

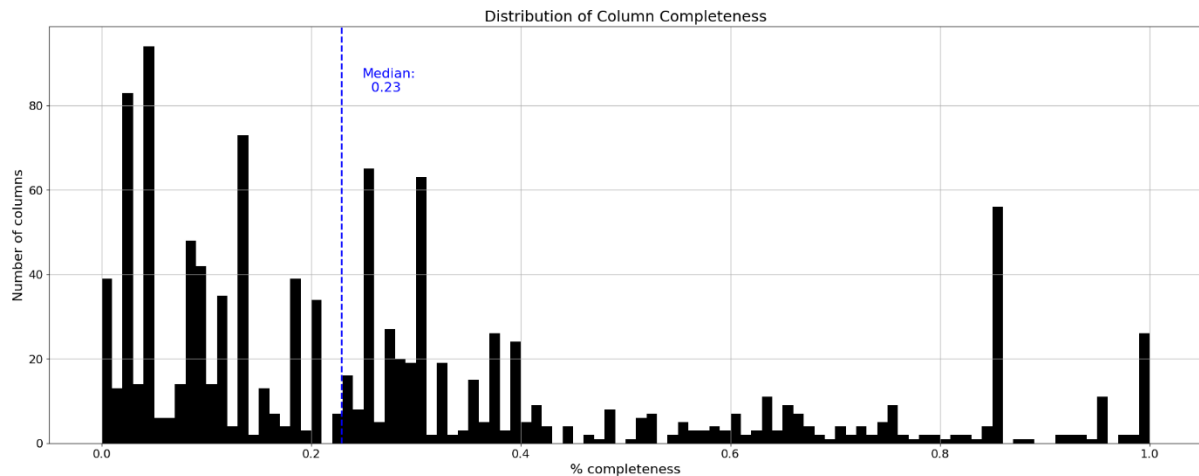
French Tax Revenue Exploration and Prediction

Written by Aleksandr Lazutin

Dataset Analysis

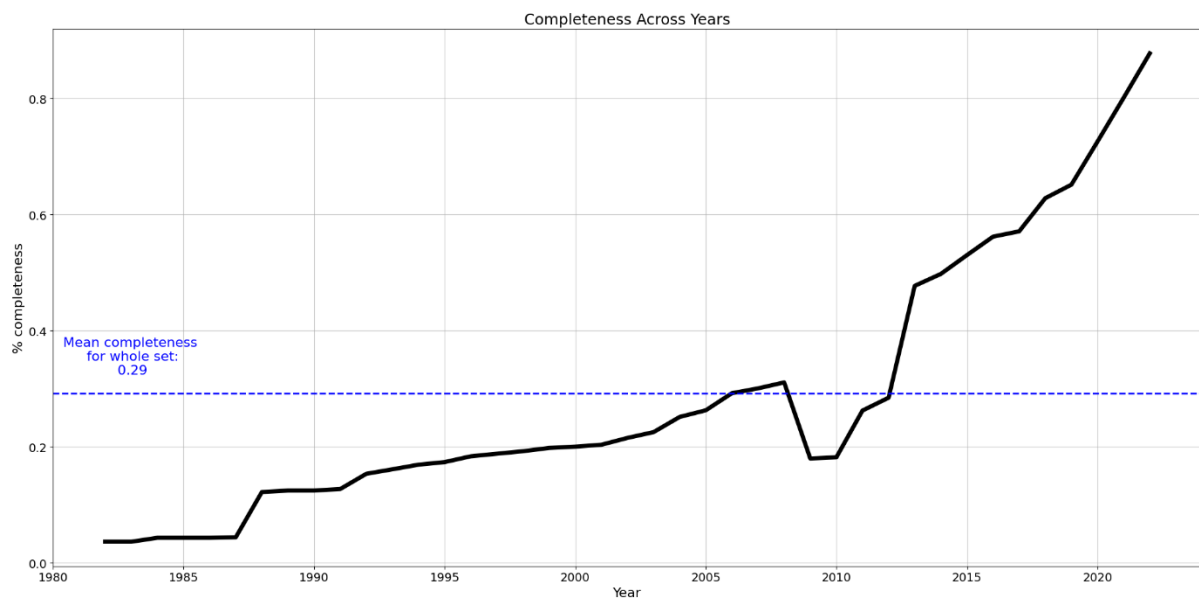
Given the large dataset, it is important to assess the completeness and quality of the data.

The dataset contains 1,181 columns. By excluding values such as "", ".", None, and blank cells, and then dividing by the total number of rows (1,528,922), we can determine the percentage of columns with data. The distribution of this percentage is as follows:



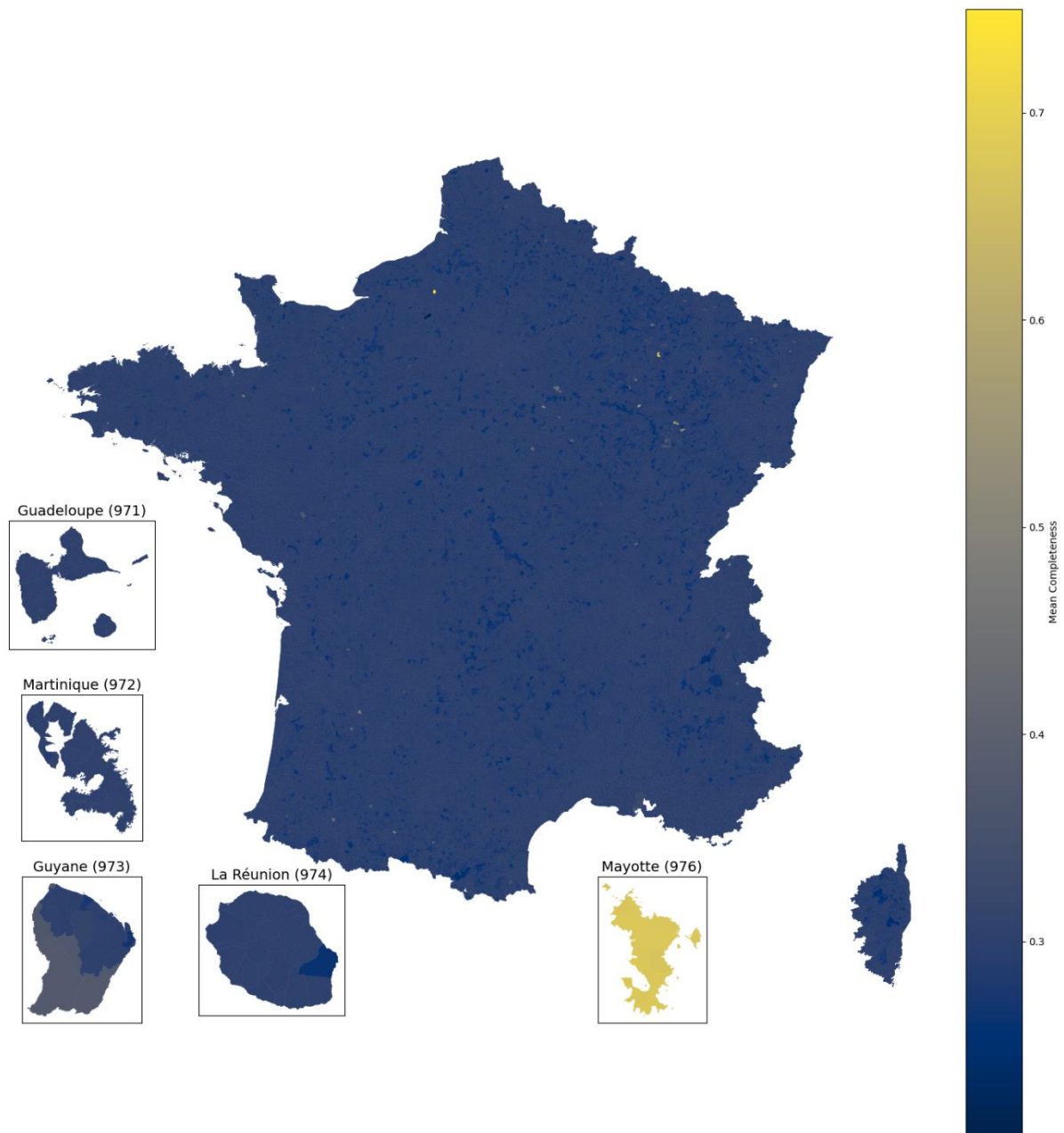
Median of 0.23 means that half of the columns are filled less than 23%.

