

# Zachary Thomas Cruz



Sacramento, CA | Phone: +1 916.764.6916

Email: [zacharytcruz@gmail.com](mailto:zacharytcruz@gmail.com)

## DATA ANALYST TURNED BUDDING DATA SCIENTIST

A young professional and recent graduate of a Master of Science in Data Science, looking to bring my experience with and passion for data to a dynamic venture. As an adaptive learner, I've gained reputé for improvised and creative solutions to overcome the challenges that data often brings. Combining ad hoc programmatic decision-making with user empathy, I've learned that data should be delivered as intuitively and pleasantly as possible to stakeholders and customers, even if it means stepping beyond company conventional practice and expectation.

## PROFESSIONAL EXPERIENCE

### Predictive Analyst

September 2018 – Present

*Marketing Evolution, Inc. – Rancho Cordova, CA*

Provide descriptive analytics, conduct research, and deploy predictive models for internal inquiry and client deliveries.

Job duties:

- Utilize AWS EC2 and EMR clusters to deploy predictive models on big datasets using R, PySpark
- Collaborate across teams to validate data that enters machine learning model pipeline
- Monitor models and distill results into insights for stakeholder interpretation
- Produce ad hoc and periodic descriptive analytics to understand additional trends in data

Sample project:

#### Internal Project – Automated Data Validation

2019

Utilize Python libraries to script SQL queries for data validation against various sources in the AWS ecosystem, including: Hive, Athena, Postgres RDS, and S3 objects. The goal of this project is to standardize validity checks for data that enters the modeling stage, in turn maximizing integrity of results. The process is made scalable and portable by its implementation in a series of Jupyter notebooks that can run either locally or in AWS EMR (Elastic Map Reduce), depending on the data size and storage location of the data being validated. The notebooks produce summary reports, plots, and warning messages to help the analyst in diagnosing issues in the data at various stages in the modeling pipeline.

### Lead Developer

October 2017 – August 2018

*Simpler Systems, Inc. – Santa Barbara, CA*

Create processes for automating development, for internal use and customer needs.

Job duties:

- Create generalizable and configurable reporting applications
- Predictive analytics using scikit-learn
- Dynamic SQL queries through Python scripting

Sample project:

#### Internal Project – Predictive Metadata

2018

Initiated an experimental project where I developed a classification model for predicting column schema from imprecise or undefined schema. This was an internal project designed to automate column metadata definitions for reducing reporting application development. Using primarily Python's scikit-learn libraries, created a KNN model that trains on thousands of existing columns across Simpler Systems' data warehouses. A new dataset is passed through this model, and the model predicts metadata for a given column based on that column's features.

## **Lead Data Analyst**

**January 2015 – October 2017**

*Simpler Systems, Inc. – Santa Barbara, CA*

Led teams in the development of enterprise-grade reporting applications and delivered software services for public and private sector organizations.

Job duties:

- Lead the development and implementation of various concurrent ETL projects
- Use creative leeway to enhance user-experience and productize design
- Technical support lead and point of contact with end-users

Sample projects:

### **Internal Project – Dynamic Reporting Applications**

**2017**

Created dynamic scripting templates embedded within Simpler Systems' platform, which automated most reporting application development, saving myself and colleagues from hours of development time. These dynamic applications also instantly called SQL objects with a preconfigured naming convention, delivering functioning applications upon the creation of a SQL view.

### **County of Tulare, Information Technology – Property Tax Data Conversion**

**2016**

Aided the county's IT department with data conversion from their mainframe property tax system to a modern one. Using the Simpler platform, I developed a reporting website for regression testing and data validation, following pre-converted source schema to crosswalk and import schema.

### **University of Georgia, Auditor-Controller – Human Resources & Financial**

**2015**

Collaborated closely with UGA's IT department to leverage the historical value of their legacy financial and human resources systems, stored in an Oracle SQL database. I led in building a reporting website for navigating this data more intuitively, ultimately replacing their SAP Crystal reporting. When development was complete, I curated technical documentation for their specific configuration, and eventually conducted a series of training sessions in which several programmer-analysts were taught to build upon and maintain the new application themselves.

## **Data Analyst**

**January 2013 – December 2014**

*Simpler Systems, Inc. – Santa Barbara, CA*

Assisted in the development of reporting applications and delivered software services for public and private sector organizations.

Job duties:

- Design ETL processes using SQL programming
- Web application development
- Technical writing and documentation

Sample project:

### **City of Tustin, Community Development Department – Permits and Code Enforcement**

**2014**

Provided ETL processes to merge access of three siloed systems managed by the Community Development Department into a single reporting application. The Permits web application now seamlessly pools data derived from the historical system, alongside data from the current system and an additional auxiliary system into a single platform.

Collected data and assisted with front-end application development for non-profits and small businesses.

Job duties:

- Data research, collection, consolidation, and organization
- Online marketing and research
- Front-end web application development

A select project developed for dataPreneur include:

**Notes for Notes – Non-profit U.S. Tax Code 501(c)(3)****2012**

I developed a web application for processing membership information and daily studio attendance logs for the studio branches of the non-profit organization. The data was stored in a custom-built database in SQL Server, then pooled into a reporting website for analysis of the data collected.

**ACQUIRED SKILLS & KNOWLEDGE**

- SQL and NoSQL programming (MSSQL, Oracle, MySQL, Postgres, Hive, Spark SQL, MongoDB)
- Python
- R
- Machine learning (clustering, regression, natural language processing, neural networks, random forests)
- Unix/Linux OS commands
- Web application development (HTML, CSS, JavaScript, C#-extended XML)
- User-experience (UX) design
- AWS ecosystem including: S3, Athena, Glue, RDS, EC2, EMR
- Git
- Data visualization including Tableau, matplotlib, plotly, ggplot
- Census data, public sector data, mapping data
- Collaboration with marketing teams to enhance user experience
- User empathy through face-to-face/webinar client demonstrations, end-user training, and project management
- Turning complexity into meaningful insight

**EDUCATION****University of the Pacific**

*Master of Science, Data Science – Degree conferred December 2019*

*The M.S. in Data Science provides a framework for solving real-world machine learning and predictive analytics problems, from gathering data, transforming it through reproducible pipelines, modeling the final features, and validating the model. Taught bilingually in Python and R, a deep understanding of data science principles provides both a foundation and practical execution of supervised and unsupervised learning techniques.*

- Master's program bilingually taught in Python and R
- Extensive use of core Python libraries, including:
  - Pandas, Matplotlib, SciPy, scikit-learn, seaborn, PyViz, NLTK, Tensorflow
- Use of core R packages, including:
  - broom, ggplot, plotly, statistical package, data.table

## Online Self-Education

*Additional supplemental auxiliary courses online:*

*Codecademy.com:*

- Python
- Javascript
- Responsive Design

*CodeSchool.com:*

- R

*LinkedInLearning:*

- <https://www.linkedin.com/in/zachary-cruz/> (see linked in profile)

## Santa Barbara City College

*Supplemental Computer Science courses to advance career including:*

*CIS 232: SQL Server Database Administration*

- Backups, Restores, Transaction Logs
- Storage and design principles (best practices)
- OS Engine/Infrastructure concepts
- Real-world events solving and system administrator crisis management

*CS 101: Computer Concepts*

- Fundamental computer science concepts taught in Python

## University of California, Santa Barbara

*Bachelor of Arts, Political Science – Degree conferred 2012 with Honors*

## Universidad de Córdoba, Language and Culture Program, Córdoba, Andalucía, Spain

*UC Education Abroad Program (Fall 2010)*

*UCOIDIOMAS - (immersive courses taught entirely in Spanish)*