


## Hi, I'm Katie Sipos.

I'm a curious data/analytics engineer and analyst and I *excel* at turning raw data into actionable insights. I love making a manual process more efficient with automation and am always looking for ways to improve workflows. I work hard to create an environment where coworkers can self-serve data reasonably. Itake joy in creating cleaned up, straight-forward data sets for business users.

In my non-work time I love exploring public datasets, volunteering for [causes that are important](#) to me, [drawing](#), and playing tabletop RPGs and board games. 

---

## Core skills & technologies

- **Technologies:** SQL (advanced), Git (advanced), Python (intermediate)
- **Data tools:** Fivetran, Airflow, dbt, Hightouch, Segment, BigQuery
- **Marketing & Sales tools:** Hubspot, Salesforce, Madkudu
- **Data workflow & visualization:** Hex, Looker, Mode, Jupyter
- **DataOps:** GitHub, CI/CD, testing best practices
- **Data modeling:** dbt, best practices with a variety of data modeling techniques (star schema, OBT)
- **Executive presence:** engaging with leadership team directly, presenting and putting together board & investor slides

🌟 **Currently learning:** containerization and orchestration from [data engineering zoomcamp](#)

---

## Employers & experience

### PlanetScale, Analytics Engineer & Analyst

*August 2021 - April 2024*

- **Built the company's data infrastructure from 0 to its current state**
  - Implemented data warehouse (BigQuery) to have a combined source of truth for Sales, Marketing, Finance, and Product data. Managed access through Terraform with BigQuery service accounts which followed the principle of least privilege permission model. Worked with Security and Legal to ensure only business-critical PII was stored
  - Implemented and maintained a suite of data tools that enabled a completed end-to-end data process for the business (Fivetran, dbt, Hightouch, Hex)
  - Maintained existing Airflow setup by troubleshooting logs, fixing errors, and adding new tables & columns to schema as needed
- **Created and maintained dbt project**
  - Macros for reusability of code and basic cleaning of data
  - Followed a structure of source > staging > intermediate (when necessary) > production marts. Production marts were a mix of reporting tables for dashboards and cleaned up business entity tables for deeper analytics use
  - Wrote generic and custom tests, used dbt-utils and dbt-expectations packages to help maintain data quality
  - Wrote and socialized a pull request process for creating & reviewing dbt changes
  - Installed a number of packages to help more efficiently clean up and model ad, marketing, and finance data
  - Created incremental syncs, when necessary, to help with cost & performance of builds
  - Set up nightly snapshots for tables with changing dimensions/hard deletes and daily business metrics rollups
- **Put together a number of scheduled reports in Hex to help identify data issues**

- **Fivetran usage report:** a daily Slack report that notified me of any sharp increase in our Fivetran usage so that I could get ahead of billing surprises
- **dbt usage report:** a report using dbt-artifacts package to help monitor dbt project health
- **Mode usage report:** a monthly Slack report that highlighted when a new report was generated that put us in jeopardy of exceeding our query compute
- **Enabled Sales & Marketing with product data that helped identify new leads and better qualified current customers**
  - Scoped Segment track events for key customer conversions across the website which were then sent to other tools (Hubspot, Salesforce) for additional qualification and follow up
  - Rolled out Hightough and used that to send more complex product data to Salesforce to enrich contacts, campaign members, and accounts
    - Enriched contacts with their Segment uuid so that contact website activity could be rolled up to the account level and used in sales and customer engineering calls with customers
    - Sent all-time and session referral data and UTM fields to contacts & campaign members to better understand what initially drove a user to PlanetScale and what referred them to a specific campaign
    - Sent key product conversion events to a Slack channel that notified sales team members when a customer hit a specific set of requirements that we deemed worthy of targeted outreach
- **Followed data modeling best practices to create a set of easy to use tables for the business (and myself)**
  - **Sales and marketing enablement:**
    - **All people:** a combination of contacts & leads, de-duped, that joined ids from other sources so PlanetScale users could be more easily joined to their website and product activity
    - **All-time referral & first path:** created all-time referral and all-time first path fields that joined to a Segment uuid to understand first touch data better
    - **Campaigns with campaign members:** used Salesforce field history data for a custom Salesforce object (campaigns\_over\_time) to put together a table of a user's complete campaign history
    - **Relationships:** created a set of five tables that tied Salesforce contacts and account entities with their corresponding product and website uuids
    - **Organization and user activation moments:** two tables to represent key moments in a user or organization's journey. Example stages - product setup, product experimentation, product adoption, product expansion
    - **Database read and write estimator:** A read and write estimator built with Python that looked at a database's usage and predicted when they might need additional resources for their database
  - **Product and engineering enablement**
    - **canonical business tables:** a set of 10 foundational tables created to represent the building blocks of our business (databases, organizations, branches, invoices, users)
    - **all\_events:** a rollup table of all the different key events from the MySQL database, joined and transformed to look like an event stream. Example events - created a database, completed a deploy request, confirmed email
    - **organizations\_enhanced:** created 10 different signals that could be joined to the topmost business entity of PlanetScale (an organization). Example columns - total invoiced dollars, number of active databases, last invoiced dollars, number of active members, self-serve or sales-serve customer status
    - **Feature utilization score:** defined and operationalized a new value to represent a customer's feature utilization score, representing how much of the product they have used in a meaningful way
- **Put together analyses & reports to help keep the business informed about data**
  - Defined and socialized company-level core metrics to help measure the success of the business (monthly active or engaged users, monthly active or engaged databases, confirmed users)
  - Created a customer explorer which allowed stakeholders to enter a customer's name as an input field in Hex which would dynamically change the report. Included metrics like invoice amounts, database usage, recent website activity, and more
  - Created an end-of-month billing report for engineering and finance to use during end-of-month invoice reconciliation. Joined data from our MySQL database and Stripe to make sure everything was settled properly

