Problématique

Le problème de stationnement est complexe car elle est le reflet d'un conflit de différent usager d'un même espace public .Cet espace est rare car il est souvent confiné dans des zones denses telles que les zones d'aéroport , des grands supermarchés, Avenu ... Il a toujours été mis à la disposition des parking où le client peut avoir une place parking et payer à la fin en fonction de la durée d’occupation de la place et la deuxième méthode de réservation est d'acheter des cartes de stationnement par laquelle le client peut acheter un espace pour une durée bien définie.  
Regardons les deux solutions maintenant, c'est des solution qui ne répondent pas aux problèmes de notre vie quotidienne, trouver une place de parking libre en temps réelle en ville à grenoble et surtout le week end devient une des choses presque impossible.

Ce service OuiDispo sera réalisé pour répondre à une grande problématique qui de trouver une place libre à proximité en temps réelle

### **Les problèmes de stationnement!**

* Les embouteillages rallongent inéluctablement le temps de trajet.

Certains immeubles ne disposent pas de parking pour les habitants.

* Le nombre de voitures se multiplie, ce qui entraîne de maintes difficultés de  
  se garer.

### **Problèmes de gestion**

* Difficulté de l'accès à l'information.
* indisponibilité de la mise à jour d'une information utile à terme:  
  \* perte de temps dans la recherche d'information.  
  \* Traitement assez lourd.  
  On remarque aussi que la plupart des tâches ne sont pas informatisées ou elles font recours à l'intervention manuelle exagérée du responsable.

Plan marketing

Notre projet consiste à créer une application mobile qui permettra à l'entreprise CO-voiturage de proposer un nouveau service de réservation et de location de courte durée de places de parking en temps réel, avec comme objectif de faciliter le stationnement ce qui va garantir une bonne mobilité dans les grandes villes de France.

la clientèle ciblé est à la fois une clientèle professionnelle on leur permettant de gérer facilement le stationnement des location temporaires pendant les rendez vous professionnels, et aussi les particuliers pour réserver dès place de parking dans les grande ville de france

L'objectif de l'application est de faciliter le stationnement dans les grandes villes de France pour les clients de l’entreprise ce qui est souvent un problème de trouver des places libres et en parallèle ça va faciliter pour les particulier qui veulent stationner à trouver une place libre en temps réel et aussi de répondre à des exigences de mobilités.

sur le plan organisationnel :

faire des partenariat avec les responsables des parkings pour permettre d’avoir toujour des disponibilité de place pour les clients de l’entreprise,

- mettre en avant le gant de l’utilisateur en terme de temps et d'argent et aussi la proximité géographique que l’application permet d’assuré

sur le plan technologique :

une application mobile qui sera à la portée de main , facile a y accéder, en tout moment et n'importe ou, avec une interface simple et très parlante pour l’utilisateur , avec une possibilité de reconnaissance vocale pour bien l'utiliser tout en conduisant.

Dès promotion du nouveau service seront mise en place pour les client et dans les période des fêtes et aussi pour tous les utilisateurs, des abonnement disponible avec une réduction qui dépend de la période d’abonnement.

La possibilité des feedbacks sera possible sur l’application pour pouvoire amélioré les remarque signalé et des promotions serron proposé régulièrement au person déjà utilisé l’application .

Analyse Jeux de données

* 1. Choix technologique

Pour accomplir notre mission, nous avons choisi d’utiliser les technologies suivantes:

* + 1. **Talend data integration :** cet outil nous a permis de filtrer et d'intégrer les données.
    2. **PhpMyAdmin :** pour créer une base de données et stocker nos données.
    3. **Talend Data Quality:** pour vérifier la qualité des données
    4. **Power BI:** pour afficher des données et créer des dashboards
    5. **Python**: Pour automatiser le flux de données, faire du data preprocessing
  1. Analyse de notre jeu de données

Nous avons commencé par sélectionner les colonnes pertinentes par rapport à notre problématique (la disponibilité de places dans les parkings). Ainsi les colonnes suivantes ont été retenues: (nom parking, adresse, lat & long, nbr\_places - nb\_pr, nb\_pmr, nb\_voitures\_electriques, nb\_autopartage?, hauteur\_max ?, tarifs(1h,2h,3h,4h, 24h)) ouverts à tous

Grâce à ces données nous pouvons avoir la localisation des parkings dans les grandes villes, et les tarifs par heure, et le nombre de places total ou par typologie.

Cependant l’étape de data preprocessing nous montre qu’il y a beaucoup de données manquantes et d’autres sur un format non conforme sur le dataset. Que faire avec ces données:

* Normaliser
* Supprimer : Fortement déconseillé
* Enquêter

Concernant notre problématique générale, il nous faut la possibilité d’avoir le nombre de places libres à temps réel, ce qui n’est pas le cas sur ce Jeu de données

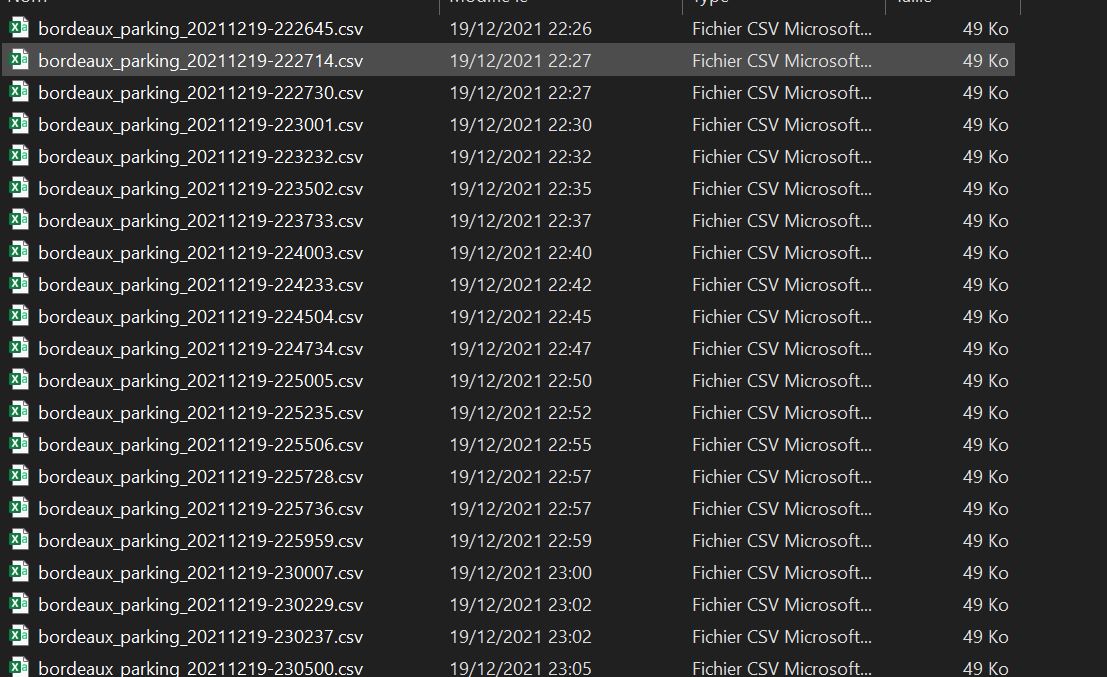
Acquisition d’un jeux de données complémentaire

* 1. Jeu de données Bordeaux

Ce jeu de données représente la localisation des principaux parkings payants ouverts au public et parcs relais présents sur le territoire de Bordeaux Métropole, qu'ils soient de gestion publique ou privée. Il fournit des données sur leurs caractéristiques administratives, techniques et tarifaires (gestionnaire, nombre et types de places, tarifs types).

