**Final Report**

**Background and Aims**

Facts:

* Idaho and Washington are adjacent US states. These two states account for nearly 70% of the total production.
* Over 95% of the potatoes planted in Idaho and Washington are Russet potato.
* In Idaho and Washington, farmers seed the potato in April through the first two weeks of May, and harvest the potato between Sept and Oct the same year.
* Temperature impacts on potato production:

Potato Crop Yield can be severely reduced by temperature fluctuations outside **5-30 °C**. This is out of farmers’ control, when compared with water availability, fertilization and pest control.

Questions our data may answer:

(1) In Idaho and Washington states, "Exceptional Temperature Days Count (between April 1 and Oct 31)" could be one of the major determinant of “Potato Crop Yield (cwt/acre)"? Does the history data prove this theory? Can we use weather data for the prediction of potato crop yield in Idaho and Washington?

(2) Did “Yearly New Crop Yield (cwt)” in Sep and Oct determine “the lowest retail price in the following two months”, or “the highest price in the following year” (normalized with inflation rate)?

**Data Extraction, Cleaning and Transformation**

Temperature Data - Bennet Farrington

(to be updated)

Potato Yearly Production Data – Carlton Lewis

Data Source: <https://usda.library.cornell.edu/concern/publications/fx719m44h?locale=en>

Downloaded csv files that contained the production and yield per acre by state. These files were chosen because the data could be used to analyze the effect on the total production and acreage yield that weather extremes had on the annual potato crop over a period of time. The data can also be used to graph the annual production and acreage yield by state. These graphs can used in conjunction with the pricing and temperature graphs to determine if there is a correlation between the weather and these variables.

The csv files were converted into dataframes and all of the unnecessary rows were dropped. The dataframes were converted back to csv files for loading into the database. These rows were removed because either they contained no data or information that was not relevant to our data analysis.

Retail Price Data - Mubbasheer Ahmed

Data Source: <https://data.bls.gov/timeseries/APU0000712112?amp%253bdata_tool=XGtable&output_view=data&include_graphs=true>

We were interested in the price of potatoes as experienced by consumers over a 10year period. We found both a table format in HTML and excel download file via Bureau of Labor Statistics Data website. This data was pertinent given that it was monthly and our weather patterns were based on monthly trends. We downloaded the excel file and cleaned the data in excel and saved this as a CSV file. From there we imported this into a pandas data-frame. From here—the data was graphed for the purpose evaluating the price and temperature relationship.

**Data Uploading** -Li Gao

We uploaded the above-mentioned datasets into MongoDB for further analysis.

**Discussion**

One use for our database could be to predict what effect current weather conditions should have on the price of potato futures. If temperatures were outside of the optimal range for potato growth for x number of days, we should be able to make an accurate forecast of the future prices of potatoes for that season. This information would allow us to determine if the future prices of potatoes are high or low and we would be able to buy and sell futures accordingly. We have eight years of production, pricing and weather data to analyze and use for our forecasting.

Limit:

- We assume that in the past 10 years, the farmers utilized similar farming technology including fertilizer, watering and pest control.

- We assume that the soil conditions are similar no matter the area was farmed or not in previous years.

- Not only the weekly/monthly supply amount, but also the fluctuating demand will lead to potato price change. The supply and demand of fresh potato (good for 1 month in refrigerator) vs. frozen potato (good for up to 1 year in freezer) will regulate the market price simultaneously. Purchasing power of big buyers such as Mc Donald and Burger King may affect the price significantly. Potato consuming data could be a big plus to this project for prediction purpose.