LAURA ELLIS

LauraEllis@Outlook.com
(617) 751-2800 (home)
(301) 254-9520 (cell)
10 Plateau Pl, Unit T; Greenbelt, MD 20770

SCAN to see my DATA PORTFOLIO with links to my GitHub, LinkedIn, and Tableau Public profile located at https://lauraellis36.wixsite.com/dataportfolio

Excellent communication skills.

Enjoys working as part of a team or independently. Skill with database design, migration, and implementing functionality.

Able to look at problems from multiple perspectives to find the best solution after considering all options.

Outstanding time-management skills leading to deadlines met, often ahead of schedule.

Ready to jump in and help others when needed.

PROFILE

I love working with data and I enthusiastically take on new challenges. I'm looking for a position where I can put my data skills to work in a role that presents data related challenges, and their resolutions benefit the company. I prefer working with large data sets and complex problems.

SKILLS

Languages/scripting
- SQL
- Python
- R

Programs
- SAS EM
- SPSS
- Microsoft Excel/Access
- IBM Watson/Cognos

Visualization/Dashboards/ Presentations
- Tableau
- Excel
- Power BI
- PowerPoint
- Data Studio

Strategies
- DMAIC
- Six Sigma
- Agile
- NPS
- Process mapping

Large data sets
- US Census, ACS
- MIMIC III
- NIH/NSF Grants
- Amazon (API)

Processes
- Data collection
- ETL/Virtualization
- Data cleaning and exploration
- Model development
- Statistical analysis
- Predictive analytics
- Forecasting, trends
- Reporting (creation, modification, automation)
- Natural Language Processing

Big Data
- Hadoop
- HDFS
- Spark
- HBase
- Hive
- MapReduce

Cloud Computing
- Azure
- Google
- AWS
- Oracle

Machine Learning – Supervised
- Regression
- Classification
- Naïve Bayes
- Random Forest
- Neural Networks
- Support Vector Machines
- Pattern recognition

Machine Learning – Unsupervised
- Clustering (K-means, Hierarchical, Probabilistic)
- Compression (PCA)
- Anomaly detection
- Apriori
- SVD
- Association mining

Governance and Ethics

Project documentation and training

Certificates/MOOCs
- Six Sigma Green Belt (MSI)
- Data Science Foundations (JHU)
- Micro Masters, Bioinformatics (UMGC)
- Data Science Essential Statistics (Microsoft)
- VBA

EXPERIENCE

Director, Research Analytics, Government Funders Division 2021-present
Digital Science, Cambridge, MA

- ❖ Provided support for government and funding organizations. Create custom analyses to gauge collaboration, impact, influence of scientific research including researcher/program/organization level measures. Use of attributes including bibliometrics, scientometrics, and altmetrics.
- ❖ Analysis initiatives including bibliometrics (citation, altmetrics, and NLP evaluation) used to identify key growth areas, international and organization level collaboration in multiple fields of research including multiple category designations as identified by Digital Science or custom categories determined by client.
  - Google BigQuery (SQL), SDK, Data Studio
  - Python, Gephi, VOSViewer

Contractor 2016-2021
Various positions including data work for the DC Forensic Nurse Examiners, legal transcription, technical support for IT company. (During this time, my primary focus was education.)

Project Manager (Project Assistant from 2002 to 2006) 2002-2015
IMV, Limited, Columbia, MD

- ❖ Lead generation project for manufacturers of lab equipment. Mail and email campaigns sent to over 500,000 individuals semi-annually. Created database and structure. Merged large mailing lists from multiple sources by creating ID based on name, zip, and grant recipient data. Generated hundreds of leads delivered on a biweekly basis to over 50 customers. End of year report summarizing lead generation by equipment for purpose of determining spending trends.

- ❖ Customer satisfaction survey for healthcare imaging equipment manufacturer. Received 5 to 10 data files today that were merged and matched. Created databases, merged large datasets, filtered based on project requirements (ex. limited number of contacts per site per quarter, quotas for modalities, and 87 US regions. Reported urgent matters daily. Monthly and year to date reports issued monthly and distributed to key personnel.
- ❖ Census study collecting hospital equipment usage at US Hospitals (based on AHA data). Created database, modified for additional data collected, reported on quarterly basis in custom software.
  - Database design and modification (RDBMS)
  - ETL (CSV, XLS), ID assignment, deduping
  - Data creation, selection, modification
  - Provide end-user technical support
  - Progress tracking and monitoring, quota management
  - Data analysis using appropriate models
  - Report design, automation, generation, maintenance

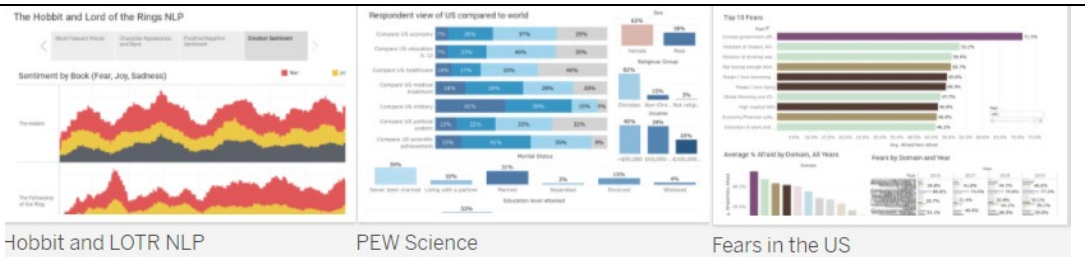
EDUCATION

Master of Science, Data Analytics, 4.0 GPA — 2019-2021
University of Maryland Global Campus
Recipient of President's Scholarship; Member: Phi Kappa Phi, Upsilon Pi Epsilon

Master of Science, Bioinformatics, 4.0 GPA (1/2 completed) — 2018-2019
University of Maryland Global Campus

Bachelor of Arts, Classical Studies — 1995-1999
University of Florida

Projects in my data portfolio…

	Algorithm	Topic
Supervised	Naïve Bayes	Heart Attack Survival *Classification to identify the most influential factors and how those factors influence the predicted survivability of a heart attack at the hospital and up to a follow-up appointment.*
	Decision Tree	Image Classification *The decision tree groups observations by variable values and predicts their membership in classes.*
	Neural Network	Heart Attack Survival *Neural network to predict the survival of heart attack patients admitted to the hospital. Rather than focusing on the attributes (previously done with Naïve Bayes Classification), the emphasis of this analysis will be on the accuracy of predictions of the outcome (living or deceased).*
Unsupervised	Association Rules/Apriori	NIH Spending Categories *Goal is to determine common groupings of spending categories assigned to grants that may indicate how funding could be reallocated to other spending categories upon the dissolution of an NIH entity.*
	Clustering	Turkiye Student Evaluation *Three unsupervised learning clustering methods were used in this analysis: agglomerative hierarchical clustering/nesting (AGNES), divisive hierarchical (DIANA), and k-means clustering.*
	Dimension Reduction	Heart Attack Survival *A new data set was created using PCA (DS1). Principal component 1 (PC1) accounts for 43.6 of the data variance.*
Natural Language Processing		Myriad Genetics *Data patterns and recommendations based on annual reports and Lexis Advance newspaper articles.*
Predictive Analytics		Public Library eBook adoption *This project focuses on public libraries' integration of eBooks into their catalog. Using Census data this analysis predicts the expected increase in digital materials and explores public libraries' ability to meet those goals.*
Tableau		

DATA PORTFOLIO located at https://lauraellis36.wixsite.com/dataportfolio