



Data Engineering & Data Visualization

🔍 « Visualiser, c'est transformer des terabytes en décisions — la data viz est l'interface entre l'ingénieur et le décideur. »

Pourquoi chaque Data Engineer doit maîtriser la data viz

La visualisation de données n'est pas réservée aux analystes. Elle devient un outil essentiel dans l'arsenal du Data Engineer moderne, transformant des processus techniques complexes en insights actionnables.

Qui utilise réellement vos données ? Cette question fondamentale révèle pourquoi la maîtrise de la data viz est devenue indispensable pour concevoir des pipelines efficaces et maintenables.

La data viz : un levier stratégique pour les Data Engineers



Validation des pipelines

QA visuelle instantanée pour détecter anomalies, drifts et patterns suspects dans vos flux de données



Diagnostic rapide

Identifier en un coup d'œil latence, erreurs et variations de volume lors d'incidents



Collaboration facilitée

Partager le travail de façon reproductible avec les équipes produit et métier



Monitoring qualité

Interface de surveillance pour la qualité des données et le respect des SLAs

Exemples concrets : visualiser un backfill raté, détecter un pic d'ETL inattendu, ou identifier une dérive de distribution en temps réel.

Les 5 principes clés d'une data viz efficace



01

Simplicité

Un message par graphique — éviter la surcharge cognitive

02

Lisibilité

Titres explicites, unités claires, légendes immédiatement visibles

03

Hiérarchie visuelle

KPI en haut/gauche, détails en bas/droite — guider le regard

04

Constance

Mêmes couleurs et formats pour mêmes mesures à travers tous vos dashboards

05

Contextualisation

Comparaisons vs baseline, annotations pour événements clés

💡 Adapter le niveau de détail technique selon l'audience — moins de jargon pour les non-techniques.

Règles rapides : choisir le bon type de graphique

Le choix du graphique doit toujours être guidé par la question métier que vous cherchez à répondre. Voici un guide de référence basé sur les cas d'usage les plus fréquents.



Tendance temporelle

Line chart / Area chart — Idéal pour suivre l'évolution dans le temps, identifier patterns et saisonnalité



Comparaison catégories

Bar chart / Grouped bars — Comparer des valeurs discrètes entre différentes catégories



Distribution

Histogram / Boxplot — Analyser la répartition et identifier outliers



Composition

Stacked bar / Donut — Part du tout (utiliser avec prudence pour ne pas surcharger)



Corrélation

Scatter plot — Relations entre variables, avec régression si nécessaire



Données géospatiales

Map / Choropleth — Visualiser distributions et patterns géographiques

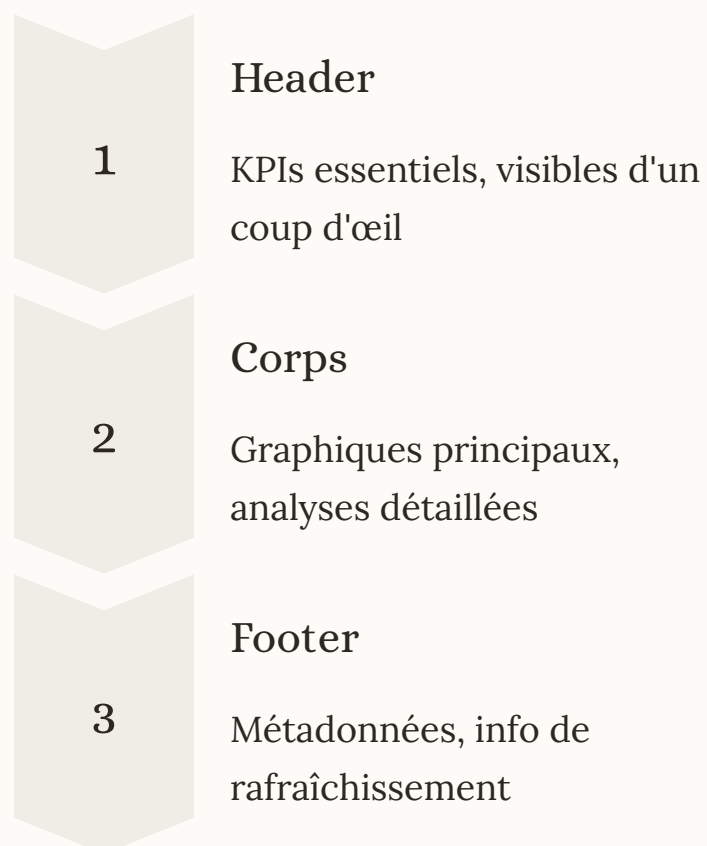
Anatomie d'un dashboard bien organisé

Patterns de lecture


Les utilisateurs suivent naturellement un **Z-pattern** ou **F-pattern** lors de la lecture. Exploitez ce comportement :

