Springboard

Springboard Capstone 2 project report 1

# How to promote the honey production?

Yuanchun Wang

06/12/2020

# Honey and Honeybee

1. Honey is an important food source and the production is decreasing.
2. Honeybee colony collapse disorder is getting worse.
3. How to use neonics pesticide to save the honeybee and

Increse the honey production

# Outline

- Data acquiring
- Data wrangling
- Data visualization
- Statistical analysis
- Summary
- What’s next

# Data acquiring

- Honey production data is from Kaggle website:

https://www.kaggle.com/jessicali9530/honey-production

- Honeybee neonics data is from Kaggle website:

https://www.kaggle.com/kevinzmith/honey-with-neonic-pesticide#vHoneyNeonic_v03.csv

- Related references are downloaded from the google scholar.

# Who is this project for?

- The beekeepers: give the suggestions on how to increase the honey production and how to apply neotics to promote honeybee colony number increasing
- the customers who consume the honey: provide historical data analysis and production prediction to let everyone know that how the honey production will develop in future.
- Finally, this project will give suggestions on how to promise the enough honey providing in the market for consumers.

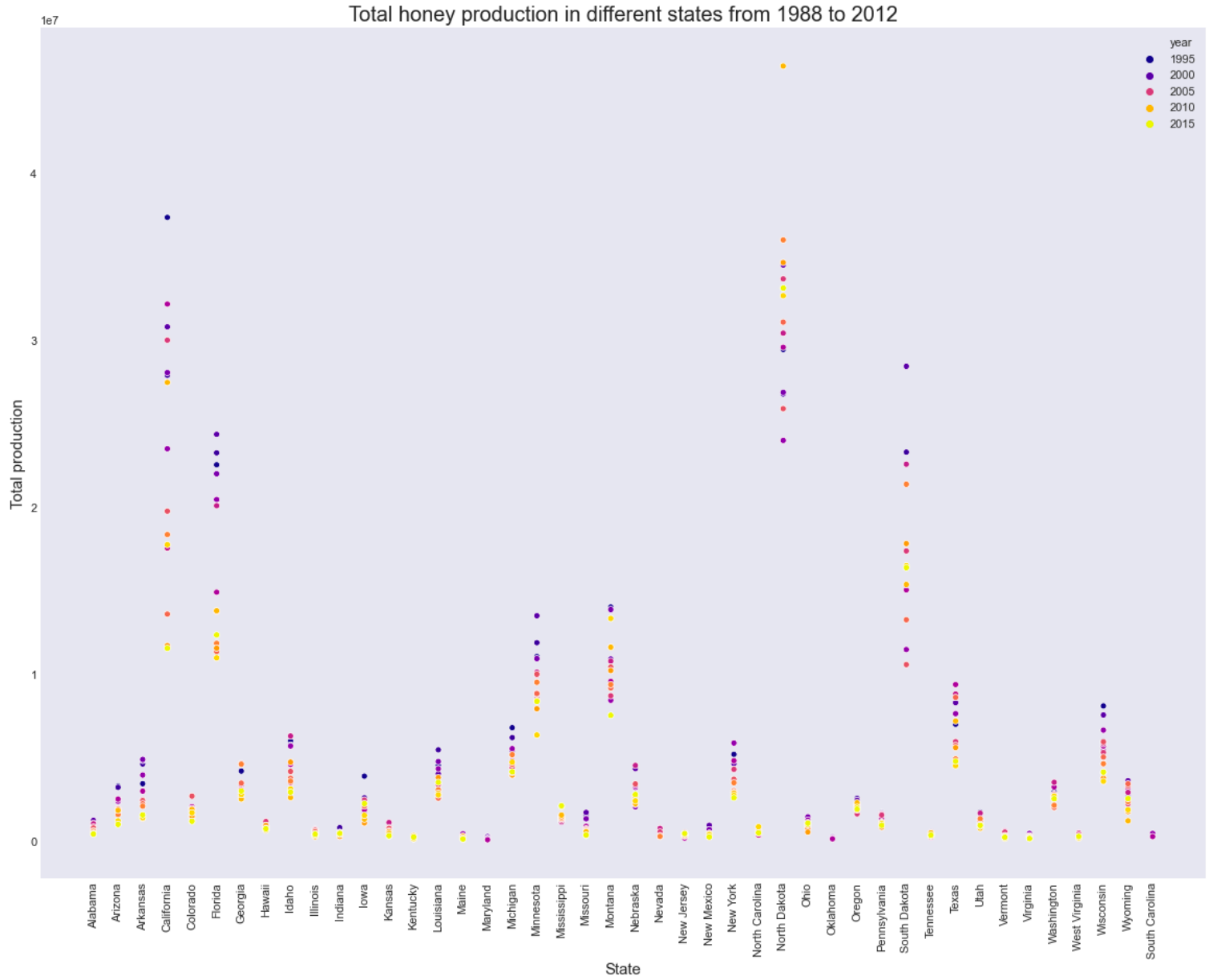
# Data Wrangling

1. Rename the columns to make the data easy to read.
2. Drop useless column ‘FLPS’.
3. Fill the empty space with 0.
4. Replace ‘state_code’ with full state name.

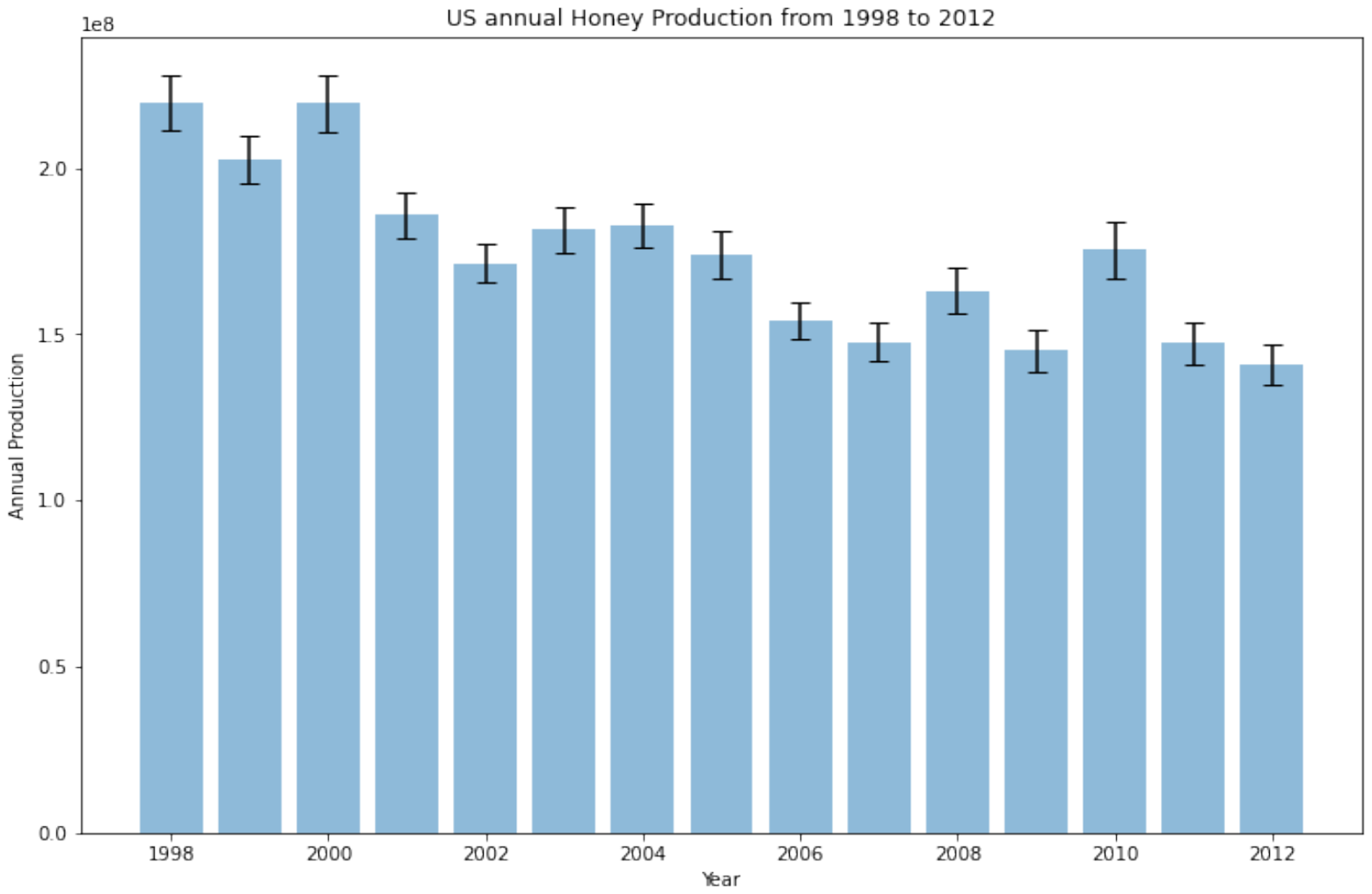
After applying the above steps, there are 626 rows left; for neonics data, after wrangling, there are 1132 rows left.

