lot of different parameters that we could weight to our liking. But there is no ground truth. Plus, as you might imagine, people don't choose the neighbourhood they live in based upon the Baltimore Neighbourhood Indicators Alliance Vital Signs report. The theory that underlies this model is the one of efficient markets. The price of the home in each neighborhood should encode a lot of information about the area that the home is in. Succinct Perhaps we should stop here then. That's it. The market has told us everything we need to know.

Well, another bit of information that the market provides is the time on the market, or TOM. The time on the market is a useful indicator because it provides us with a sense of popularity. Areas with a higher time on the market are less popular for new homebuyers. Therefore, we use the negative sign. Combined, this simple model should give us some notion of the most desirable neighbourhoods. And it does. We've done some validation with lists available to us (aggregated neighborhood reviews and the like) that appear to confirm this initial baseline.



And as plotted geo-spatially -





Art and Diversity

We also wanted to look at neighborhoods from a different perspective- rather than direct measures like prices, education standards, and transportation we wanted to evaluate eah neighborhood based on how it ranks in Art and diversity. People have become more sensitive and aware about these things when they choose a neighborhood to live in. Places with high creativity index are usually preferred by people, and we would like to test his hypothesis by comparing the ranking obtained by art index with other models.

We have chosen the following parameters to calculate art index for each neighborhood:

1. Businesses involved in creative economy- this takes into accont all the profit and non profit businesses involved directly or indirectly in arts- industiries that support artisitc and cultural skillsets.

2. Employment in creative businesses- total number of people employed in the businesses we took into account. This does not include peopl who identity themselves as artists or are involved in part time art related activities.

3. Total public art- museums, murals etc

4. Events permits requested- art related events organized to support such activities

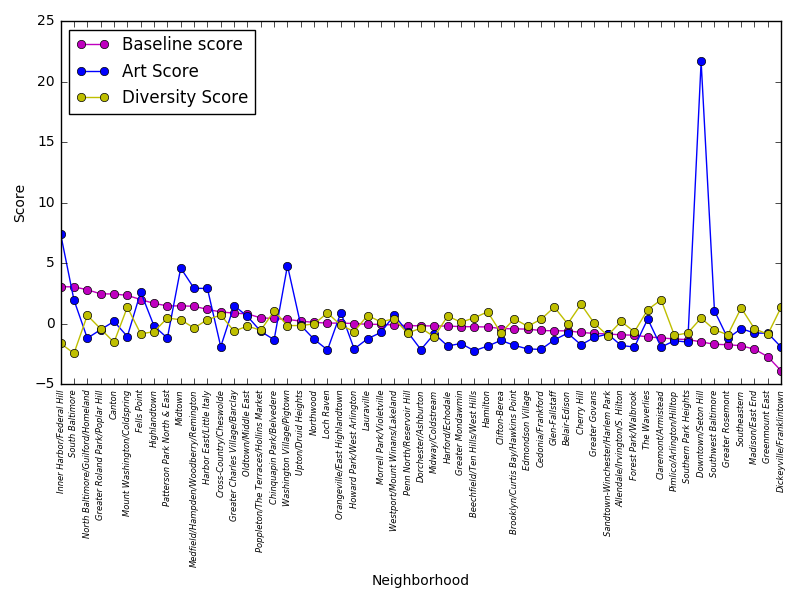
Ranking by ART index is calculated by summing the abovementioned parameters after they have been normalized.

Enhancements: We can asssign ratings to art places- museums, monuments, theatres etc to come up with a better and robust ranking system. This will equire us to scour the web for such ratings as a comprehensive list of such places is difficult to get by per neighborhood basis.

Ranking by DIVERSITY index

Diversity is one thing which has different meaning to it. For our puspose we planend to consider the following parameters:

1. Racial diversity- this takes into account the fact that when two people are picked randomly what are the chances that they belong to different races- WHite, Asian, Black, Hispanic and non Hispanic



2. Female population- this shows percentage of females in the area.

Interesting observation

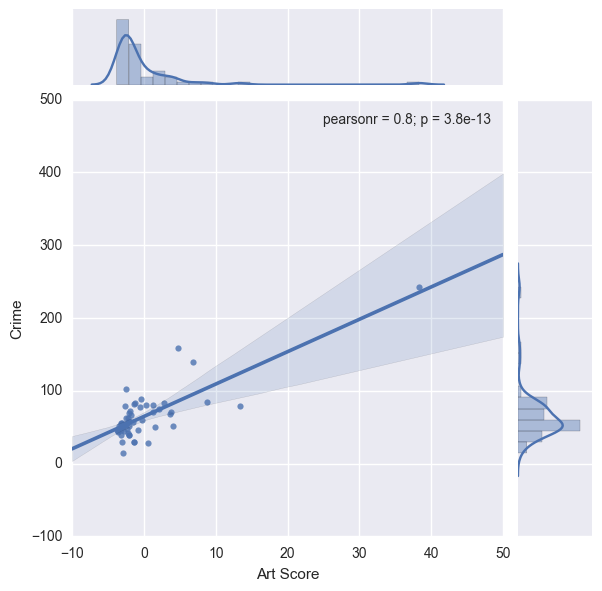
We can see that Federal Hill ranks the highest according to our baseline model, but it does not score very well in terms of art and diversity index. Further scouring the web for its demographics reveals that majority of people in this neighborhood are white (87.3%) even though it has almost equal representation from both the genders.

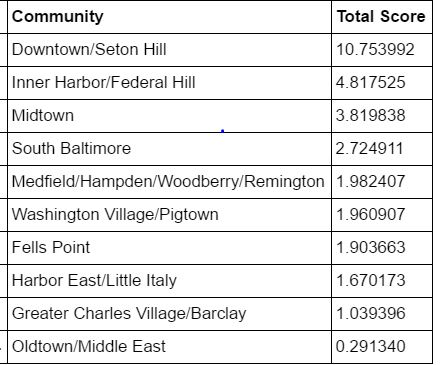
From the graph one striking thing we observe is that Downdown/Seton hill surpasses all other neighborhoods in terms of its art score. This place is in close proximity with Mount Vernon which is famous for its nightlife, cultural excitement, and LGBT hangout places. It is also famous for shows put up by galleries and emerging artists. A recent poll on Nextdoor shows that people love Seton hill for its architectural charm and historic appeal.

In our baseline model we have taken house prices as a proxy for quality and places which rank high according to that model have a bad diversity score. From the graph fist 10 places have below average diversity score. Toward the middle of the graph we can see that all the three graphs come close in their rankings and as we move further right places which are ranked lower as per our baseline model have either average or higher diversity score. Even though this is something we expect because we know that white community in Baltimore is richer than its counterparts, but when the data reveals such statistics it gets interesting!

Specifically, Arts

We further narrowed down in to using only the number of public murals, the population involved in the creative economy in an area, number of arthouses and population earning their income from the arts industry per thousand population as a ranking function., and we saw a rather interesting correlation. The art score of an area, is highly correlated with the crime rate!





Let’s now look at the population distribution across these areas

In the figure below, we can see, that there exists a greater population belonging to the working class in areas such as Seton and Federal Hill. We can infer from this, that a vast population of youth does prefer these two areas as opposed to the others. Federal Hill seems to rank higher in most rankings and seems to be a good combination of diversity, art and culture and quality of living.

