* ~~Project title with student’s names~~
* Problem definition (be specific)
* The primary challenge is informing the California Insect Barcoding Initiative where to collect data in order to efficiently utilize their limited sampling resources
* Description of background:
  + Why is this meaningful?
  + What does it solve?
  + Related work?
  + Contribution?
* Description of dataset
  + Quantity (adequate size?)
  + Quality
  + Explain important attributes
* Description of methods
  + Description
  + Why was it used? (fine if saying it is the one you know…)
* Experiment setup and analysis of results
  + Select and describe *the most important* result
* Observation and conclusion
  + Good or bad, describe your understanding of the conclusion
* References
  + Half page is enough
  + Follow some standard way
  + Should be referenced in text

**Final Report (65% of project)**  
Due on **Dec 4, 11:59 PM**  
Max 8 page report + codes Submit a single zip file with the name, Final-proj-“team#”, which include report an codes.  
Late penalty: 30% deduction each day   
**Project Final Report** should include the followings   
Project title with student’s names   
Problem definition  
Description of background (why is this meaningful? What to solve? Related work?   
Contribution?)  
Description of dataset (quantity, quality)   
Description of methods used  
Experiment: experiment setup and analysis results  
Observation and Conclusion   
References

Ideas:

* Linear regression between # of individuals and human population, show relationship and explain how this can be used to encourage exploration of a causal relationship (use correlation coefficient and r^2 metric) (basically want to see if there’s a sampling bias by actually investigating if mantis biodiversity is caused by human density) **ARIELLE**
* Decision tree based on features (one tree per species) (one tree introduced, one tree native) **CHARLES**
* Nearest Neighbor - start with complete linkage and assess: (cluster by introduced vs. native) **DYLAN**
* Color by native vs. introduced
* Compare centroids
* Try to figure out which features are most explanatory
* Null result (no clustering) could imply further research needs to be done into competition
* ER Diagram for dataset description **BRIAN**
* Most of the writing: **BRIAN**
* Find text and references to work with here: <https://docs.google.com/document/d/12VbrE-YWvrlu3IxT_5lekDCCcTvvDTN6IUaSIISdXRM/edit>
* Apocalypse citations and other citations: <https://docs.google.com/document/d/1_MGv_APSpMvFkHJlPXYtec0vtNR1i7vwHe1lByIBFok/edit>

Problem Definition: The primary challenge is informing the California Insect Barcoding Initiative where to collect data in order to efficiently utilize their limited sampling resources