

AGRICULTURAL DATA AND AI AT THE USDA

US Department of Agriculture | Local and Regional Foods Division Research Branch

Ashley Chu, Data Fellow | *UC San Diego, Cognitive Science & Machine Learning*

Galiba Anjum, Data Fellow | *Cornell University, Computer Science & Information Science*

Keywords:

Financial analytics, automated reporting, grants awards dashboards, data cleaning, visualization

Summary:

To periodically report grant spending, Ashley designed and developed dashboards and visualizations summarizing Grants Division data in **Tableau** and **Python**. They wrote **SQL** queries and Python code to pull the data from the database in the Enterprise Data Analytics Platform and then parse, clean and analyze it for reporting. The work culminated in an automated report in the form of a **Quarto** notebook to continuously explain more detailed views of the obligations, disbursements, and trends over time.

AGRICULTURAL DATA AND AI AT THE USDA

US Department of Agriculture | Local and Regional Foods Division Research Branch

Ashley Chu, Data Fellow | *UC San Diego, Cognitive Science & Machine Learning*

Galiba Anjum, Data Fellow | *Cornell University, Computer Science & Information Science*

Keywords:

Data cleaning, Python, geographic blocking, data standardization, record linkage

Summary: Galiba standardized and consolidated over 200,000 Local Food Purchase Assistance program records submitted by all U.S. states, territories, and tribal partners through quarterly spreadsheets. Using Python libraries such as pandas and fuzzywuzzy, they built automated scripts for entity resolution by normalizing text, applying fuzzy similarity scoring, and implementing state-level geographic blocking to improve match accuracy and reduce processing time. Location fields were validated and standardized using U.S. Census Bureau datasets. Purchase and distribution records were then linked to track the flow of goods from vendors to recipients. Excel pivot tables were used for quality assurance to identify and correct any remaining errors or inconsistencies in the mapping table. This resulted in a clean and consistent dataset that will support accurate program analysis and reporting in the future.

AGRICULTURAL DATA AND AI AT THE USDA

US Department of Agriculture | Local and Regional Foods Division Research Branch

Ashley Chu, Data Fellow | *UC San Diego, Cognitive Science & Machine Learning*

Galiba Anjum, Data Fellow | *Cornell University, Computer Science & Information Science*

Keywords:

Retrieval-augmented generation, large language models, classification, natural language processing

Summary:

To streamline the **NEPA environmental review process**, Ashley developed a **retrieval-augmented generation model** and **pipeline** that could ingest and reference policy documents and guidelines to determine what level of review a construction project proposal requires. They wrote **Python** scripts to chunk and process the reference texts and embed them into a vector database. The relevant policy texts could then be retrieved and used as context for the **LLM** to quote and use in the classification and explanation it generates.

AGRICULTURAL DATA AND AI AT THE USDA

US Department of Agriculture | Local and Regional Foods Division Research Branch

Ashley Chu, Data Fellow | *UC San Diego, Cognitive Science & Machine Learning*

Galiba Anjum, Data Fellow | *Cornell University, Computer Science & Information Science*

Keywords:

Data visualization, R programming, geospatial mapping, survey analytics, stakeholder feedback

Summary: Ashley and Galiba are analyzing data from the 2019 - 2020 National Agritourism Survey, conducted by the University of Vermont in collaboration with the USDA, which includes responses from over 1,600 U.S. farm, ranch, and vineyard operators. Using R and Python, they are developing reproducible geospatial and statistical visualizations - such as interactive maps to highlight regional participation patterns and comparative charts to reveal relationships between key business factors - to uncover trends, challenges, and opportunities in agritourism. These preliminary visuals are iteratively refined based on stakeholder feedback to ensure they are accessible to producers, partners, and the public, with the goal of informing research, outreach, and policy to strengthen agritourism operations.

