

Projet Vélib'

**Amélioration de la
qualité du service
grâce à la donnée**



Jordan S. & Abdelmalek B.

RNCP 7 – Machine Learning & IA – DataScientest





Vélib' Métropole – En quelques mots

Proposition de valeur





- ✓ Mobilité 24/7 dense & flexible
- ✓ Alternative écologique crédible

⚠ Tient si disponibilité OK





Segments clients

-  Franciliens quotidiens (commuters)
-  Touristes & occasionnels
-  Entreprises (abonnements groupés)
-  Autorités publiques (IDFM)





Ressources clés

-  ~20k vélos (40% électriques)
-  1450 stations connectées
-  Données temps réel (GBFS)
-  Équipes terrain & IT




Activités clés

-  Maintenance préventive
-  Rééquilibrage (jour/nuit)
-  Supervision IT temps réel
-  Analyse data → optimisation

Structure de coûts

-  Maintenance vélos/stations
-  Logistique (coût variable)
-  Personnel & IT
-  Amortissement flotte

Flux de revenus

-  Abonnements + usage
-  Subventions publiques
-  Partenariats (potentiel)



P1 Disponibilité imprévisible

Stations vides (stock-out) ou pleines (dock-out) aux heures de pointe (8h-10h, 18h-20h)

Impact : Frustration usagers, détours, retards, perte de temps, risque de désabonnement

P2 Rééquilibrage inefficace

Approche réactive au lieu de prédictive → coûts logistiques élevés et efficacité variable

Impact : Coûts opérationnels élevés (carburant, personnel nuit), km inutiles, efficacité non optimale

P3 UX dégradée en cas d'anomalie

Messages d'erreur génériques, pas prescriptifs → vélos en panne restent en station

Impact : Hausse tickets support, NPS bas, abandons de session, confusion utilisateurs

Alice - Navetteuse

Alice, 32 ans

Commuter quotidien

Usage :

Domicile-travail

2x/jour (8h15 et 18h30)

Pain points :

- Station vide le matin
- Station pleine le soir
- Pas de VAE dispo

Besoins :

- Fiabilité
- Prédications

Marco - Touriste

Marco, 27 ans

Touriste occasionnel

Usage :

2-3 trajets/week-end

Visiteur Paris

Pain points :

- Inscription complexe
- Messages pas clairs
- Station pleine

Besoins :

- Simplicité
- Assistance

Julien - Régulateur

Julien, 41 ans

Logisticien Smovengo

Usage :

Planification tournées

Travail de nuit

Pain points :

- Pas de priorisation
- Tournées inefficaces
- Pas de visibilité

Besoins :

- Prédications ML
- Optimisation



EXPERIENCE MAP

Parcours utilisateur — Trajet domicile-travail

Alice - Navetteuse quotidienne (32 ans)

| Étape | Planification | Départ | Trajet | Arrivée |
|--------------|--|---|--|--|
| ACTIONS | Consulte l'app Vélib' avant de partir Vérifie la disponibilité à sa station | Se rend à la station habituelle Scanne le QR code | Pédale vers son lieu de travail Trajet habituel 11e → 2e | Arrive à destination Cherche une borne libre pour restituer le vélo |
| PAIN POINTS | <ul style="list-style-type: none">❌ Données pas toujours fiables❌ App indique 3 vélos mais 0 sur place | <ul style="list-style-type: none">❌ Station vide (stock-out)❌ Aucun vélo disponible → retard❌ VAE jamais disponibles aux heures de pointe | <ul style="list-style-type: none">❌ Vélo défectueux (freins, vitesses)❌ Borne de départ défaillante → facturation incorrecte | <ul style="list-style-type: none">❌ Station pleine (dock-out)❌ 0 borne libre → détour vers autre station❌ Facturation continue même si pas de place |
| OPPORTUNITÉS | <ul style="list-style-type: none">✅ Prédications fiables à 15 min✅ Notification push "Station critique dans 10 min" | <ul style="list-style-type: none">✅ Alertes push si risque de stock-out✅ Suggestions de stations alternatives proches✅ Réservation VAE prioritaire pour abonnés réguliers | <ul style="list-style-type: none">✅ Signalement pannes en temps réel✅ Pré-check vélo (état général) avant départ✅ Compensation automatique si incident | <ul style="list-style-type: none">✅ Prédications disponibilité bornes à l'arrivée✅ Suggestions stations proches avec bornes libres✅ Arrêt facturation automatique si dock-out constaté |



Insight clé :

Le problème principal d'Alice est le **manque de prédictibilité**. Elle a besoin de savoir **à l'avance** s'il y aura un vélo disponible et une borne libre à destination pour planifier sereinement son trajet quotidien.



EXPERIENCE MAP

Parcours utilisateur — Découverte touristique



Marco - Touriste occasionnel (27 ans)

