Improving Processing/Analysis Time on Large Datasets Using Public Data and On-Demand Computing Methods with the Fairfax County Economic Development Agency (FCEDA)

A Professional Readiness Experiential Program (PREP) Project Effort

----- Authors / Student Project Team Members -----

Rema Alharthi is a student at George Mason University majoring in Management Information Systems with a minor in Operations and Supply Chain Management. She served as the Communication Lead for the FCEDA project, working closely with stakeholders to align the team's technical progress with client needs. Rema is particularly interested in cloud solutions and cybersecurity, and she plans to pursue a master's degree after graduation.

Victor Luque is a student at George Mason University graduating with a bachelor's degree in Management Information Systems and Business Analytics. He has developed strong skills in data analysis, cloud computing, and automation, applying tools like Python, SQL, AWS, and R to real-world projects. With hands-on experience in machine learning models such as neural networks and logistic regression, Victor is passionate about using technology to solve business problems and is eager to contribute meaningfully in data-driven roles after graduation.

Hailey Ortega is a student at George Mason University graduating in August 2025 with a bachelor's degree in Business with a concentration in Management Information Systems. Through this project, she has gained experience navigating scope, cost, and time constraints, as well as a greater understanding of cloud services and geospatial data. After graduation, she will actively seek client-facing opportunities to continue growing her business process management (BPM), business intelligence (BI), and systems thinking skills, with the long-term goal of becoming a PMP-certified project manager.

Chau Phan is a Business Analytics student at George Mason University with a strong foundation in data analytics, cloud computing, and business operations. She is proficient in Python, R, SQL, and PowerBI. Chau aims to pursue a data analyst role where she can apply her analytical and problem-solving skills to real-world business challenges. During an experiential project with FCEDA, she has gained hands-on experience building AWS S3 and Lambda pipelines while optimizing data workflows. Her roles as a Finance Teaching Assistant and Project Manager Intern at Costello College of Business reflect her leadership, technical ability, and commitment to student success and professional development.

Ellie Yang is a student at George Mason University graduating with a bachelor's degree in Business Analytics with a minor in Data Analysis. She has real-world experience as a Market Analyst Intern at Becton, Dickinson and Company and has completed several academic projects focused on data analysis and machine learning using Python, R, SQL and Tableau. Through her PREP project, she developed strong skills in Python programming and data analysis, particularly in working with large datasets and transforming raw data into clean, actionable formats. Given her passion in the field of analytics, Ellie will pursue a Master's degree in Applied Data Science at the University of Chicago after graduation. Her longterm goal is to become a Data Analyst or Data Scientist, using data-driven insights to solve complex problems. ----- Industry Participant / Mentor -----

John Blair Director of National Business Investment Fairfax County Economic Development Authority (FCEDA)

----- Faculty Member -----

Brian K. Ngac, PhD Instructional Faculty, Dean's Teaching Fellow, & FWI Corporate Partner Faculty Fellow George Mason University's Costello College of Business bngac@gmu.edu

Interested in being an Industry Participant and or PREP Sponsor? Please reach out to <u>bngac@gmu.edu</u>, Thanks!

---- Client Testimonial ----

"The PREP team went above and beyond to really understand our needs and delivered a professional solution that impressed us and our senior leadership. It will save us time and also provide us with more intelligence on the business community in Fairfax County."

- John Gustavo Blair | Director, National Business Investment | Fairfax County Economic Development Agency

Introduction

The Fairfax County Economic Development Authority (FCEDA) is an agency that supports economic growth by assisting American and international businesses in establishing, expanding, or relocating operations in the county. Business investment divisions and marketing intelligence specialists rely on accurate, timely data to identify prospective businesses, provide a range of insights and information to support their growth, and analyze trends in business activity and expansion in the area. Public datasets contain valuable information regarding newly registered businesses in Fairfax County; however, their utility is diminished by the manual and timeintensive data processing methods required to generate clean, location-ready datasets.

