Dairy Herd Assessment Report

Beidel Brother’s Dairy Farm | January 8th, 2018

Brandon Beidel
Email: bcbeidel@gmail.com
Linkedin: linkedin.com/in/brandonbeidel/
Github: github.com/bcbeidel

Report Objective

This report is intended to provide insight and guidance aimed at improving herd health and milk production volumes of the Beidel Brother's Dairy farm, located in Franklin County, Pennsylvania. The current scope of analysis includes data regarding the facilities and weather, herd composition, and nutrition of the herd. This report focuses on milk production data for calendar years 2016 and 2017.

Recommendations

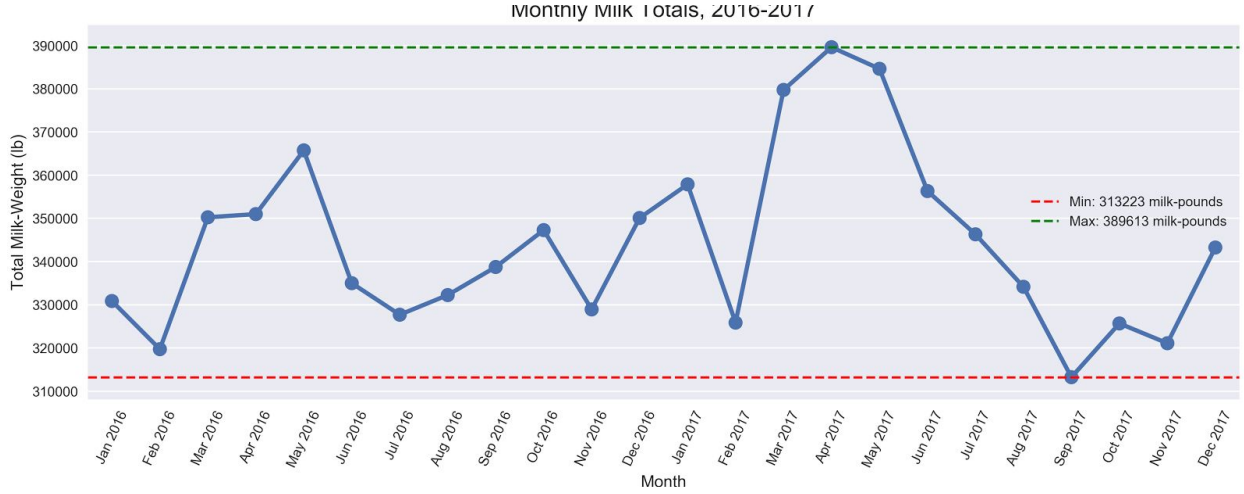
Based upon the findings below, the following high-level actions are recommended to improve future operations:

- Mitigate the risk of further data loss for milk production information through the use of an online data storage provider such as DropBox or Google Drive
- Expedite future analyses by receiving Pedigree and Classification Reports from Holstein USA in a digital format (CSV, PDF, etc.)
- Prioritize the replacement of cattle identified as low-performing from the herd

