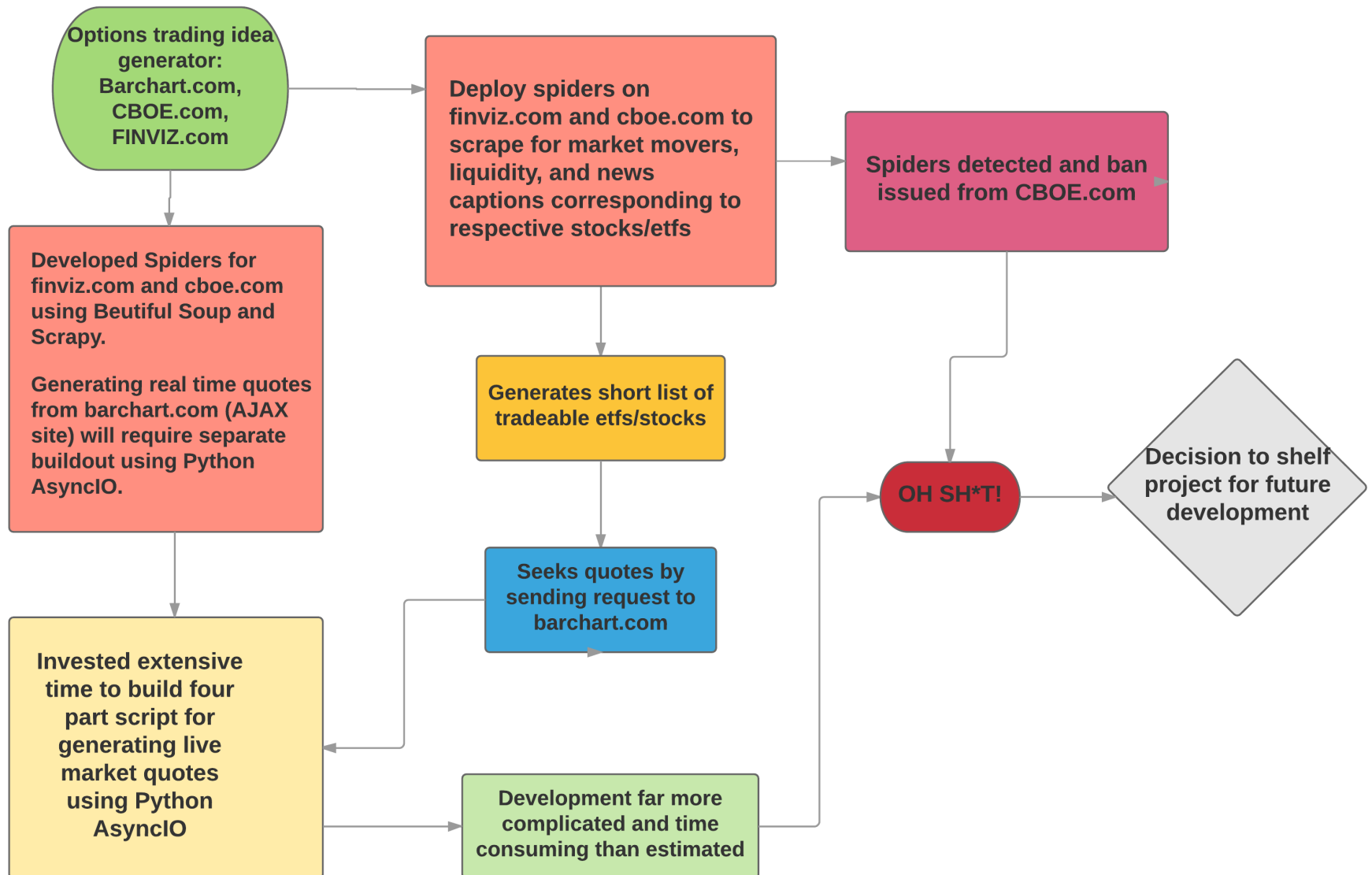


A Tale of Two Projects

“...It was the best of times, it was the worse of times...” - Charles Dickens (*unconfirmed if he was referring to a premonition of Michael Zolla’s NYC Data Science Academy web-scraping project)



What is AJAX and AsyncIO?

