

Rapport de stage «Data Analyst»



Ibtissam YADANE

MACS 3

Tuteur pédagogique : Olivier LAFITTE

Ingénieur responsable : Rémy PEYRET

2018-2019

RAPPORT DE STAGE «DATA ANALYST»

Table des matières

Remerciements.....	3
Introduction	4
Présentation de l'entreprise	5
Hxperience	5
L'offre de la société dans le domaine	5
Contexte actuel.....	6
Présentation de SMATI.....	7
Architecture du produit.....	8
Mon expérience	10
L'objectif et le déroulement du stage	10
Description des missions	11
Méthodes de travail	11
Technologies et outils utilisés	13
Travail effectué.....	16
Retour personnel sur le stage.....	28
Conclusion.....	29
Bibliographie.....	30
Acronymes	31
Annexes	33
Partenaires de Hxperience.....	33

Remerciements

Je tiens à remercier tous les employés de l'entreprise, pour l'expérience enrichissante et pleine d'intérêt qu'ils m'ont fait vivre durant ces six mois au sein de l'entreprise Hxperience.

J'aimerais également remercier mes supérieurs, Patrick FICHOU (CEO d'Hxperience), Rémy PEYRET (Data Scientist) pour leur confiance et leur soutien lors de cette collaboration, j'ai beaucoup appris pendant cette période. Ils m'ont vraiment bien intégré à toute l'équipe existante et encadré pendant ce stage. Je poursuis en remerciant toute l'équipe d'Hxperience pour leur accueil et leurs précieux conseils; j'ai pu évoluer dans un cadre agréable et dynamique rythmé d'échanges et de partages de connaissances.

En addition, j'aimerais remercier mon tuteur de stage, Olivier LAFITTE, pour son engagement dans le déroulement de mon stage ainsi que tous mes professeurs pour m'avoir conseillée et aidée dans ma démarche professionnelle.

Finalement, j'aimerais remercier le corps enseignant de Sup Galilée pour m'avoir aidée dans ma démarche professionnelle.

Introduction

Dans le cadre de ma formation d'ingénieur en Mathématiques Appliquées et Calcul Scientifique (MACS), j'ai effectué mon stage technicien de six mois chez Hxperience une startup qui incube à Boucicaut à Paris, dans le 15ème arrondissement.

L'idée de ce rapport se concentre sur la démonstration et l'explication du travail que j'ai réalisé tout au long de mon stage et les compétences acquises. Ainsi que son influence sur moi et sur mon parcours professionnel souhaité. Ce rapport est divisé en trois parties.

Dans la première partie, je donnerai une introduction du contexte de mon stage en passant par la présentation de l'entreprise et du domaine de travail ainsi que le produit développé par l'entreprise.

Dans la deuxième partie, je fournirai une description du travail à réaliser initialement dans l'entreprise au sein de son domaine d'activité. Je parlerai également des méthodes de travail utilisées au sein de l'équipe data, et notamment les technologies adoptées. Et finalement j'expliquerai en détails les tâches effectuées, avec le travail fourni tout en donnant des exemples concrets.

Dans une troisième partie, mon retour personnel sera décrit.

Enfin, je terminerai le rapport par la conclusion qui résume ma progression et ma vision globale du stage. Vous trouverez également à la fin du rapport: un résumé exécutif et une annexe.

Présentation de l'entreprise



Hxperience est une Startup créée en 2015, par des cadres issus de Nokia Networks, considéré comme le premier opérateur de services numériques pour le bâtiment. Elle a été sélectionnée pour rejoindre Paris&Co, l'agence de développement économique et d'innovation de Paris.

L'organisation structurelle de l'entreprise s'articule autour des fondateurs, d'une équipe « sales » composée de commerciaux et d'un Product manager. Une équipe technique complète le tout. Cette dernière est composée d'une équipe de développeurs (front end et back end) ainsi que d'une équipe data au sein de laquelle j'ai réalisé mon stage. L'entreprise conçoit une suite logicielle d'applications en mode SaaS à destination des exploitants et gestionnaires d'immeubles (résidentiels, tertiaires, industriels) dont le but est d'optimiser l'exploitation du bâtiment et d'en maximiser l'usage. Elle travaille dans la proposition et l'application des solutions pour ses clients dans les domaines du Smart building et de l'industrie. Elle s'appuie sur les technologies IoT, une plateforme ouverte qui se connecte aux équipements du bâtiment, récolte et analyse les données transmises grâce à des algorithmes.

L'offre de la société dans le domaine :

Les bâtiments, leurs équipements et leurs usages constituent une source de données inestimable, et l'internet des objets (IoT) permet aujourd'hui de récupérer ces données en temps réel, de les faire parler et de les rendre intelligentes. L'idée est de fédérer l'ensemble des acteurs du bâtiment autour d'un outil commun, en vue de partager les données émises et de les transformer en valeur ajoutée. Hxperience s'appuie notamment sur une plateforme qui collecte, sécurise et traite les données du bâtiment. Les services proposés permettent aux entreprises exploitant un bâtiment de suivre son état opérationnel en temps réel, d'échanger des informations associées et de déclencher des actions à distance.

Concrètement, l'offre Hxperience est la suivante :

- Numérisation

Hxperience connecte les équipements et les bâtiments du client afin d'en extraire des données pour suivre leurs états et analyser leurs usages.

- Plateforme de services

La plateforme Hxperience, se connecte aux équipements du client quels qu'ils soient. Elle récolte, traite et sécurise les données transmises et déclenche des actions en conséquence.

- Applications suite

A travers une suite logicielle ou la création d'applications sur mesure, Hxperience accompagne le client dans son optimisation opérationnelle et la création de nouveaux services tels que la maintenance prédictive, la gestion des actifs, l'exploitation des bâtiments.

Contexte actuel :

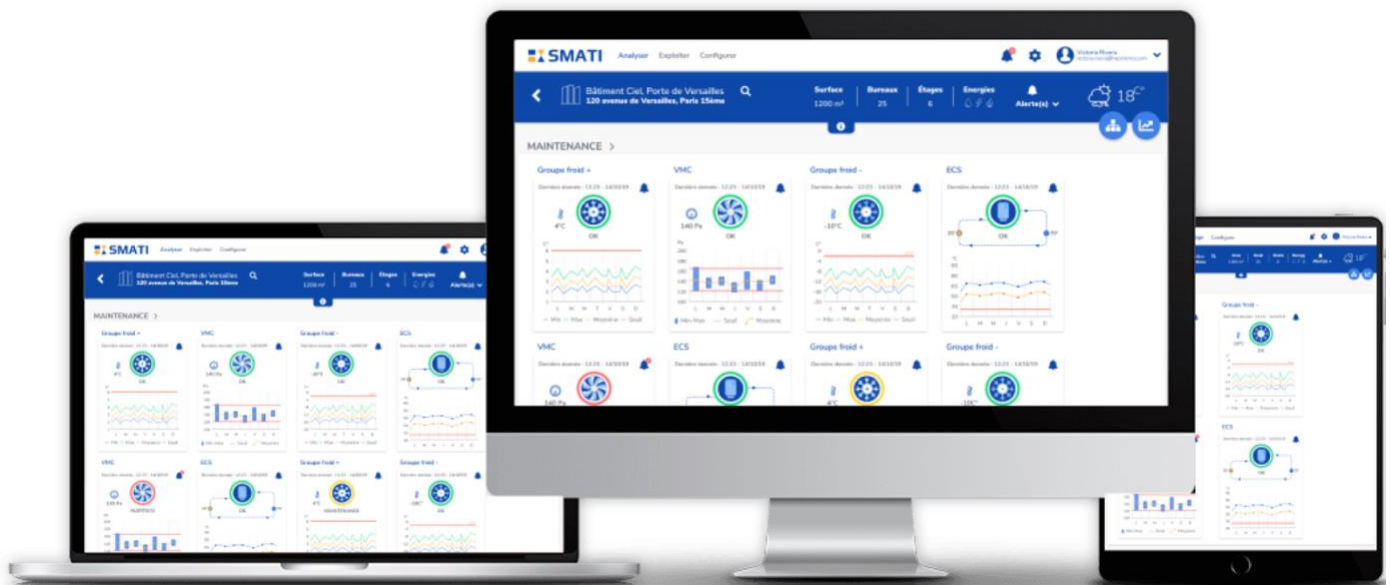
L'IoT est un terme très tendance depuis quelques années mais il n'est pas toujours facile de savoir de quoi il s'agit exactement. Afin que ce concept soit plus compréhensible du grand public, les médias créent un buzz autour de l'engouement pour certains objets connectés tels que les montres, les vêtements ou encore la domotique et il est vrai que la portée de ces différentes applications est importante.

Cependant, pour la majeure partie des entreprises, l'enjeu essentiel de l'IoT est tout autre, à la fois plus stratégique et plus complexe. En effet, au cours de la prochaine décennie, selon le cabinet de conseil McKinsey, 70% de la valeur créée par l'IoT proviendra des applications dans le B2B (Business to Business) tels que la production industrielle, l'agriculture, le domaine de la santé dans son sens le plus large, la construction, les transports, les bureaux etc. À cette création de valeur viendront s'ajouter de profonds changements structurels dans de nombreuses industries.

