

# Matt Wheeler

## Customer-Centric Data Scientist

---

Denver, Co

303.408.3351  
matt.s.wheeler@gmail.com

Experienced Data Scientist with a Master's Degree in Statistics and over 15 years of expertise in business analytics, leveraging machine learning models, utilizing distributed computing, building visualizations, and employing cloud applications. Deeply passionate about problem definition and data-driven storytelling.

---

---

### Skills

- \* Languages: Python, Spark, SQL, SAS
  - \* Frameworks: Tensorflow, PyTorch, scikit-learn, Prophet, DeepAR, Catboost, sklearn/NLTK/Spacy, OpenAI/Langchain/LLM,
  - \* Platforms: Databricks Orchestration, Snowflake, Jupyter, AWS, Sagemaker, EMR, GCP, Git/CLI, Azure DevOps, Hive / Hadoop, Terraform
  - \* Visualization: Tableau Design/Online, Power BI, Google Analytics
  - \* Project Management: Agile Methodology, Jira, People Management
- 

---

### Experience

#### Hellofresh / Senior Data Scientist, Marketing/CRM

NOVEMBER 2021 - PRESENT, BOULDER, CO

- \* Created a robust Prophet time-series model pipeline with hyperparameter optimization for customer campaign value and discounts across marketing channels, then ensembled these predictions with heuristic layers to produce maximum CAC (Customer Acquisition Cost) estimates across Brands & marketing channels, enabling optimization of future expenditures by subchannel.
- \* Developed POC Quasi-Experimental Evaluation Design (QED) for customer touchpoint hypothesis testing based on KD-tree and nearest neighbor customer feature vector matching, e.g. can we evaluate the profitability of a customer downloading our App for their first order, prior to rolling out a formal test?
- \* Harnessed and transformed Google Analytics session data to provide Multi-Touch Attribution customer insights for VP of Data Analytics, driving 2024 strategic initiatives and project scoping for Attribution Modeling.
- \* Designed automated Databricks workflows for ingesting and visualizing 3rd party Direct Mail API data to display marketing campaign results and save Marketers hundreds of hours per year.
- \* Built Order Count and Discount gradient boosted tree models for VP of Good Chop; implemented predictions and monitoring pipeline at scale, including 300+ weekly training jobs, model performance monitoring, and versioning.
- \* Leveraged pre-trained neural networks (SBERT) for NLP Topic Modeling around cancellation responses as well as Reddit brand sentiment and potential voucher fraud with weekly refreshed front-end.
- \* Developed a large-scale customer duplicate matching pipeline to detect instances of customer discount manipulation, while also creating

cross-brand customer matching algorithms, across all brands, to reveal profitable customer paths and facilitate effective cross-selling initiatives, all hosted in self-service Tableau layer.

### **Slalom Consulting/ Solution Architect**

MARCH 2020 - NOVEMBER 2021, DENVER, CO

- \* Healthcare C-Level: Built data pipelines and XGBoost models to predict customer churn and inform important features for go-to-market strategies around targeting customer cohorts.

- \* Government Agency: Implemented Computer Vision solutions around image object classification for highway cameras using GCP/AutoML and YOLO frameworks ([here's a blog](#)).

- \* Major Telecom Player: Led quality assurance and data governance for OTT Streaming Platform product launch. Built LDA Topic Model for customer tickets to assist CareOps with front end analytics portal. Developed automated API engineering scripts for downstream analytics.

### **AirDNA / Senior Data Scientist**

OCTOBER 2019 - MARCH 2020, DENVER, CO

- \* Assembled and managed the Data Science team. Worked alongside several teams to facilitate daily and monthly updates to AirDNA's web services

- \* Uncovered vulnerabilities in data scraping systems, data pipelines, and modeling codebases. Improved monitoring around data collection and model performance, utilizing Tableau Online for daily refreshes.

### **Slalom Consulting/ Consultant, Solution Architect**

OCTOBER 2017 - OCTOBER 2019, DENVER, CO

- \* Major Telecom Player: Orchestrated AWS machine learning frameworks around time series forecasts for advertisement impressions utilizing AWS Sagemaker for model management, DeepAR Algorithm, and EMR for compute. Designed and integrated anomaly detection framework via median absolute deviation in Spark for improving machine learning model performance.

### **PricewaterhouseCoopers/ Associate, Senior Associate**

OCTOBER 2010 - AUGUST 2016, WASHINGTON, DC / DENVER, CO

- \* Statistical programmer for Tax group. Implemented sampling methodologies in SAS language.

### **US Census Bureau/ Survey Statistician, Mathematical Statistician**

OCTOBER 2006 - OCTOBER 2010, WASHINGTON, DC

- \* Survey analysis and methodology. Publication support.

---

## Education

---

**George Washington University / Masters of Science, Statistics**  
AUGUST 2007 - MAY 2009, WASHINGTON, DC

**Virginia Polytechnic University and State University/ Bachelors  
of Science, Mathematics / Bachelor of Arts, Economics**  
AUGUST 2001 - MAY 2006, BLACKSBURG, VA

---

## Languages

---

English, Italiano (intermedio), Francais (base)