# Data Visualization

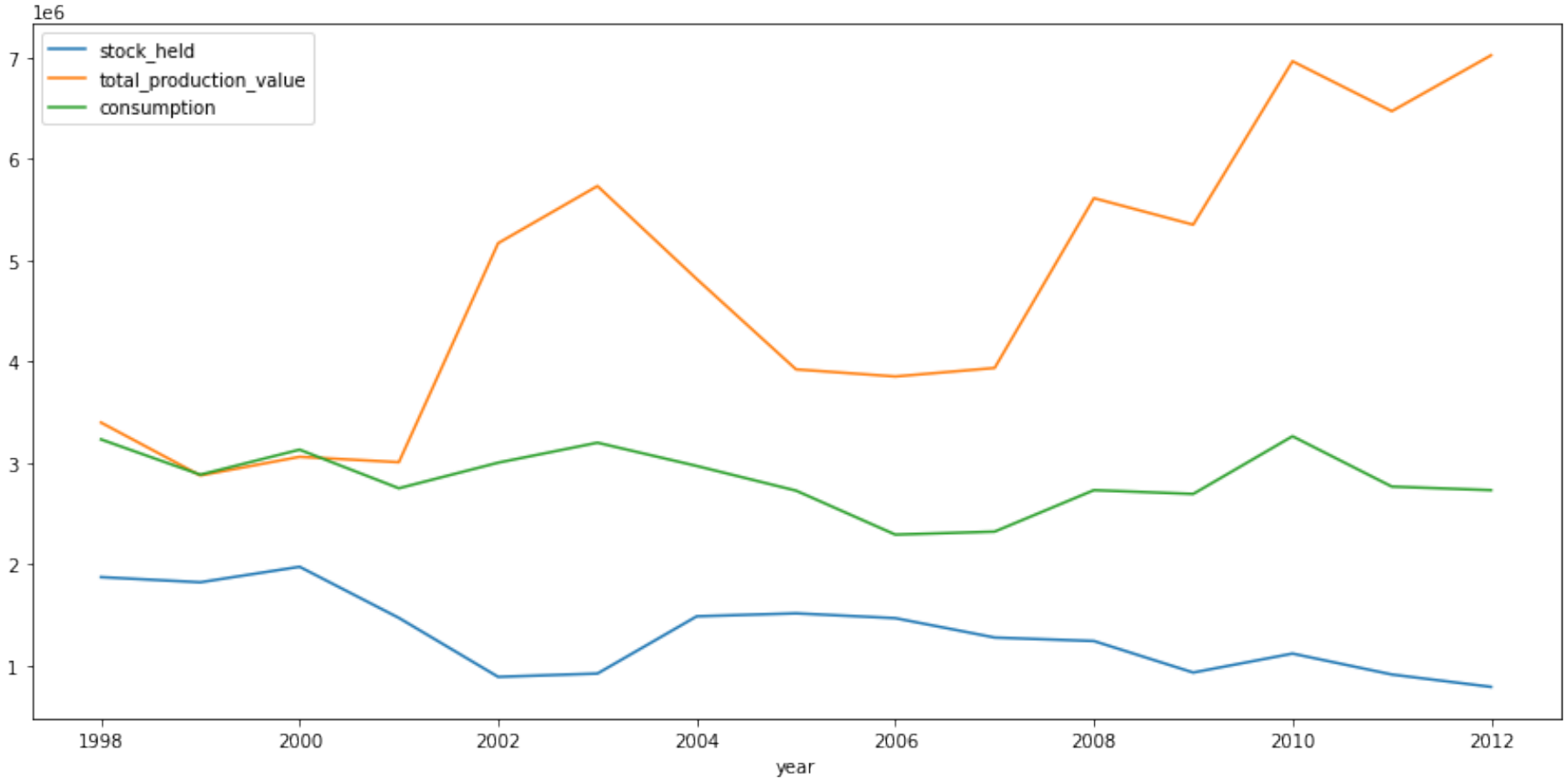
# The honey production decrease in most of the states.