On peut classer les bénéfices attendus de l'IoT en deux grandes catégories :

- L'amélioration de l'efficacité opérationnelle (amélioration de la maintenance des équipements, optimisation de la durée de vie des équipements, transformation de la gestion d'inventaire)
- La transformation de la relation client (meilleure connaissance des usages et fonctions des produits, création de produits améliorés et de nouveaux services)

Présentation de SMATI:



SMATI est une solution logicielle qui collecte, traite, analyse et valorise les données du bâtiment. Les applications métiers disponibles dans SMATI permettent d'optimiser l'exploitation de l'immobilier. Il est commercialisé sous forme d'abonnement (Software as a Service) et il s'adapte à l'évolution du besoin et le cas d'usage du client. Il permet de superviser et pilote les bâtiments en temps réel.

De plus, il aide les clients à identifier rapidement leurs dérivés de fonctionnement ou de leur consommation afin de les corriger. Grâce aux fonctionnalités analytiques de SMATI, les données sont analysées afin de comprendre les usages. Ces données aident à créer de nouveaux services pour les clients.

Les fonctionnalités de SMATI sont :

- Configuration :
 - Configuration des entités, équipements, capteurs, alertes et notifications
 - Gestion des utilisateurs et possibilité de définir différents droits d'accès pour les utilisateurs internes
- Exploitation :
 - Création automatique de tableaux de bord et indicateurs prédéfinis par type de bâtiment (entités)
 - Visualisation sous forme de widgets suivant différents thèmes (maintenance, énergie, gestion d'espaces, confort)
 - Alertes et notifications à destination des équipes de maintenance et utilisateurs

- Analyse :
 - Création et personnalisation des rapports d'analyse
 - Analyse et modélisation des usages et consommations
 - Identification des dérives de fonctionnement et maintenance prédictive

Architecture du produit :

L'automatisation de l'envoi et la réception des données :

En ce qui concerne l'accès aux capteurs des différents équipements des différents logements assez distants, plusieurs technologies réseaux sont mises en place pour garantir cet accès à travers certaines plateformes et APIs.

La plateforme de Sigfox est utilisée comme un réseau où la connexion des objets connectés est établie, ce qui aide la collection de données. Parmi les protocoles de communications utilisées dans cette étape, il y a LoRa. Elle offre aussi des points d'accès. Un autre exemple comme «Objenious», filiale de Bouygues Télécom, accompagne les entreprises dans leur mission dans le contexte de l'IoT. Cette entreprise offre un cloud qui reçoit les données de tous les capteurs liés, les données sont ensuite collectées par SMATI et le client de Hxperience peut les visualiser sur son espace dédié.

L'architecture des composants associés :

Techniquement, pour avoir accès aux données des capteurs, on doit étudier le trajet de ces données. Tout d'abord, les capteurs sont intégrés ou installés dans l'équipement qui peut être une chaudière, un appareil pour traquer la consommation d'énergie. Afin de regrouper plusieurs capteurs ensemble selon un certain type ou selon un certain lieu, ils doivent être attachés ou connectés à un transmetteur-émetteur des données. Pour nous, on utilise souvent le mot « transmetteur » dans notre produit de version française, mais il s'agit d'envoi et aussi de réception des données, donc, c'est un transmetteur-émetteur pour être plus précis. Enfin, ces transmetteurs sont attachés à un point d'accès « Access point » qui est la passerelle. C'est à partir de la passerelle les données sont envoyées directement à la plateforme suivant un certain réseau qui dépend de l'entreprise de réseaux partenaires tel que montré sur la figure suivante.

RAPPORT DE STAGE «DATA ANALYST»



- Équipement : equipment
- Capteur : device
- Transmetteur : transceiver
- Passerelle : gateway
- Utilisateur : user
- Un groupe des utilisateurs : user group
- Rôle : role

Mon expérience

L'objectif et le déroulement du stage :

Le sujet de stage est porté sur l'enrichissement des solutions mises en place pour la détection et la correction automatisées d'anomalies dans les données IoT collectées.

Ma mission portera principalement sur :

- Détection d'anomalies à l'aide des techniques d'intelligence artificielle (outlier détection)
- Analyse des données temporelles et création de filtres intelligents pour la complétude des données manquantes
- Enrichissement des rapports existants pour le diagnostic de la qualité des données avec de nouvelles fonctionnalités
- Développement de nouvelles fonctionnalités liées à l'analyse et au traitement des données
- Création d'une application Django qui consiste à créer des rapports pour analyser des données temporelles et les afficher sur différents graphes (histogramme, camembert ...)

J'ai intégré l'équipe Technique de Hxperience en tant que Stagiaire Data Analyst. La première semaine était une semaine d'intégration. Dans un premier temps, j'ai fait connaissance avec les membres de l'équipe et par la même occasion je me suis informée sur le travail de chacun d'entre eux, tout en profitant pour leur poser des questions sur l'entreprise, le logiciel qu'ils développent de façon à ce que j'aie une vision plus claire sur mes missions. Entre-temps quelques membres de l'équipe ont consacré un peu de leur temps, en vue de m'expliquer l'utilité et le fonctionnement de leur produit, en outre les technologies utilisées et le travail que je dois effectuer.

Par ailleurs j'ai commencé à installer mon environnement de travail Linux avec tous les outils nécessaires pour travailler convenablement sur la partie Backend du logiciel. Malgré les difficultés que j'ai rencontrées sur cette partie, j'ai réussi à installer tout ce qu'il faut pour commencer à travailler.

Pour bien me préparer pour la suite, j'étais amené à lire des cours sur python, Django, Git et une documentation portée sur les scripts data utilisés dans l'entreprise.

Toutefois, j'ai rencontré quelques lacunes pour bien comprendre les technologies utilisées (Python, Django, Git), à ce moment, j'avais le droit de solliciter l'aide des membres de l'équipe technique qui restaient disponibles pour répondre à mes questions, voir m'accompagner jusqu'au bout dans mes tâches. Parfois je pouvais écrire mes questions sur l'application qu'on utilise pour communiquer qui s'appelle Microsoft Teams.

Description des missions :

Durant mon stage j'ai réalisé plusieurs missions, qui étaient dans un premier temps centrées sur la correction des scripts existants permettant le diagnostic de la qualité des données, l'analyse et au traitement des données pour le besoin interne de Hxperience. Par la suite, l'objectif était de créer une application Django permettant de générer un rapport énergétique ayant pour cible les clients d'Hxperience. Ces rapports analysent les données temporelles des compteurs électriques et de gaz et les affichent sur différents graphes.

En ce qui concerne la correction des rapports existants pour le diagnostic de la qualité des données, les étapes à suivre étaient précises :

- Premièrement, lire et comprendre la documentation portée sur les scripts data.
- Une fois le fonctionnement des scripts compris, les adapter avec les changements que l'équipe de développements ont fait et les tester.
- Finalement, développer des fonctionnalités liées à l'analyse et au traitement des données.

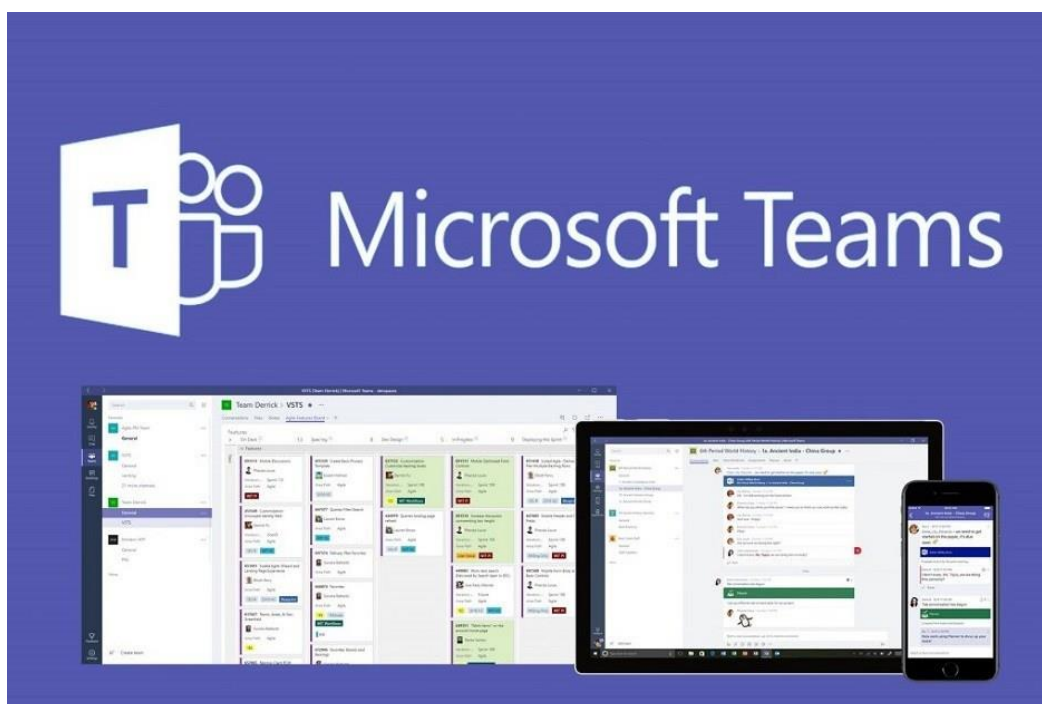
Au fur et à mesure de l'avancement de mon stage ma mission s'est orientée vers une partie importante du produit, qui est la création d'une application Django qui consiste à créer des rapports énergétiques. Cette dernière demande une connaissance plus importante, non seulement en python Django et HTML Template mais également dans l'architecture de l'application.

