Project Assignment 4: Research Paper Review and Summary

Lina Saade

G01019181

AIT 580-003

Abstract: In this research project, I used the United States Department of Agriculture's (USDA) avian influenza outbreak data to create visualizations and statistical analyses of how the avian flu could potentially be affecting egg prices in the United States. [1] I also drew from egg price data from the Bureau of Labor Statistics (BLS) to get the average price of eggs in the United States. [2] From these two datasets, I was able to draw a couple of conclusions. Avian flu cases are not increasing exponentially. The outbreak rate ebbs and flows but is not increasing over time. Egg prices ebb and flow but show a more positive trend towards the end of March 2025 and exceed their other peak in March 2023.

Introduction:

Surging egg prices are a hot topic that have cracked their way into political discussions and everyday Americans' lives. Highly pathogenic avian influenza (HPAI H5N1) outbreaks are one of the most discussed factors associated with rising egg prices. HPAI H5N1 can spread via migratory birds to commercial poultry flocks, such as chickens, and even cross species barriers to infect cattle and humans.

Some of the research questions I decided to tackle were:

i. Have we seen an uptick in avian flu cases recently within the United States?

ii. How long does an outbreak typically last?

iii. How might the avian flu affect egg prices?

iv. What is the movement of the avian flu within the United States?

Literature Review:

The literature review for this project consisted of three separate scholarly articles to help set the stage for this research. Each article confirms the research that I was performing but adds a different method or view to paint a more holistic picture of what could be affecting egg prices in the United States.

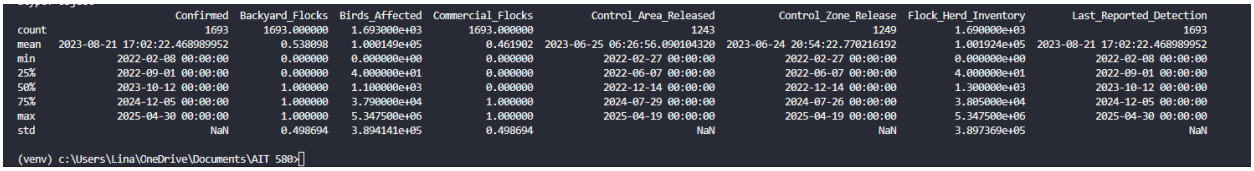
The first article I used, *Why Are Eggs so Expensive? Understanding the Recent Spike in Egg Prices,* gets straight to the point with its title. It explores different factors that go into the 2023 spike in egg prices, such as feed, deaths, and the costs associated with getting new chickens established. Since HPAI H5N1 is highly contagious and can potentially spread to humans through intermittent species, the USDA culls all flocks inflicted with the virus. This results in bird loss, effectively driving down the egg supply. The article suggests that the avian flu is a factor in surging egg prices, but not the only one. Other factors such as feed cost and natural gas costs also contribute to egg prices that seem like they are spiraling out of control. [3]

The second article I used, *The Impact of Highly Pathogenic Avian Influenza H5N1 in the United States: A Scoping Review of Past Detections and Present Outbreaks,* concentrates on the virus without any egg price data involved. It is imperative to understand how the virus moves through different species to ultimately decide how this might directly and indirectly affect humans. The article also provides a historical aspect of how the 2014 – 2015 HPAI H5N1 outbreak traversed. The avian flu was spreading through migratory birds and even spreading to new hosts like cattle. Through the consumption of animal products contaminated with the avian flu, humans could potentially become infected with HPAI H5N1 as well. [4]

Finally, the third article I used for my project was, the *Final Report for the 2014–2015 Outbreak of Highly Pathogenic Avian Influenza (HPAI) in the United States.* This article was created by the dataset originators that I used my report for. This article is more focused on the HPAI H5N1 outbreak itself, how it moves, what hosts the virus is spread through, and how long outbreaks are lasting. I used this article to see what methods USDA used the data for and how I might be able to add to that research with the BLS' egg price data. [5]

Dataset

COLUMN_NAME	NOIR DATA TYPE	ANALYSIS
Confirmed	Interval	Date that case was confirmed- falls into interval scale, because the difference in dates matter.
Incident Site Name	Nominal	State Name.
Backyard Flocks	Ratio	Hot encoding for if the flock affected was a backyard flock or not.
Birds Affected	Ratio	Number of birds affected, can be 0.
Commercial Flocks	Ratio	Hot encoding for if flock affected was a commercial flock or not.
Control Area Released	Ordinal	Date value, sooner is better in relevance to Confirmed date.
Control Zone Release	Ordinal	Date value, repeat of Control Area Released.
County Name	Nominal	County name, nominal.
County, State	Nominal	County and state concatenated.
Flock Herd Inventory	Ratio	Flock numbers can be 0.
Last Reported Detection	Ordinal	Last time virus was detected, relative to Confirmed.
Production	Nominal	Text value for commercial.
Production Type Name	Nominal	Text value for commercial.
Species Name	Nominal	Species of poultry affected.
State	Nominal	State name.