Data integrity also differs from year to year:



Data completeness is growing over the years and peaks in 2022 with 87%, while on average dataset is 29% full (or 61% empty if you are pessimist)

Data completeness for municipalities



Steps taken to increase the quality of data:

- Filling missing values
- Cross-checking data with other datasets
- Correcting wrong values

Municipality Rankings

Total tax amount is counted by using the following columns from the dataset:

- FB - COMMUNE / MONTANT REEL
- FNB - COMMUNE / MONTANT REEL
- TAFNB - COMMUNE / MONTANT REEL NET
- TH - COMMUNE / MONTANT REEL DONT THP/E AU PROFIT DE L ETAT
- CFE - COMMUNE / PRODUIT REEL NET

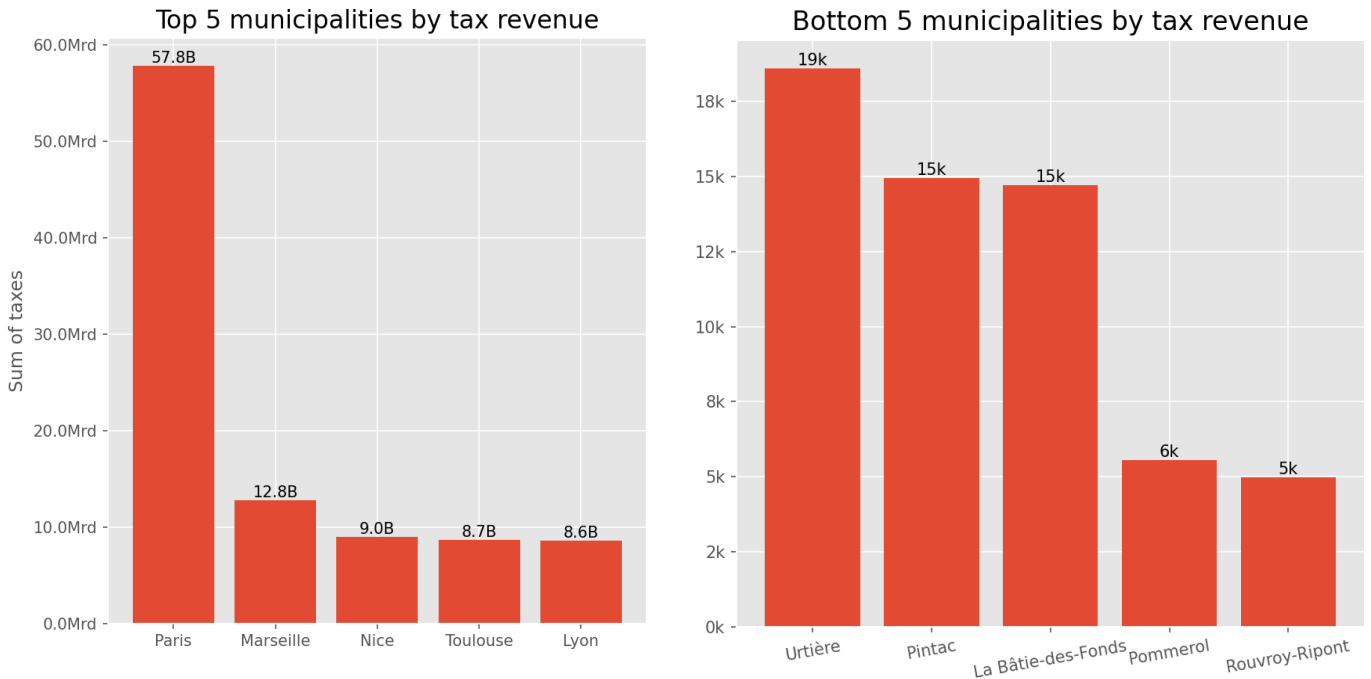
These columns cover four direct local taxes (the two TF, the TH and the CFE):

- TF - Taxe Foncière: Property tax which includes the taxes FB and FNB.
 - FB - Foncier Bâti: The property tax on developed land and buildings, paid by owners.
 - FNB - Foncier Non Bâti: The property tax on undeveloped land, paid by owners.
 - TAFNB - Taxe Annuelle sur les Locaux à usage de Bureaux, de Locaux Commerciaux et de Locaux de Stockage: An annual tax on office, commercial, and storage premises in certain regions.
- TH - Taxe d'Habitation: A residential tax paid by occupants of a property, whether owners or tenants.
- CFE - Cotisation Foncière des Entreprises: A local tax paid by businesses based on the rental value of their business premises.

Other types of taxes are excluded from calculation because they have small impact on the total sum of taxes (<0.2% from total).

Top 5 / bottom 5 municipalities by tax revenue

For bottom 5 municipalities I picked only those, which exists for 40 years (present in the whole dataset). It removes zero values and ensures fair comparison between top and bottom.

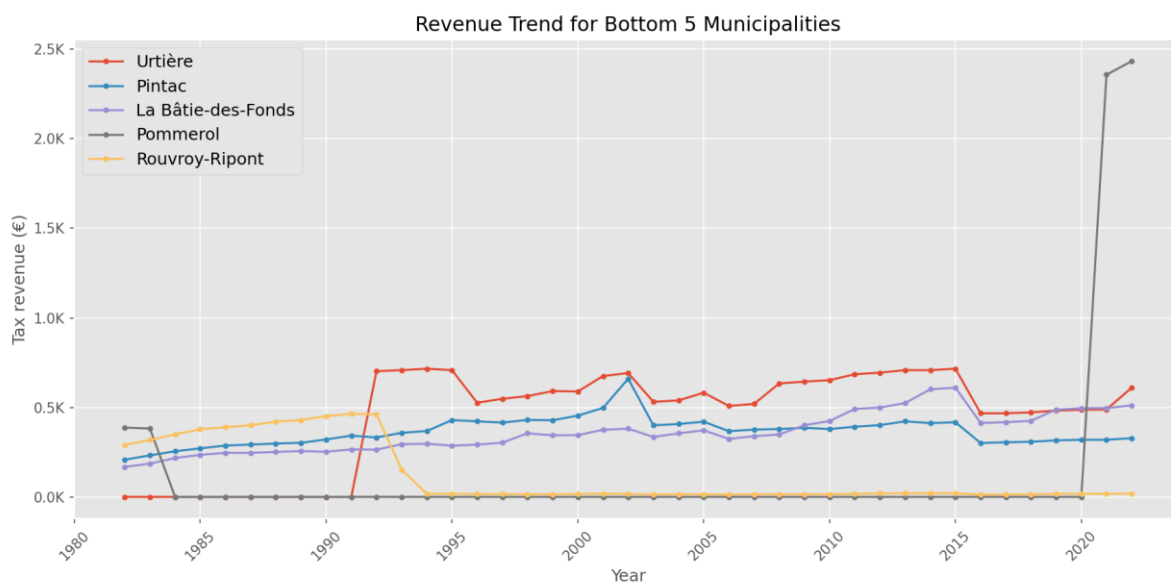
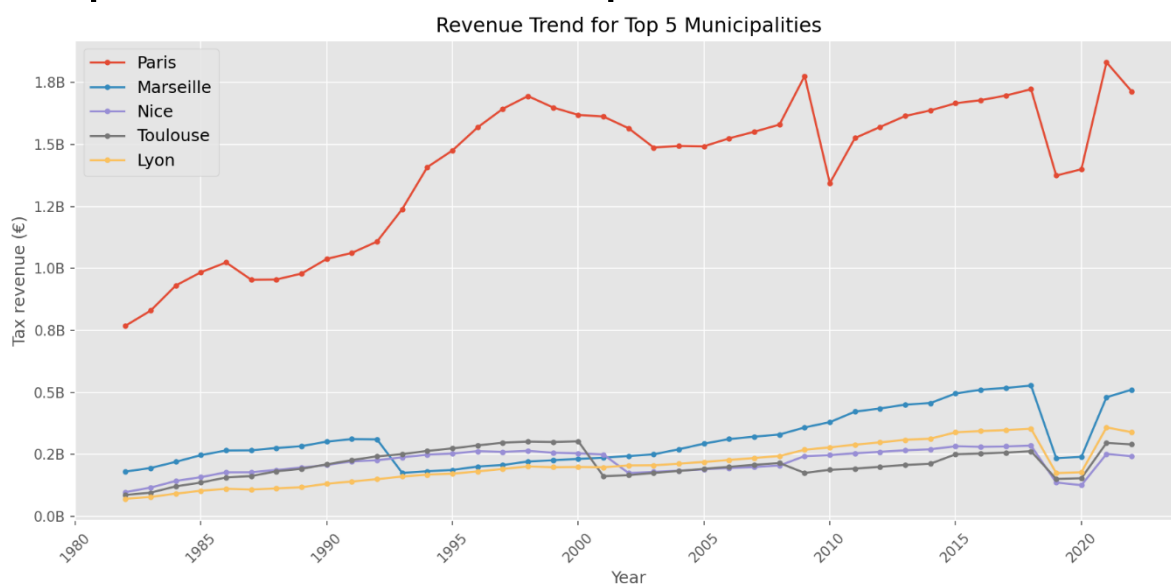


The disparity in tax revenues between the top and bottom municipalities is substantial. Paris, the highest revenue generator, collects €57.8 billion, which dwarfs the €19,000 collected by Urtière, the highest among the bottom five.

Major cities like Paris, Marseille, Nice, Toulouse, and Lyon are significant economic hubs, contributing the highest tax revenues. This can be attributed to their larger populations, higher business activity, and more robust economic infrastructure.