Méthodes de travail :

L'organisation et les méthodes de travail, font partie des choses qui m'ont le plus intéressé dans une entreprise. En effet on ne les a pas toutes vues durant notre formation, et encore moins les avoir pratiquées dans un environnement professionnel. Dans cette partie je citerai les méthodes de travail adoptées dans l'entreprise, ainsi que les outils les plus pertinents.

a. Microsoft Teams :

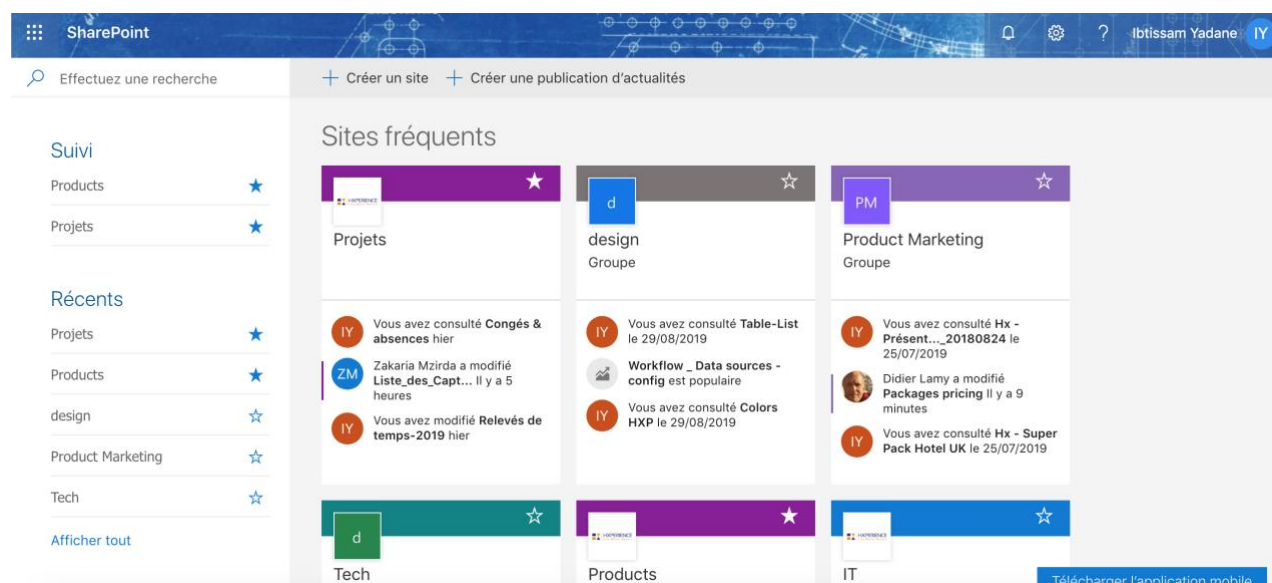
Microsoft Teams est une plate-forme de communication, ainsi qu'un logiciel de gestion de projet que l'on utilise dans Hxperience. On peut éventuellement utiliser d'autres moyens de communication, mais Microsoft Teams reste le plus utilisé, surtout pour communiquer entre les membres de l'équipe data. Soit pour poser des questions, signaler des problèmes, ou même communiquer avec l'équipe si on est en télétravail. On dispose de plusieurs chaînes sur Microsoft Teams, dédiées aux différents usages. Notamment une chaîne pour des questions techniques, data, une chaîne pour des bugs et pleins d'autres.



b. SharePoint :

SharePoint est un ensemble de logiciels développé par Microsoft. Les produits que propose SharePoint, offrent plusieurs fonctionnalités aux utilisateurs, notamment la gestion de contenu, la gestion électronique de documents, les forums et aussi la possibilité de faire des formulaires et des statistiques. Hxperience utilise SharePoint principalement pour stocker les documents importants.

Voici la structure de SharePoint de Hxperience :



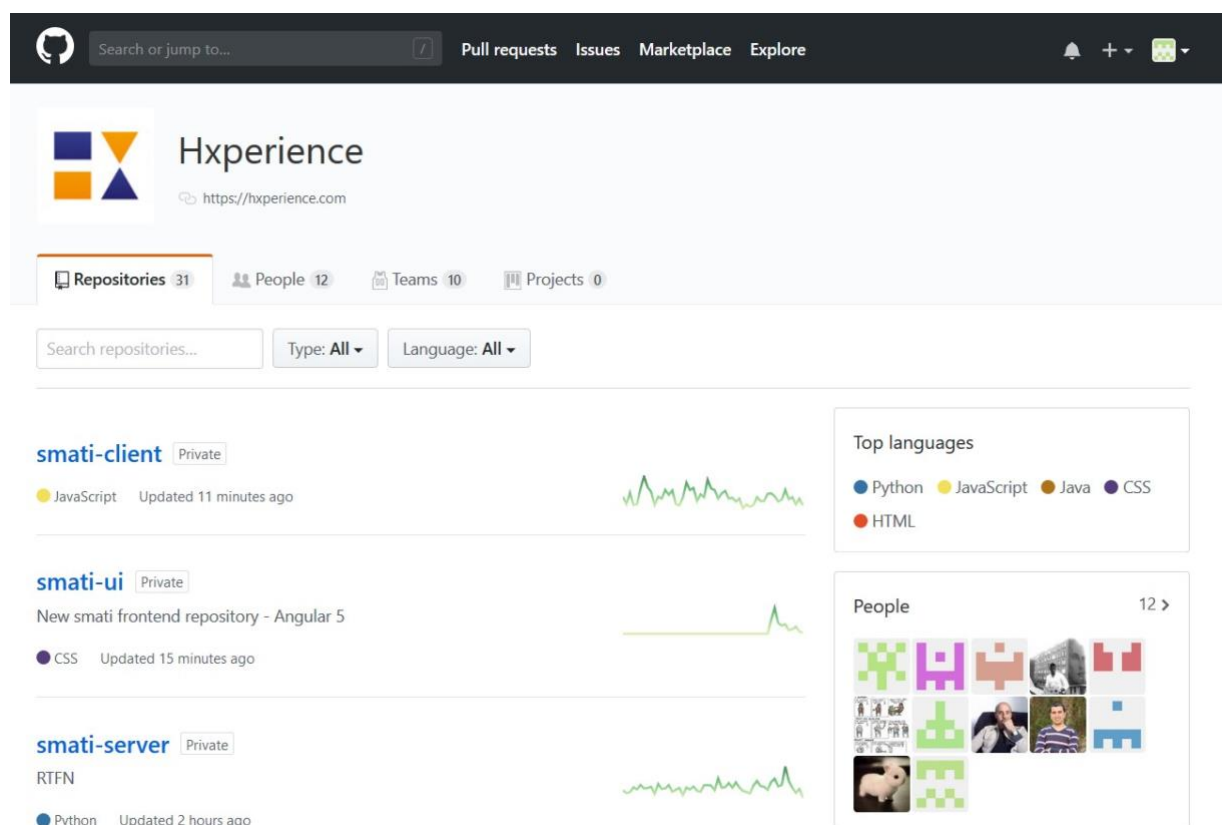
Il y a plusieurs sites, chacun est dédié pour une certaine fonctionnalité.

- Projets : ce site contient des documents sur les projets que Hxperience a avec plusieurs clients, comme SNCF, Accor, Logista et pleins d'autres.
- Fournisseurs : ce site contient la documentation des capteurs de tous les fournisseurs, ainsi que des documentations utiles sur les Gateway, en outre les plateformes utilisées.

c. *GitHub* :

GitHub est une plateforme d'hébergement de code et de gestion de développement de logiciels. Il permet à plusieurs personnes, de travailler ensemble sur des projets de n'importe où.

Hxperience utilise GitHub, afin d'héberger le code de divers projets, ci-joint une capture d'écran de quelques projets de Hxperience sur GitHub :



Je travaille essentiellement sur la partie Backend de SMATI, autrement dit le projet « smati-server ». Plus précisément sur les modules « data_quality » et « data_reporting », que j'expliquerai dans une autre partie.

Technologies et outils utilisés :

L'internet des objets connectés, qui est le domaine d'expertise de Hxperience, représente l'extension de l'internet que nous utilisons tous au quotidien. En revanche, l'internet ne se propage pas dans le monde électronique. D'où l'intérêt de l'IoT, qui représente les échanges

de données et d'informations émises par des dispositifs connectés dans le monde à travers le réseau internet.

Ainsi, l'IoT est considéré comme la troisième évolution de l'internet. En réalité l'IoT est responsable de l'accroissement exponentiel de volume de données, générées sur le réseau, via nos téléphones, nos ordinateurs, ou bien n'importe quel dispositif connecté à l'internet. Hxperience s'intéresse aux objets connectés, plus précisément les capteurs.

Or, durant mon stage j'ai eu l'occasion de manipuler des données pour créer des rapports. Dont le but de les intégrer dans le logiciel SMATI.

Cependant, l'intégration de ces rapports, demande des connaissances sur les différentes technologies utilisées par Hxperience. Par conséquent, il est indispensable d'avoir les bases nécessaires en Python 3, Django et HTL Template, ainsi que les outils et les logiciels qui ont un rapport avec ces technologies.

