Paris Gentrification Presentation

Paul Joahny and Matei Vasile

1. Research Question

This report attempts to answer the question **how does gentrification in the Paris area affect the "Tissu comercial"?** The objective of this report is to assess how the development and structure of businesses changes from gentrified neighbourhoods to those that are not.

Gentrification is a social phenomenon , as many social phenomenon it has different definitions we'll define it as the process by which a poor neighborhood in a city is replaced by a new population. We can evaluate this by looking at the development of the population of a neighbourhood overtime. This is completed via a process in which a poor neighborhood in a city is changed by people who have money, including especially the improvement or replacement of buildings and businesses.

The idea is to see how gentrification can reshape neighbourhoods in terms of retail/service supply. As higher income populations move into formerly lower income or mixed income areas, we expect to observe a shift in the available businesses and more generally the commercial mix: growth of cafes, restaurants, shops, boutiques, coworking spaces etc with a subsequent decrease in traditional proximity retail. We will also attempt to observe whether new types of businesses emerge specifically tailored to a wealthier population.

We will do this by firstly identifying areas (IRIS units) in which clear signs of gentrification have occurred. To do this we will create some indicators for price per sqm, median income levels, poverty rates, income disparity etc. Then using these to identify areas that for example have been previously below the average price per sqm and are now above that, we can note them as gentrified areas.

Once we have identified which IRIS units have undergone gentrification, we will observe changes in the "tissu commerical" of these areas using the BPE (Base Permanente des Equipements). The BPE dataset is a record of the quantity and type of establishments. By analysing changes in the BPE for our identified gentrified areas we can see whether new business categories emerge as the distribution of income changes, which type of businesses increase in quantity, which types decline.

This method of first identifying gentrified areas, then analysing the change in types and quantities of businesses, allows us to observe the effect of gentrification on the "tissu commerical" of Paris.

2. Dataset Descriptions

a)FILOSOFI

The filosofi database, offers specific estimates of household income, poverty, and inequality at area levels such as IRIS. It combines fiscal data and social benefit data to estimate income before and after redistribution, adjusted per household unit to account for household size and structure. Filosofi replaces older local income datasets and provides useful indicators like income deciles, medians, poverty rates and inequality measures like the Gini coefficient.

b)IRIS Historical Change

The "Historique des codes IRIS" dataset shows a cross-reference of IRIS codes in France from 1999-2022, revealing how these neighbourhood-level statistical units have changed over time. It offers a "table de passage" that links old and new IRIS identifiers following boundary modifications, splits, or regroupings. This resource allows for researches to align socio-economic data over the years despite changes in IRIS definitions.

c)IRIS Shapefile

The IRIS shapefile provides the official polygon boundaries for all IRIS units in France. The 2020 "France entiere" shapefile offers total coverage of France and is the standard geographic reference for mapping IRIS-level data. It is broadly used to link datasets (like FILOSOFI and DVF) to their corresponding neighbourhoods.

d)DVF

The "Demandes de Valeurs Foncieres" (DVF) dataset, which is available through data.gouv.fr, contains property transaction records for France and includes geographic information when available. We have data sets from 2014-2018 and 2020-2025. DVF provides valuable information such as transaction dates, sale values, property types, surface areas. It can be used to analyse real estate information, such as change in price per square metre. In our project, DVF enables us to construct indicators of housing price changes at the IRIS level.

e)BPE

The "Base Permanente des Equipements" (BPE) dataset for Ile de France. It provides an inventory of public and private facilities, amenities and commercial establishments for the region of France.

f)Activity of residents

This base « Activité des résidents » gives data on different characteristics of the workers of a given neighborhood. This could be age , sex , CSP category.

3. Data Analysis

The main part of our analysis was finding a way to measure something you can't directly observe: gentrification. To be able to do that we chose to measure it trough 3 different dimensions. Revenue, Population and Prices.

Population : This whole dimension is looking at who are the people living in said neighbourhoods. In france there's a good indicator called CSP(Catégorie Socio Profesionelle) created and monitored by INSEE this works as a statistical categorisation of jobs of a similar social environment.These categories are

Employees-Employés : This is Ce groupe socioprofessionnel rassemble des professions aux fonctions très variées (administratives, commerciales, de services aux particuliers, de surveillance et sécurité, etc.) dont il est difficile de trouver une définition commune si ce n'est qu'elles n'ont pas ou peu de responsabilité d'encadrement. Ce groupe socioprofessionnel est composé uniquement de salariés, à de rares exceptions près.

Workers-Ouvriers This represents that work in the industrial secrors Ce groupe socioprofessionnel regroupe des personnes qui exercent des fonctions d'exécution dans le cadre d'une division poussée du travail dans les secteurs industriels, de services à l'industrie (nettoyage, maintenance, tri, expédition, etc.) ou des tâches manuelles dans les secteurs artisanaux ou agricoles. Il ne comprend que des salariés, qui peuvent être employés par des établissements de nature publique ou privée.

