BIBOR ILDIKO SZABO
Data Scientist
Vienna, VA | 615.525.6826 | szabo.bibor@gmail.com | GitHub | Portfolio

I am a sociologist turned data scientist with a strategic eye for finding the real root of a problem. I drive to find the optimal data solution for complex projects that is highly accurate and technically sound. With a multicultural background, I bring a wealth of interpersonal experience and perspective to reach across boundaries of communication with clients and team members.

SKILLS
- **Inference:** data collection, qualitative and quantitative data analysis, hypothesis testing.
- **Prediction:** data mining, supervised and unsupervised machine learning, predictive modeling, regression, classification, natural language processing/ NLP, data visualization, recommender systems.
- **Tools:** Python, Pandas, Numpy, Sklearn, Gensim, SQL, PostGreSQL, Matplotlib, Seaborn, Tableau, SPSS, Command Line, Git, MS Excel.

RELEVANT EXPERIENCE
Teaching Assistant – Data Science Immersive | General Assembly, Washington, D.C. | 08/2020–Present
- Teaching assistant to 35 students in the 12-week long remote full-time immersive data science program.

Data Science Immersive Program | General Assembly, Washington, D.C. | 03/2020–06/2020
Completed 5 projects and 20 labs in a 500-hour full-time course, and presented my work to both technical and non-technical audiences.
- **Science or science fiction?** – Using the Pushshift API, Natural Language Processing, and logistic regression, I predicted the category of a social media post with 94.5% accuracy.
- **Hunting for features that matter** – I analyzed and visualized trends, used feature engineering, and fitted over 15,000 machine learning models to find the most viable one that was capable of predicting housing prices from a large number of features with the least possible error.
- **Coronavirus** – As a group, we collected data from 3 different websites, used NLP to conduct sentiment analysis on COVID 19 and social distancing-related twitter posts, and analyzed daily changes in new coronavirus cases on the state level. To provide an easily understandable visual display, we used Tableau to create an interactive map.
- **Capstone** – Using the Doc2Vec model of the Gensim library, I created a Django application, a search engine that recommends federal contract opportunities based on topic similarity.

Research Assessor | Peabody Research Institute, Nashville, TN | 03/2016–05/2017
- As part of a team, I supported two projects that parallelly focused on identifying potential areas of enhancement in the public education system to foster longer-term academic performance among under-served pre-Kindergarten-aged students.
- Conducted full-day classroom observations of the learning and social-emotional environment, teacher attitude, and quality of instruction. I coded and recorded the observation data.
- Assessed the cognitive development of students, using the Peabody Picture Vocabulary Test, Peg Tapping, and the Woodcock-Johnson III. psychoeducational battery. I coded and recorded the assessment data.

Survey Research Methods Seminar | Vanderbilt University, Nashville, TN | 01/2009–05/2009
- Used SPSS to conduct a multivariate regression analysis and hypothesis testing on the 2008 Nashville Health Study dataset to assess the impact of acculturation and discrimination experiences on the mental health of immigrants.

EDUCATION
General Assembly, Data Science Immersive, Washington, D.C. | 03/2020–06/2020
Middle Tennessee State University, M.A. in Sociology, Murfreesboro, TN | 12/2007
Eötvös Loránd University, J.D., Budapest, Hungary | 06/2002