Rapport de stage   
HostnFly

Julien PETOT

Maitre de stage : Guilhem Sarcy

Enseignant référant : François Husson





Table des matières

[**Présentation de HostnFly**](#_j3jffuxbiw8h) **3**

[Définition des termes de l’entreprise](#_x8u2ao78xe4v) 3

[Client vs guest](#_4bgwmwp0veb9) 3

[Listing](#_eecwfs7qmob) 3

[Booking vs Réservation](#_poaibzbo311v) 3

[Revenu garanti vs revenu réel](#_v8ntnnhr28kt) 4

[Objectif de HostnFly](#_n1tztryrj47t) 5

[Organisation de l’entreprise.](#_1ayk1bjndgy4) 5

[**Mission 1 : Business Intelligence**](#_y9w792yucxme) **6**

[Objectifs, contexte et outils](#_1hme2tpach00) 6

[Exemple d’une requête](#_qpwm9lfsiiax) 7

[La requête](#_bjy7ezkydpyu) 7

[Execution de la requete en SQL](#_xl7p1r9ysuvc) 7

[Mise en forme du graphique sur Superset.](#_evg1e265t3fm) 8

[**Mission 2 : Projet d’analyse**](#_ildxqvj9lmir) **8**

[Contexte et objectif](#_hykc3dimeeq0) 8

[Exemple D’analyse : LifeTime Value (LTV)](#_y2qp8fxbbpz2) 9

[Contexte et problématique](#_6ycm13ag27no) 9

[Situation globale de nos listings.](#_5sgtrjial51x) 9

[LifeTime Value](#_22bnth46k9gh) 11

[LTV par tranche d’age.](#_xwiwd8vtz20j) 12

[LTV par marché.](#_2lk3vkqz8xmx) 12

[LTV par catégorie de capacité d’accueil.](#_8et0rkxy2fci) 13

[**Bilan sur le stage**](#_1vmpxwpxrhcw) **13**

[Autres missions](#_e5jhkgr48a69) 13

[Attentes sur le stage](#_3pbfiztkjz4t) 14

[Compétences en Data Science](#_jms16wccrf24) 14

[Les points positifs et négatifs](#_aak96m4swvwv) 14

# 

# Présentation de HostnFly

Hostnfly est une start up créée en 2016 a l’initiative de 3 ingénieurs (Quentin CEO, Guillaume COO, Baptiste CTO). Aujourd’hui, HostnFly compte plus de 50 employés et continue de grandir.

**Activité** : HostnFly se définit comme une “**Conciergerie AirBnb**”. Ce terme reste encore très général et peut être interprété de différentes façons. HostnFly s’occupe de louer les appartements de ses clients et propose un service clé en main : prise de photo, création de l’annonce sur les sites de distribution comme Airbnb, Homeaway etc. Elle sélectionne les voyageurs, s’occupe de la remise des clés, et assure un ménage professionnel. Elle gère aussi la variation des prix grâce à des algorithmes établis par l’équipe Data.

Aujourd’hui, Hostnfly s’installe comme un leader du marché français avec plus de 5000 appartements en gestion. Des marchés sont ouverts dans 6 grandes villes de France (Paris, Lyon, Marseille, Lille, Bordeaux et Nice) ainsi que quelques stations balnéaires ou de montagne. D’autres projets sont en cours comme l’ouverture de nouvelles villes et la création d’une nouvelle équipe à Madrid.

## Définition des termes de l’entreprise

Quelques définitions afin de mieux comprendre le but de l’activité HostnFly ainsi que son fonctionnement :

### Client vs Guest

Les **clients** sont ceux qui nous confient leur appartement, ceux à qui on “achète” les nuits.

Les **guests** sont ceux qui séjournent dans nos appartements, ceux à qui on “vend” les nuits.

### Listing

Un listing est un appartement dont on possède des informations comme le nombre de pièces, l’adresse etc. Il peut être enregistré en tant que **client** ou bien **lead** (le contrat a été signé ou le contrat n’est pas encore signé). Les informations de chaque listing sont précieuses pour le pôle data. Elles permettent d’évaluer le prix de vente d’une nuit.

