



# VANESSA KENNICHE SANOCKA

## Data Science Engineer



[KENSAN.VAN@GMAIL.COM](mailto:KENSAN.VAN@GMAIL.COM) | +33 7 68 14 01 74 | [vanessa-kenniche-sanocka.github.io](https://vanessa-kenniche-sanocka.github.io)

Paris – Native French, Native Polish, English C1 (TOEIC)

### Expertise

Specialized in quantitative economics and public policy analysis. Project leadership: econometric modeling, large-scale and sensitive data processing, machine learning, translating complex analyses into accessible recommendations for stakeholders. Programming and innovative visualization solutions for evidence-based decision-making.

### Skills

**Modelling** : Time series, panel data, event studies, causal inference, ML, geospatial analytics, networks

**Languages & Data** : Python, R, SQL, Dplyr, Pandas, Plotly, Scikit-learn, PyTorch, TensorFlow, Airflow, NLP

**Tools** : Shiny, Dash, Power BI, Excel, Quarto, Git, Docker, Apache Spark, MySQL, MongoDB, QGIS, CASD

### Professional Experience – French Court of Audit, 2023 - Present

- **Franco-Swiss cross-border geospatial analysis (Swiss Court of Audit)** : Harmonization of heterogeneous GIS data across France-Switzerland to identify cross-border regulatory discontinuities. Translation of geospatial analyses into visualizations and public policy recommendations for institutional teams.
- **Urban transport pricing policy evaluation** : Panel models & event studies (national GART, CEREMA datasets) and survey research (confidential user sampling, propensity score adjustment, margin calibration, bootstrapping), enhanced with FILOSOFI database through complementary research, to formulate transport pricing policy recommendations with interactive delivery via Shiny.
- **NLP and RAG pipeline** : Full-stack integration for semantic exploration and querying of Court of Audit documents through Elasticsearch indexing, embeddings, LLM, Sentence Transformers and API to Next.js platform, including migration of Court data to new internal archiving system.
- **Predictive modeling on national confidential registries (CASD)** : ML algorithms on judicial data for survival analysis and risk scoring. Statistical modeling on tax and employment data (EEC) for workforce policy evaluation.

### Academic and Personal Experience – Master's in Data Science Engineering for Social Sciences, University of Paris Nanterre, 2020 - 2025

- **Banking fraud detection** : Supervised modeling on imbalanced data, SMOTE/B-SMOTE techniques, cross-validation for financial risk scoring.
- **Algorithmic market pricing detection – with EconomiX laboratory** : High-frequency scraping to data lake, classification and estimation of price adjustments through fixed-effects models and robustness testing.
- **Innovation determinants benchmarking (South Asia)** : Cross-country econometric modeling (logit/Poisson) on survey data with robustness testing through Lasso regularization.
- **Parliamentary voting network analysis** : Government open data scraping, Manhattan distance matrices and spring layout clustering to identify cross-party coalitions.
- **Demographic Power BI pipeline** : MySQL star schema data warehouse, windowing queries (moving averages), territorial dashboard for vacant private housing policy indicators.
- **PATSTAT patent analysis** : Competitive risk mapping and sectoral strategic positioning with statistical and geospatial visualizations for industrial policy recommendations.

### Additional Information

- STEM Ambassador for "Filles et Maths" program : promoting scientific careers to high school students.
- Development of training tools and collaborative platform for Court of Audit Data Scientists.