	Confirmed	Backyard_Flocks	Birds_Affected	Commercial_Flocks	Control_Area_Released	Control_Zone_Release	Flock_Herd_Inventory	Last_Reported_Detection
count	1693	1693.000000	1.693000e+03	1693.000000	1243	1249	1.690000e+03	1693
mean	2023-08-21 17:02:22.468989952	0.538098	1.000149e+05	0.461902	2023-06-25 06:26:56.090104320	2023-06-24 20:54:22.770216192	1.001924e+05	2023-08-21 17:02:22.468989952
min	2022-02-08 00:00:00	0.000000	0.000000e+00	0.000000	2022-02-27 00:00:00	2022-02-27 00:00:00	0.000000e+00	2022-02-08 00:00:00
25%	2022-09-01 00:00:00	0.000000	4.000000e+01	0.000000	2022-06-07 00:00:00	2022-06-07 00:00:00	4.000000e+01	2022-09-01 00:00:00
50%	2023-10-12 00:00:00	1.000000	1.100000e+03	0.000000	2022-12-14 00:00:00	2022-12-14 00:00:00	1.300000e+03	2023-10-12 00:00:00
75%	2024-12-05 00:00:00	1.000000	3.790000e+04	1.000000	2024-07-29 00:00:00	2024-07-26 00:00:00	3.805000e+04	2024-12-05 00:00:00
max	2025-04-30 00:00:00	1.000000	5.347500e+06	1.000000	2025-04-19 00:00:00	2025-04-19 00:00:00	5.347500e+06	2025-04-30 00:00:00
std	NaN	0.498694	3.894141e+05	0.498694	NaN	NaN	3.897369e+05	NaN

(venv) c:\Users\Lina\OneDrive\Documents\AIT 580>

Figure 1: Descriptive statistics of the dataset for initial review.

Strategy & Methods & Tools:

To effectively represent the data, I hatched a data cleaning plan by using Amazon Web Service's (AWS) S3 bucket and Glue DataBrew services to remove columns, remove special characters, and redefine some string columns to be a date data type instead. Once I finished, I created a job to push the cleaned data as a parquet file back to my S3. From my S3 bucket, I downloaded the parquet file and read it into my Python console to start working on my research questions.

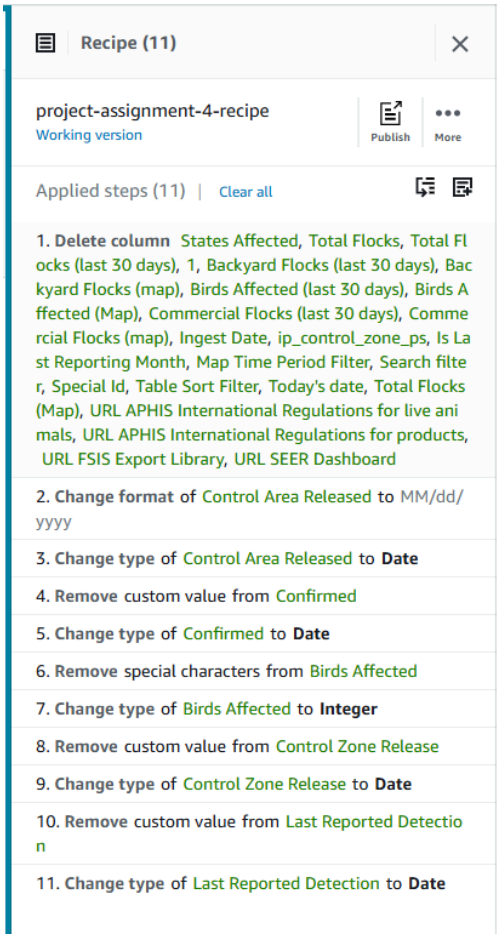


Figure 2: AWS Glue DataBrew recipe for cleaning the dataset.

In Python, I imported my packages, created my data frame, and had to change some of the data types since the parquet file did not carry over the data type conversion I did in Glue. From there, I created a new column, Month_Year, to store date values as month_name Year for visualization and grouping purposes.