### Booking vs Réservation

Deux concepts liés aux dates de disponibilité d’un listing.

Un **booking** est un intervalle de temps durant lequel un client a mis son listing en disponibilité. Il peut donc y avoir plusieurs bookings pour un même listing si le client décide de confier son appartement à Hostnfly plusieurs fois.

Une **réservation** est un intervalle de temps durant lequel un guest a réservé un listing sur Airbnb: les dates d’une réservation sont donc nécessairement incluses dans les dates d’un booking.

### Revenu garanti vs revenu réel

Deux types de calcul de revenu entre lesquels le client choisit lors de la signature du contrat :

Le **revenu garanti** est le revenu calculé en fonction des caractéristiques de l’appartement (capacité d’accueil, localisation, nombre de salles de bain…) et est versé au client le jour de son départ, que l’appartement soit finalement loué ou non. HostnFly prend le risque à la place du client de ne pas le louer.

Le **revenu réel** est un revenu calculé en fonction du revenu réellement généré par la location de l’appartement et sur lequel HostnFly prend une commission. Si le bien n’est pas loué, le client ne touche donc rien. Le revenu est versé à la fin du booking.

Le but d’HostnFly est donc de maximiser les réservations afin que le booking soit au maximum remplit, tout en assurant un prix par nuit suffisamment élevé. Autrement dit le but est de maximiser le taux d’occupation et le prix de vente de nos nuits en gestion.

Pour le revenu garanti, l’objectif pour Hostnfly est d’acheter les nuits moins chères que le prix auquel elles seront revendues à nos guest, mais “assez cher” pour arriver à satisfaire le client.

## **Objectif de HostnFly**



À la suite de la crise liée au coronavirus, l’activité de HostnFly a été compromise pendant quelques mois, les objectifs de 2020 ont donc été revus pour essayer de maintenir les chiffres de 2019.

## **Organisation de l’entreprise.**

L’entreprise est organisée pour optimiser les process et satisfaire au maximum les clients ainsi que les guests.

L’équipe data que j’ai intégrée pendant 6 mois est inscrite dans le pôle Tech de l’entreprise. Elle est composée de 4 personnes : Guilhem (Lead data scientist), Olivier (Lead data scientist), Fiona (Junior data scientist) et moi-même.

# Mission 1 : Business Intelligence

## Objectifs, contexte et outils

L’objectif premier de mon stage était d’épauler l’équipe data d’HostnFly et aider toutes les équipes pour la visualisation des données (Business Intelligence) afin d’évaluer les performances de l'entreprise et d’aider au pilotage de l’activité des différentes teams.

Hostnfly utilise la business intelligence (BI) autrement appelée informatique décisionnelle. Afin de prendre les meilleures décisions stratégiques, les données de l’entreprise sont affichées sous forme de dashboard. Chaque pôle possède ses propres dashboards. Ils sont remplis de divers types de graphique (histogrammes, tableaux, chiffres etc). L’objectif est de visualiser en temps réel les données de l’entreprise afin de prendre les meilleures décisions pour atteindre les objectifs de chaque pôle. Ainsi, c’est grâce à ces dashboards que les managers de chaque équipe peuvent consulter leurs *Key Performance Indicator* (KPI) pour atteindre leurs *Objectifs and Key Results* (OKR).

Mon rôle chez HostnFly était de maintenir et développer la visualisation des données pour l’ensemble des équipes de l’entreprise.

Afin de rendre ces données accessibles pour tout membre de l’entreprise, HostnFly utilise l’outil “**Apache Superset**” appelé Superset.

Superset est une plateforme d'exploration de données conçue pour être visuelle, intuitive et interactive, historiquement développée par AirBnb. L'objectif principal de Superset est de faciliter la visualisation des données. Superset dispose d’une trentaine de types de visualisation de données et est facile d’accès pour une personne qui ne sait pas coder. Superset est un logiciel open source gratuit, ce qui est un avantage par rapport à d’autres plateformes tel que Tableau ou Qlikview.

