



NOTARY Transactions DATABASE

Etalab

ETALAB is a Open Source Data
listing all the Real Estate Transactions
done in France during the past 5 Years.

<https://app.dvf.etalab.gouv.fr/>

That initiative have been developed in France few years after the UK.

Few RE business were developed using that data

Business Opportunities

Famous UK Company using similar data

ZOOPLA

<https://www.zoopla.co.uk/>

Famous French Companies

Meilleurs Agents

<https://www.meilleursagents.com/>

Se Loger

<https://www.seloger.com/>

Département

Commune

Section cadastrale

Parcelle cadastrale

Date de la mutation

01/01/2015 - 31/12/2019

etalab



Etalab - GitHub/API



Why GitHub? Team Enterprise Explore Marketplace Pricing

Search

Sign in

Sign up

etalab / DVF-app

Watch 26

Star 277

Fork 55

Code

Issues 16

Pull requests 7

Actions

Projects 1

Security

Insights

Join GitHub today

GitHub is home to over 50 million developers working together to host and review code, manage projects, and build software together.

Sign up

master

5 branches

0 tags

Go to file

Code



marion-paclot Update faq.html

d983948 on 11 Aug 289 commits



db

Update build_db.sh

5 months ago



lib

Add script to generate a communes mapping from COG

17 months ago



static

Update faq.html

2 months ago



.env.sample

Add minimal Node.js server for front developers

17 months ago



.gitignore

Fix build_db

17 months ago



LICENSE

Create LICENSE

17 months ago



MAJ_semestrielle.md

Update MAJ_semestrielle.md

5 months ago



README.md

Update README.md

6 months ago



app.json

Add Heroku generated app.json

17 months ago



app.py

Passage du nom de la db en paramètre

11 months ago



config.csv.sample

Ajout de db_name en paramètre

11 months ago

About

Exploration des données DVF

app.dvf.etalab.gouv.fr

cadastre

dvf

ventes-immobilières

open-data

cartography

visualisation

Readme

Apache-2.0 License

Releases

No releases published

Packages

No packages published

Etalab data = CSV or JSON

http://api.cquest.org/dvf?code_commune=94068
<http://api.cquest.org/dvf?section=94068000CQ>
http://api.cquest.org/dvf?numero_plan=94068000CQ0110
<http://api.cquest.org/dvf?lat=48.85&lon=2.35&dist=200>
http://api.cquest.org/dvf?code_postal=89110&type_local=Maison

Demande de Valeur Foncière

Département
75 - Paris

Commune
Paris 9e Arrondissement

Section cadastrale
AA

Parcelle cadastrale

Date de la mutation
01/01/2015 - 31/12/2019

Télécharger les mutations de la section

etalab

JSON Données brutes En-têtes

Enregistrer Copier Tout réduire Tout développer (lent) Filtre le JSON

```
{
  "source": "DGFIP / Demande de Valeurs Foncières",
  "dernière_maj": "2020-04",
  "licence": "http://data.cquest.org/d_merales-dutilisation.pdf",
  "nb_resultats": 149,
  "results": [...]
}
```

DataBase

For Paris IXe we focused only on 'Appartements' transactions.

```
array(['Appartement', 'Local industriel, commercial ou assimilé', 'Dépendance', 'Maison', 'None'], dtype=object).
```

Check of Data types

Check non values and unique values for the most interesting columns

Deep dive in the 'classifications' to update and organize = 'Dépendance' & 'None'

Convert data types to an usable type - from string to int64 / float64

Order columns / drop non required columns / creation new columns for data analysis

Based on our preliminary study drop all non required rows

```
index1 = db[(db['type_local'] == 'Appartement') & (db['surface_carrez_total'] == 0) & (db['surface_reelle_bati'] == 0)].index
```

```
index2 = db[(db['type_local'] == 'Dépendance') & (db['surface_carrez_total'] == 0) & (db['surface_carrez_total'] < 9)].index
```

```
index3 = db[db['nature_mutation'] == 'Expropriation'].index
```

```
index4 = db[db['type_local'] == 'Local industriel, commercial ou assimilé'].index
```

```
index5 = db[db['type_local'] == 'Maison'].index
```

```
# Warning : we will have to consider surface reelle bati if Carrez =0
```

Start data analysis in Python before Tableau

Code

Let's go to Jupyter Notebook

DataBase

From our data cleaning exercise and organisation we extracted 6 differents database

db_general.to_csv

db_surface_carrez.to_csv

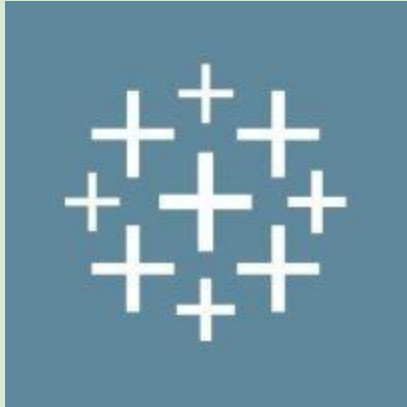
db_surface_carrez_not_outliers.to_csv

db_surface_bati.to_csv

db_surface_bati_not_outliers.to_csv

Paris9e_Expropriations.csv

Tableau



Tableau

Tableau - Notary-2020-09-25

FileDataServerWindowHelp

Connections

Add

Paris9e_Transactions_Carrez

Text file

Files

Use Data Interpreter

Data Interpreter might be able to clean your Text file workbook.

75109_000AA.csv

Paries9e_General.csv

Paris9e_Expropriations.csv

Paris9e_SeLogger.csv

Paris9e_Transactions.csv

Paris9e_Trans...ions_Bati.csv

Paris9e_Trans...s_Carrez.csv

New Union

Paris9e_Transactions_Carrez

Paris9e_Transactions_Carr...

Need more data?