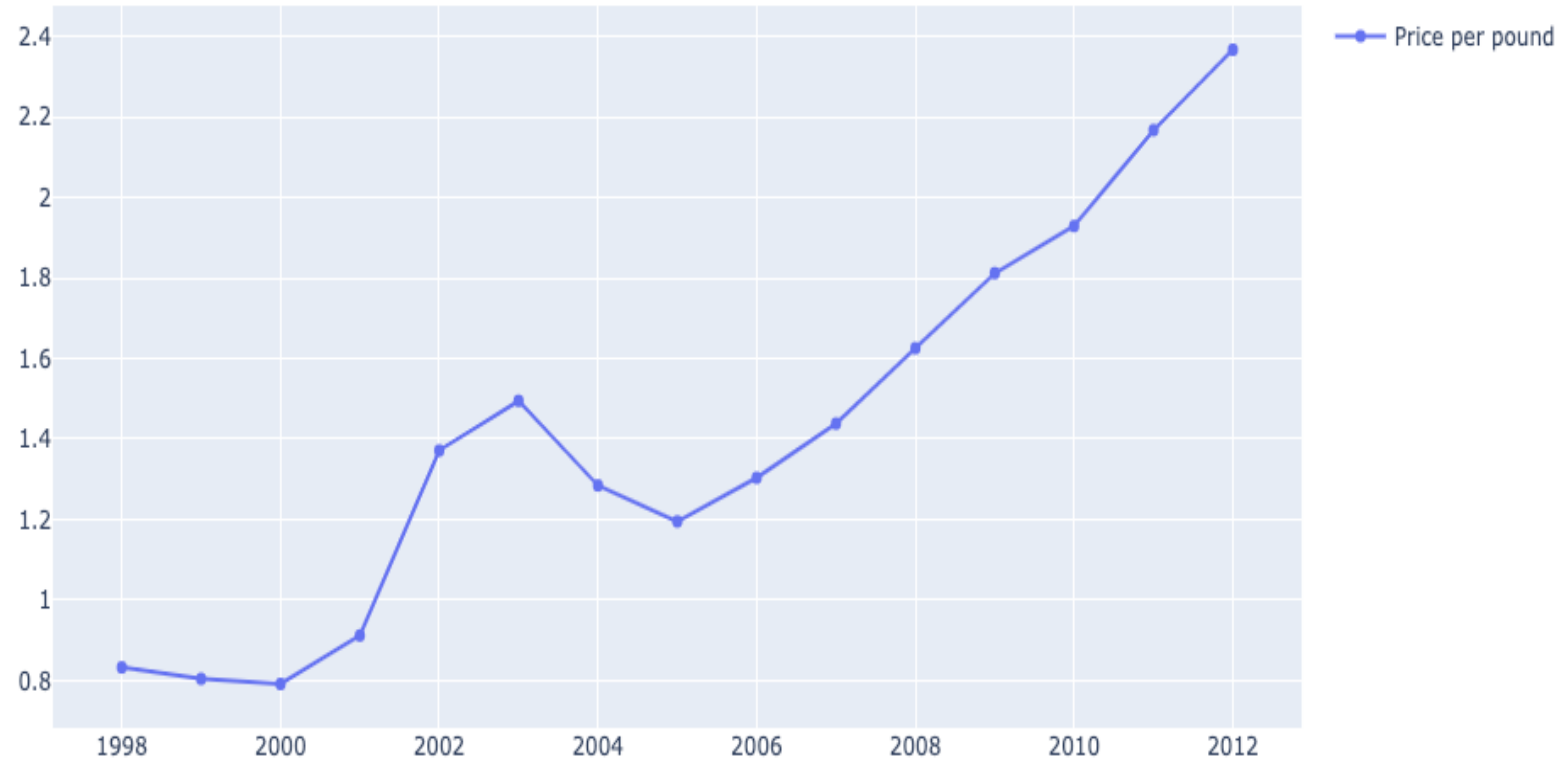
# The USA produces the maximum honey in 2000



