Pavas Krishnan Shanmugasundaram
Charlotte, NC|508-365-9299|pshanmugasundaram@clarku.edu
[LinkedIn|Portfolio|TableauProfile |GitHub](#)

Education

Clark University	**Worcester, MA**
Master of Science–Information Technology (CGPA:3.9/4.0)	Aug 2023-May 2025
Sri Sivasubramaniya Nadar College of Engineering	**Chennai, India**
Electrical and Electronics Engineering (CGPA:8.2/10)	Aug 2017 – May 2021

Technical Skills

Programming & Data Management: Python, SQL (MySQL, Oracle, SQL Server), C#, PySpark, HTML, CSS, Fast API

Visualization & BI Tools: Power BI • Tableau • Tableau Prep • Streamlit • Excel • DAX • Google Sites

Machine Learning & Model Management: Regression • Classification • Time Series Analysis • Anomaly Detection • Predictive Modeling • A/B Testing • XG Boosting • Scikit-learn • TensorFlow • CNN • MLFlow• DagsHub • DVC • Azure ML Studio

Data Engineering & Automation: Microsoft Fabric • Databricks • Snowflake • Airflow • Apache Spark • Kafka • Alteryx • DBT (Data Build Tool) Prefect.io (Data Orchestration Tool) REST API Git • GitHub

AI/GenAI Ecosystem: Hugging Face • Ollama • Lang Chain • Chroma DB • Agents (Crew AI, Autogen) RAG (Classic, Graph)

Experience

Daimler Truck North America	**Gastonia, NC**
Data Analytics Intern	Jun 2025- Till now

- Engineered and Launched a Power BI dashboard enhancing quality assurance monitoring and streamlining manual reporting efforts by 40% in automotive manufacturing.
- Orchestrated real-time machine sensor data integration via API connections, analyzed with Python, and automated the pipeline using Alteryx, resulting in a 35% reduction in data processing time.
- Engineered a Fast API-driven TPM platform with embedded QR code scanning and real-time alert notifications, enhancing maintenance efficiency and proactive issue resolution on the factory floor.
- Constructed an interactive Streamlit application, storing processed data in Snowflake, enabling real-time visibility and accelerating decision-making on production quality.

Ernst &Young (EY LLP)	**Chennai, India**
Associate Data Consultant (Client: Merck Animal Health)	Aug 2021-Aug 2023

- Refined SQL queries, stored procedures, and views in SQL Server, lowering query execution time by 42% and enhancing data processing efficiency for advanced data mining and analysis.
- Crafted interactive Power BI reports using DAX calculations and Power Queries, visualizing key business metrics for financial and sales reporting, improving executive decision-making by 70%.
- Automated Power BI report updates and incorporated email alerts to proactively notify stakeholders of critical insights.
- Spearheaded Data Warehouse and Lakehouse integrations using Snowflake, unlocking high-performance data workflows and uplifting analytical efficiency by 35%.
- Engineered and optimized data pipelines using Alteryx, enabling seamless data extraction, transformation, and loading (ETL) to enhance data preparation and analysis efficiency.
- Partnered with Business Unit heads & sales reps across the EMEA region to develop predictive analytics models and market segment strategies, driving data-driven decision-making that elevated regional sales performance and optimized go-to-market tactics.
- Utilized advanced Excel functions and formulas like Pivot tables, charts, V-Lookups, and VBA to streamline data analysis processes.
- Engaged in data stewardship efforts to manage data quality and resolve data issues, ensuring alignment with organizational policies and standards while maintaining robust data modelling practices.

Yogeesh Trading Pty Limited	**Durban, ZA**
Business Data Analyst	Jan 2020- Jul 2021

- Designed interactive Tableau dashboards to visualize logistics data, enhancing decision-making for international trade.
- Optimized procurement and cross-border trading efficiency by managing SAP sales & distribution and material management modules, reducing order processing time by 20%.
- Developed custom SQL queries to monitor raw materials & machinery shipments between countries, providing delivery insights.

SITEK INC	**Lexington, KY**
Summer Intern	Jul 2024-Aug 2024

- Participated in the development of applications using ASP.NET MVC, supporting the design and implementation of features-based requirements.
- Crafted Power BI reports to present key business metrics and track project progress, providing stakeholders with actionable insights into application performance, increasing stakeholder engagement by 60%.

Certifications

Microsoft Power BI Associate |Alteryx Designer Core | SQL for Data Science | Data Bricks Gen AI Fundamentals | Python for Data Science| Azure AI Fundamentals| Microsoft Fabric Data Engineer Associate | Azure Data Engineer Associate

Technical Projects

Factory Safety Chatbot — AI-Powered RAG with Small Language Model   Jun 2025 -Jul 2025

- Developed an AI-powered Retrieval-Augmented Generation (RAG) chatbot using local LLMs (Ollama/Gemma3), Chroma DB vector search, and Streamlit, delivering secure, document-grounded answers from factory safety manuals.
- Designed and deployed a full-stack private semantic search pipeline: PDF parsing and text chunking, embedding with Sentence Transformers, persistent vector database retrieval, and interactive web UI.
- Demonstrated advanced AI engineering by integrating PDF-to-vector ingestion, similarity-based retrieval, and context-constrained language modeling for real-time, confidential industrial knowledge delivery.

Seismic Insights: Earthquake Data Pipeline & Visualization using Microsoft Fabric   Mar 2025-Apr 2025

- Developed a real-time data pipeline to ingest seismic data from the USGS API into a Microsoft Fabric Lakehouse.
- Implemented medallion architecture (bronze, silver, gold) with geolocation enrichment and magnitude classification.
- Built an interactive Power BI dashboard to visualize earthquake trends by date, magnitude, and region.

Unlocking Boston Housing Loan Trends   Jan 2025- Feb 2025

- Utilized statistical and data science techniques for predictive modelling and analysis
- Enhanced housing market predictions through Linear Regression on the Boston Housing dataset, optimizing features for accuracy.
- Improved loan approval processes using logistic regression, one-hot encoding, and careful feature selection for better predictive performance. Python and statistical libraries for regression analysis in complex financial scenarios offer data-driven insights.

End-to-End Data Integration and Visualization with Snowflake and Tableau   Oct 2024-Nov 2024

- Built a comprehensive data pipeline incorporating data engineering, transformation, & analytics processes, improving data processing efficiency by 60% to support cloud-native applications.
- Utilized Snowflake for efficient data processing, facilitating the integration of datasets into a cloud data lake for analysis and reporting, reducing data processing time by 50%.
- Designed dynamic Tableau dashboards to visualize insights and complex data trends, aligning with cloud-native data architecture needs.

Racing Insights: An Exploratory Data Dive into Formula 1   Sep 2024- Oct 2024

- Conducted comprehensive EDA on a Formula 1 2023 dataset to uncover insights on driver performance, team statistics, and race outcomes, identifying 55% more trends in race dynamics compared to previous analyses.
- Utilized Python libraries such as Pandas, NumPy, and Matplotlib to manipulate and visualize data, highlighting key trends and patterns, which led to 80% more accurate predictions on race outcomes.

E-Commerce Database Management System   Jan 2024 – Apr 2024

- Developed ERD and logical schema to create a robust relational database system in Oracle SQL, ensuring referential integrity and data consistency. Optimized SQL queries, improving database performance and reducing execution time by 30%.
- Incorporated advanced data modelling techniques to design and implement database structures for managing customer accounts, orders, and product inventory.
- Designed and implemented a relational database system in Oracle SQL, incorporating advanced data modeling techniques with ERDs and logical schemas to manage customer accounts, orders, and product inventory effectively.