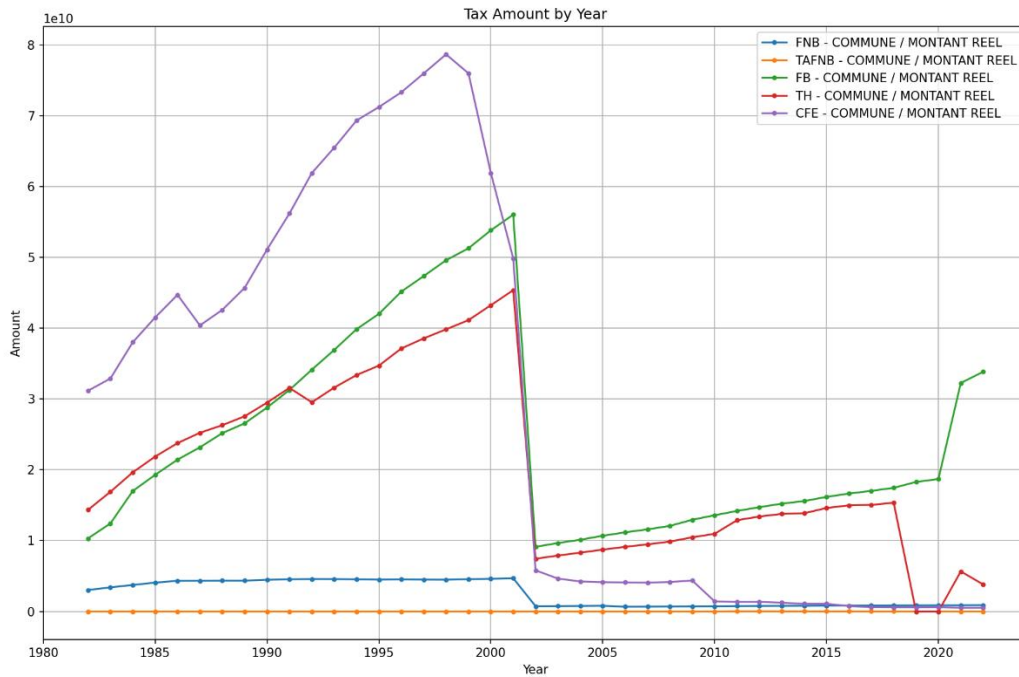
The disparity in tax revenues between Paris and the top five municipalities compared to Urtière and the bottom five municipalities is striking. Paris alone generates approximately €57.8 billion in tax revenue, vastly overshadowing even the other top cities like Marseille (€12.8 billion) and Nice (€9.0 billion). In contrast, Urtière, the highest among the bottom five, collects only €19,000, with the lowest, Rouvroy-Ripont, at €5,000.

Top 5 / bottom 5 municipalities revenue trends

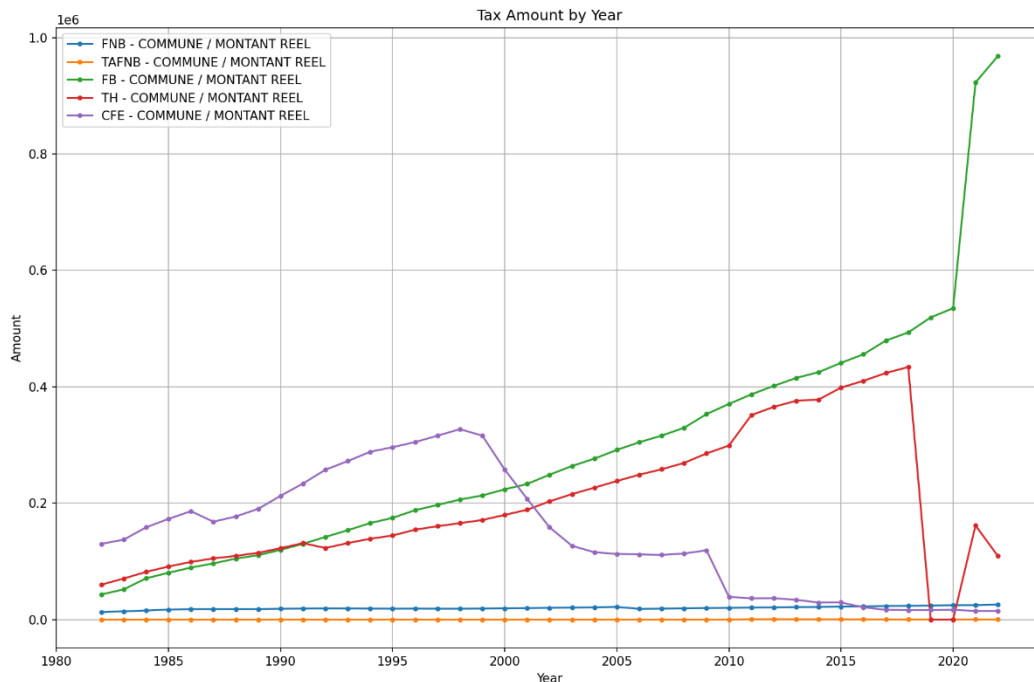


Top 5 revenues municipalities are more affected compared to the bottom 5 municipalities by the 2018-2019 removal of the TH - Taxe d'Habitation and the following COVID-19 crisis.

Revenue Growth



This drop between 2001-2002 is caused by France changing their currency from franc to euro with conversion rate of 6.55957.

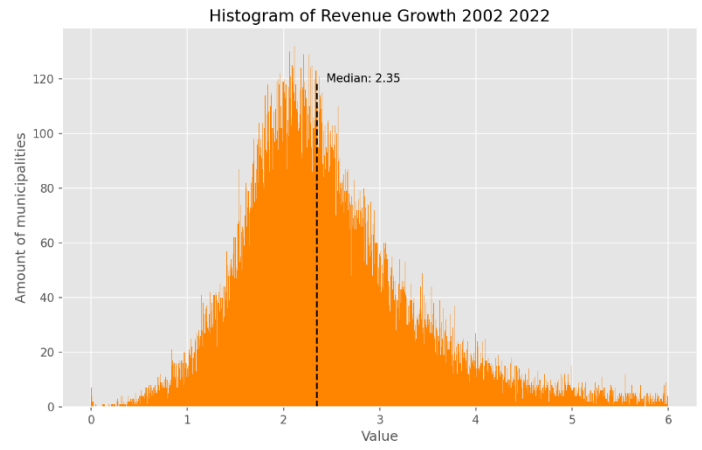
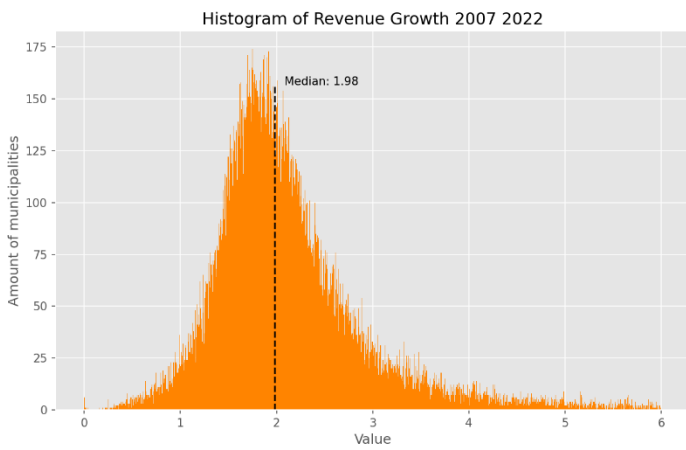
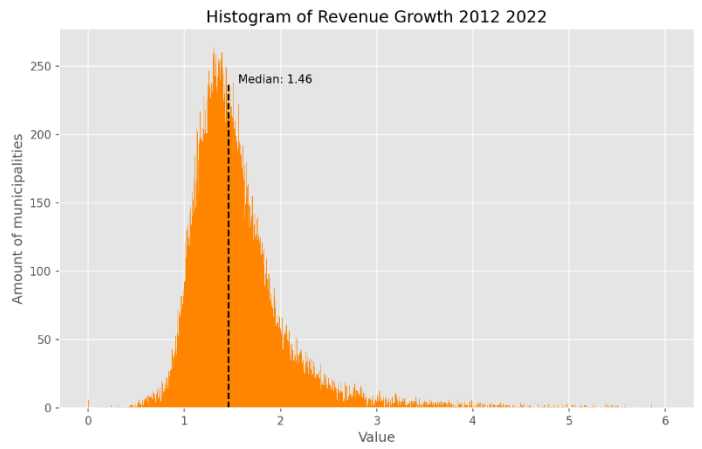
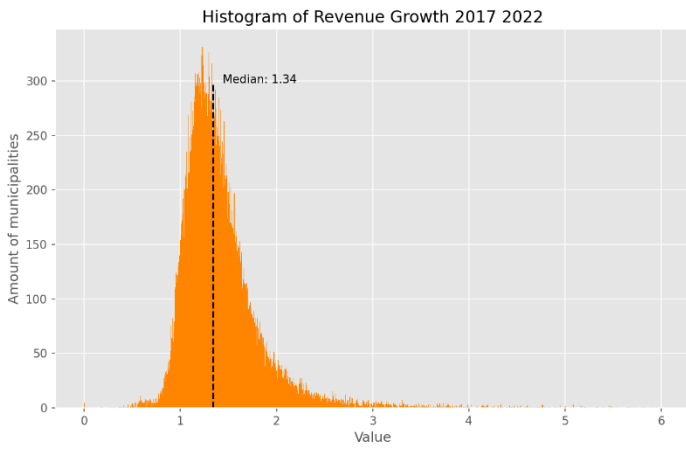


After converting everything to euro

TH drop could be explained by the following law:

<https://www.connexionfrance.com/practical/is-tax-dhabitation-still-due-for-2019/437221>

Distribution of revenue growth for municipalities over the past 5, 10, 15, and 20 years.