Overview of Herd, 2016-2017

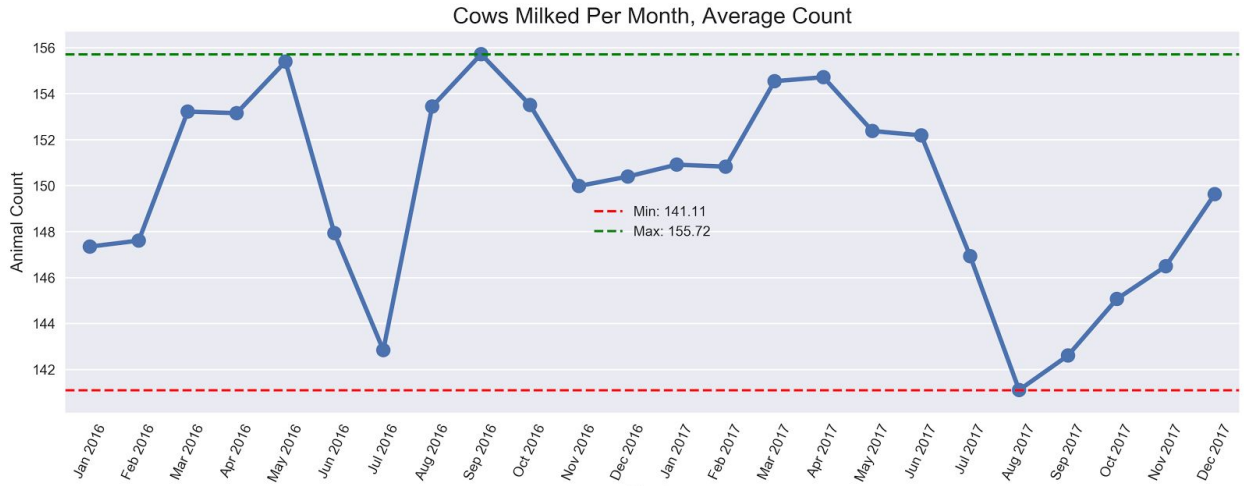
Total Milk-Weight Production

The herd produced an estimated 8,111,872 milk-pounds (943,241 gallons milk) between January 1, 2016, and December 31, 217. The on-premise, DeLaval - ALPRO™ herd management system provided the data for this calculation. Notable in this data was the corruption or loss of data for at least 20 days of records in the two year period. In September 2107, 12 of sets of daily records were missing. If the measurements for a single day did not surpass the 10,000 milk-pound level, the mean daily milk-weight value of 11,032 milk-pounds per day was assigned. The chart below provides a monthly summary milk-weight production for the two year period.



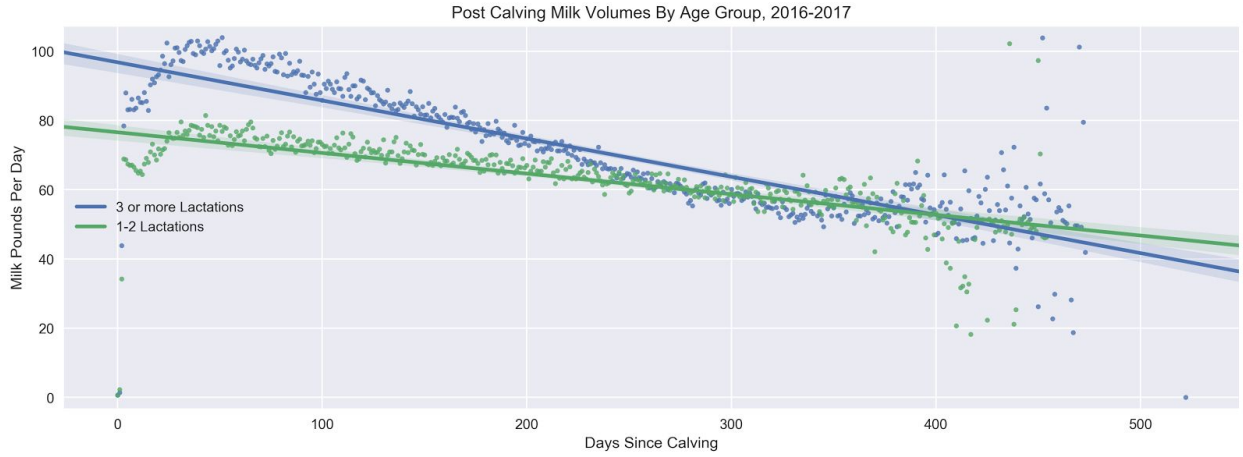
Utilization of Milking Facilities

For calendar years 2016 and 2017, the total number of unique animal milked per month ranged from 141 to 155 cows, or 83.1% to 92.2% of the maximum capacity of the current 168 stall free-stall facilities. The plot below provides a visualization of the total number of cows milked per month.