The FCEDA reached out to the GMU Professional Experiential Readiness Program (PREP) because they had a real need. They were looking for a low-cost, secure ETL solution to automate the cleaning and processing of large business datasets, as well as a dashboard to visualize data intuitively. To meet their needs, we built an ETL pipeline using AWS Services, including AWS Lambda, Amazon S3, and Amazon Location Services. This pipeline takes raw data and transforms it into clean, mappable CSVs, ready to deploy in FCEDA software systems.

Additionally, we created an exploratory dashboard featuring an interactive map that allows users to visualize business locations, apply relevant filters, and quickly extract important insights from the cleaned datasets.

Business Challenge

Our team was given datasets from the Virginia SCC as a test case for our pipeline. We utilized records from various business entity types, including corporations, limited liability companies (LLCs), general partnerships (GPs), and limited partnerships LPs. In total, we were tasked with processing over **1.85 million records**. Before our project, these datasets were manually cleaned and filtered manually using Microsoft Excel. However, this was not only time-consuming due to the volume of data (not to mention that the combined dataset records exceeded Excel's 1,048,576 Row limit), but the dataset also contained several quality issues. The main concern areas we discovered during the requirements gathering and analysis phase were the following:

- Datasets contained incorrect address fields, including unclean street names, city names, and zip codes.
- Necessary fields contained missing values.
- Datasets do not contain geographical coordinates (longitude/latitude) to support map visualizations.
- Several business addresses within the county exhibit errors or overlap with those of neighboring jurisdictions. Businesses registered under specific ZIP codes may not be located within Fairfax County, which complicates efforts to accurately and quickly identify businesses within the county.
- Cleaning and filtering business records is currently done manually, which is laborious and negatively impacts the ability to conduct outreach and analysis effectively.
- FCEDA cannot accurately label data and distinguish between industrial-based or officebased business records to support and improve business targeting and analysis efforts.

We created the following **AS-IS process** to highlight the concern areas that we sought to address.



Figure 1 FCEDA AS-IS Process

Steps:

- 1. Download data from public websites and other data sources.
- 2. Manually clean, filter, and label VA SCC business record datasets on Microsoft Excel and store data in a local file. (Format inconsistencies are not fully addressed)
- 3. Retrieve "cleaned" data.
- 4. Upload cleaned data to FCEDA software systems.
- 5. Users review and use it for analytical and outreach purposes.

Activities Done to Address the Business Challenge

We created the following TO-BE process to highlight the areas that we addressed with our proposed solution.



Figure 2 FCEDA TO-BE Process

Steps:

- 1. Staff uploads the latest Corp, LLC, GP, LP CSVs from Virginia SCC into the landing AWS S3 bucket.
- 2. A serverless Lambda pipeline ingests the raw files from S3, standardizes addresses, geocodes new records, filters them to Fairfax County, and enriches each record with officer details and building-type codes.
- 3. An Amazon Location Services Place Index calculates the coordinates (latitude/longitude) for new, incoming addresses.
- 4. The clean dataset is saved to an S3 bucket.
- A cleaned snapshot of business records in S3 is available, ready for deployment to FCEDA software systems including Tableau connects to S3 for automatic dashboard refreshes, ArcGIS, and Salesforce.
- 6. Users explore dashboards, maps, and trends.

Once the data is processed, we integrate the clean data into an exploratory dashboard tailored for FCEDA business investment and market intelligence teams. The dashboard features an interactive map with address points, drill-down exploration, keyword and spatial filters, and business trends analysis to support decision-making and outreach efforts.