Notre projet pilote sera donc mis en place sur la ville de Bordeaux, puis étendu par la suite à d’autres villes.

Le jeu de données étant rafraîchi toutes les 150 secondes, il faut alors mettre en place une solution de collecte automatisée de nos données. Nous avons optés pour l’option Python, en créant un script qui s'exécute à intervalle de temps réguliers pour récupérer la dernière version sur le site, et de faire l’historisation:



* 1. Analyse du Jeux de données de Bordeaux

Les colonnes suivantes nous seront utiles pour notre problématique:

**Geo Point :** donner géographique

**insee** : code commune

**nom**: nom du parking

**propr:** les propriétaires

**adresse :** adresse du parking

**etat** : exemple (libre, complet,ouvert , ferme, abonné)

**libres** : nombre de place libre en temps réel

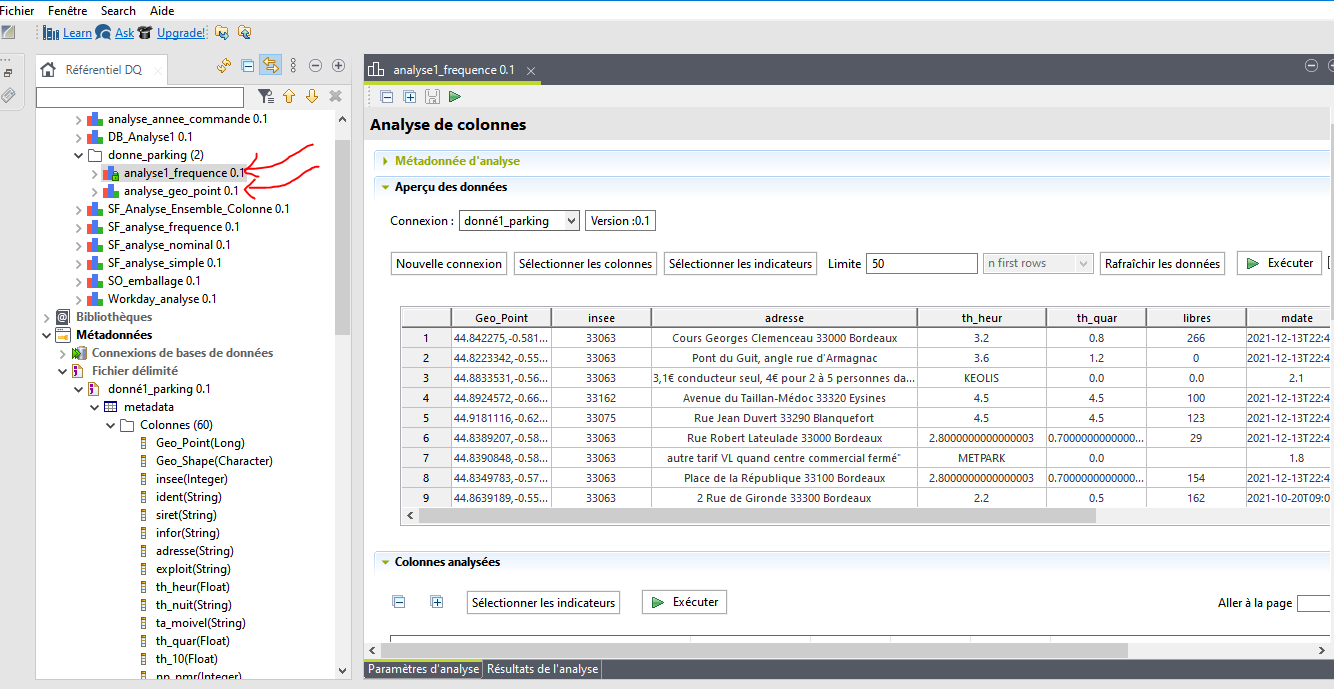
**np\_total**: nombre de place total du parking

**secteur**: secteur (centre , périphérie, hyper\_centre)

**mdate**: date de modification des données (donnée chargé chaque 2 min)

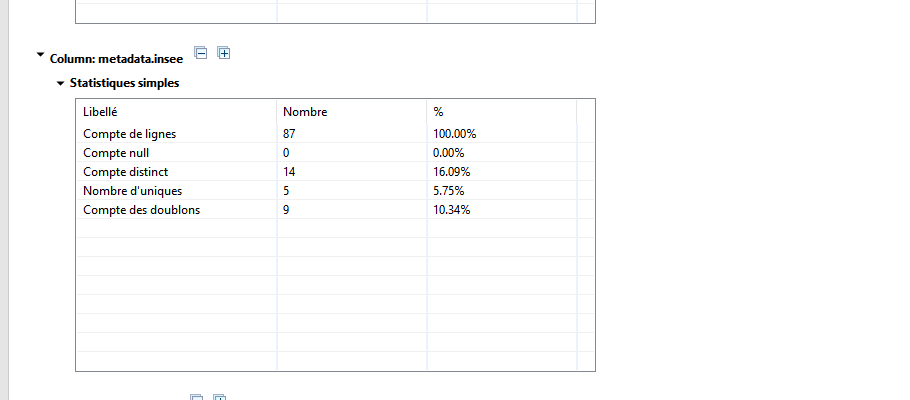
* 1. Data Quality

**Aperçu des données**

****

pour auditer nos donné on a créé de different analyse pour nos donné avec TALEND data QUALITY

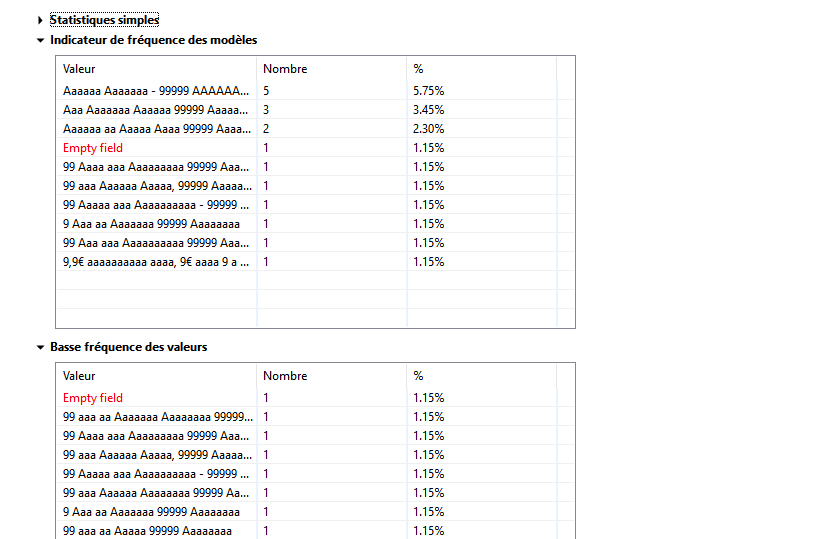
* analyse simple :



remarque: on constate que il ya 9 doublon

solution:

* analyse de fréquence de modèle:



remarque: il ya des modèles Empty field car ça convient pas au type

* 1. Partie ETL

En utilisant l' ETL TALEND Data Integration on va charger notre jeu de donné dans une base de donné ODS qui servira comme base donnée temporaire pour charger les informations de notre fichier CSV (donnée brute) et à partir de la base de donné ODS on va créer la base de donné de datawharhous qui va contenir notre table de fait et les table de dimension (on a choisit un modèle en étoile) qui sera utilisé pour la réalisation de notre dashboard.

