# Peter Yonka
## DATA SCIENTIST

San Francisco, CA | [peteryonka@gmail.com](mailto:peteryonka@gmail.com) | 917-602-1693 | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

## SUMMARY

I'm a data scientist dedicated to exploring the hidden stories our data can tell us about what has been and what may be to inform our decisions and connections in the present. Driven by a passion for learning, understanding, connecting, and sharing these insights, I offer diverse experience in collaboration across disciplines to shape observations into compelling narratives in areas ranging from predictive pricing models to marketing the next big Broadway show.

## SKILLS

Python, Pandas, NumPy, Matplotlib, Seaborn, SciKit-Learn, TensorFlow/Keras, StatsModels, SciPy, Google Colab, Amazon Web Services, Jupyter, HTML, CSS, APIs, SQL, Git/GitHub, Apache Spark, Collaboration, Project Management, Marketing (digital and print), and Communications Strategy

## EXPERIENCE (Selected)

**General Assembly Data Science Immersive**, Remote                                                       **2021**
Data Science Fellow
- Leveraged the power of Python and various support libraries to find the value hidden in complex, real-world data by performing analyses and statistical modeling ranging from regression to neural networks.
- Received focused, daily, hand-on instruction to reinforce best practices in programming, usage, and interpretation of machine learning models.

**Berkeley Repertory Theatre**, Berkeley, CA                                                       **2014 - 2019**
Earned Income Data Manager
- Modeled attendance and ticket pricing utilizing Tessitura (CRM and ticket sales platform), SQL, proprietary access tools, and Excel that resulted in the top four grossing shows in the theatre's 50-year history each exceeding their goals by more than 20%.
- Monitored and analyzed single-ticket inventory for dynamic pricing/promotion opportunities that resulted in an average of over $730,000 in additional income each season, setting a single-season record for ticket revenue of more than $9.9 million.

**Magic Theatre, Inc.**, San Francisco, CA                                                       **2011 - 2014**
Data and Analytics Manager
- Implemented a new ticketing system built on Salesforce that decreased ticket-processing expenses by over 40% and streamlined bookkeeping improving data quality and reporting efficiency.
- Collaborated with the board treasurer to design a standardized financial reporting package, budgeting package, and forecasting tool using QuickBooks and Excel that increased transparency and communication with the board.

**LesConcierges, Inc.**, San Francisco, CA                                                       **2011**
Global Operations Analyst
- Executed daily analytical assessments and reporting using custom SQL data pulls, SharePoint, and Excel for LesConcierges' largest client (a top financial services provider), providing timely, actionable data to account managers and executives along with recommendations to enhance and maintain service levels.

**Barter Theatre**, Abingdon, VA                                                       **2007 - 2010**
Marketing Data Manager
- Furnished data analysis in Excel to all administrative departments to aid in revenue forecasting, measurement of program efficacy, and grant preparation.
- Provided strategic planning, branding, audience development, and marketing support for all earned-revenue programs resulting in a 24% increase in earned revenue.

## PROJECTS

**Reddit, White, and Blue:** In Python, I scrapped Reddit posts from three different political subreddits using the pushshift.io API and utilized natural language processing to train a model to predict the correct subreddit for a set of test posts. (Tools used: Python, SciKitLearn, Pandas, BeautifulSoup, Matplotlib, NumPy)

**COVID-19 Symptoms and Predictions:** Utilizing Carbon Health's clinical data repository, my contributions to this project were focused on building an interpretable model to assess the influence of patient-reported symptoms in prioritizing potential positive cases. (Tools used: Python, SciKitLearn, Pandas, StatsModels, Matplotlib, NumPy, imbalanced-learn)

## EDUCATION/CERTIFICATIONS

**General Assembly**, Data Science Immersive, Completed
**The University of Michigan, Ann Arbor, MI,** BFA