Drag tables here to relate them. [Learn more](#)

Sort fields

Data source order

Show aliases

Show hidden fields

1000

rows

Abc	#	#	#	#	#	Abc	Abc	Abc	#	Abc	Abc
Paris9e_Transactions_Carrez_id_mutation	Paris9e_Transactions_Carrez_valeur_fonciere	Paris9e_Transactions_Carrez_nombre_pieces_pr...	Paris9e_Transactions_Carrez_surface_carrez_to...	Paris9e_Transactions_Carrez_surface_reelle_bati	Paris9e_Transactions_Carrez_surface_terrain	Paris9e_Transactions_Carrez_type_local	Paris9e_Transactions_Carrez_date_mutation	Paris9e_Transactions_Carrez_nature_mutation	Paris9e_Transactions_Carrez_adresse_numero	Paris9e_Transactions_Carrez_adresse_suffixe	Paris9e_Transactions_Carrez_adresse_i
2014-1059692	82 000,00	2	12,60	25,000	0,00	Appartement	07/01/2014	Vente	41,000	None	RUE D
2014-1059693	599 100,00	2	66,41	55,000	0,00	Appartement	06/01/2014	Vente	18,000	None	RUE D
2014-1059694	200 000,00	1	24,00	23,000	0,00	Appartement	07/01/2014	Vente	14,000	None	RUE LI
2014-1059698	90 000,00	1	10,79	10,000	0,00	Appartement	06/01/2014	Vente	34,000	None	RUE PI
2014-1059701	1 065 000,00	3	110,40	90,000	0,00	Appartement	16/01/2014	Vente	15,000	None	RUE VI
2014-1059702	382 000,00	2	45,70	46,000	0,00	Appartement	10/01/2014	Vente	32,000	None	RUE RI
2014-1059704	1 322 306,89	10	128,44	310,000	0,00	Appartement	07/01/2014	Vente	1,000	None	SQ LA
2014-1059707	138 000,00	2	40,50	41,000	0,00	Appartement	14/01/2014	Vente	6,000	None	RUE H
2014-1059714	540 000,00	3	59,75	60,000	0,00	Appartement	10/01/2014	Vente	10,000	None	RUE M
2014-1059719	1 198 000,00	5	143,75	140,000	0,00	Appartement	08/01/2014	Vente	56,000	None	RUE D
2014-1059733	231 000,00	2	30,07	31,000	0,00	Appartement	15/01/2014	Vente	13,000	None	RUE G
2014-1059738	250 000,00	2	30,28	30,000	0,00	Appartement	13/01/2014	Vente	7,000	None	RUE D
2014-1059739	340 000,00	2	37,16	32,000	0,00	Appartement	06/01/2014	Vente	43,000	None	RUE N
2014-1059743	1 424 400,00	6	153,30	148,000	0,00	Appartement	15/01/2014	Vente	14,000	None	RUE BI
2014-1059749	665 000,00	4	130,55	100,000	0,00	Appartement	15/01/2014	Vente	17,000	None	BD DE

Data Source

1 - Outliers

2 - Surf/Prix

Map

TransactionVSMarket

Map-HighTransaction

Map-LowTransaction

RangeOFFTransaction

Map Highest Transactions

Price Evolution/Year

Price Evolution/T

TransactionsVSPrice

Price Distribution

Vente Typologies-V2

Map Typologies

Arthur Lapey...

Tableau

DILEMMA

what database to use ?

