Kanish Godani

872-279-4605 | kgodani2@illinois.edu | LinkedIn | Github | Portfolio

Education

University of Illinois Urbana-Champaign — Champaign, IL
Master of Science in Information Management | GPA - 4.0 — Aug 2024 - May 2026
Coursework: Machine Learning & Cloud, Methods of Data Science, Data Warehousing & Business Intelligence

SVKM's NMIMS University — Mumbai, MH
Bachelor of Technology in Computer Engineering | GPA - 3.7 — Aug 2020 - May 2024

Experience

Data Science Intern — Jan 2026 - May 2026 | Jun 2025 - Aug 2025
Second Wind Pro — New York, NY

- Developed Python-based data validation processes using Pandas and NumPy on 8,500+ NFL player-season records, applying missing value imputation and Z-score outlier treatment to reduce noise and invalid records by 20%
- Engineered features across 15,000+ FBS player records, developing per-opportunity rates, position composite scores, and percentile rankings using Scikit-learn, creating college-to-NFL feature bridge with $\geq 95\%$ completeness
- Trained an XGBoost regression model for predictive modeling on NFL player data to predict cap_hit_pct using time-series cross-validation and hyperparameter tuning, achieving R^2 of 0.78 and MAE $\leq 1.5\%$
- Validated valuation predictions through backtesting against external benchmarks, achieving rank correlation ≥ 0.65, and applied SHAP-based interpretability to identify position-level value drivers for stakeholder decision-making

Data Analyst Intern — Jan 2024 - Jun 2024
Symbo SouthAsia Enterprises Private Limited — Mumbai, MH

- Extracted 20,000+ raw customer profiles from Snowflake using SQL window functions and aggregations, applying deduplication and standardization logic to improve data quality by 40% for customer segmentation analysis
- Automated anomaly detection and claim frequency trend reporting across 50,000+ policy records using Python (Pandas, NumPy, Matplotlib), replacing manual Excel workflows and saving 10+ hours of weekly analysis time
- Applied K-Means clustering to segment insurance customers into 4 behavioral cohorts (Silhouette Score: 0.78), translating insights into targeted product recommendations that enabled a 35% uplift in customer engagement
- Built 10+ interactive Power BI dashboards with real-time KPI tracking and scheduled refresh cycles, delivering actionable policy performance insights to product and marketing teams that improved policy adoption rates by 15%

Projects

Vibe Recipes - AI Recipe Recommendation App — Jan 2026 - Mar 2026

- Implemented a data pipeline to ingest, clean, normalize 150k+ recipes, applying feature engineering (vibe tagging) and generating 384-dimensional embeddings using Sentence-Transformers for semantic search in pgvector database
- Constructed a RAG-based recommendation system using FastAPI, vector similarity search with LLM generation (Groq, Cerebras) and post-retrieval filtering to improve relevance by 75% and achieve sub-2s query latency

Forecasting U.S. Electricity Generation Efficiency — Aug 2025 - Dec 2025

- Forecasted U.S. state-level electricity efficiency (RMSE: 1.96) by building a machine learning pipeline to ingest 4.3M+ time-series records from EIA API, using AWS (S3, SageMaker) to train and deploy TimeGPT and Prophet

Exploratory Analysis and Visualization of Music's Impact on Mental Health — Jan 2025 - May 2025

- Visualized 5,000+ MxMH survey responses in Tableau using correlation and regression analysis across music and mental health variables, surfacing genre-specific impact patterns that improved music therapy accuracy by 30%

Operational Analytics and Forecasting for Multi-Facility Optimization — Sep 2024 - Dec 2024

- Processed 30,000+ operational records across 20 facilities using Power Query, developing time-series trend analysis and Power BI KPI dashboards that identified cost inefficiencies and improved production efficiency by 10%

Technical Skills

Programming Languages: Python, SQL, R, C++, Java
Libraries & Frameworks: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, TensorFlow, PyTorch, PySpark
BI & Visualization Tools: Power BI, Tableau, Microsoft Excel (VBA, Macros), Looker Studio, Power Query, Git
Cloud & Databases: PostgreSQL, Snowflake, AWS (S3, Redshift), BigQuery, MySQL, MongoDB