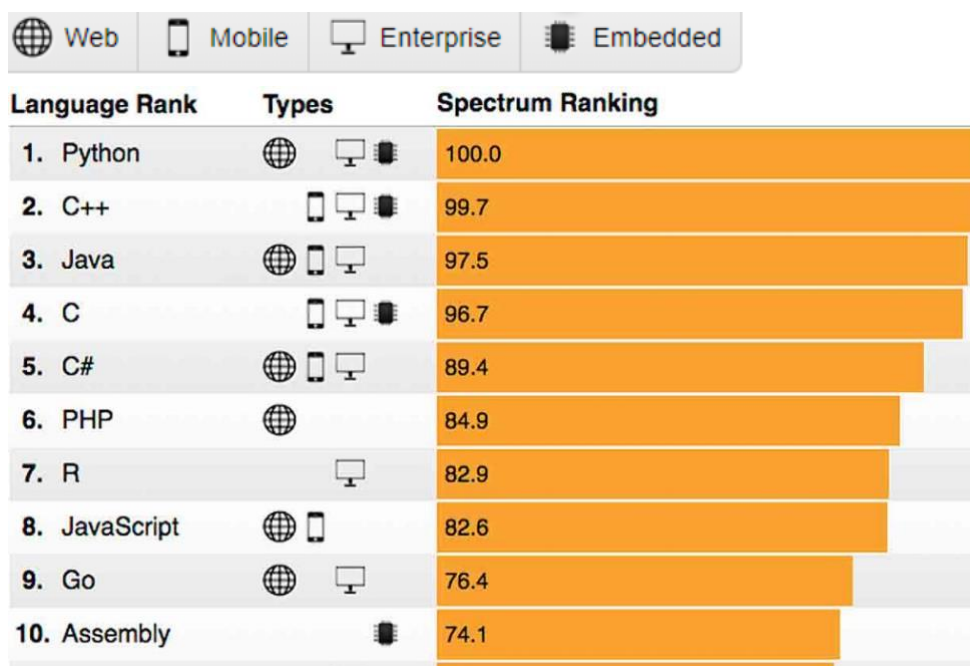
a- Python

Python est un langage de programmation de haut niveau, il est à la fois structuré et open source. Il est notamment, conçu pour être orienté objet, cependant il dispose de plein d'outils, qui ne le rendent pas seulement utilisable dans la programmation fonctionnelle, mais également dans la programmation impérative. Il est notamment multi-usage et multiparadigme, ce qui signifie qu'on peut considérer les programmes créés avec python comme une collection d'objets en interaction, dans le cas d'une programmation objet. Tandis que l'utilisation de python dans la programmation fonctionnelle, change la vision des développeurs sur les programmes, ainsi ils sont vus comme une suite d'évaluations de fonctions sans états.

Toute la partie Backend de SMATI est faite avec python, donc il était nécessaire pour moi d'acquérir les bases importantes sur ce langage pour que je puisse commencer mes missions.



Aujourd'hui, la popularité de python est assez importante. Surtout avec la montée en croissance dans les domaines de big data et les objets connectés. D'une part, python est le langage de programmation le plus compatible avec Hadoop qui est la plateforme open-source la plus populaire du Big Data. D'une autre part, c'est le langage le plus utilisé avec la Raspberry Pi, qui est l'un des microcontrôleurs les plus populaires sur le marché. L'internet des objets (IoT) révolutionne la manière dont les appareils / objets interagissent les uns avec les autres. Et lorsqu'on a l'IoT avec Python on peut créer et programmer des objets interactifs. Voici ci-dessous, le top 10 des langages de programmation en forte croissance en 2018/2019.



b- Django

Django est un Framework de développement Web open-source écrit en python. L'objectif principal des Framework, est de rendre le développement Web plus rapide et plus simple. Il est utilisé avec python pour ajouter des fonctionnalités importantes dans le produit SMATI, comme les modèles.

Or, pour mes missions, il n'était pas obligatoire de connaître et de comprendre beaucoup de fonctionnalités sur ce Framework. Je me suis limité juste à la partie qui concerne les « Queryset », en effet cette partie m'a permis de comprendre comment je peux communiquer avec la base de données, afin de récupérer des données selon le besoin.

C- HTML Template

En HTML, on ne peut pas mettre directement du code Python car les navigateurs seraient incapables de le comprendre. Les navigateurs ne connaissent que le HTML. HTML est du genre statique, alors que Python est bien plus dynamique.

Django fournit des balises de gabarit (**Template tags**) qui sont intégrées au Framework. Donc les balises de Template Django permettant de transférer des choses ressemblant à du Python dans du HTML afin de construire des rapports.

D- CouchDB

CouchDB est un système de gestion de base de données, orienté document. À savoir que la base de données gérée par CouchDB, est du type NoSQL.

Cependant lorsqu'on travaille à distance un autre système de gestion de base de données NoSQL est utilisé, c'est « Cloudant » d'IBM.

E- PostgreSQL

C'est un système de gestion de base de données relationnelles. On l'utilise dans SMATI pour stocker les données dans des tables.

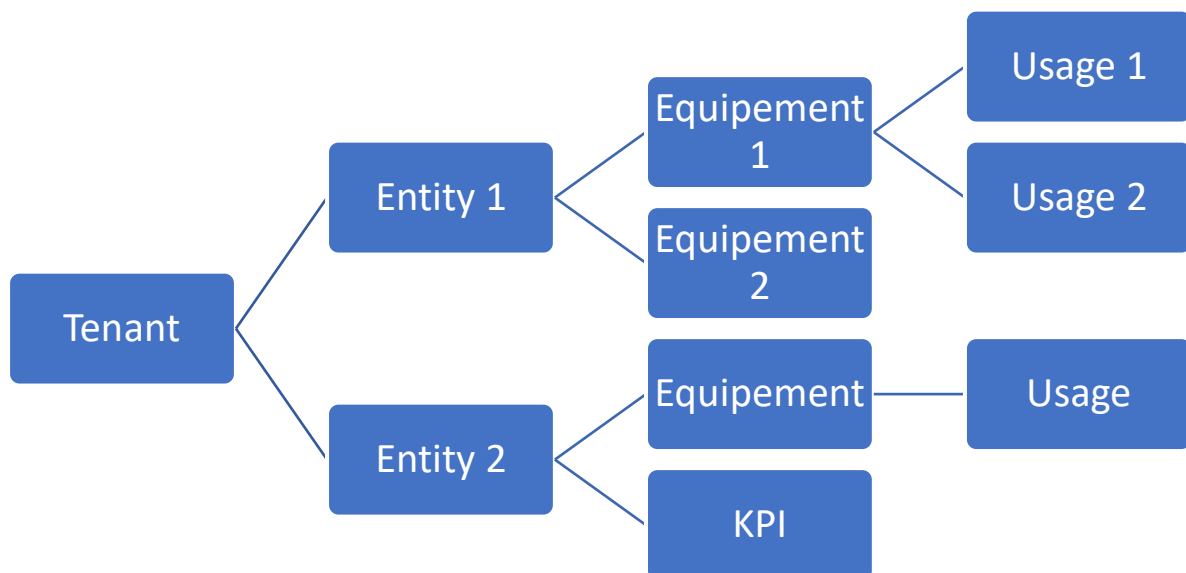
A savoir que, PostgreSQL possède de nombreuses caractéristiques, tel que :

- La possibilité de gérer les tables avec une interface graphique.
- Des bibliothèques pour plusieurs langages de programmation.
- Il possède aussi, une API ODBC (Open Database Connectivity), permettant à n'importe quelle application supportant ce type d'interface d'accéder à une base donnée PostgreSQL.

Travail effectué :

Dans cette partie je vais parler, du travail que j'ai effectué durant mes 6 mois de stage chez Hxperience. Dans un premier temps, je présenterai la partie Backend sur laquelle j'ai travaillé, c'est-à-dire les scripts « data_quality » que j'ai améliorés et les adapter avec les changements que l'équipe de développement ont faits. Ensuite dans un second temps, je parlerai de l'application Django « data_reporting » que j'ai créé.

La figure ci-dessous représente l'architecture de la base de données



- Tenant : désigne un client.
- Multi-Tenant : une seule instance d'application va servir plusieurs clients. Chaque client est appelé locataire (tenant en Anglais). Elle s'oppose à une architecture Single-Tenant où chaque organisation cliente à sa propre instance d'installation logicielle ou matérielle

- Entité : un tenant peut avoir plusieurs entités qui peuvent être des hôtels, parking, Bureau ...
- Équipement : mesurer et remonter des données à SMATI à l'aide des capteurs, dans une entité on peut avoir un équipement ou plusieurs.
- KPI : Indicateur clé de performance, il est utilisé dans SMATI pour proposer des services au client en se basant sur ses données. Par exemple, à partir des données de consommation on peut créer un KPI qui calcule la consommation par personne dans un hôtel.
- Usage : correspond à un point de mesure.

a. Les scripts « data_quality »

Ma première tâche consiste à adapter le script « monitoring_report » avec les changements que l'équipe de développement ont faits. Les outputs de ce script est un fichier .pdf avec le rapport complet.

Le rapport se présente sous la forme d'un tableau résumant les métadonnées de chaque équipement ainsi que la date de dernière transmission, la fréquence de transmission et une mesure de l'état de transmission permettant de catégoriser chaque usage en quatre catégories:

- Fonctionnement normal (Vert) : Dernière valeur transmise
- Fonctionnement perturbé (Orange)
- Équipement en coupure de transmission (Rouge)
- Équipement n'ayant jamais transmis (Violet) : Aucune données transmise

Fréquence de transmission divisée par la dernière donnée reçue.

Ce rapport permet :

- La vérification de l'état de transmission instantané d'un ensemble d'usages
- La comparaison l'évolution de l'état de transmission d'un équipement au cours d'une période à fixer.

