**Honeybees and Neonic Pesticides Data**

**Background:**

Honey is an important food source. The consumption of honey and bee larvae likely provided significant amounts of energy, supplementing meat and plant food. In 2006, beekeepers globally were struck by honey bee colony collapse disorder (CCD). The best way to to kill CCD is the use of a family of pesticides called neonics; however, the excess neonics may kill bees over extended periods. Thus, predicting the honey production and track the correlational evidence between the usage of neonics and honeybee colonies are very useful.

1. **What is the problem you want to solve?**

Two problems: one, predict the honey production.

Two, find out the correlational evidence between the usage of neonics and the numbers of honeybee colonies.

1. **Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that** they wouldn’t have done otherwise?

The beekeepers and the customers who consume the honey would be my clinet. My production prediction would give suggestions to beekeepers to obtain the highest value of sales; and the prediction on how to use neonics to obtain the most honey colonies would benefit the beekeeps to keep the bee healthy and maximize the honey production; finally, All these factors will promise the enough honey providing in the market for consumers.

1. **What data are you using? How will you acquire the data?**

**I am using the honey bees colonies and neonicotinoids data, which comes from Kaggle website:** <https://www.kaggle.com/kevinzmith/honey-with-neonic-pesticide#vHoneyNeonic_v03.csv>

1. **Briefly outline how you’ll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.**

My approach outline is data importing, data wrangling, data visualization and model predicting; The detail are as follows:

1. For the honey production prediction: Based on the production data and the usage of neonics data during 1998-2016, three different inference tools, frequentist interference, bootstrap interference and Bayesian interference would be employed and evaluated to find the best model to predict the production.
2. For the correlational evidence between the usage of neonics and the numbers of honeybee colonies: the kendall correlation method would be used to predict the correlation between five neonics and the number of honeybee colonies.
3. **What are your deliverables? Typically, this includes code, a paper, or a slide deck.**

There are mainly two deliverables for this project:

1. Jupyter notebook that includes all my raw code and reasoning for the decisions I made.
2. PowerPoint presentation that summarizes the key results from the project and future directions that would be interesting to pursue.