

# Zach Schumacher

New York, New York

618-980-9882  
[zschu15@gmail.com](mailto:zschu15@gmail.com)

[zschumacher.github.io](https://zschumacher.github.io)

## Profile

As a data engineer with demonstrated experience in a variety of industries and technologies, I can provide value to any organization through my proficiency in **leadership, python, and data architecture.**

## Experience

**Senior Manager Data Engineering, Simplebet;** New York, New York – 2021-Present

**Senior Software Engineer, Simplebet;** New York, New York – 2019 Nov-2021

- Lead and mentor team of 5 engineers of varying skill levels
- Groom and prioritize backlog and help team set weekly goals
- Lead architect in **building data streaming solution for click stream data and real time user betting behavior**
- Wrote bash script for all teams to use in CI pipelines for publishing Avro schemas to the confluent schema registry
- Wrote ETL library in python for ensuring reusable, modular and well tested data pipelines
- Developed end to end CI pipeline for deploying pyspark code to databricks
- Developed an error handling mechanism using python and postgres that is **currently part of a patent application**
- Used Apache Airflow for task orchestration, deployed on Kubernetes
- Develop and maintain REST API on top of data warehouse using DRF
- Automated database migrations using Django
- Helped machine learning engineers deploy models by supporting MLFlow

**Software Engineer, Quest Analytics;** Kansas City, Missouri – 2019 June-Nov (Contract)

- Rewrote legacy ruby REST api using Flask and SQLAlchemy
- Collaboratively designed new schemas for data warehouse and automated migrations using alembic
- Rewrote legacy data aggregation platform from ruby to pyspark deployed on databricks; **saved roughly 80% execution time**

**Data Engineer, SelectQuote,** Kansas City, Missouri – 2019 Jan-June (Contract)

- Worked closely with data scientists to integrate predictive and optimization models into Flask web services to minimize spend on customer leads in real time
- Wrote python script that would periodically run a machine learning model and use the google AdWords REST API to programmatically adjust keyword allocation
- **Increased uptime of core machine learning Flask web service from ~75% to ~98%** by refactoring, adding celery, and increasing test coverage dramatically

**Data Engineer, American Century Investments,** Kansas City, Missouri – 2018-2019

- Wrote python script to query Salesforce objects and convert them to JSON format; empowered the Informatica ETL developers to ingest Salesforce data using schema on read to eliminate downtime of data pipeline due to sales team changes; **saved company ~\$45,000 a year by eliminating need for Salesforce Informatica connector**

- Wrote python CLI using click for automating a lot of repetitive tasks for informatica ETL developers
- Reduced critical stock rating process **from ~two-day manual effort to a scheduled python script that ran in ~20 seconds**
- Led effort to refactor legacy CRUD application to a flask admin + sqlalchemy app
- On call support of python and legacy java code base
- Replaced manual deployment processes using Jenkins
- Built companies first graph database in neo4j to help marketing identify potential new customers by relating accounts to custodians and secondary account owners

**Business Intelligence Developer, Amsted Rail, Kansas City, Missouri – 2016-2018**

- Built email scheduling web application using Flask and KendoUI to allow non licensed Tableau users to still receive emails of Tableau dashboards on a scheduled basis; **saved ~\$50,000 due to reduced license needs**
- Administered company's Tableau environment (Desktop, Server and Mobile)
- Python data pipelines for moving data from user defined SAP business objects reports into the data lake for consumption by data scientists
- Built out and maintained data warehouse for 5 manufacturing locations including all data marts and ETL processes.

Education

**University of Kansas** – BSB Information Systems Technology, 2016

Skills

python, flask, django, pyspark, databricks, git, circeci, kafka, sql, golang, postgresql, oracle, microsoft sql server, bash, celery, leadership, agile