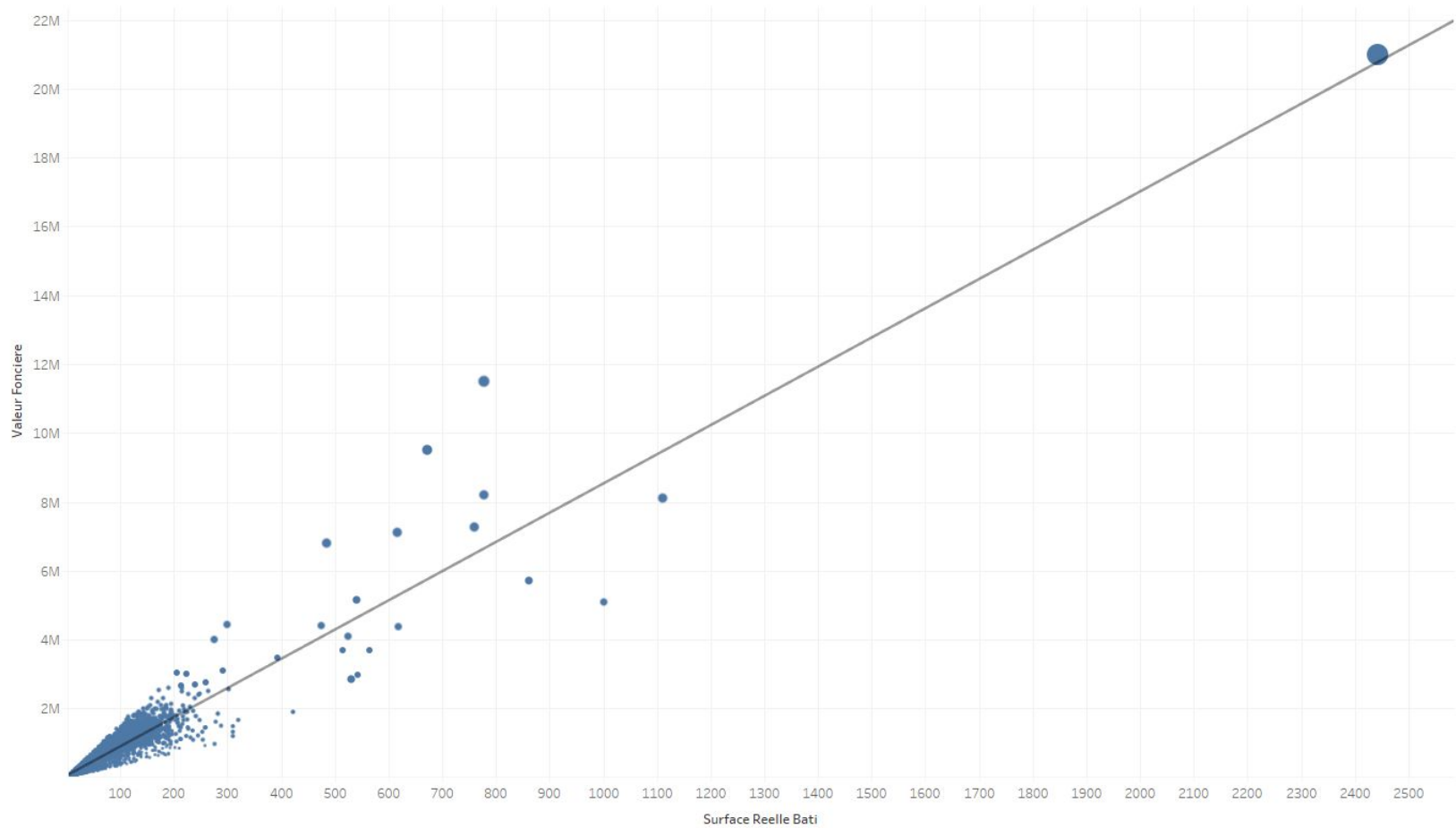
db_surface_carrez_not_outliers.to_csv

or

db_surface_bati_not_outliers.to_csv

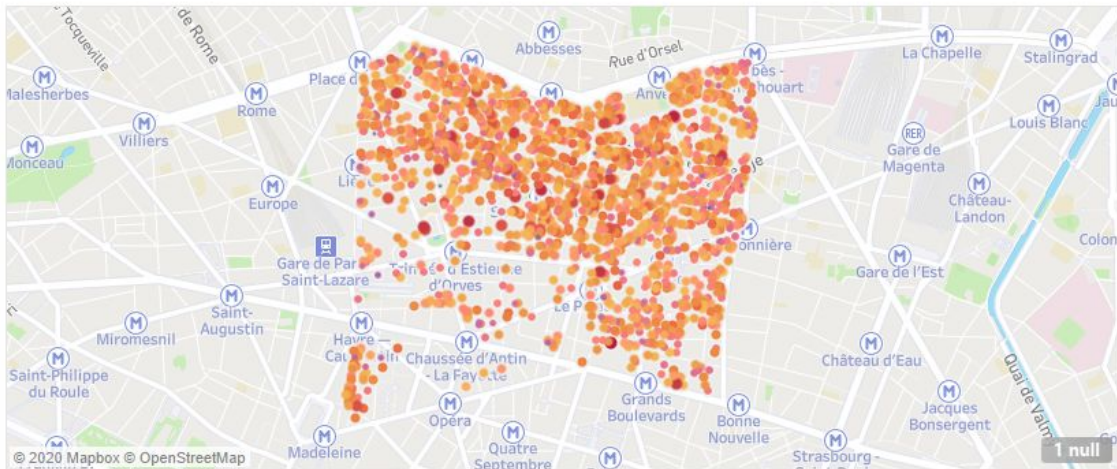
Correlation

1 - All Data



Market & Location

Map



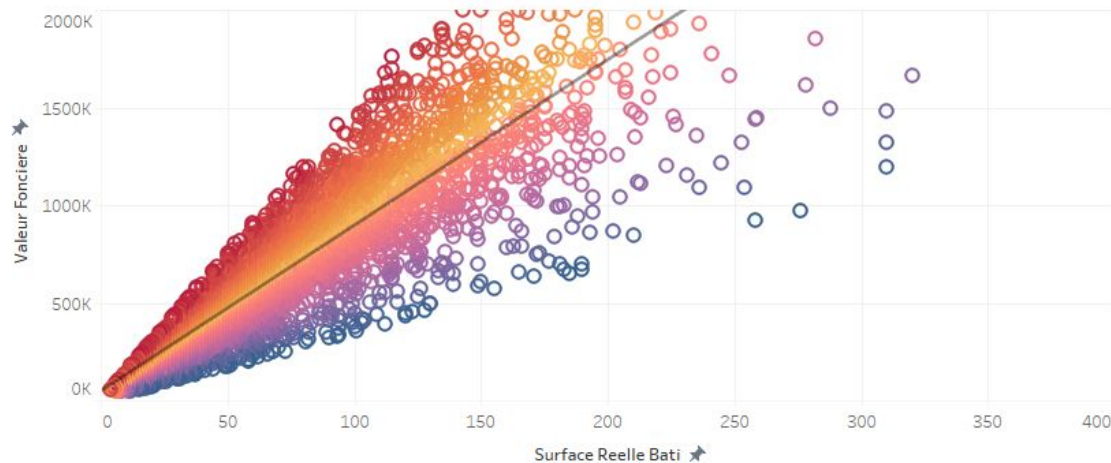
Avg. valeur_fonciere/surf..



Avg. valeur_fonciere/surf..

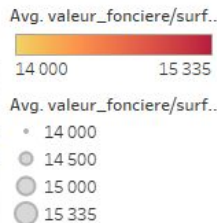
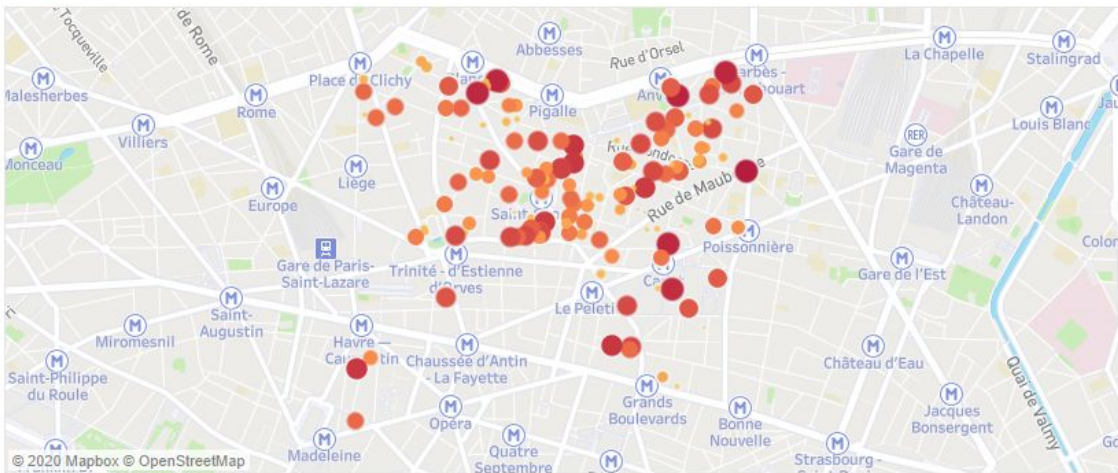
- 3 544
- 6 000
- 8 000
- 10 000
- 12 000
- 14 000
- 15 160

2 - Surf/Prix

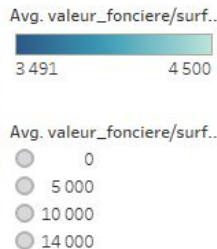
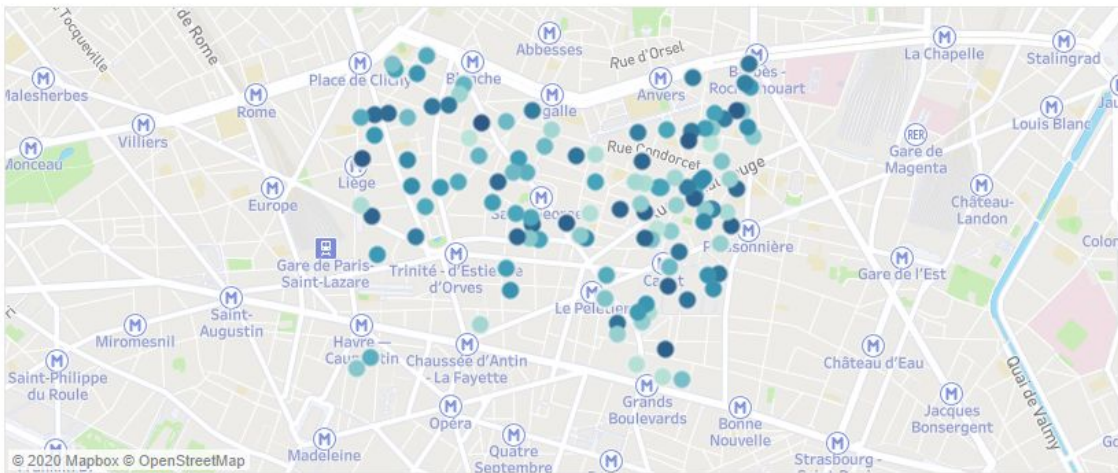