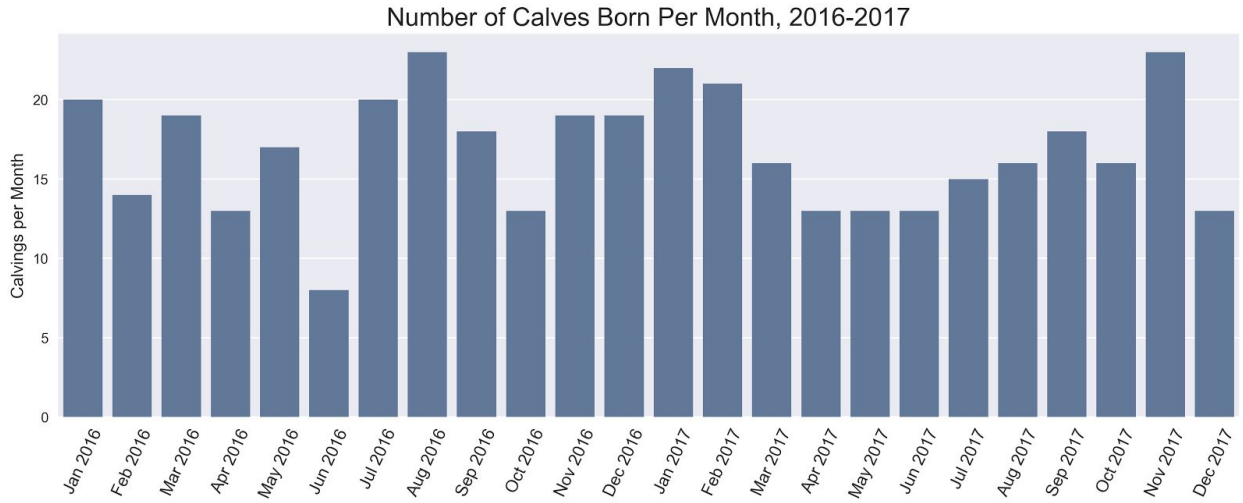
Aged Cow Performance

Older cattle, or those having gone through more than two lactations, outperformed younger cattle by 17.5% for the first 305 days after calving, with an average of 24,435 milk-pounds (2,841 gallons) compared to 20,789 milk-pounds (2,417 gallons). The performance gap is most pronounced in the 20-100 day post-calving range but gradually decreased to zero until approximately 400 days after calving. The variance for milk-weights beyond 400 days is believed to be the result of incomplete calving data.



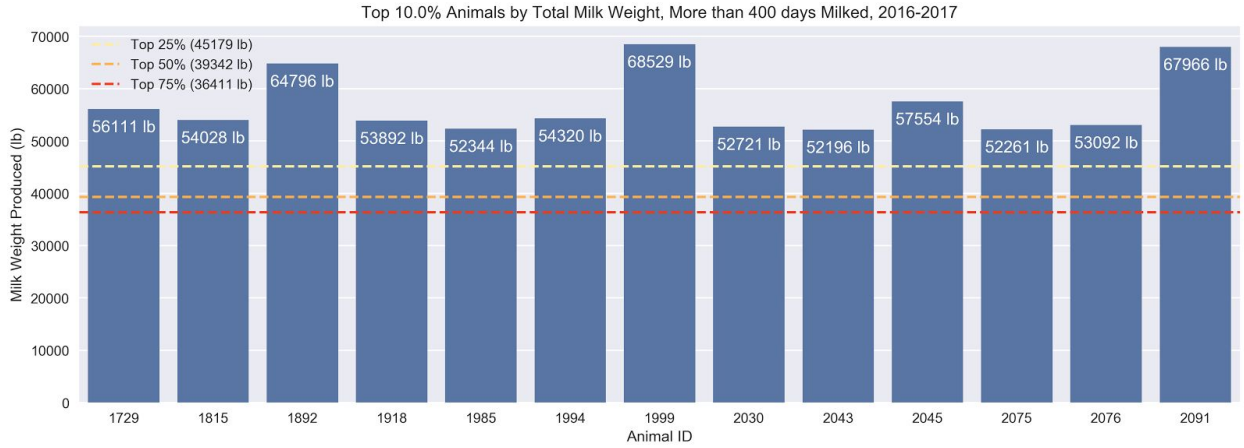
Calving Statistics

Calvings occurred regularly throughout the 2016 and 2017. The monthly average number of calves born was 16.75, with 203 calves born in 2016 and 199 calves born in 2017. The plot below indicates that a steady stream of calves allowed for cows to have to a regular dry period 2-3 month dry period after each lactation while permitting the herd overall to keep a relatively consistent output for year-round income.



