

Keith Chamberlain, M.S.

GitHub /[keithchamberlain](#)

LinkedIn /[keithchamberlain-](#)

[kchamberln@gmail.com](#)

(720) 300-9745

Greeley, CO

Skills

Implemented and validated predictive and prescriptive models. Created statistical models. Incorporate a variety of statistical and machine learning techniques in projects. Wrote programs to cleanse and integrate data in an efficient and reusable manner. Used leading edge and open-source tools including Python, R, and TensorFlow. Communicated with internal and external clients to understand and define business needs and appropriate modelling techniques to provide analytical solutions. Evaluated modeling results and communicated the results to technical and non-technical audiences.

Data Projects

Tweet Sentiment Analysis of Music Genre

[GitHub](#)

- Harvested 60k+ Tweets over 2 weeks using the Twitter API 2.0 meeting question criteria
- Engineered clean dataset from initial set of duplicates and multiple overlapping records
- Ensured data accuracy/consistency for question of interest

Supervised Learning & NLP Textual Analysis of New Patient Leads

[GitHub](#)

- Added NLP step for text fields & feature engineered; improved F1 Score by >7%.
- Employed data science techniques to measure performance of classifier models with new NLP features by F1 Scores and feature importance by Shapley values in Python.
- Developed models and software tools for data cleaning, URL decomposition, model comparisons & data interpolation by fitted distribution.
- Developed project in an Agile environment.

Supervised Learning of New Patient Leads for a Marketing Company

- Cleaned and performed EDA on ad-click transaction and communication log from 2012-2021.
- Measured performance of classifier models: Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, Gradient Boosting by AUC and F1; and Feature Importance by drop 1 F1 in Python in an Agile environment.

To Loan or Not (to Loan): EDA of Lending Tree Data

[GitHub](#)

- Performed and documented exploratory analysis of Lending Tree data for accepted and rejected loan requests from 2009 to 2018.
- Communicated results of analyses in written, presentation and oral form
- Performed tests of population proportions
- Developed project in an Agile environment.

Experience

Research Statistician & Data Engineer | Freelance

Sep – Oct 2020

- Generated data pipelines and coordinated data from different databases, including from usafacts.org, CDC, NYT for daily and weekly COVID-19 deaths and hospitalizations supporting a Data Science team.
- Verified data between the institutions from a pool of 32 source municipalities, resulting in a dataset with one unreconciled pair of data between USAFacts and NYT.
- Generated univariate and bivariate statistical tests and data visualization

Keith Chamberlain, M.S.

GitHub /[keithchamberlain](#)

LinkedIn /[keithchamberlain](#)

kchamberln@gmail.com

(720) 300-9745

Greeley, CO

Experience (Cont.).

Quality Control Analyst II | *ChromaDex Inc.*

Aug 2014 - Sep 2020

- Setup, maintained, verified, ran analytical instrumentation and analyzed signals from chemical processes in the electromagnetic spectrum & validated data collection systems before and after use, while verifying measurement systems daily.
- Developed and validated controlled data entry processes that reduced scientific data processing time from 30 min per datafile to 5 min per file, freeing up 4-6 hours per sequence.
- Chamberlain, K. A. (2017). **70-point Acetic acid curve by GC/FID demonstrates heterogeneity of variance in analytical data.** DOI:[10.17632/dwf4ddww3w.2](https://doi.org/10.17632/dwf4ddww3w.2).

Technical Editor, R Programming | *Freelance*

Feb 2016 – Apr 2016

- Robbins, N. B., Robbins, J. (March 2016). **Effective Graphs with Microsoft R Open:** Data Treatment White Paper. Microsoft. 53 pp. Retrieved from <https://github.com/nbrgraphs/mro/raw/master/EffectiveGraphsMRO.pdf>.
- Edited R code for this R graphics white paper expose as part of the Microsoft R Open release.
- See acknowledgements on p. 52.

Contract Analytical Chemist | *ChromaDex, Connection*

Apr 2011 - Aug 2014

- SME for Residual Solvents Analysis by Headspace analysis in a GMP lab
- Investigated internal standard recoveries based on headspace volume and analyzed the results in R by selecting an appropriate analysis technique, generated inferences based on statistical analyses, communicated results with stakeholders, revolutionized our house method for half a decade and reached Pharmacopoeia requirements.

Chemist | *ChromaDex Analytics*

Jul 2008 - Aug 2010

- Analyzed chemical reference standards by chromatographic methods, wrote CoAs.
- Ensured the quality of measurements & identified root cause of malfunctions
- Hoekstra, B., Chamberlain, K., Traub, J., Baugh, S., Venkataraman, S. (Aug 2009). **Comparative study of HPLC methods for the Analysis of Diterpene Glycosides from Stevia rebaudiana.** 57th International Congress and Annual Meeting of the Society for Medicinal Plant and Natural Product Research, Geneva, Switzerland. Retrieved from: <https://www.thieme-connect.com/products/ejournals/abstract/10.1055/s-0029-1234701>

Education

Data Science Infinity | *Remote*

Present

Galvanize, Inc. Data Science Immersive | *Remote Full Time Pacific; 420 hours*

Certificate: Aug 2021

MS | Business Intelligence and Data Analytics | *Saint Mary's Univ. of Minnesota, Winona*

Dec 2020

www.TheAnalysisFactor.com | *Remote - New York*

2016

- Analyzing Count Data: Poisson, Negative Binomial, and Other Essential Models; 6-weeks
- Analyzing Repeated Measures Data: ANOVA and Mixed Model Approaches; 8-week workshop
- Logistic Regression for Binary, Ordinal, and Multinomial Outcomes; 4-week workshop

BA | Psychology | *University of Colorado, Boulder, CO*

May 2009