Uniform Distribution

Map-HighTransaction

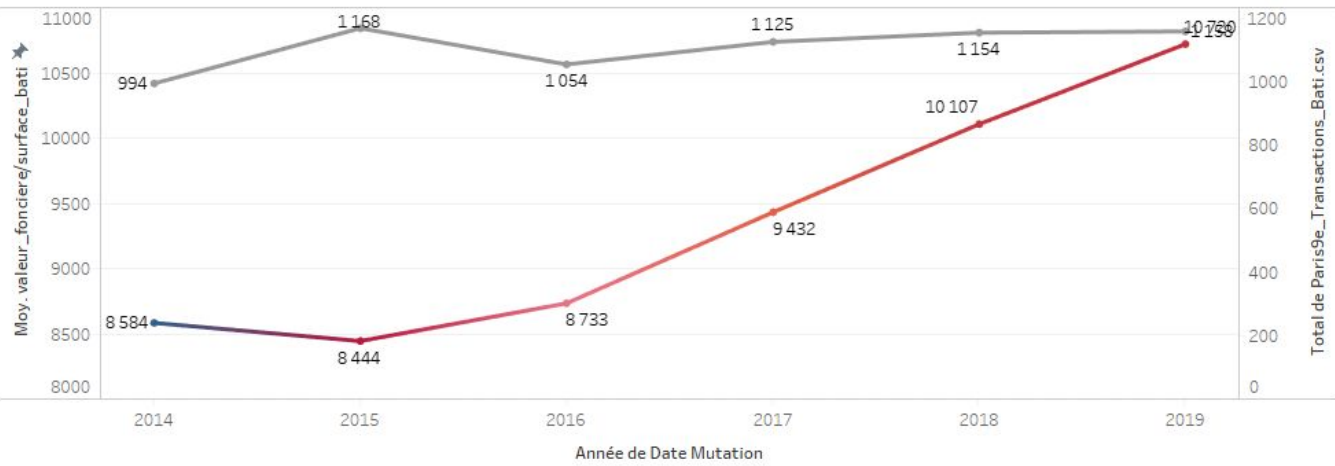


Map-LowTransaction

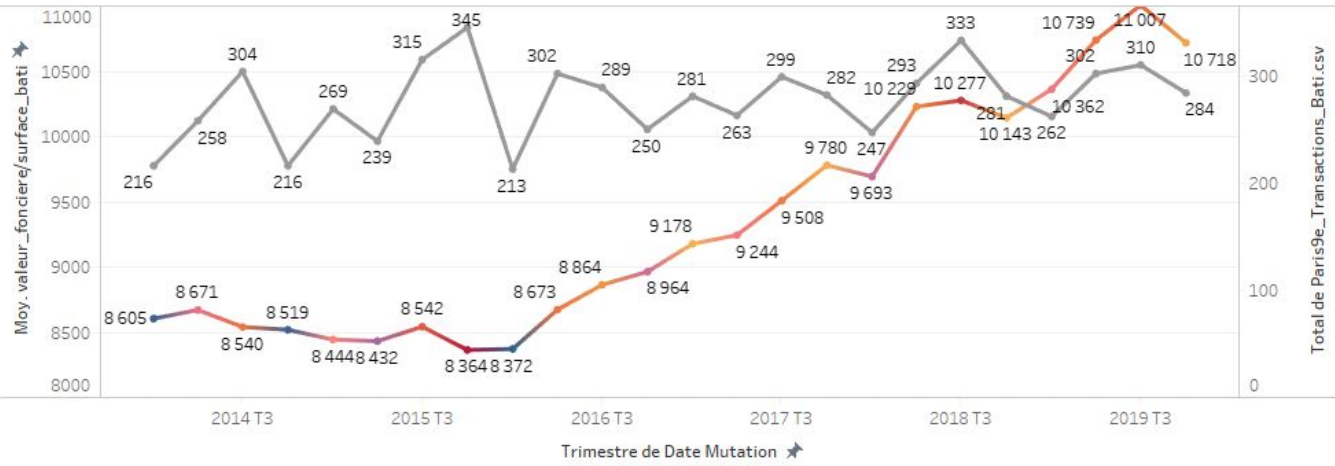


Prices & Transactions

Price Evolution/Year

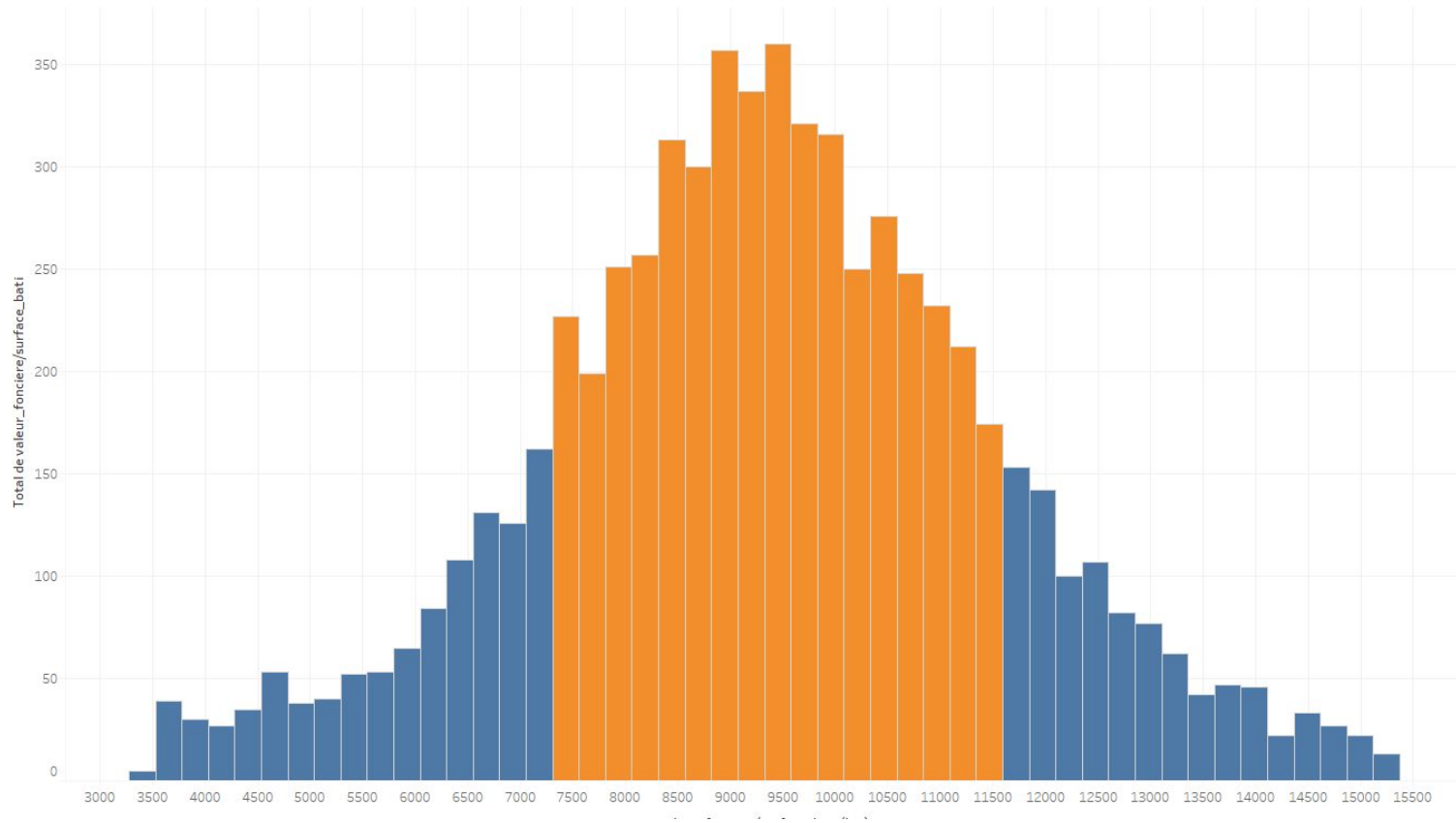


Price Evolution/T



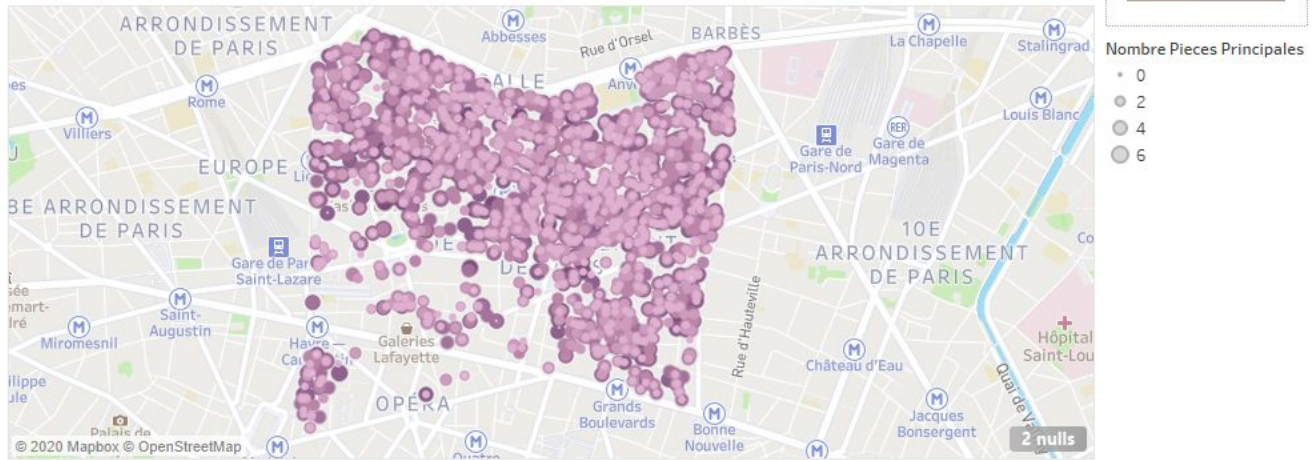
Normal Distribution

Price Distribution

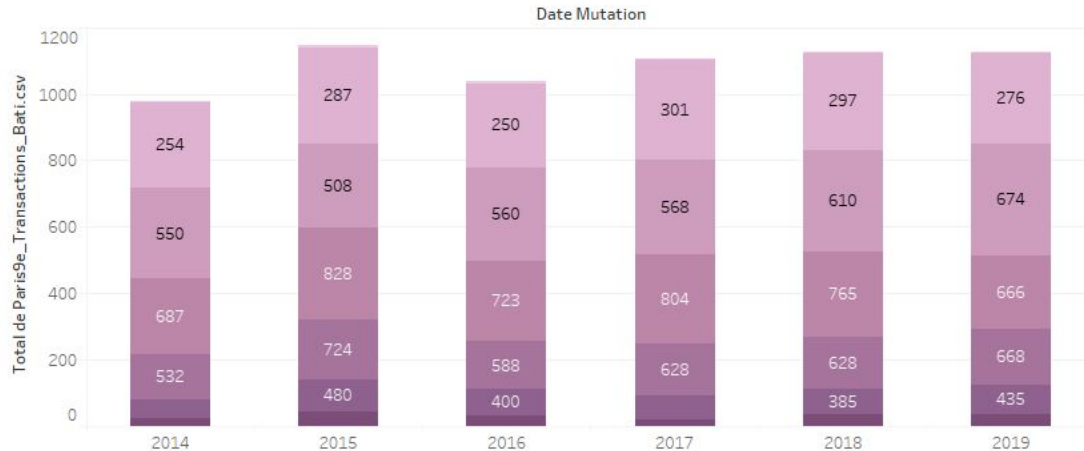


Flat Typology Breakdown

Map Typologies



Vente Typologies-V2

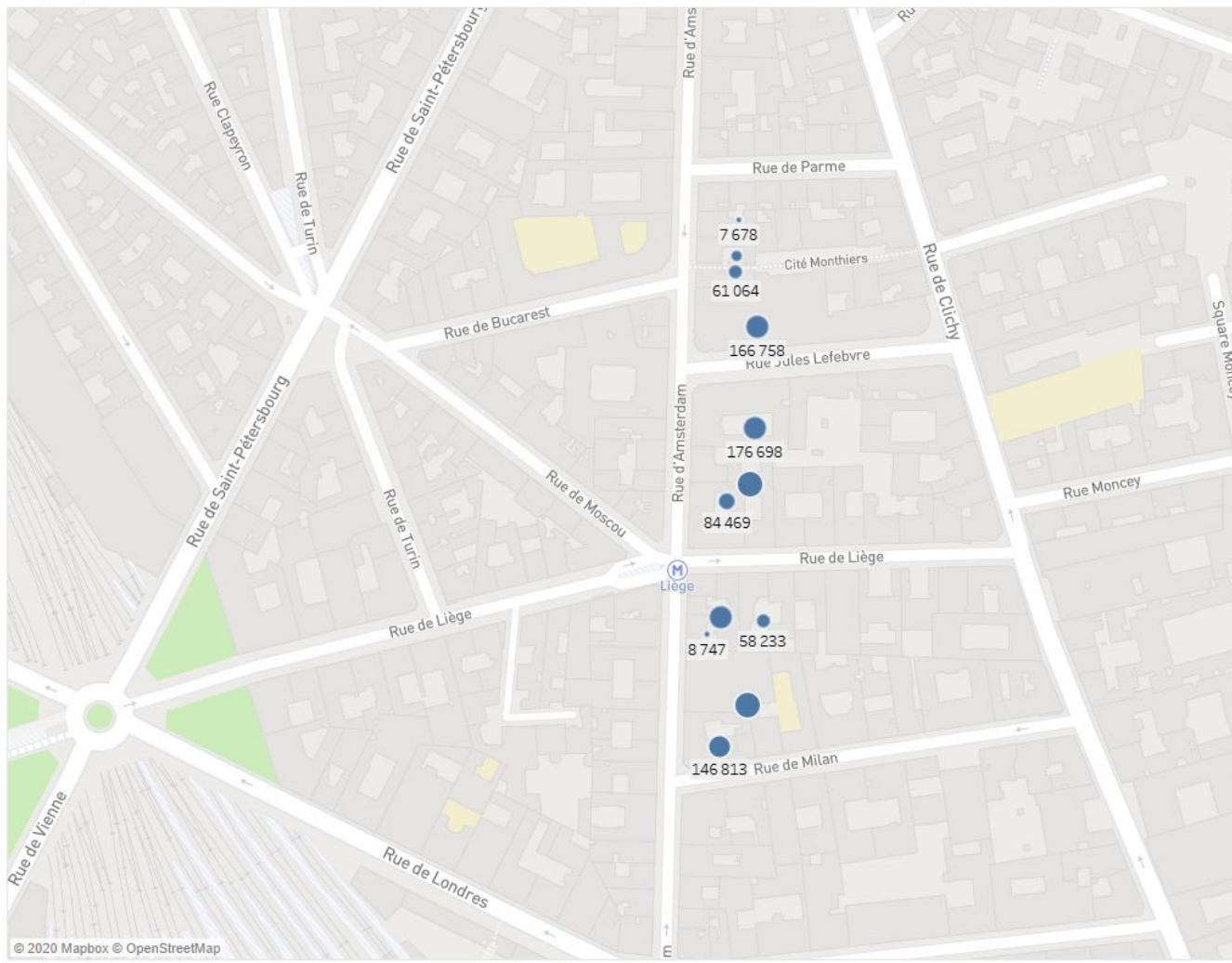


Expropriation

No Surface data provided
We cannot compare
if the 'Valeur Foncière' defined was :

- Under Market
- At Market
- Above Market

Expropriation



Database VS Web Scraping

In order to compare our 6 months / 1 year old database

We Web Scraped data for the same area from the Website SeLogger.com

se 484 annonces de vente et viager +

seloger.com/list.htm?m=initial&enterprise=0&idtypelib=1&idtt=2.5&natureben=1.24&c=750109&m=search_hp_new

seLoger + Déposer une annonce