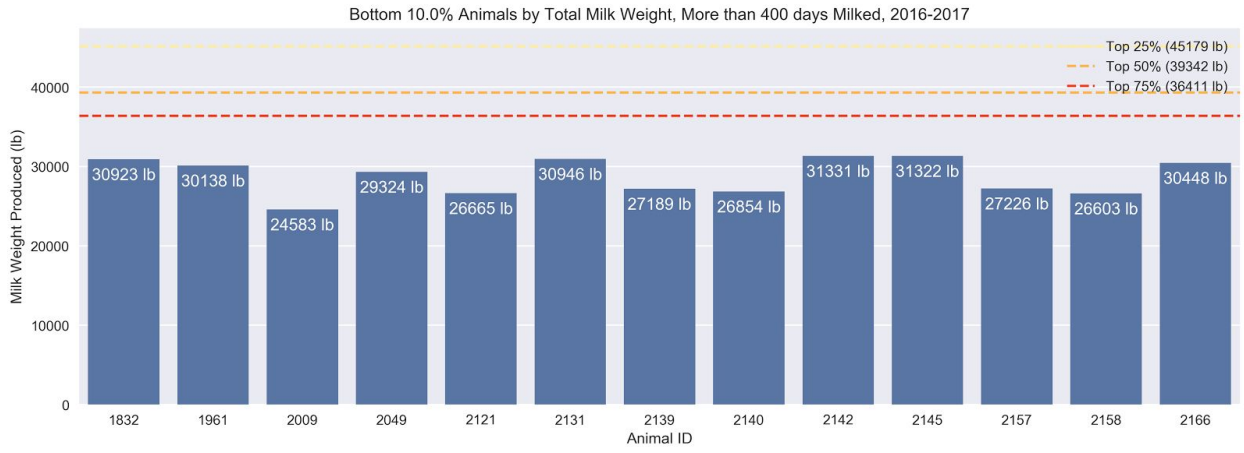
Top Producers By Milk Weight

Of cows that milked more than 400 days in that two year period at the top 25% produced at least 44,050 pounds of milk and the top 50% of the herd produced at least 39,343 pounds of milk in the same two year period. The plot below highlights the top 10 producing animals by total milk volume in that two year period. As high producers, these animals should continue to part of the herd, and animals with similar profiles should be added to the herd to ensure higher milk volumes better and associated financial success.



Under Performers By Milk Weight

For 2016 and 2017 calendar year, the bottom 10% of the animals by total milk volume are pictured below. These animals were milked for more than 400 days in that two year period, and produced at most 30,398 milk-pounds. These animals should be evaluated for medical conditions, and prioritized for replacement with cows that are capable of producing higher milk weights.



Facilities and Climate

In the years 2015, 2016, and 2017, Franklin County experienced a total of 6 days of where the average of 3 weather stations recorded maximum temperatures higher than 90 degrees, and 33 days where a low of temperature less than 10 degrees occurred. The volume of milk produced during and immediately after the extreme temperature days did not see a statistically significant impact on production volumes. As a result, current data suggest that existing facilities and practices have proven sufficiently effective for heat and cold abatement. The current recommendation is to maintain existing ventilation, cooling, insulation and heating strategies. Additional capital investment to improve these facilities and practices beyond regular maintenance may not lead to improved milk volumes.

Identifying Clusters of Low Performers

Identification of low and high performing animals is essential to making well-informed decisions regarding the retention and removal of animals from the milking herd. Finding clusters of low performers based on an assortment of available metrics may give the herd manager additional tools to help decide whether or not to retain an animal in the milking herd.

Collecting Data

The following table demonstrates the data elements gathered for this analysis.

	Milking Data			Genetics Data		Classification Data		
Animal ID	Days Milked	Per Day Average (lb)	Total Milk-Pounds (lb)	Milk Score	CTPI	Final Score	Dairy Form	Udder Score
1997	576	76.65	44148.3	-134	1547	77	22	73
2058	537	67.4	36193.7	-353	1423	80	35	82
2018	604	72.24	43630.6	-313	1663	85	35	86
2083	603	69.1	41666.4	-2	1650	73	27	70
1933	544	70.12	38143.1	-530	1382	81	25	78
1954	480	77.13	37023.9	456	2004	83	24	83
2109	585	73.32	42894.6	430	2118	76	22	77
2031	536	79.75	42747.9	-122	1504	81	31	81
2052	483	76.75	37068.3	189	1626	80	42	82

Visualizing the Herd

After reducing the dimensions of the dataset from eight dimensions to two, through Principal Component Analysis (PCA). A methodology known as k-means analysis was applied to identify the groups within the herd. The lowest performing animals are concentrated in a single identified group.