- Conducted a cost-impact analysis to help our CFO understand the implications of sunseting a paid tier. Identified a number of different edge cases in these paid customers to help with communication strategy when we did eventually sunset the plan
- Analyzed customer storage, usage, and pricing data to help inform a new pricing model for existing plans
- Defined success metrics for engineering feature launches and created reports for monitoring over time
- Analyzed free and paid user retention and retention patterns of successful users to help inform onboarding nurture campaigns
- **Organized the data process from the ground up**
  - Organized a monthly product data review with CEO, CFO, and CTO to discuss trends in our product and invoice data
  - Responsible for weekly reporting and metrics for Marketing, Engineering, & GTM teams to help them understand what's working, key opportunities, and what could be improved upon
  - Created a #data channel to field ad hoc data requests from the leadership team (and eventually the whole company) and a complementary data request intake process within project management tool
  - Helped to elevate a culture of curiosity around data and drove a meaningful change in how folks thought and used data

## **GitHub, Senior Data Analyst**

*October 2020 - August 2021*

- **Data analysis & reporting**
  - Conducted free-to-paid conversion analysis to help GitHub's Growth team understand upgrade indicators to identify experimentation opportunities
  - Developed quantitative user profiles for GitHub purchasing personas to help support qualitative research
  - Conducted TAM analysis using Crunchbase API & GitHub data to determine a market fit for a new product line
  - Categorized and analyzed open source software (OSS) communities to understand project trends, usage, and retention to help identify opportunities where engineering teams could help maintainers or contributors
  - Helped define and operationalize the GitHub Community Health metric, a quantitative measurement of an OSS community's health
- **Data process**
  - Helped build curriculum and run GitHub's Analytics Primer, a course aimed at increasing data literacy and self-service of data via SQL and tool training
  - Helped refine and manage the Data team's data request and intake process, reducing the average time to ticket response and closure. Used GitHub Actions to automate notifications, issue responses and closes

## **GitHub, Senior Product Manager**

*November 2015 - October 2020*

**Product Manager** specializing in data analysis, company strategy, and user research. As a PM I worked across many different functions of the business - Enterprise, GitHub Education, the user profile, user authentication, OSS communities.

## **MagmaLabs, Technical Product Lead**

*November 2014 - November 2015*

**Product Lead** specializing in client management & team mentorship. Worked mainly with GoPro and assisted in the PM side of their GTM function for new product lines.

## **ModCloth, Technical Project Manager**

*November 2009 - November 2014*

**Project Manager** specializing in large scale project implementation & remote team management. Project examples - move from Quickbooks to Oracle for the finance team and the rollout of a new warehouse management system for 250+ warehouse employees