Recommandations Mes favoris Mon compte Menu

Vous êtes investisseur ?
Achetez avec la loi Pinel et économisez
jusqu'à **63 000 € d'impôt.**

Je réalise ma simulation gratuitement

Achat Paris 9ème Appartement

484 annonces de vente et viager d'appartements à Paris 9ème (75)
Accueil > Annonces > France > Île-de-France > Paris > Paris 9ème

Liste (484) Carte (333)

Trier par Sélection

Appartement
2 p • 1 ch • 55,21 m²
Paris 9ème
Lorette-Martyrs

Appartement 2 pièces de 55m² | Rue Lamartine | Paris 09 Libereys vous propose cet appartement T2 de 55m² (33m² ill Carrez),

Exclusivité

Libre keys

À voir sur Belles Demeures →

575 000 €
ou 2 378 €/mois**

Assurance de prêt ? Changez pour Cardif Libertés Emprunteur

Activer l'alerte

Économiser sur son assurance de prêt

DevTools failed to load SourceMap: Could not load content for https://static-seloger.com/commons/ajax/libs/styled-components/4.3.2/styled-components.min.js.map: HTTP error: status code 403, net::ERR_HTTP_RESPONSE_CODE_FAILURE

```
1 import pandas as pd
2 import numpy as np
3 import requests as r
4 from bs4 import BeautifulSoup
5 import re
```

```
1 headers="""accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,*/*;q=0.8,app
2 accept-encoding: gzip, deflate, br
3 accept-language: fr-FR,fr;q=0.9,en-US;q=0.8,en;q=0.7
4 cookie: __uzma=l2efda3d-e52b-fed2-659e-3921e3abf325; __uzmb=1601042769; __uzmc=510521944541; __uzmd=1601042918; visitId
5 referer: https://www.seloger.com/
6 sec-fetch-dest: document
7 sec-fetch-mode: navigate
8 sec-fetch-site: same-origin
9 sec-fetch-user: ?1
10 upgrade-insecure-requests: 1
11 user-agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/85.0.4183.121 Safari
```

```
1 headers=dict(i.split(': ') for i in headers.split('\n'))
2 url='https://www.seloger.com/list.htm?tri=initial&enterprise=0&idtypebien=1&idtt=2,5&naturebien=1,2,4&ci=750109&m=search
3 html=r.get(url,headers=headers).content
4 soup=BeautifulSoup(html)
```

