Charlie Baker

Data Scientist - Icahn School of Medicine

New York, NY - Email me on Indeed: indeed.com/r/Charlie-Baker/661198ea63d0e561

Energetic, early career data scientist currently working in the health space. Passionate about organizing and harnessing the world's health data. Expert in analysis and visualization in Python, SQL, and VBA.
Willing to relocate to: New York, NY
Authorized to work in the US for any employer

WORK EXPERIENCE

Data Scientist

Icahn School of Medicine - New York, NY - November 2015 to Present

Established team's first data science practice via the following steps:

● Met with a variety of stakeholders (including MDs, PhDs, IT specialists, project managers, and compliance personnel) to gather requirements and desired analysis outcomes of analytics pipeline.

● Worked with IT personnel to access the organization's data warehouse and establish data pipelines for both host team and partner teams.

● Referenced data dictionary to write complex joins in SQL to merge patient, diagnosis, treatment, and outcome tables. Wrote functions to identify incorrect and incomplete data.

● Using Python, generated reports with calculations and data visualizations, and presented reports at weekly meetings to audience of 10 healthcare professionals.

● Documented pipeline to transfer knowledge to peers and superiors, and trained new hires in methodology and data science programming languages. Set up GitHub repository to ensure singular distribution of data science processes.

● Selected publication: "A web-based tool to facilitate shared decision-making regarding neoadjuvant chemotherapy use in muscle-invasive bladder cancer." Journal of Clinical Oncology (pending)

● References: William K. Oh, MD | Matthew D. Galsky, MD

Research Assistant

Lowery Lab at Boston College - Boston, MA - September 2013 to May 2015

Conducted entire experimental pipeline in study of growth cone kinetics, on up to 6 conditions in tandem:

● Designed and executed wet lab experiments and collected image data via time-lapse microscopy.

● Performed image processing in MATLAB to generate time-series microtubule dynamics dataset.

● Created and distributed templates in Excel to automate analysis of image processing output.

● Selected publication: "Xenopus TACC1 is a microtubule plus-end tracking protein that can regulate microtubule dynamics during embryonic development." Cytoskeleton

EDUCATION

Research Education and Engagement for Data Science

Icahn School of Medicine - New York, NY
June 2017

Bachelor of Science in Biochemistry in Jazz guitar

Boston College - Boston, MA
May 2015

SKILLS

data science (1 year), Excel (1 year), MATLAB (1 year), Python (1 year), SQL (1 year)

ADDITIONAL INFORMATION

Skills
Software and Programming Languages: Python (NumPy, pandas, matplotlib, seaborn, scikit, Jupyter notebooks, SymPy), R, MATLAB, SQL, Oracle SQL Developer, VBA, UNIX shell, HTML/CSS/JavaScript, GitHub, MS Applications (Access, Excel, PowerPoint, Word)
Data Science Skills: Data Gathering, ETL, Data Cleaning, Visualization, Analysis, Machine Learning, Publishing, Presentation