PRASHANT SONI

DATA GOVERNANCE | MEDIA MEASUREMENT | ANALYTICS ENGINEERING | DATA QUALITY

Jersey City, NJ | +1 551 362 9944
prashantsonibps@gmail.com | LinkedIn | GitHub | Portfolio

SUMMARY

Data engineering and analytics professional building reliable data systems for research and media analytics. I hold a Master's in Artificial Intelligence from Yeshiva University and a Postgraduate Diploma in Data Science from IIIT Bangalore. As AI/Product Lead at OrphaNova Labs, I lead the data platform work end to end, from ingestion and schema design to quality gates and release decisions. My work runs on the media side too: I ran the targeted outreach that onboarded 500+ researchers at OrphaNova, and grew tech and promotional content on X to 5 million impressions in 30 days, among the first 10,000 creators monetised on the platform in 2024. Skilled in designing data dictionaries, taxonomies, metadata structures, and validation workflows that improve data quality, consistency, and trust.

HIGHLIGHTED PROJECT

MarketTrust 96 - Global Media Data Governance & Marketing Mix Modeling Nov 2026

Python, SQL, Snowflake, dbt, GCP, Streamlit, MMM

- Architected a global media data governance platform spanning 96 simulated markets, harmonizing campaign metadata, channel taxonomies, and source-to-target mappings across 3 advertising-platform formats.
- Developed 24 automated data-quality controls for media spend, naming conventions, schema integrity, and reporting consistency; in the simulated incident run, 100% of corrupted records were quarantined before certification through validated remediation workflows.
- Built traceable data pipelines, governed data dictionaries, and stewardship workflows to support certified campaign reporting, issue resolution, and auditable data changes.
- Developed an MMM-ready analytical dataset spanning 6 pilot markets, 4 media channels, and 104 weeks, integrating media spend, sales, and business controls for channel contribution analysis and marketing measurement.

EXPERIENCE

AI/Product Lead, OrphaNova Labs

Jan 2026 - Present

New York, NY

- Lead the end-to-end data ingestion and processing workflows for a rare-disease research platform, integrating multi-source data into standardized, governed structures.
- Designed data schemas and metadata structures supporting standardized representation, retrieval, and quality validation across research datasets.
- Implemented data-quality validation and pre-release evaluation checks to detect inconsistencies and support reliable research outputs.
- Ran targeted media outreach and onboarding campaigns that brought 500+ researchers onto the platform, driving 2,400+ research workflows.

Research Assistant, Yeshiva University

Nov 2024 - Dec 2025

New York, NY

- Developed OrphanAtlas, curating and structuring biomedical research data using standardized schemas, metadata, and controlled vocabularies.
- Established data provenance and evidence-traceability workflows, linking research findings and retrieved evidence to their original sources.
- Validated structured datasets for consistency, completeness, and quality, supporting reliable retrieval and analysis.
- Led lab reach and growth on the school's communication team, increasing lab reach 30% compared to the prior team's results.

PROJECTS

SpaceGuard 2026

Python, FastAPI, Firebase, REST APIs

- Developed multi-source data ingestion workflows integrating orbital telemetry, environmental conditions, and satellite data covering 25,000+ tracked objects into structured datasets for analytics and monitoring.
- Structured heterogeneous data feeds for downstream analytics, event monitoring, and risk assessment, improving data consistency and accessibility.

TPO360 2026

Python, Structured Data Validation, Streamlit, Machine Learning

- Built a healthcare claims review prototype combining structured claims data, configurable policy rules, and supporting documentation for governed, auditable decisions.
- Ran every claim through a 7-step audit pipeline with a full execution trace on 100% of claims; clean claims auto-resolve at 90%+ confidence, and edge cases route to explainable human review.

Multimodal Stock Predictor 2025

Python, Pandas, Time-Series Analytics, Machine Learning

- Built a data processing pipeline combining historical financial prices, technical indicators, and unstructured news data for time-series analysis.
- The multimodal model reached 56.8% average directional accuracy across AAPL, TSLA, and MSFT, beating Random Forest (51.0%), LSTM (53.0%), and FinBERT-only (54.6%) baselines.

TECHNICAL SKILLS

Measurement & Experimentation Marketing Mix Modeling (MMM), RROI Attribution, Response Curves, Budget Optimization & Forecasting, Incrementality Analysis, A/B Testing, Hypothesis Testing, Causal Inference, Regression, Time Series

Modeling & ML Python, R, SQL, PyTorch, XGBoost, Random Forest, Ensemble Methods, Neural Networks, NLP, LLM Fine-Tuning (LoRA/QLoRA), Machine Learning, Feature Engineering, Anomaly Detection

Data Governance & Management Data Dictionaries, Schema Design, Metadata Management, Taxonomies, Ontology Mapping, Data Provenance, Data Quality & Validation

Data & Cloud Snowflake, AWS (Glue, QuickSight, SageMaker), Google Cloud Platform (GCP: BigQuery, Vertex AI), dbt, Airflow, Hive, MySQL, MS SQL, PostgreSQL, ETL, Data Ingestion Pipelines, Data Transformation, REST APIs

BI & Visualization Tableau, Looker, Power BI, Domo, Looker Studio, Alteryx, Excel, Streamlit

Media & Marketing Technology Datorama (Salesforce Marketing Cloud Intelligence), Salesforce Data Cloud, Claravine, Adobe Analytics, Google Analytics (GA4), Google Ads, Meta Ads, TikTok Ads, Amazon Ads, Display & Video 360, Campaign Manager 360, The Trade Desk

Productionization Docker, Terraform, CI/CD, Kubeflow, KServe, MLflow, FastAPI, Firebase

EDUCATION

Yeshiva University | M.S., Artificial Intelligence | GPA 3.8/4.0 — Dec 2025

IIIT Bangalore | PG Diploma, Data Science | GPA 3.8/4.0 — Jun 2024

GLA University | B.Tech., Computer Science | GPA 3.1/4.0 — May 2023

RECOGNITION

• **Best R&D Award, Yeshiva University Symposium** - for OrphanAtlas, the rare-disease research hub.

• **Top 5 Solo Developer, YC Bio x AI Hackathon** - for multi-step biomedical research agents.

• **1st Place, Microsoft Quantum Hackathon** - Python/Pulser pipeline for neutral-atom pulse schedules; beat the published benchmark in 11-atom simulation.