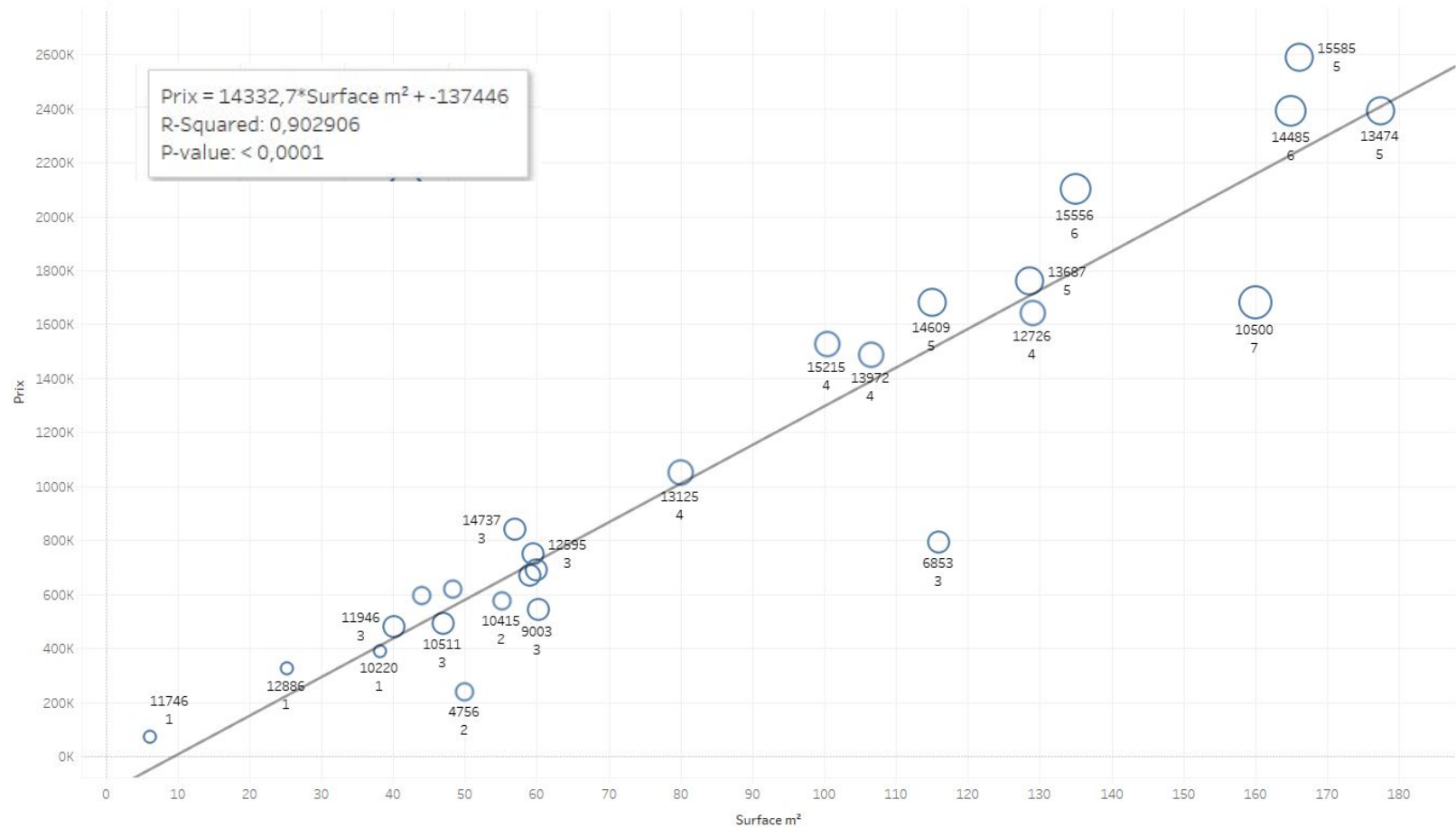
```
1 rooms=[int(i.strip()[0]) for i in re.findall(r'\d [p]', str(soup.select('ul.ContentZone__Tags-wghbmy-8')))]
2 surface_m2=[float(i.replace(',','.')) for i in re.findall(r'li>{[0-9,]+} m²<\/li>', str(soup.select('ul.ContentZone_
3 links=[i.get('href') for i in soup.select('.Card__ContentZone-sc-7insep-3 a.CoveringLink-a3s3kt-0')]
4 price=[int((''.join(i.text.split()[:-1])).strip('bouquet')) for i in soup.select('div.Price__Label-sc-1g9fitq-1')]
```

```
1 df=pd.DataFrame(rooms)
2 df.rename(columns={0: 'Nombre de pièces'}, inplace=True)
3 df['Surface m²'] = surface_m2
4 df['Prix'] = price
5 df['Lien'] = links
```

```
1 #export
2 df.to_csv('Paris9e_SeLogger.csv', sep='/', index=False)
```