| Étape | Découverte | Inscription | Premier trajet | Restitution |
|--------------|--|--|---|---|
| ACTIONS | <p>Recherche "vélo Paris" sur Google</p> <p>Découvre Vélib' sur Google Maps</p> <p>Compare tarifs avec trottinettes</p> | <p>Télécharge l'app Vélib'</p> <p>Remplit le formulaire d'inscription</p> <p>Ajoute sa carte bancaire</p> | <p>Cherche une station proche de son hôtel</p> <p>Prend son premier vélo</p> <p>Direction Tour Eiffel</p> | <p>Arrive à destination touristique</p> <p>Cherche où restituer le vélo</p> <p>Reçoit message d'erreur</p> |
| PAIN POINTS | <p>❌ Tarifs peu clairs (abonnement vs ticket 1 jour)</p> <p>❌ Comparaison difficile avec autres options</p> | <p>❌ Formulaire d'inscription long</p> <p>❌ Pas d'option "touriste" simplifiée</p> <p>❌ Interface uniquement en français</p> | <p>❌ Pas de tutoriel rapide "première utilisation"</p> <p>❌ Vélo défectueux (pas détecté à la prise)</p> <p>❌ Messages app en français uniquement</p> | <p>❌ Station pleine à Tour Eiffel</p> <p>❌ Message "Erreur technique" pas explicite</p> <p>❌ Ne sait pas quoi faire → panique</p> <p>❌ Surfacturation si ne trouve pas de borne</p> |
| OPPORTUNITÉS | <p>✅ Comparateur tarifaire clair et visuel</p> <p>✅ Page d'accueil "Touristes" dédiée</p> <p>✅ Vidéo explicative multilingue</p> | <p>✅ Onboarding gamifié (3 étapes max)</p> <p>✅ Option "Inscription rapide touriste"</p> <p>✅ Interface multilingue (EN, ES, IT, DE)</p> | <p>✅ Tutoriel interactif au premier usage</p> <p>✅ Pré-check vélo guidé ("Testez les freins")</p> <p>✅ Mode "Touriste" avec itinéraires suggérés</p> | <p>✅ Messages prescriptifs multilingues : "Station pleine. 3 stations proches :"</p> <p>✅ Navigation GPS vers borne libre</p> <p>✅ Arrêt facturation automatique si problème</p> <p>✅ Chat support multilingue instantané</p> |



Insight clé :

Marco a besoin de **simplicité** et de **clarté**. En tant que touriste, il ne connaît pas Paris et a besoin de **messages prescriptifs multilingues** qui lui disent exactement quoi faire en cas de problème.



EXPERIENCE MAP

Parcours opérationnel — Tournée de rééquilibrage nocturne

Julien - Logisticien régulation Smovengo (41 ans)

| Étape | Analyse | Planification | Exécution | Bilan |
|--------------|---|---|--|---|
| ACTIONS | <p>Consulte les données de fin de journée (18h-20h)</p> <p>Identifie les stations critiques</p> <p>Évalue les besoins de rééquilibrage</p> | <p>Planifie les tournées de nuit</p> <p>Définit l'ordre des stations</p> <p>Affecte les équipes et camions</p> | <p>Équipes partent en tournée (22h-5h)</p> <p>Rééquilibrage vélos méca + VAE</p> <p>Signallement pannes découvertes sur place</p> | <p>Bilan matinal (6h)</p> <p>Analyse efficacité tournées</p> <p>Rapport pour direction</p> |
| PAIN POINTS | <p>❌ Pas de visibilité sur prévisions matinales</p> <p>❌ Analyse basée sur données passées uniquement</p> <p>❌ Pas de score de criticité automatique</p> | <p>❌ Priorisation au "feeling"</p> <p>❌ Pas d'outil d'optimisation de tournées (VRP)</p> <p>❌ Contraintes trafic/horaires difficiles à anticiper</p> <p>❌ Pas de visibilité sur pannes bloquantes</p> | <p>❌ Bornes HS découvertes sur place → temps perdu</p> <p>❌ Vélos bloqués non signalés à l'avance</p> <p>❌ Stations rééquilibrées se re-vident en 1h (8h-9h)</p> <p>❌ Efficacité variable (65%)</p> | <p>❌ Difficile de mesurer le ROI des tournées</p> <p>❌ Pas d'indicateurs d'efficacité standardisés</p> <p>❌ Reporting manuel chronophage</p> |
| OPPORTUNITÉS | <p>✅ Dashboard prédictif : criticité prévue à 8h-10h</p> <p>✅ Score de priorité automatique par station</p> <p>✅ Alertes pannes détectées en temps réel</p> | <p>✅ Optimisation tournées (VRP) avec contraintes trafic</p> <p>✅ Suggestion itinéraires optimaux</p> <p>✅ Prévisions de demande par créneau horaire</p> <p>✅ Carte interactive avec criticité prédite</p> | <p>✅ App terrain avec mises à jour temps réel</p> <p>✅ Notifications si station devient critique pendant tournée</p> <p>✅ Pré-check bornes avant déplacement</p> <p>✅ Ajustement dynamique de la tournée</p> | <p>✅ KPI automatisés (Rebalancing Efficiency, km économisés)</p> <p>✅ Reporting automatique</p> <p>✅ Analyse avant/après par station</p> <p>✅ Recommandations ML pour améliorer futures tournées</p> |



Insight clé :

Julien a besoin d'**outils d'aide à la décision** basés sur des prédictions. Passer d'une approche réactive (réagir aux données passées) à une approche **prédictive** (anticiper la criticité matinale) permettrait d'améliorer l'efficacité des tournées de 65% → 75%+ et réduire les km inutiles.

synthese

BACKLOG D'ANALYSES QUANTITATIVES

8 questions pour mesurer l'ampleur des problèmes identifiés



| Axe | Question | Hypothèse à tester |
|------------------------|--|---|
| A DISPONIBILITÉ | Q1. Taux de criticité par station et créneau Quelle part du temps chaque station est en stock-out (0 vélo) et dock-out (0 place) par créneau ? | H1 : Les pics 8h-10h et 18h-20h génèrent une hausse significative de stock-out/dock-out sur les zones bureau et résidentiel. |
| | Q2. Clusters de stations Quels clusters de stations (quartiers) partagent la même dynamique de criticité ? | H2 : Les patterns de criticité sont répétables (ex. stations près des gares ont toutes un pic stock-out le matin). |
| B RÉGULATION | Q3. Efficacité du rééquilibrage Après une tournée, quelle proportion de stations ciblées sortent durablement de l'état critique ? | H3 : L'efficacité est meilleure la nuit (moins de contraintes logistiques, demande plus stable). |
| | Q4. Temps de retour en service (MTTR) Quel est le temps moyen de retour en service d'une station après blocage/panne ? | H4 : Longue inactivité détectable via GBFS (station avec 0 mouvement >30 min). Idéalement, logs maintenance nécessaires. |
| C USAGE & MIX | Q5. Turnover et criticité Quel est le turnover (prises/dépôts par heure) par station et sa corrélation avec criticité ? | H5 : Turnover élevé → risques de stock-out/dock-out plus fréquents (volatilité). |
| | Q6. Impact du mix VAE/mécaniques Le mix VAE/mécaniques influe-t-il sur la criticité ? | H6 : Pénurie de VAE aux heures de pointe (forte demande pour trajets longs/vallonnés). |
| D FACTEURS EXTERNES | Q7. Impact météo Quel impact de la météo (pluie/température) sur la demande et criticité ? | H7 : La pluie diminue la demande globale mais accentue les déséquilibres localement. |
| | Q8. Impact événements Quel impact des événements (concerts, manifestations) sur les stations alentour ? | H8 : Pics ponctuels prédictibles (ex. concert Stade de France → stations vides avant, pleines après). |

