U.S. Meat Production Analysis and Forecasting

Caleb Hart

STAT442 Unsupervised Machine Learning

1. Abstract

This report examines U.S. meat production statistics and attempts to forecast meat production for the year. The sanitized dataset includes variables such as beef, veal, pork, lamb and mutton, broilers, turkey, total poultry production, and total meat production. The analysis leverages both traditional statistical techniques and machine learning models, specifically a Variational Autoencoder (VAE) and a transformer-based forecasting model, to analyze trends and forecast future production.

[1] Some understanding and techniques have been gained from Hunter Heidenreich's tutorial on VAE's which is available on their personal blog though none of their original code has been used for this project.

Data Source

The original data comes from the USDA, available through their website: https://www.ers.usda.gov/data-products/livestock-and-meat-domestic-data

Although there were many possible datasets to choose from, even from within the USDA, I chose to analyze the 'meat statistics table, historical'. Contained within the excel file is a table cataloging all meat production in the US by month dating all the way back to 1921. I chose to crop the data to the period of December 2024 – January 2001 as this was the period that lacked any missing values. I additional removed the columns for federally inspected meat as I simply wanted to look at total production.

Red meat and poultry production (million pounds)

Type 1/	Commercial 2/					Federally inspected									
	Beef 3/	Veal 3/	Pork 3/	Lamb and mutton 3/	Total red meat 3/ 4/	Beef 3/	Veal 3/	Pork 3/	Lamb and mutton 3/	Total red meat 3/ 4/	Broilers 5/	Other chicken 5/	Turkey 5/	Total poultry 4/ 5/ 6/	Total red meat and poultry 4/
Jan 2025	2,370.1	2.5	2,501.9	10.9	4,885.3	2,335.6	2.3	2,492.6	9.7	4,840.2	4,137.7	48.0	415.2	4,612.7	9,452.9
Jan 2024	2,280.8	3.9	2,472.7	10.6	4,768.0	2,246.1	3.8	2,462.4	9.6	4,721.9	4,051.4	47.4	435.2	4,547.6	9,269.5
Jan-2025	2,370.1	2.5	2,501.9	10.9	4,885.3	2,335.6	2.3	2,492.6	9.7	4,840.2	4,137.7	48.0	415.2	4,612.7	9,452.9
Dec-2024	2,200.7	3.1	2,328.7	12.0	4,544.4	2,168.6	2.9	2,319.9	10.8	4,502.1	3,882.4	42.0	376.3	4,311.2	8,813.3
Nov-2024	2,216.6	3.1	2,334.4	10.4	4,564.5	2,185.0	3.0	2,325.7	9.2	4,522.9	3,651.4	36.9	394.1	4,093.1	8,616.0
Oct-2024	2,465.5	3.3	2,543.5	11.7	5,024.0	2,425.8	3.1	2,532.2	10.3	4,971.4	4,370.9	52.6	493.9	4,929.6	9,901.0
Sep-2024	2,204.9	3.1	2,232.6	10.6	4,451.2	2,170.7	2.9	2,221.8	9.4	4,404.8	3,892.6	49.2	413.4	4,366.2	8,771.0
Aug-2024	2,288.1	3.1	2,289.9	10.5	4,591.7	2,256.1	3.0	2,278.5	9.2	4,546.8	4,038.7	49.6	435.0	4,533.2	9,080.0
Jul-2024	2,286.8	3.3	2,252.8	11.3	4,554.3	2,255.9	3.2	2,242.5	10.0	4,511.7	4,072.1	50.5	433.5	4,566.2	9,077.9
Jun-2024	2,135.8	3.1	2,117.7	10.4	4,267.0	2,103.6	3.0	2,109.2	9.1	4,225.0	3,724.9	47.7	408.9	4,190.7	8,415.7
May-2024	2,326.4	3.5	2,278.6	11.5	4,619.9	2,292.3	3.4	2,269.5	10.2	4,575.4	4,011.0	50.3	446.8	4,518.9	9,094.3
Apr-2024	2,303.4	3.5	2,317.4	11.4	4,635.6	2,269.3	3.3	2,308.2	10.1	4,590.9	3,917.7	47.2	450.7	4,427.5	9,018.4
Mar-2024	2,110.4	3.6	2,250.1	12.3	4,376.4	2,076.5	3.4	2,240.7	11.1	4,331.7	3,636.8	42.9	409.5	4,100.3	8,432.0
Feb-2024	2,168.6	3.6	2,371.2	10.9	4,554.3	2,135.0	3.4	2,361.4	9.8	4,509.7	3,741.9	45.0	423.9	4,221.4	8,731.1
Jan-2024	2,280.8	3.9	2,472.7	10.6	4,768.0	2,246.1	3.8	2,462.4	9.6	4,721.9	4,051.4	47.4	435.2	4,547.6	9,269.5

Variables: Beef, Veal, Pork, Lamb and Mutton, Total Red Meat, Broilers, Other Chicken, Turkey, Total Poultry, and Total Red Meat and Poultry.

