Extraction, Transformation, and Load Technical Report

Chicken Feed versus Sports

**TABLE OF CONTENTS**

1. [Introduction 3](#_bookmark0)
   1. [Summary 3](#_bookmark1)
   2. [Scope 3](#_bookmark2)
   3. [Technologies and resource contributions 3](#_bookmark3)
   4. [Definitions, Acronyms and Abbreviations 3](#_bookmark4)
2. [ETL Details 4](#_bookmark5)
   1. [Data Import/Extract Sources and Method 4](#_bookmark6)
   2. [Data Acquisition 4](#_bookmark7)
   3. [Data Transform 4](#_bookmark8)
   4. [Data Integrity 4](#_bookmark9)
   5. [Data Refresh Frequency 4](#_bookmark10)
   6. [Data Security 4](#_bookmark11)
   7. [Data Loading and Availability 5](#_bookmark12)
3. [Data Quality 6](#_bookmark13)

**1.** **INTRODUCTION**

*The purpose of the Extraction, Transformation, and Load (ETL) Technical Report is to capture details that pertain specifically to ETL portion of the data pipeline that is to be used in a data science project. This however does keep in mind the final target objective while performing the ETL.*

# Summary

We intend to argue a correlation between sports and chicken feed as it relates to the price of chicken wings.

# Scope

We will be using data sources for 5 years from Sports Media Watch for the sports statics. We are also using Urner Barry for the historical data of the chicken wing prices. And we are also using Barley statistics from Index Mundi for the prices of chicken feed.

# Technologies and resource contributions

Michael Coates – Creating the data sets and the correlation between the data sets.

Tim Veal – Creating the extraction and the narrative among the data.

# Definitions, Acronyms and Abbreviations

Chicken feed – The food that the chicken eats.

Chicken wings – Prices as defined by the northeast. They are several different sizes such as small, medium, and jumbo; in reference to this dataset for this ETL we are using jumbo priced wings in price per pound.

**2. ETL DETAILS**

*This section outlines a more detailed description of the processes utilized/proposed to achieve the objectives of this initiative.*

# Data Import/Extract Sources and Method

As mentioned above, the data is coming from three sources: Sports Media Watch, Urner Barry, and Barley statistics from Index Mundi.

Sports Media Watch – Sports Media watch is a free for all information URL that brings the famous Nielson ratings all in one place to provide the detailed information of sports including but not limited to: the teams that competed, viewers in millions, the channel that aired the game, the date of the game such as week and month, and other information.

Urner Barry – Urner Barry is a paid / subscription-based platform for any vendor that requires the data for their business or other reasons. It includes the prices for all sorts of fresh foods at wholesale prices such as: red meat, poultry, seafood, egg, and grains; but it also includes weather statistics that could affect those prices.

Barley from Index Mundi – Index mundi is a wisely open data site that organizes several types of data sets by regions, by state, by industry, and other separations in a very organized manner and easy to use format.

# Data Acquisition

This data set was static for the most part. We were able to use the previous historical data for the last 5 years. Because we are looking at how prices were affected by sports over the years and not currently, we do not need to constantly update the data. There were no pre-requisites that were needed to obtain the data since it is openly available.

# Data Transform

Before we could immediately use the data, we had to change it a little. The sports data was organized into weeks, while the chicken data was organized into months and quarters. Therefore, we had to easily group the sport games into months instead of weeks.

# Data Integrity

In this section discuss the reliability of the extraction source data (e.g., missing data, dates stored as text, invalid code values, text fields with odd characters, etc.). Address the frequency with which the data sources are updated and if it is necessary to update the local data at the same frequency. Lastly, how if any notification can be received when the source data is updated; and what if any notification will be sent to the internal team when the local dataset is updated.

# Data Refresh Frequency

Aligning with the dataset of the chicken prices, I believe that it should be updated on a monthly basis to continue to watch the trends of how sports affects chicken wing prices.

# Data Security

There are no data security concerns that need to be addressed as this is open data.

# Data Loading and Availability

We will be using Jupyter Notebook to present the data.

**3. DATA QUALITY**

Address in this section success criteria for this project. Summarize the parameter KPIs such as Totals and expected counts. What user acceptance testing was performed and what were the outcomes. What is the recommended site acceptance testing that your client can perform to ensure the expected outcomes meets their expectations?