**Nuts & Oils Analysis Slide Bullet Points**

* Almonds ~ Honey:
  + -0.8131253 Correlation
  + y = -15.02(Honey) + 1737890.72. *(Negative Correlation)*
  + 51% confident that Honey accounts for roughly 63% of Almond production value on any given year.
* Walnut ~ Honey:
  + -0.7421160 Correlation
  + y = -5.381(Honey) + 750350.122 *(Negative Correlation)*
  + 37% confident that Honey accounts for roughly 52% of Almond production value on any given year
* Rapeseed Oil ~ Honey:
  + -0.7345804 Correlation
  + y = -7.709(Honey) + 1,028,000 *(Negative Correlation)*
  + 36% confident honey can account for 50% of the production value of Rapeseed Oil on any given year
* Soybean Oil ~ Honey:
  + -0.6745064 Correlation
  + Y = -32.06(Honey) + 11,290,000 *(Negative Correlation)*
  + 24% confident honey production significantly impacts the production of Soybean oil by 41%

**Slide 20 Script:**

Category 4 of our commodity deep-dive analysis is the nuts and oils group.

This includes:

Almonds, Filberts, Pistachios, Walnuts, Peanut Oil, Olive Oil, Rapeseed Oil, Cottonseed Oil, & Soybean Oil

Upon running the correlation matrix here, we found four high correlators with Honey:

Almonds at 81%, Walnuts at 74%, Rapeseed Oil at 73%, and Soybean Oil at 67% correlation

Here is the linear trend plot with Honey Production Value as our x axis and commodity production value as our y. We had to create an individual axis for soybean oil to make it plottable as the production values far exceed the other commodities by tens of millions.

Something interesting to note here is that among the top correlators, they are all showing as negatively correlated with honey production.

At the bottom are the equations received from each linear regression analysis.

The level of confidence we have that honey influences each of these four nut and oil production values, falls in the same order as our correlation percentages.

* + 51% confident that Honey accounts for roughly 63% of Almond production value on any given year.
  + 37% confident that Honey accounts for roughly 52% of Walnut production value on any given year
  + 36% confident honey can account for 50% of the production value of Rapeseed Oil on any given year
  + 24% confident honey production significantly impacts the production of Soybean oil by 41%

Note: Interesting trend similarity

Filbert Trivia!!

**Slide 16 Script:**

Upon attempting to wrangle and analyze our data, we quickly realized how extensive the dataset was,

Wrangling alone was difficult, but trying to run an analysis with every commodity all at once just wouldn’t work because of multicollinearity. Altogether, the commodities were overfitting the data. So, we divided our data up into 5 groups that each of us could manage.

Category 1 is the Meats & Dairy group

Category 2 is Fruits & Vegetables

Category 3 is Grains & Oats

Category 4 is Nuts & Oils

And Category 5 is all the miscellaneous commodities that didn’t quite fit into any of the other groups

All of the values of commodity production you will see are listed in Metric Tons

Once we had our categories, the process of our preliminary analysis was to get a correlation matrix,

Single out the commodity production values that correlated highly with honey (we defined high correlation to be .65 or above). We then ran linear regressions on each high correlator to get the linear equation for each commodity as it relates to honey, to all be saved for use in the final predictive analysis. In our y=mx+b equations, our x variable is honey and y is our high correlator.

**Slide 26 Script:**

- Finding Better Data

- Reading the many 400 page texts Professor Raetano gave us on ARIMAs

- Developing deep understanding of ARIMAs to run them all manually &

fine-tune each parameter

- Looking into improving R2 scores in ARIMAs

- Fixing the Homogeneity issues common across many of the linear regressions we did on each commodity

- Bigger effort on removing outliers

- Looking into other factors that effect production values and incorporating them to find causation/better predictive measures

- Looking at the global data in the USDA