BENCHMARK - SOLUTIONS EXISTANTES

A1 — Réduire criticité des stations aux heures de pointe

 CONCURRENT

Citi Bike (NYC)

Programme "Bike Angels" : gamification pour inciter usagers à rééquilibrer eux-mêmes les stations (points, récompenses).

Prédictions temps réel de disponibilité sur app mobile.

✓ **Pertinence** : Modèle répliquable à Paris. Système de points + prédictions combinés.


 CONCURRENT

Santander Cycles (Londres)

API publique avec prédictions de disponibilité à 15-30 min.

Intégrations tierces (Citymapper, Google Maps) affichent les prévisions.

✓ **Pertinence** : API ouverte = adoption massive. Modèle à suivre pour Vélib'.

 RECHERCHE

Paper : "Predicting Bike Availability"

Études académiques (Stanford, MIT) sur prédiction disponibilité avec Random Forest, LightGBM.

Précision 75-85% à horizon 30 min.

✓ **Pertinence** : Modèles ML éprouvés, datasets publics disponibles pour benchmarking.

A2 — Optimiser les tournées de régulation

 CONCURRENT

Motivate (opérateur Citi Bike)

Algorithmes de réoptimisation dynamique : tournées ajustées en temps réel selon demande.

Réduction de 15-20% des km parcourus vs approche statique.

✓ **Pertinence** : Résultats prouvés. Nécessite investissement IT (API temps réel, algo OR).

 PRODUIT DIGITAL

UPS/DHL - Optimisation tournées

Algorithmes VRP (Vehicle Routing Problem) avec contraintes temps, trafic, priorités.

OR-Tools (Google) open-source.

⚠ **Pertinence** : Transposable mais contexte différent (livraison vs rébalancing). Adaptation nécessaire.

 RECHERCHE

Paper : "Dynamic Rebalancing"

Recherche académique (Barcelona, Paris) sur algorithmes heuristiques pour rébalancing vélo-partage.

Comparaison Greedy vs Optimization exacte.

✓ **Pertinence** : Contexte identique (vélo-partage). Codes sources parfois disponibles (GitHub).

A3 — Améliorer la gestion des anomalies

CONCURRENT

Lime / Bird (trottinettes)

Diagnostic guidé in-app : "Véhicule défectueux ? Suivez ces 3 étapes..."

Photo du problème + signalement géolocalisé.

✔ **Pertinence** : UX simple, répliquable pour Vélib'. Réduit tickets support.

PRODUIT DIGITAL

Uber - Support contextuel

FAQ dynamique selon contexte (en course, après course, paiement).

Messages prescriptifs : "Problème de paiement ? Voici 3 actions..."

✔ **Pertinence** : UX éprouvée pour mobilité. Adapté aux situations urgentes (vélo bloqué).

RECHERCHE

Chatbots ML pour support

Papiers sur NLP pour support client automatisé (classification intention, réponses automatiques).

Réduction 30-40% tickets niveau 1.

⚠ **Pertinence** : Nécessite dataset training (historique tickets). ROI à valider pour Vélib'.

A4 — Recommander des stations alternatives

PRODUIT DIGITAL

Google Maps - Affluence

Prédictions "Usually not too busy" pour lieux publics (restaurants, commerces).

Basé sur historique + temps réel.

✔ **Pertinence** : UX familière pour utilisateurs. Modèle transposable aux stations Vélib'.

PRODUIT DIGITAL

Citymapper - Routes alternatives

Suggestions d'itinéraires alternatifs en temps réel selon disponibilité transports.

Combine plusieurs modes (vélo, métro, bus).

✔ **Pertinence** : Citymapper intègre déjà GBFS Vélib'. Partenariat potentiel pour reco stations.

RECHERCHE

Recommender Systems - Mobilité

Papiers sur systèmes de recommandation pour mobilité urbaine (stations alternatives, itinéraires optimaux).

Approches : Collaborative Filtering, Content-Based.

⚠ **Pertinence** : Complexe pour un MVP. Commencer par règles simples (distance, disponibilité).

💡 Synthèse :

Concurrents directs (Citi Bike, Santander Cycles) offrent les benchmarks les plus pertinents → contexte identique, résultats prouvés.

Produits digitaux grand public (Google Maps, Uber, Citymapper) proposent des UX éprouvées facilement transposables.

Recherche académique fournit méthodes rigoureuses (ML, OR) mais nécessite adaptation au contexte opérationnel.