After answering my first research question, I chose to answer the question, "How long does an outbreak typically last?" For this question, I chose to use SQL. I created a new database in SQL Server Management Studio (SSMS) and imported my table as a flat file from my S3 bucket. After initializing my schema, I ran my query.

DESKTOP-IR8ADBD\...u - dbo.main_data | SQLQuery1.sql - DE...IR8ADBD\Lina (53))*

Column Name	Data Type	Allow Nulls
Confirmed	date	☐
Incident_Site_Name	nvarchar(50)	☐
Backyard_Flocks	bit	☐
Birds_Affected	int	☐
Commercial_Flocks	bit	☐
Control_Area_Released	date	☑
Control_Zone_Release	date	☑
County_Name	nvarchar(50)	☐
County_State	nvarchar(50)	☐
Flock_Herd_Inventory	int	☑
Last_Reported_Detection	date	☐
Production	nvarchar(50)	☐
Production_Type_Name	nvarchar(50)	☐
Species_Name	nvarchar(50)	☐
State	nvarchar(50)	☐

Figure 3: Database schema for SQL query in SSMS.

To complete the next research question, I utilized R to create a line plot for my egg price data from the BLS.

For the last question, I downloaded shape files from Census.gov to get the outline of my United States map and merged my flock data on the Site Incident Name column, which was the state, to match to the State column from the shape files.

Results:

From my analysis, I came to these conclusions:

i. **Have we seen an uptick in avian flu cases recently within the United States?**
Avian influenza is ebbing and flowing but does not show exponential growth to justify surging egg prices. The United States is not experiencing an uptick in HPAI H5N1 outbreaks. Figure 4 provides a visualization of confirmed avian flu cases in the United States from 2022 through 2025.

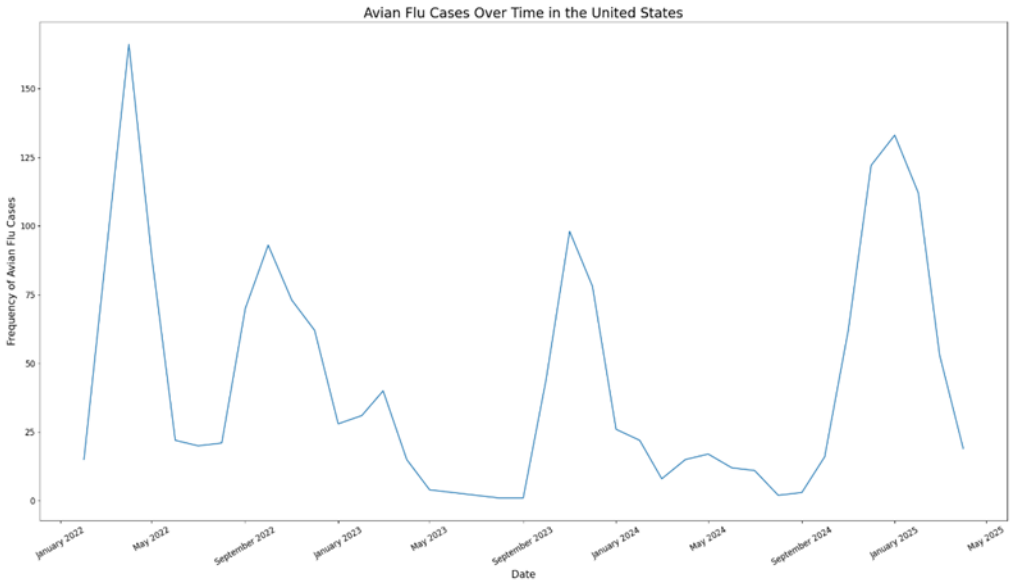


Figure 4: HPAI H5N1 outbreaks in the United States over time.

ii. **How long does an outbreak typically last?**

Outbreaks typically last for an average of 28 days in each control zone. To get this calculation, I took the date from the Confirmed column and used the DATEDIFF function to calculate how many days each outbreak lasted until the control zone was confirmed clear of HPAI H5N1. Null values mean that the outbreak is still ongoing and HPAI H5N1 is still active within the control zone being recorded. Figure 5 shows a snippet of the SQL query's output individually. To get the average amount, I nested the DATEDIFF function in the AVG function and removed the other fields.

