

# Présentation du Funathon

## Éléments généraux

Aurélien Fortin, Lucas Malherbe et François Sémécurbe (SSP  
Lab)

25 mai 2021

## Une présentation en deux parties

- Présentation générale du funathon
- Quelques premiers exemples (non limitants !) de ce que vous pourriez faire

## Qu'est ce qu'un «funathon» ?

Funathon = évènement de formation, collaboratif, offrant aux participants l'occasion de s'entraîner sur des techniques de data science. C'est donc :

- ① Un évènement de formation
- ② Un moment de travail collaboratif, fait pour aider à progresser
- ③ Un moment sans enjeu autre que personnel

# Qu'est ce que n'est pas un funathon ?

Un hackathon. Car :

- ❶ Pas de gagnant, pas de compétition
- ❷ Pas d'enjeu pour un sponsor

## Pourquoi organiser cet évènement ?

Trois constats :

- ❶ Les techniques de datascience peuvent sembler obscures, alors qu'il est tout à fait possible d'apprendre à commencer à les employer
- ❷ L'offre de formation actuelle permet d'apprendre, parfois de se perfectionner, mais est trop ponctuelle ce qui empêche la bonne assimilation (problématique de formation continue)
- ❸ Même quand on sait déjà faire, la pratique est essentielle pour l'entretien

## C'est fait pour qui ?

- Tout agent du SSP, Insee ou non, qui souhaite se former aux techniques de datascience, qu'il ait déjà entamé cette démarche ou non
- Prérequis : Être déjà familier avec un langage de programmation adapté (R ou Python), car pas possible de profiter du funathon pour une formation partant de zéro. A minima, être capable de suivre et/ou reproduire un notebook ou des posts de blogs.
- Sur les packages de datascience, pas forcément de prérequis (possible d'apprendre à les utiliser pendant le funathon)

## Que puis-je en attendre ?

- Attente réaliste : pouvoir, pendant (jusqu'à) 2 jours, m'entraîner à l'utilisation de techniques innovantes avec un encadrement en appui. Progresser ainsi sur le sujet, ou ne pas perdre d'éventuels acquis existants.
- Attente irréaliste 1 : devenir un expert sur le sujet
- Attente irréaliste 2 : devenir autonome sur le sujet si c'est votre première confrontation.
- Pour ces deux attentes "irréalistes", voir l'offre de formation complémentaire

## Les principales craintes potentielles

- ❶ Je n'ai pas le niveau. Si vous n'avez jamais fait de R ou Python, plus prudent en effet de partir sur une formation d'initiation au langage. Sinon, l'objectif du funathon est justement d'aider à progresser !
- ❷ Je n'ai pas le temps d'assister aux deux jours. Les deux jours sont une possibilité, pas une obligation. Il n'y a pas d'attente sur le résultat final.

Seul point d'attention : les attentes doivent être en cohérence avec votre point de départ.



## Si on veut se lancer, cela se passe comment ?

Avant l'évènement :

- 1 Envisager de constituer une équipe : pas pour être compétitif, mais parce qu'on fait souvent plus et mieux à plusieurs (et que c'est plus sympa)
- 2 Réfléchir à ce que vous avez envie de faire (du machine learning ? de l'analyse géographique ? de l'analyse de texte ? autre chose ?) et nous en parler pour qu'on essaie de mieux préparer. Utile également de nous indiquer votre niveau de départ.
- 3 Prévoir d'assister aux séances de présentation du datalab et des données :)

## A quoi cela va-t-il ressembler ?

- Le SSP Lab fournit des données au potentiel d'exploitation assez large en termes de datascience (machine learning, analyse de texte, dataviz,...), en l'occurrence des données issues d'Airbnb
- Fourniture également d'exemples de code, ainsi que possiblement des exemples de mise en pratique sur d'autres types de données (type post de blog) pour vous aider à vous lancer dans le sujet
- Pendant les deux jours, travail de chaque équipe en autonomie, mais avec échanges très réguliers avec nous pour que nous puissions aider à avancer, et débloquer si besoin
- A la fin du 2e jour, séance d'échange collective. Les groupes qui le souhaitent (et eux seuls, s'il y en a) pourront présenter leurs résultats.