The animals in the low performing cluster should be prioritized for replacement and removal from the herd, however, considering the relatively weak nature of the clustering further metrics and analysis should be included before removing an animal from the herd.

Future Considerations

The current analysis considered only the milk production data available in years 2016 and 2017. Future considerations for this analysis include:

- Comparison of animal production in the first lactation period
- Offspring Survival Rates
- Health Records

Estimating Future Milk Weights

The ability to predict future milk production translates into a single tangible business objective, modeling future cash flows. The ability to predict the total dollar value of milk produced during a single cows lactation cycle requires some additional components.

- Expected future calving date
- Planned length of dry periods
- Projected sale price for milk produced

These elements, coupled with the total predicted milk volumes over a given location period can produce an estimate of income from a given animal over the projected period. When applied across the herd, it can give a sense of predictability to overall income. Prerequisite to building this estimate of financial success is a reliable model of milk volume produced on a given day by an animal. The following steps were taken to produce such a model.

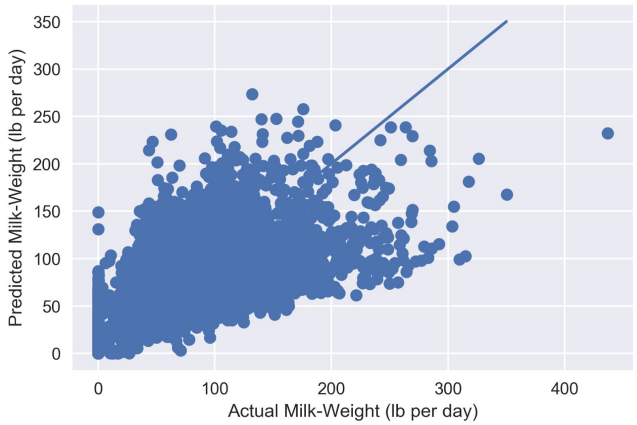
Collecting The Data

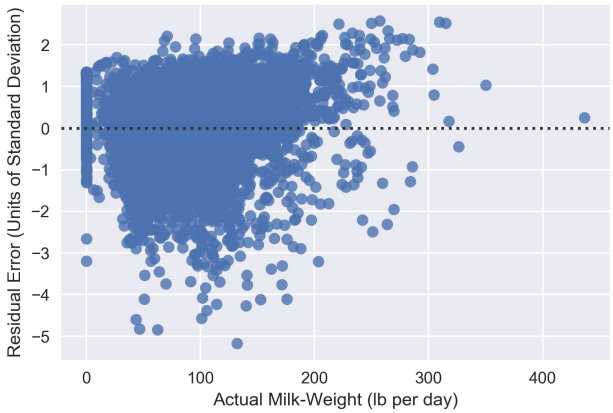
A series of single-day production profiles include data such as milk-weight, the age of the animal, days after calving, prior year production, classification scores, and genetics metrics. The following image is an example of this data.

Milk Weight		Age and Reproductive Data		Prior Year Performance	Classification Data					Genetics Data	
Predicted Milk Weight (lb)	Acutal Milk Weight (lb)	Days Since Calving	Age (Months)	Weight Per Day Milked	Dairy Form	Udder Score	Dairy Score Aggregate	Final Score	Breed Age Average	Milk Score	CTPI
112.2	156.7	173	45	78	25	83	83	82	106.2	418	1883
69.7	62.4	83	55	97	37	87	88	86	107.2	76	1510
80.4	80.7	84	75	86	34	84	87	85	106.2	-554	1349
64.0	65.0	262	58	35	35	81	88	82	102.2	138	1626
39.3	41.0	270	49	90	33	87	85	85	109.4	-427	1598
94.7	97.9	111	55	82	42	90	92	88	109.1	885	1744
32.9	12.8	281	57	83	35	82	87	85	107.6	-60	1734
94.9	95.2	125	50	61	21	82	86	82	102.5	-460	1674
47.5	47.2	276	45	55	26	86	83	85	107.4	-585	1756
80.6	82.7	162	42	73	23	85	82	82	103.6	486	1835

Estimating Milk Production

Provided aggregated data, a series of regression algorithms were applied. Of those tested, a random forest regressor presented the best results. However, the model had limited success, explaining only approximately 28% of the variability in milk production. The following plots provide a visualization of the results.