Un exemple de rapport « monitoring_report » :

Raw : donnée brute

Processed : donnée traitée

Rapport de surveillance de l'état des dernières transmissions

Ce rapport a été établi le : 02/27/19 11:22:43

Entité(s) : ['kpi-data-2']

Nombre total de dispositifs vérifiés: 805

Database : kpi-data-2

Ce rapport résume l'état instantané de transmission des capteurs

Transmission normale
Fonctionnement perturbé
Arrêt de transmission
Aucune transmission

Tableau des dispositifs examinés:

parent	entity	name	freq	last_date	transmission_quality
H2057 MERCURE-LYON SAXE LAFAYETTE	Indicateurs	ENAC	1440	02/26/19 00:00:00	Transmission normale
H6188 MERCURE-PARIS BOULOGNE	Indicateurs	ENCA	1440	02/26/19 00:00:00	Transmission normale
H0414 NOVOTEL-NICE CAP 3000	Indicateurs	DHAL	1440	02/26/19 00:00:00	Transmission normale
H6188 MERCURE-PARIS BOULOGNE	Indicateurs	ENCW	1440	02/26/19 00:00:00	Transmission normale
H2057 MERCURE-LYON SAXE LAFAYETTE	Indicateurs	ENCA	1440	02/26/19 00:00:00	Transmission normale
H6188 MERCURE-PARIS BOULOGNE	Indicateurs	ELAC	1440	02/26/19 00:00:00	Transmission normale
H2057 MERCURE-LYON SAXE LAFAYETTE	Indicateurs	DHAC	1440	02/25/19 00:00:00	Fonctionnement perturbé
H2057 MERCURE-LYON SAXE LAFAYETTE	Indicateurs	DHCA	1440	02/25/19 00:00:00	Fonctionnement perturbé
H0356 IBIS-BORDEAUX LE LAC	Indicateurs	DHAL	1440	02/25/19 00:00:00	Fonctionnement perturbé
H0736 NOVOTEL-LYON GERLAND	Indicateurs	DHAL	1440	02/25/19 00:00:00	Fonctionnement perturbé
H2057 MERCURE-LYON SAXE LAFAYETTE	Indicateurs	DHAL	1440	02/25/19 00:00:00	Fonctionnement perturbé
H2597 IBIS BUDGET-PARIS PORTE DE MONTMARTRE	Indicateurs	DHAL	1440	02/25/19 00:00:00	Fonctionnement perturbé

RAPPORT DE STAGE «DATA ANALYST»

H1409 IBIS-PARIS VERSAILLES CHATEAU	Indicateurs DHAL	1440	02/21/19 00:00:00	
H1409 IBIS-PARIS VERSAILLES CHATEAU	Indicateurs DHAL/S10	1440	02/21/19 00:00:00	
H3296 IBIS-PARIS LA DEFENSE COURBEVOIE	Indicateurs ELAL/S10	1440	01/31/19 00:00:00	
H0356 IBIS-BORDEAUX LE LAC	Indicateurs FUAL/S6	1440	01/31/19 00:00:00	
H0356 IBIS-BORDEAUX LE LAC	Indicateurs FUAL/S10	1440	01/31/19 00:00:00	
H0356 IBIS-BORDEAUX LE LAC	Indicateurs ENAL	1440	01/31/19 00:00:00	
I0557 A&L MULTI-MARQUE-INSTITUT QUIBERON INS	Indicateurs WAAL/S10	1440	None	
H2217 POC MERCURE PARIS GARE DE LYON	Indicateurs WAAL	1440	None	
H2217 POC MERCURE PARIS GARE DE LYON	Indicateurs ENAL/S6	1440	None	
H2217 POC MERCURE PARIS GARE DE LYON	Indicateurs WACA	1440	None	
H2217 POC MERCURE PARIS GARE DE LYON	Indicateurs WACA/S10	1440	None	
H2217 POC MERCURE PARIS GARE DE LYON	Indicateurs WACA/S14+S153	1440	None	

Nombre total de dispositifs vérifiés: 805

Nombre de dispositifs en état de fonctionnement normal : 26

Nombre de dispositifs avec un fonctionnement perturbé : 60

Nombre de dispositifs en arrêt de transmission: 685

Nombre de dispositifs n'ayant jamais transmis: 34

J'ai également modifié le scripte « usage_diagnostic ». Les outputs de ce script est un fichier .pdf avec le rapport complet.

Ce rapport se présente sous la forme d'un fichier détaillant les informations concernant les métadonnées, la transmission, les valeurs transmises par un usage et les potentiels traitements spécifiques à certains usages type.

Les informations sur la transmission concernent :

- La répartition des périodes de transmission de la donnée par le capteur
- La période de transmission théorique enregistrée dans le système
- L'estimation d'une période réelle de transmission au vu des données reçues
- La détection de "trous" de transmission

Les informations concernant les valeurs sont :

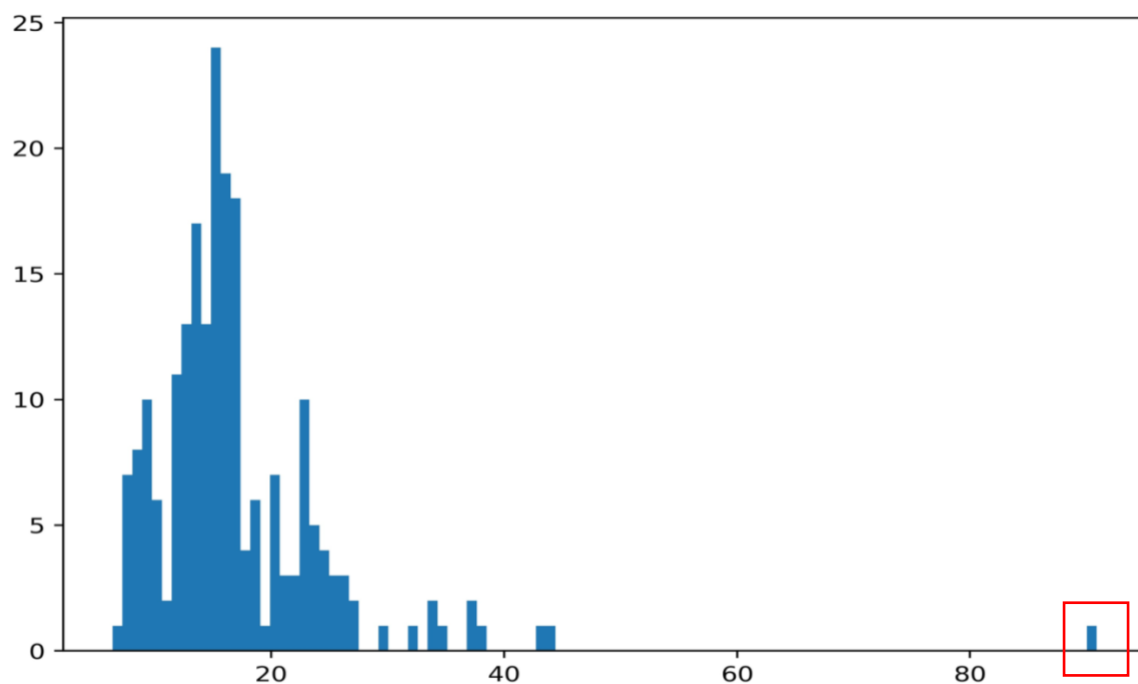
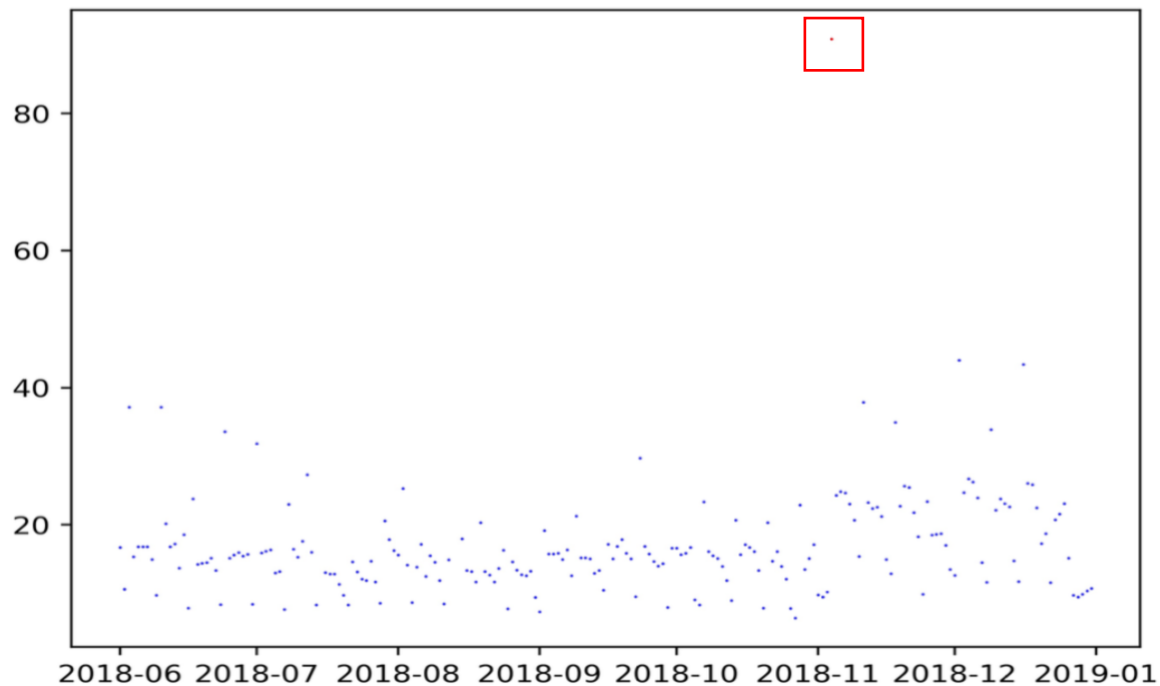
- La répartition des valeurs des données
- La détection des données aberrantes

****Une valeur aberrante est une valeur extrême, anormalement différente de la distribution d'une variable. En d'autres termes, la valeur de cette observation diffère grandement des autres valeurs de la même variable.**

Il est important de détecter les valeurs aberrantes, car ils influencent sur certains paramètres statistiques, comme la moyenne. Cela peut fausser notre compréhension du jeu de données et nous conduire à émettre des hypothèses erronées sur ce dernier.

