Data Science Portfolio Presentation

Lessa Fleming

Syracuse University

lgersten@syr.edu

Agenda

- Introduction
- Course Overview/Learning Objectives
- Projects and Business Applications
- Conclusion

Lessa Fleming

- Born in Syracuse, NY
- Associates in Paralegal Studies
- Bachelors in Computer Science
- Masters of Science in Applied Data Science (2024 Graduation)
- 2 years in analytics at United Healthcare
- Currently a Senior Claims Business Process consultant at UHC

Syracuse University's Applied Data Science program is a program that works with two schools of study to prepare a student upon graduation to gain insights from data and provide actionable feedback on those insights.

After choosing a track and completing all the required courses upon graduation a student should be able to demonstrate a range of skills including:

- Collect, store, and access data by identifying and leveraging applicable technologies.
- Create actionable insight across a range of contexts (e.g., societal, business, political), using data and the full data science life cycle.
- Apply visualization and predictive models to help generate actionable insight.
- Use programming languages such as R and Python to support the generation of actionable insight.
- Communicate insights gained via visualization and analytics to a broad range of audiences (including project sponsors and technical team leads.
- Apply ethics in the development, use, and evaluation of data and predictive models (e.g., fairness, bias, transparency, privacy).

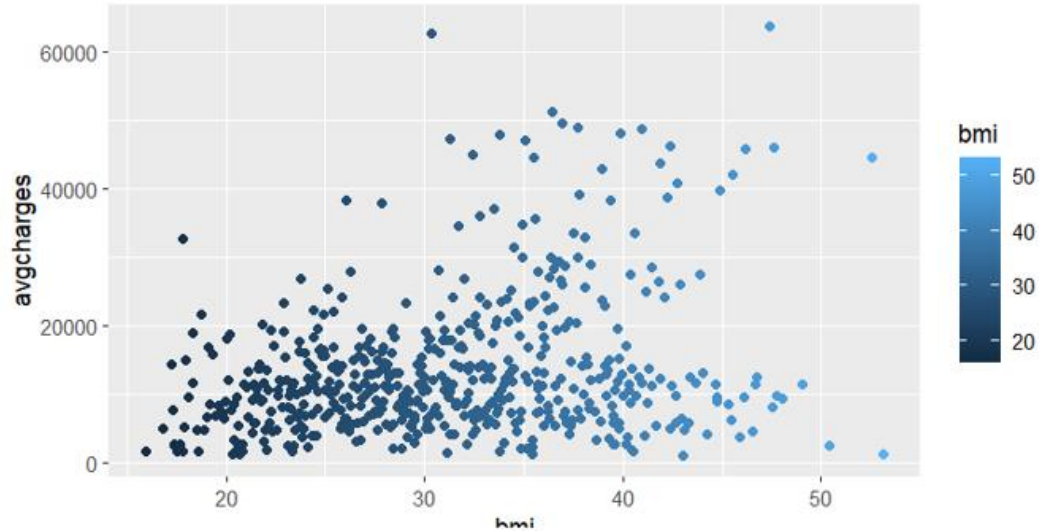
Courses	
Data Administration Concepts and Database Management IST 659	Introduction to Data Science IST 687
Applied Machine Learning IST 707	Big Data Analytics IST 718
Quantitative Reasoning for Data Science IST 772	Business Analytics SCM 651
Scripting for Data Analysis IST 652	Advanced Big Data Management IST 769
Cloud Management IST 615	Introduction to Information Security IST 623
Information Visualization IST 719	Applied Data Science Portfolio IST 782

