PhishTank API Data Pipeline

Storage User Ingestion Processing **Data Sources** Extract: - Pull JSON response from PhishTank API - Decode base64 URLs PostgreSQL: Load: - Python script - Raw layer: raw API Analytics: query data and identify insights
Visualizations - Insert raw JSON into (PhishTank_API_Extract _Load_raw.py) - GitHub Actions pull raw.phishtank results - Staging layer: table (PostgreSQL) - PhishTank API staging.phishtank - Machine learning / Transform: (daily scheduled pipeline) results predictive modeling - Classify phishing status - Warehouse layers - Clean data - Insert transformed records into staging table