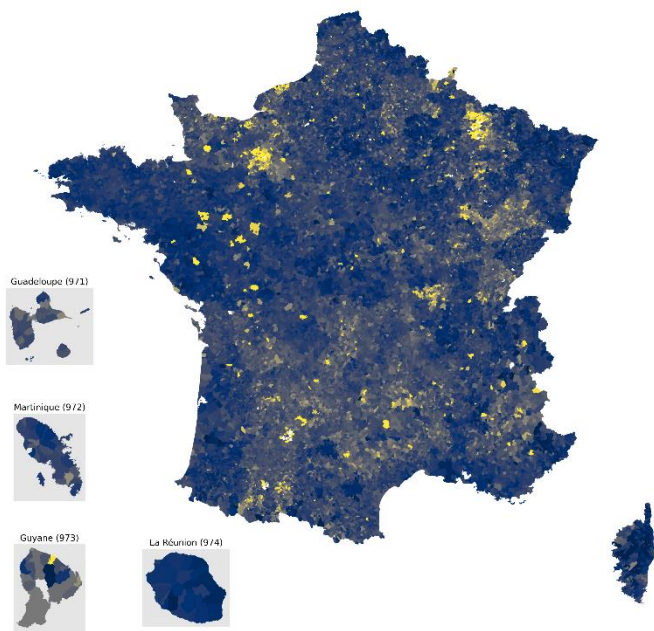
Revenue growth is calculated as:

$$\text{Revenue growth} = \frac{\text{Revenue in 2022}}{\text{Revenue in } N},$$

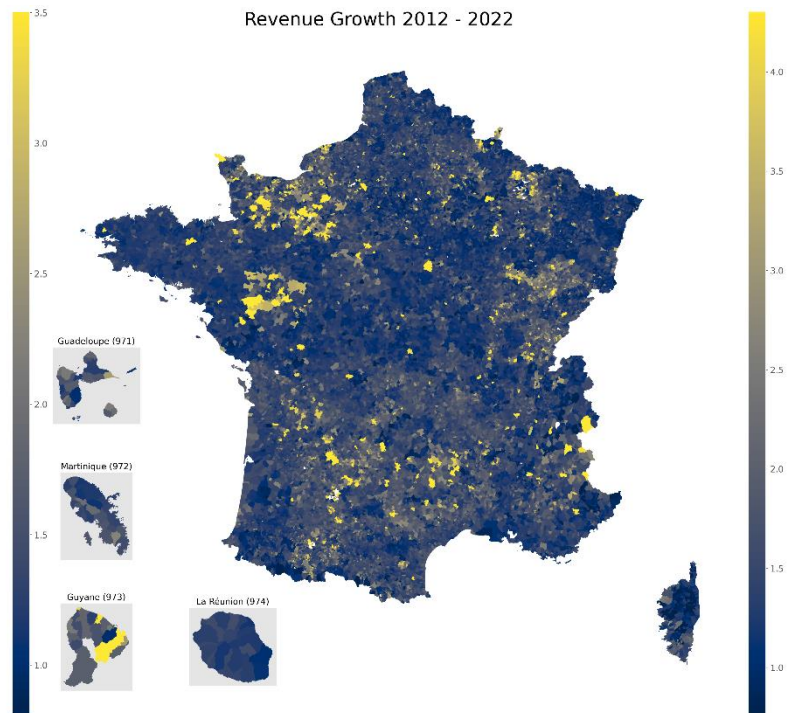
where N corresponds to the respective years in the past.

The key observation is that the median of these distributions increases as the number of years increases. This suggests that municipalities have experienced consistent revenue growth over time. The longer the period considered, the greater the overall revenue growth observed.

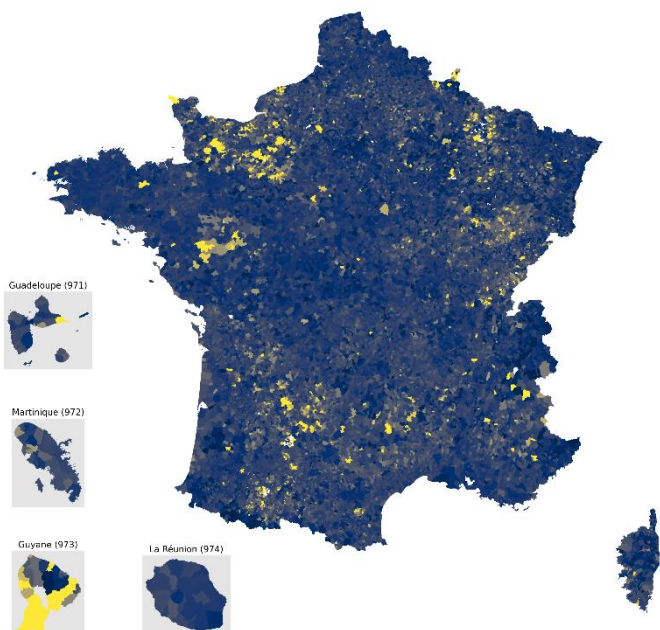
Revenue Growth 2017 - 2022



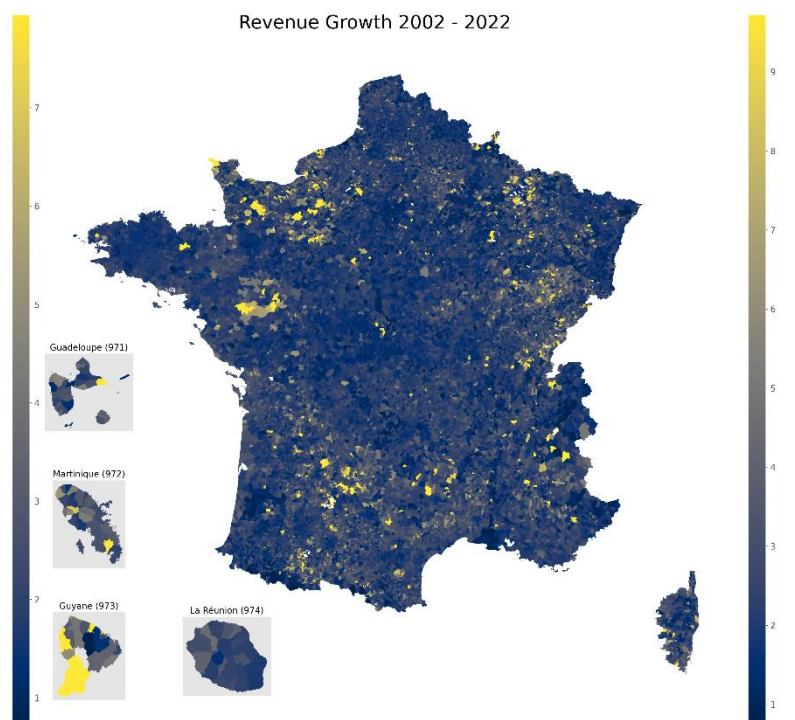
Revenue Growth 2012 - 2022



Revenue Growth 2007 - 2022



Revenue Growth 2002 - 2022



The areas with the highest revenue growth are consistently located in the same regions on the map, indicating that certain departments are continually generating more revenue. Some of these differences can be attributed to factors such as population size and geological characteristics.

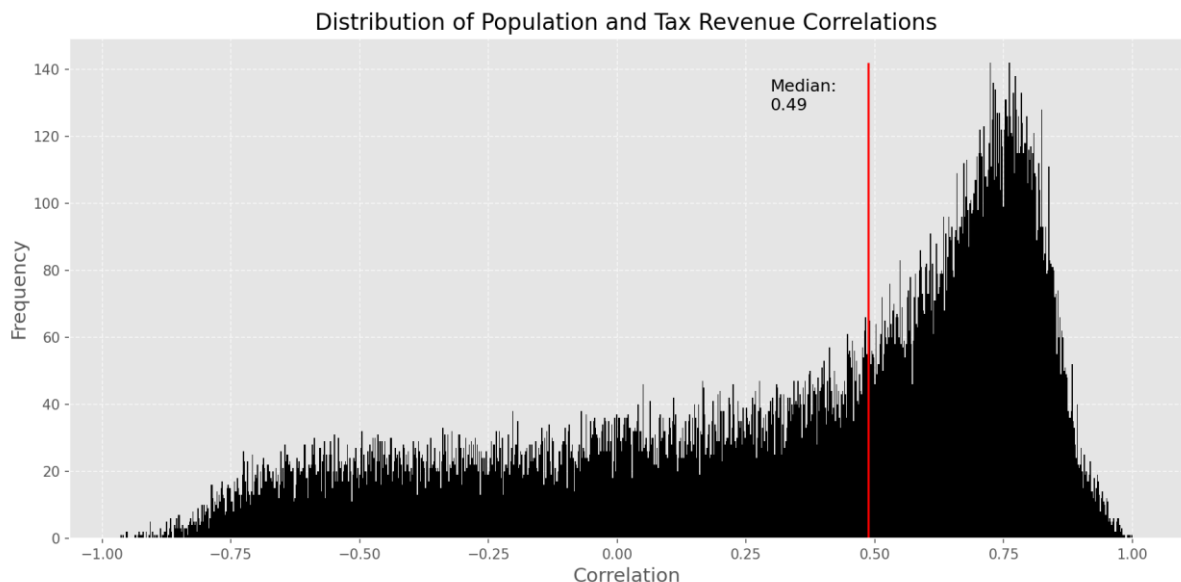
Municipalities and departments with the most significant revenue growth over the past 20 years.