## Techniquement ?

- Utilisation du datalab, plateforme la plus flexible et complète existante
- Utilisation a priori de gitlab pour mettre à jour le code, et faciliter le travail en équipe
- Utilisation de zoom a priori pour les échanges réguliers pendant les deux jours (création d'une sous-salle par équipe en plus d'une salle globale)
- Utilisation également de tchap (un salon par équipe, salon global)

## Les prochaines étapes

- Inscription, individuelle ou par équipe, auprès de votre correspondant formation : pas de limite de place, mais le plus tôt sera le mieux pour que nous puissions préparer
- Présentation du datalab le 7 juin à 10h par la DIIT
- Présentation plus détaillée des données le 15 juin à 14h
- 21 et 22 juin : fête de la data :)

## Le funathon, un élément de l'offre de formation du SSP Lab

Grands principes de la formation tels qu'ils se profilent :

- Mise à disposition de notebook régulièrement, plutôt d'initiation sur différentes techniques pour autoformation
- Présentation trimestrielle de ces notebooks, et disponibilité pour échange
- Pour un vrai perfectionnement, recherche avec la division formation d'une offre plus complète, sur le modèle de la formation "initiation à Python" réalisée avec Datascientest
- Mise en place (régulière ?) de Funathon pour permettre la mise en pratique des connaissances

## Les données mises à disposition

- Données présentant l'ensemble des locations réalisées via Airbnb dans quelques grandes villes de France, avec un certain nombre d'informations associées (localisation du bien, description détaillée, prix, commentaires du client)

## Pourquoi ces données ?

- Thématique de l'immobilier a priori intéressante - même à titre personnel - pour tout le monde
- Possibilité de croiser ces données avec d'autres données utiles (données des notaires,...)
- Permet de mettre en place différentes techniques de datascience : machine learning, analyse de texte, dataviz, et également de travailler sur des données géographiques fines.

## 'Le droit à la ville' est-il menacé par Airbnb ?

Les plateformes numériques (Airbnb, Uber) transforment les villes et inquiètent les acteurs publics qui évoluent sur une ligne de crête entre favoriser l'attractivité de leur territoire et garantir le bien-être de leurs citoyens.

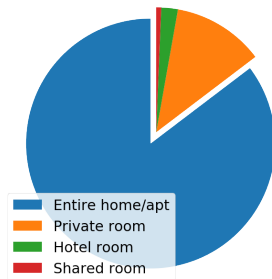
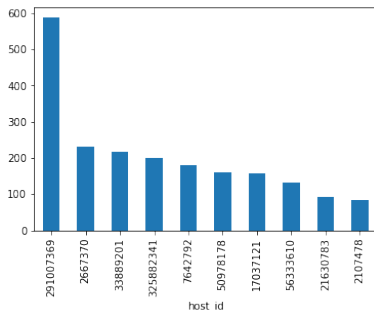
- Le site Inside Airbnb est une démarche citoyenne qui vise à ouvrir et rendre accessible les informations disponibles sur le site Airbnb (Webscrapping pour Paris, Lyon et Bordeaux).
- Les données ont été analysées, nettoyées et agrégées pour faciliter le débat public.

Les économistes, agences d'urbanistes, les sociologues utilisent ses données pour mesurer les conflits d'usages entre les fonctions résidentielles et touristiques, analyser la discrimination... et plus généralement pour décrire les transformations en cours.



# Connaître l'offre, un peu de stat desc

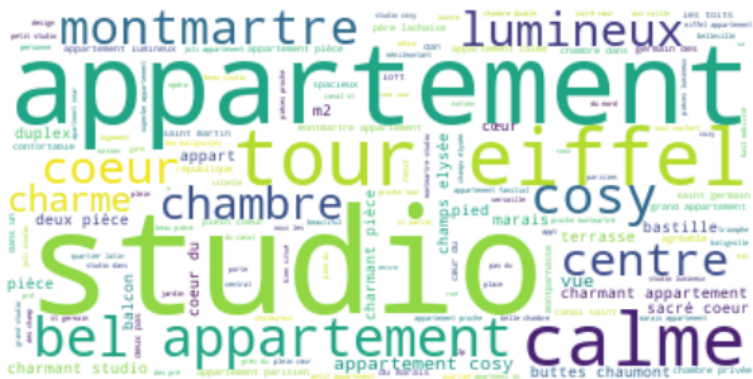
Airbnb, c'est 63 000 logements rien que sur Paris.



Le principal hôte à plus de 600 logement, c'est la société de conciergerie Luckey. L'étudiant qui partage son canapé pour rencontrer l'altérité est un hôte marginal...