La détection des outliers (valeurs aberrantes) nous permettra de faire des suppositions plus aguerries.

Les figures ci-dessous détectent des données aberrantes :



Les traitements spécifiques aux usage_type sont :

- La détection d'index décroissants

**Un index est une série de chiffres qu'on peut retrouver sur le compteur d'électricité. L'index se lit en kWh (Kilowattheure) et permet de mesurer la consommation d'électricité en indiquant avec précision combien de kWh d'électricité ont été consommés à date.

- La détection de consommations négatives

Un exemple de rapport « usage_diagnostic » :

Report for KPI: ENAL/S10

This report was done on: 19/02/2019 18:10:14

Description:

Entity: Indicateurs

Parent: H5240 SUITE NOVOTEL-LILLE EUROPE

Linked devices: []

Transmitting frequency: 1440

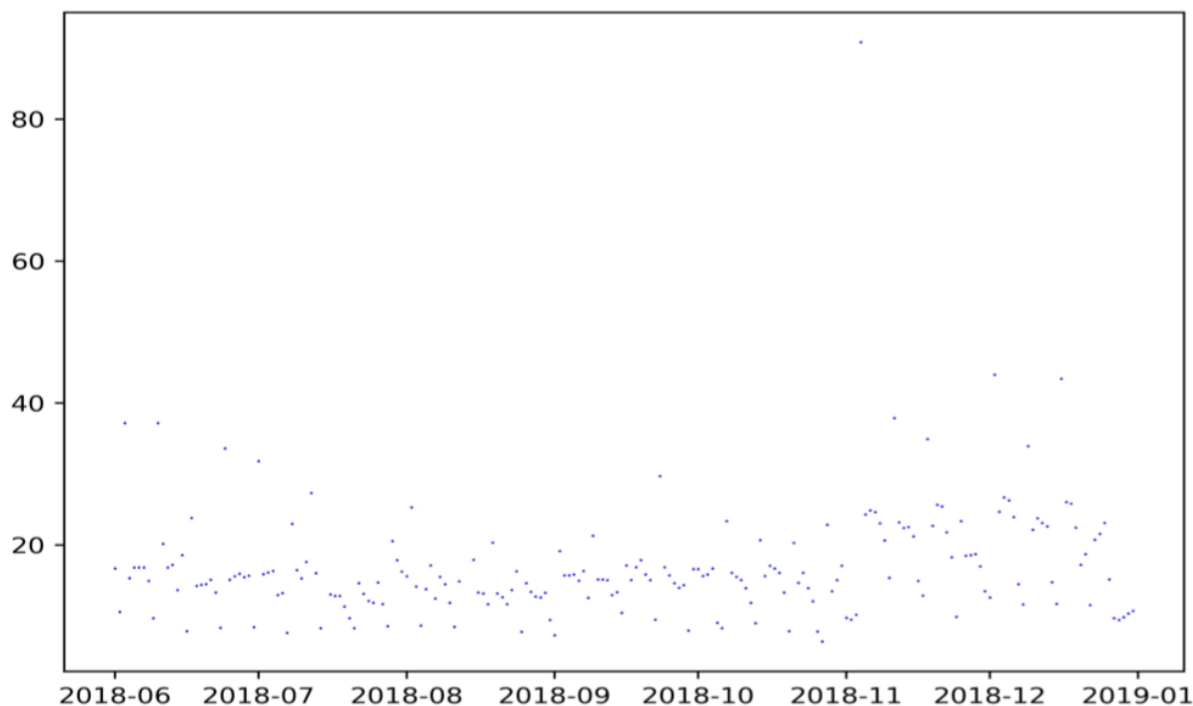
Last computation: 31/12/2018 00:00:00

Database: kpi-data

Time analysis

First document computed date: 01/06/2018 00:00:00.

Last document computed date: 31/12/2018 00:00:00.



2 gaps were detected.

List of the 5 largest computation gaps

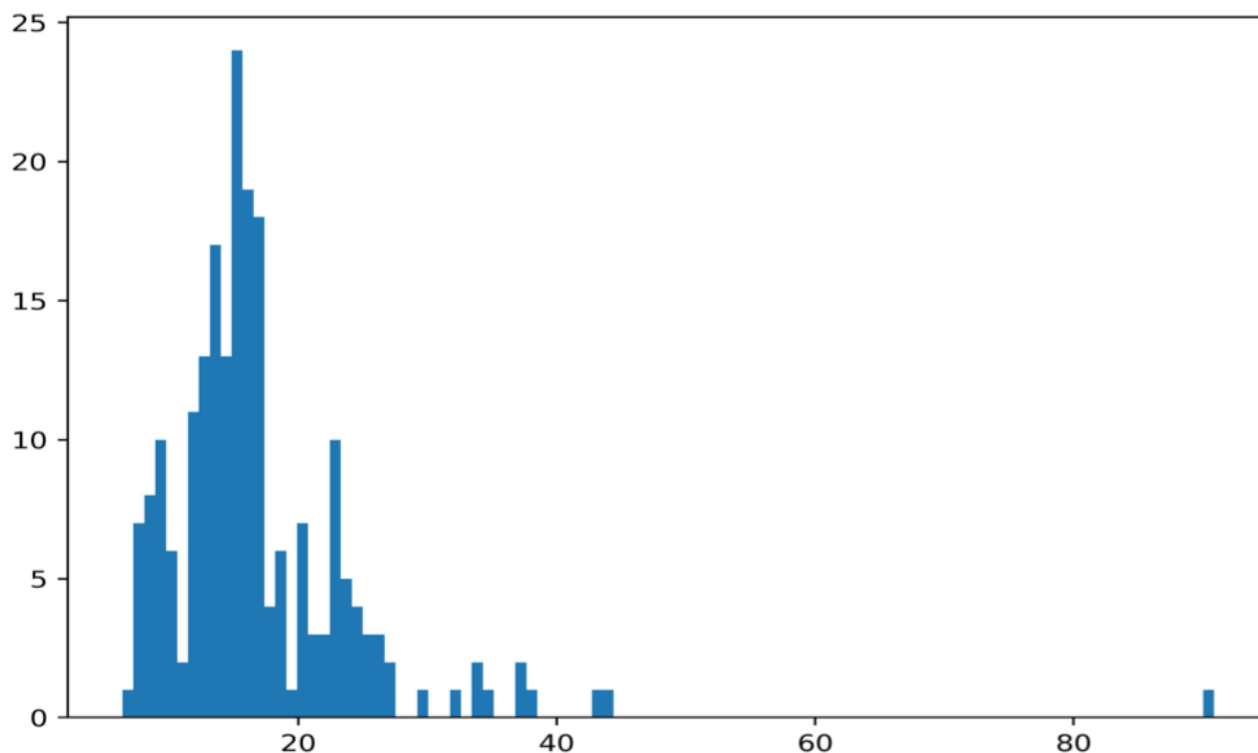
1. A transmission gap of 4320.0 minutes (or 3.0 jrs 0.0 min) was observed from 12/08/2018 00:00:00 to 15/08/2018 00:00:00
2. A transmission gap of 2880.0 minutes (or 2.0 jrs 0.0 min) was observed from 14/07/2018 00:00:00 to 16/07/2018 00:00:00

