

EDA - Desk

Plateforme d'Analyse Exploratoire de Données

Exploratory Data Analysis

Présentation de l'Application





Vue d'ensemble

EDA - Desk PRO HYBRID est une application desktop sophistiquée dédiée à l'analyse exploratoire de données (EDA). Conçue avec une architecture modulaire en 4 zones, elle offre une expérience utilisateur fluide et professionnelle pour l'analyse de fichiers CSV et Excel.

Philosophie de conception : Architecture hybride combinant l'efficacité d'une interface multi-zones avec un système d'accumulation progressive des résultats, permettant la construction incrémentale de rapports d'analyse complets.

Architecture à 4 Zones

L'interface principale s'articule autour d'une organisation spatiale optimisée :

- | | |
|--|---|
|  Zone 1 — Barre supérieure
Chargement de fichiers et métriques clés |  Zone 3 — Analyses et contrôles
Panneau d'outils d'analyse statistique |
|  Zone 2 — Aperçu des données
Prévisualisation des 50 premières lignes |  Zone 4 — Résultats rapides
Affichage synthétique des résultats |

Onglets et Fonctionnalités

L'application propose 6 onglets spécialisés :

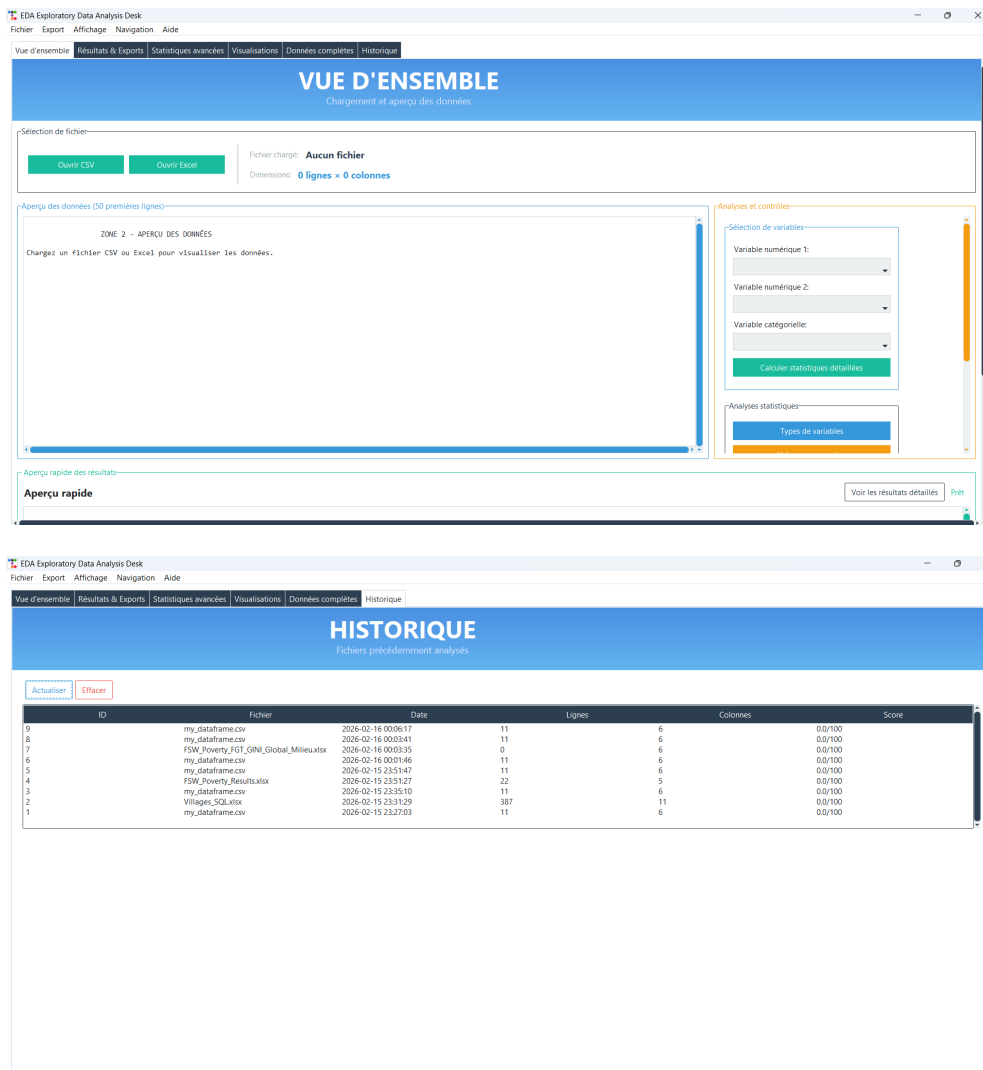
- Vue d'ensemble — Interface principale avec architecture 4 zones intégrée
- Résultats & Exports — Accumulation progressive des analyses avec exports professionnels (Word, PDF, Excel)
- Statistiques avancées — Support complet des variables numériques ET catégorielles avec calculs détaillés
- Visualisations — Générateur de graphiques (histogrammes, boxplots, nuages de points, matrices de corrélation)
- Données complètes — Explorateur tabulaire avec vue complète du jeu de données
- historyHistorique — Traçabilité des analyses avec base SQLite persistante

