Colin Middleton

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

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
Software Tools	Visual Studio Code, Anaconda, Jupyter Notebooks, Excel, Power BI
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   **September 2019 — June 2021**
Eastern Washington University   *Cheney, WA*

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

Bachelor of Science in Mathematics   **September 2013 — December 2017**
Western Washington University   *Bellingham, WA*

- GPA: 3.5
- Minors: Chemistry
- Honors Program Graduate
- Math Fellow (2015 - 2017)
- Presidential Scholarship (2013)

PROJECT EXPERIENCE

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

- Acquired list of words from Wordle website source code and word frequencies from Kaggle dataset.
- Built 6 component data filter using regular expressions in Python.
- Published user-friendly Streamlit website to suggest Wordle words for users.
- Words sorted and displayed by user preference: Word Frequency, Letter Frequency, Letter Position Frequency.
- https://share.streamlit.io/middlec000/wordler/main/src/main.py

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

- Published a User-friendly Streamlit web interface where users can use an example dataset or upload data.
- Website uses an ANOVA test to determine if there is a significant difference in grades between groups of students with groups from chosen characteristic such as race or neighborhood.
- Automatically performs normality and skedasticity hypothesis tests and informs user of results.
- https://github.com/middlec000/grades_vs_student_characteristic

Colin Middleton

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

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

Entry Level Data Scientist

Data Science Lead / Homelessness Prediction
June 2020 — June 2021
Spokane Predictive Analytics and Master's Thesis Project
Spokane, WA

- Received, matched on composite key, and preprocessed 4,785,131 rows of de-identified data from 8 csv files.
- Communicated with data providers to improve datasets over 5 iterations.
- Performed exploratory data analysis on the monthly billing data representing 91,591 people over a 5-year period using visualization techniques and 4 types of correlation to discover relationships and trends.
- Established data homogeneity across years using ranks of variable-outcome correlations.
- Engineered 2 new cumulative features and 3 new aggregate features.
- Investigated longitudinal prediction approaches using Linear Regression, Cox Proportional Hazards, and LSTM models.
- Investigated non-longitudinal prediction approaches using Logistic Regression and Vanilla ANN models.
- Evaluated and compared models based on Receiver Operator Characteristic curves and Area Under the Curve. The Logistic Regression model performed best with an AUC of 0.81 - similar to current research that incorporated richer data sources.
- Implemented K-Folds and minority class oversampling to make model performance more reproducible and combat extreme class imbalance - only 0.39% of cases were of the positive class.
- 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 - retaining only 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 find the number of clusters to form using the Elbow Method.
- Currently developing a fast hierarchical text document clustering Python package as an extension of this work.
- `https://github.com/middlec000/fhdc`

Data Scientist / Traffic Prediction
January 2017 — December 2017
Whatcom Council of Governments and Senior Project, Western Washington University
Bellingham, WA

- Predicted traffic flow arriving at the Peace Arch and Blaine US / Canadian border crossings based on upflow traffic sensors using averaging, Principal Component Analysis, and curve fitting methods.
- Presented findings to Whatcom Council of Governments and Western Washington University faculty.

WORK EXPERIENCE

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

- Create architectural drawings using AutoCAD LT: floor plans, roof plans, sections, exterior elevations, window and door schedules.
- Assemble drawing sheet set: dimensions, code compliance, viewport setup.
- Communicate work schedule and technical project details.
- Site measure existing buildings for remodel.
- Managed and shared files with Dropbox.

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.
- Communicated work schedule and technical project details.
- Provided feedback and suggestions on design.
- Researched building code and construction product specifications.

Colin Middleton

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

Entry Level Data Scientist

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

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

- Trained as a university level mathematics teacher.
- Completed a final project on making lessons more applicable to the lives of the students by incorporating familiar contexts into mathematical situations.
- Developed 1 lesson plan and taught 2 live remote class sessions (linear algebra).
- Graded homework and exams fairly and systematically.
- Provided feedback and suggestions to students about assignments.
- Provided feedback to professors about students' comprehension of class material.

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

- Graded homework and exams fairly and systematically.
- Provided feedback and suggestions to students about assignments.
- Provided feedback to professors about students' comprehension of class material.

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.
- Communicated work schedule and technical project details.
- Researched building code.
- Managed and shared files with Dropbox.

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, 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 (A)
- Advanced Topics in Statistics (A)
- Applied Linear Statistical Modeling (A)
- Data Mining (A)
- Independent Study: Time Series Analysis (A)
- Relational Database Systems (A)
- Data Structures (A)

Western Washington University   **September 2013 — December 2017**
- Probability and Statistical Inference (A)
- Nonlinear Optimization (B+)
- Linear Optimization (B-)
- Linear Algebra (A-)
- Mathematical Modeling (A)