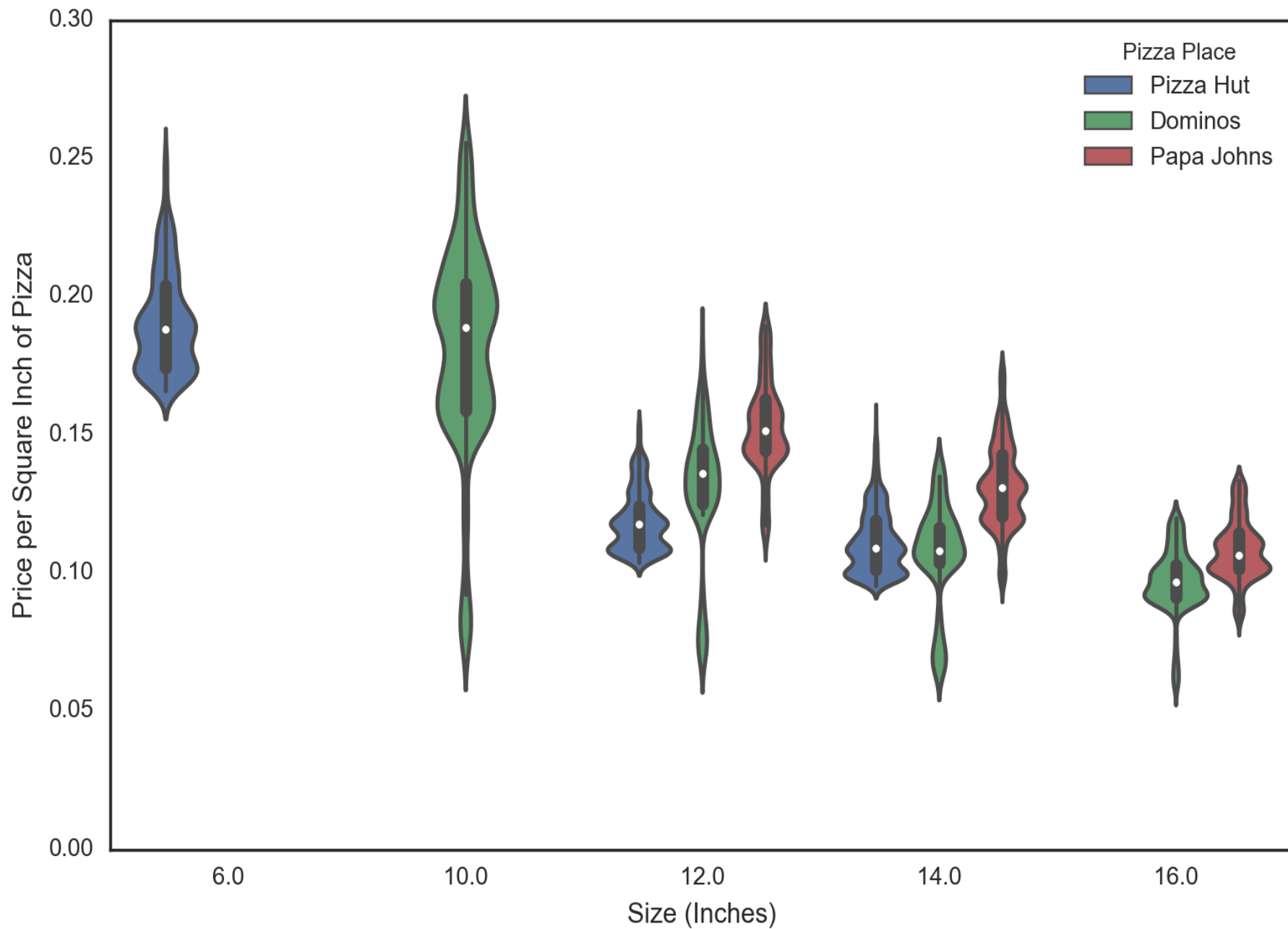
- Set of web dev techniques that improves efficiency during website interaction.
- Allows websites to send requests to the server and updates w/o need to reload the page.
- Improved efficiency comes with a cost of complexity, particularly for scrapers.
- AsyncIO enables ability to perform coroutines and event loops:
 - Coroutine – generator that consumes data, but doesn't generate it
 - Tasks – schedulers for coroutines
 - Event Loop – think of it as the central executor