AGRICULTURAL DATA AND AI AT THE USDA

Local and Regional Foods Division Research Branch

U.S. Department of Agriculture

Kamran Zendehtdel — Branch Chief

coding it forward >



ASHLEY CHU
UC San Diego
Cognitive Science

GALIBA ANJUM
Cornell University
Computer Science

The Local and Regional Foods Division (LRFD) supports local food producers, markets, and communities by providing easy access to the **ideas, innovations and research** necessary to grow and sustain a productive business.

Local Food Research and Development

Regional Food Business Centers

Local and Regional Food Systems Research and Publications

Local Food Directories

Food Facility Design

The Local and Regional Foods Division (LRFD) supports local food producers, markets, and communities by providing easy access to ideas, innovations, and research in order to grow and sustain a productive business. This information ensures that opportunities for U.S. food producers are readily available and communities are equipped to successfully grow and sell regionally produced foods, while also supporting increased access to locally produced foods.

The [USDA Local and Regional Food Systems Resource Guide \(pdf\)](#) offers a comprehensive summary of USDA programs that support local and regional food systems.

OVERVIEW OF PROJECTS



Grants Division Dashboards and Reporting



Local Food Purchase Assistance Cooperative Agreement Program Data Wrangling



LLMs and RAG for Qualitative Data



Agritourism Data Visualizations

GRANTS DIVISION DASHBOARDS

- Write SQL queries to pull data from the USDA's Enterprise Data Analytics Platform into Tableau and Python notebooks → data wrangling
- Create proof-of-concept dashboards for:
 - Summary statistics to report awards by program, by year, etc
 - Maps of grants over the years
- Automated report in Quarto notebook
- More detailed view of spending over time track obligations, disbursements and trends

USDA Grant Awards Summary Statistics

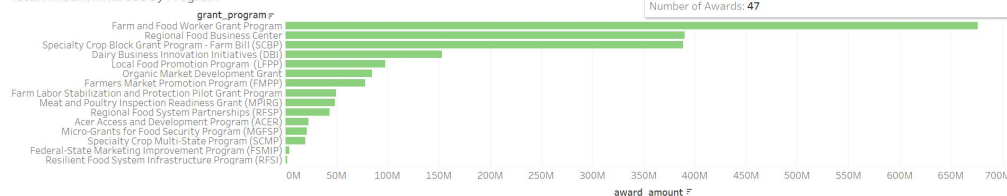
Total Amount Awarded and Number of Awards by Program

grant_program	Award Amount	Number of Awards
Acer Access and Development Program (ACER)	22,826,629	47
Dairy Business Innovation Initiatives (DBI)	153,193,777	20
Farm and Food Worker Grant Program	675,984,056	16
Farm Labor Stabilization and Protection Pilot Grant Program	49,780,000	142
Farmers Market Promotion Program (FMPP)	78,291,301	256
Federal State Marketing Improvement Program (FSMIP)	4,118,707	19
Local Food Promotion Program (LFPP)	97,732,419	254
Meat and Poultry Inspection Readiness Grant (MPIRG)	48,475,148	245
Micro-Grants for Food Security Program (MGFSF)	21,379,930	18
Organic Market Development Grant	85,000,000	107
Regional Food Business Center	390,000,000	13
Regional Food System Partnerships (RFSP)	43,605,458	85
Resilient Food System Infrastructure Program (RFSI)	2,070,799	1
Specialty Crop Block Grant Program - Farm Bill (SCBP)	388,601,341	276
Specialty Crop Multi-State Program (SCMP)	19,434,149	25

Number of Awards by Program

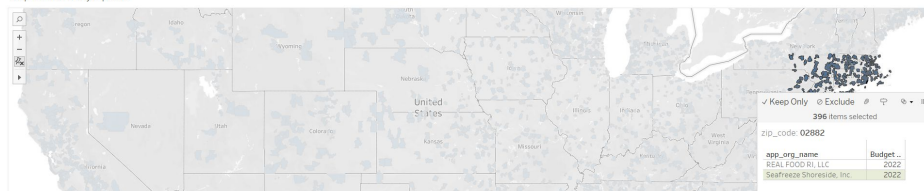
276 Specialty Crop Block Grant Program - Farm Bill (SCBP)	254 Local Food Promotion Program (LFPP)	142 Farm Labor Stabilization and Protection Pilot Grant Program	107 Organic Market Development Grant
256 Farmers Market Promotion Program (FMPP)	245 Meat and Poultry Inspection Readiness Grant (MPIRG)	85 Regional Food System Partnerships	25 Dairy
		47	19 16
			18 13