Future Considerations

Given the relatively low level of accuracy, additional experimentation should be conducted to improve the reliability of the model, before leveraging it for business decisions. Considerations for future iterations include:

- Prediction of milk-weight per week, or per month
- Prediction of milk-weight at herd level, using the 'average' animal
- Prediction of total milk-weight produced during entire lactation
- Collection of additional data and development of features such as
    - Number of lactations
    - Changes in Classification Score
    - Animal Weight
    - Milk Production Performance in 1st, 2nd, and 3rd lactations

As other years of data become available, these data-sets should be included in the current analysis and experimented upon to improve the predictive accuracy of the model.

Selected Terms

Milk Weight (milk-pounds) - The amount of milk produced by an animal. Measured in pounds of milk. For reference, a gallon of milk weighs approximately 8.6 pounds.

Dry Period - The period when a cow is not producing milk. Often serves as a time of rest following a lactation period.

Lactation Period - The period when a cow is producing milk.

Days Since Calving - The number of days that have passed since a cow has given birth.

Linear Classification Score - An integer score between 50-99 given to a milk cow, providing a numerical representation of how well a the physical attributes of an animal fits the profile of an 'ideal' milking cow. A weighted summarization of 18+ assessments of a given animal.

Data Sources

Data Source: Milk Weight

Daily milk-pound production data was derived from the on-site storage from the DeLaval - ALPRO™ herd management system from log files. Daily log files were collected for a date range spanning from July 2015 to December 2017. Approximately 15 files were corrupted, and no log files were retained. The results of these milking sessions are captured daily system logs in a series of text files from the local storage. The following lines provide an example of relevant data elements:

04:52:14 R 1831 Cow Duration1 6:25
04:52:14 R 1831 Cow AverFlow1 3.6
04:52:14 R 1831 Cow PeakFlow1 4.8
04:52:14 R 1831 Cow MilkToday1 23.2

The lines above suggest that Cow #1831 produced 23.2 pounds of milk, in six minutes and twenty-five seconds with an average flow rate of 3.6 lb/min and a peak flow of 4.8 lb/min. Also, this milking occurred at 04:52:14 am.

Acquisition

The system logs were manually retrieved from the herd management system and uploaded into secured private storage utilizing Amazon Web Services (AWS) for on-demand, repeatable retrieval, processing, and backup.

Data Wrangling

Before analysis, the contents of each log file are downloaded via script from AWS and brought into local storage. Each file is processed individually and put into local storage for future analysis.

Future Improvements

This process can be improved through an automated retrieval, ingestion, and cleansing of daily milk production data. This process would be enabled by the connection of the herd management system to an active network connection and the creation of automated scripts to conduct daily uploads of production data.

Data Source: Linear Classification Score

Linear Classification Scores provide a periodic assessment of the physical attributes of a given animal. Animals are classified on a scale from 50-99 based on some measured characteristics for comparison against the 'ideal' milking cow. These Linear Classification Reports were conducted by a representative of Holstein Association USA between August 8, 2014, and July 10, 2017.

BARN_ID,AGE,LAC,DATE_CALVED,ST,SR,BD,DF,RA,RW,LS,RL,LO,FA,FU,UH,UW,UC,UD,TP,RT,TL,UT,CS,FC,DS,RP,FL,MS,FS,E,%BAA
1485,7-Jul,6,7/10/14,50,45,44,42,25,35,29,25,25,25,14,26,35,50,5,35,35,35,35,25,92,92,92,82,80,85,,106
1542,9-Jun,4,8/6/13,50,35,35,42,15,35,50,25,,25,35,36,35,35,40,25,25,25,26,17,93,93,90,84,93,91,2,113.5

The example above shows the scoring for cows number 1485 and 1542. The animals received a final linear score of 85 and 91 respectively. The assessment occurred on August 5, 2014.

Acquisition

They were retrieved in the form of paper reports. The contents of the reports were scanned to PDF and parsed into CSV files using the program PDF Element by Wondershare. The resulting CSVs were uploaded to a private AWS S3 bucket for on-demand, repeatable retrieval via script

Future Improvements

This process can be improved through integration with the Holstein USA online systems. A software integration was explored early in the process but was abandoned due to cost prohibitive pay-per-drink model per-animal per classification. In the event of further automation, an alternative data acquisition process would be required to prevent the analysis from becoming too costly.