Variable Types: Continuous (e.g., production volume), Nominal (e.g., date).

Cleaned Data:

```
[7]: print(df.sample(n=10))
```

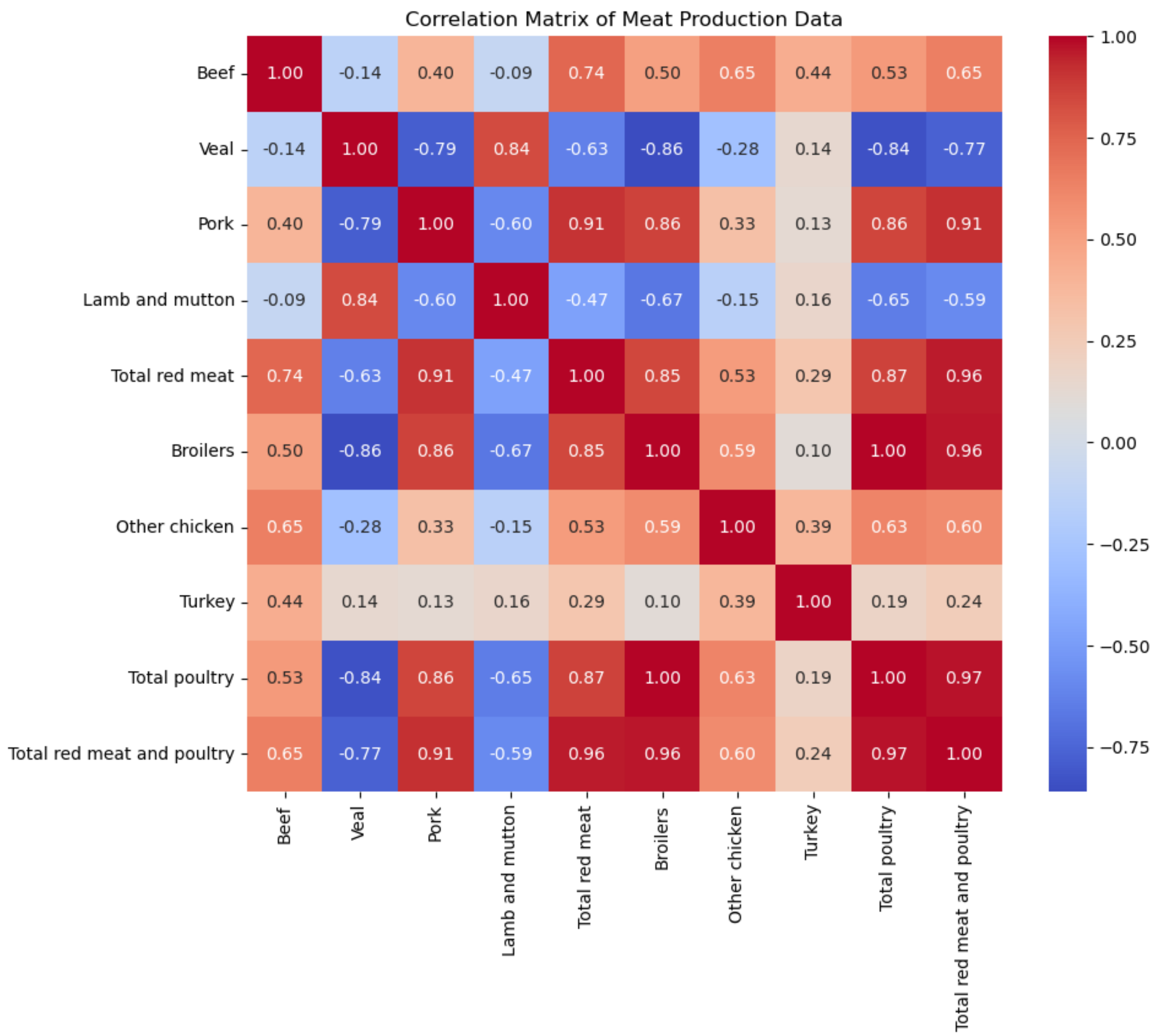
```
        Date    Beef  Veal    Pork  Lamb and mutton  Total red meat  Broilers  \
4     24-Sep  2204.9   3.1  2232.6             10.6          4451.2    3892.6
268    2-Sep  2201.0  16.3  1638.0             17.6          3873.2    2596.4
217    6-Dec  2049.6  13.1  1796.4             15.4          3874.5    2737.8
105   16-Apr  1964.0   6.0  2000.2             12.9          3983.0    3302.1
0     25-Jan  2370.1   2.5  2501.9             10.9          4885.3    4137.7
122   14-Nov  1849.5   6.5  1890.2             11.5          3757.7    2939.9
277    1-Dec  2110.0  16.0  1668.0             19.0          3813.6    2464.8
139   13-Jun  2157.0   8.4  1676.3             12.7          3854.5    3002.0
210    7-Jul  2256.7  10.7  1659.5             13.5          3940.4    3055.0
285    1-Apr  1939.0  15.0  1533.0             20.0          3507.5    2515.7

     Other chicken  Turkey  Total poultry  Total red meat and poultry
4             49.2   413.4         4366.2                      8771.0
268           46.4   444.6         3098.3                      6918.3
217           36.9   429.1         3215.5                      7043.0
105           43.8   484.8         3842.1                      7787.9
0             48.0   415.2         4612.7                      9452.9
122           36.1   480.8         3467.3                      7190.1
277           40.9   419.8         2935.4                      6694.4
139           42.9   473.4         3527.6                      7345.1
210           44.6   505.2         3616.0                      7512.8
285           42.2   428.8         2997.9                      6455.9
```

