# Yash Kantharia

+1-551-344-8949 • ykantharia27@gmail.com • linkedin.com/in/yashkantharia • yashkantharia.github.io

## SKILLS

**Programming Skills:** Python, SQL, Java, JavaScript, C, C++, HTML, CSS
**Data Analytics:** Advanced SQL, Tableau, Snowflake, BigQuery, Matplotlib, Seaborn, Google Analytics, Power BI, Excel, Pandas, NumPy
**Cloud:** Amazon Web Services, Google Cloud Platform, Azure, Dremio, Kubernetes, Docker, Databricks, DBT, Airflow, Firebase
**API & Integrations:** RESTful APIs, OAuth, JWT, SAML, Webhooks, SDK Integration & Debugging, Postman, Swagger, Flask, FastAPI
**Machine Learning & AI:** TensorFlow, PyTorch, Scikit-learn, NLP, LLM Fine-Tuning, ChatBots, Classification, Regression, Clustering, Time Series Forecasting, GPT, Gemini, Llama, BERT, TFIDF, Word2Vec, Doc2Vec, RAG, Vertex AI, n8n, LangChain
**Software Tools:** Jupyter Notebook, Git, Android Studio, Unity, JIRA, Splunk, CleverTap, Linux, Unix, Shell Script, Putty, SCP, Trello, Xcode
**Web Scraping:** Selenium, BeautifulSoup
**Data Security & Compliance:** IAM, GDPR, CCPA

## WORK EXPERIENCE

### Data Engineer Academy, New York, NY                    November 2024 - Present
*EdTech platform delivering industry-aligned curriculum and projects in Data Engineering, Analytics, and Artificial Intelligence.*
#### Data Engineer
- Spearheaded technical coaching for aspiring data professionals, delivering hands-on training in **SQL, Python, AWS, Azure, GCP,** and **System Design**, resulting in a 90% student project completion rate.
- Designed and produced learning materials and capstone projects centered on **Data Analysis, ETL pipelines, and Cloud Architecture, Artificial Intelligence** (Fine-tuning LLMs, RAGs, Agents), directly enhancing student technical proficiency.
- Led live debugging sessions to troubleshoot complex code issues in **Data Models, Python scripts, Analytical Dashboards,** and **Data Pipelines**, improving student problem-solving velocity by 40%.

### Filmy AI, New York, NY                    August 2024 - December 2025
*Pre-seed startup building an AI-powered talent portal catered to the film and entertainment industry*
#### Data Scientist
- Developing a **job matching system** using context vectors with **Gemma Embeddings**, aiming to **improve job search accuracy by 40%.**
- Implementing a **Natural Language-to-SQL** querying mechanism using **NLP** and **AI** to enable querying through simple Natural Language.
- Architected and managed cloud infrastructure on **AWS**, leveraging **RDS**, **EC2**, **Lambda**, **S3**, and **IAM** to deploy a scalable website.
- Trained junior interns with **NLP**, **LLM fine-tuning**, **word embeddings**, and **vector database** (chroma db).

### Fractal, Mumbai, India                    August 2021 - August 2022
*Worked with Crux Intelligence, an NLP-powered analytics platform by Fractal, a global AI and data solutions provider.*
#### Senior Customer Solutions Architect / Data Engineer
- Built **ETL** pipelines using **Python**, **SQL**, **Azure**, **Dremio**, **Databricks**, and **Airflow** to process client data (2-12 GB) with **98%** accuracy.
- Optimised **SQL** procedures and **data models,** resulting in a decrease in process execution time by **45%**.
- Orchestrated **Analytical Dashboards** for **Finance (Capital Markets)** and **CPG** clients, boosting platform usage to 97%.
- Gathered client requirements and **mapped use cases to business KPIs** by collaborating with technical and non-technical stakeholders.
- Led **upselling** and **retention** efforts, designing tailored solutions that addressed business needs, demonstrated value through **POCs**, and secured a **$120K deal** and a **3-year contract extension** for two clients.