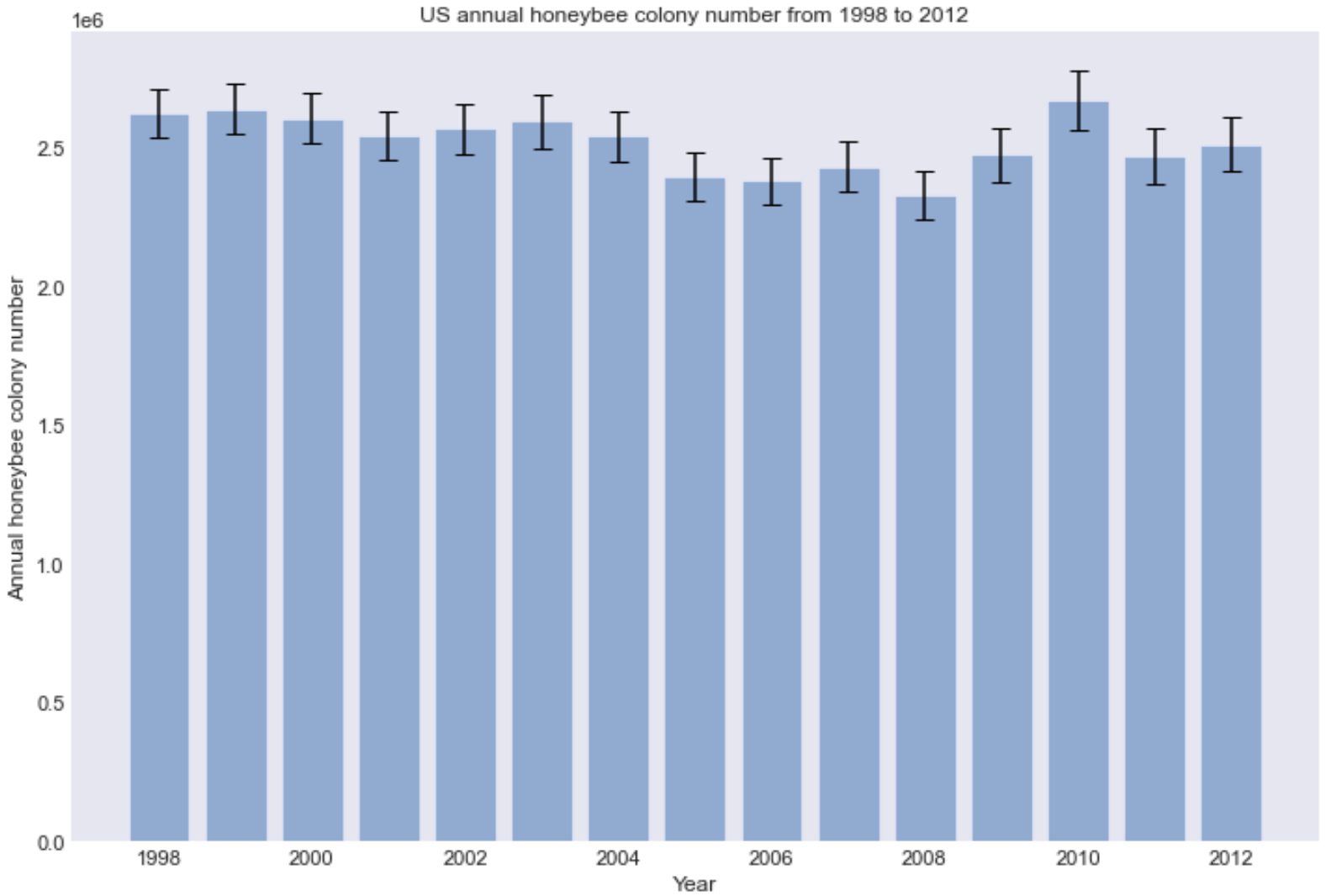
# The honey stock, total production value and consumption changing trend