Municipality name	Code INSEE	Revenue in 2002	Revenue in 2022	Growth 2002 2022
Demandolx	04069	763	171445	224.698558
Bourniquel	24060	218	39228	179.944954
Silly-en-Gouffern	61474	5501	958694	174.276313
Auvillers-les-Forges	08037	1230	180855	147.036585
Beaumont-Hague	50041	169440	21536376	127.10325

Department name	Code	Revenue in 2002	Revenue in 2022	Growth 2002 2022
Guyane	973	1720752	4509086	11.63
Alpes-de-Haute-Provence	4	395932	672905	4.69
Hautes-Pyrénées	65	162727	331508	4.39
Lot	46	134155	346386	4.25
Haute-Saône	70	76743	186778	4.2

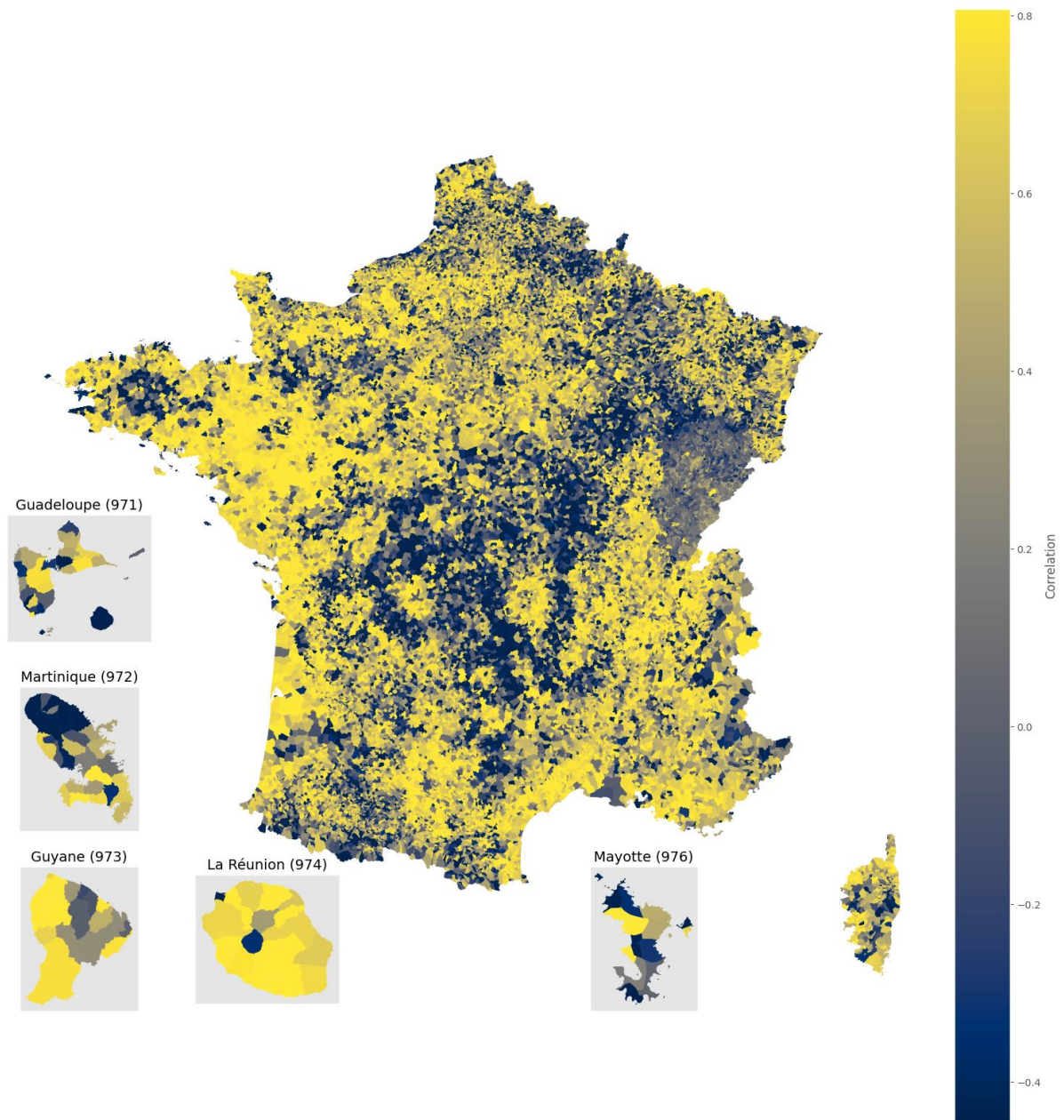
Correlations

To determine if there's a correlation between population size and tax revenue, I created this chart.



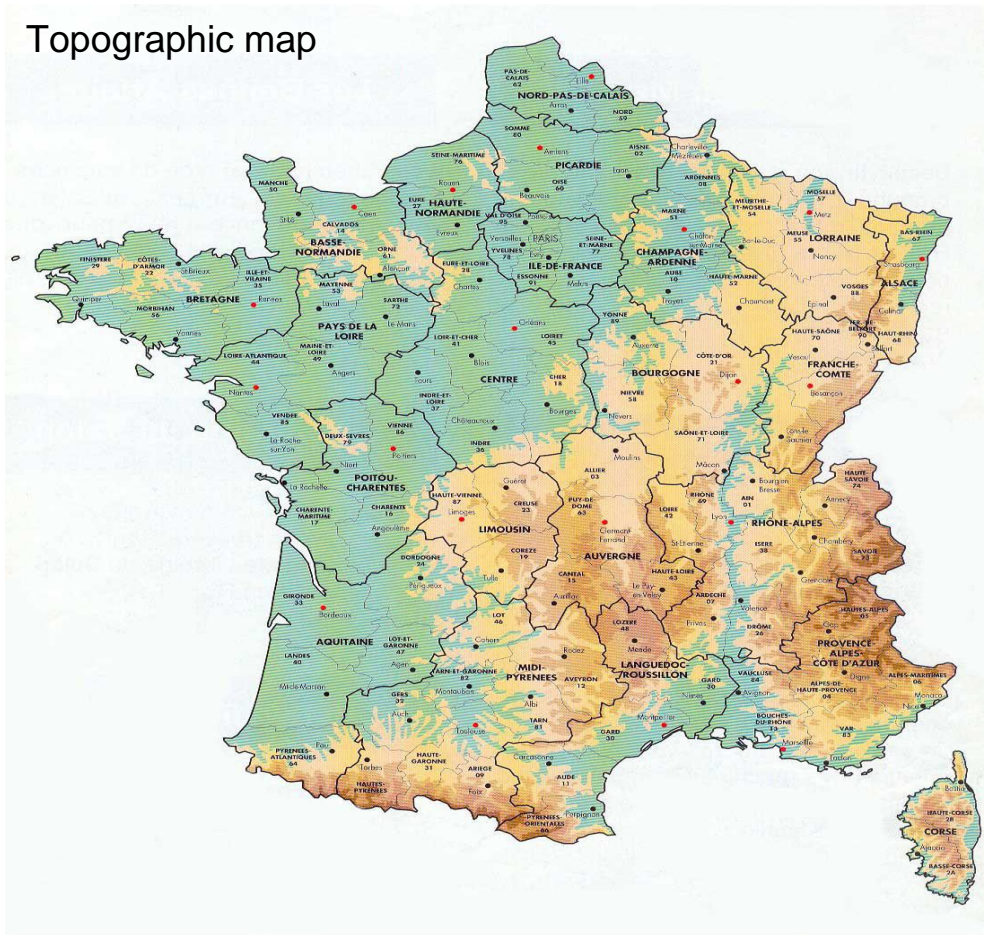
The chart shows that most of the correlation values are greater than 0, indicating that there is a positive correlation between population size and revenue for most municipalities. This suggests that as the population of a municipality increases, its revenue tends to increase as well.

Correlation map

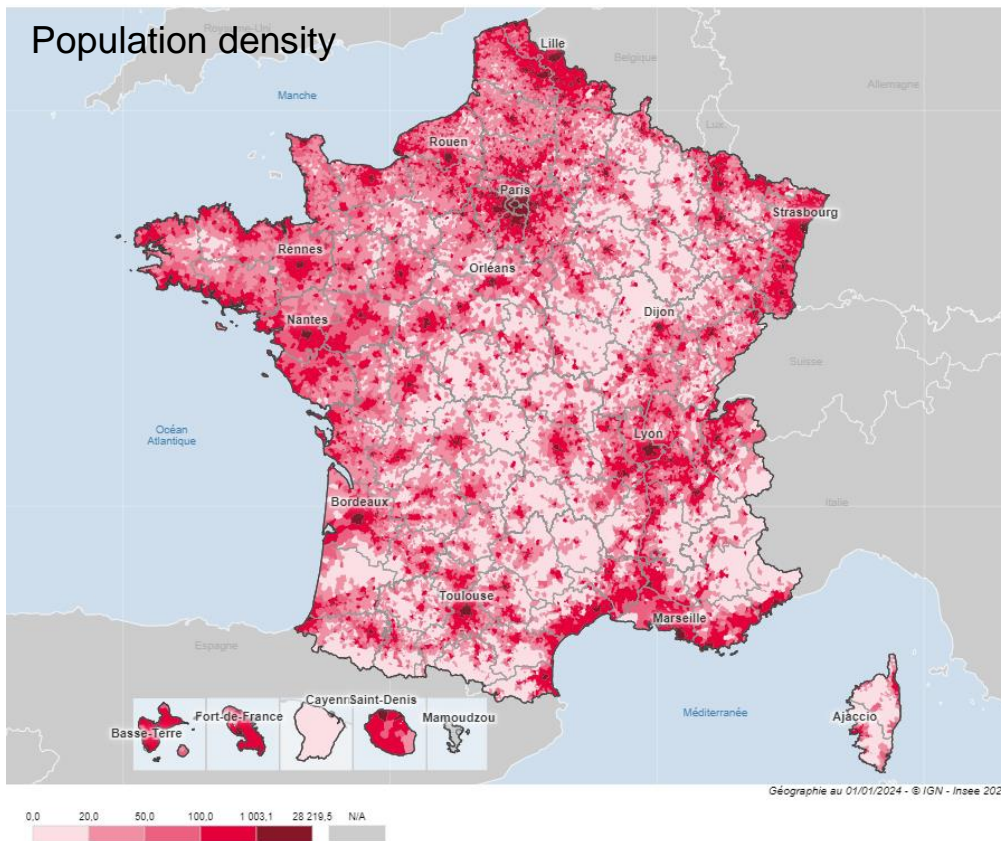


The differences in correlation values can be attributed to the geographical features of France. The central, eastern, and south regions are characterized by mountainous terrain, which can impact population distribution and economic activities, thereby affecting the correlation between population size and revenue in these areas.

