Lifestyle Analysis of Paris

A.Boittin

11/05/2020

Agenda

Problem Statement

Project Overview

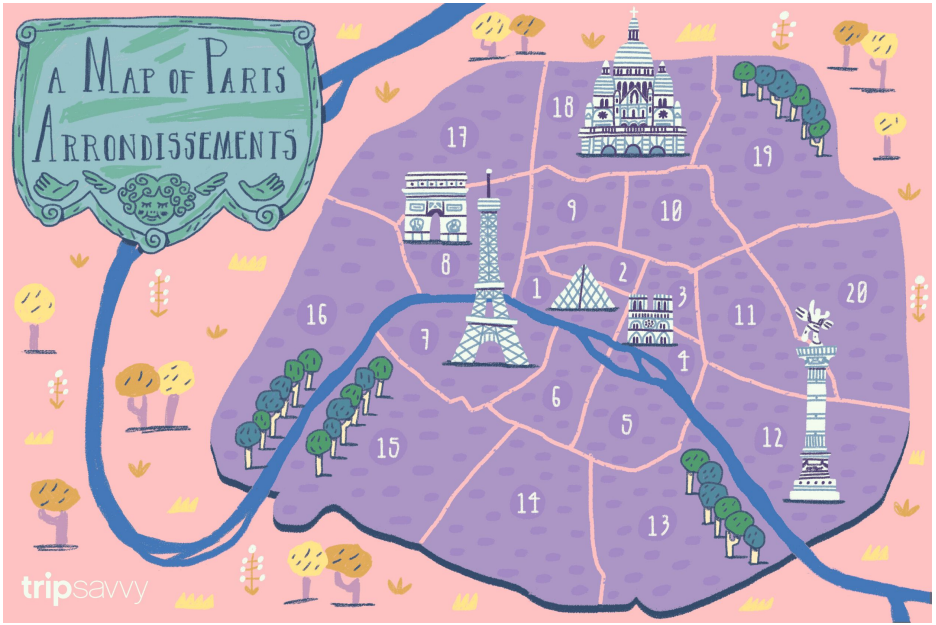
Problem Solution

Benefits

Way Forward

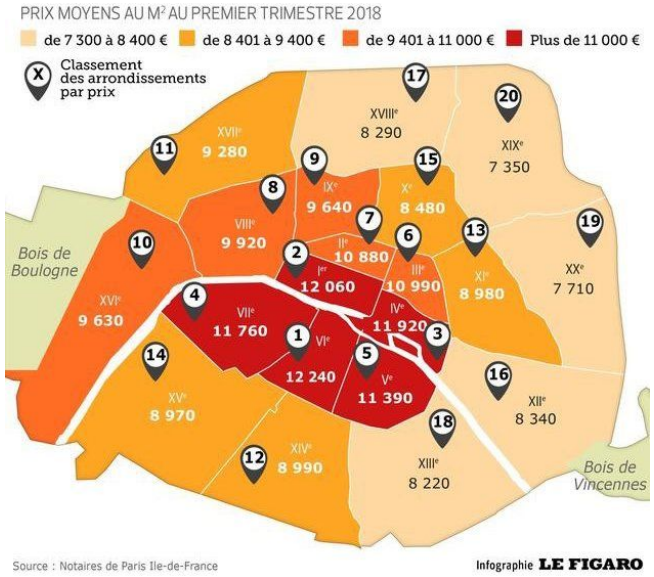
Introduction

Paris is divided in 20 arrondissements.



Problem Statement

- What does explain the cost differences between arrondissements?