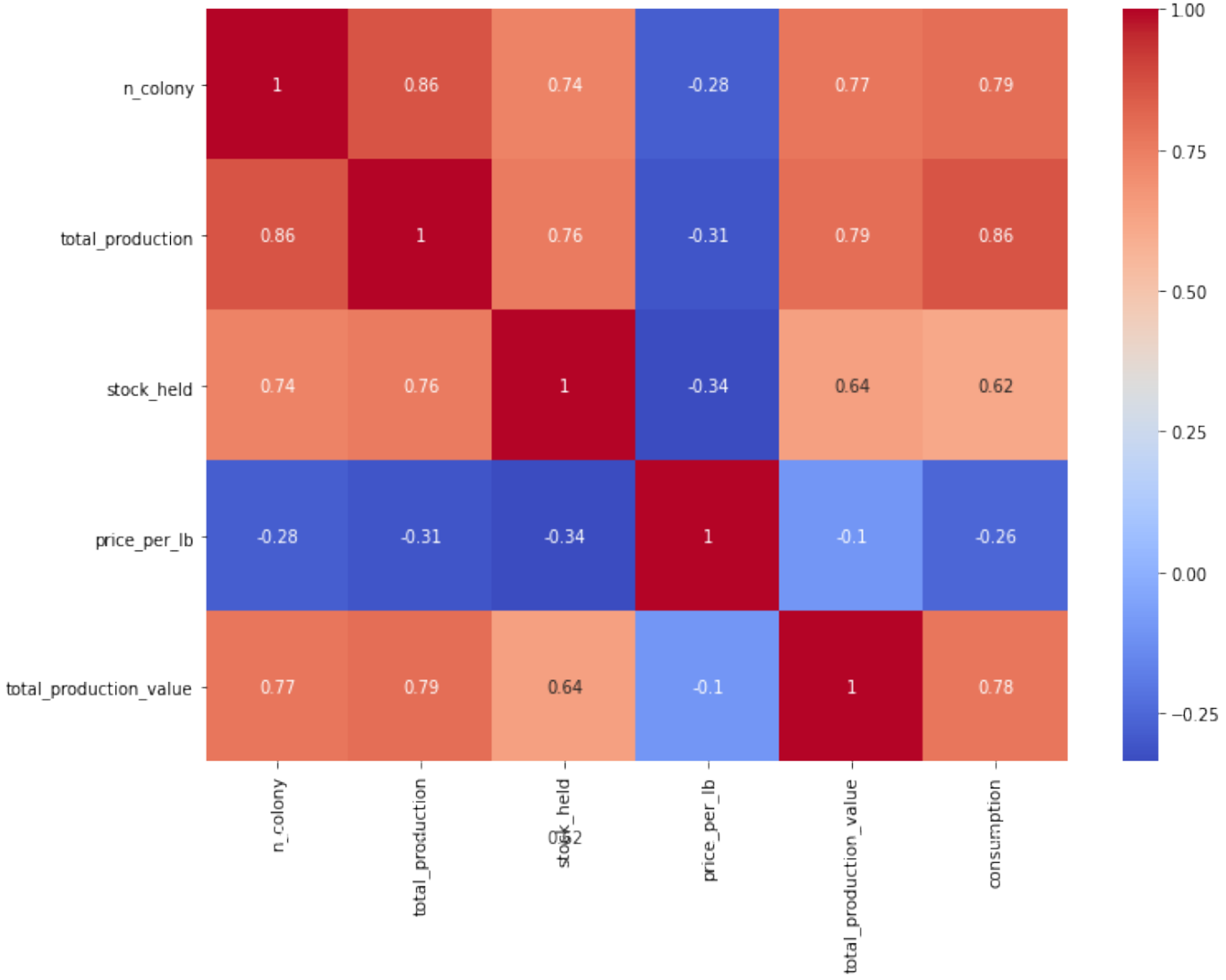
# The honey price exhibits the almost increasing trend



