

# OMAR BAIG

DATA SCIENTIST

✉ o\_baig@live.com 🌐 omarbaig.net 📞 770-595-8628 📍 Atlanta, GA [in linkedin.com/in/omarbaig2](https://www.linkedin.com/in/omarbaig2) 📄 obaig47

## SUMMARY PROFILE

---

I am a curious, solution-oriented Data Scientist seeking to assist end-to-end decision making. After earning a degree in Computer Science at the University of Georgia, I began using my programming skills at Griffin & Strong. I communicate with expert lawyers and economists who specialize in Disparity Studies and automate their research methodology using Python. So far, I've created unique data pipelines to clean and analyze three different government's procurement data. I am currently building a Natural Language Processing model for categorizing data which reduces the need for difficult manual work. My focus is always on making sure any software or analysis platform is unit tested and can be deployed on any computer, locally and on the cloud. Please visit my website to see examples of my work.

## WORK EXPERIENCE

---

### Griffin & Strong P.C., Data Analyst

Nov. 2018 - Current

- Performs multifaceted analysis on collected data from governmental organizations
- Translates research methodology into software and adapts models to each organization's constraints
- Isolates data gaps and anomalies and cleans data for reproducible analysis
- Communicates with clients to collect data and present findings
- Leads efforts to take on time consuming and costly problems using automation

## PROJECTS

---

### Work Category Text Classifier

An NLP machine learning algorithm which classifies procurement data into one of five work categories. NLTK is used to process the manually categorized data records and then train the classification model.

### Pandas Profiler

A Python software allowing non-programmers to take advantage of the open-source "pandas\_profiling" library which creates comprehensive HTML summaries of raw data files. This summary includes data warning, missing values, and correlations. The front-end was built using PyQt and can run on any operating system.

### Equity Lib

A Python library containing pre- and post-processing algorithms for wrangling common problem patterns in procurement data. The primary problem solved is resolving entities in joined data systems.

### Forecasting Avocado Prices

An analysis of historical avocado price data implemented using the open-source Prophet library to predict future prices

### PASSNYC: Data Science for Good

An analysis of NYC public schools which provides a recommendation to PASSNYC for how they should distribute their services to improve the diversity of the specialized high schools

## TOOLS

---

**PROGRAMMING:** Python, R, C, Java

**DATA ANALYSIS:** Pandas, Plotly, Dash, Altair, iPyWidgets, Superset, Voila

**DATABASE/BIG DATA:** SQL, NoSQL, Spark

**MACHINE LEARNING:** Scikit-learn, NLTK, Tensorflow

**WORKFLOW:** Jupyter Lab, Emacs, Org-mode, Git, Unix, Unix

**CLOUD:** AWS, GCP

**DEVOPS:** Docker, Kubernetes, Terraform

## EDUCATION

---

### University of Georgia

Bachelor of Science Computer Science 2018

- Cumulative GPA: 3.55 (Cum Laude)