Mark Roepke

markroepke.me • github.com/markroepke • mroepke5@gmail.com • 419-889-0724 • Fort Myers, FL

PROFESSIONAL EXPERIENCE

September 2019 – Present   *Databricks*   *Remote*
Curriculum Engineer - Data Science and Machine Learning
- Design and develop instructor-led data science and machine learning courses for Apache Spark, Python and Scala
- Designed and developed the Databricks Certified Associate Developer for Apache Spark certification exam
- Designed and developed the Databricks Certified Associate ML Practitioner for Apache Spark certification exam
- Co-maintain a Python and Scala library of internally-facing course development tools
- Co-designed a self-paced data science and machine learning curriculum for Apache Spark, Python and Scala
- Writing internally-facing question-writing guide to assess the application of code in different question formats
- Review self-paced and and instructor-led data science and analytics courses for instructional and content quality

June 2018 – September 2019   *84.51°*   *Cincinnati, OH*
Senior Data Scientist - Digital Personalization
- Designed and developed a complementary recipe search system for Kroger digital platforms utilizing gensim's Doc2Vec in Python and delivered as a service API via Flask
- Developed and deployed production-grade Shiny app to collect feedback from 50+ co-workers for recommendation systems to gather training data and baseline metrics prior to placement on digital platforms
- Co-designed a production-path blueprint to accelerate the delivery of sciences within an OKR/Agile framework
- Co-designed a personalized recipe recommendation system utilizing layered collaborative filtering and knowledge-based systems to inspire 2,000,000+ Kroger customers with new recipe ideas
- Co-developed an internal data science training philosophy to be detailed in an upcoming co-written white paper
- Co-planned and co-delivered semi-annual intermediate R and introductory Python data science workshops
- Constructed and developed an opinionated and consistent R package ecosystem for company-wide use to assist data scientists with I/O, branded data visualization and communication, and automation with a team of data scientists

February 2019 – September 2019   *University of Cincinnati*   *Cincinnati, OH*
Python Instructor
- Designed, developed, and delivered data-focused Python workshops for the Center for Business Analytics

June 2016 – June 2018   *84.51°*   *Cincinnati, OH*
Data Scientist - Customer Data Foundations
- Created, automated, and communicated a suite of customer-level supervised machine learning models to provide foundational Kroger customer preference data assets to 100+ data scientists and 100+ clients
- Utilized Oracle PL/SQL, R, and DataRobot and migrated to Apache Spark and Python prior to platform change
- Actualized a comprehensive customer-level continuous customer price sensitivity measure using XGBoost in R
- Designed and delivered introductory R and Tidyverse workshops to 100+ data scientists
- Authored an internal company R package for use by 100+ data scientists with another data scientist
- Evaluated data science platforms like Domino and IBM Data Science Experience for company use as a team
- Inherited and maintained analytical methodology for a customer-level segmentation in SAS
- Created, evaluated, and expanded company-branded ggplot2 theme and color palettes

May 2015 – August 2015   *U.S. Bancorp*   *Cincinnati, OH*
Quantitative Financial Analyst (Intern)
- Conducted econometric time series analyses in R for financial stress testing purposes
- Identified determinants of demand and effects of policy changes for various financial products
- Developed financial statement report automation macros in VBA for branch offices

September 2014 – March 2014   *Beech Acres Parenting Center*   *Cincinnati, OH*
Fundraising Data Analyst (Intern)
- Selected for volunteer project utilizing fundraising analytics for non-profit organization
- Analyzed over ten years of donation data in Stata to identify effective fundraising techniques via logistic regression

PERSONAL PROJECTS, TALKS, AND OPEN-SOURCE CONTRIBUTIONS

Personal Projects
- *Introduction to Data Science with Python* - github.com/uc-python/intermediate-python-datasci
- *Intermediate Data Science with Python* - github.com/uc-python/intro-python-datasci
- *Personal website in Jekyll* - markroepke.me
- *Butler County Donut Trail Shiny app* - markroepke.shinyapps.io/butler_county_donut_trail

Talks
- *Word Embeddings with Python and Spark* - github.com/markroepke/presentations/tree/master/word-embeddings

Open Source Contributions
- *flxr* - Developed an R package to scrape basic IMDB movie data
- *skimr* - Contributed whitespace calculation to EDA package on CRAN (Github PR #380) - 298,000 downloads
- *cattonum* - Added median encoding to feature-encoding package on CRAN (Github PR #10) - 11,000 downloads

---

EDUCATION

2018 – 2019   *Johns Hopkins University*   *Remote*
Non–degree-seeking, Graduate Computer Science
Courses: Introduction to Programming, Data Structures, Computer Organization, Software Engineering (GPA: 3.925)

2012 – 2016   *Xavier University*   *Cincinnati, OH*
Bachelor of Arts
Major: Economics; *Concentration*: Mathematics; *Minors*: Statistics & Business (GPA: 3.628; Quantitative GPA: 3.86)
Relevant Projects:
- Built causal impact model to estimate the economic impact of minor professional ice hockey
- Built gravity-influenced migration model to test the political dissimilarity hypothesis among refugees

---

AWARDS, SERVICE & CONFERENCES

- 2020 CodingNomads #StayAtHome Mentorship Program - Mentor two aspiring data scientists
- 2019 RStudio::conf Conference Attendee and Advanced RMarkdown Workshop Completion
- 2018 RStudio::conf Conference Attendee and Tidyverse Train-the-trainer Workshop Completion
- 2017-2019 Board Member - Xavier University Economics Department Board of Executive Advisors
- 2018-Q2 Extraordinary Contribution Award - The Kroger Company
  - Selected from all Kroger employees for enabling organization's data science teams
- 2016 Craig Menderson Memorial Award - Xavier University Economics Department
  - Awarded to an economics major demonstrating excellence in the classroom and leadership potential
- 2016 Xavier University Mortar Board

SKILLS & CERTIFICATES

Data Science and Machine Learning
- Skills: ML workflow, natural language processing, recommendation systems, deep learning, and ML operations
- *Databricks Certified Associate ML Practitioner for Apache Spark*

Data/Software Engineering
- Skills: Data pipelines, data streaming, functional programming, internal tool/library design
- *Databricks Certified Associate Developer for Apache Spark*

Education and Instructional Design
- Course and lesson design, mind maps, instruction, technical assessment
- *RStudio Tidyverse Instructor*

Technologies
- Languages: Python, Scala, SQL, R, Java, Oracle PL/SQL, SAS, Stata
- Data Technologies: Apache Spark, Delta, Databricks, RStudio, Jupyter, Oracle SQL Developer

- Other Technologies: Unix/Linux, AWS, GCP, Azure