100 %

Results | Messages

	Incident_Site_Name	(No column name)	Confirmed	Control_Zone_Release
1	Wyoming	NULL	2024-12-06	NULL
2	Oregon	NULL	2024-03-14	NULL
3	Wyoming	22	2022-10-12	2022-11-03
4	Wyoming	NULL	2025-02-28	NULL
5	Indiana	27	2025-02-14	2025-03-13
6	Wyoming	NULL	2025-02-12	NULL
7	Ohio	NULL	2025-04-14	NULL
8	Wyoming	NULL	2024-12-27	NULL
9	Indiana	NULL	2025-01-28	NULL
10	North Carolina	28	2022-04-07	2022-05-05
11	Washington	22	2022-06-09	2022-07-01
12	California	20	2023-12-05	2023-12-25
13	North Carolina	NULL	2022-11-18	NULL
14	Minnesota	22	2022-09-14	2022-10-06
15	South Dakota	16	2024-12-10	2024-12-26
16	Oregon	NULL	2022-12-13	NULL
17	Idaho	NULL	2025-01-03	NULL
18	California	46	2024-12-31	2025-02-15
19	Maryland	22	2025-02-04	2025-02-26
20	Washington	31	2022-05-10	2022-06-10
21	Texas	NULL	2024-03-20	NULL
22	Nebraska	14	2024-12-17	2024-12-31
23	Nebraska	17	2022-09-19	2022-10-06
24	Iowa	12	2022-04-26	2022-05-08
25	Minnesota	25	2023-12-11	2024-01-05
26	South Dakota	21	2022-09-20	2022-10-11
27	North Carolina	20	2024-02-20	2024-03-11
28	California	46	2024-12-09	2025-01-24

Figure 5: DATEDIFF() output for individual incident sites. Null values mean that HPAI H5N1 is still ongoing.

SQLQuery1.sql - DE...IR8ADBD\Lina (53))*

```
SELECT AVG(DATEDIFF(DAY, Confirmed, Control_Zone_Release))
FROM main_data;
```

Figure 6: SQL query for average calculations.

iii. How might the avian flu affect egg prices?

From the analysis and visualization run on the BLS egg data prices, we can infer that the avian flu impacts egg price data through other factors such as feed, natural gas costs, and bird loss, but we cannot directly conclude that the avian flu is the only factor driving increased egg prices. Strictly speaking about HPAI H5N1, the trends from the egg price data ebb and flow similarly to the avian flu outbreak data. Bird loss and the costs associated with culling the entire flock, having to follow USDA protocol to clear the virus, and having to start over from scratch is a more reasonable conclusion to make from the data. The avian flu affects egg prices, but it also feeds into other factors that trickle down to create a higher egg price for Americans.