* **Modélisation de notre datawarehouse :**

On a choisi le modèle en étoile pour modéliser le Datawherhous qui va contenir une table faitDisponibilité et 3 table Dimenssion :

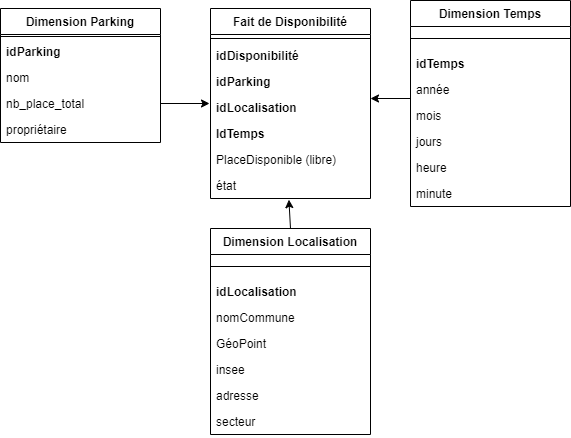
**DimensionTemps**(idTemps,année,mois,jours,heurs,minute)

**DimensionLocalisation**(idLocalisation, GéoPoint, insee, adresse, secteur)

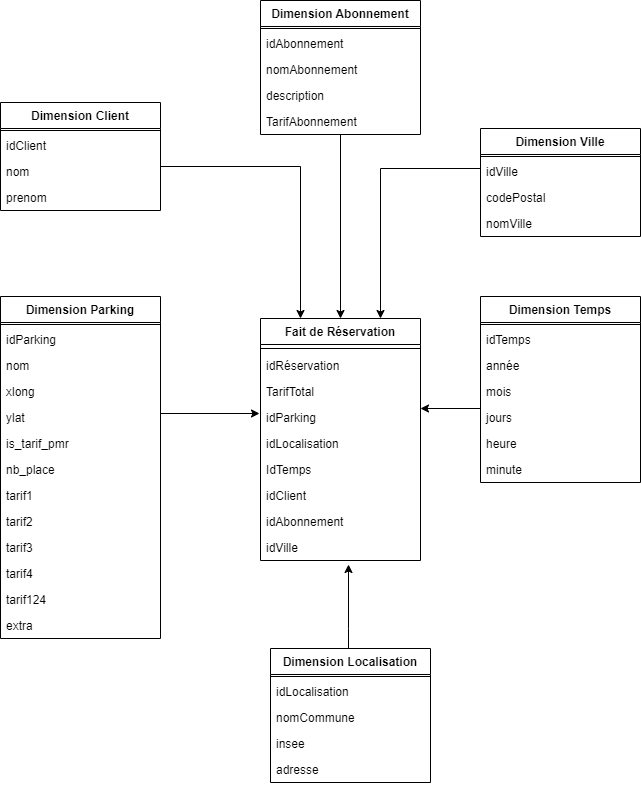
**DimentionParking**(idParking, nom, nb\_place\_total, propriétaire)

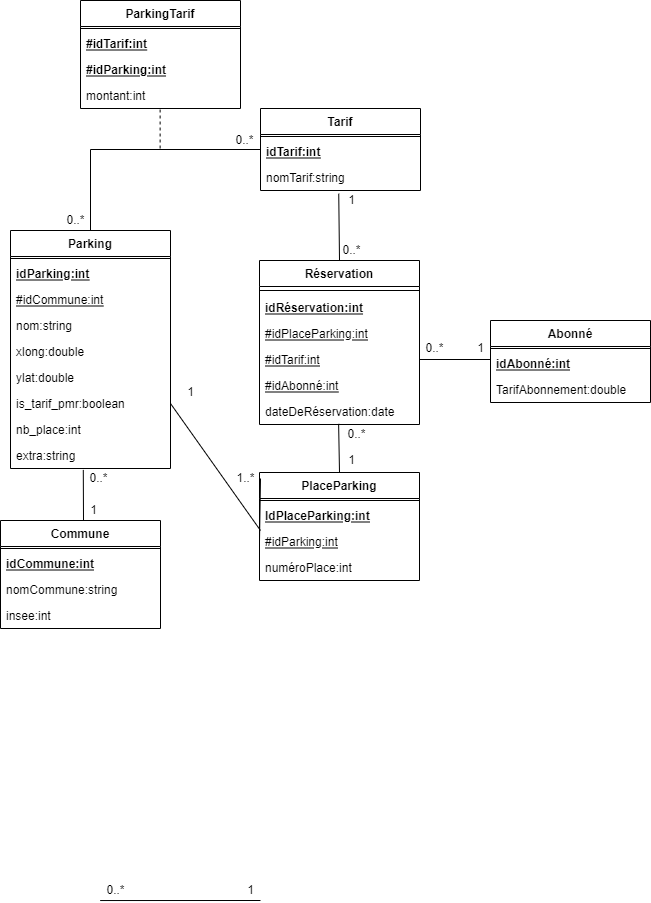
**FaitDisponibilité**(idDisponibilité, idParking, idLocalisation, idTemps, PlaceDisponiblie, état)

**le schéma du datawherhous**



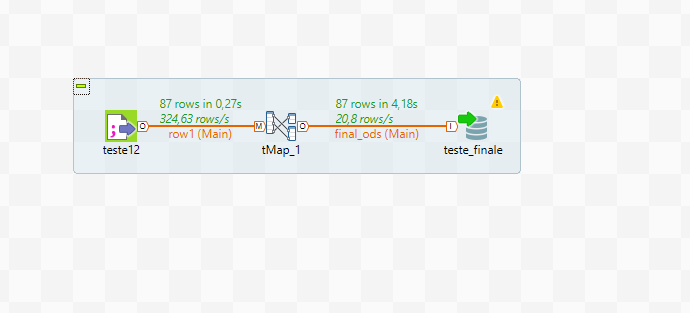
Notre service va évoluer vers ce datawarehouse et va comporter finalement ce diagramme de classe. Il prendra donc en compte La réservation et les abonnements:

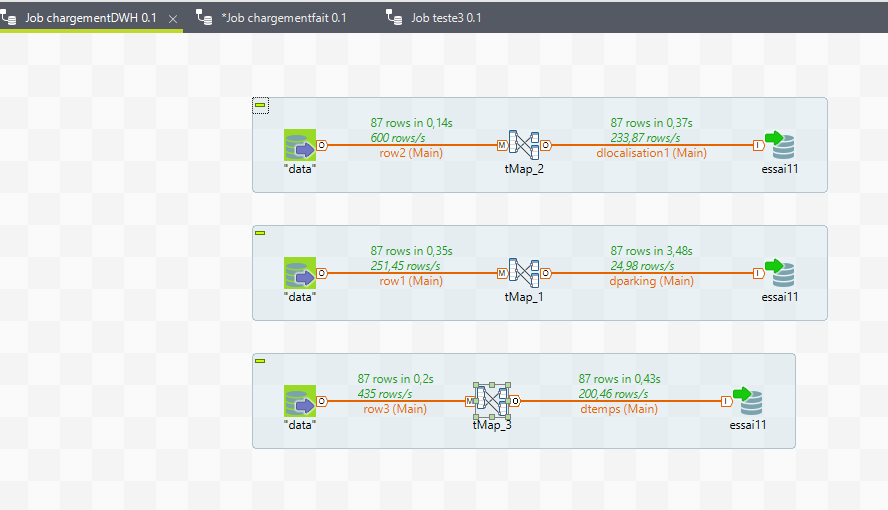


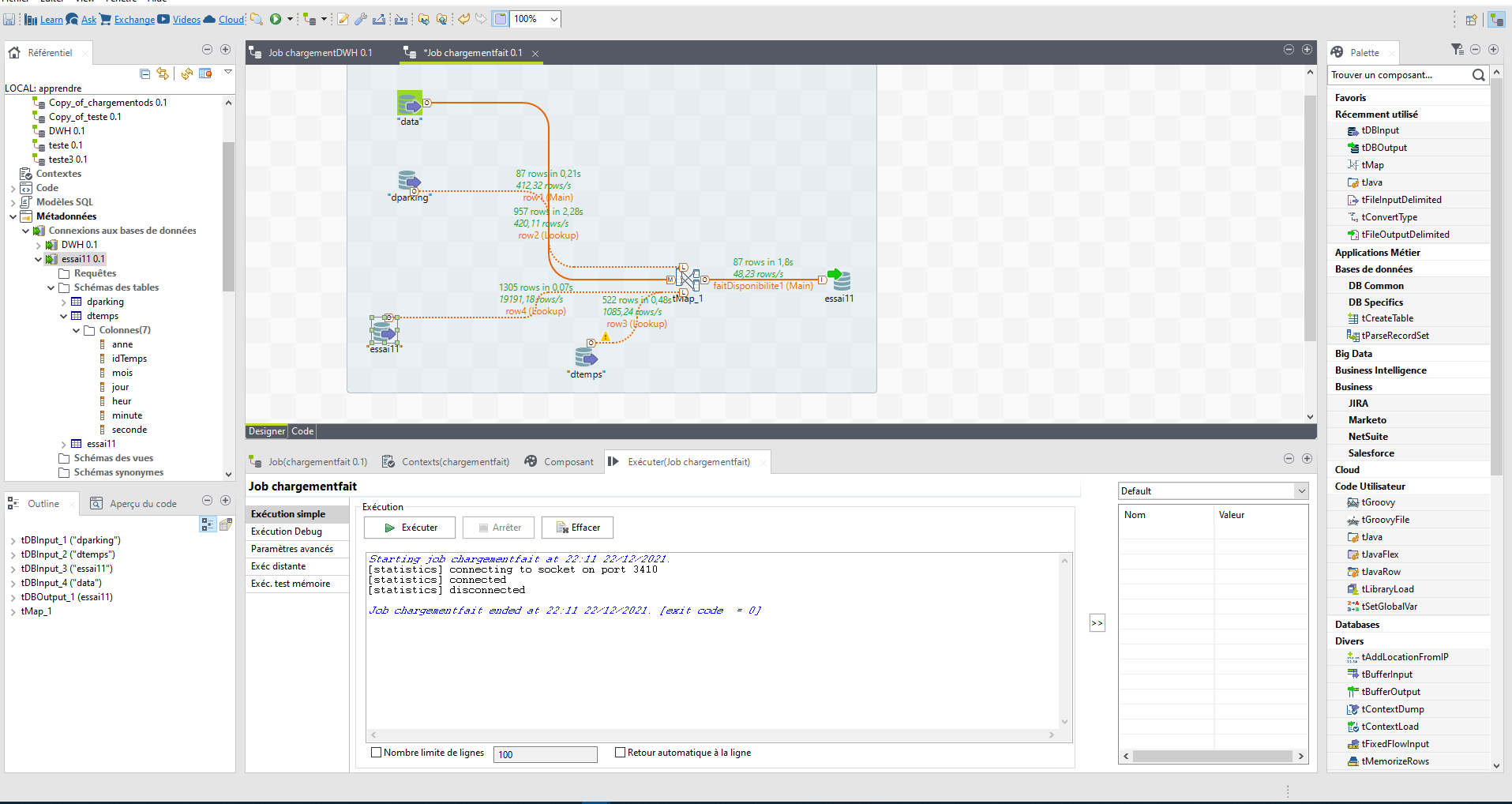


* **chargement de données et Création**

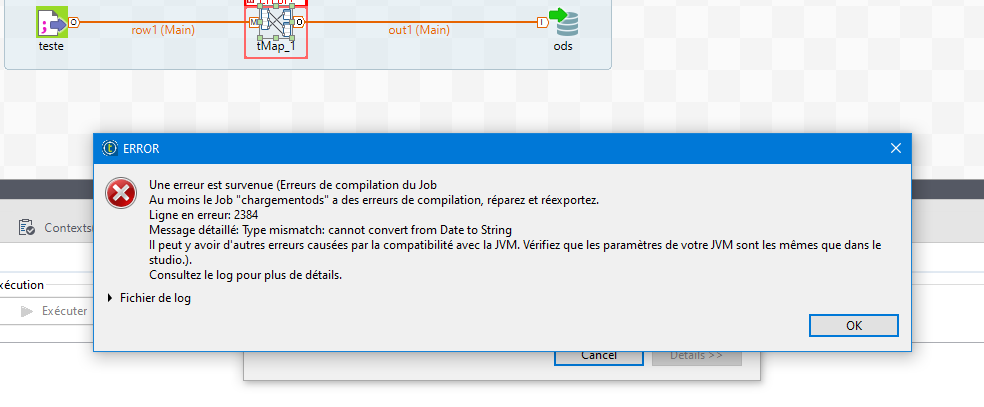
Chargement des données à partir de notre fichier csv qui contient les données brutes (teste12) vers la base de donnée temporaire ODS (teste\_finale)

