**SOPHIA BAHAR ROOT**

sophiabaharroot@gmail.com | <https://www.linkedin.com/in/sofishka/> | +1 (341) 400 1438 | <https://github.com/sophiabahar>

**Personal Statement:** I am a highly motivated and analytical professional with a strong foundation in data science, applied statistics, and public health research. My diverse experience, from managing academic research projects to developing machine learning models, has honed my ability to extract actionable insights from complex datasets. With hands-on proficiency in Python, SQL, and R, as well as expertise in statistical modeling and data visualization, I have successfully delivered data-driven solutions in both academic and professional environments. I am currently pursuing an MSc in Applied Statistics with Data Science, where I am expanding my skills in big data analytics and statistical analysis. My goal is to leverage my technical expertise and problem-solving skills to contribute to impactful data science projects that drive innovation and informed decision-making. I thrive in collaborative settings and am eager to apply my knowledge to real-world challenges in a dynamic data-focused role.

**Technical Skills:** Python, SQL, R, SPSS, Qualtrics, Google Suite, Microsoft Office, Zoom, Slack, Smartsheet

**Soft Skills:** problem solving, critical thinking, communication skills, collaboration/teamwork, curiosity, adaptability/flexibility, attention to detail, time management, creativity/innovation.

**Languages:** English (native), French (B2 – DELF), Turkish (B2), Arabic (A1)

**PROJECTS & PUBLICATIONS**

* *Hotel booking cancellation model*: Built a hotel booking cancellation prediction machine learning model using logistic regression, support vector machine (SVM), decision tree, and random forest in Python, achieving an accuracy of a range from 79 - 99% on a dataset of 36,275 bookings.
* *Netflix movie recommendation system*: Built a movie recommendation system using knowledge/rank based recommendation system, similarity-based collaborative filtering, and matrix factorization based collaborative filtering in Python on a dataset of 100,004 observations.
* *Exploratory data analysis*: Analyzed variables related to diabetes in the Pima tribe on a dataset of 768 entries by (1) plotting graphs, (2) finding the mean, median, and mode, (3) identifying the interquartile range, and (4) creating a correlation matrix.
* Bahar Root S, Clarey E. Regulation of antidepressant prescription for the effective treatment of mental disorders in Turkey: A narrative review. *Psychiatry Clin Psychopharmacol*. 2022;32(3):250-261. DOI: [10.5152/pcp.2022.22408](https://www.psychiatry-psychopharmacology.com/en/regulation-of-antidepressant-prescription-for-the-effective-treatment-of-mental-disorders-in-turkey-a-narrative-review-133180)
	+ This non-systematic narrative review aimed to summarize the results of clinical studies evaluating the effectiveness of antidepressants used to treat mental illnesses in Turkey.
	+ Conclusions drawn from this article can guide ongoing efforts by Turkish health policymakers to improve public health development in the country by further regulating the prescription of antidepressants.
* Root, Sophia, "The Relationship Between Language, Emotional Intelligence, and Cultural Sensitivity" (2019). Honors Theses. 55. DOI: <https://doi.org/10.33015/dominican.edu/2019.HONORS.ST.14>
	+ The present study looked at emotional intelligence and cultural sensitivity levels between monolinguals and multilinguals by collecting data through an electronic survey.
	+ Using SPSS, a statistically significant positive correlation was found between cultural sensitivity subscales and emotional intelligence subscales. That is to say that an individual that has a high level of cultural sensitivity is also bound to have a high level of emotional intelligence.

**EXPERIENCES**

Research Assistant | Boğaziçi University Computational Linguistics Lab | (remote) Istanbul, Türkiye | September 2024 – present

* Assisting Professors Şaziye Betül Özateş, PhD and Emine Sümeyra Turaliemre, PhD on the project titled *Development of a Deep Learning-Based Data Expansion Tool for Predicting the Efficacy of Antibacterial Nanoparticles Used in Nanomedicine*.

Public Policy Analyst | UC San Francisco | San Francisco, California, USA | October 2021 – present

* Managing 3 Academic Senate committees and supporting 6 by: conducting administrative tasks; writing letters, minutes, agendas, reports, and grant proposals; and developing surveys and conducting data analysis to improve faculty experience.
* Drafting cover letters for University of California systemwide reviews as a part of the university policy review process to ameliorate university function and adherence to evolving state and national regulations.
* Collaborated on the development of pilot projects: (1) a learning module to educate laboratories on optimal sustainable practices; (2) writing services to help faculty with manuscript and grant writing; and (3) currently developing a task force dedicated to standardizing data and metadata in repositories.
* Developed surveys on Qualtrics, analyzed target populations for data collection, analyzed data and created graphs for various projects, including a report on writing services recommendations.
* Created a committee management system on Smartsheet to streamline workflows.

Behavior Technician | Individualized ABA Services For Families| San Francisco, California, USA | January 2022 – August 2022

* Implemented applied behavior analysis (ABA) principles to provide therapy to children with autism spectrum disorder.
* Managed 2 patients’ data with CentralReach (ABA data collection software) to document and graph patient behavior.

English Language Instructor | American Language Center | Meknes, Morocco | September 2019 - July 2020

* Applied diverse teaching techniques to enhance over 50 students’ English language proficiency.
* Built a digital grading system in Excel and digitized language learning using Google classroom.

**EDUCATION**

MSc Applied Statistics with Data Science, University of Strathclyde | September 2024 - present

* Modules: Foundations of Probability & Statistics, Data Analytics in R, Statistical Modelling & Analysis, Big Data Fundamentals, Big Data Tools & Techniques.

AS Mathematics, Peralta Community College District (Laney College) | June 2023 – June 2024

* Modules: Calculus I, Calculus II, Calculus III, Linear Algebra, Genetics Lab

Certificate Data Science and Machine Learning, Great Learning | November 2022 - March 2023

* Modules: Foundations of Data Science, Making Sense of Unstructured Data, Regression & Prediction, Classification & Hypothesis Testing, Deep Learning, Recommendation Systems, Networking & Graphical Models.

MSc Public Health and Health Promotion, Brunel University London | September 2020 – September 2021

* Modules: Implementing Change in Public Health and Health Promotion, Global Public Health, Health & Society, Health Policy, Politics & Social Justice, Health Economics, Health Promotion, Epidemiology.

BS Psychology, BA International Studies Honors, Dominican University of California | August 2015 – May 2019

* Relevant Modules: Statistics for Health/Behavior Sciences, Advanced Experimental Design, Macroeconomics.