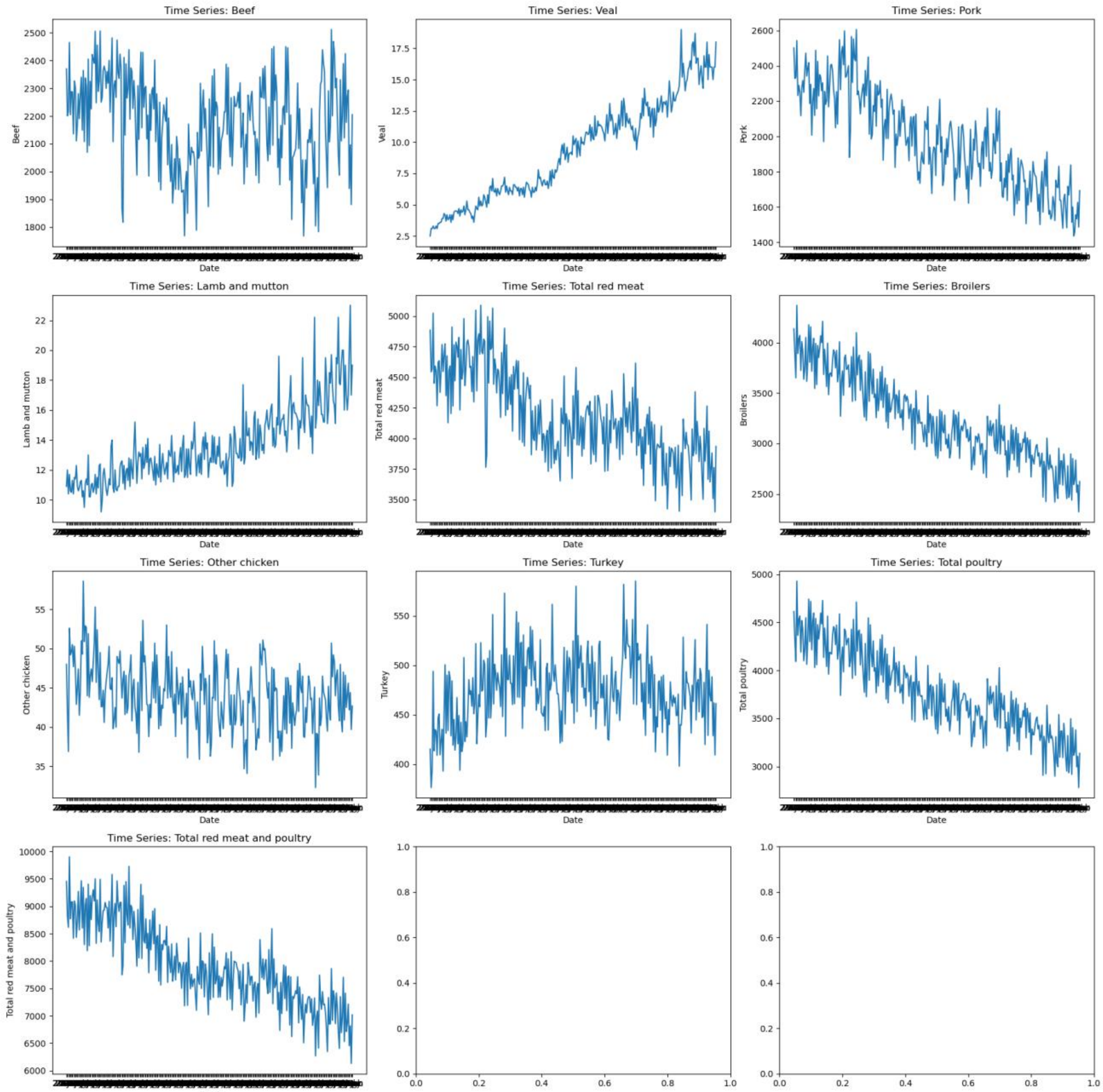
Statistical Summary: Mean, median, mode, and skewness are assessed through boxplots and 2-dimension scatterplots. A correlation matrix was also generated to assess inter-variable relationships.