This is very useful for our analysis some of theses categories represent a

We chose to measure trough an indicator with a score, gentrification at least as we see it isn't a binary variable it's a relative notion. It makes way more sense to see it as which neighbourhood is more gentrified than an other instead of simply saying which one is or isn't. To do this we built an indicator that itself is composed of 3

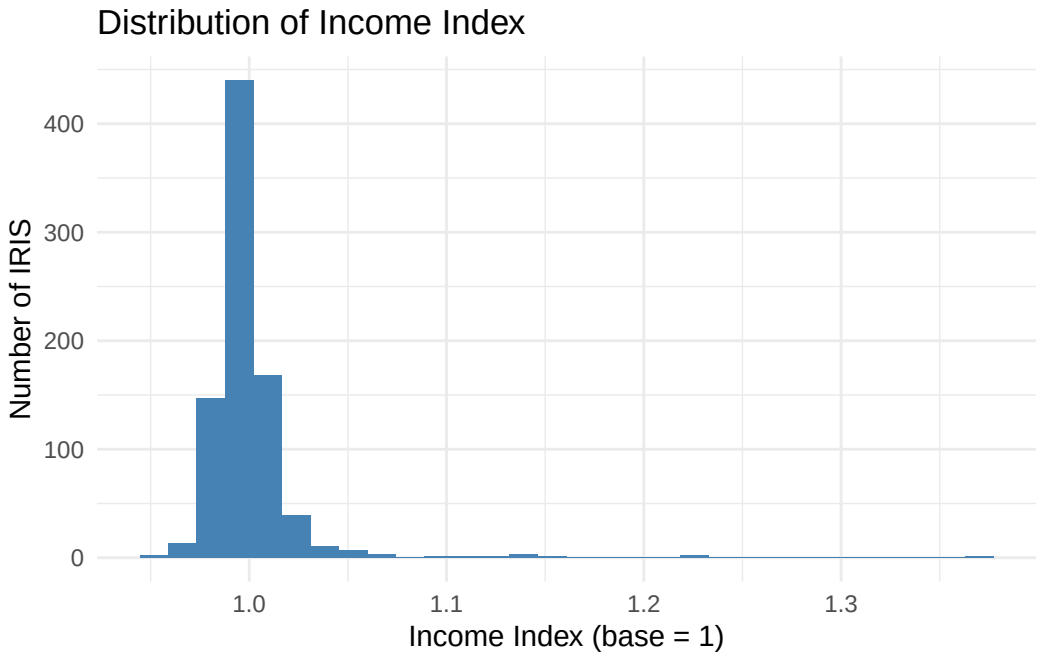
Prices: From the DVF dataset we observed changes in housing prices over time. Using price per sqm we can create an indicator attempting to rank iris codes by the distance at which their growth in price sqm per year diverges from the mean. To do this first we created an

Revenue: From the Filosofi dataset we constructed another indicator that captures changes in come structure at the IRIS level. This indicator is composed of two parts: a median income component and a poverty composition component.

We first look at mdeian income as a measure of economic change that can reveal gentrification. For each IRIS and year, we observe the median income. We compute annualised income growth rates between observed years, this produces a per-year growth rate that we can compare across

time. From this we can create for each year a Paris wide average income growth. We define relative income growth as the distance of an IRIS from this city wide average. Positive values indicate that income in an IRIS has grown faster than the Paris average in that year, with negative values showing underperformance comapred to this average. Finally, we aggregated these distances over time into a cumulative relative income index with a base value of 1. An index value of 1 indicates that the median income in the IRIS followed the Paris wide average exactly over the period of time, values over 1 indicate overperformance and those udner 1 indicate underperformance. This accounts for half of the total indicator created from FiloSofi.

```
# A tibble: 10 x 2
   CODE_IRIS index_price
   <chr>           <dbl>
 1 751156009        1.37
 2 751166124        1.22
 3 751135015        1.22
 4 751207910        1.15
 5 751103801        1.14
 6 751156010        1.14
 7 751207906        1.14
 8 751197317        1.13
 9 751186925        1.11
10 751207801        1.09
```



The second part is constructed from the percentage of the population who have an income deemed under the "poverty line". Each IRIS has a value of its poverty share each year. For each year we can rank all the IRIS codes on this poverty share. From this we can observe how this changes from year to year and create a cumulative score of change in this poverty share each year. Those IRIS that exceeded the average score highly on the final poverty rank change index.

Finally create a final index based on the two parts "revenue" and "poverty share". We do this by standardising each indicator such that they are both Z distributed. Thus we create a weighted linear combination concluding with the final index S = 0.6Z(inc) + 0.4Z(pov), a higher S indicates a stronger improvement than average of income changes and poverty share changes.

Final Construction of the GentriScore: (How all were summed)

4. Conclusion