Figure 3 Exploratory Dashboard on Tableau

Main Dashboard		PDF		Powerpoint			Image
ame	Company Directory						
	Name1	Street1	City	Zipcode	Business Type	Year of Incorp Date	
eti	1R Technologies, Llc	6564 Loisdale Court	Springfield	22150	LLC	2014	
	1Tech Bargains Llc	7382 Stream Way	Springfield	22152		2024	
corp Date 1910 2022	3CI Technology, Inc.	1709 Pebble Beach Dr	Vienna	22182	Corp	2007	
	3D Technology Inc.	14100 Park Meadow Dr	Chantilly	20151	Corp	2020	
	3Links Technologies, Inc.	8641 Hillside Manor Dr	Springfield	22152	Corp	1999	
	3Rd Eye Technologies Inc.	13800 Coppermine Rd	Herndon	20171	Corp	2016	
th, Year of Incorp Date	3Towerstech, Llc	9497 Lynnhall Pl	Alexandria	22309			
)	 4Ntech Llc 	11604 Holly Briar Ln	Great Falls	22066		2022	
ipcode (All) -	4S Technologies Llc	2466 Leyland Ridge Rd	Herndon	20171	LLC	2018	
	4Xgroup-Techlaw Jv, Llc	14840 Conference Center Dr	Chantilly	20151	LLC	2019	
	5 Points Tech Llc	1640 Boro PI FI	McLean	22102			
nere Tune	5Town Technologies, Llc	1939 Roland Clarke Pl	Reston	20191		2024	
(All) -	9-Tech Connect, Llc	3428 Sleepy Hollow Rd	Falls Church	22044		2018	
	10Forward Technologies .	. 3077 Holmes Run Rd	Falls Church	22042	Corp		
	24X7 Network Technologi	L. 8300 Greensboro Dr Ste L1-	McLean	22102	Corp	2008	
μη .	108 Technologies, Llc	7401 Parkwood Ct	Falls Church	22042		2014	
	127 Technologies, Llc	1519 Woodacre Dr	McLean	22101		2015	
	703Techops, Llc	2552 Huntington Dr	Herndon	20171	LLC	2022	
	1010 Technology Llc	8711 Running Fox Ct	Fairfax Station	22039		2018	
	1911 Technologies, Llc	1911 Virginia Ave	McLean	22101		2023	
	4200 Tech Court, Llc	481 Carlisle Drive	Herndon	20170	LLC	2013	
	5010 Tech, Llc	114 James Dr Sw	Vienna	22180		2020	
	13525 Dulles Technology	277 Windover Ave. Nw	Vienna	22180		1998	
	A Star Cyber Technology L	12234 Fort Buffalo Cir	Fairfax	22033	LLC	2024	
	A Tech Heating & Air Cond	L. 2931-A Eskridge Rd.	Fairfax	22031	Corp	2014	
	A Tech Hvac Services Llc	11851 Monument Dr	Fairfax	22030		2020	
	A True Tech Llc	7852 Birch Branch Ter	Alexandria	22315	LLC	2025	
	A3 Tech Service Incorpora	L. 3812 Pineland St	Fairfax	22031	Corp	2024	
	A-1 Autotech, Inc.	13977 Metrotech Dr	Chantilly	20151	Corp		
	A-Tech Services Company	11596B Cavalier Landing Ct	Fairfax	22030	Corp	1993	
	A/E/C Training Technologi		Great Falls	22066	LLC	1997	
	Aaa Tech Solutions Llc	1900 Reston Metro Piz	Reston	20190		2021	
	Aarka Technologies, Inc.	12799 Thacker Hill Ct	Herndon	20171	Corp	2021	
	Aarntech Llc	12783 Oak Farms Dr	Herndon	20171		2024	
	Aatech Inc.	3127 Wayne Rd	Falls Church	22042	Corp	2020	
	Aayega Technologies Inc.	2513 Wheat Meadow Ct	Herndon	20171	Corp	2020	
	Ab Information Technolog	. 13911 Valley Country Dr	Chantilly	20151	Corp	2007	
	Abbatech Solutions Inc.	6554 Loindala Ct		22160	Corn		

Fairfax County Business Registrations

Figure 4 Business Registration Directory

Results & The Positive Impact

By automating and standardizing the processing of over a million business records with AWS, our tool significantly improved data reliability and usability. We began with messy, unstructured datasets and transformed them into a single, easily searchable and visualizable dataset, ready to support FCEDA's objective of identifying and supporting Fairfax County's fast-growing industries and early-stage businesses, among other use cases.

