

Mathew Perez

(469) 964-6005 | matperez755@gmail.com | mathewperez.github.io | GitHub: MathewPerez | LinkedIn: mathew-perez

Education

University of Texas at Dallas | **MS** — *Computer Science*

Richardson, TX | **Aug 2019 - Dec 2020**

- **Concentration:** Intelligent Systems

University of Texas at Austin | **Honors BBA** — *Management Information Systems*

Austin, TX | **Aug 2016 - Jun 2019**

- **Certificate:** Computer Science

Experience

Facebook | **Incoming Data Engineer**

Menlo Park, CA | **Dec 2020**

Facebook | **Data Engineer Intern**

Menlo Park, CA | **May 2019 – Aug 2019**

- Designed, built, and tested Data Swarm pipelines that synthesized over 6 months of data from multiple organizations providing over 10 new specialized and aggregate metrics
- Created custom Jinja template to implement intuitive data dashboard with four bottom line metrics and over 12 graphs
- Worked closely with mobile platform engineers to correctly source more than six client and server side logging data tables for ETL design and implementation
- Formulated three different Presto-generated metrics and graphs with ML Engineers and Product Managers to prove value of adding my created metrics to current ML models

University of Texas at Austin | **Undegraduate Research Assistant (Dr.Rajiv Garg)**

Austin, TX | **Feb 2019 – May 2019**

- Applied Python NumPy library and csv module to clean and transform over 5 GB of image tensor data to feed into Convolutional Neural Network
- Built Java application through restfb API and login dialog through Facebook JavaScript SDK to automate social media research survey

Facebook | **Facebook University Data Analytics Intern**

Menlo Park, CA | **Jun 2018 – Aug 2018**

- Created Data Swarm pipeline improving accuracy of over two years of data to produce over 30 valuable metrics
- Overhauled highly-used data dashboard with over 120 aggregate and rolling metrics and over a dozen charts processed daily

Skills

Languages Python, SQL, Java, Scala, C/C++, R, Swift, PHP, HTML/CSS

Tools & Tech UNIX/LINUX, Git, Spark, Hive, Hadoop, Presto, DataSwarm, NodeJS, Flask

Projects

Austin Environmental Directory Website | **Senior Capstone Project** — *Team of 2*

February 2019 - May 2019

- Overhauled the MariaDB backend point to decrease web page load times and simplify website maintenance for stakeholders
- Modernized the Directory website by rewriting over 20 PHP and CSS files

Podcast Search and Summary Algorithms | **Graduate Research** — *Thesis Project*

May 2020 - Nov 2020

- Using Deep Learning NLP techniques in PyTorch on over 13GB of podcast data to create relevant summaries of podcast episodes
- Pairing classical IR methods with Deep Learning in Lucene inverted index to build podcast excerpt search engine