MATRICE DE PRIORISATION

Scoring des 4 améliorations selon 4 dimensions business

| Amélioration | Satisfaction (Impact NPS, satisfaction client) | Nouveau business (Acquisition nouveaux clients) | Efficience interne (Réduction coûts, processus) | Croissance revenu clients actuels (Rétention, usage accru) | Score de priorisation |
|--|---|--|--|--|--------------------------|
| A1 Réduire criticité stations (SOR/DOR) | 3 | 1 | 2 | 3 | 9 |
| A2 Optimiser tournées régulation (RE) | 1 | 0 | 3 | 1 | 5 |
| A3 Améliorer gestion anomalies (MTTR) | 3 | 1 | 2 | 2 | 8 |
| A4 Recommander stations alternatives (JAR) | 2 | 1 | 1 | 2 | 6 |



Priorité 1

A1
Réduire criticité stations

9



Priorité 2

A3
Améliorer gestion anomalies

8



Priorité 3

A4
Recommander stations alternatives

6



Priorité 4

A2
Optimiser tournées régulation

5



MÉTRIQUES & FORMULES

Indicateurs opérationnels pour mesurer la criticité et l'efficacité



Disponibilité

SOR

Stock-Out Rate (Taux de rupture vélos)

$$\text{SOR} = (\text{Minutes où bikes_total} = 0) / (\text{Minutes observées})$$

Exemple : Si République est à 0 vélo pendant 120 minutes sur 1440 minutes (1 jour), alors $\text{SOR} = 120/1440 = 8,3\%$

DOR

Dock-Out Rate (Taux de saturation bornes)

$$\text{DOR} = (\text{Minutes où docks_free} = 0) / (\text{Minutes observées})$$

Exemple : Si Bastille est à 0 borne libre pendant 90 minutes sur 1440 minutes, alors $\text{DOR} = 90/1440 = 6,25\%$



Usage & Turnover

MTTR

Mean Time To Repair (Temps moyen de résolution)

$$\text{MTTR} = \text{avg}(t_{\text{résolution}} - t_{\text{incident}})$$

Proxy : Définir un incident par règle (ex. bornes libres = 0 & vélos inchangés >30 min)

TOH

Turnover per Hour (Mouvements par heure)

$$\text{TOH} = \text{Nb prises} + \text{Nb dépôts par heure}$$

Note : Déduit des deltas d'inventaire si on historise les snapshots GBFS

Key Performance Indicators

A1 — Réduire criticité stations



KPI #1

Réduction Stock-Out Rate
(% temps sans vélo)



KPI #2

Réduction Dock-Out Rate
(% temps sans borne libre)



KPI #3

Réduction Criticality Minutes
(minutes en état critique)






KPI #4

Amélioration Rebalancing
Efficiency
(+10 points d'efficacité)



COMPARAISON AVANT / APRÈS

| KPI |  Baseline (État actuel) | |  Objectif (Après amélioration) |  Amélioration |
|---|---|---|--|--|
| KPI 1 : SOR Stock-Out Rate | 8,5% | → | 6,8% | -20% |
| KPI 2 : DOR Dock-Out Rate | 6,2% | → | 5,3% | -15% |
| KPI 3 : CM Criticality Minutes/jour | 185 min | → | 139 min | -25% |
| KPI 4 : RE Rebalancing Efficiency | 65% | → | 75% | +10 pts |

Exemple concret : Station République (AM Peak 8h-10h)

Station République — 11e arrondissement

Contexte : Station très fréquentée, 45 bornes, zone bureau, forte demande matin

AVANT (Baseline)
35%
SOR 8h-10h

APRÈS (Cible)
28%
SOR 8h-10h

GAIN
-7 pts
-20% relatif

Traduction concrète :

Sur 120 minutes (8h-10h), la station passe de **42 minutes à 0 vélo** (baseline) à **34 minutes à 0 vélo** (cible).
→ **8 minutes gagnées** où les usagers trouvent un vélo disponible.

Station Opéra — 2e arrondissement

Contexte : Station destination travail, 50 bornes, zone bureau, saturation soir

AVANT (Baseline)
22%
DOR 18h-20h

APRÈS (Cible)
18,7%
DOR 18h-20h

GAIN
-3,3 pts
-15% relatif

Traduction concrète :

Sur 120 minutes (18h-20h), la station passe de **26 minutes à 0 borne libre** (baseline) à **22 minutes à 0 borne libre** (cible).
→ **4 minutes gagnées** où les usagers peuvent restituer leur vélo sans détour.



Business Intelligence

Dashboard pour visualiser criticité des stations, alerter équipes ops et suivre KPIs

MVP ✓



Pourquoi commencer par la BI ?

1

Quick win opérationnel

Les équipes ops ont besoin **maintenant** de visibilité sur les stations critiques. Pas besoin de ML pour détecter qu'une station est à 0 vélo.

2

Données déjà disponibles

GBFS fournit données temps réel (vélos disponibles, bornes libres). Pas besoin d'attendre historique pour calculer SOR/DOR actuels.

3

Complexité maîtrisée

Dashboard BI = 4-6 semaines dev. ML prédictif = 3-6 mois (data collection, training, validation, déploiement).

4

Base pour le ML

Le dashboard BI permet de **valider les hypothèses** et de comprendre les patterns avant d'investir dans du ML coûteux.



Roadmap d'intégration des 3 approches

T0-T1 : BI (MVP)

0-3 mois

Data Eng : Pipeline GBFS → PostgreSQL

BI : Dashboard Metabase avec SOR/DOR temps réel, alerting, Top 10 stations critiques

Objectif : Visibilité opérationnelle

T2-T3 : ML Prédictif

3-9 mois

Data Eng : Historisation 30-90 jours

ML : Modèle prédictif (LightGBM) pour criticité H+30

BI : Intégration prédictions dans dashboard

Objectif : Anticipation

T4 : Optimisation

9-12 mois

Data Eng : Logs tournées opérationnelles

ML : Algorithme VRP (Vehicle Routing Problem) pour optimiser tournées

Objectif : Efficacité RE +10-15 pts

✓ Conclusion :

On choisit **BI comme MVP** car c'est le meilleur compromis **impact/effort** pour T0. Le ML viendra en Phase 2 une fois qu'on aura validé que le dashboard BI est adopté et utile. Les 3 approches (Data Eng, BI, ML) sont **complémentaires** et s'intègrent progressivement dans la roadmap.