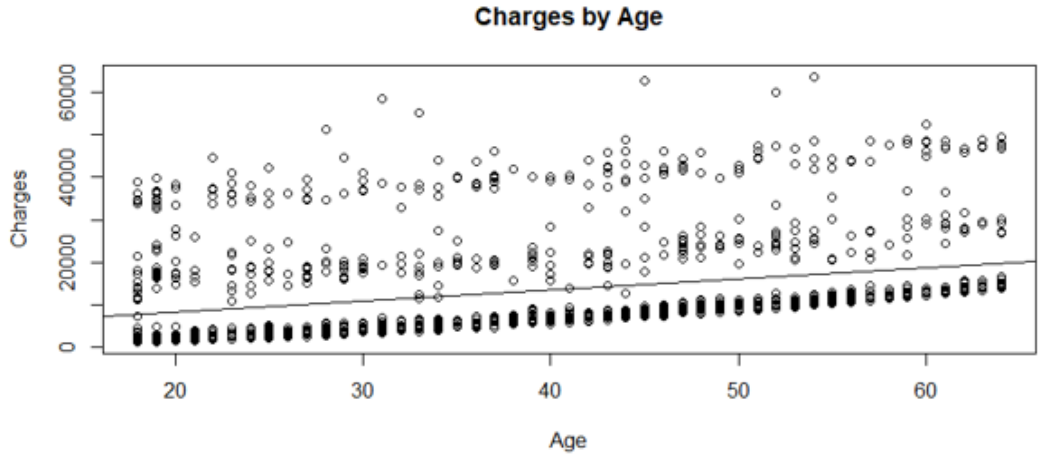
During my time in the Applied Data Science Masters program, I was able to highlight the skills and requirements in these 5 course projects

- Introduction to Data Science – Variables that affect health insurance charges
- Scripting for Data Analysis- AirBnB Prices and what affects them
- Applied Machine Learning- Data Scientist Salaries
- Big Data Analytics- Honey Bee Production
- Advanced Big Data Management- SC weather Analysis

Introduction to Data Science

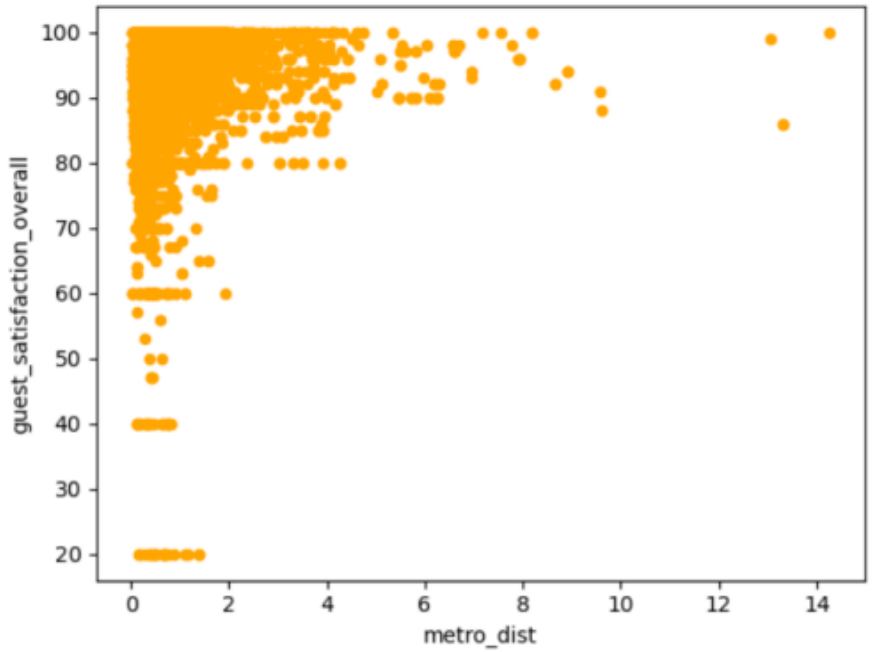
- Identify a problem and the data needed for addressing the problem.
- Perform basic computational scripting using R and other optional tools.
- Transform data through processing, linking, aggregation, summarization, and searching.
- Organize and manage data at various stages of a project life cycle.
- Determine appropriate techniques for analyzing data.