# US annual honeybee colony number are relatively stable during 1998 and 2012.



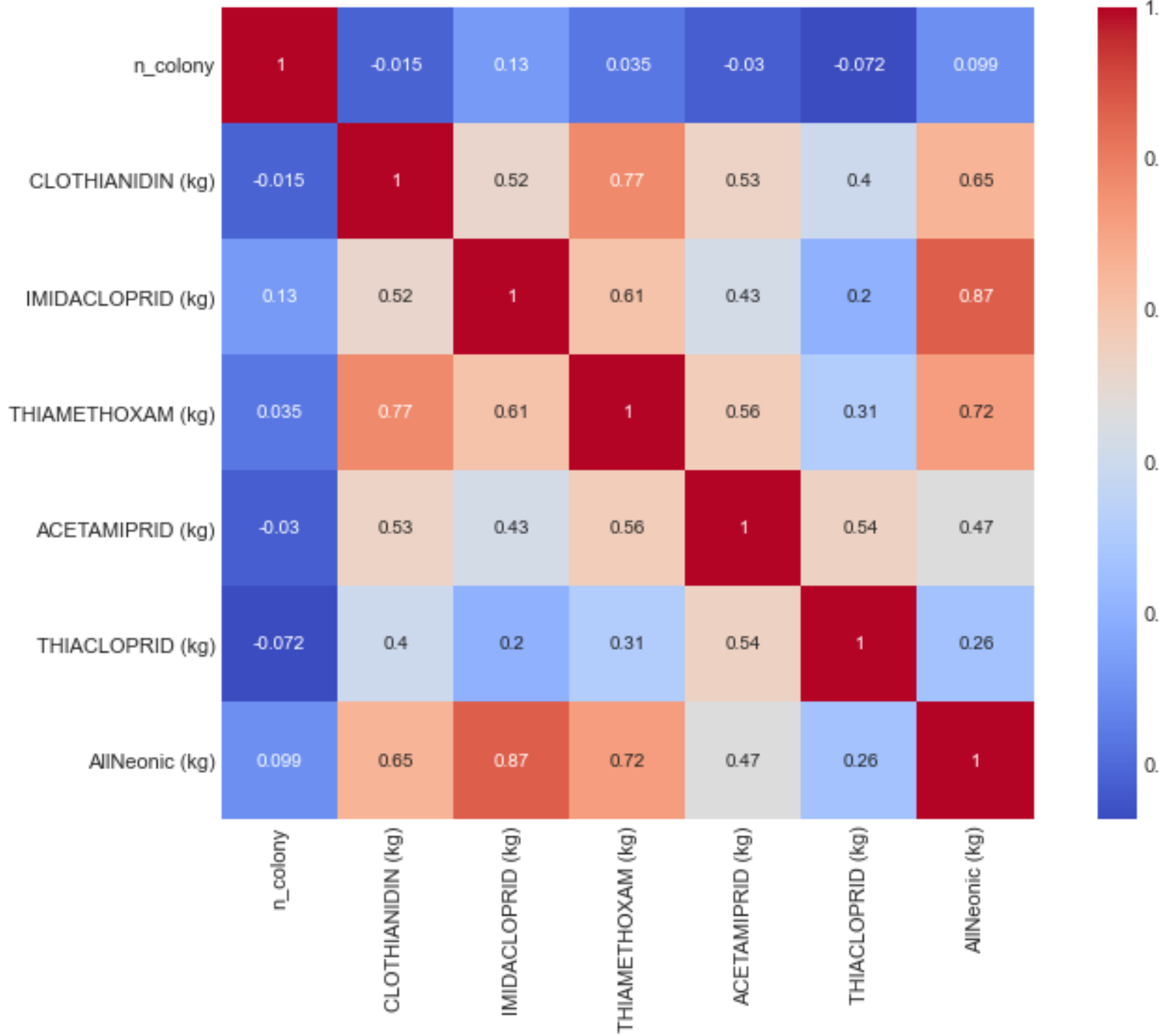
# The kendall correlation between honey price and production



# Summary I

- Honey price per pound has negative correlation with 'number of colony', 'total production' and 'stocks' at the correction value of '-0.28', '-0.31', '-0.34', which indicates that when the honey colony become less or total production goes down or stocks decreases, the honey price per pound increases.
- Colony number has strong correlation with 'total production'(0.86), 'stock held'(0.74), 'total production value'(0.77) and 'consumption'(0.79); with the colony number increasing, the honey production, stock and total production value all goes up.
- Consumption has strong correlation with 'total production'(0.86), too; it indicates that the more total production, the more consumption; if we want to increase the consumption, we have to improve production.
- The colony number plays a key in role in influencing the production. The effect of neonics need to be understood more deeply.
- Let' s check how to use neonics properly to increase the colony number.

# The kendall correlation between neonics application and honey production



# Summary II

- The neonics are applied in USA since 2003 to control the colony collapse disorder (CDD). Then we analyze the correlation between five kinds of neonics and the colony number.
- All the five kinds of neonics exhibit different correlation trend with honeybee colony number; however, allneonic has positive correlation with the colony number at the value of 0.099, which indicates the application of neonic pesticide could promote the honeybee developing.
- Among the neonics, IMIDACLOPRID (corr=0.13) plays a key role in promoting honeybee developing; it also show strongest correlation with allNeonic at 0.87. Thus, Imidacloprid is the most import neonics in promoting honey propagation.
- The second important one is THIAMETHOXAM (corr=0.035). The rest of neonics all affect the honeybee colony negatively.

# What’s next

- We will focus on using supervised learning related method to train predictive models and employ cross validation to evaluate the model’ s metrics to find out the best model.
- Summary the data analysis and present the final report.

# Thanks for your attention!