Dashboard BI

De la vision à la conception détaillée

Dashboard de monitoring de criticité

Un outil BI pour donner aux équipes opérationnelles une **visibilité temps réel** sur l'état de criticité des stations Vélib' et permettre un **rééquilibrage data-driven**.



KPI Cards temps réel

Visualisation instantanée : SOR, DOR, nombre de stations critiques, évolution vs hier



Carte géographique

Localisation des stations avec code couleur selon criticité (● Normal / ● Attention / ● Critique)



Top 10 stations critiques

Classement des stations prioritaires à traiter, triable par SOR/DOR et créneau horaire



Heatmap temporelle

Visualisation patterns de criticité par heure et jour de la semaine

1

Data Engineering

Pipeline GBFS
→ PostgreSQL

2

Dashboard BI

Metabase
Visualisations

3

Déploiement

Production
Équipes ops

🎯 Objectifs de succès du MVP

-20%

Stock-Out Rate

-15%

Dock-Out Rate

4-6 sem

Délai développement

5 users

Chefs d'exploitation



JUSTIFICATION & DÉCISIONS

Pourquoi ces visualisations ? Quelles décisions ?



Justification des choix de visualisation

1. KPI Cards (Number)

Type : Cartes de chiffres clés

Pourquoi : Scan instantané de l'état global en 2 secondes

Best practice : Position haute gauche (zone prioritaire), delta vs hier pour contexte

2. Tableau Top 10

Type : Table triable

Pourquoi : Priorisation actionnable : savoir immédiatement quelles stations traiter

Best practice : Code couleur (●●●) pour détection rapide + tri multi-colonnes

3. Carte géographique

Type : Carte interactive avec pins

Pourquoi : Contextualisation spatiale : identifier clusters géographiques de criticité

Best practice : Pins colorés par criticité + zoom/filtres pour exploration

4. Courbe temporelle (Line Chart)

Type : Time series (30 jours)

Pourquoi : Détecter tendances (hausse/baisse) et anomalies (pics inhabituels)

Best practice : 2 courbes (SOR + DOR) sur même graphe pour comparaison directe

5. Heatmap (heure × jour)

Type : Heatmap 2D

Pourquoi : Révéler patterns temporels (8h-10h = pics récurrents)

Best practice : Palette séquentielle (vert → orange → rouge) pour intensité



PROTOTYPE DASHBOARD - MAQUETTE

Dashboard de monitoring de criticité des stations



Vélib' — Criticité Stations

Filtres : Créneau [Tous ▼] | Arrondissement [Tous ▼] | Dernière MAJ : 14h35

Stock-Out Rate (SOR)

8,5%

▲ +1,2 vs hier

● Objectif: 6,8%

Dock-Out Rate (DOR)

6,2%

▲ +0,8 vs hier

● Objectif: 5,3%

Stations Critiques

47

▲ +5 vs hier

3,1% du réseau

Rebalancing Efficiency

68%

▲ +3 pts vs hier

● Objectif: 75%

📍 Top 10 Stations Critiques (SOR)

| Rang | Station | Arrond. | SOR | DOR | Status |
|------|--------------|---------|-------|-------|--------|
| 1 | Bastille | 11e | 12,5% | 8,2% | ● |
| 2 | Gare du Nord | 10e | 11,8% | 9,1% | ● |
| 3 | République | 11e | 11,2% | 6,5% | ● |
| 4 | Châtelet | 1er | 10,8% | 7,3% | ● |
| 5 | Opéra | 2e | 10,5% | 12,1% | ● |

— Voir les 100 stations critiques

🗺️ Carte Géographique

🗺️ Carte interactive

Pins colorés par criticité

● Normal | ● Attention | ● Critique



Évolution SOR/DOR (30 jours)



Courbe temporelle (line chart)
SOR et DOR sur 30 derniers jours
Tendance : Hausse progressive



Heatmap Criticité (heure × jour)



Heatmap (heure en Y, jour en X)
Pics confirmés : 8h-10h et 18h-20h
Lundi et Vendredi = pires jours

🎯 Décisions métier permises par le dashboard



Décision 1 : Priorisation tournées

💡 INSIGHT

"Bastille, Gare du Nord et République ont SOR > 12% (Top 3)"

✅ ACTION

Planifier tournée de rééquilibrage ciblée sur ces 3 stations dès ce soir (priorité 1)

👤 Responsable : Chef d'exploitation



Décision 2 : Ajout de capacité

💡 INSIGHT

"8e arrondissement : 5 stations dans le Top 10, toutes critiques lundi 8h-10h"

✅ ACTION

Étudier ajout de 10-15 bornes sur ces 5 stations (investissement capex)

👤 Responsable : Directeur opérations



Décision 3 : Ajustement horaires régulation

💡 INSIGHT

"Heatmap montre pics 8h-10h (AM) et 18h-20h (PM) tous les jours semaine"

✅ ACTION

Renforcer régulation 7h-8h (avant pic AM) et 17h-18h (avant pic PM)

👤 Responsable : Responsable régulation



Décision 4 : Mesure de succès

💡 INSIGHT

"SOR actuel = 8,5%, objectif = 6,8%. Écart = 1,7 points, tendance stable"

✅ ACTION

Décider de continuer les actions actuelles OU pivoter si pas d'amélioration en 2 semaines

👤 Responsable : Directeur opérations + Équipe produit

💡 Principe "Insight to Action" :

Chaque visualisation du dashboard répond à la question "Et alors, qu'est-ce que je fais ?"

Un dashboard BI efficace ne se contente pas de montrer des chiffres. Il doit **permettre des décisions concrètes** en transformant les insights en actions claires avec des responsables identifiés.



