Brainstorming
What interesting questions might this dataset address? Brainstorm a handful of interesting questions, and scope them. Describe steps, methodology, and level of effort that would be required to answer each question. Additional, enriching, datasets are allowed, but not required.

Answer:
A number of questions can be asked, regarding the "as-is state", including the following:

1. Which job categories are the most prevalent in the VISA sponsoring data?
    a. nationwide
    b. by state of the sponsoring institution's headquarters

    To answer this question exists in the given data one will have to first collect the relevant rows from the existing data (i.e. "lca_case_soc_name", which provides title of the occupational group. and 'lca_case_employer_state', which provides the state of the employer). Then for (a) they will need to group all the rows by "lca_case_soc_name", count the VISA cases with each group and take the one with the largest count. For (b) one will have to group by both "lca_case_employer_state" and "lca_case_soc_name" before counting. Then they will need to choose the ones with the max count for each state.

2. By job category how do employers' average proposed salaries compare against the average prevailing wages for the same jobs?

    This question is similar to the one above, except that for each group, defined by the job category, one has to compute the mean proposed and prevailing salaries (after possibly transforming all the salaries to the same unit) and then define a measure of difference (i.e compute the mean of the percentage differences)

3. Is there a relationship between the number of VISA requests and number of VISA granted
    a. overall
    b. by state
    c. by employer
    d. by both employer and state

4. Is there a relationship between time to process a VISA and its final outcome?

5. Is there a relationship between proposed wage and approval rates?

    For the above three questions (3, 4, 5) one has to first decide how to measure relationship (e.g. through visualization or computing a measure of relevance, for example the correlation coefficient on the raw or properly transformed data)

6. What job categories have better chances to get approved
    - nationwide
    - by state
    - by industry where the sponsoring institution belongs

    For this question the existing data are not enough to provide answer. One will need to get collect data on the industry where each employer belongs, before following a process similar to the other questions above.

In addition, one can attempt to predict future events, based on the information provided in the existing dataset. A prevalent one is whether one can predict VISA approval from this dataset.: additional data (candidate's education, demographics, employer size, employer's industry).

This question will requires:

A thought process to decide what variate to use (i.e. what columns may have predictive power)

Possibly augment the data with additional data that can affect the accuracy of the prediction (e.g. educational level of the sponsored employee)

A prediction algorithm (e.g. logistic regression)

A decision on the evaluation metrics and the testing process (e.g. perform cross validation)

A process similar to the other question which will collect the relevant data and transform them into the form that is required by the prediction algorithm and methodology