# Introduction

## Problem Goal

Nowadays, when we talk about real estate business, it seems clear that is at the center of a lot of expectations:

* For project owners:
  + The choice of a neighborhood to implement a business will be critical for its future success (competition, potential customers, attractivity of the neighborhood…) ;
  + When it comes to invest in a house, the decision of the neighborhood must be thoroughly deliberate as it will impact people quality of life (activities in the neighborhoods, school, shops, parks, distance to their work…) .
* For investors, this choice will have a direct impact on their investments and the profitability.

In other terms, real estate has a direct financial impact combined with a strong social impact on population lives.

Our goal here is to provide insight on the different neighborhoods that will help decider to choose the correct place to invest.

Choosing a correct business emplacement will depend on several aspects: the type of business, the target, the population density, the competition, price per square meter of the local …).

Choosing a correct place to live will depend also on several aspects : the age of the buyer, the family structure (single, couple, kids…) , their hobbies, the place of their work, commodities, transport services, price per square meter, type of housing facilities ….

Finally, investors will be mainly interested in the capacity of the borrowers to pay of their loan, but they will also be interested in the potential price trends of the neighborhood in order to secure their investment.

Our goal here, will be to aggregate neighborhoods with information’s about :

* Real estate price
* Most common type of real estate properties (apartments, houses…)
* Principal venues of the neighborhoods

real estate in France is becoming co

When it comes to find the best place to start a business or to buy a house/apartment,

Our objective is to help a family that must move in a city that they do not know. Based on several dataset, we will propose them some neighborhoods in which they should start looking for a house.

Projection prix immonbilier en fonction du quartier par m2

Quelqu’un cherche à s’installer en ville et connaitre les prix immobiliers :

Données nécessaires :

* FOURSQUARE POUR INFORMATION SUR QUARTIERS (commerces, écoles, bars, parcs, sports, creches…)
* Obtenir les quartiers de bordeaux : <https://opendata.bordeaux-metropole.fr/explore/?disjunctive.publisher&disjunctive.frequence&disjunctive.territoire&sort=title>
* Data tourisme pour les activités familiales (facultatif) <https://diffuseur.datatourisme.gouv.fr/fr/flux>
* Qualité de l'air  pour le degré de pollution : csv téléchargeable <https://www.data.gouv.fr/fr/datasets/jdd-federe-qualite-de-lair-france-donnees-a-date/>
* Arbres : <https://www.data.gouv.fr/fr/datasets/arbres-en-open-data-en-france-par-nam-r/>
* Impôts locaux : <https://www.data.gouv.fr/fr/datasets/les-impots-locaux-en-france-2004-a-2013/>
* Nombres ventes dans le quartier sur les dernières années (est ce que beaucoup de gens partent ??) (pas la, pc, va mourir sinon)
* Prix habitation - base de données dvf soit en téléchargement soit api (dl c’est sans doute mieux)== > <https://www.data.gouv.fr/fr/datasets/demandes-de-valeurs-foncieres/>
  + <http://data.cquest.org/dgfip_dvf/LISEZ_MOI.txt>
* Temps de trajet travail / école/

Type logements :

Propose a neighborhood Utiliser la clusterisation pour faire une recommandation de quartier en fonction des centres d’intérêts à destination des personnes qui cherchent à emménager dans la ville

Proposer des types de business qui semblent manquants dans les quartiers en fonction des venues que nous avons retiré

Conseiller des quartiers pour implémentation de Business

1. Qu’est ce qui fait la qualité de vie ?
2. Trouver les informations par ville :
   1. Lieux de sorties : restaurants, café, bars tous les combiens de mètre
   2. Des services (banques, journaux, postes)
   3. Transports : métro, bus stops
   4. Des commerces (alimentaires, habits….)
   5. Des soins (médecins, dentistes, ophtalmologistes …)
   6. Culture : théâtre, cinémas, salles de concerts, stades
   7. Sports et nature : parcs, terrain/salles de sport

Vérifier si je trouve ces informations dans les venues de foursquare

Ou idée 2 qui emploiera de la clusturisation ou de la labelisation :

Trouver ou ouvrir son commerce de la manière la plus optimisé

Dire où est le meilleur emplacement pour ouvrir un business

Comparer qualité de vie entre les villes :

*Clearly define a problem or an idea of your choice, where you would need to leverage the Foursquare location data to solve or execute. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem.*

*This submission will eventually become your****Introduction/Business Problem****section in your final report. So I recommend that you push the report (having your Introduction/Business Problem section only for now) to your Github repository and submit a link to it.*

## Objectives – How tackle the problem

## Analytic approach

This means identifying what type of patterns will be needed to address the question most effectively.  
If the question is to **determine probabilities** of an action, then a **predictive model** might be used.  
If the question is to show **relationships**, a **descriptive approach** maybe be required. This would be one that would look at **clusters** of similar activities based on events and preferences.  
**Statistical analysis** applies to problems that require **counts**. For example if the question requires a yes/ no answer, then a classification approach to predicting a response would be suitable.  
**Machine Learning** is a field of study that gives computers the ability to learn without being explicitly programmed. Machine Learning can be used to **identify relationships and trends** in data that might otherwise not be accessible or identified.  
In the case where the question is to **learn about human behaviour**, then an appropriate response would be to use **Clustering Association** approaches.

# Data

## Data requirements & collection :

Think of this section of the data science methodology as cooking with data. Each step is critical in making the meal. So, -**if** the problem that needs to be resolved is the **recipe**, so to speak, and data is an ingredient, then the data scientist **needs to identify: which ingredients are required, how to source or the collect them**, how to understand or work with them, and how to prepare the data to meet the desired outcome.

**Prior** to undertaking the data collection and data preparation stages of the methodology, it's vital to **define the data requirements** for decision-tree classification. This includes **identifying the necessary data content, formats and sources for initial data collection**

*Describe the data that you will be using to solve the problem or execute your idea. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea. You can absolutely use other datasets in combination with the Foursquare location data. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using, even if it is only Foursquare location data.*

*This submission will eventually become your****Data****section in your final report. So I recommend that*

# Methodology:

## Data understanding

Which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

## Data preparation

## Modeling

# Results

## Evaluate

where you discuss the results.

# Discussion

where you discuss any observations, you noted and any recommendations you can make based on the results.

# Conclusion

where you conclude the report.

# Objectif

# Review criteria

For this week, you will be required to submit the following:

1. A description of the problem and a discussion of the background. (**15 marks**)
2. A description of the data and how it will be used to solve the problem. (**15 marks)**

For the second week, the final deliverables of the project will be:

1. A link to your Notebook on your Github repository, showing your code. (**15 marks**)
2. A full report consisting of all of the following components (**15 marks**):
3. 3. Your choice of a presentation or blogpost. (**10 marks**)