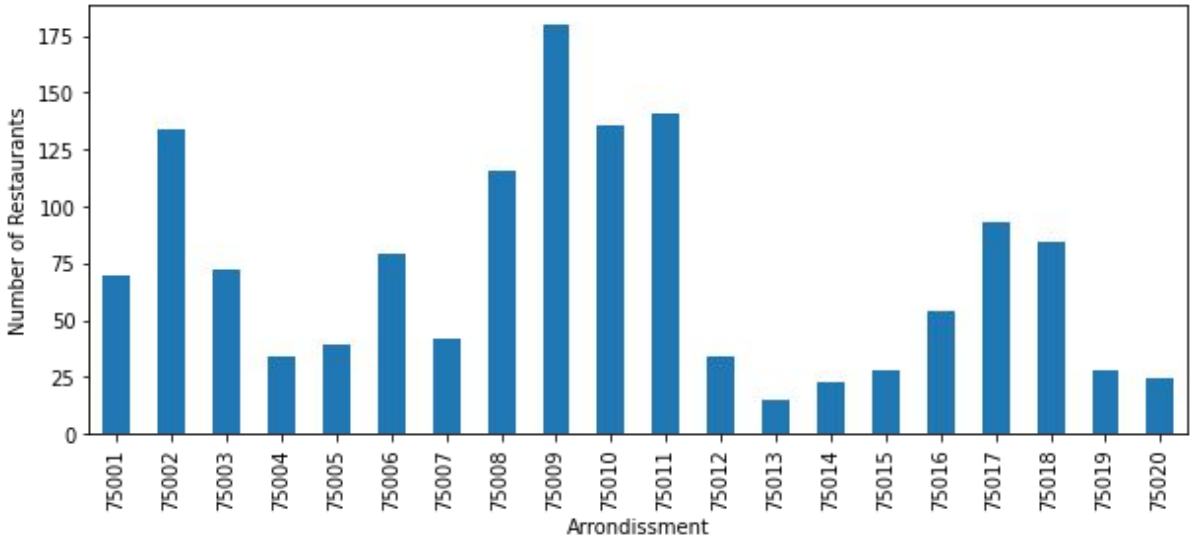


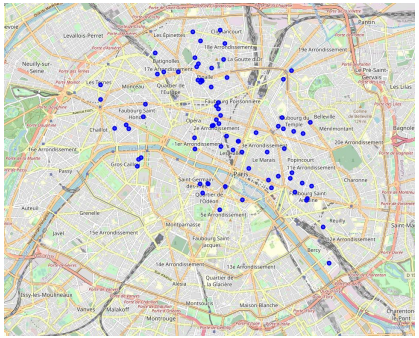
Data Description

Source	Description	Type
Newtable	Last opened restaurants	Website + Personal Notebook
Paris Data	Arrondissements + Administratives Neighborhoods coordinates	API + API
	Houses rental cap 2019	API
	Events and activities	API
	Parks and green spaces	API
	Opened Markets	API
RATP	Global transportation offer (schedules, localisations, lines)	API
Inside Airbnb	Available hosts list for temporary rents	API
Overpass	OpenStreetMap API	API
Etalab	Past real estate buying prices (last 5 years)	Source Code
INSEE	Population	Datasets

