+1-360-206-7812
Leavenworth, WA
colindmiddleton@gmail.com

Colin Middleton

Entry Level Data Scientist

GitHub: middlec000
LinkedIn: colin-middleton-000
Website: https://middlec000.github.io/

Recent Eastern Washington University graduate with an MS in Applied Mathematics. Completed thesis project on predicting likelihood of homelessness based on an individual's utility billing history. Strong background in applied mathematics and statistics with practical experience in data exploration, manipulation, analysis, and prediction. Seeking to launch career with innovative and learning-focused company.

Skills

Software Languages	Python (pandas, numpy, statsmodels, scikit-learn, matplotlib, seaborn, scipy, tensorflow, streamlit), R, SQL, Java, Git, LaTeX, Markdown
Data Analysis	Correlation, Box plot, Histogram, Q-Q Plot, Hypothesis Testing
Data Preparation	Reformatting, Filtering, Imputation, Class Balancing
Supervised Learning	Linear Regression, Logistic Regression, Cox Proportional Hazards, Vanilla ANN, LSTM, Decision Trees
Unsupervised Learning	K-Means Clustering, Hierarchical Clustering, Principal Component Analysis
Performance Analysis	Performance Metrics, Parameter Interpretation

Education

Master of Science in Applied Mathematics / Eastern Washington University **September 2019 — June 2021**

- GPA: 4.0
- Total Credits: 79
- First graduate of reopened program (2021)
- Outstanding Graduate Award (2021)
- Graduate Service Appointment (2020 — 2021)

Bachelor of Science in Mathematics / Western Washington University **September 2013 — December 2017**

- GPA: 3.5
- Total Credits: 211
- Minors: Honors, Chemistry
- Math Fellow (2015 - 2017)
- Presidential Scholarship (2013)

Project Experience

Data Science Lead / Homelessness Prediction **June 2020 — June 2021**

Spokane Predictive Analytics and Master's Thesis Project *Spokane, WA*

- Established specific research questions.
- Received, matched, and preprocessed de-identified data from multiple files using Python.
- Performed data exploration on the 91,591 rows using multiple types of correlation and data visualization to determine relationships within the data with Python.
- Engineered new cumulative features.
- Investigated longitudinal approaches with Linear Regression, Cox Proportional Hazards, and LSTM models as well as non-longitudinal approaches with Logistic Regression and Vanilla ANN models using Python.
- Evaluated and compared models based on their Receiver Operator Characteristic curves. The Logistic Regression model performed best with an Area Under the Curve of 0.81 - similar to current research.
- Utilized K-Folds and minority class oversampling to combat extreme data imbalance (0.39% positive cases).
- Presented findings to the City Council of Spokane and defended as master's thesis.
- Paper pending publication in a peer reviewed journal.
- https://github.com/middlec000/SPA_predict_homelessness

Software Engineer / Academic Paper Clustering by Topic **January 2021 — March 2021**

Big Data Analytics Class Project, Eastern Washington University *Spokane, WA*

- Established a data preprocessing pipeline that pulled the body of each paper, detected the language and only retained English documents, tokenized and lemmatized each text, then converted all documents to sparse vectors of TF-IDF scores using Python.
- Assisted with development of a customized K-Means algorithm optimized for sparse, largely disjoint TF-IDF vectors.
- Developed component to determine number of clusters to form using the Elbow Method.
- Currently working on publishing a fast hierarchical text document clustering Python package as an extension of this work.
- https://github.com/middlec000/fhdc

+1-360-206-7812
Leavenworth, WA
colindmiddleton@gmail.com

Colin Middleton

Entry Level Data Scientist

GitHub: middlec000
LinkedIn: colin-middleton-000
Website: https://middlec000.github.io/

Data Analyst / The Wordler **March 2022 — Present**
Personal Project *Leavenworth, WA*

- Acquired data from Kaggle and Wordle website source code.
- Created data filters using regular expressions in Python.
- Created user-friendly Streamlit website to suggest Wordle words for users.
- https://share.streamlit.io/middlec000/wordler/main/src/main.py

Software Engineer / Grades vs Student Characteristic **December 2021**
Personal Project *Alpine, WY*

- Used an ANOVA test to determine if there is a significant difference in grades between groups of students where groups are based on chosen characteristic such as race or neighborhood.
- User-friendly Streamlit web interface.
- Automatically performs normality and skedasticity hypothesis tests.
- https://github.com/middlec000/grades_vs_student_characteristic

Data Scientist / Traffic Prediction **January 2017 — December 2017**
Senior Project, Western Washington University *Bellingham, WA*

- Predicted traffic flow arriving at the US / Canadian border crossing based on upflow traffic sensors using averaging, Principal Component Analysis, and curve fitting methods.
- Presented findings to Whatcom Council of Governments and WWU faculty.

Work Experience

Architectural Drafter **January 2022 — Present**
Alison Miller Architect *Leavenworth, WA*

- Create drawings using AutoCAD LT: floor plans, roof plans, sections, exterior elevations, window and door schedules.
- Final drawing set: dimensioning, code compliance, viewport scaling.
- Site measure existing buildings for remodel.
- Use Dropbox to share files.

Architectural Drafter **October 2018 — December 2021**
Brooks Middleton Architect *Anacortes, WA*

- Created drawings using AutoCAD LT: site plans, floor plans, roof plans, electrical plans, sections, interior elevations, exterior elevations, window, door, and finish schedules.
- Made design adjustments and suggestions.
- Researched building code and construction product specifications.

Graduate Service Appointment **September 2020 — June 2021**
Eastern Washington University *Cheney, WA*

- Trained as a university level teacher.
- Completed a final project on student learning and incorporating familiar contexts into lesson plans.
- Developed lesson plans and taught two live remote class sessions (linear algebra).
- Graded homework and exams fairly and efficiently.
- Provided feedback and suggestions to students about assignments.
- Provided feedback to professors about their students' progress.

Teacher Assistant **September 2019 — September 2020**
Eastern Washington University *Cheney, WA*

- Graded homework and exams fairly and efficiently.
- Provided feedback and suggestions to students about assignments.
- Provided feedback to professors about their students' progress.

Architectural Drafter **November 2018 — May 2019**
Beacon Rock Architecture Group *La Center, WA*

- Created drawings using AutoCAD LT: floor plans, roof plans, sections, exterior elevations, window and door schedules.
- Researched building code.
- Used Dropbox to share files.

+1-360-206-7812
Leavenworth, WA
colindmiddleton@gmail.com

Colin Middleton

Entry Level Data Scientist

GitHub: middlec000
LinkedIn: colin-middleton-000
Website: https://middlec000.github.io/

R & D Intern / Liquid Gas Mixing System **July 2016 — August 2016**
ESS Inc. *Portland, OR*

- Assisted with preparation and maintenance of battery testing stations.
- Assisted with fabrication and improvement of liquid-gas mixing system as component of large scale battery.

Publications

Prediction of first-time homelessness risk based on utility payment history **Pending**
Eastern Washington University *Cheney, WA*

- Advisor: Dr. Andrew Oster, Associate Professor, Eastern Washington University
- Paper based on Homelessness Prediction project with Spokane Predictive Analytics - see projects

Related Coursework

Eastern Washington University **September 2019 — June 2021**

- Big Data Analytics
- Advanced Topics in Statistics
- Applied Linear Statistical Modeling
- Data Mining
- Independent Study: Time Series Analysis
- Relational Database Systems

Western Washington University **September 2013 — December 2017**

- Probability and Statistical Inference
- Nonlinear Optimization
- Linear Optimization
- Linear Algebra
- Mathematical Modeling