Ameya Damle

+44 7767956960 | ameyadamleuk@gmail.com | LinkedIn | GitHub | Medium | Reading, UK

Professional Summary

Data Scientist and Engineer with experience in designing and optimizing geospatial data pipelines, spatial analytics, and enterprise-level data solutions. Proven ability to integrate geospatial intelligence into **insurance risk modelling, catastrophe analysis, and exposure assessment**. Adept in Python, PostGIS, and cloud computing, delivering efficient and scalable geospatial solutions. Passionate about leveraging spatial data for **risk mitigation, pricing optimization, and strategic decision-making in the insurance industry**.

EDUCATION

University of Reading, Reading, UK **September 22 – September 23**
Master of Data Science and Advanced Computing Merit (2:1)
Coursework: Data Analysis, Machine Learning, Big Data, Cloud computing, Artificial Intelligence

University of Mumbai, India **June 16 – November 20**
Bachelor of Science in Statistics CGPA: 6.13
Coursework: Hypothesis Testing, Probability Theory, Bayesian Theory, Estimation Theory, Actuarial Science

TECHNICAL SKILLS

Geospatial Analytics and Tools: GDAL, GeoPandas, Shapely, Rasterio, QGIS, ArcGIS, FME
Programming & Scripting: Python, SQL, JavaScript (Leaflet.js, D3.js), Shell Scripting
Cloud & Big Data: AWS (S3, Lambda, RDS, Athena), Google BigQuery, Apache Spark
Database Management: PostgreSQL/PostGIS, MySQL, MongoDB
Machine Learning: Scikit-learn, XGBoost, Spatial Clustering, Kriging, Spline Interpolation
Visualization Tools: Power BI, Tableau, Matplotlib, Plotly, Folium
Other Tools: Docker, Git, Apache Airflow, Agile/Scrum

PROFESSIONAL EXPERIENCE

Technical Author, Medium **July 24 – Present**
Technologies: Python, Git, Power BI, MS Excel

- Authored detailed, beginner-friendly tutorials and technical articles on advanced data science and actuarial topics, like **Spatial Analysis** for Insurance data, **Actuarial Price Modeling**, Time Series forecasting with **Facebook API**, and Claim Reserving with **Chain ladder**.
- Currently developing content for **Mapping popular tourist attractions in London using DBSCAN**, detailing how to integrate API for scalable and automated machine learning models.
- Articles featured advanced use cases, clear visuals, and practical insights, receiving positive feedback for clarity and actionable value from readers.

Geospatial Data Engineer, Blackmont Consulting **April 24 – June 24**
Technologies: OSM, LiDAR, AWS, BigQuery

- **Developed automated data pipelines** to process, clean, and validate geospatial datasets (e.g., flood zones, OpenStreetMap, LiDAR).
- Enhanced **geospatial risk modeling accuracy** by developing spatial smoothing and Kriging-based methods, improving predictive power by 20%.
- **Developed efficient ETL pipelines** by leveraging GDAL utilities to preprocess, reproject, and merge large scale raster datasets, including the creation of virtual rasters (VRT) for optimal file management.
- **Implemented advanced compression techniques** with GDAL to reduce map tile file sizes, significantly enhancing storage efficiency and processing speed.
- **Automated complex geospatial workflows** by integrating GDAL with Python, enabling streamlined data manipulation and analysis for high-impact projects.
- Automated the **containerization of machine learning models with CI/CD**, speeding up deployment and improving model consistency across environments.
- Created a **geospatial data validation framework** that detects **topology errors**, validates CRS consistency and data completeness, auto-reprojects datasets, and generates visual error reports—supporting formats like Shapefiles and GeoJSON.
- Worked on **cloud-based GIS architectures**, using AWS S3, RDS, and BigQuery to handle large-scale geospatial datasets.

Data Engineer, Cognizant **January 24 – April 24**

Technologies: Spark, Scala, CI/CD, Git, DevOps, Airflow, Bash, Kubernetes, databricks, MLFlow

- **Designed and implemented ETL** pipelines using Spark/Scala, processing over 10 million bank records per week. Streamlined data integration from diverse sources, optimizing efficiency and ensuring data accuracy.
- **Developed a streaming API** using databricks to serve real-time predictions powered by **MLFlow models**, enabling quicker and more reliable decision-making across teams.
- **Achieved a 10% increase** in team productivity by introducing CI/CD pipelines for building and deployment automation as part of the Dev Practices.
- Established a **Git strategy that reduced the number of merges** to the master branch by 3, ensuring a smoother development process and higher code quality.
- **Implemented Bash scripts for environment configuration** across development, UAT, and production, contributing to seamless deployment. Utilized Kubernetes for efficient application deployment and management.
- Managed **artifact storage and automation for infrastructure components**, ensuring efficient version control and deployment consistency.
- Collaborated **with actuaries and underwriters** to integrate geospatial insights into pricing models and claims analytics.
- Implemented **geospatial fraud detection algorithms**, identifying high-risk claim clusters and improving fraud detection rates by 20%.