Features Selections

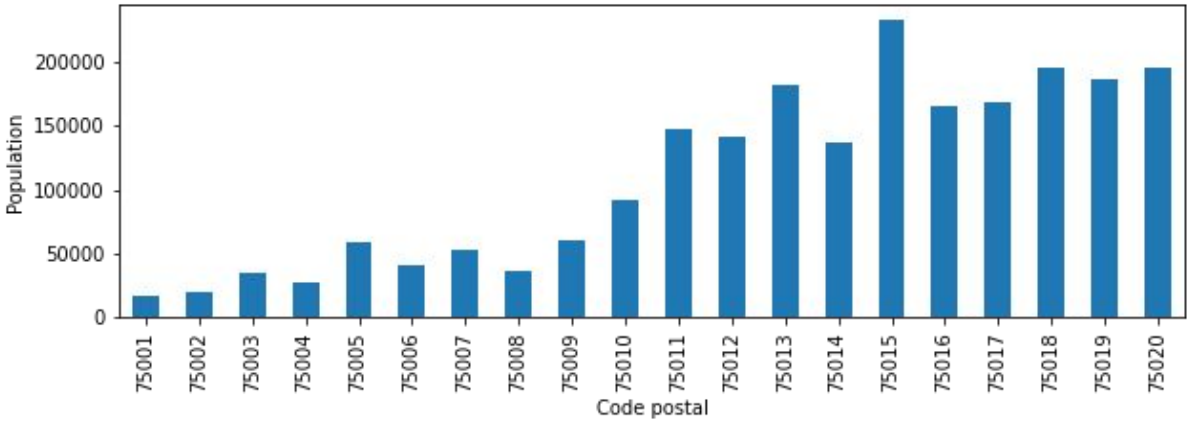
Restaurants





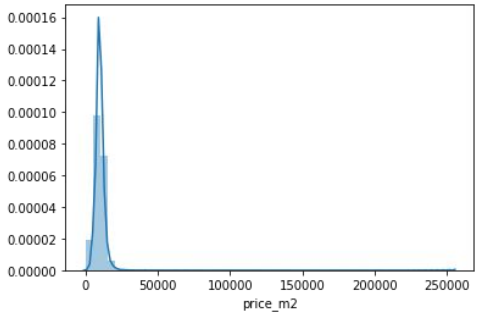
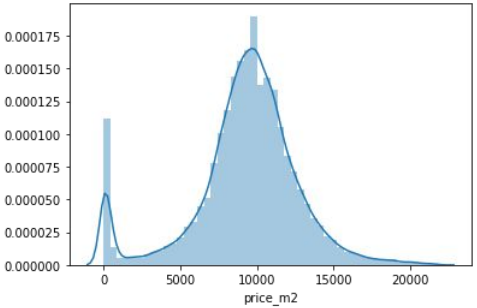
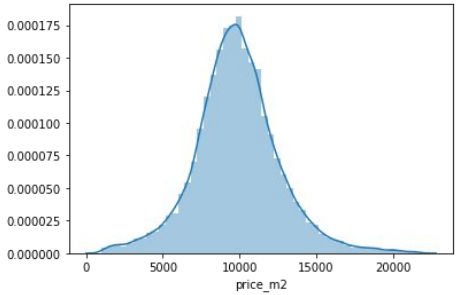
Features Selections