Data Source: Pedigree Scores

Holstein Association USA conducts additional analysis on individual animal genetics based on available pedigree data, genomic sequencing, as well as actual production information from the animal and its genetic siblings where available. CTPI and Milk are two values from this report that represent the CTPI as an aggregated indicator of milking performance and Milk as an indicator focused solely on the likelihood of higher volumes of milk production. In both cases, higher values are more favorable.

ANIMAL_ID,NAME,FS,PRO,%P,Fat,%F,Rel,Milk,SCS,PL,DPR,TYPE,REL,UDC,FLC,CTPI
1999 ," BELSHWAY PLANET 1999
USA 71404944100-NA12/12/2012",86 ,49,-0.02,40,-0.09,50 ,1772,2.94 ,3.4,0.2,1.34,53
,0.55,-0.29 ,2198
2043 ," BELSHWAY MASSEY 2043
USA 72758233100-NA 06/26/2013",79 ,36,0.01,38,-0.01,47 ,1132,2.76 ,3.6,-0.2,0.94,53
,0.66,0.94 ,2150

The example above indicates that animal with the ID of 1999 had a Milk Indicator of 1772 a CTPI of 2198. Cow #2043 had a milk indicator of 1132.

Acquisition

They were retrieved in the form of paper reports. The contents of the reports were scanned to PDF and parsed into CSV files using the program PDF Element by Wondershare. The resulting CSVs were uploaded to a private AWS S3 bucket for on-demand, repeatable retrieval via script

Future Improvements

This portion of the data pipeline can be improved through integration with the Holstein USA online systems. A software integration was explored early in the process but was abandoned due to cost prohibitive pay-per-drink model. The expected format of the CTPI report is available online. In the event of further automation, an alternative data acquisition process would be required to prevent this analysis from becoming too costly.

Data Source: Weather

The weather data set consists of daily summaries of weather measurements for Franklin County, Pennsylvania such as low temperature, high temperature, and total precipitation. The following provides an example of the CSV file format.

```
STATION,NAME,LATITUDE,LONGITUDE,ELEVATION,DATE,PRCP,PRCP_ATTRIBUTES
,SNOW,SNOW_ATTRIBUTES,SNWD,SNWD_ATTRIBUTES,TMAX,TMAX_ATTRIBUTES,
TMIN,TMIN_ATTRIBUTES,TOBS,TOBS_ATTRIBUTES,WESD,WESD_ATTRIBUTES,WE
SF,WESF_ATTRIBUTES,WT01,WT01_ATTRIBUTES,WT03,WT03_ATTRIBUTES,WT04,
WT04_ATTRIBUTES,WT06,WT06_ATTRIBUTES,WT11,WT11_ATTRIBUTES
USC00361354,"CHAMBERSBURG, PA
US",39.9353,-77.6394,195.1,2016-01-01,0,",,7,2100",0,",,7",0,",,7",38,",,7",34,",,7",34,",,7,2
100",,,,,,,,,,,,,,
USC00361354,"CHAMBERSBURG, PA
US",39.9353,-77.6394,195.1,2016-01-02,0,",,7,2100",0,",,7",0,",,7",42,",,7",28,",,7",29,",,7,2
100",,,,,,,,,,,,,,
```

Acquisition

The CSV files were requested from the NOAA Online Climate Data Online Search for full calendar year 2014, 2015, and 2016, and then again for all available data in 2017. The resulting CSV files were uploaded to AWS S3 to be programmatically retrieved by the script get_data.py. The raw files are processed by in the script parse_weather.py to produce daily weather summaries.

Future Improvements

Automated scripts with the NOAA weather API would reduce the workload required to collect this data.

Data Source: Calving Records

The calving data set consists of daily records of animals born, their identification number and their mother's identification number. These records were transposed to CSV format to collect the birthdates of each animal in the herd, as well as the calving dates for animals between January 1st, 2015 to December 31st, 2017.

Acquisition

The CSV files were stored on AWS S3 to be programmatically retrieved by the script get_data.py. The raw files are processed by in the scripts parse_birthdates.py and parse_calvings.py.

Future Improvements

This data is also collected by Holstein USA. Scripted collection of this information should be investigated for future iterations.