**UNIT 3: Little Data to Big Data**

**Project 3.2.4 Making Meaning from Data**

**Enduring Understandings**  
*EU-DAT-2 : Programs can be used to process data, which allows users to discover information and create new knowledge.*  
*EU-AAP-4: There exist problems that the computer cannot solve, and even when the computer can solve a problem, it may not be able to do so in a reasonable amount of time.*

**\*\*Instructions:** You may work in pairs to complete this Project.  Please change the text color of your responses to red text.  Please organize the endings to each page.

STEP 11: Test that the data cleansing worked by printing the Value column before and after your cleaning algorithm. Paste your solution in the space provided.

|  |
| --- |
|  |

Document with pseudocode the data-cleansing algorithm you used and include a brief description of the cleaning effect.

|  |
| --- |
|  |

STEP 13: Iterate over your unique states and experiment with the following two algorithms that use pandas, with grouping and without. Each retrieves data from the CSV file and stores the data in two arrays. Observe the print results to see how data becomes grouped.

all\_honey = []

all\_states = []

# without grouping

for state in unique\_states:

honey\_data = df[df['State'] == state])['Value']

print (state, honey\_data.sum())

all\_honey.append(honey\_data.sum())

all\_states.append(state)

# with grouping

for state in unique\_states:

honey\_data = df[df['State'] == state].groupby('Year')['Value']

print (state, honey\_data.sum())

all\_honey.append(honey\_data.sum())

all\_states.append(state)

Describe how your data collection algorithm works using the groupby method.

|  |
| --- |
|  |

STEP 15: Add comments to your code, summarizing how your data collection algorithm works.

|  |
| --- |
|  |

STEP 19: Use three graphs to visualize the data. One way to do this is to show large honey producers, mid-level honey producers, and small producers, each on their own graph. Insert your code or a link to your code below.

|  |
| --- |
|  |

STEP 22: Use a new groupby statement to get yearly totals.  Add Comments and insert your code or a link to your code below.

|  |
| --- |
|  |

STEP 23:  Validate either an average or a total value using a process similar to the state-by-year validation. Show your work (image would suffice).  What is the value of manually validating the data?

|  |
| --- |
|  |

STEP 25:  Improve the look of your data visualizations by formatting the legend.

|  |
| --- |
|  |

STEP 26: With your data visualizations complete, you should now be ready to make some conclusions about the data. What does the data say about honey production? Do you think this correlates to problems with bee colonies in the U.S.? Note any unusual patterns or outliers in your data.

|  |
| --- |
|  |

STEP 30:  You will not need to submit your code for Step 29.  Make a conclusion about the operational facilities.  Then correlate the operational facilities to the honey production and revisit your conclusion.  Have operational facilities increased or decreased? How does the increase or decline relate to your conclusion about production rates?

|  |
| --- |
|  |