

---

# Projet professionnel

Daniel DOS SANTOS  
Abdourahmane BALDE  
Simon FANI

# Sommaire

- 1- Introduction & Contexte
- 2- Objectifs
- 3- KPI
- 4- Outils Utilisés
- 5- Gestion de projet
- 6- Matrice SWOT
- 7- Pipeline de donnée
- 8- Machine Learning
- 9- Data Visualisation
- 10- Conclusion



# Introduction & Contexte

- Notre client est un grand compte de vente des vélos qui produit différents types de vélos. Depuis le début de la pandémie en mars 2020, les ventes de vélos ont explosé.
- Les ventes de vélos d'une année sur l'autre pour les 12 mois se terminant en avril 2021 ont augmenté de 57 %, totalisant 6.5 milliards de dollars chez les grands détaillants



# Objectifs

1. Améliorer la rentabilité : Maximiser le chiffre d'affaires et minimiser les coûts.
2. Mettre en place une plateforme Data Lake House : Pour traiter, stocker et analyser les données.
3. Optimisation de l'inventaire : Assurer que les stocks de vélos sont alignés avec la demande.
4. Ciblage des clients et des magasins : Identifier les segments de clientèle et de magasins les plus rentables.
5. Optimisation de la gamme de produits : Identifier les types de vélos qui sont les plus populaires et rentables.
6. Exploiter les tendances saisonnières et géographiques : Identifier les périodes et les lieux où la demande est forte.
7. Création de tableau de bords : Visualisations interactives sur les ventes totales, les ventes par produit, par région.
8. Analyse prédictives des ventes : prédire les tendances futures des ventes pour aider à la planification de la production et de la distribution

# KPI

## 1. Ventes

- a. Ventes Totales
- b. Ventes par produit
- c. Ventes par Région
- d. Croissances des ventes Mois après Mois
- e. Ventes annuelles
- f. Ventes par Trimestre
- g. Ventes par Saison

## 2. Clients

- a. Quantité acheté par client
- b. Montant moyen par commande
- c. Le client qui dépense le plus
- d. Chiffres d'affaire par sexe