Values analysis

The highest value received from the device is: 90.86428571428571

The lowest value received from the device is: 6.387352941176471

Histogram of the values

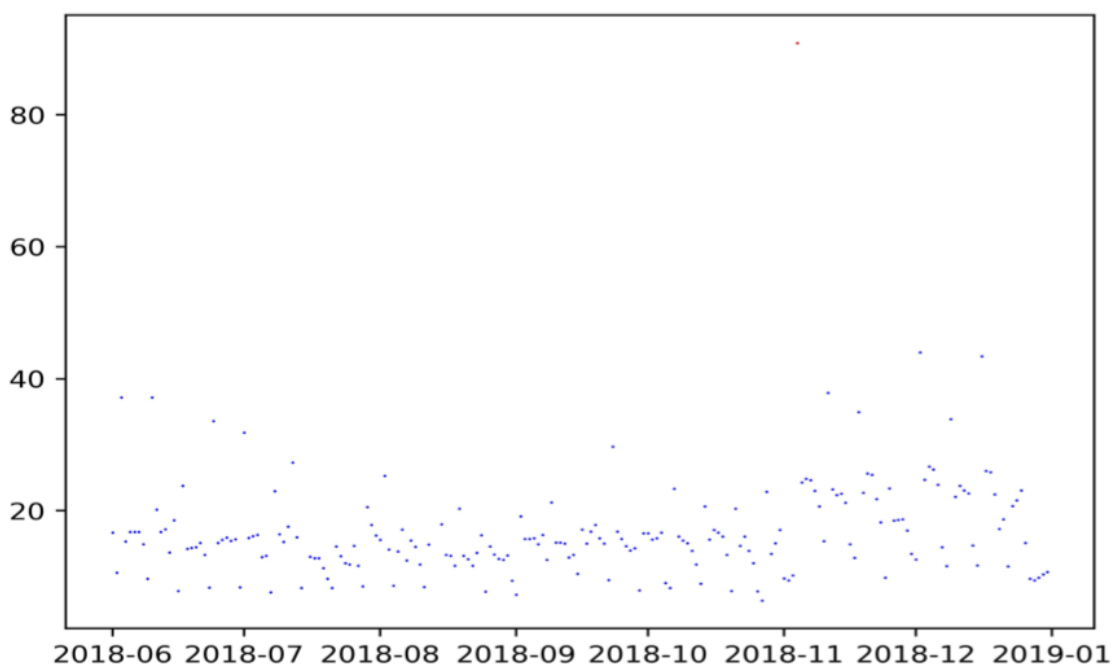


1 outlier was detected

List of the 5 largest outlier values

1. An outlier was detected on 04/11/2018 00:00:00+0100 with a value of 90.86428571428571

Plot of the outliers



Proposed warning thresholds

The proposed maximum value is: 44.00013513513513

The proposed minimum value is: 6.387352941176471

List of negative consumptions

No negative consumption was detected

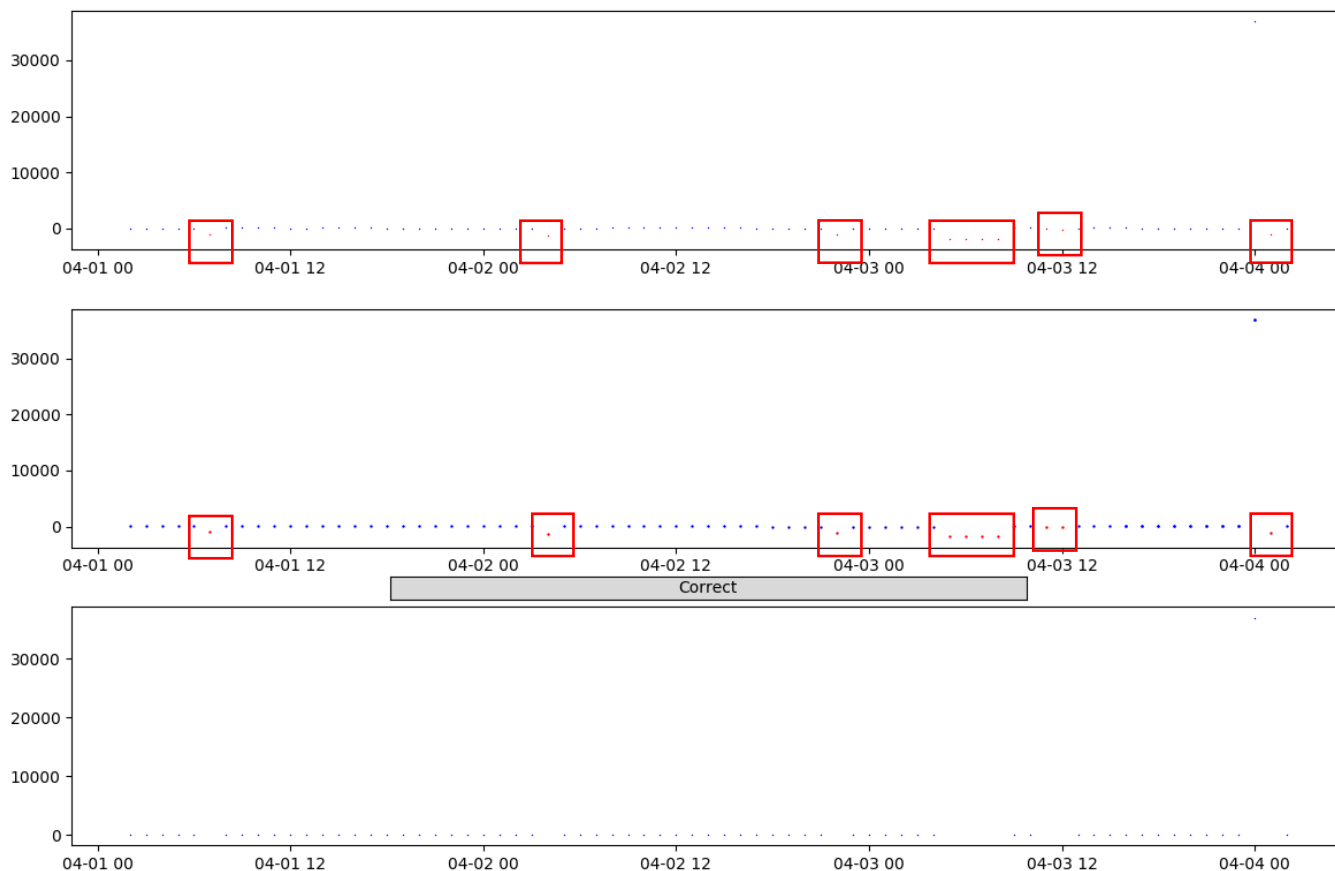
Ensuite, j'ai modifié le scripte « data_cleansing ». Ce script, appliqué à un usage, permet de détecter les données à supprimer en fonction du type de traitement souhaité et de les afficher à l'utilisateur. Ce dernier a ensuite le choix d'appliquer le traitement en supprimant les données correspondantes sur la base Cloudant grâce à une interface en cliquant sur le bouton 'correct'.

L'interface affiche 3 rangées de figures. La première montre l'ensemble du signal avec tous les points de données tels qu'ils sont stockés sur la base Cloudant. Sur la deuxième rangée, se trouvent les défauts détectés par l'algorithme. Les données détectées par le traitement réalisé sont affichées en rouge et un zoom est exécuté sur le signal autour de ces valeurs. La 3e rangée montre le signal lorsque l'ensemble des données classifiées comme défauts sont supprimées.

Différents traitements sont possibles :

- Suppression des valeurs décroissantes pour les index,
- Suppression des consommations négatives,
- Seuillage,
- Suppression des données aberrantes – avec différentes techniques de suppression d'outliers possibles.

La figure ci-dessous supprime les consommations négatives :



b. Application Django « data_reporting »

Au fur et à mesure de l'avancement de mon stage ma mission s'est orientée vers une partie importante du produit, qui est la création d'une application Django qui consiste à créer des rapports énergétiques ayant pour cible les clients d'Hxperience. Ces rapports analysent les données temporelles des compteurs électriques et de gaz et les affichent sur différents graphes.

Un exemple de rapport énergétique :

Rapport énergétique

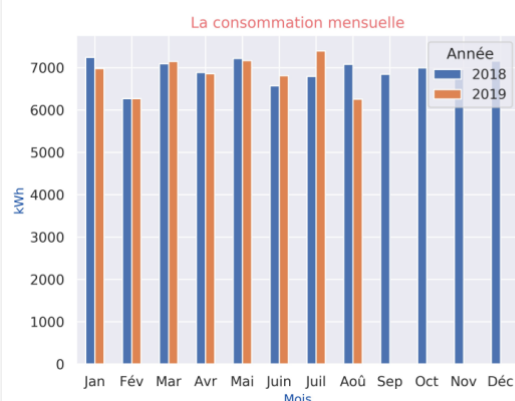
Ce rapport a été effectué le: 09/10/19 10:14:29

Entité: Hxperience

Consommations d'énergies

La somme des consommations mensuelles des compteurs suivants :

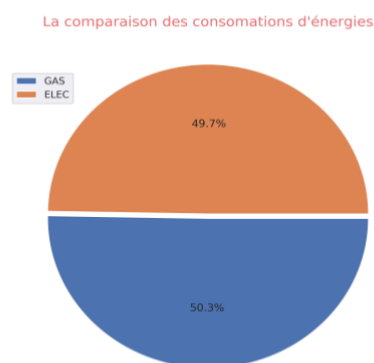
- Compteur Gaz HX Lourmel
- Compteur ELEC HX Lourmel



La comparaison des consommations des compteurs d'année en cours:

- Compteur Gaz HX Lourmel
- Compteur ELEC HX Lourmel

La consommation total des compteurs : 54889 kWh



Consommations de Gaz

La consommation des compteurs en kWh :

- Compteur Gaz HX Lourmel

	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre	Total
2018	3503	3099	3502	3449	3549	3222	3461	3439	3285	3325	3310	3558	40702
2019	3423	3149	3600	3409	3654	3474	3776	3115					27600
Moyenne	3463	3124	3551	3429	3601.5	3348	3618.5	3277	3285	3325	3310	3558	34151