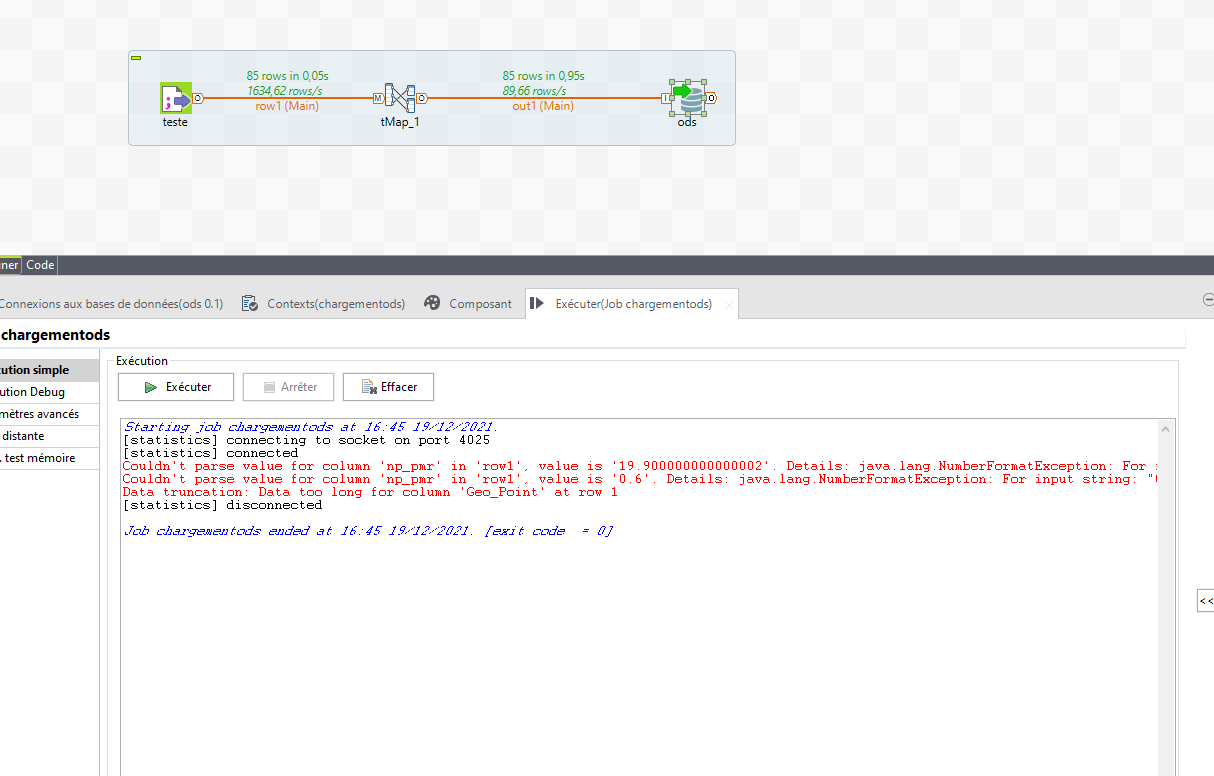
****

****

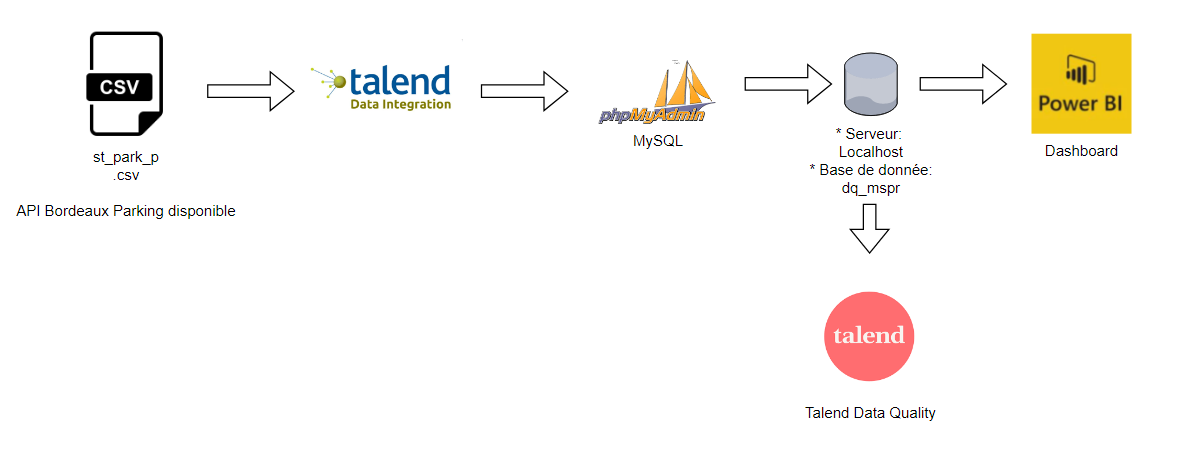
****

**les problème rencontré:**

****

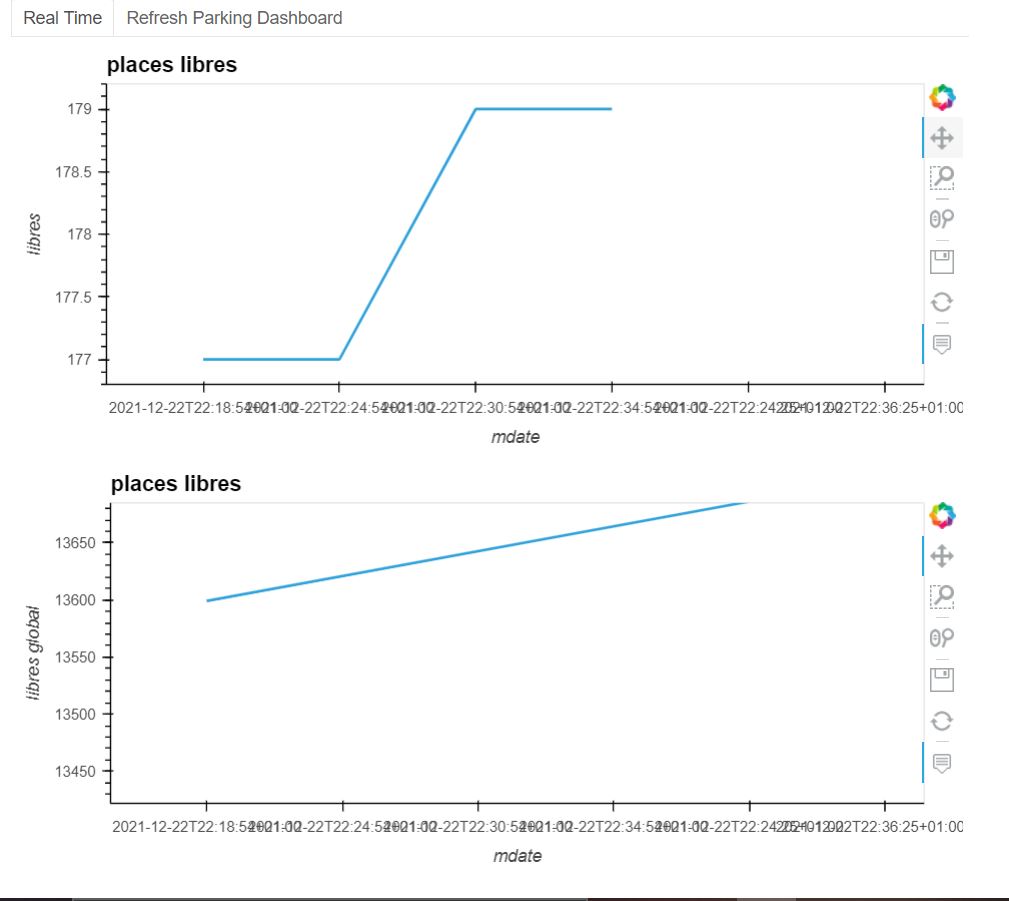
****

* 1. **Partie visualisation**

Schéma: Processus des flux

**-Visualisation à temps réelle:**

Grâce à la puissance de python, nous avons observé à temps réel l’évolution du nombre de place d’un parking, du nombre libre total sur bordeau et suivi à temps réels le nombre de place. Ci-joint le une vue du Dashboard à temps réelle:



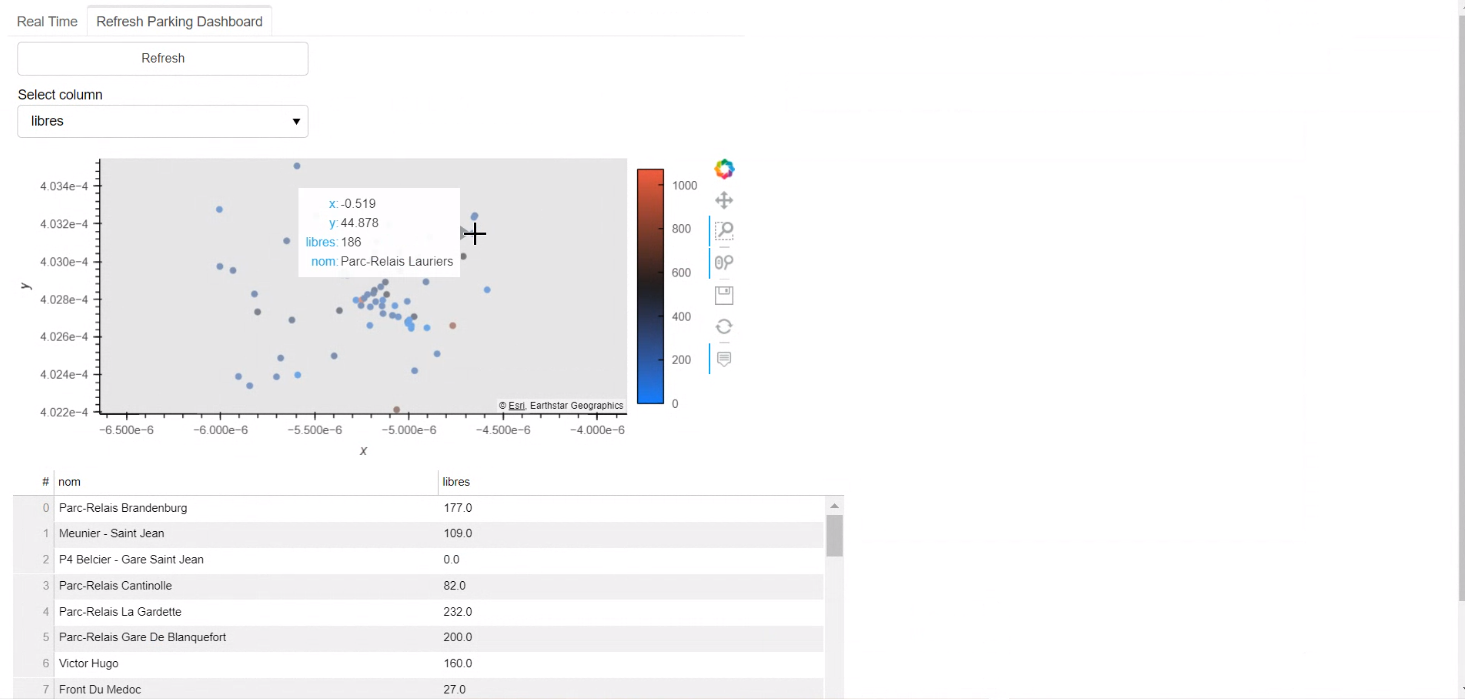


Schéma: Modélisation de la Base de donnée

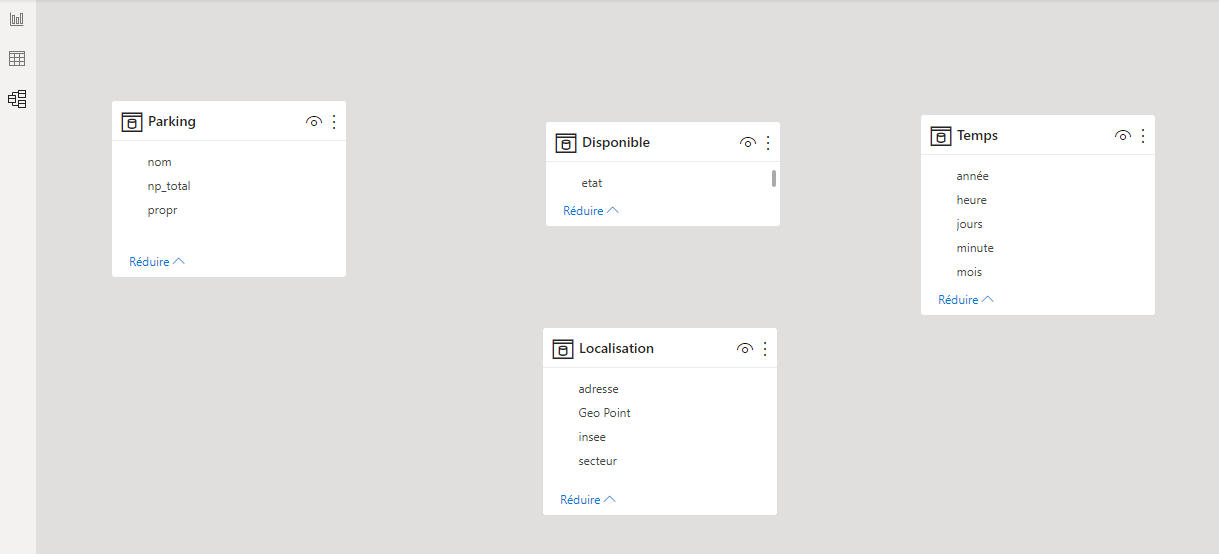


Schéma: Graphique montrant la moyenne des places libres par jour

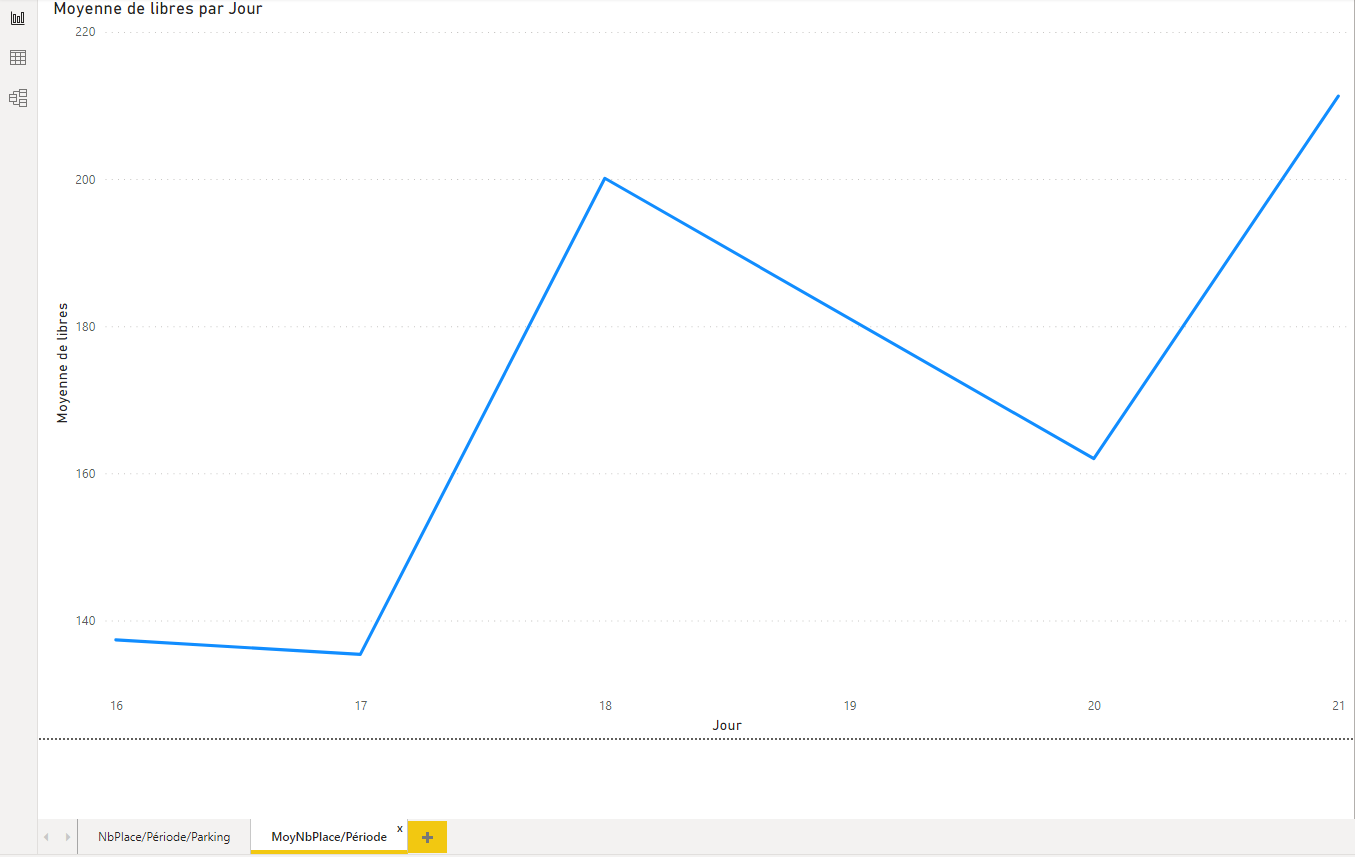
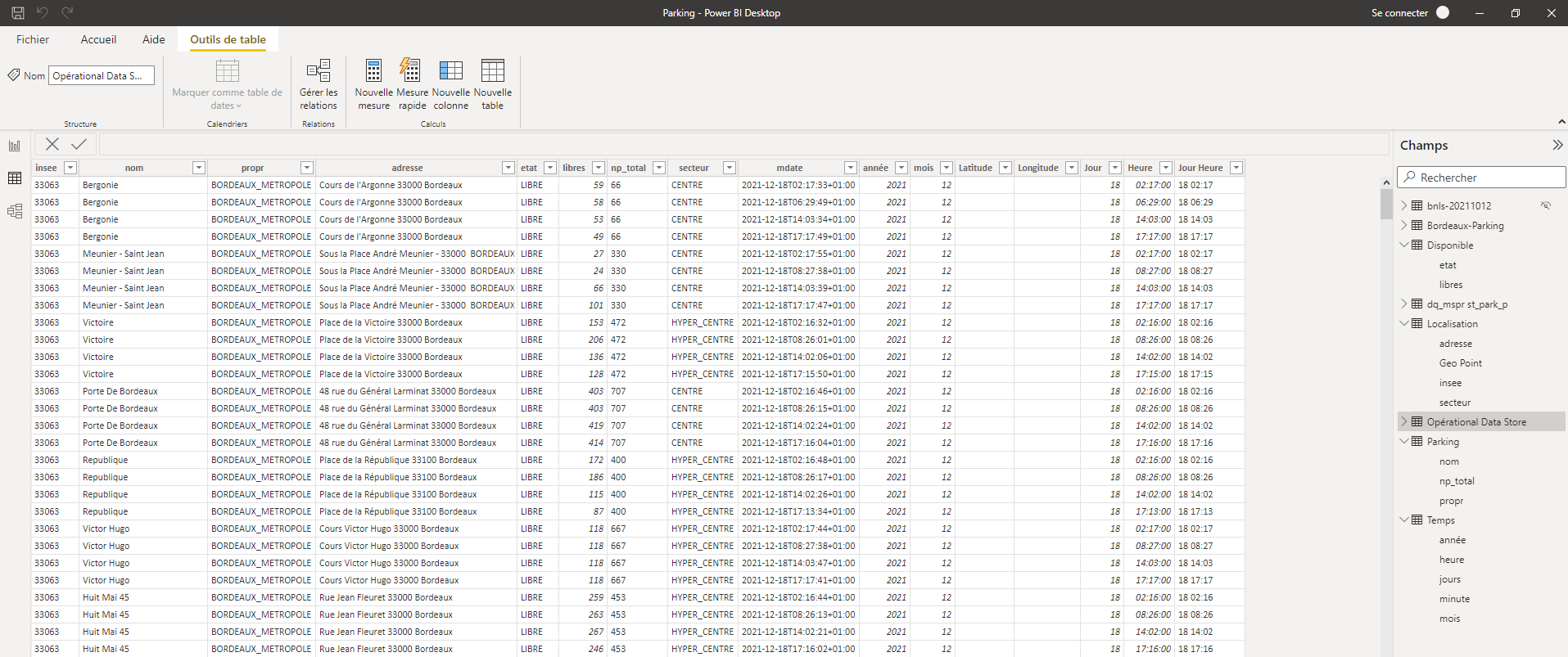
Schéma: Graphique montrant la moyenne des places libres par jour

Schéma: Graphique montrant les places libres par période pour un parking choisit



Schéma: Tableau montrant nos jeux de donnée



* 1. **Modèle statistique et Data Science**

L’équipe data devra faire face à un flux constant de données à traiter, et des méthodes de la data science seraient d’une grande utilité. Ces méthodes doivent permettre d’abord d’automatiser le processus de collecte et de traitement des données, mais aussi de prédire le comportement d’une ou de variable(s) donnée(s), et d’anticiper sur les opportunités.