BACKLOG QUESTIONS MÉTIER

Les questions auxquelles le dashboard doit répondre

? 1. QUOI ?

Q1 — État global

Quel est le niveau de criticité actuel du réseau ?

→ KPI cards : SOR, DOR, nb stations critiques

Q2 — Évolution

Sommes-nous meilleurs ou pires qu'hier / semaine dernière ?

→ KPI cards avec delta (1 +1.2% ou 1 -0.5%)

Q3 — Objectifs

Sommes-nous en ligne avec nos objectifs (-20% SOR) ?

→ Indicateur visuel (● atteint / ● non atteint)

📍 2. OÙ ?

Q4 — Stations prioritaires

Quelles sont les 10 stations les plus critiques à traiter maintenant ?

→ Tableau Top 10 triable (SOR/DOR/Criticality Index)

Q5 — Localisation géographique

Où sont localisées les stations critiques sur la carte ?

→ Carte géographique avec pins colorés (● ● ●)

Q6 — Clusters géographiques

Y a-t-il des quartiers entiers en difficulté ?

→ Carte avec agrégation par arrondissement

🕒 3. QUAND ?

Q7 — Évolution temporelle

Comment évoluent SOR/DOR sur les 30 derniers jours ?

→ Courbe temporelle (line chart)

Q8 — Patterns horaires

À quelles heures et quels jours la criticité est maximale ?

→ Heatmap (heure × jour de semaine)

Q9 — Créneaux critiques

Les pics 8h-10h et 18h-20h sont-ils confirmés ?

→ Bar chart (criticité par tranche horaire)

🔍 4. POURQUOI ?

Q10 — Corrélation turnover

Les stations critiques sont-elles celles avec fort turnover ?

→ Scatter plot (Turnover vs Criticality Index)

Q11 — Impact rééquilibrage

Nos actions de rééquilibrage sont-elles efficaces ?

→ Bar chart (Rebalancing Efficiency par équipe/créneau)

Q12 — Comparaison zones

Quels arrondissements performant mieux/moins bien ?

→ Bar chart horizontal (SOR moyen par arrondissement)



IN SCOPE (MVP)

KPI Cards temps réel

SOR, DOR, nb stations critiques, RE avec delta vs hier

Top 10 stations critiques

Tableau triable par SOR/DOR/Criticality Index

Carte géographique

Pins colorés par criticité (● Normal / ● Attention / ● Critique)

Évolution temporelle

Courbe SOR/DOR sur 30 derniers jours

Heatmap criticité

Patterns heure x jour de semaine

Filtres de base

Par créneau (matin/soir/journée) et arrondissement

Export Excel

Top 100 stations critiques pour analyse externe



OUT SCOPE (Phase 2+)

Prédictions ML

Criticité prévue à H+15/H+30 → nécessite historique 30-90 jours

Optimisation tournées (VRP)

Algorithme Vehicle Routing Problem → Phase 3

Intégration app mobile

Notifications push usagers finaux → nécessite validation équipe produit

Alerting automatique

Slack/Email si seuil dépassé → Phase 1.5 (quick win post-MVP)

Drill-down détaillé

Vue par station individuelle avec historique complet

Données météo

Corrélation criticité x météo → nécessite API externe

Reporting automatisé

PDF hebdomadaire envoyé par email → Phase 2



Stratégie de déploiement progressif

Phase 1

Pilote restreint

🚀 Semaines 1-2

Utilisateurs :

- 2 chefs d'exploitation
- 1 responsable régulation

Objectif :

Valider utilisabilité, recueillir retours, identifier bugs

Phase 2

Déploiement étendu

🚀 Semaines 3-4

Utilisateurs :

- 5 chefs d'exploitation
- 3 responsables régulation
- 1 directeur ops

Objectif :

Usage quotidien, ajustements UI/UX, formation équipes

Phase 3

Production complète

🚀 Semaine 5+

Utilisateurs :

- Toutes équipes ops (30-50 personnes)
- Direction Smovengo
- Équipes IDFM (reporting)

Objectif :

Déploiement complet, monitoring KPIs, itérations



RISQUES LIÉS AUX DONNÉES

Analyse des risques data alimentant le MVP

R1 — Qualité des données GBFS

CRITIQUE

DESCRIPTION DU RISQUE

L'API GBFS peut fournir des données **incohérentes ou erronées** : • Stations avec bikes_total < 0 ou > capacité

- Coordonnées GPS incorrectes (0,0 ou hors Paris)
- Timestamps dupliqués ou manquants

IMPACT

Dashboard affiche des KPIs **faux** → décisions erronées → perte de confiance des équipes ops

MITIGATION

- **Data quality checks** à l'ingestion (règles métier)
- Alerting si anomalie détectée (ex: SOR > 50% sur 100 stations)
- Logs détaillés pour debugging
- Fallback : exclure données aberrantes du calcul

R2 — Disponibilité API GBFS

ÉLEVÉ

DESCRIPTION DU RISQUE

L'API GBFS peut être **indisponible** (erreurs 5xx, timeout) : • Maintenance programmée non communiquée

- Surcharge serveur aux heures de pointe
- Problèmes réseau ou firewall

IMPACT

Dashboard ne se met plus à jour → équipes ops **aveugles** → retour au "feeling"

MITIGATION

- **Retry logic** avec backoff exponentiel (3 tentatives)
- Cache des dernières données valides (afficher "Dernière MAJ: 10 min")
- Monitoring uptime API (Pingdom, UptimeRobot)
- SLA avec fournisseur API (IDFM) : 99,5% uptime minimum

R3 — Complétude historique

MOYEN

DESCRIPTION DU RISQUE

Historique GBFS **incomplet** (trous de données) : • Pipeline d'ingestion arrêté pendant maintenance