2. Analytical Objectives

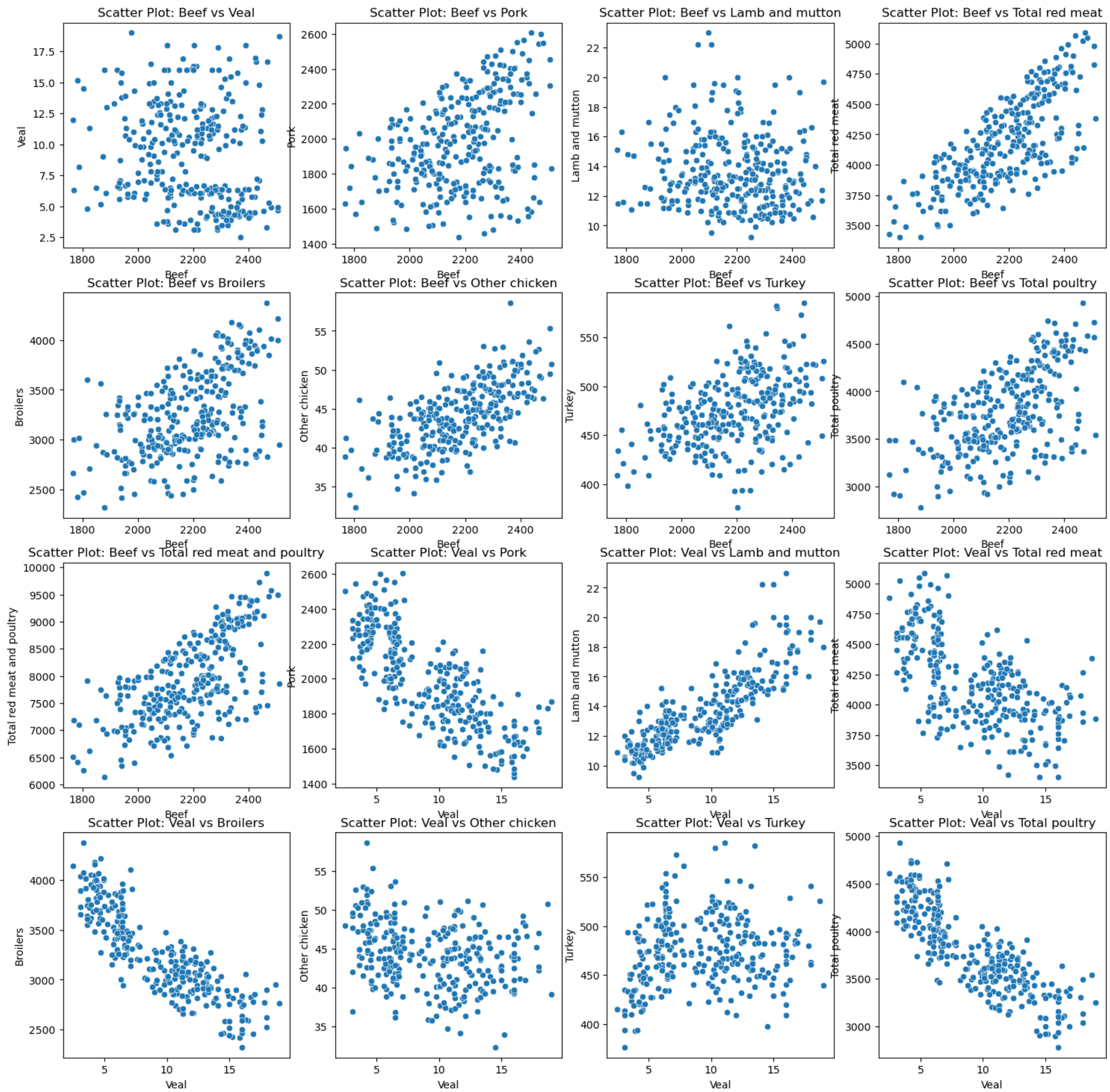
Correlation:



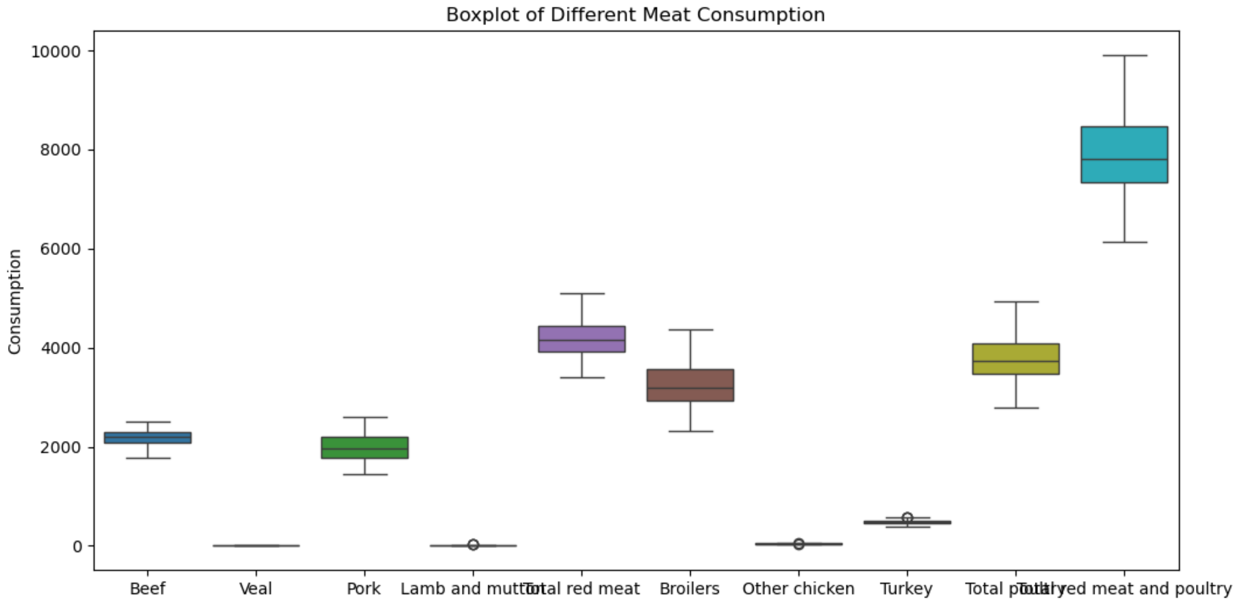
Time-series trends:



2-Dimension analysis:



Scatter Plot: Beef vs Veal
Scatter Plot: Beef vs Pork
Scatter Plot: Beef vs Lamb and mutton
Scatter Plot: Beef vs Total red meat
Scatter Plot: Beef vs Broilers
Scatter Plot: Beef vs Other chicken
Scatter Plot: Beef vs Turkey
Scatter Plot: Beef vs Total poultry
Scatter Plot: Beef vs Total red meat and poultry
Scatter Plot: Veal vs Pork
Scatter Plot: Veal vs Lamb and mutton
Scatter Plot: Veal vs Total red meat
Scatter Plot: Veal vs Broilers
Scatter Plot: Veal vs Other chicken
Scatter Plot: Veal vs Turkey
Scatter Plot: Veal vs Total poultry
Beef
Veal
Pork
Lamb and mutton
Total red meat
Broilers
Other chicken
Turkey
Total poultry
Total red meat and poultry



3. Analytical Plan

Data Cleaning

The original data sourced from the USDA has data going back to 1967, but I trimmed this forward to 2001 in order to reduce the amount of missing data.

Removal of commas and conversion to numeric values.

MinMaxScaler is used to normalize the data while preventing negative numbers.

Statistical and Machine Learning Procedures

Descriptive Analysis: Correlation heatmap, time series visualization, 2-dimension analysis, and boxplots.

Predictive Analysis: Variational Autoencoder (VAE) for time series forecasting. This was done in Python using Keras and Pytorch's neural network libraries.

Software and Tools

Python (Pandas, NumPy, Seaborn, Matplotlib, Keras, Pytorch)

PyTorch under Keras' framework was used for VAE implementation

Rationale for VAE

The VAE allows capturing latent patterns in the data and generates realistic future production samples all while using the original data source as its training data.

