

# Pavan Venkata Manjunath Mallipudi

Tempe, AZ, 85281 | (623) 499-5275 | [pvmmallipudi@gmail.com](mailto:pvmmallipudi@gmail.com) | [LinkedIn](#) | [Github](#) | [Portfolio](#)

## PROFESSIONAL SUMMARY

**Data analyst and BI professional** with experience building SQL-driven reporting systems, data-quality workflows, and stakeholder dashboards across healthcare, SaaS, and operations datasets. Skilled in SQL, Python, Power BI, Tableau, and PostgreSQL, with projects covering CMS healthcare analytics, cohort analysis, KPI reporting, ETL pipelines, and executive-ready insights.

## EDUCATION

### Arizona State University

Aug 2022 - May 2026

B.S., Computer Science | Data Science Minor | Magna Cum Laude | GPA: 3.76 | 5x Dean's List

Tempe, AZ

• **Coursework:** Machine Learning, Linear Algebra, Probability and Statistics, Statistical Modeling and Inference, Data in R & Python

## TECHNICAL SKILLS

**Programming Languages:** Python, R, SQL

**Data Engineering & Platforms:** PostgreSQL, MySQL, Apache Spark (PySpark), ETL/ELT Pipelines, Data Modeling, Dimensional Modeling, Data Warehousing, Databricks, AWS, Docker, Git/GitHub, dbt, Snowflake, Airflow

**Data & BI Technologies:** Pandas, NumPy, Excel, Power BI, Tableau, Tableau Prep, Streamlit, Plotly, Matplotlib, Jupyter Notebook

**Data Analytics:** Exploratory Data Analysis (EDA), Statistical Analysis, Data Visualization, A/B Testing, Cohort Analysis, KPI Reporting, Ad Hoc Analysis, Root Cause Analysis, Stakeholder Management, Database Management

## EXPERIENCE

### ASU Social Embeddedness (Office of University Affairs)

Sep 2024 - May 2026

Community Engagement Data Specialist (Data Analytics & Platform Administration)

Tempe, AZ

- Managed, validated, and standardized university-community engagement data on the **Collaboratory** platform, improving data accuracy and consistency for cross-functional, data-driven reporting and stakeholder decision-making.
- Automated reporting workflows and developed **Tableau** and **Power BI** dashboards for **10+ ASU departments**, surfacing engagement trends and **KPIs** while expanding self-service analytics adoption.
- Led trainings and **stakeholder demos** for faculty/staff, translating complex data processes into actionable insights and promoting consistent data governance practices.

### DigiClips (Capstone)

Aug 2025 - May 2026

Database & Analytics Developer

Tempe, AZ

- Developed **MySQL** relational workflows for high-volume media-transcript datasets, improving reporting readiness through normalized schemas, indexed tables, and validation checks.
- Designed structured backup, database maintenance workflows to support long-term storage reliability, monitoring, and analytics readiness.

### Food Forest AI

Jun 2025 - Jul 2025

Jr. Data Analyst Intern

Philadelphia, PA

- Built Python (Pandas, Regex) data cleaning and enrichment pipelines to standardize **500+ B2B company records**, reducing manual processing effort by **~75%** and improving data quality for downstream analytics and reporting.
- Developed executive **KPI** reports and Power BI dashboards in **Excel** and **Power BI** to visualize supplier capabilities, certifications, geographic trends, and partnership opportunities, enabling stakeholders to identify market gaps and prioritize outreach strategies.
- Implemented **SQL**-based data quality checks and data validation workflows to detect **duplicates, null values, and schema inconsistencies**, increasing data reliability and supporting business intelligence initiatives.

## PROJECTS

### ReadmitScope US | Python, Pandas, SciPy, CMS Data, React [Deployed Project](#)

- Analyzed **CMS Medicare readmission data** across **3,055 hospitals** and **6 clinical conditions**, integrating ownership, star-rating attributes (99.7% match) to identify high-risk facilities; found **77% exceed expected 30-day readmissions**, a key driver of Medicare penalties.
- Built reproducible Python (Pandas) cleaning, enrichment, and QA workflows for suppression-prone public datasets, then used statistical testing (SciPy) to confirm hospital volume, ownership, and CMS star rating as significant readmission drivers (**p < 0.001**).
- Designed and deployed an interactive **React dashboard** with a searchable 3,000-hospital explorer and geographic breakdowns, translating complex CMS quality metrics into facility-level comparisons and **actionable insights**.

### SaaS Revenue & Churn Intelligence Platform | Python, PostgreSQL, SQL, Streamlit, Plotly [Deployed Project](#)

- Engineered a data pipeline and **PostgreSQL MRR** state machine using window functions and **CTEs** to classify **50K+** customer-months into revenue movement categories, enabling **NRR, GRR, churn**, and **cohort retention** reporting.
- Designed analytics models and materialized views covering **revenue waterfalls**, retention **cohorts**, **churn** segmentation, and customer health scoring, surfacing that payment failure predicts churn **3x better** than any other signal.
- Built **Streamlit dashboards** with interactive filters and drill-down analytics, validating results through **23 automated data quality assertions** across the analytics pipeline.

### Enterprise Data Lakehouse Integration Platform | Databricks, PySpark, Delta Lake, AWS S3 [Github](#)

- Built an end-to-end **Lakehouse** data platform in **Databricks** using **Bronze, Silver, and Gold** layers to integrate and standardize data from multiple business systems, enabling a unified analytics environment for reporting and decision-making.
- Developed **PySpark ETL** pipelines with **Delta Lake MERGE** operations for historical backfills, incremental processing, schema harmonization, data quality validation, and automated ingestion from **AWS S3** into curated analytical datasets.
- Designed interactive Databricks dashboards and analytics-ready Gold layer views, leveraging **Unity Catalog** governance and external S3 integrations to deliver scalable, **reliable business insights** across sales, customers, products, and revenue metrics.