J’ai pu lors de mon stage découvrir l’accessibilité et la facilité que propose Superset, mais aussi bien, et très rapidement, les limites de cette plateforme. Aujourd’hui, cet outil est peu mis à jour et ne propose que très peu de nouvelles fonctionnalités.

Une vingtaine de dashboards sont utilisés sur l’espace de travail d’HostnFly. En tout, on compte plus de 150 graphiques consultés régulièrement.

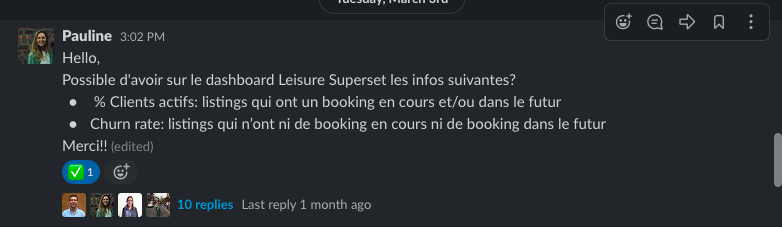
Afin de préparer mes requêtes SQL pour créer les graphiques sur Superset, j’utilisais **Dbeaver,** un logiciel de gestion de base de données, plus performant et intuitif que l’espace SQL editor de Superset.

## **Exemple d’une requête**

Comme indiqué précédemment, mon rôle était de dialoguer avec les équipes HostnFly afin de maintenir et d’assurer une bonne visibilité des données. Nous allons voir grâce à un exemple le déroulement d’une **Data request.**

### La requête

L'entreprise possède un channel de discussion slack dédié aux Data request. On y trouve les différentes demandes des employés. Comme ici, Pauline (Team Leisure), qui demande une nouvelle visualisation pour pouvoir connaître la répartition des listings actifs et passifs.



### Exécution de la requête en SQL

Après tous les détails recueillis pour effectuer la requête, je travaille sur DBeaver.

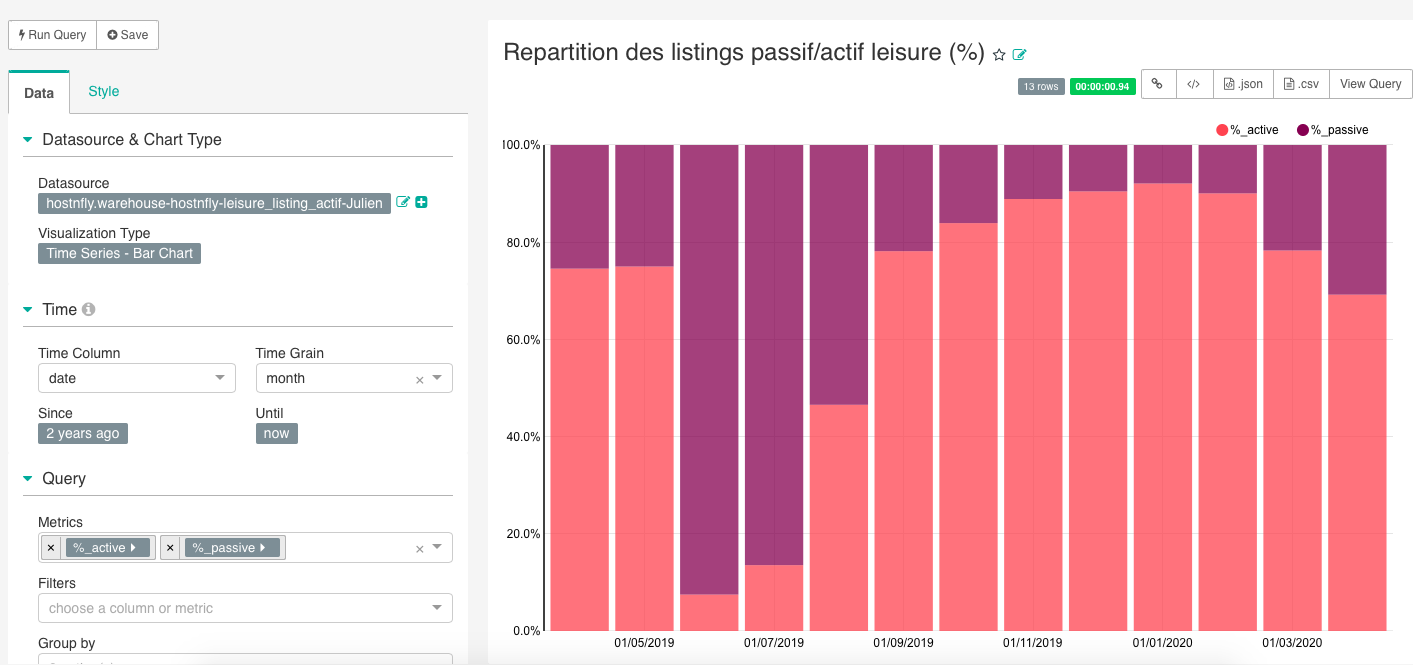
Voici l’exemple de la requête utilisée pour répondre à la demande de Pauline:

|  |
| --- |
| **select date ,   avg (case when max\_date >= date then 1::FLOAT else 0::FLOAT end ) as avg\_active,   (1 - avg\_active) as avg\_passive from (select date, listings.first\_airbnb\_date\_added\_at, listings.max\_date , market  from dates  left join listings on listings.first\_airbnb\_date\_added\_at <= date  where segment = 'client'  and market in ('Leisure\_fr',  'Saint-Tropez')) where date <= Current\_date and first\_airbnb\_date\_added\_at <= date and date >='2019-04-12' group by date order by date** |
|  |

### Mise en forme du graphique sur Superset.

J'exécute ensuite la requête sur Superset dans l’espace SQL editor. Les différents paramètres Superset me permettent d’assurer une bonne visualisation des données. Une fois la visualisation effectuée, le graphique est enregistré dans le dashboard correspondant à la team.

Dans notre exemple, le graphique a donc été ajouté au dashboard *Customer Success - Leisure*.



# Mission 2 : Projet d’analyse

## Contexte et objectif

Hostnfly possède un projet “analyse-data” où chaque membre de l'équipe tech (développeur et data) peut ajouter ses analyses sur les données HnF. La majorité de ces analyses sont codées en python et sont présentées sous forme de notebook grâce à **Jupyter Notebook**.

L’objectif de ces analyses est d’explorer les données que possède Hostnfly. Ces analyses permettent ensuite d’aider la prise de décisions sur les stratégies de l’entreprise. Cela peut par exemple être une analyse technique sur les algorithmes de pricing, ou bien marketing comme le comportement de nos clients.

Lors de mon stage, particulièrement depuis le début de la crise du coronavirus, j’ai pu réaliser quelques analyses de données afin d’établir des bilans sur certaines thématiques de l’entreprise. Nous allons ici présenter un exemple d’analyse exploratoire de données.

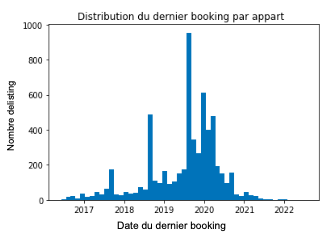
## Exemple D’analyse : **LifeTime Value (LTV)**

### Contexte et problématique

Dans un objectif d’une meilleure rentabilité d’acquisition client, l'équipe marketing voudrait connaître la LTV des clients HostnFly. La LTV est un terme utilisé en marketing pour désigner la valeur vie d’un client. Chez HostnFly deux dimensions ont une forte influence sur la valeur de la vie d’un client : Le nombre de nuits que l'on a pu acheter au client et le prix de vente des nuits aux guests. Elle peut être exprimée en en chiffre d’affaire ou bien en bénéfice.