- Panne serveur PostgreSQL
- Partitions manquantes (erreur admin)

IMPACT

Courbes temporelles avec **gaps** → analyses tendances biaisées → fausses conclusions

MITIGATION

- **Alerting** si pas de données reçues pendant > 5 min
- Backfill automatique des données manquantes (récupération a posteriori)
- Redondance infrastructure (DB répliquée)
- Visual warning sur dashboard si données incomplètes

R4 — Fiabilité calculs KPIs

ÉLEVÉ

DESCRIPTION DU RISQUE

Erreurs dans formules SQL de calcul SOR/DOR : • Bug dans requête (division par zéro, window fonction mal utilisée)

- Changement définition KPI non propagé
- Données filtrées incorrectement

IMPACT

KPIs **faux** affichés pendant des semaines → décisions basées sur données erronées

MITIGATION

- **Tests unitaires** sur requêtes SQL (dbt tests, Great Expectations)
- Validation croisée avec calculs manuels (échantillon 10 stations)
- Versionning définitions KPIs (Git, documentation)
- Code review obligatoire avant déploiement requêtes



CONFORMITÉ RÉGLEMENTAIRE

Analyse RGPD, sécurité et réglementations secteur



RGPD - Règlement Général sur la Protection des Données

1. Collecte de données personnelles

Données collectées :

- Aucune donnée personnelle d'utilisateurs finaux
- Seules données agrégées : stations, vélos disponibles, timestamps
- Pas de tracking individuel des trajets

✓ CONFORME

2. Base légale du traitement

Fondement juridique :

- Mission d'intérêt public (art. 6.1.e RGPD)
- Service public de mobilité (IDFM)
- Optimisation service → intérêt légitime

✓ CONFORME

3. Minimisation des données

Principe appliqué :

- Collecte uniquement données nécessaires (stations, dispo)
- Pas de sur-collecte (âge, genre, trajet usager)
- Conservation 90 jours max (rolling window)

✓ CONFORME

4. Droit d'accès et rectification

Non applicable :

- Pas de données personnelles identifiables
- Données agrégées anonymes par nature
- Aucune demande d'accès/rectification possible

✓ N/A

5. Analyse d'impact (AIPD)

Statut :

- AIPD non obligatoire (pas de risque élevé)
- Données non personnelles, pas de profilage
- Registre de traitement à tenir (obligation)

⚠ REGISTRE À CRÉER

6. Sous-traitance et transferts

Vérifications nécessaires :

- Contrat DPA avec hébergeur cloud (AWS/GCP)
- Si hébergement EU : conforme
- Si hébergement US : vérifier Data Privacy Framework

● ACTION REQUISE

Sécurité des données - Mesures techniques



Chiffrement

- Données en transit : TLS 1.3
- Données au repos : AES-256
- PostgreSQL : encryption at rest
 - Backups chiffrés



Contrôle d'accès

- Authentification SSO (OAuth2)
- RBAC : rôles utilisateurs (ops, admin, viewer)
 - MFA obligatoire pour admins
 - Logs d'accès complets



Infrastructure

- Firewall applicatif (WAF)
- Segmentation réseau (VPC)
- Patch management automatisé
- Backup quotidien (rétention 30j)



Monitoring

- Détection anomalies (SIEM)
- Alerting tentatives intrusion
- Audit logs immuables
- SOC 24/7 si cloud managé



Continuité

- Plan de reprise (PRA)
- RTO : 4h | RPO : 1h
- DB répliquée (multi-AZ)
- Tests disaster recovery trimestriels



Conformité

- ISO 27001 (si hébergeur certifié)
- Audits sécurité annuels
- Pentest externe (si budget)
- Documentation PSSI complète



Régulations spécifiques au secteur mobilité

1. Contrat IDFM (Île-de-France Mobilités)

Obligations contractuelles :

- SLA disponibilité service (99,5%)
- Reporting trimestriel à IDFM
- KPIs de performance définis
- Pénalités si non-respect

 VÉRIFIER CONTRAT

2. Open Data (Loi République Numérique)

Obligation de transparence :

- Données agrégées doivent être ouvertes
- API GBFS déjà publique (conforme)
- Dashboard interne OK (pas d'obligation ouverture)

 CONFORME

3. Accessibilité numérique (RGAA)

Standard RGAA 4.1 :

- Service public ... obligation accessibilité
- Dashboard doit être RGAA niveau AA minimum
- Audit accessibilité nécessaire

 AUDIT RGAA REQUIS

4. Marchés publics

Si développement externalisé :

- Procédure d'appel d'offres (Code marchés publics)
- Cahier des charges technique détaillé
- Clause de réversibilité (exit provider)

 SELON BUDGET



CONSIDÉRATIONS ÉTHIQUES

Biais, discrimination, vie privée et impact social



1. Biais algorithmique & décisionnel

Biais dans les données historiques

RISQUE FAIBLE

Problème potentiel :

Les données GBFS reflètent l'état actuel du réseau, qui peut contenir des biais structurels (ex: moins de stations dans certains quartiers défavorisés).

✓ MITIGATION

- Le dashboard ne fait que **refléter** l'état, pas décider de l'allocation
- Transparence totale : les chefs d'exploitation voient toutes les stations
- Pas de modèle ML prédictif (Phase 1) → pas de biais amplifié

Priorisation des quartiers aisés

RISQUE MOYEN

Problème potentiel :

Si les équipes ops concentrent leurs efforts sur les stations "rentables" (quartiers touristiques, bureaux) au détriment des quartiers populaires.

✓ MITIGATION

- Politique de **service public** : toutes stations égales
- KPIs mesurés sur **Top 100 stations** (pas seulement centres touristiques)
- Audit annuel : vérifier répartition géographique des interventions

Biais dans les KPIs choisis

RISQUE FAIBLE

Problème potentiel :

Mesurer uniquement SOR/DOR peut favoriser optimisation des stations à fort trafic (effet Goodhart : "quand une mesure devient un objectif, elle cesse d'être une bonne mesure").