- KPIs clés en haut pour impact immédiat
- Filtres à gauche pour accès facile
- Détails et graphes secondaires à droite


Architecture en 3 zones



Types de filtres

 **Filtres globaux**

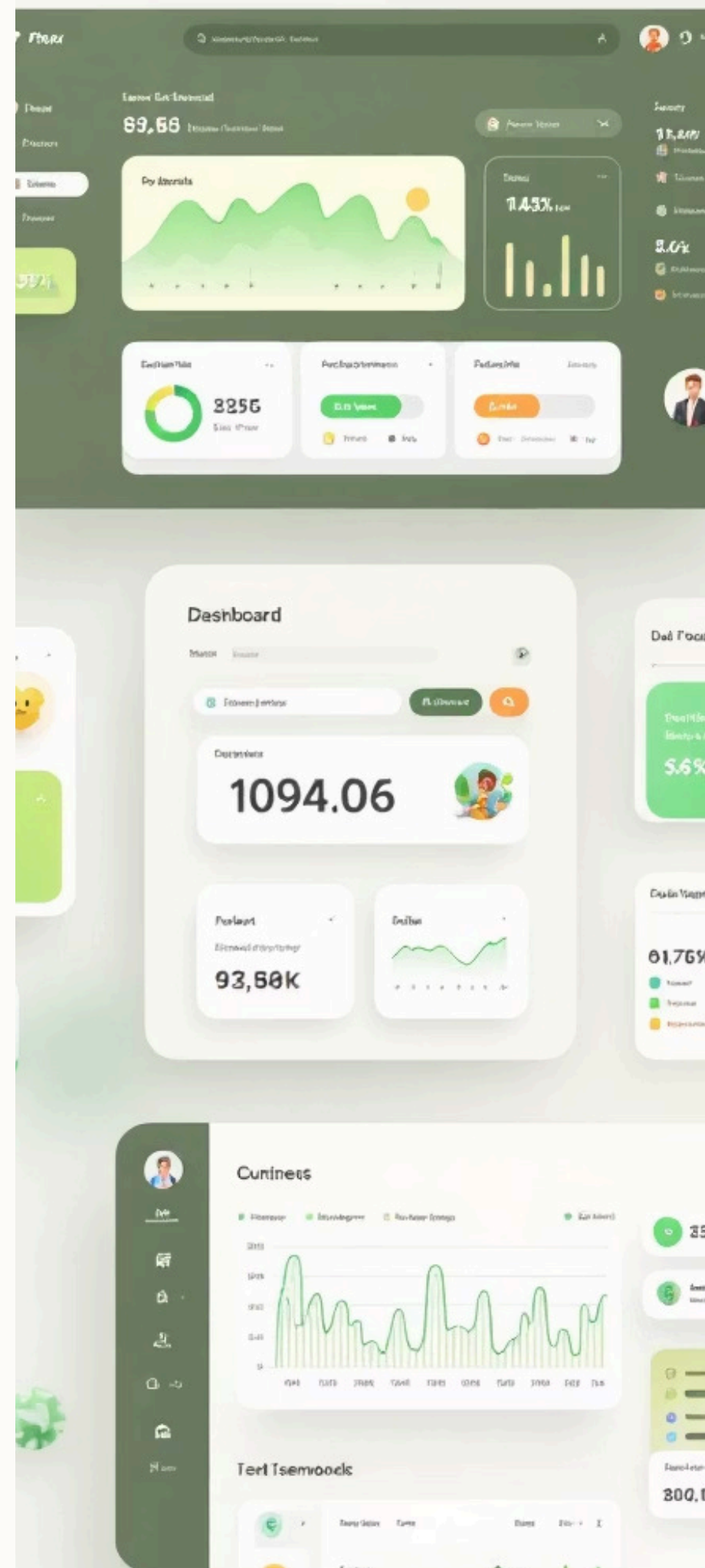
Impactent l'ensemble du dashboard — période temporelle, environnement, région

 **Filtres locaux**

Affectent un seul graphique — zoom sur données spécifiques

Interactivité

- Drill-down pour explorer les détails
- Hover tooltips pour contexte additionnel
- Time range picker pour analyses temporelles



KPI Engineering : quoi afficher et comment

6

Indicateurs essentiels

Métriques critiques à monitorer en permanence pour l'analyse et les modèles

KPIs clés pour Data Analysts & Scientists

- **Performance des Modèles** — Accuracy, Precision, Recall, F1-score
- **Métriques Business** — Taux de conversion, Churn, LTV (LifeTime Value)
- **Indicateurs d'Exploration** — Distribution des features, Corrélation entre variables
- **Qualité Prédictive** — Détection de drift, Importance des features
- **Validation des Résultats** — Résultats d'A/B tests, Signification statistique
- **Coût & Complexité** — Coût de calcul/stockage des features, temps d'inférence

Présentation optimale

Chaque KPI doit inclure :

- **Valeur actuelle**
Chiffre principal mis en avant
- **Variation**
Comparaison vs période précédente (%, delta)
- **Sparklines**
Micro-graphique de tendance contextuelle

❏ **Important** : Pour les KPI dérivés, documenter la formule de calcul et la fréquence de rafraîchissement. Implémenter des seuils colorés (rouge/orange/vert) pour des alertes visuelles instantanées, essentielles pour monitorer la santé des dashboards business ou des modèles prédictifs.

Mauvaises pratiques : les pièges à éviter

❌ Surcharge cognitive

Trop de graphiques sur une seule page — l'utilisateur ne sait plus où regarder. **Solution** : maximum 4-6 visualisations par vue, organiser en onglets si nécessaire.

❌ Graphiques inadaptés

Graphiques 3D, camemberts pour comparaisons fines, gauges décoratives. **Solution** : privilégier simplicité et clarté — bar charts pour comparaisons.

❌ Palette de couleurs problématique

Manque de contraste, trop de couleurs différentes, pas de considération pour daltonisme. **Solution** : utiliser palettes accessibles, 5-7 couleurs max.

❌ Absence de métadonnées

Pas de source, date de rafraîchissement inconnue. **Solution** : footer systématique avec timestamp et source de données.

❌ KPI mal définis

"Erreurs" sans préciser le périmètre ou la méthode de calcul. **Solution** : documenter chaque métrique avec définition, formule, et business logic.

Priorisation des types de graphiques

Adapter la complexité visuelle au niveau d'expertise de votre audience. Commencez simple, augmentez la sophistication uniquement si nécessaire.



Priorité 1 : Universels

Line, Bar, Stacked bar (modération), Table, KPI card

Simple, immédiatement compréhensibles par tous. À utiliser en priorité pour 80% des cas.



Priorité 2 : Intermédiaires

Heatmap, Scatter, Histogram, Area chart

Nécessitent une familiarité avec l'analyse de données. Parfait pour audiences techniques.



Priorité 3 : Avancés

Sankey, Network graph, Small-multiples complexes, Interactive choropleth

Réservés aux experts et cas d'usage spécifiques. Requièrent formation et contexte.

❏ **Règle d'or :** En cas de doute, choisir la visualisation la plus simple qui répond à la question métier. La sophistication visuelle ne doit jamais compromettre la clarté du message.

Panorama des outils BI & frameworks

Choisir l'outil adapté à votre contexte technique, vos contraintes budgétaires et le niveau d'expertise de votre équipe.

Outil	Usage typique	Courbe apprentissage	Scalabilité / Coût	Quand l'utiliser
Power BI	Dashboards entreprise, intégration MS	Facile → Intermédiaire	Licence payante, robuste corporate	Rapports internes, écosystème MS365
Tableau	Analytique visuelle avancée	Intermédiaire	Payant, très robuste	Analysts, exploration visuelle poussée
Superset / Metabase	Open-source, déploiement rapide	Faible → Moyen	Faible coût, auto-hébergé	Projets internes, prototypage, POCs
Streamlit / Dash	Apps Python custom, prototypes	Moyen (code Python)	Flexible, scale via infra	Logique personnalisée, intégration ML
R Shiny	Apps R interactives	Moyen (stack R)	Flexible (écosystème R)	Data scientists R-first, stats avancées

Intégrations essentielles

- Connexions DB : Postgres, BigQuery, Snowflake, Redshift
- Versioning Git pour dashboards as code
- CI/CD pour déploiement automatisé