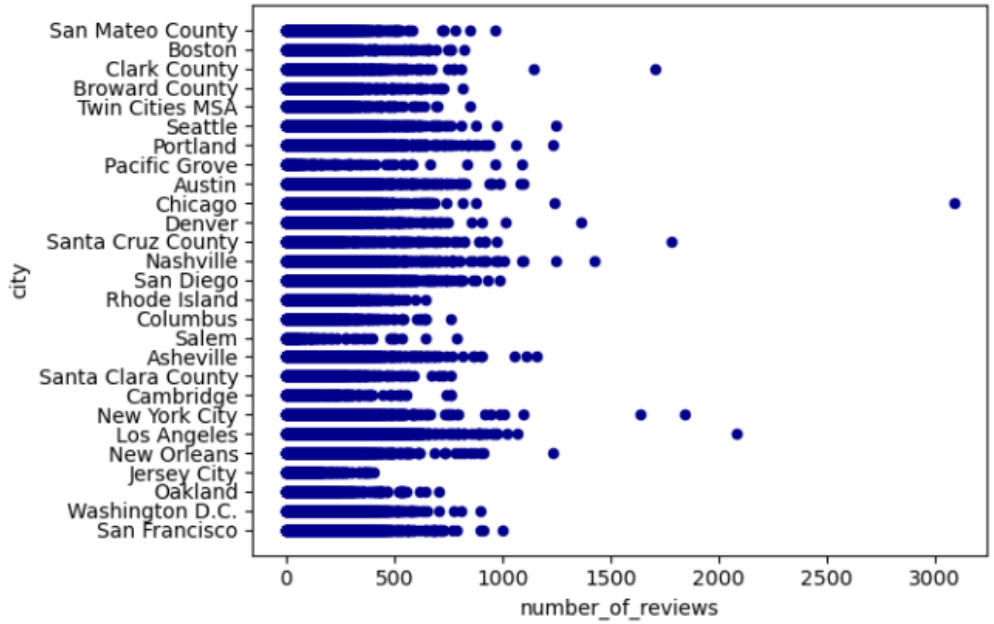




Scripting for Data Analysis

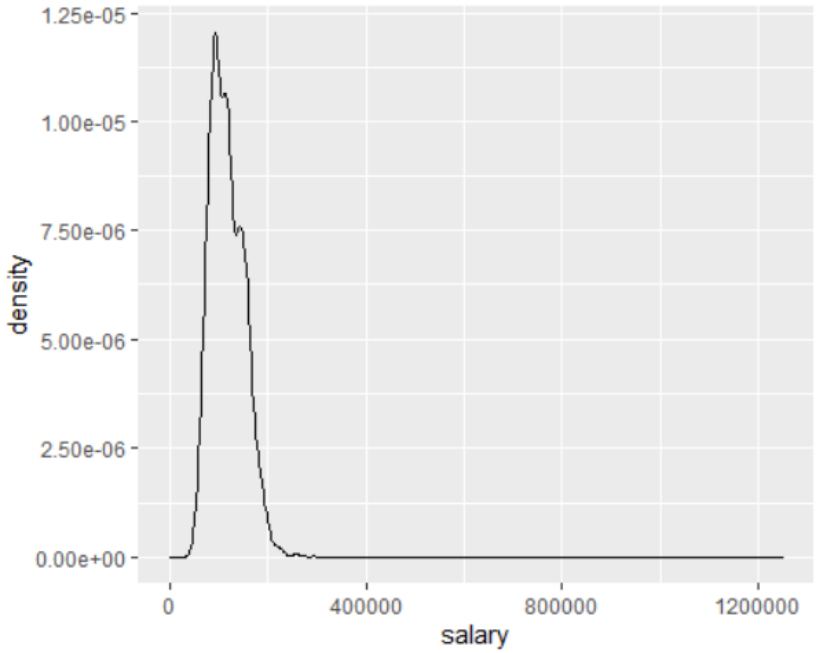
- Obtain a dataset
- Create Business questions
- Organize, clean, and analyze data
- Provide visualizations
- Actionable feedback and data results