GIS Data Analyst, British Airways **September 23 – December 23**

Technologies: Python, R, SQL, SAS, Leaflet.js, Folium, Power BI, Tableau

- **Performed geospatial risk assessment** by analyzing insurance claims data and overlaying it with hazard maps.
- Developed **geospatial reports and dashboards** using Power BI and Tableau to help insurers visualize high-risk zones.
- **Built interactive maps and geospatial applications** for internal teams using **Leaflet.js** and **Folium**.
- Mapped **fraudulent claims using clustering techniques**, identifying **15% increase** in fraud-prone areas.
- **Automated data cleaning scripts** to ensure the accuracy of location-based datasets in insurance underwriting.

Junior GIS Analyst, Data Glacier **June 23 – September 23**

Technologies: Python, R, SQL, SAS, Folium, Power BI, Tableau

- **Designed spatial machine learning models** to predict claim severity based on geographic risk factors.
- **Developed interpolation techniques** (Kriging, IDW) for geospatial hazard prediction.
- Implemented **spatial feature engineering techniques,** improving model performance by 20%.
- Contributed to **geospatial fraud detection research,** identifying patterns in claim submissions.

Actuarial Intern, HDFC ERGO General Insurance **February 22 – July 22**

Technologies: Python, R, SQL, scikit-learn, statsmodel, QGIS, SAS, MS Excel

- Applied statistical tests, such as Chi-Square goodness-of-fit-test, determine and flag fraudulent claims if the observed distribution significantly differs from the expected Benford's Law distribution.
- Utilized expertise in writing **SQL queries** and **SAS** to **build macros**, and in building and testing GLM models.
- **Analyzed and validated model performance** through residual analysis and cross-validation techniques, leading to 30% improved accuracy in predicting claim severity.
- Conducted severity modeling for a motor third party insurance portfolio, utilized a Gamma distribution to model claim severity, ensuring accurate prediction of claim payouts.
- Calculated **relativity factors for risk segmentation**, improving model accuracy by 20% through effective feature engineering and validated model performance against real-world claims data.
- Developed and performed a **Spatial Smoothing exercise to improve district cluster accuracy by 30%**, enhancing the assessment of motor insurance risk in neighboring geographical areas. The result of this clustering exercise was used in the pricing model.
- Designed **extract, transform, load pipeline**, performed **spatial joins** and summary statistics by using processing toolbox and automated workflows using model designer in **QGIS**.
- Implemented machine learning algorithms such as **logistic regression (sigmoid function), decision trees** for fraudulent claim detection, and reduced false positives by 50%.
- Designed and implemented **GLM** and **CANN** to model claim frequency for **motor pricing** insurance portfolio using **python, R** and **SAS** for model building and validation.
- Experience using actuarial WTW software like Radar and Emblem.

ACADEMIC PROJECTS

Catastrophe Risk Modeling Pipeline | Python, PostGIS, GDAL, Rasterio, Power BI

- Built a **geospatial data pipeline** to ingest and preprocess satellite imagery, flood maps, and historical claims data.
- Developed a **clustering algorithm** to identify flood-prone regions, integrating results into pricing models.

Spatial Smoothing on UK Traffic Accidents/Claims Data | R, Python, Geopandas, shapely, scipy.spatial

- **Developed a spatial smoothing model** for UK accident data using advanced techniques such as centroid-based KDTree analysis, Gaussian weighting with bandwidth optimization, and geostatistical approaches like Kriging and Markov Random Fields.
- **Enhanced accident hotspot identification** by incorporating spatial dependencies and neighbor influence, leading to more robust and interpretable clustering and spatial analysis results.

Property Exposure Analysis for Insurers | Python, QGIS, Tableau, PostgreSQL

- Analyzed geospatial datasets (building footprints, flood zones) to assess property exposure across **Europe** regions.
- Automated **risk factors** extraction for **10 million** policies, enhancing underwriting efficiency.

Large-Scale Geospatial ETL Pipeline | Apache Airflow, AWS Lambda, PostgreSQL/PostGIS

- Designed an **automated data pipeline** for processing government hazard maps and insurance claim datasets.
- Achieved a **50% improvement in processing time** by optimizing geospatial data ingestion.

CERTIFICATIONS

Cognizant - Artificial Intelligence Job Simulation
British Airways - Data Science Job Simulation
Coursera - Introduction to Structured Query Language (SQL)
Free Code Camp - Legacy JavaScript Algorithms and Data Structures
Ligency – Python A – Z: Python for Data Science with Real Exercises
Ligency – R Programming: Advanced Analytics in R for Data Science
Ligency – R Programming A – Z: R for Data Science with Real Exercises