Capacités d'Analyse

L'application offre un ensemble complet d'outils d'analyse :

- **Détection automatique des types** — Classification intelligente en variables numériques, catégorielles et booléennes
- **Valeurs manquantes** — Identification et quantification avec seuils de criticité
- **Variables constantes** — Détection des variables quasi-constantes (>95% d'une modalité)
- **Outliers (IQR)** — Détection par méthode des quartiles avec comptage et pourcentages
- **Statistiques catégorielles** — Modalités dominantes, effectifs, pourcentages
- **Score de qualité** — Évaluation synthétique sur 100 points

Point clé : Le système d'accumulation permet de construire un rapport unique contenant toutes les analyses successives, éliminant la nécessité de répéter les opérations.



⚙️ Fonctionnement Détaillé

Workflow d'Utilisation

1. **Chargement** — Importation d'un fichier CSV ou d'un fichier Excel avec détection automatique du séparateur (virgule, point-virgule, tabulation, etc.)
2. **Détection automatique** — Classification des variables et affichage des métriques de base (dimensions, types)
3. **Analyses multiples** — Exécution successive des analyses depuis la Zone 3, chaque résultat s'ajoutant automatiquement au rapport cumulatif
4. **Consultation** — Navigation entre les onglets pour explorer données brutes, statistiques et visualisations
5. **Export** — Génération de rapports professionnels incluant l'intégralité du contenu accumulé

Technologies Utilisées

- Python 3.10+
- tkinter (GUI native)
- ttkbootstrap (thèmes modernes)
- pandas (manipulation données)
- numpy (calculs numériques)
- scipy (statistiques avancées)
- matplotlib / seaborn (viz)
- SQLite (historique)
- python-docx (export Word)
- ReportLab (export PDF)
- openpyxl (export Excel)

[Insérer ici capture d'écran de l'onglet Résultats avec analyses accumulées]

! Limites et Axes d'Amélioration

Limites Actuelles

- **Format de données** — Supporte uniquement CSV et Excel; pas de support natif pour , JSON, Parquet ou bases de données format Stata
- **Volume** — Performance non optimisée pour les jeux de données massifs (>500k lignes) en raison de l'absence de traitement par chunks
- **Visualisations** — Palette limitée de graphiques; pas d'interactivité avancée (zoom, filtres dynamiques)
- **Statistiques** — Absence de tests d'hypothèses formels (normalité, t-test, ANOVA, χ^2)
- **Export** — Les exports PDF/Word n'intègrent pas encore les graphiques générés
- **Collaboration** — Aucune fonctionnalité de partage ou de travail collaboratif

Pistes d'Amélioration

Court terme

- Support formats multiples
- Intégration graphiques dans exports
- Sauvegarde de sessions
- Templates d'analyse

Moyen terme

- Tests statistiques avancés
- Visualisations interactives (Plotly)
- Gestion big data (Dask)
- API REST pour automatisation
- Machine Learning intégré

Roadmap Stratégique

Vision à long terme : Transformer EDA-Desk en plateforme complète d'analyse de données avec capacités de nettoyage automatisé, feature engineering, et intégration de modèles prédictifs, tout en conservant la simplicité d'utilisation actuelle.