We achieved the following outcomes:

- Faster Data Processing The new system reduced the time to ~ 30 minutes needed to clean and prepare business data. What used to take hours or days of manual Excel work can now be completed in minutes with automated AWS Lambda functions. This provides FCEDA faster access to accurate data for better outreach and analysis.
- Better Data Accuracy Our tool ensures a clean, standardized structure for downstream functions (geocoding and Tableau dashboards). Additionally, by integrating Amazon Location Services and GIS datasets, the final outputs more accurately reflect business locations specifically within Fairfax County.
- Scalability and Cost-Efficiency The solution operates on a serverless infrastructure that utilizes AWS's pay-as-you-go cost structure. Our solution enables long-term sustainability, allowing FCEDA to easily adapt the tool to changing business needs.
- 4. Empowered FCEDA Staff— With user-friendly Tableau dashboards and CSV export capabilities, non-technical staff can filter, search, and act on business intelligence without needing IT support.
- 5. Stronger Outreach & Economic Insights FCEDA can now proactively engage new businesses aligned with its strategic sectors, such as technology, healthcare, finance, and corporate headquarters, supporting its mission of economic diversification.

Conclusion

Our PREP team delivered a low-cost, cloud-based ETL pipeline that gives FCEDA faster, more reliable access to clean, location-ready business data. By automating what was once a manual, time-consuming process, we helped reduce processing time from days to under an hour and improved data accuracy using AWS Lambda, S3, and Amazon Location Services. The solution filters and geocodes over 1.85 million records and enriches them with valuable business details that empower FCEDA teams to work more efficiently with tools like Tableau, ArcGIS, and Salesforce.

In addition to speeding up data workflows, the system provides a scalable foundation for future enhancements, such as automated updates, more granular filtering, and expanded visualizations. The use of serverless architecture ensures sustainability while keeping costs low, currently under \$100/month. The dashboard equips non-technical users with actionable insights and makes business trends easier to identify and communicate across teams.

Ultimately, this project enhances FCEDA's capacity for proactive outreach and strategic analysis, directly supporting its mission to foster a diverse and thriving economy in Fairfax County.

PREP Student Reflection

Participating in the experiential program and working on the FCEDA project over the past few months was an invaluable experience that challenged us to apply classroom knowledge to solve a real-world business problem. Our team was exposed to a wide range of technologies, including cloud services like AWS Lambda and S3, geospatial tools like Amazon Location Services and ArcGIS, data visualization technologies like Tableau, and process modeling tools like Figma. Most of us had limited prior experience with these platforms, so we had to quickly adapt, self-teach, and collaborate closely to build an effective solution. This hands-on learning helped us develop critical technical skills in data engineering, ETL processes, and cloud architecture.

Beyond the technical components, we also gained experience navigating client expectations, aligning our work with stakeholder needs, and communicating project progress clearly and professionally. We also learned how to create compelling storytelling presentations to communicate our efforts to executives, our professor, and the agency project team. Working with the FCEDA Team also provided insight into how public-sector organizations manage large-scale data and how modern cloud tools can improve efficiency and impact.

Most importantly, this project helped us grow as professionals. It pushed us to think critically, work as a team under tight deadlines and budget constraints, and present our work confidently to faculty, industry mentors, and senior leadership. The PREP experience showed us the value of team collaboration and reinforced our shared passion for data-driven problem-solving. It also highlighted areas for improvement where we can continue to learn and grow as we transition into our future careers. We'd like to thank the FCEDA team, John Blair (National Business Investment), Theresa Rhodes (Market Intelligence), Stephen Tarditi (Market Intelligence), and John Hoeveler (Business Investment) for guiding us through this challenging yet rewarding project. We are proud of what we accomplished and grateful for the opportunity to make a meaningful contribution to Fairfax County's economic development efforts.