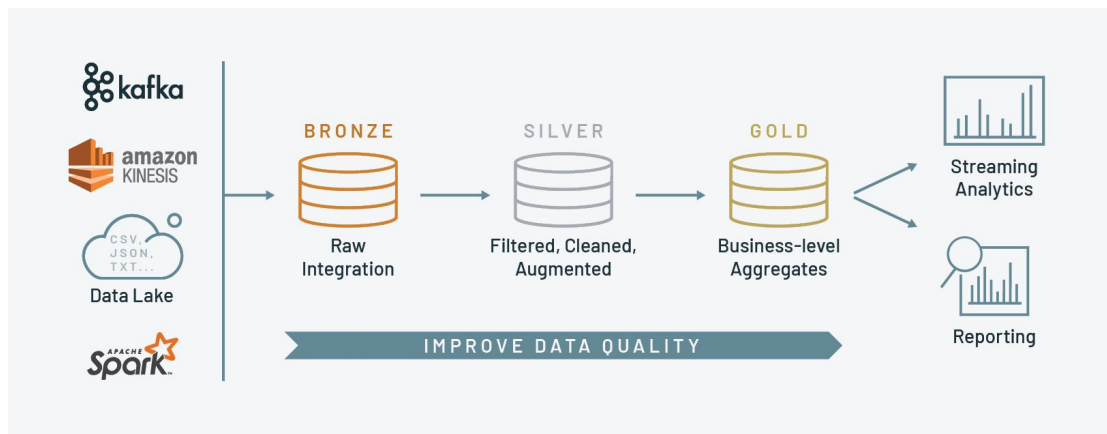
## 3. Produits

- a. Produit le plus vendu
- b. Prix moyen par produit



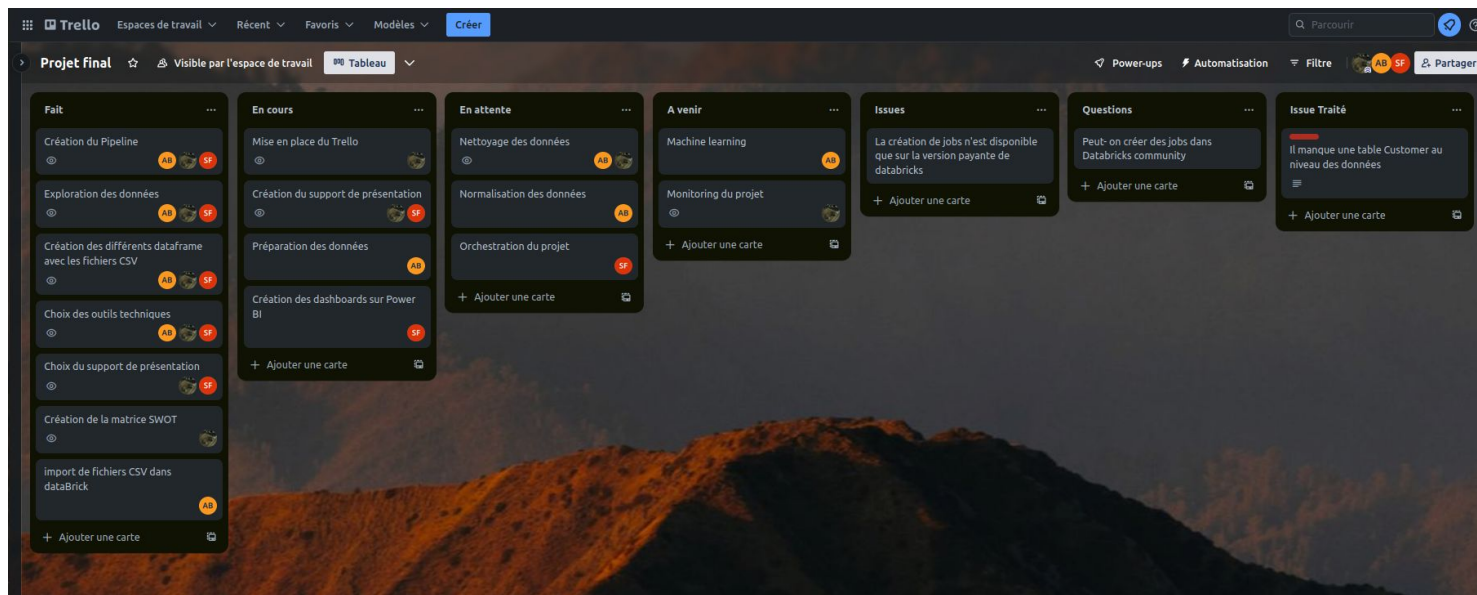
# Outils utilisés

- Databricks (Delta Lake)
- Power BI
- Langage (Python, Spark, SQL)
- Trello
- Teams
- GIT



# Gestion de projet

Lien du Trello : <https://trello.com/invite/b/GO3lQY7b/ATTI5daabbf4f8f898a378cdd7ed99fe12b8C0697777/projet-final>





# Matrice SWOT

## Forces

- Solutions open source limitant les coûts
- Identification améliorée des tendances actuelles du marché

## Faiblesses

- Il existe des solutions payantes plus complètes
- Les données fournies par le clients sont de faible qualité et assez anciennes.

## Opportunités

- Explosion du marché du vélo
- Possibilité d'étendre cette solution au monitoring des employés

## Menaces

- La qualité des données fournis pourrait nuire à la qualité du résultat de nos prédictions



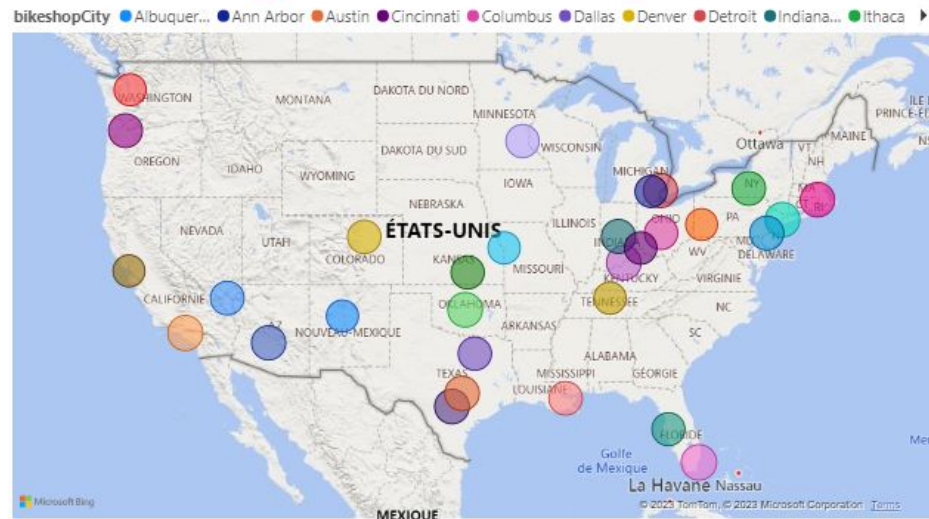
# Démonstration

Pipeline de données et  
machine Learning

# Data visualisation

1. Visualisation sur Power BI
2. Création de différents DashBoards :
  - a. Ventes par région
  - b. Ventes par clients et chiffre d'affaires
  - c. Ventes par périodes
  - d. Prévisions des ventes futures

## Emplacement des magasins



# Conclusion

- Import des fichiers CSV
- Exploration des données
- Nettoyage des données
- Machine Learning
- Data Visualisation avec Power BI