✓ MITIGATION

- KPIs multiples (SOR + DOR + RE + satisfaction)
- Revue trimestrielle des KPIs avec comité éthique
- Garde-fous : pas de pénalités pour stations "difficiles"

Transparence décisions ops

RISQUE FAIBLE

Problème potentiel :

Dashboard utilisé comme "boîte noire" pour justifier décisions opaques (fermeture stations, réallocation bornes).

✓ MITIGATION

- **Explicabilité totale** : requêtes SQL documentées, formules KPIs publiques
- Accès dashboard étendu (élus, associations usagers sur demande)
- Reporting public trimestriel IDFM



2. Protection de la vie privée

Anonymat des usagers

RISQUE FAIBLE

Situation actuelle :

Le dashboard n'utilise aucune donnée personnelle identifiable. Seules données agrégées par station (vélos disponibles, bornes libres).

✓ GARANTIE

- Pas de tracking individuel des trajets
- Pas de lien avec comptes utilisateurs Vélib'
- Anonymat par design (privacy by design)

Risque de réidentification

RISQUE FAIBLE

Problème théorique :

Même avec données agrégées, des patterns temporels pourraient théoriquement permettre de déduire des habitudes individuelles (ex: "Station République toujours vide à 8h05 → quelqu'un prend vélo régulièrement").

✓ MITIGATION

- Agrégation minimale : 5 min ou 10 vélos (k-anonymity)
- Pas d'affichage temps réel à la seconde
- Données individuelles jamais exposées dans dashboard

Évolution vers ML prédictif

RISQUE MOYEN (Phase 2)

Risque futur (Phase 2) :

Si on ajoute prédictions ML, risque de créer profils comportementaux même sans données personnelles (ex: "Usagers type bureau vs résidentiel").

✓ MITIGATION FUTURE

- AIPD obligatoire avant déploiement ML
- Pas de segmentation discriminatoire
- Comité éthique validation modèles ML

Partage données tiers

RISQUE FAIBLE

Situation :

Dashboard interne Smovengo/IDFM uniquement. Pas de partage avec acteurs privés (assurances, publicité, etc.).

✓ GARANTIE

- Accès restreint : équipes ops + IDFM
- Pas de revente données (interdit service public)
- Open data : seules données agrégées publiques (GBFS déjà ouvert)



STRATÉGIE DE LANCEMENT



Communication

- 1 Présentation projet aux stakeholders**
Réunion kickoff avec Direction Ops + IDFM pour présenter objectifs, timeline, ROI attendu
 Chef de projet Data + Directeur Ops
- 2 Campagne de teasing interne**
Email + affichage bureaux : "Nouveau dashboard ops arrive !" avec mockup et bénéfices clés
 Équipe Communication interne
- 3 Newsletter hebdomadaire post-lancement**
Partager insights découverts, success stories, tips d'utilisation
 Product Owner



Formation

- 1 Formation pilote (Phase 1)**
Session 1h30 avec 2-3 early adopters : navigation, filtres, interprétation KPIs
 Data Analyst + Chef de projet
- 2 Formation groupe (Phase 2)**
Sessions 1h avec 5-8 users : cas d'usage réels, Q&A, best practices
 Champions formés en Phase 1
- 3 Documentation & tutoriels vidéo**
Guide utilisateur PDF + 3 vidéos courtes (2-3 min) : "Comment filtrer ?", "Comprendre SOR/DOR", "Export Excel"
 Product Owner + UX Designer



Préparation Technique

- 1 Tests end-to-end (pré-lancement)**
Valider pipeline data, calculs KPIs, performance dashboard (< 3s chargement)
 Data Engineer + QA
- 2 Setup monitoring & alerting**
Pingdom pour uptime, Sentry pour erreurs, alertes Slack si pipeline fail
 Data Engineer + DevOps
- 3 Backup & rollback plan**
Snapshot DB avant lancement, procédure rollback documentée (< 30 min)
 DevOps + SRE



Suivi & Mesure

- 1 Dashboard d'adoption**
Tracker : nb connexions/jour, temps moyen session, features les plus utilisées
 Product Owner + Data Analyst
- 2 Survey utilisateurs (Semaine 2 & 6)**
NPS, questions ouvertes : "Qu'est-ce qui vous aide le plus ?", "Qu'est-ce qui manque ?"
 Product Owner
- 3 Reporting mensuel Direction**
KPIs métier (SOR -X%, DOR -Y%), adoption (Z users actifs), backlog améliorations
 Chef de projet Data



CALENDRIER DE LANCEMENT

| | S-1 Prépa | S1 Phase 1 | S2 Phase 1 | S3 Phase 2 | S4 Phase 2 | S5 Phase 3 | S6 Phase 3 | S7-8 Suivi |
|----------------------------|--------------|-----------------|---------------|-----------------|---------------|---------------------|----------------|---------------|
| ⚙️ Tests techniques | Tests E2E | | | | | | | |
| 👤 Formation pilote | Prépa docs | Session 1h30 | | | | | | |
| 👥 Pilote (2-3 users) | | Usage quotidien | Feedback | | | | | |
| 🔧 Ajustements UI/UX | | | Corrections | Améliorations | | | | |
| 👤 Formation groupe | | | | Sessions 1h | Q&A | | | |
| 👥 Déploiement (8-9 users) | | | | Usage quotidien | Adoption | | | |
| 🛑 Go/No-Go Phase 3 | | | | | Décision | | | |
| 🔥 Production (30-50 users) | | | | | | Déploiement complet | Adoption large | Monitoring |
| 📊 Reporting Direction | | | | | Mois 1 | | | Mois 2 |