Topographic map



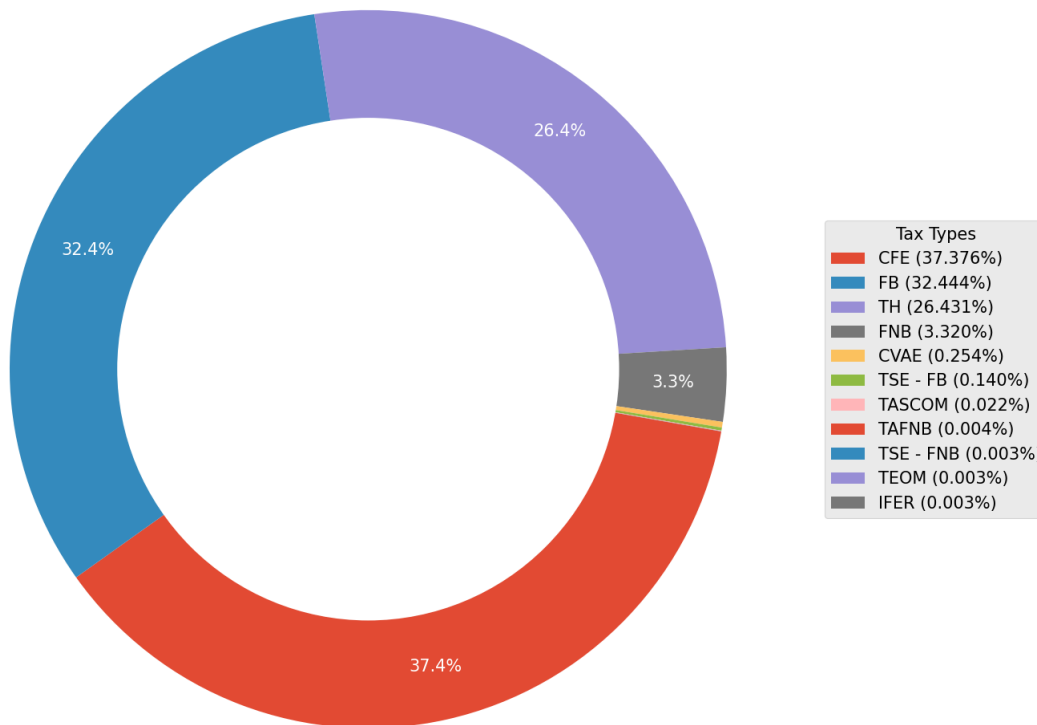
Population density



Significant Taxes

The most significant taxes in municipalities by total value:

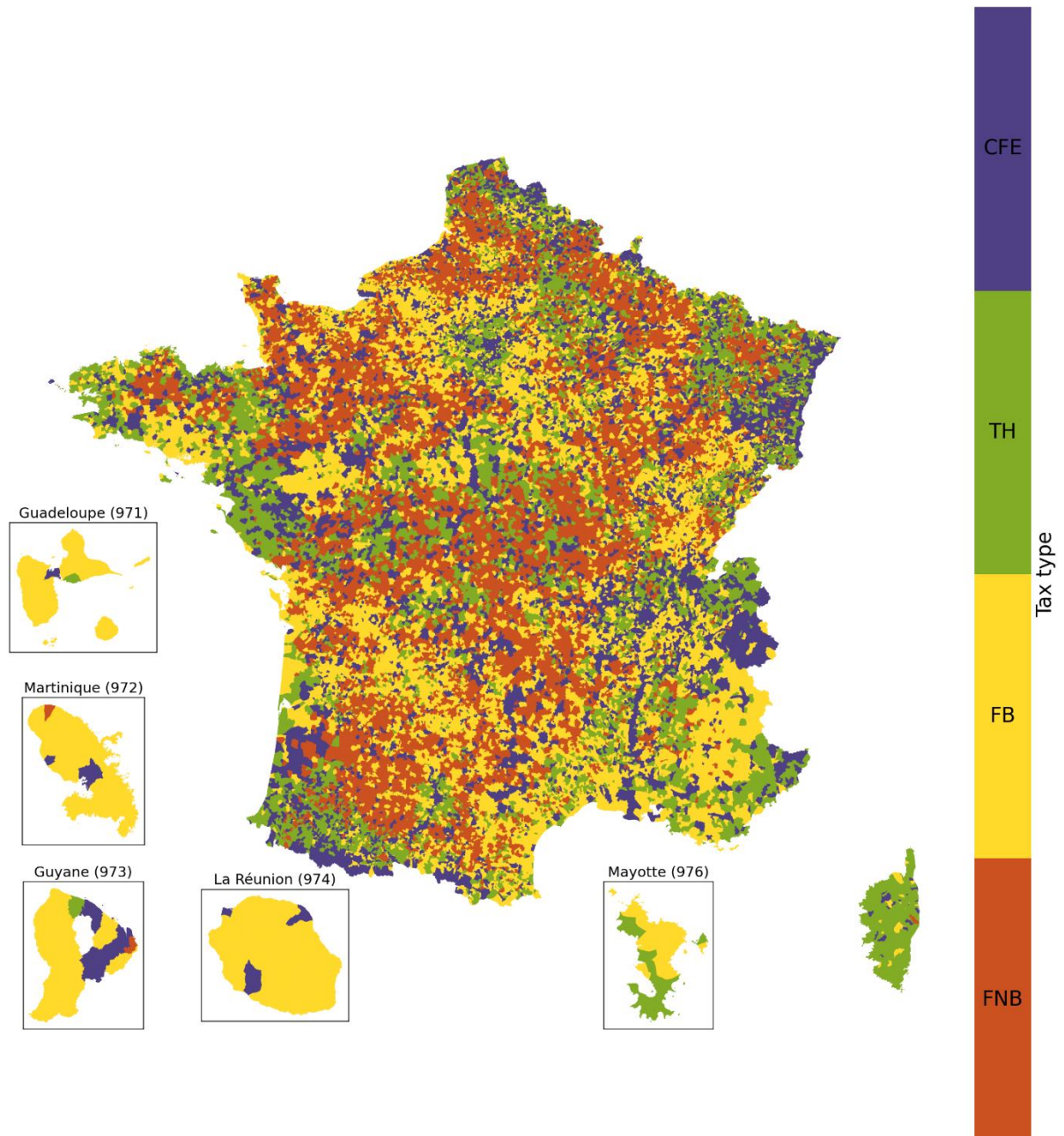
Municipalities Tax Types



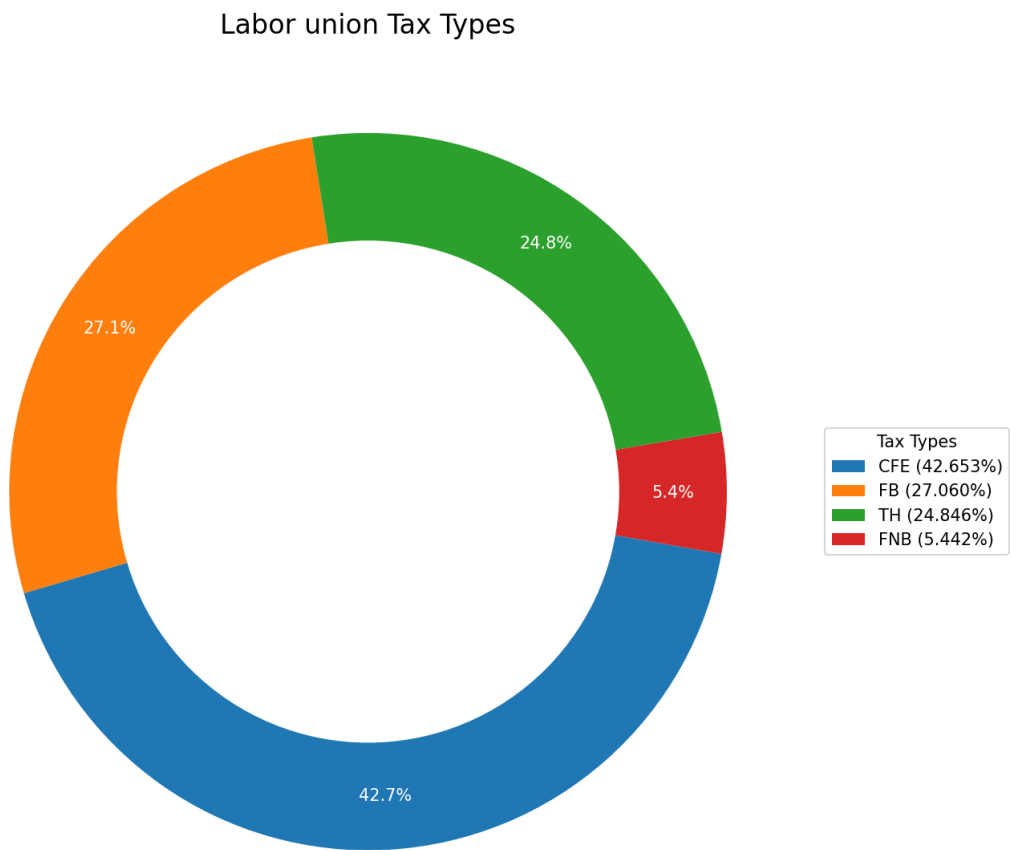
Although the CFE tax generates the highest revenue overall, it is the largest tax in only 17% of municipalities.