Population



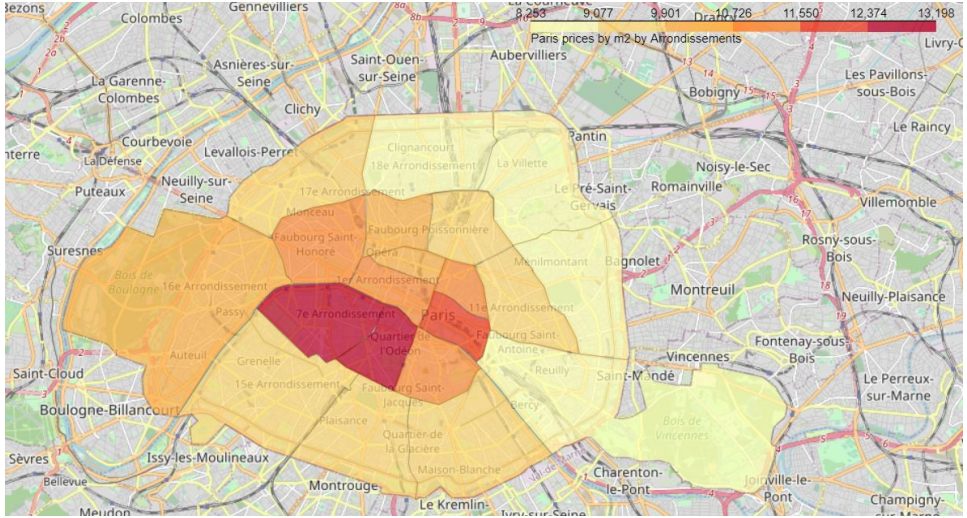
Target

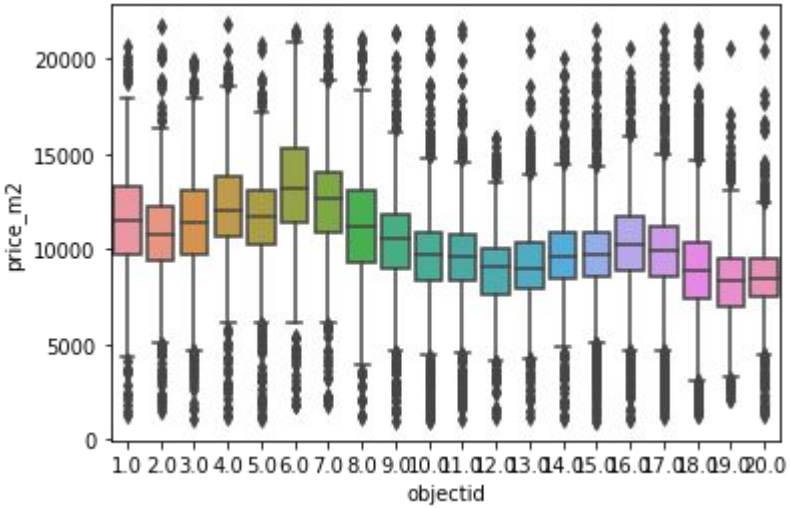
Price

First extraction	Cleaning with Z-Score	Distribution peak Cleaning
		

Target

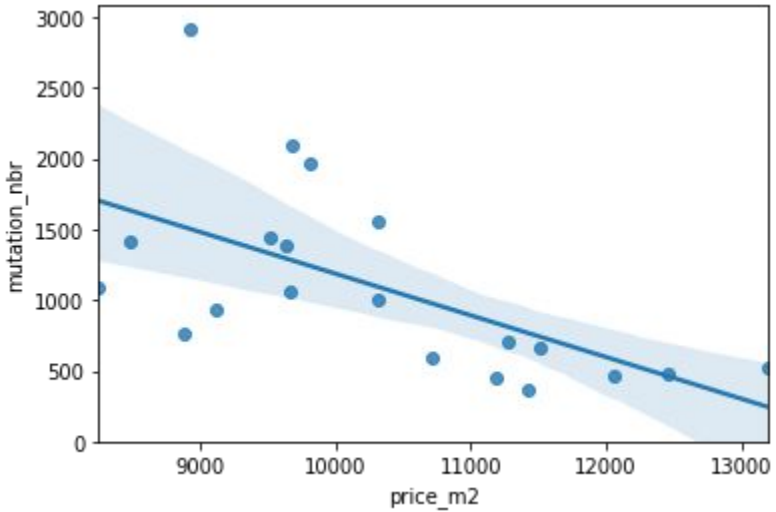
Price





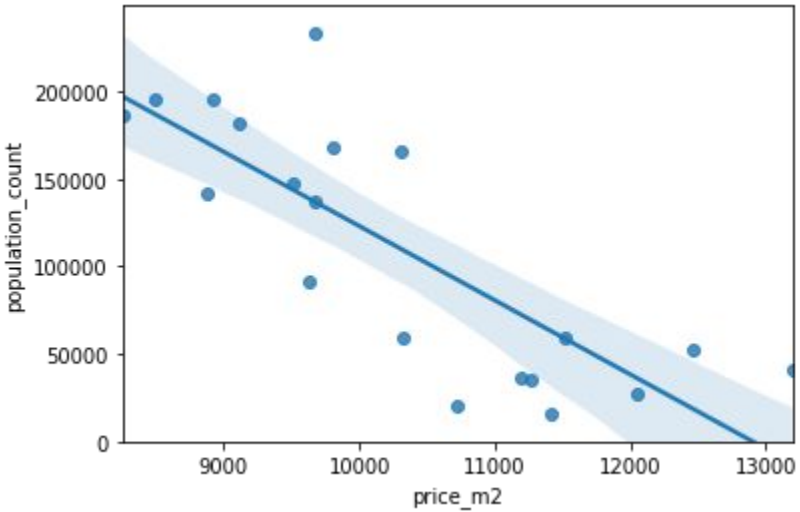
Methodology

Correlation: Number of mutation vs Price



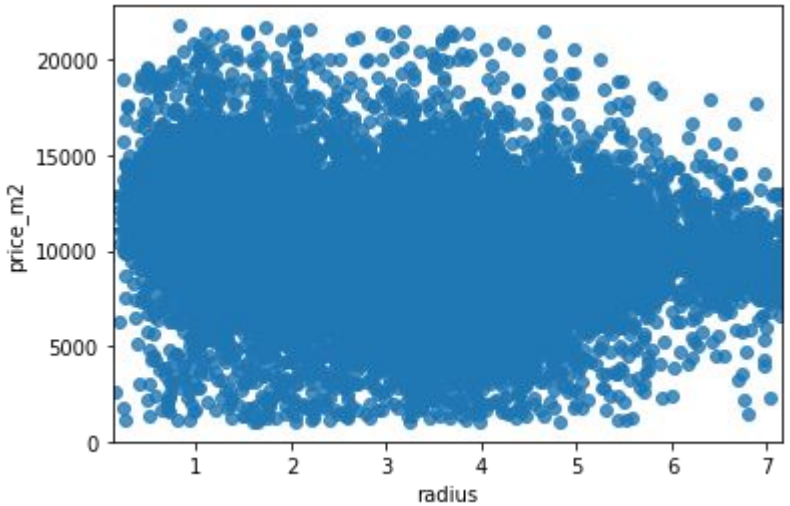
Methodology

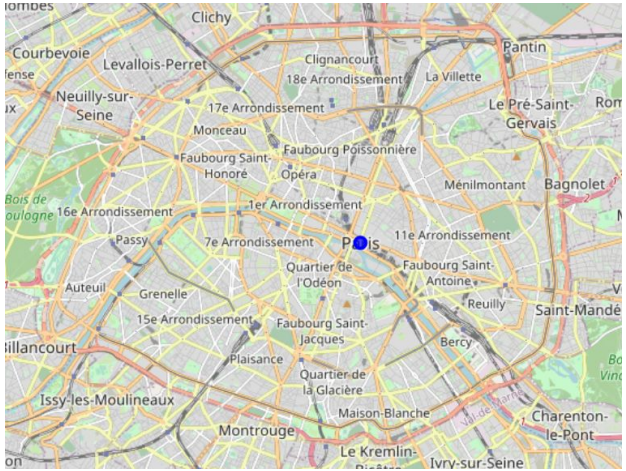
Correlation: Population vs Price



Methodology

Correlation: Radius from Paris center vs Price