Hypothétiquement, on suppose qu’il est plus rentable d'acquérir des listings de grande capacité d’accueil. En effet, nous pouvons vendre les nuits plus chères et ainsi générer plus de profit pour l’entreprise. Mais les listings à forte capacité se remplissent moins bien que les petits listings ; ils sont aussi surement moins disponibles à l’année que des petits listings qui sont parfois disponibles plusieurs week-ends dans l’année.

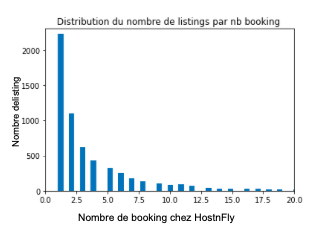
L’objectif est donc de découvrir s’il y a une catégorie de client qui est plus ou moins rentable sur l’investissement marketing. Autrement dit, y-a-t-il une tranche d'âge, une catégorie d’appartement ou bien une ville sur laquelle nous devrions apporter plus d’investissement marketing plutôt qu’une autre.



### Situation globale de nos listings.

Nous pouvons voir sur ce graphique la distribution du dernier booking par listings en fonction du temps. On peut voir qu’une grande majorité de nos clients ont mis des dates disponible leur bien dans la dernière année (entre début 2019 et aujourd”hui).

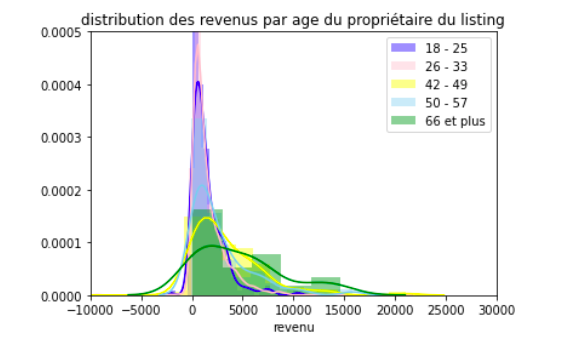
Un client est catégorisé comme ***churned (*plus intéressé)** s’il n’a pas de booking dans le futur et qu’il n’a plus mis de booking en ligne depuis un an. Sur un total de 5949 listings en gestion, on peut voir qu’une partie importante (1820) de listing ne sont plus intéressés par HostnFLy, c’est à dire que leur dernier booking est antérieur à avril 2019.



Dans la deuxième figure, on présente ici la distribution des listings en fonction du nombre de bookings mis en ligne par listing.  
Ce graphique est intéressant car l’on peut voir qu’une majorité de client nous ont confié qu’une seule fois leur logement. Plusieurs hypothèses peuvent être déduites de ce constat :

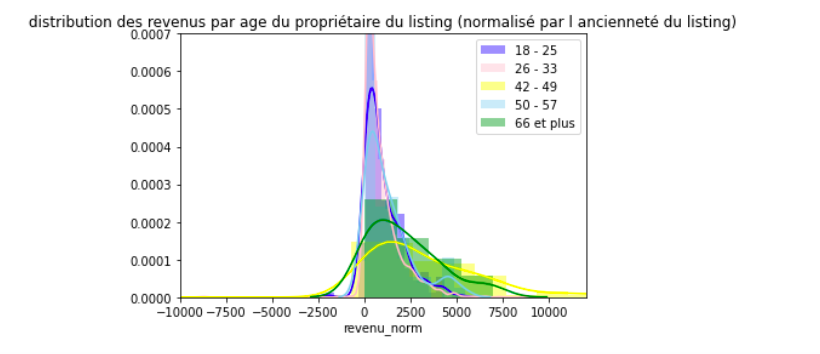
* Les clients n’ont pas apprécié leurs expériences avec HostnFLy.
* Les clients n’ont tout simplement pas eu le temps de mettre deux bookings (l'âge du client est trop jeune), ajouter des nuits sur HostnFly n’est pas encore une habitude pour eux. Cette hypothèse a été privilégiée car HostnFly a une forte acquisition client depuis un an.