### CleverTap, Mumbai, India                    July 2019 - August 2021
*A mobile analytics, customer retention, and multi-channel engagement Ad Tech platform.*
#### Senior Customer Success Engineer
- Automated root cause analysis and reporting for **data discrepancies** with **Python** and **SQL**, reducing resolution time by **20%.**
- Onboarded **80+ clients** with platform training and **cross-platform SDK integration** (Android, iOS, React Native, Unity, Web, REST APIs).
- Supported **pre-sales efforts** with **90%** deal closure via **PoCs**, **SDK integration support**, and **stakeholder demos**.
- Closed a **$150K deal** by addressing the Android X and Unity integration challenge through custom SDK integration.
- Mentored a **7-member team**, increasing productivity by **30%** via structured training and knowledge-sharing.
- **Collaborated** with Sales, Engineering, Data Science, and Product teams to **build features** and mitigate **cross-functional tasks**.
- Authored **technical documents**, **FAQ's** on the product website, and an **internal knowledge base** on **JIRA** and **Confluence**.
#### Customer Success Engineer
- Managed **120+ global accounts**, achieving **97% CSAT** by optimizing ad campaigns and resolving **15,000+ AdTech queries**.
- **Boosted ad campaign performance by 15%** through collaboration with client marketing teams on ad flows and conversion tracking.
- Facilitated as **SME** on **GDPR, CCPA, Cloud,** and **multi-channel marketing** (Push, Email, SMS, Webhooks, WhatsApp, AWS, GCP).
- Developed an **Android** app with **Java**, **CleverTap SDK** & **Firebase**, improving debugging and internal training efficiency by 30%.
- Delivered resolutions to customer queries related to campaign design, data-driven analytics, API integrations, webhooks, and debugging.

## EDUCATION

**Pace University,** New York, NY - Master of Science (**M.S.**) in Computer Science
**University of Mumbai**, Mumbai, India - Bachelor of Engineering (**B. Eng.**) in Information Technology

## PUBLICATION

**Success Prediction using Random Forest, CatBoost, XGBoost, and AdaBoost for Kickstarter Campaigns**      August 2019
**Link:** https://ieeexplore.ieee.org/document/8819828

## PROJECTS *(More projects available on GitHub)*

**Multi-Agent AI Browser** | Python, Electron, Google Gemini, Claude, Antigravity, Selenium
- Built a privacy-first Electron/Chromium browser using AI-assisted rapid prototyping, implementing a local multi-agent AI architecture with on-device models served via Ollama (Llava and Llama 3 for vision/reasoning, Qwen2.5 Coder for code generation), designed using Gemini and Claude in Antigravity IDE, Selenium-based autonomous browsing, and a Tailwind CSS resource-monitoring dashboard, ensuring zero data exfiltration and no API costs.

**Stella Chatbot** | *Python, Google AI Studio, Vertex AI, StreamLit*
- Fabricated an LLM-powered chatbot using Python, StreamLit, and Gemini API to assist users with career guidance and tailoring resumes.

**Fine-tuning LLM for English-to-French Translation** | *Python, Sloth, Gemma 2b, PyTorch, Seq2Seq, LORA, Pandas*
- Enhanced English to French sentence translation quality using the Gemma 2b Large Language Model with Unsloth, achieving a 40% improvement in BLEU score over a baseline Seq2Seq model.

**Data Pipeline Automation** | *Python, SQL, Dremio, Azure Blob Storage, REST APIs, Pandas, Numpy*
- Implemented Python scripts to automate periodic data extraction from a private network's SQL database, executing complex calculations, formatting data into CSV and Parquet files, uploading to Azure Blob Storage, and triggering an HTTP API to refresh data in Dremio.

**WebScraper API** | *Python, Selenium, Flask, BeautifulSoup,  HTML, DOM, Pandas*
- Built a Flask API with Selenium, Beautiful Soup, Gensim, and basic authentication for capturing metadata, icons, contact details, a screenshot of the home page, and a website summary into a CSV file.

**Real-time Uninstall Tracking for Android Applications** | *Java, Firebase Analytics, Node.js, REST APIs, Android, CleverTap*
- Integrated a Java-based Android app with Firebase Cloud Functions, REST APIs, CleverTap SDK, and Firebase Analytics SDK for real-time uninstall tracking, achieving nearly 100% accuracy and eliminating reliance on unreliable silent push notifications.

**Vector Representations of Words using Glove and Word2Vec** | *Python, Glove, Word2Vec, Pandas, Seaborn, matplotlib*
- Explored semantic relationships between words using pre-trained Glove and Word2Vec embeddings. Visualized word vectors and performed vector arithmetic to analyze analogies and similarities.

**Cat vs Dog Classification with Transfer Learning** | *Python, Pandas, Keras, PyTorch, MobileNetV2, Tensorflow, Seaborn,*
- Developed an image classification model using transfer learning with MobileNetV2, comparing the performance of TensorFlow (Keras) and PyTorch frameworks. Evaluated standard classification metrics and TensorBoard, visualizing key performance insights.

**Power Consumption Prediction and Analysis for New York** | *Python, Pandas, Numpy, Scikit-learn, FBProphet, Seaborn*
- Analyzed time series data of power consumption in New York City boroughs to identify trends and seasonality. Built a forecasting model using FBProphet to predict future power consumption, visualize results, and evaluate model accuracy, achieving an MAPE of 10%, to predict future power consumption for 2 to 5 years.