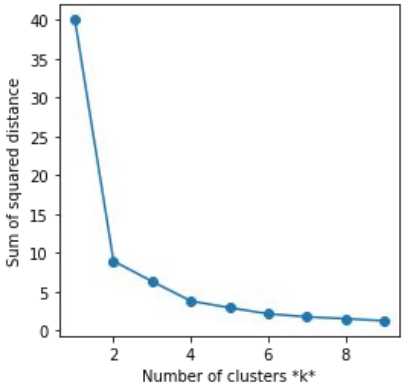


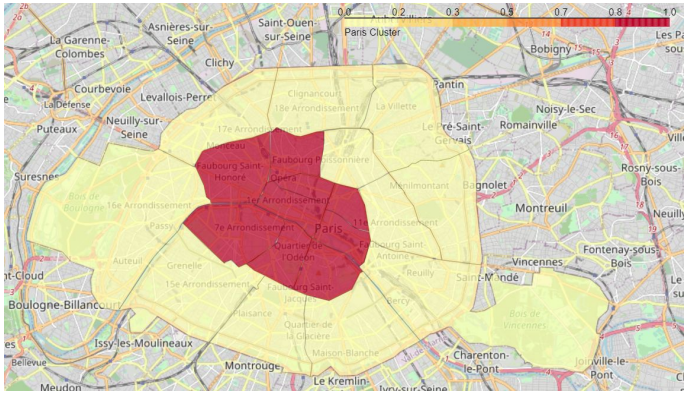


Methodology

Clusterization using K-Means

- Center: rich (11572 €/m2) and less populated (38,600
- Periphery: cheaper (9296 €/m2) and more populated (167,539).





Results

- Hard to find real correlations between chosen features and price
- More data is needed to refine the model

Conclusion

- Good opportunity to apply python and data science skills
- Lot of data available on French administration websites