

EDUCATION

COLUMBIA UNIVERSITY, Mailman School of Public Health

New York, NY

Master of Public Health (MPH), **Epidemiology/Biostatistics**. Honors: Heilbrunn Scholar

Expected May 2019

Coursework included: Applied regression I & II, Data science I & II, Relational database & SQL for data science, Python for data analysis, Latent variable and structural equation modeling, Research methods & application, Analysis of categorical data, Design & conduct of observational studies, Application of epidemiologic research methods.

FIRST MOSCOW STATE MEDICAL UNIVERSITY

Moscow, Russia

Doctor of Medicine (MD)

July 2017

- Clinical Knowledge in Oncology, Cardiovascular & Infectious diseases
- Graduated Summa Cum Laude

SKILLS

Proficient in R, SAS, Python, SQL, Git, GitHub

R tools: tidyverse, shiny, plotly, caret

Python tools: NumPy, pandas, matplotlib, seaborn, scikit-learn, keras

SAS Tools: SAS/Base, SAS/Stat, SAS/Graph, SAS/ SQL, SAS/ODS, SAS/Report

Database: MySQL, MS Access

Programs: Microsoft Office Suite (Excel, PowerPoint, Access and Outlook), Mplus

Machine Learning: Regression, Feature engineering, SVM, Decision Trees, Principal Component Analysis and K-means

CERTIFICATIONS

-
- SAS Certified Base Programmer for SAS 9 (Verification number: B431ETW12MBQQEKV)

DATA ANALYSIS / RESEARCH EXPERIENCE

Graduate Research Assistant

September 2018 - Present

Database Shared Resource, Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center

- Serve as SAS programmer for 2 large-scale cancer studies and collaborate with statistician/data managers in analyzing initial data sets and generating tables, listings and figures (TLFs) for clinical studies using EHR, clinical and claims data
- Develop SAS code for modeling data using SAS/STAT procedures such as PROC LIFETEST, PROC LIFEREG, PROC PHREG, PROC REG and PROC GLM for Survival analysis, logistic regression and others.
- Analyze and validate data sets and SAS outputs with other programmers' outputs using PROC COMPARE, PROC CONTENTS, and PROC FREQ. Created formats for the coded data and used PROC SQL for data validation, significantly increasing accuracy for complex, previously inconsistent data.
- Perform and manage various duties such as data transformation and manipulation processes; creating and modifying new and existing SAS programs; and producing Ad hoc reports for cancer studies to contribute to publication of articles.

Graduate Research Fellow – Data science

May 2018 – August 2018

Global Healthy Living Foundation, New York, NY

- Conducted analysis of data from various research for publications, and patient-reported outcome data of over 18,000 observations to support advocacy efforts and comparative effectiveness studies using SAS
- Used statistical techniques including regression and chi-squared for hypothesis testing to validate data and interpretations
- Social media (Facebook & Twitter) data mining/scraping using Python for qualitative analysis (Sentiment analysis, word cloud)
- Successfully analyzed and interpreted web traffic (painspot.com) data using data visualization and dashboards to draw conclusions for managerial data intelligence and strategy using SAS, SQL and R
- Designed scalable dataflows and ETL (Extract, Transform, Load) logic for data collection and management
- Maintained study datasets, authored codebooks, and managed data distribution to researchers

Evaluation Intern - Health Research Training Program

September 2018 – October 2018

New York City Department of Health and Mental Hygiene (NYCDOHMH)

Bureau of Chronic Disease Prevention and Tobacco Control

- Conducted intercept surveys and observational studies speaking to general public and community leaders in neighborhoods through NYC.
- Accurately compiled and organized data and observations into an electronic database

TEACHING EXPERIENCE/OTHER

Graduate Teaching Assistant, SAS programming (Statistical Methods)

January 2019 – Present

Columbia University Mailman School of Public Health

- Competitively selected to facilitate student learning in topics including linear regression, logistic regression, ARRAYS and DO LOOPS, statistical testing, ANOVA, correlation, debugging and confounding
- Conduct lab sessions and provide guidance to graduate students on statistical data analysis in SAS and collaborate with a team of 6 teaching assistant and faculty

Community Teaching Assistant - Data Science: Wrangling

Provided by Harvard University (HarvardX)

January 2019 – Present

- Provide training and guidance to students from diverse technical and non-technical backgrounds on data wrangling in R programming language.
- Engage in dialogue with a vibrant community of global learners on programming methods such Data import, Web scraping, String processing with regular expressions (regex), dates and times and Text mining

Graduate Teaching Assistant - Determinants of Health

Columbia University Mailman School of Public Health.

September 2018 – December 2018

- Ensured understanding of complex science and social science topics for 40 graduate students from a broad range of backgrounds.
- Collaborated with numerous stakeholders including faculty, teaching assistants and administration to execute curriculum

Medical Assistant

August 2016 – June 2017

MPC Social Services, Moscow, Russia

- Managed patients by taking history and vitals for diagnosis
- Counselling and educating patients on healthy lifestyles and medications to improve their health and general wellbeing
- Ensured the place is ready to receive patients for treatment and consultation

LEADERSHIP EXPERIENCE

- National President, National Union of Ghana Students, (NUGS) Russia. 2015 - 2016
Served as the leader of international organization with 1000 + members
- Founder, MedApp – Android and Apple mobile app for student of First Moscow State Medical University 2014

CONFERENCE PARTICIPATION & PRESENTATIONS

- Cambridge University International Model United Nations Cambridge, UK. 4th – 6th November 2016
- The European International Model United Nations, The Hague, Netherlands. 10th – 18th July 2016.

PROJECTS (More projects and code available at <https://github.com/aminyakubu>)

- **Histologic Cancer Detection** – Kaggle data science competition. Challenge was to create an algorithm to identify metastatic cancer in small image patches taken from larger digital pathology scans. Ranked in the **top 8% worldwide**. 1st Place Area Under ROC 0.9822, my score 0.9754
- **Santander Bank Customer Transaction Prediction** (\$65,000 prize Kaggle competition). Challenge was to identify which customers will make a specific transaction in the future. Methods used: Ridge Regression, Lasso, PCR, XGBoost, LightGBM. Currently in the top 51% worldwide. **1st place has Area Under ROC score of 0.907, my score 0.896.**
- **NYC Crime data analysis** - Led a team of 4 to transform a 1.8 GB of real-world dataset of incident crimes to the NYPD into meaningful analysis. Created an interactive visualization by using Shiny, flex dashboard and presented results by making a dedicated website. Website can be found: <https://aminyakubu.github.io/nyc-crime>