Total Amount Awarded by Program



grant_number	grant_name	grant_state	award_amount	year
23ACERCT0031	Yale University	Connecticut	499,965	2023
23FMPPMA1454	TRUSTEES OF TUFTS COLLEGE	Massachusetts	497,397	2023
23FMPPNY1177	Cornell Cooperative Extension of Broome County	New York	97,870	2023
23FMPPNY1186	Wildaqua.com, Inc.	New York	74,779	2023
23FMPPRI0106	Rhode Island Commerce Corporation	Rhode Island	185,600	2023
23FMPPCT1244	CTFS BRED INC.	Connecticut	497,996	2023
23FMPPMA1184	Growing Places Garden Project, Inc.	Massachusetts	660,710	2023

Map of Awards by Zipcode



LFPA DATA WRANGLING PIPELINE

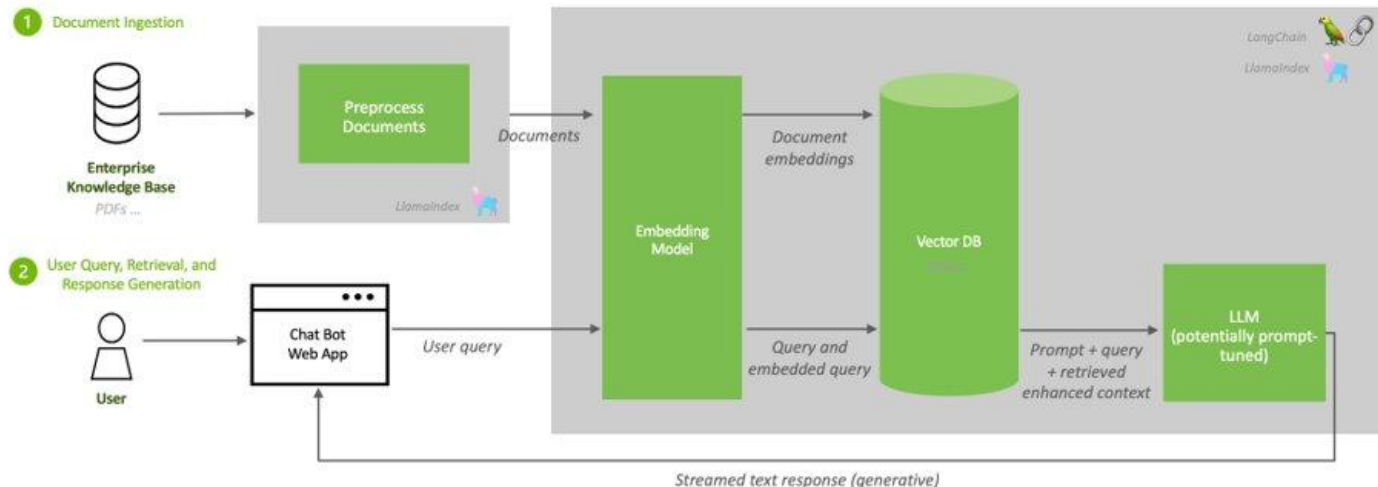
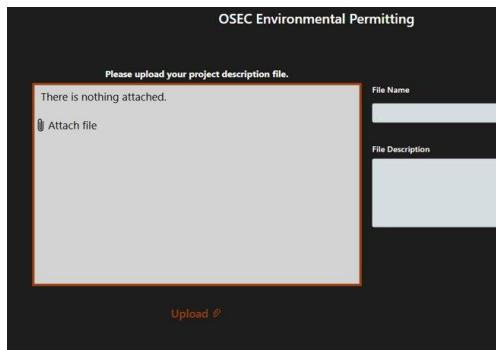
The LFPA dataset (2022–2024), manually submitted by local partners across all 50 states and territories, had minimal data entry standards - leading to data-quality issues which complicated aggregation and analysis of the data