Predominant types of taxes in municipalities

FNB	32%
FB	27%
TH	24%
CFE	17%



The most significant taxes in labour unions by total value:

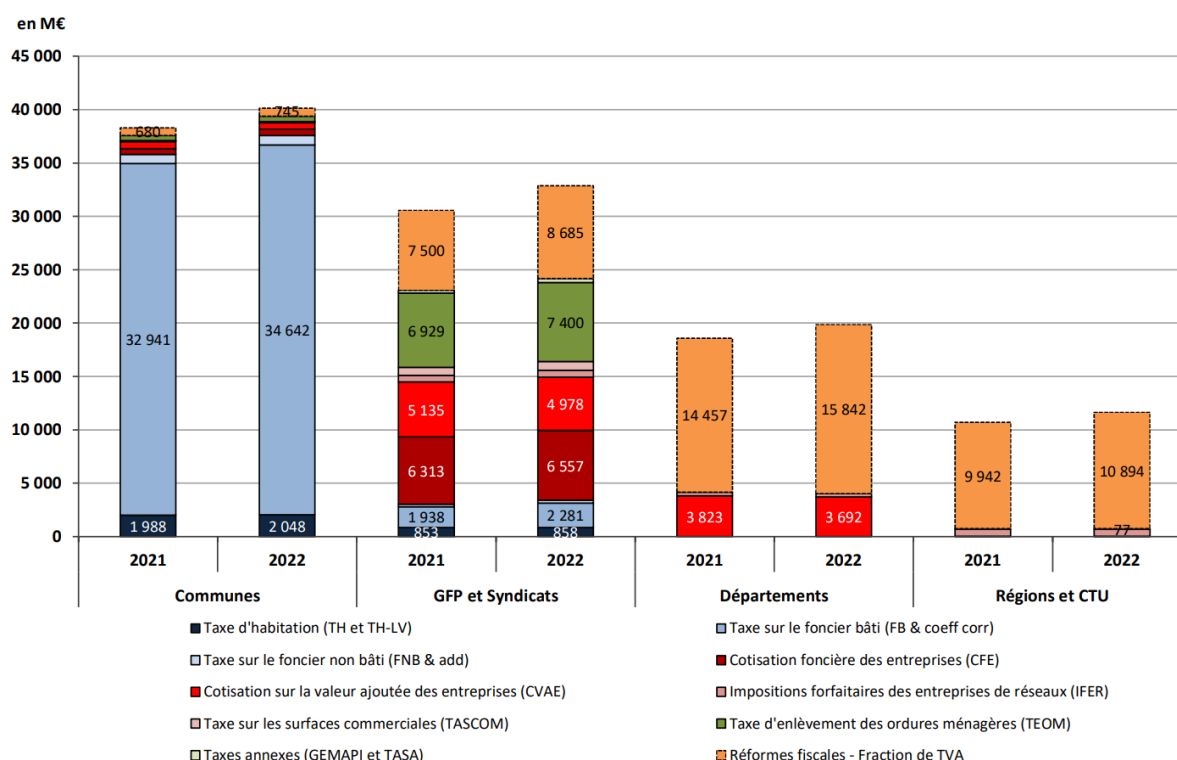


Predominant types of taxes in labour unions

FNB	35%
CFE	30%
TH	20%
FB	15%

Although the CFE (Cotisation Foncière des Entreprises) tax generates the highest revenue overall, it is the primary source of revenue in only 30% of all labour unions.

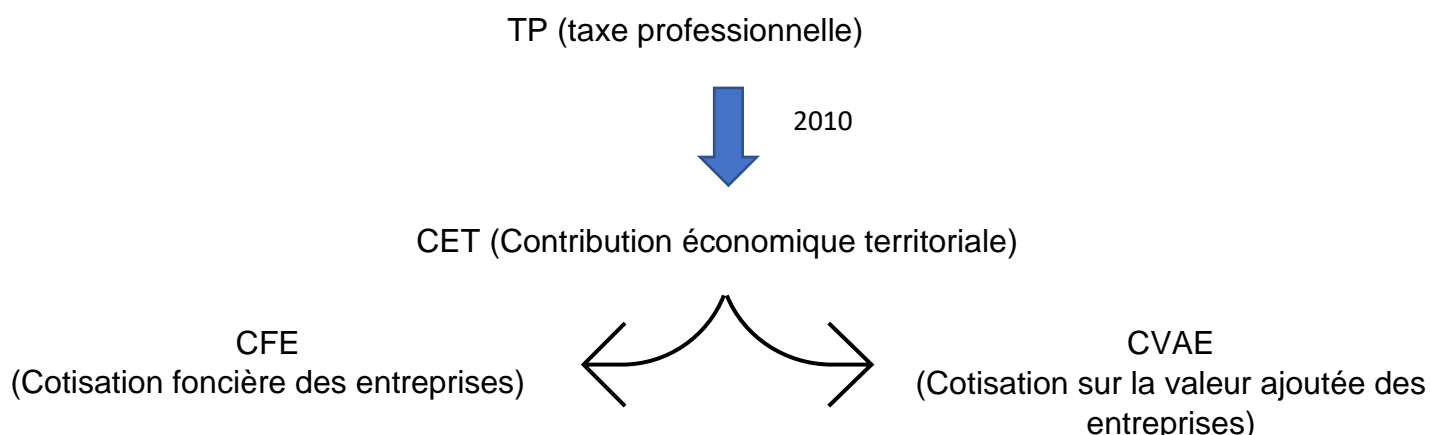
Taxes in 2021 and 2022

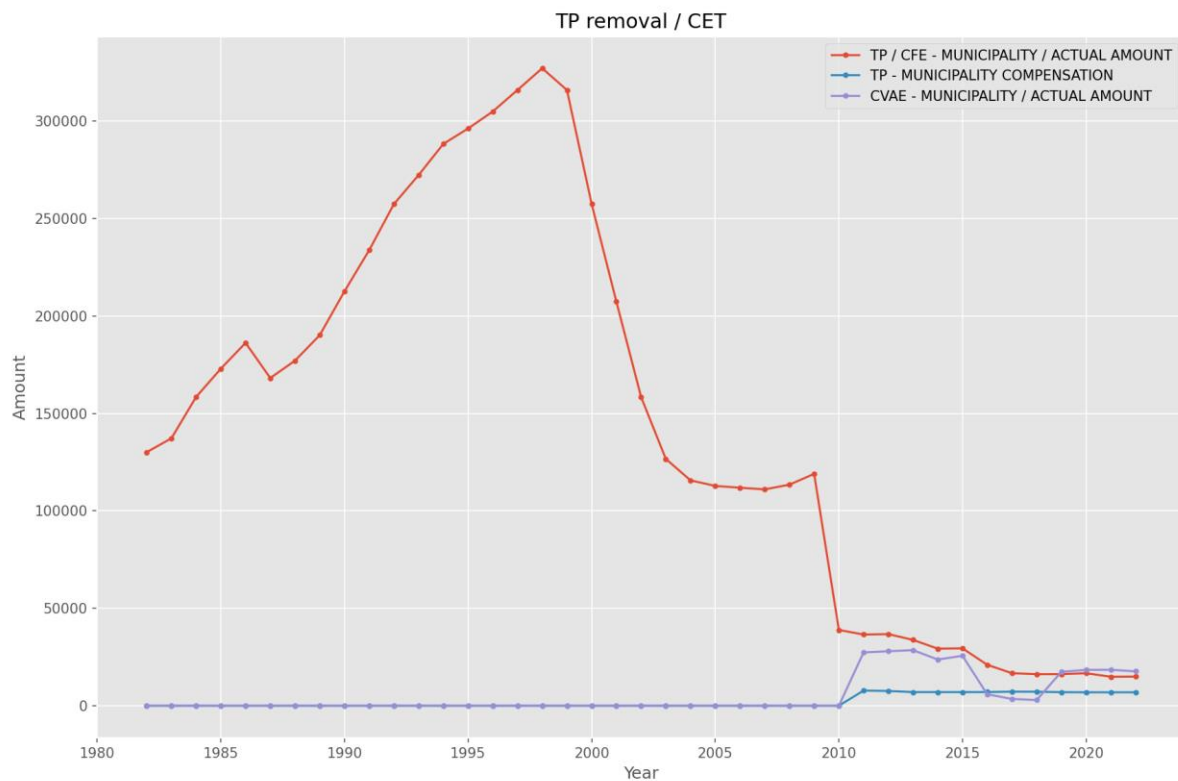


Picture from: https://www.collectivites-locales.gouv.fr/files/Accueil/DESL/BIS_175%20Fiscalité%20locale%202022%20VF_web.pdf

Professional Tax Reform

Since the data set does not have a dedicated column for TP, I will conduct the analysis under the assumption that pre-2010 CFE earnings are indeed TP. This is due to the fact that TP has been replaced by CET: CFE + CVAE.