Model Architecture:

```
VAE(
 (encoder): Encoder(
  (lstm): LSTM(10, 128, batch_first=True)
  (fc_mean): Linear(in_features=128, out_features=20, bias=True)
  (fc_log_var): Linear(in_features=128, out_features=20, bias=True)
 )
 (decoder): Decoder(
  (fc1): Linear(in_features=20, out_features=128, bias=True)
  (lstm): LSTM(128, 128, batch_first=True)
  (fc2): Linear(in_features=128, out_features=10, bias=True)
 )
)
```

Model Training:

```
Epoch [1/20], Train Loss: 12.4279, Test Loss: 6.7641
Epoch [2/20], Train Loss: 8.9746, Test Loss: 4.2488
Epoch [3/20], Train Loss: 6.0170, Test Loss: 2.8316
Epoch [4/20], Train Loss: 4.4828, Test Loss: 3.5307
Epoch [5/20], Train Loss: 4.5161, Test Loss: 4.4063
Epoch [6/20], Train Loss: 4.3073, Test Loss: 4.0044
Epoch [7/20], Train Loss: 3.9038, Test Loss: 3.6971
Epoch [8/20], Train Loss: 3.6064, Test Loss: 3.4347
Epoch [9/20], Train Loss: 3.5371, Test Loss: 3.2422
Epoch [10/20], Train Loss: 3.5407, Test Loss: 2.9917
Epoch [11/20], Train Loss: 3.4958, Test Loss: 2.9350
Epoch [12/20], Train Loss: 3.3564, Test Loss: 2.6883
Epoch [13/20], Train Loss: 3.1736, Test Loss: 2.7235
Epoch [14/20], Train Loss: 3.1244, Test Loss: 2.5810
Epoch [15/20], Train Loss: 3.0892, Test Loss: 2.6776
Epoch [16/20], Train Loss: 3.0346, Test Loss: 2.5432
Epoch [17/20], Train Loss: 3.0131, Test Loss: 2.5721
Epoch [18/20], Train Loss: 2.9728, Test Loss: 2.5750
Epoch [19/20], Train Loss: 2.8821, Test Loss: 2.6214
Epoch [20/20], Train Loss: 2.9507, Test Loss: 2.5096
```

Additional Models

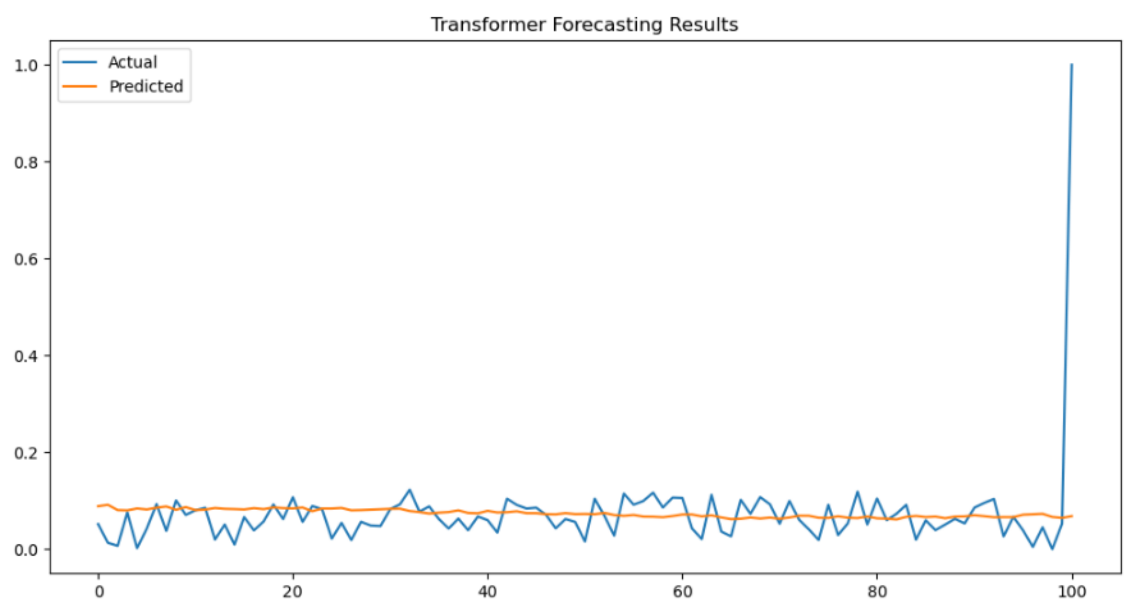
A transformer-based forecasting model was later developed after the lack-luster results of the VAE. This included a slightly different method of cleaning the data with the hopes that this would also improve the model. The main difference in data cleaning was the inclusion of data going back to 1977.

```
class TransformerForecastingModel(torch.nn.Module):
    def __init__(self, input_dim, embed_dim, num_heads, ff_dim, num_layers, dropout=0.2):
        super().__init__()
        self.embedding = torch.nn.Linear(input_dim, embed_dim)
        self.relu1 = torch.nn.ReLU()
        self.dropout1 = torch.nn.Dropout(dropout)

        encoder_layer = torch.nn.TransformerEncoderLayer(
            d_model=embed_dim, nhead=num_heads, dim_feedforward=ff_dim, dropout=dropout
        )
        self.transformer_encoder = torch.nn.TransformerEncoder(encoder_layer, num_layers=num_layers)

        self.relu2 = torch.nn.ReLU()
        self.dropout2 = torch.nn.Dropout(dropout)
        self.fc_out = torch.nn.Linear(embed_dim, input_dim)
        self.relu3 = torch.nn.ReLU()

    def forward(self, x):
        x = self.embedding(x)
        x = self.relu1(x)
        x = self.dropout1(x)
        x = self.transformer_encoder(x)
        x = self.relu2(x)
        x = self.dropout2(x)
        x = self.fc_out(x[:, -1, :])  # Take the last time step output
        x = self.relu3(x)
        return x
```