Core Tasks	Approach	Result
Standardize Organization & Producer Names	Used Python libraries for basic normalization of text and applied matching to consolidate variant entries.	Consolidated all duplicates into standardized names.
Geospatial Validation & Normalization	Standardized ZIP codes to 5-digit format and validated city/state consistency via U.S. Census Bureau data.	Corrected ZIPs and validated place assignments.
Link Vendors and Distributors	Performed joins between Purchases and Distributions datasets to align vendor and distributor identities.	Linked vendor and distributor records accurately

RETRIEVAL AUGMENTED GENERATION

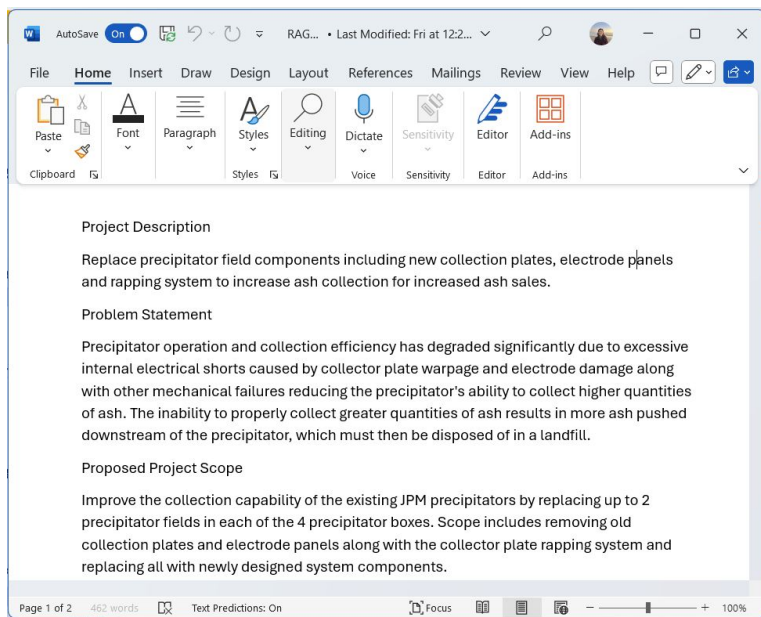
- Grants Division chatbot ↔ classification model for USDA Accelerator Hackathon
- Power App interface ↔ HTTP request via Power Automate ↔ notebooks and models in Databricks
- Embedding model: GTE Large En
- LLM: Claude Sonnet 3.7