1. Automatisation de la collecte de données

Notre jeu de données est accessible sur un serveur distant et actualisé à intervalle de 2 minutes 30s, l’automatisation permettrait de configurer un Cron qui exécute à intervalle de temps régulier, une collecte du jeu de données courant, fait les vérifications nécessaires et met à jour notre base de données en prenant le soin d’historiser. Cette tâche est possible grâce à un script python préalablement écrit par notre équipe et testé pour vérifier son intégrité.

1. Calcul d’une fréquence de rafraîchissement

Le temps réel peut être coûteux en ressource, donc il faut se poser la question sur la nécessité de tourner le script toutes les 150 secondes. La fréquence de rafraîchissement doit être représentative, celà ne servirait à rien de rafraîchir notre dataset toutes les 150 secondes si en général le nombre de place ne change pas à la même fréquence. Une analyse à l'œil nu permettrait de fixer une valeur qui peut être représentative, mais au vu de la quantité de données, celà peut vite être éprouvant. La méthode consisterait à trouver une fonction linéaire proche de la variation du nombre de place libre en fonction du temps, x, représentant le coefficient de la fonction linéaire. Cette méthode peut être utilisée sur l'ensemble des parkings, mais aussi sur chaque parking si jamais, la nécessité d’obtenir une fréquence propre à chaque parking.

1. Clustering

Grâce au clustering, les lieux pourront être classés suivant des caractéristiques propres. Celà offre des opportunités qui permettraient de comprendre par exemple quels lieux ont une probabilité d’avoir des parkings pleins à des périodes, et anticiper peut-être sur la recherche de parking relais.

1. Régression

La régression (linéaires ou non linéaires) permettrait aussi de prédire une variable

cible en fonction de nos variables explicatives. Celà se déroulera en ces étapes:

Collecte d’un volume suffisant de données

Préparation et standardisation de données: cette étape est très importante, et permettra de réduire au mieux les biais, et mettre nos données au bon format

Calcul de la corrélation entre variables et sélection des variables pertinentes: cette étapes compare la variation de deux variables d’un même individu.

Séparation de notre jeu de données (entrainement, test)

Entraînement d’un modèle de régression sur nos données de test: Dans cette étape, notre modèle prend chaque individu enregistre la variable cible et met à jour un fonction associant nos variables explicatives à des coefficients. A chaque itération, elle est mise à jour. A la fin, cette fonction prédira sur le même jeu de données notre variable cible avec une certaine précision.

Prédiction de notre cible sur le jeu de données test: cette fois notre modèle sera exécuté sur des données qu’elle ne connaît pas, et c’est là qu’on peut avoir une idée sur la précision réelle de notre modèle.

**g. Sécurité**

* **Autorisation de contrôle d’accès:** Un compte utilisateur unique et un mot de passe complexe permettront d'accéder aux ressources et services informatiques de l'entreprise
* **Responsabilités des utilisateurs:** Tous les utilisateurs doivent veiller à ne laisser aucune information sensible ou confidentielle autour d'eux.
* Tous les utilisateurs doivent garder leurs mots de passe secrets et ne pas les partager.
* **Accès aux informations confidentielles et restreintes:**L’accès aux données classées comme " confidentielles " ou « restreintes » doit être limité aux personnes autorisées dont les responsabilités professionnelles l’exigent, tel que déterminé par la Politique de sécurité des données ou la direction.Le service de sécurité informatique est responsable d’instaurer les restrictions d’accès.
* **Directives techniques:**
* Droits d’authentification Web
* Droits d’accès et listes de contrôle d’accès aux bases de
* Chiffrement des données au repos et en transit
* Modèle d’accès basé sur les rôles
* Restreindre l'accès aux outils et interfaces d'administration aux seules personnes autorisées.
* Utiliser un compte moins privilégié pour les opérations
* Appliquez une stratégie de mot de passe spécifique à l'administrateur.
* Installez sans délai les mises à jour critiques pour le système d'exploitation ou pour les applications, programmez des vérifications hebdomadaires automatiques
* utiliser le compte de nomination pour accéder à la base de données et créer un compte spécifique à l'application ;
* prendre des mesures contre les attaques par code SQL, scripts, etc.
* Faites des sauvegardes et vérifiez-les régulièrement.
* Implémentez le protocole TLS sur tout le Web, utilisez uniquement les dernières versions, vérifiez l'implémentation correcte.
* Si le cookie n'est pas nécessaire à l'utilisation du service, le consentement de l'internaute suit cette information et le dépôt du cookie.
* Limitez le nombre de composants déployés en mode veille et mettez-les à jour.
* Définir le processus de gestion du stockage : que faut-il stocker, comment et où sont-ils stockés, comment sont gérées les données descriptives ?
* Mettre en œuvre des méthodes d'accès spécifiques aux données car l'utilisation d'un référentiel doit être ponctuelle et ad-hoc.
* Pour la destruction d'archives, sélectionnez un mode qui garantit que l'intégralité de l'archive est détruite.
* Enregistrer les interventions de maintenance dans une main courante.
* Insérer une clause de confidentialité dans les contrats conclus par les prestataires de services.
* Encadrer par un responsable de l’organisme les interventions par des tiers.
* Écrire et exécuter la procédure de suppression des données.
* Supprimer de façon sécurisée les données des matériels avant leur mise au rebut, leur envoi en réparation chez un tiers ou en fin du contrat de location.
* **Exigences de reporting**
* Les rapports d'incident quotidiens doivent être générés et traités par la sécurité informatique ou l'équipe d'intervention ci-dessus.
* Des rapports hebdomadaires détaillés sur les incidents doivent être générés par la sécurité informatique et envoyés au DSI.
* Les problèmes hautement prioritaires détectés par le service informatique doivent être signalés immédiatement.Le DSI doit être contacté aussi vite que possible.

**h.Scénario**

Notre scénario parle d’une personne (homme/femme) d’une trentaine d’année avec un véhicule léger qui souhaiterai trouver une place de parking rapidement sur Bordeaux. Pour cela, il utilise notre application pour repérer une zone géographique près de sa destination et voir si des places y sont encore disponibles.

**i. Evolution future**

Permettre la réservation d’une place depuis notre application.

Permettre la sélection d’une autre ville que Bordeaux.

Voir un indicateur des évolutions possibles sur les places d’un parking en particulier selon une période.

**CONCLUSION**

Notre projet s’inscrit dans le cadre d’un projet Data, pour la création d’un nouveau service pour une entreprise. Cette expérience nous a permis de nous imprégner de l’ampleur d’un projet Data et mettre en œuvre les techniques utilisées dans le domaine du Data. L’étape de préparation de données nous a permis de détecter les problèmes de notre jeu de données de départ, et d’être en mesure de les intégrer avec Talend, et de mener des analyses avec les outils Power BI. Nous avons aussi utilisé les techniques de data science notamment avec les librairies pandas de python.

Cependant nous avons rencontré quelques problèmes, et celà nous a permis d’apprendre. En fin de compte, nous retenons qu’un projet Data implique la participation de diverses profils à des étapes différentes du projet.