4. Results

Descriptive Analysis

The correlation matrix revealed strong correlations between all categories. This isn't surprising as meat production will follow a global trend.

Time series plots indicated seasonal trends in production like annual peaks and falls. Other trends are shown like a decrease in total poultry production year over year.

Forecasting Results

The VAE successfully generated 12-month forecasts showing that it was able to capture some of the historical trends in the data, though there is still much improvement to be made in the model.

```
Predicted future trends (normalized):
[[0.406  0.2314 0.3859 0.1909 0.4021 0.3579 0.3504 0.3878 0.3994 0.388 ]
 [0.5515 0.2984 0.4811 0.2184 0.4862 0.442  0.4384 0.4298 0.498  0.4898]
 [0.5996 0.3158 0.5096 0.2209 0.5072 0.465  0.4677 0.425  0.5273 0.5168]
 [0.6167 0.3202 0.5183 0.2205 0.513  0.4719 0.4792 0.4174 0.5365 0.5233]
 [0.6233 0.3213 0.5212 0.2204 0.5148 0.4743 0.4845 0.4124 0.5397 0.5244]
 [0.6263 0.3216 0.5223 0.2206 0.5154 0.4755 0.4873 0.4095 0.5409 0.5242]]

Predicted future trends (original scale):
     Date    Feature_1  Feature_2    Feature_3  Feature_4    Feature_5  \
0  Jan-25  2069.683594   6.779370  1886.989136  11.834916  4078.976562
1  Feb-25  2178.073975   7.843806  1998.528687  12.213923  4221.109863
2  Mar-25  2213.849854   8.120829  2031.905029  12.247731  4256.477539
3  Apr-25  2226.565186   8.191932  2042.193604  12.242837  4266.317871
4  May-25  2231.521729   8.209249  2045.560181  12.241272  4269.277344
5  Jun-25  2233.714844   8.213400  2046.828735  12.244476  4270.285645

     Feature_6  Feature_7   Feature_8    Feature_9   Feature_10
0  3055.531250  41.515602  457.274445  3639.246826  7594.373535
1  3227.756592  43.829285  466.046814  3851.158691  7978.141602
2  3274.892822  44.600796  465.035797  3914.104736  8079.801270
3  3289.002441  44.902718  463.459747  3933.799561  8104.331543
4  3293.989746  45.041573  462.399292  3940.562988  8108.364258
5  3296.393799  45.115040  461.794006  3943.285156  8107.657715
```

Actual:

Type 1/	Commercial 2/					Federally inspected									
	Beef 3/	Veal 3/	Pork 3/	Lamb and mutton 3/	Total red meat 3/ 4/	Beef 3/	Veal 3/	Pork 3/	Lamb and mutton 3/	Total red meat 3/ 4/	Broilers 5/	Other chicken 5/	Turkey 5/	Total poultry 4/ 5/ 6/	Total red meat and poultry 4/
Jan-2025	2,370.1	2.5	2,501.9	10.9	4,885.3	2,335.6	2.3	2,492.6	9.7	4,840.2	4,137.7	48.0	415.2	4,612.7	9,452.9

Transformer-based model results

Unscaled Future Predictions:

	0	1	2	3	4	5	6	7	\
0	1741.0094	32.8152	1309.6686	31.1271	2822.1050	1405.9308	380.1586	1956.3875	
1	1886.8129	26.1994	1275.5325	24.3964	3365.1445	1713.3722	391.3523	2178.2158	
2	1960.1750	21.9284	1373.6909	21.6222	3632.3477	1988.3779	406.1049	2433.3838	
3	2003.7330	18.0578	1505.2623	18.4523	3837.4453	2326.9846	429.4685	2763.1968	
4	2068.4082	14.1441	1656.1318	16.3812	3932.5452	2686.8577	449.2589	3133.7710	
5	2093.5562	10.8822	1793.6283	14.9633	3918.9602	2940.5046	460.7080	3417.1995	
6	2113.3562	8.9785	1896.0061	13.3408	3949.0161	3109.6438	468.7467	3625.2012	
7	2117.9890	8.1316	1961.2762	12.0100	4004.7803	3216.7402	474.0439	3765.1921	
8	2123.5330	7.9407	1994.4252	11.6604	4055.2300	3270.8459	475.9704	3838.4641	
9	2125.2400	7.8247	2006.6818	11.5188	4073.0515	3291.5183	475.9724	3863.3645	
10	2125.9209	7.7744	2012.1145	11.4549	4081.2314	3300.5908	475.9526	3874.3167	
11	2126.2141	7.7519	2014.5176	11.4265	4084.8560	3304.6008	475.9412	3879.1487	

	8
0	5425.0786
1	5530.1543
2	5942.7886
3	6459.6069
4	7017.6904
5	7406.4473
6	7693.3213
7	7864.8477
8	7981.8794
9	8027.1855
10	8047.4150
11	8056.3550

First 12 rows of the original dataset:

	Beef 3/	Veal 3/	Pork 3/	Lamb and mutton 3/	Total red meat 3/ 4/	Broilers 5/	Turkey 5/	Total poultry 4/ 5/ 6/	Total red meat and poultry 4/
0	2335.6000	2.3000	2492.6000	9.7000	4840.2000	4137.7000	415.2000	4612.7000	9452.9000
1	2246.1000	3.8000	2462.4000	9.6000	4721.9000	4051.4000	435.2000	4547.6000	9269.5000
2	2335.6000	2.3000	2492.6000	9.7000	4840.2000	4137.7000	415.2000	4612.7000	9452.9000
3	2168.6000	2.9000	2319.9000	10.8000	4502.1000	3882.4000	376.3000	4311.2000	8813.3000
4	2185.0000	3.0000	2325.7000	9.2000	4522.9000	3651.4000	394.1000	4093.1000	8616.0000
5	2425.8000	3.1000	2532.2000	10.3000	4971.4000	4370.9000	493.9000	4929.6000	9901.0000
6	2170.7000	2.9000	2221.8000	9.4000	4404.8000	3892.6000	413.4000	4366.2000	8771.0000
7	2256.1000	3.0000	2278.5000	9.2000	4546.8000	4038.7000	435.0000	4533.2000	9080.0000
8	2255.9000	3.2000	2242.5000	10.0000	4511.7000	4072.1000	433.5000	4566.2000	9077.9000
9	2103.6000	3.0000	2109.2000	9.1000	4225.0000	3724.9000	408.9000	4190.7000	8415.7000
10	2292.3000	3.4000	2269.5000	10.2000	4575.4000	4011.0000	446.8000	4518.9000	9094.3000
11	2269.3000	3.3000	2308.2000	10.1000	4590.9000	3917.7000	450.7000	4427.5000	9018.4000

5. Discussion

Strengths and Limitations

Strengths: Capable of generating predictions that capture some of the information. Output is consistent with latent variables as they go through the VAE.

Limitations: The current model is under-preforming and lacks sensitivity to monthly trends.

Personal Learning

Gained insights into the practical application of VAEs for time series forecasting. Learned that VAEs are very finicky and take a lot of trial and error to get accurate results.

Need to further explore hyperparameter tuning for improved accuracy.

6. Conclusion

This project successfully applied business analytics techniques to understand and apply time-series forecasting techniques. The use of a VAE provided an innovative approach to time series forecasting, making the results suitable for strategic decision-making when implemented properly. The transformer-based model also provided sufficient forecasting. Further testing and research are needed to more fully understand both model's lack of improvement.

7. References

[1] Heidenreich, H. (2024, March 3). Modern Pytorch techniques for Vaes: A comprehensive tutorial. Hunter Heidenreich, Machine Learning Engineer. https://hunterheidenreich.com/posts/modern-variational-autoencoder-in-pytorch/

Su, L., Zuo, X., Li, R. et al. A systematic review for transformer-based long-term series forecasting. Artif Intell Rev 58, 80 (2025).

https://doi.org/10.1007/s10462-024-11044-2