De plus, il est important de noter que les clients qui ont confié leur bien plusieurs fois (plus de 4/5 fois) à HostnFly sont fortement susceptibles de continuer leur aventure avec HnF : il s’agit probablement des clients convaincus par le service

Dans cette figure, on affiche la distribution des listings en fonction du revenu généré dans la vie du client. Nous affichons cela pour différentes tranches d'âges du propriétaire. 

Toutes les tranches d'âge n’ont pas été affichées dans un souci de lisibilité.

On peut voir que les jeunes clients (courbe bleue et rose) apportent globalement un revenu uniforme proche de 800e/an. Au contraire, les clients les plus âgés (courbe verte et jaune) ont tendance à générer un revenu plus important et plus variable durant leur vie client. Néanmoins, ce graphique est biaisé par la durée de vie des clients. Il est donc important de normaliser le revenu généré par an par l’ancienneté du client.

Les conclusions apportées à ce graphique sont similaires au graphiques précédent. Les tranches d'âges les plus âgés ont tendance à générer un revenu plus important. 

### LifeTime Value

Pour notre analyse nous avons choisi de définir la LifeTime Value par :

**LTV = Chiffre d’affaire annuel x durée d'âge client**

**Chiffre d’affaire annuel =** Chiffre d’affaire du client / (today - created date)

Le chiffre d’affaire du listing est aussi appelé Hosts\_payout chez HostnFLy.

Nous avons fait le choix de travailler sur le chiffre d’affaire plutôt que sur le revenu pour éviter d’être biaisé par le type de revenu (garanti ou pourcentage) choisi par le client.

**Durée d’âge client** = last\_booking\_end\_date + 1 an - created date (si le client n’est plus intéressé par HnF)

= Today - Created\_date (si le client est toujours chez HnF)

Autrement dit, l'âge d’un client est la période depuis sa création de compte plus un an s’il n’est plus intéressé par HnF. On ajoute un an car on estime qu’un client est intéressé pendant au moins un an après la fin de son contrat. En effet une grande partie des clients confient leur logement une fois par an. Au contraire, si le client est toujours intéressé, on dira que sa durée de vie court de la date de création de compte jusqu'à aujourd'hui.

Afin d’assurer une fiabilité dans les résultats, nous avons choisi de sélectionner uniquement les clients qui ont plus d’un an d’ancienneté chez HostnFly. Ceci évite d’accorder une grande importance aux clients qui nous ont confié leurs appartements récemment (car today\_date - created\_date serait bien trop petit).

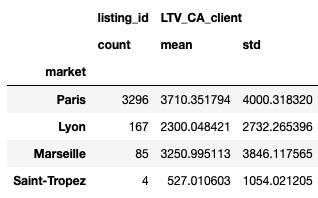
Ainsi nous avons pu calculer la LTV sur différentes catégories : l'âge, le marché et la capacité d’accueil.

#### LTV par tranche d’Age.

Le tableau de droite présente les statistiques sur la LTV par catégorie d'âge. En moyenne, on peut voir que la LTV a tendance à être plus grande avec l’âge. Seules les catégories de 50-57 et 58-65 ans font exception à cette tendance, elles montre une légère baisse par rapport au 42-49 ans.

On peut expliquer la tendance générale par une différence de richesse des populations. En effet on peut supposer que les tranches d’âge jeunes logent dans des catégories d’appartements semblables et plus petit. Au contraire, les populations les plus âgés possèdent des logements plus spacieux, avec une plus grande capacité d'accueil voir même des résidences secondaires. Ainsi il est préférable d’investir dans des campagnes marketing pour attirer des populations plus âgées car elles génèrent plus de chiffres durant leur vie de clients.

#### LTV par marché.



On réalise la même analyse mais cette fois ci on regarde si certains marchés méritent un plus grand investissement que d’autres.

Il est difficile d’apporter des conclusions à cette analyse pour plusieurs raisons :

* Les prix des nuits à Paris sont globalement plus élevés, il est donc normal de constater que les listings parisiens ont une plus grande LTV.
* Les marchés ont ouvert à des intervalles de temps différent, ainsi les algorithmes de pricing ne performent pas de la même manière en fonction du marché.
* L’effectif des villes est assez faible, il est donc difficile de comparer Paris aux 85 listings de Marseille.