Taxe Professionnelle (TP)

Definition and History

- Direct local tax paid by businesses to local authorities (communes, departments, and regions)
- Created in 1975 to replace the older "patente" tax
- Existed until 2010

Key Features

- Applied to individuals and legal entities conducting regular business activities in France, with some exceptions (e.g., agricultural businesses)
- Initially based on company's payroll and rental value of assets
- After 2002, primarily based on rental value of company's assets
- Represented about 50% of tax resources for local authorities

Reasons for Removal

- Penalized investment and capital-intensive industries
- Complex, leading to high management costs for businesses and administration
- Increasingly costly for the state due to various capping and reduction mechanisms

Transition to New System (2010)

The TP was replaced by the Territorial Economic Contribution (CET), consisting of two new taxes:

- Cotisation Foncière des Entreprises (CFE)
- Cotisation sur la Valeur Ajoutée des Entreprises (CVAE)

Comparison of TP, CFE, and CVAE

Aspect	Taxe Professionnelle (TP)	Cotisation Foncière des Entreprises (CFE)	Cotisation sur la Valeur Ajoutée des Entreprises (CVAE)
Tax Base	Rental value of company's fixed assets	Rental value of properties	Added value produced by the business
Calculation Method	Tax rate (set by commune) x Tax base	Similar to TP, but only on property values	Based on company's turnover and added value, with progressive rates
Scope	All businesses	Most businesses, some exemptions for small enterprises	Businesses with turnover exceeding €500,000
Impact on Investment	Criticized for penalizing investment in equipment and assets	Less impact on investment decisions	Designed to be less detrimental to investment
Local Authority Control	Significant control over tax rate	Some control over rates	Rate set nationally, less local control

Impact of the Reform

On Municipal Finances

- Varied impact across municipalities
 - Larger decreases for areas with strong industrial bases or many large businesses
 - Less effect on smaller municipalities with fewer large businesses
 - Potential benefits for urban areas with high property values (due to CFE)
- New dynamics in business tax revenue allocation due to CVAE redistribution
- Government implemented compensation mechanisms for most affected municipalities

On Business-Local Government Relations

- Reduced local control over tax rates, altering economic development strategies
- Aimed to reduce tax burden on businesses, especially in manufacturing
- Shifted more tax burden onto other local taxes, potentially affecting residents

The elimination of the Professional Tax led to significant changes in the relationship between businesses and local governments. Prior to the reform, municipalities had some control over their tax rates, which could lead to tax competition between areas to attract businesses.

The new system reduced this local control, as the CVAE rate is set nationally. This change altered the dynamics of local economic development strategies. Additionally, the reform aimed to reduce the tax burden on businesses, particularly in manufacturing sectors, to boost competitiveness. However, it also shifted more of the tax burden onto other local taxes, potentially affecting residents.

The long-term effects included changes in municipal investment patterns, as some local governments found their revenue streams less predictable or robust under the new system. This reform thus not only impacted municipal finances but also had broader implications for local governance, economic development strategies, and the distribution of tax burdens between businesses and residents.

Conclusion

The transition from the Taxe Professionnelle to the new system represented a significant shift in French local business taxation. While aiming to address the shortcomings of the TP, it introduced new complexities and challenges in balancing municipal finances, business competitiveness, and local economic development.

Prediction Model

TASK: Develop and train a machine learning model capable of forecasting the yearly municipal tax revenues.

To accurately predict yearly municipal tax revenues, we need to forecast the individual tax revenues that contribute to the annual total: FNB, TAFNB, FB, TH, CFE, and CVAE.

Data for other tax types (TEOM, IFER, TASCOT) is incomplete, with more than 95% of values missing in some cases.

To solve this task, I will use linear regression and decision tree models.

I'll build several variants: to predict revenue based on the data from same year and to predict revenue up to 3 years to the future.

Data preparation:

For each type of tax I took related columns, removed all filler data “.”, removed columns with more than 80% zeros and more then 95% NULL

Train / Test / Validation – 0.6 / 0.25 / 0.15

Results:

	Linear regression				Decision tree			
	Test		Validation		Test		Validation	
	MAE	R2	MAE	R2	MAE	R2	MAE	R2
FNB	2260	97.26%	2255	97.62%	873	98.75%	868	98.79%
TAFNB	358	32.34%	342	22.90%	4	99.65%	4	99.39%
FB	155069	98.25%	158241	95.49%	27838	99.73%	41171	90.89%
TH	29672	99.98%	27670	100.00%	77756	99.80%	67225	99.89%
CFE	25782	98.91%	7672	92.42%	4225	99.97%	3968	93.13%
CVAE	16	100.00%	22	100.00%	690	99.58%	278	99.33%

These are the results for predicting revenue within the same year using all tax-related columns. Linear regression performs worse at predicting TAFNB revenue but is more accurate for CVAE and TH.

On average, decision trees yield better results than linear regression.

The strong performance can be attributed to the high correlation between the input and output variables, such as the taxable base and tax revenue.

To minimize the impact of highly correlated features, I will leave only those features with a correlation between 0 and 0.7.

Without highly correlated columns:

	Linear regression				Decision tree			
	Test		Validation		Test		Validation	
	MAE	R2	MAE	R2	MAE	R2	MAE	R2
FNB	3267	95.98%	3339	95.57%	1563	96.66%	1582	97.21%
TAFNB	468	1.65%	459	2.25%	337	2.97%	347	3.95%
FB	157004	97.71%	164840	98.33%	91569	98.55%	84287	96.85%
TH	169219	95.40%	118659	98.92%	176053	98.09%	148675	99.14%
CFE	20284	99.79%	21865	99.88%	5022	99.97%	670	98.51%
CVAE	14	100.00%	15	100.00%	207	100.00%	291	99.08%

The most affected prediction is for TAFNB, where the validation R2 accuracy dropped dramatically from 99.39% to 3.95%.

Future years revenue prediction

	Linear regression						Decision tree						AVG
	Test			Validation			Test			Validation			
	+1y	+2y	+3y	+1y	+2y	+3y	+1y	+2y	+3y	+1y	+2y	+3y	
FNB	96.35	93.81	92.43	95.48	94.97	94.16	97.57	97.25	96.02	96.66	95.87	96.19	95.56
TAFNB	83.35	56.64	30.98	74.74	53.48	47.39	75.15	65.38	58.65	85.72	81.68	74.6	65.64
FB	99.96	96.09	97.79	99.88	88.03	97.35	99.96	99.9	92.04	99.98	98.72	99.77	97.45
TH	99.87	99.54	97.12	99.91	98.34	99.61	92.63	99.49	72.43	89.87	99.5	66.36	92.88
CFE	0	91	44.09	0	75.39	88.18	99.95	95.42	98.95	98.53	87.87	89.51	72.4
CVAE	22.35	74.93	83.43	51.53	0	0	99.98	85.37	100	86.41	8.24	80.17	57.7
Average	66.98	85.33	74.3	70.25	68.36	71.11	94.2	90.46	86.34	92.86	78.64	84.43	80.27

Difficulties arise when predicting TAFNB, CVAE, and CFE because data cleaning results in insufficient data to train the model effectively. While decision trees handle most predictions better than linear regression, they perform worse in predicting TH.

Specifically:

Linear Regression: Performs better for predicting TH.

Decision Tree: Performs better for predicting FNB, TAFNB, FB, CFE, and CVAE.

I chose Linear Regression and Decision Tree models because they are computationally efficient to train, which is crucial given my limited computational resources. Additionally, they demonstrated good results for datasets with sufficient data.

Conclusion

The analysis of French municipal tax revenues reveals several critical insights with far-reaching implications:

- **Revenue Growth Trends:** Most municipalities have experienced consistent revenue growth over time, though growth patterns vary considerably across regions and municipality sizes.
- **Geographic Influences:** The correlation between population, geography, and tax revenue underscores the importance of tailored economic development strategies. One-size-fits-all policies are unlikely to address the unique challenges faced by diverse regions, particularly in mountainous or rural areas.
- **Tax Reform Impact:** The 2010 removal of the Taxe Professionnelle (TP) and its replacement with new taxes (CFE and CVAE) significantly impacted municipal finances. While aimed at improving business competitiveness, these changes shifted tax burdens and altered the dynamics between local governments and businesses.
- **Tax Composition:** The relative importance of different taxes (FNB, FB, TH, CFE) varies among municipalities and labour unions, reflecting diverse economic structures across France.
- **Predictive Modeling:** Machine learning models, particularly decision trees, show promise in forecasting municipal tax revenues. However, prediction accuracy varies across tax types and time horizons, with some taxes (like TAFNB and CVAE) proving more challenging to predict.
- **Future Considerations:** Ongoing changes in tax policies, such as the phasing out of the Taxe d'Habitation, will likely continue to impact municipal finances and require adaptive strategies from local governments.

These insights highlight the need for flexible policies that can address the diverse economic landscapes across French municipalities while ensuring fiscal stability and promoting economic growth.