Applied Machine Learning

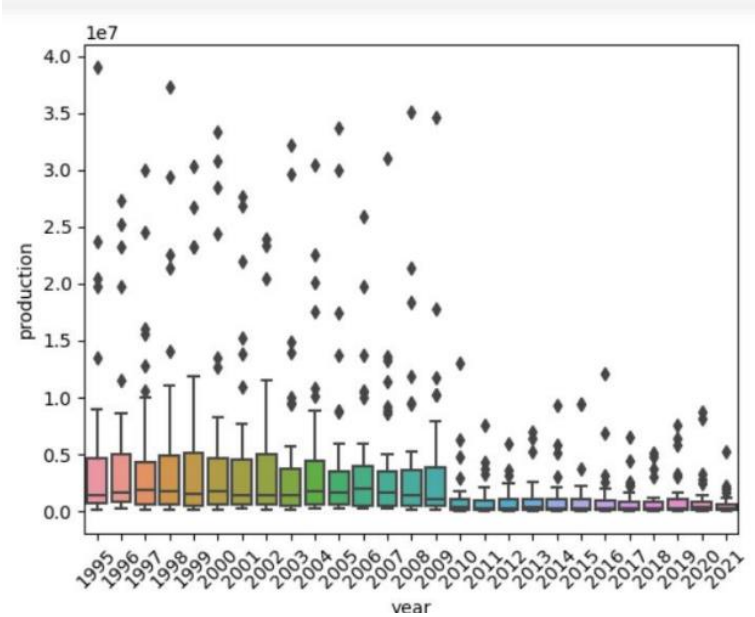
- The objective of the project is to use the main skills taught in this class to solve a real data mining problem.
- Obtain a data set
- Clean, organize, and prepare data
- Analyze and provide actionable results.

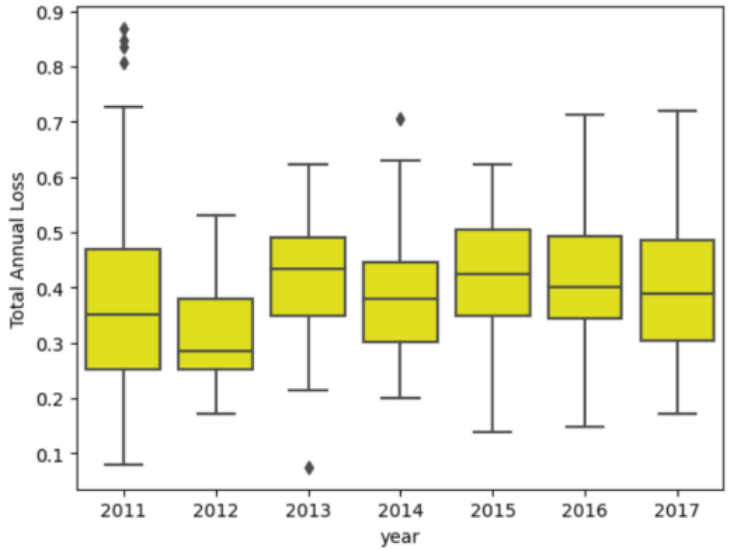


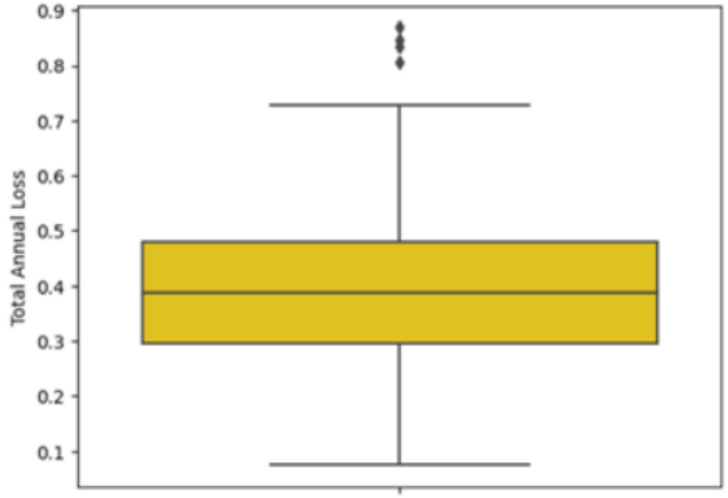
	startdate <dbl>	mean_salary <dbl>
1	2019	110556.
2	2020	115906.
3	2021	121755.
4	2022	129672.