27 Offers 28/09/2020

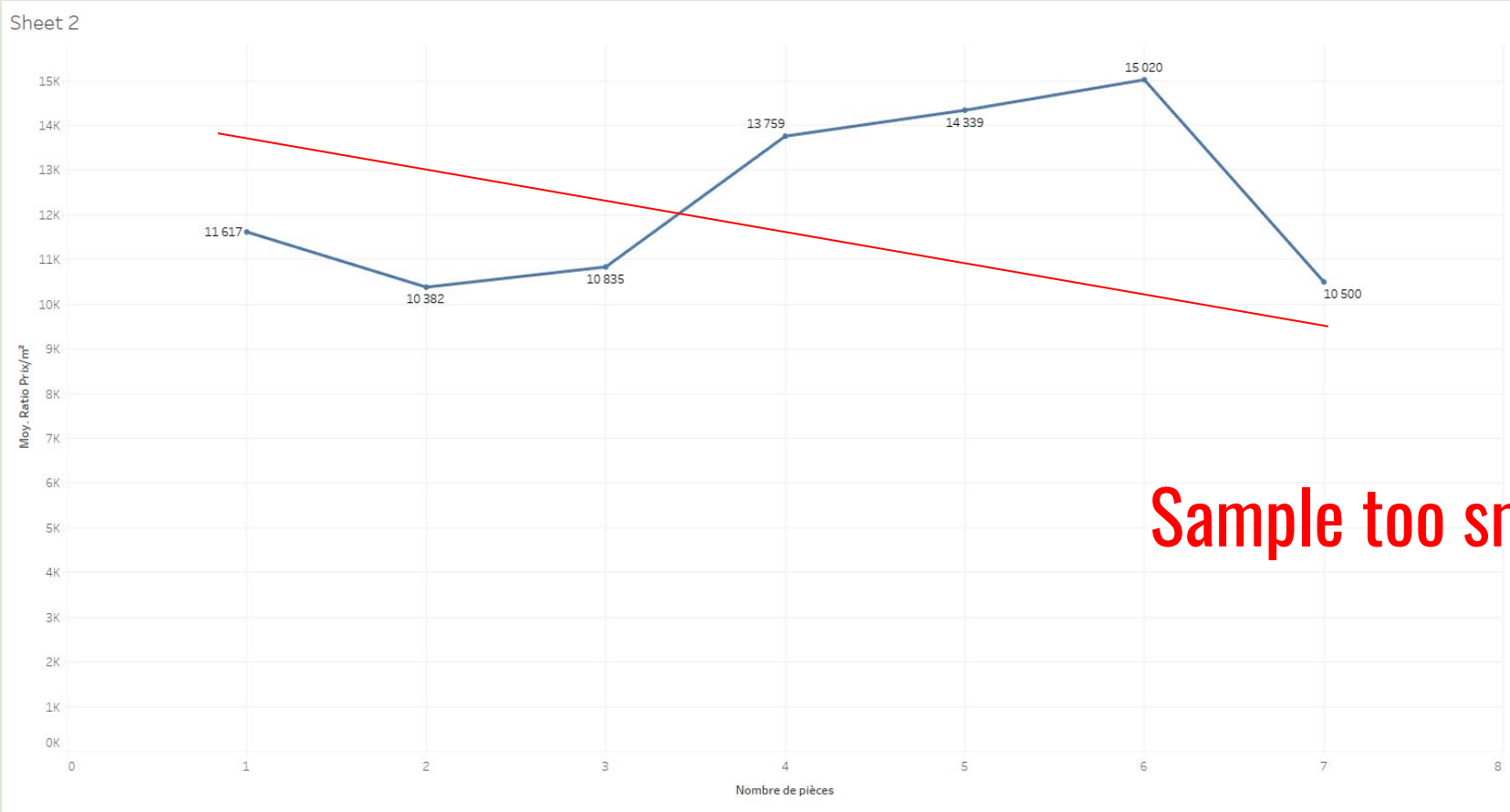
Price Analysis



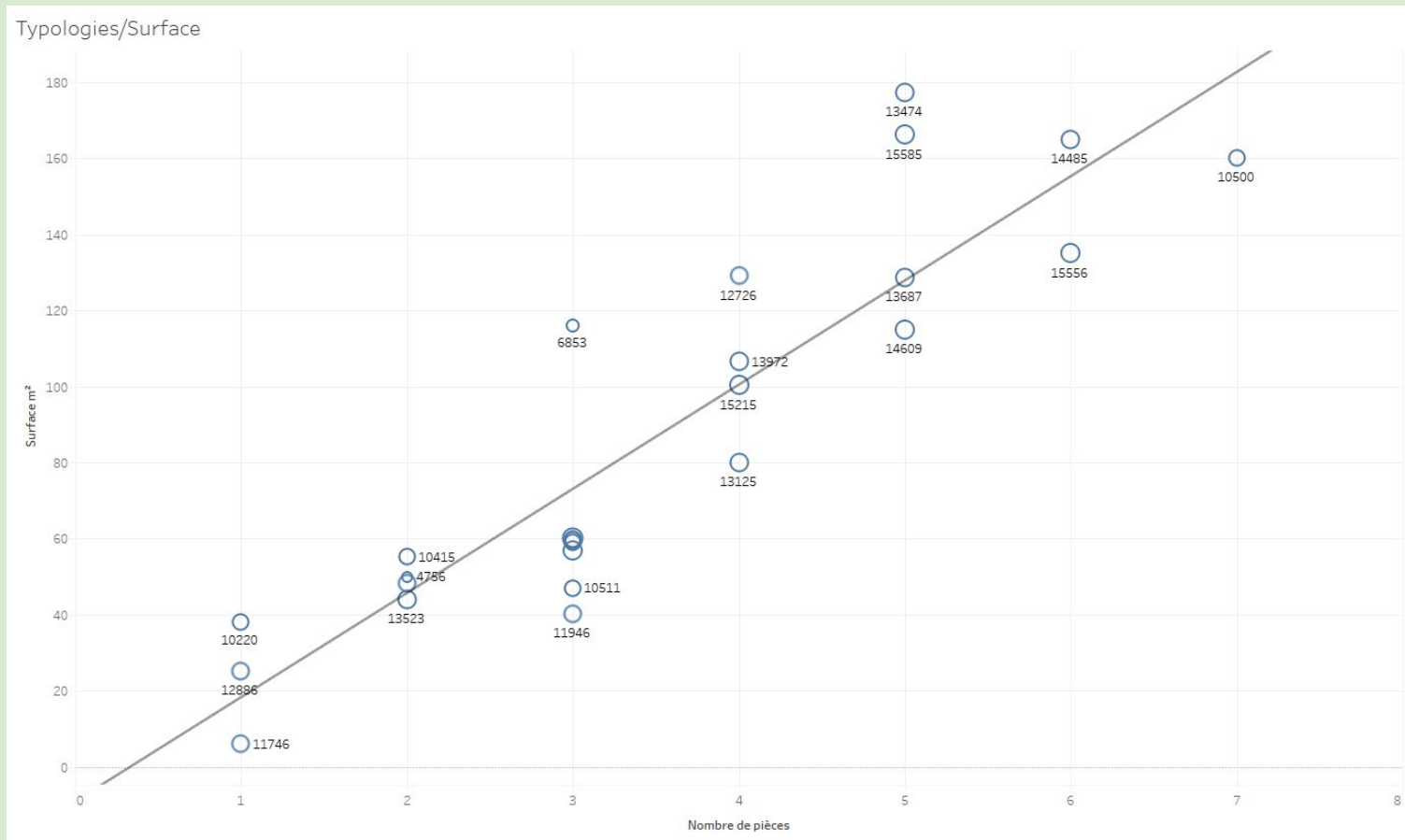
Nombre de pièces



Sample Test - Average Price per Typology



Typology VS Surface



Learnings

- Tableau viz
- Correlation in python
- Hypothesis testing in python
- Data cleaning skills
- Export data to use it in Tableau
- Web scraping skills

Potential Improvements

- Compare more areas in the same city
- Compare with other cities
- Establish a biggest database for this year 2020 and take in account the data coming from website have to be adjusted to get figures closer to transaction price.
- It will be interesting to incorporate to the study others factors (Green spaces / Street Type / Road traffic ...) Data we could get from the city website.
- Realise the study for other transaction types (Retail, Factories, Carparks ...)