#### LTV par catégorie de capacité d’accueil.

Enfin nous avons calculer la LTV pour des capacités d’accueil différentes. On peut constater qu’en moyenne, plus le logement a une capacité d’accueil grande, plus il génère du chiffre d’affaire lors de sa vie de client. Ainsi, nous confirmons notre hypothèse initiale.

**Attention :** Toutes les informations apportées sont des tendances. Comme nous pouvons le voir dans les tableaux, les écarts-types (std) prennent des valeurs extrêmement élevées par rapport à la moyenne. Ainsi les calculs de LTV montrent une grande variabilité, ce qui ne nous permet pas d’assurer statistiquement un effet de l'âge ou de la capacité d’accueil sur la LTV.

Cette analyse a été réalisé sur les données des trois premières années de l’entreprise. Ainsi, les données ne permettent pas de connaître réellement le comportement des clients HostnFly. Il serait intéressant de réaliser une analyse similaire dans quelques années pour avoir des durées de vie client plus importantes et ainsi avoir une meilleurs fiabilité des résultats.

# Bilan sur le stage

## Autres missions

Les deux grandes missions décrites précédemment ont été les deux piliers de mon stage, au-delà de ça, j’ai pu y faire d’autres missions telles que des analyses du code, des réflexions pour améliorer le fonctionnement de l’entreprise (notamment lié à la crise du covid-19) et développer des compétences.

Les analyses du code m’ont permis de me familiariser un peu plus avec la programmation python et particulièrement la Programmation Orientée Objet. Cela m’a permis aussi de me rendre compte de la manière dont est construit le code des projets à l’échelle d’une entreprise. Les réflexions d'amélioration mon permis de mieux comprendre le fonctionnement de l’entreprise ainsi que le travail en équipe. Enfin, la formation sur différents outils a été directement appliquée sur des projets, ce qui m’a permis de valider les compétences liées.

## Attentes sur le stage

Lors de mes entretiens pour le stage, j’ai pu découvrir le fonctionnement de la boite, l’ambiance qu’elle dégageait et cela me correspondait. Lors de mon arrivée j’ai pu confirmer ce que j’avais entrevu et le stage me correspondait. Des bilans d’équipes ont été réalisés toutes les deux semaines et des présentations générales ont été faites de façon hebdomadaire. Le travail d’équipe et la communication sont des valeurs ancrées dans le fonctionnement de l’entreprise, ce fonctionnement m’a fortement plu.

## Compétences en Data Science

Comme indiqué précédemment, lors de mes missions j’ai pu travailler sur des outils que je ne connaissais pas. J’ai pu utiliser un logiciel de visualisation de données dans lequel j’ai pu développer mes compétences en SQL. J’ai aussi découvert Jupyter Notebook dans lequel j’ai pu développer mes compétences en programmation Python.   
J’ai aussi découvert de nouveau outils qui sont au centre de la data science : l’environnement de développement Python PyCharm. Ou bien le service de gestion de logiciel de développement GitHub. J’ai pu découvrir Apache Airflow, une plateforme de gestion de *workflow.*

Par ailleurs, j’ai développé d’autres compétences comme la présentation de résultats ou la rédaction de rapport de projet.

## Les points positifs et négatifs

Mes attentes ont été satisfaites, et c’est un point qui a enrichi mon stage. Les points négatifs sont liés à la crise du Coronavirus. En effet, tous les CDI de l’entreprise ont été placés en chômage partiel entre 0 et 30 % d’activité. L’équipe Data a quant à elle été placé à 0%. Je me suis donc retrouvé seul à travailler quotidiennement. Malgré cela, Baptiste, le CTO m’a encadré et me donnait régulièrement des missions. Même dans ce contexte difficile, HostnFly a maintenu leurs engagements avec les stagiaires et a réussi à me donner des missions très intéressantes. Ce contexte particulier a tout de même été valorisant malgré un nombre de missions nettement inférieurs à la normale.