Big Data Analytics

- Students will identify a data-focused problem, bring together different data sources, conduct analysis, draw conclusions, and produce a report explaining the results. Students are to demonstrate the ability to select the appropriate analytical methods for the chosen problem; interpret the data, model, analysis, and findings; draw appropriate conclusions; and present the results in a meaningful way.





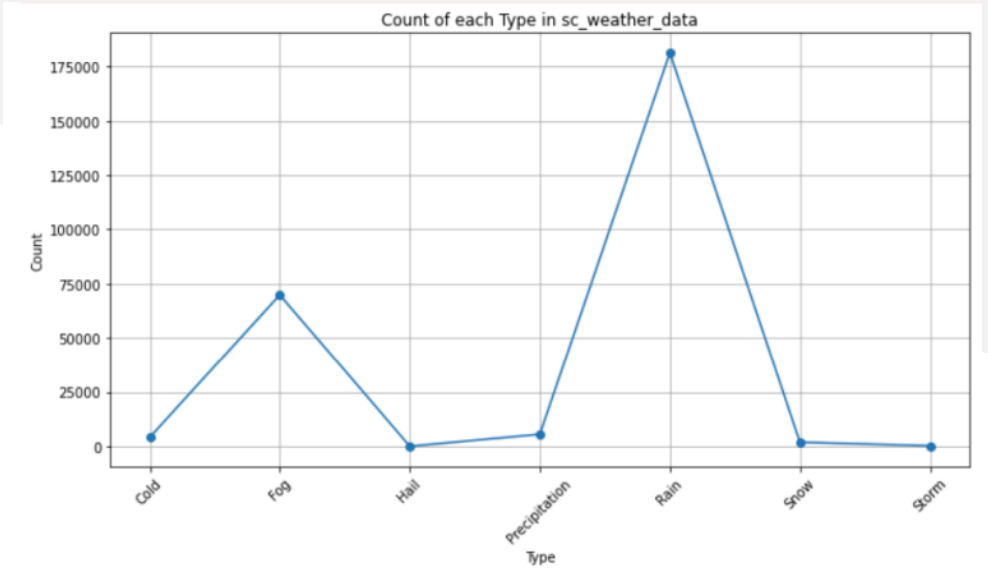


Advanced Big Data Management

summary	Type	Precipitation(in)	ZipCode
count	262938	262938	262938
mean	null	0.09266907027512254	29484.194977523217
stddev	null	0.3796076585291063	237.19485055474027
min	Cold	0.0	29020
25%	null	0.0	29360.0
50%	null	0.0	29527.0
75%	null	0.06	29649.0
max	Storm	9.99	29926

- Work in self-assembled teams of one to three students.
- Devise your own case study and data set to work on.
- Identify outcomes that you intend to do, and why.
- Data sourcing

```
2542    4.549203
2543    5.289536
2544    5.010973
2545    5.116749
2546    5.077549
2547    5.093035
2548    5.087892
2549    5.090531
2550    5.090235
2551    5.091046
2552    5.091439
2553    5.091990
2554    5.092481
2555    5.092995
2556    5.093500
2557    5.094009
2558    5.094516
2559    5.095024
2560    5.095531
2561    5.096039
2562    5.096547
2563    5.097054
2564    5.097562
2565    5.098070
2566    5.098577
2567    5.099085
2568    5.099593
2569    5.100100
2570    5.100608
           ...
2747    5.182485
2748    5.182954
2749    5.183423
2750    5.183892
2751    5.184361
Name: predicted_mean, Length: 210, dtype: float64
```



Conclusion

Having a background in computer science, exploring analytics and data science felt like an exciting transition. The courses I engaged in and the skills I refined were pivotal in shaping my personal and academic growth. The projects and coursework have equipped me to advance in my current career, enabling me to offer more profound insights into our data. By persistently practicing and refining my skills, I am preparing myself to transition into a data science role within my company.

https://lfleming21.github.io/index.html