Consommations électriques

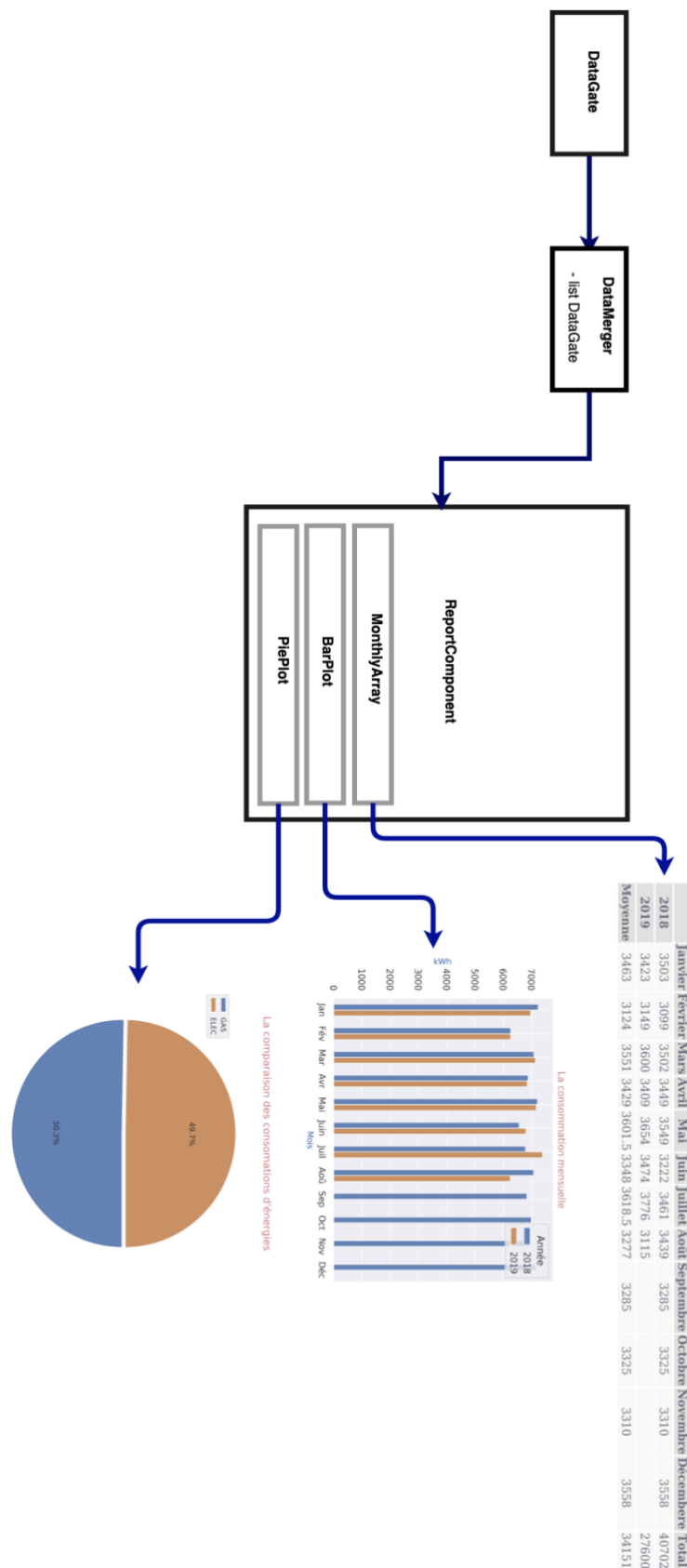
La consommation des compteurs en kWh :

- Compteur ELEC HX Lourmel

	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre	Total
2018	3741	3170	3591	3438	3670	3350	3332	3643	3560	3670	3648	3600	42413
2019	3559	3119	3547	3448	3517	3336	3620	3143					27289
Moyenne	3650	3144.5	3569	3443	3593.5	3343	3476	3393	3560	3670	3648	3600	34851

RAPPORT DE STAGE «DATA ANALYST»

La figure ci-dessous représente l'architecture de l'application « data_reporting »



DataGate : classe pour obtenir les données traitées agrégées pour une source unique (entité)

DataMerger : classe pour obtenir des données agrégées à partir de plusieurs sources (en utilisant liste de DataGate)

ReportComponent : classe pour obtenir un graphique des divers composants
MonthlyArray, **BarPlot** et **PiePlot** sont des classes qui héritent de la classe ReportComponent

Retour personne sur le stage

Le stage que j'ai réalisé chez Hxperience, m'a aidée à monter en compétences. Et cela est dû à mon travail pendant 6 mois sur plusieurs nouvelles technologies, que je n'avais pas eu l'occasion de voir en cours.

Satisfaite de la quantité du travail que j'ai fournie en 6 mois, le directeur de l'entreprise décide de me proposer un autre stage, que j'ai accepté. En effet, je voulais continuer dans le domaine de la data qui m'intéresse.

Conclusion

Je suis très satisfaite de mon stage au sein d'Hxperience. Les missions qui m'ont été confiées étaient très intéressantes et je suis heureuse de pouvoir les poursuivre. Dès le début j'ai trouvé le sujet de stage très attractif et en lien avec mon cursus et mes aspirations. En effet, les projets de l'entreprise s'inscrivent dans le domaine du bâtiment connecté ("Smart building").

De plus, je pense que le fait d'avoir effectué ce stage dans une startup m'a donné l'opportunité de témoigner du processus complet de l'implémentation d'un projet. Ceci m'a aidée à découvrir les bases et tous les détails du projet et particulièrement en travaillant avec des ingénieurs expérimentés. De même, l'environnement d'une startup est exceptionnel d'un point de vue social ce qui m'a donné la chance de bien connaître mes supérieurs et mes collègues.

Durant la période de stage, j'ai compris la différence entre l'environnement du travail d'une entreprise et celui de l'université. Pour l'instant, j'ai bien compris le déroulement du travail au sein des entreprises. De plus, j'ai bien assimilé la hiérarchie et son importance. Je peux noter aussi l'amélioration de mes compétences fonctionnelles grâce à nos discussions et nos réunions dans tout ce qui concerne la démarche du travail.

Par ailleurs, le stage m'a beaucoup aidée à développer et acquérir des compétences techniques en « python » et le Framework « Django ». Je me suis aussi améliorée, au niveau relationnel, grâce à la communication et au partage avec les membres de l'équipe et notamment ma participation à plusieurs réunions. De plus j'ai acquis plusieurs compétences en rapport avec l'organisation, la qualité du code écrit et notamment les bonnes pratiques pour réaliser un projet en équipe. Ces valeurs ajoutées participent à mon ouverture dans le monde du travail.

Ce stage de 6 mois, s'est déroulé dans de très bonnes conditions. Ce stage m'a lancé dans le monde du travail mais m'a surtout permis de me voir concrètement travailler dans un domaine qui me plaît.

Bibliographie

<https://www.element14.com/community/groups/internet-of-things/blog/2017/02/17/iot-with-python-essential-packages>

<https://www.element14.com/community/groups/internet-of-things/blog/2017/02/17/iot-with-python-essential-packages>

<http://apprendre-python.com/page-django-query-set-queryset-manager>

<https://hxperience.com/>

https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Accueil_principal

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

<http://objenious.com/>

<http://www.wi6labs.com/>

<https://www.zbre.io/>

Acronymes

Smart Building (ou Bâtiment Intelligent)

Il s'agit de bâtiments qui contiennent des systèmes intelligents. Cela induit la mise en place de capteurs et/ou d'actionneurs qui permettent la gestion automatisée des différents équipements du bâtiment. Cela induit également la mise en place de réseau de communication intra bâtiment. Le Smart Building a plusieurs objectifs :

- Offrir des nouveaux services aux utilisateurs ;
- Améliorer la gestion de l'énergie ;
- Réduire la consommation de gaz, eau, électricité ;
- Améliorer le confort des usagers ;
- Faciliter le travail des équipes de maintenance.

IoT

L'IoT, pour Internet of Things, se traduit par l'Internet des objets, ou plus communément, les objets connectés. Ces derniers sont capables d'émettre des données grâce à des capteurs. On peut prendre en exemple les nouveaux modèles de voitures équipés de capteurs qui alertent le conducteur en cas de pression pneumatique trop faible. L'intérêt étant que ces objets soient interconnectés afin d'apporter de la valeur et des solutions aux données collectées.

API (Application Programming Interface)

L'API est une interface de programmation applicative permettant aux applications de communiquer entre elles et de s'échanger des services et/ou des données. Ces interfaces de programmation d'application sont déjà utilisées dans de nombreux domaines tels que le BTP, l'informatique ou le marketing digital. La tendance tend vers des Open API car depuis quelques années, on observe un fort développement des applications mobiles, un développement d'usage B2B2C, mais également un important développement des applications en temps réel. Ainsi, les objectifs des nouvelles API sont donc de s'ouvrir aux nouvelles organisations, que ce soit en interne ou vers l'externe. L'intérêt étant d'ouvrir les infrastructures vers les nombreux canaux digitaux et les différents écrans (mobiles, tablettes...).

Cloud

De plus en plus utilisé par les entreprises pour éviter la gestion et le coût d'une infrastructure informatique, le Cloud est un système de stockage à distance. Pour y accéder, les usagers doivent s'y connecter via une liaison internet sécurisée. Ainsi, ils peuvent avoir accès à leurs données de n'importe quel endroit et sur n'importe quel support : mobile, ordinateur fixe ou portable, tablette, et même certaines applications. En résumé, il s'agit d'un réseau de stockage partagé.

SaaS (Software as a Service)

Le Software as a service (SaaS) est une catégorie de Cloud Computing. Il permet de se connecter à des applications cloud et de les utiliser via une connexion Internet. En pratique, des applications sont mises à disposition à distance par un fournisseur qui les héberge. Les applications de messagerie en ligne en sont le parfait exemple.

LoRa ou LoraWAN

LoRa (Long Range) se traduit par « longue portée » en français. Il s'agit d'un réseau IoT qui permet de faire circuler les données récoltées par les capteurs des objets connectés via les fréquences radio libre 868 MHz et Internet. Peu gourmand en énergie, ce réseau s'intègre dans le protocole réseau ouvert LoRaWan qui permet à toutes les entreprises de déployer leur propre réseau.

Sigfox

Concurrent direct du réseau LoRa, Sigfox est aussi un réseau IoT. Il s'agit plus précisément d'un réseau longue portée et à bas débit. Il permet la transmission d'informations et de données émanant des objets connectés via une technologie radio Ultra narrow band. L'objectif étant de faire baisser la consommation d'énergie des périphériques connectés ainsi que leur prix.

Annexes

Partenaires de Hxperience:



