

Ashwathy Ashokan

Data Analyst | Senior Statistical Programmer

 ashwathyreddy.97@gmail.com

 [+1 \(437\) 669 1129](tel:+1(437)6691129)

 [Ontario](#)

 [LinkedIn](#)

PROFESSIONAL SUMMARY

Senior Statistical Programmer pivoting to Data Analytics & ML with 5+ years in clinical data programming. Expert in SAS and CDISC (SDTM/ADaM, TLFs) with strong Python, SQL, and PySpark. Hands-on with Kafka, Spark Streaming, XGBoost, MongoDB, Airflow, Docker, and AWS. Deliver audit-ready outputs, automate pipelines, and communicate clearly with cross-functional teams.

SKILLS

- Programming & Data:** SAS Base/Advanced/Macros (5+ yrs), SQL (3+), Python (2+), PySpark (1+), R (2+)
- ML & Analytics:** EDA, feature engineering, regression, clustering, model validation
- Big Data & Cloud:** Spark & Spark Streaming, Kafka, Hadoop, MongoDB, Airflow, AWS (EC2), Docker, Git
- BI & Visualization:** Tableau (1+), Power BI (1+), Excel (Power Query, PivotTables)
- Clinical Standards:** CDISC SDTM/ADaM, Define.xml, Pinnacle 21, FDA submissions, GCP, 21 CFR Part 11
- Collaboration:** Agile/Jira, Git, Confluence, documentation, mentoring, SOP/QC

WORK EXPERIENCE

Senior Statistical Programmer

Agati Clinical Informatics - India

Sep 2019 – Jan 2025

- Led end-to-end statistical programming for Phase I–III trials, partnered with Biostatistics and Data Management to resolve data issues and ensure traceability.
- Designed, derived, and validated SDTM and ADaM datasets, authored programming specifications aligned to SAPs and CDISC standards.
- Produced and QC-reviewed Tables, Listings, and Figures (TLFs), ensured consistency with analysis plans and submission requirements.
- Built reusable SAS macro libraries and checks, reducing manual effort and cycle time by ~30% and improving reproducibility.
- Owned submission readiness:** Define.xml creation, Pinnacle 21 conformance.
- Mentored junior programmers, established code reviews, reusable spec templates, and documentation best practices.
- Supported global clients as primary/validation programmer, meeting quality SLAs and timelines.
- Therapeutic Areas:** Oncology, Endocrinology, Cardiovascular.

PROJECTS

iBrand - AI/NLP Big-Data Pipeline for Trend-Aware Social Posts

Tech: PySpark · Kafka · Spark Streaming · MongoDB Atlas · FastAPI · Docker · Transformers (BART, DistilBERT, Sentence-Transformers)

Repo: <https://github.com/LCM-S25-3035/iBrand>

- Engineered a streaming pipeline:** multi-source news scrapers → Kafka (news-articles) → Spark Streaming for cleaning, NER, topic/sentiment, summarization → MongoDB.
- Containerized stack with Docker Compose (Kafka, Spark master/workers, Mongo, Elastic), designed FastAPI backend, scoped React UI for brand-tone generation.
- Output enables automated, brand-voice content generation from enriched news for marketers.

Realtor's Housing Price Prediction - Canada

Tech: Python · XGBoost · MongoDB · AWS EC2 · Airflow · Twilio · Next.js UI

Repo: https://github.com/Lambton-AI-Group/house_price_prediction_project

- Automated weekly Kijiji/Realtor.ca scraping, cleaning, and model training with XGBoost to forecast prices and flag undervalued listings.
- Deployed Airflow DAGs on AWS EC2 orchestrating ingest → train → predict → notify, integrated Twilio SMS alerts and Next.js UI for real-time insights.
- Delivered actionable alerts and feature-importance signals to support pricing decisions.

EDUCATION

Postgraduation in Big Data Analytics

Lambton College, Mississauga, ON

May 2024 – Dec 2025

Bachelor of Engineering in Fire Technology & Safety

Noorul Islam University, India

Jun 2014 – May 2018

CO-CURRICULAR ACTIVITIES

- **College Editor Executive - Annual Magazine:** Led content curation and layout, used LaTeX/InDesign for publication-ready formatting.
- **Symposium Assistant Secretary:** Coordinated department symposium logistics, speaker scheduling, and communications.
- **PHUSE Associate Member (Clinical Trials Standards):** Active in SDTM/ADaM community discussions and webinars.
- **Volunteer:** Fire & Safety training initiatives, clinical data compliance awareness training.