

# DRUCILA LEFEVRE

✉ lefevre.drucila@gmail.com ☎ (646) 982 - 5347 in [linkedin.com/in/drucila-lefevre](https://www.linkedin.com/in/drucila-lefevre) 📄 <https://github.com/drucilal>

## SKILLS

**DATA ENGINEERING:** MySQL, PostgreSQL, Spark, S3, AWS Athena, Apache Cassandra, Hadoop, Azure, SSIS, MSSQL, Hadoop, Google Cloud Platform, SQL Server, ETL, TFS, Databricks, Airflow, MIG, Control M, Cloud Composer, Azure Data Factory, Kubeflow  
**PROGRAMMING LANGUAGE:** Python, SQL, R, Scala, Powershell, C#  
**FRAMEWORKS AND TOOLS:** Git, Bash, AWS, Docker, Flask, Tableau, AWS QuickSight, Dataiku, Web Scraper, Linux, CICD Pipelines  
**MACHINE LEARNING:** Scikit-Learn, NLTK, Supervised Learning, Unsupervised Learning, Deep Learning

## EXPERIENCE

**Microsoft, Software Engineer, Charlotte, NC**

May 2022 - Current

- Gain an understanding over time of Microsoft businesses and processes and collaborate with stakeholders to contribute to a cohesive set of capabilities built to deliver a frictionless experience
- Define estimates and roadmaps, contribute to strategy elements – assist my team from ideation to articulation to delivery
- Design, implement, validate, and release components/services/systems using my software engineering skills in an Agile team
- Work in a modern DevOps culture, administering my service in Azure, implementing infrastructure as code, automating manual operations processes, and supporting/ troubleshooting production live site incidents
- Design and implement accurate and actionable monitoring to maintain business Service Level Agreements and system health
- Mentor team members, pair program, participate in code reviews, and delegate work to encourage professional and technical growth of those around me
- Be constantly learning, evaluating both existing and emerging technologies, delivering POCs and evaluating needs for service modernization and enhancements
- Collaborate with other teams and disciplines and drive improvements in engineering across our organization

**Bloomingdale's, Data Engineer, New York, NY**

Aug. 2020 - May 2022

- Integrate data from multiple databases across different platforms such as Oracle, SQL Server, Teradata, and Hadoop to integrate data and build a data lake and analytical data marts.
- Assisted in development of real-time recommendation server to provide users in session & historical behaviors.
- Work with business partners on the full life cycle of data-related projects from requirements analysis and design through development.
- Managed 20 engineers and led the cloud migration from Microsoft Azure to Google Cloud Platform.
- Developed design standards and use cases for Reporting platforms.
- Trained internal data engineers to administer GCP, Azure, and Hadoop platforms and helped them performance tune these environments.
- Developed ETL architecture and dimensional data modeling for metadata design and development of self-service tools.
- Work with the Lead Data Architect to determine high-level data strategy for the Enterprise Data Warehouse within a Microsoft Azure environment.

**Columbia University, Adjunct Associate Faculty, SQL and Relational Databases, New York, NY**

Jan. 2021 - Current

- Provide subject matter expertise and supports the instructional process for a course section.
- Attend all online class sessions, assist with instruction, lead breakout sessions, facilitate discussions.
- Lead class lectures, instructional activities, and classroom discussions. Attend all class sessions.
- Evaluate, grade student work and assessments as requested by the course Lecturer.
- Monitor and address student concerns and inquiries.

**Insight Data Science, Data Engineer Fellow, New York, NY**

May 2020 - Aug. 2020

- Built a platform to provide sports industries, inclusive of various sports to aid the analyst in gaining deeper insights.
- Automated the whole batch processing pipeline with Airflow scheduling, decrease the processing run time by 35%.
- Developed a batch processing pipeline to fetch multiple sports data statistics with python, normalize data format using Dask, and stored results in PostgreSQL for web access.

**JPMorgan Chase & Co., Software Engineer Virtual Intern, Remote**

June 2020 - July 2020

- Interface with a stock price data feed and set up your system for analysis of the data.
- Implemented the Perspective open source code in preparation for data visualization.
- Used Perspective to create the chart for the trader's dashboard.

**New York City Data Science Academy, Data Science Fellow, New York, NY**

Sept. 2019 - Dec. 2019

- Developed a ETL pipeline to fetch 10 GB of google merchandise store data with BigQuery, normalized data format with MySql, and stored the results in Python for further analysis to be done.
- Implemented a data pipeline for drivers using Python for transforming data and R Shiny to visualize pipeline which showed that 50% of traffic tickets resulted in speeding traffic violations.
- Optimized machine learning algorithms (Regularized Regression, Random Forest and XGBoost) using scikit-learn in python to predict house sale prices in Iowa and obtained a 90% accuracy score.

**Columbia University Medical Center, Research Analyst, New York, NY**

Feb. 2017 - May 2019

- Decreased task friction by utilizing Python and Matlab to optimize neural data cleaning which increased preprocessing performance optimization.
- Aided 10 interns in the preprocessing of neural data by developing a simple step-by-step program in Matlab for teaching interns about ingesting and transforming multiple neural data at once.

**KindLab, Research Science Intern, Brooklyn, NY**

Sept. 2018 - Dec. 2018

- Implemented a processing batch script for incoming research data using SPSS and Python which was able to process ~500 MB of data in a short period of time compared to Excel.
- Performed statistical tests on social psychological research data which indicated a positive correlation between the number of kind acts and increases in happiness.

## EDUCATION

**Columbia University City of New York**

2016 - 2018

Masters of Arts: Psychology

-- Concentration: Applied Statistics

-- Relevant Coursework: Applied Regression Analysis, Causal Inference, Introduction to Programming in Python, Introduction to Programming in R, Multilevel Longitudinal Data Analysis, Multivariate Analysis I, Probability & Statistical Inference, Programming in MatLab

**College of Staten Island (CUNY)**

2011 - 2015

Bachelor of Science

-- Minor: Statistics

-- Deans List (3 years)