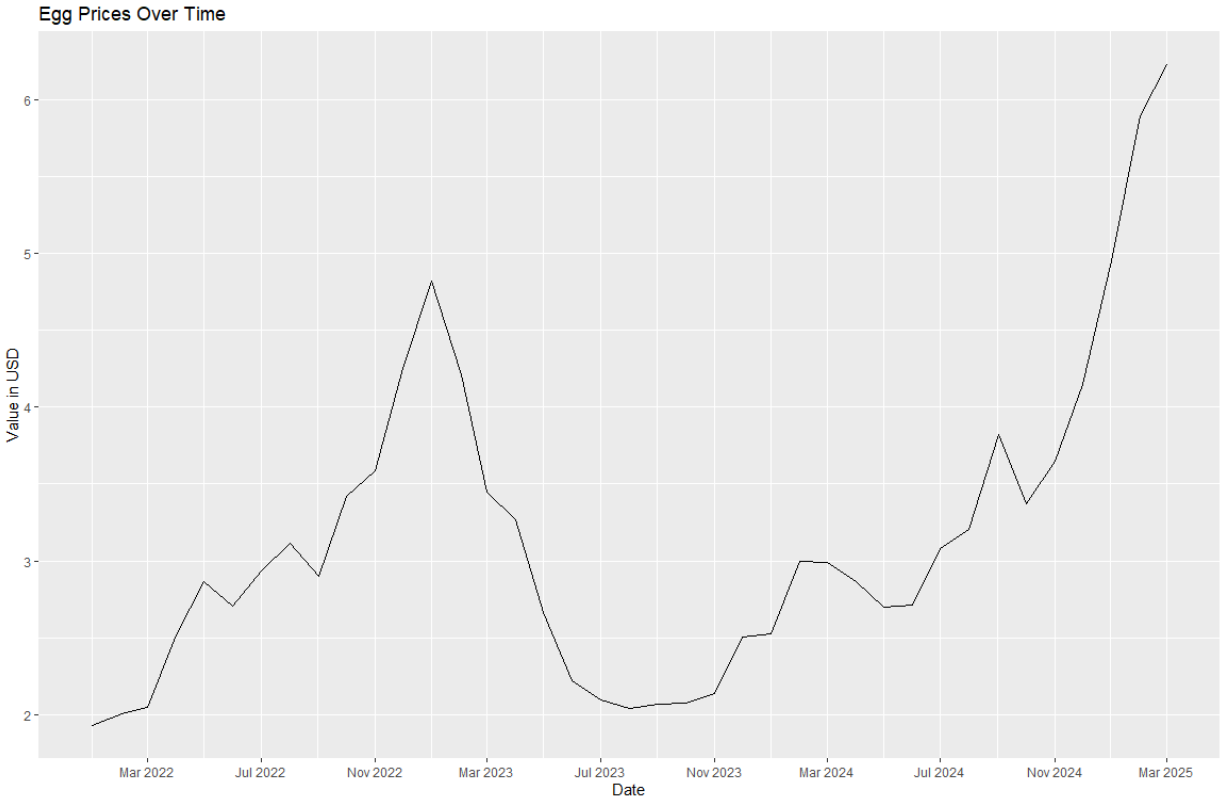


Figure 7: Egg prices in USD over time.

iv. What is the movement of the avian flu within the United States?

From some geospatial analysis, the avian flu is most prevalent in Minnesota and South Dakota. Most states that are experiencing an outbreak are coupled with other states in proximity. This could potentially mean that the avian flu is traveling through migratory birds and other species.

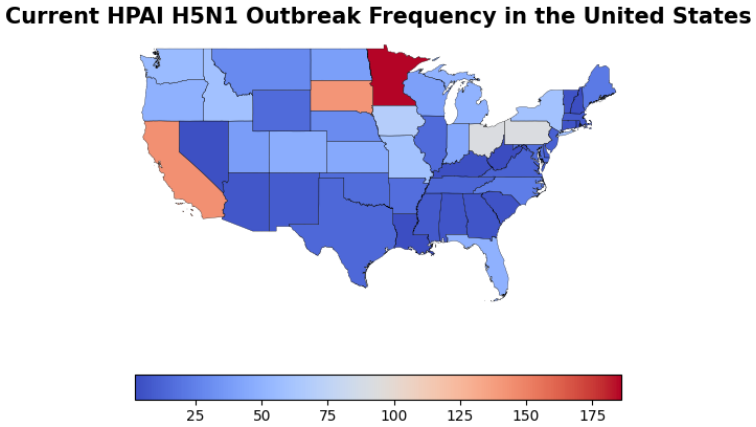


Figure 8: Current avian flu outbreak frequency by state.

Conclusion:

After creating the visualizations and running separate analyses on each factor, HPAI H5N1 is a contributing factor to surging egg prices in the United States. While it may not be the only factor, the infection of one individual bird is enough to cull the entire flock which can be devastating for commercial farms. Farmers must start from scratch with increasing feed costs and natural gas prices that make this process even more difficult. From my analysis, it would be more appropriate to say that avian influenza has a trickle-down effect which leads to bird mortality, paired with economic factors with supply and demand.

References

[1] United States Department of Agriculture, “HPAI Confirmations in Commercial and Backyard Flocks,” *www.aphis.usda.gov*, Jun. 20, 2024. https://www.aphis.usda.gov/livestock-poultry-disease/avian/avian-influenza/hpai-detections/commercial-backyard-flocks

[2] Bureau of Labor Statistics, “BLS Data Viewer,” *Bls.gov*, 2018. https://data.bls.gov/dataViewer/view/timeseries/APU0000708111

[3] M. Andrew, M. Charles, and L. Abdelaziz, “Why Are Eggs so Expensive? Understanding the Recent Spike in Egg Prices,” Choices, vol. 38, no. 2, 2023, Accessed: Mar. 24, 2025. [Online]. Available: https://ageconsearch.umn.edu/record/338531?ln=en&v=pdf

[4] A. Mena, M. E. von Fricken, and B. D. Anderson, “The Impact of Highly Pathogenic Avian Influenza H5N1 in the United States: A Scoping Review of Past Detections and Present Outbreaks,” Viruses, vol. 17, no. 3, p. 307, Feb. 2025, doi: https://doi.org/10.3390/v17030307.

[5] USDA, “Public Version,” Aug. 2016. Accessed: Mar. 24, 2024. [Online]. Available: https://www.aphis.usda.gov/media/document/2086/file