## Déchiffrer le marketing territorial des loueurs (Traitement automatique du langage)

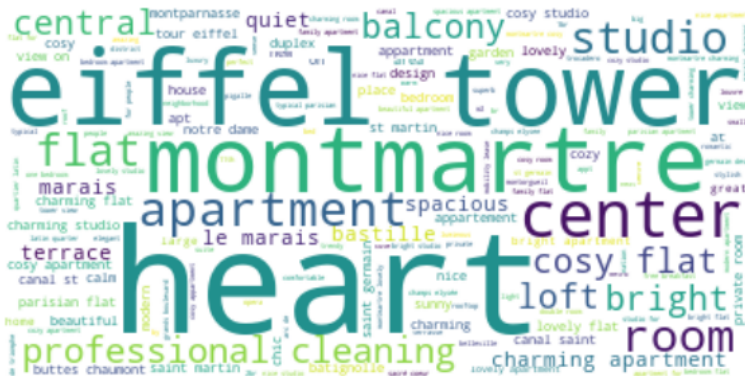
## Nuage de mots des titres des logements parisiens



## Pour aller plus loin : comment caractériser un quartier ?

## Deciphering the Place Branding (NLP)

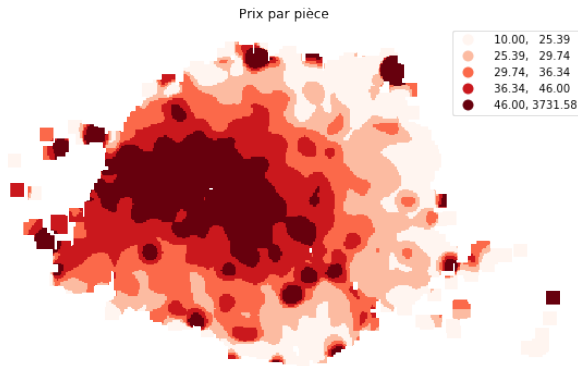
The previous frame in English ? Not really, we can try to detect the language of the name of the accommodations and make an English Wordcloud.



To go further : Describe the origin of the tourists ?

## Prix des logements

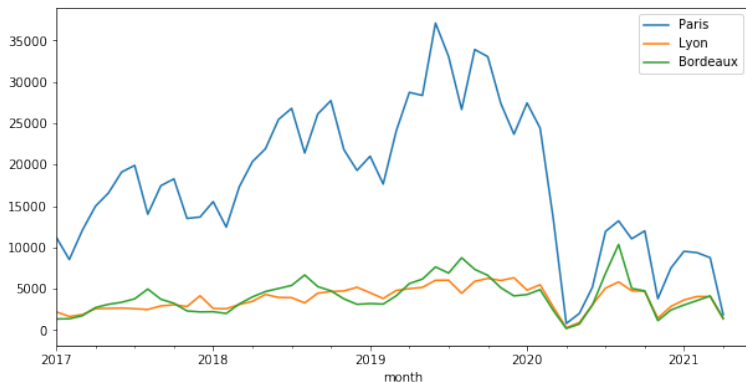
Prix des logements rapportées au nombre de pièces.



Pour aller plus loin : calculer un modèle de prix hédonique ?

## Suivi de l'activité touristique

### Commentaires mensuels laissés par les Guests



Pour aller plus loin : Quels impacts sur le secteur de l'immobilier ?

## Synthèse sur les objectifs du funathon

- Manipuler des données réalistes au lieu de classer les iris selon leur couleur alors que l'on n'est pas capable de distinguer une jonquille d'une rose !
- Expérimenter de nouvelles méthodes de Machine Learning ou de visualisations et se rendre compte que ça ne casse pas trois pattes à un canard !
- Approfondir sa connaissance de R ou de Python ;
- Regarder un code Python et se dire c'est vraiment pas compliqué pour un-e cadoR!
- Travailler avec ses ami-e-s et tester gitlab, cela fait chaud au coeur !
- Découvrir ou approfondir sa connaissance du Datalab, la plateforme de la DIIT dédiée au Machine Learning.

## Ce que vous devez retenir de tout cela

- Le funathon est une opportunité, et ne doit pas vous inquiéter
- Dès lors que vous connaissez un minimum R ou Python, n'hésitez pas à venir participer, même ponctuellement
- Possibilité de creuser tout type de sujet de datascience, mais utile de nous en parler avant
- pour nous joindre : Tchap (Canal du Funathon, ou contact direct) ou par mail (adresse Insee)