Retrieval Augmented Generation (RAG) Sequence Diagram

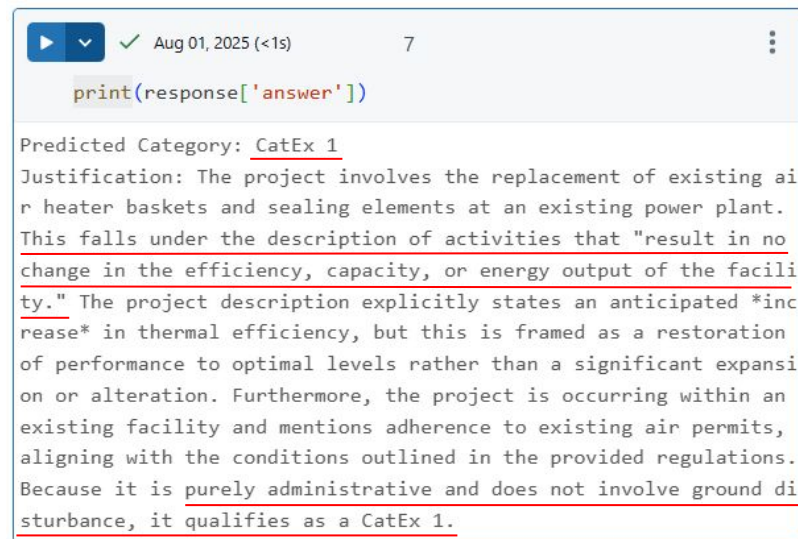


RETRIEVAL AUGMENTED GENERATION

Input

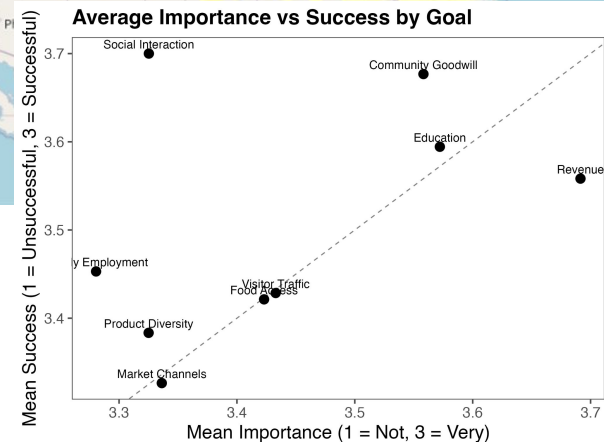
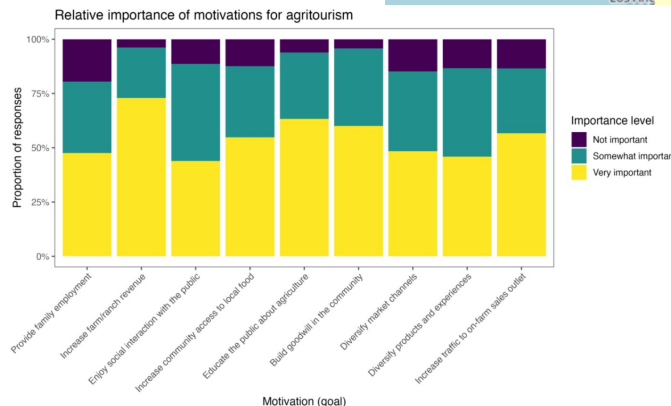
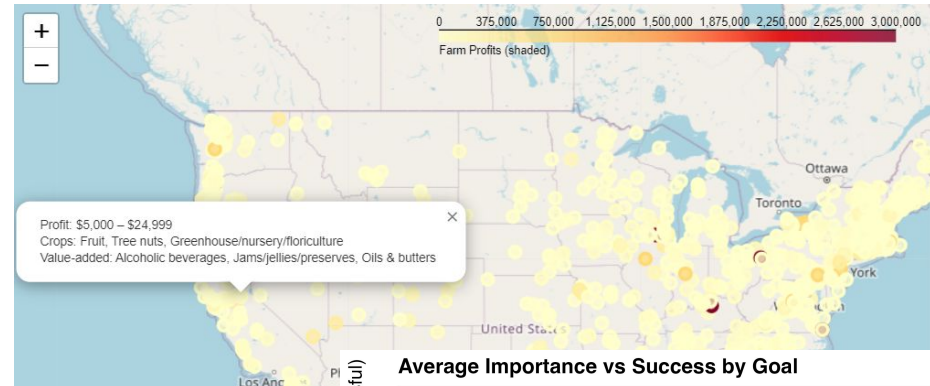


Output



AGRITOURISM STUDY

- Cooperative agreement with University of Vermont to research agritourism
- Analyze and visualize survey data to understand motivations, successes, demographics, challenges, and profitability
- Report for producer reference



LEARNINGS AND TAKEAWAYS

- Unpredictability & Adaptability
 - Glimpse into what it's like working in the federal government
- Learned so much about how data is used at the USDA across all four projects
 - Possibilities of AI to streamline processes + summarize/reference qualitative data
 - Leverage data to make an impact + better support producers across the country



THANK YOU!

Special thanks to Sara Elazan, Elliot Hohn, Kamran Zendehdel, Yuyang Zhong, Cassie Rubio, and everyone we had the privilege of working with this summer!