Navigating setbacks and roadblocks

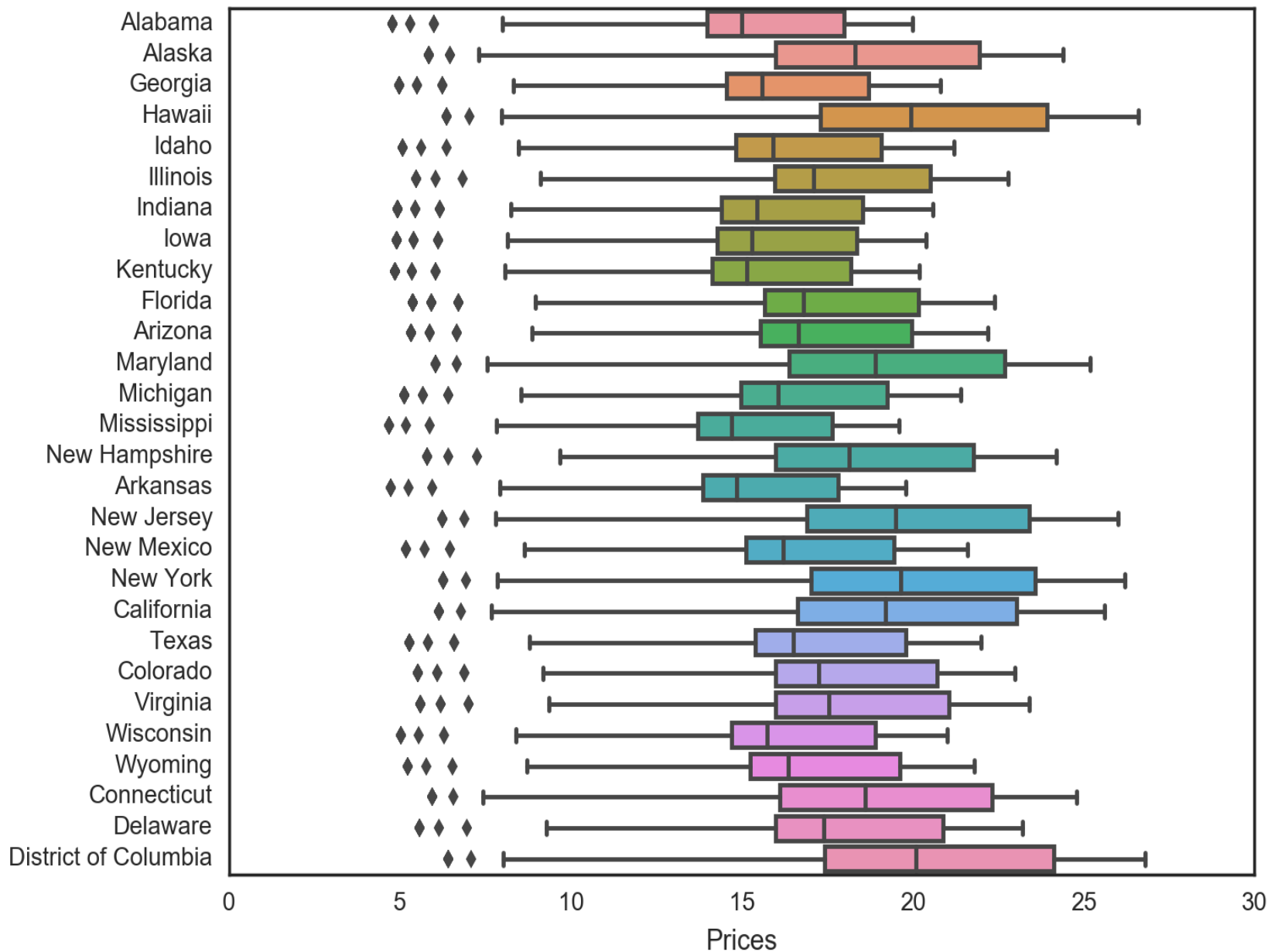
- Project Management – setting ambitious goals vs respecting deadlines
- Murphy's Law
- Aligning priorities with incentives
- Decision analysis
- Analogies to trading and investment experiences

Salvaging the opportunity

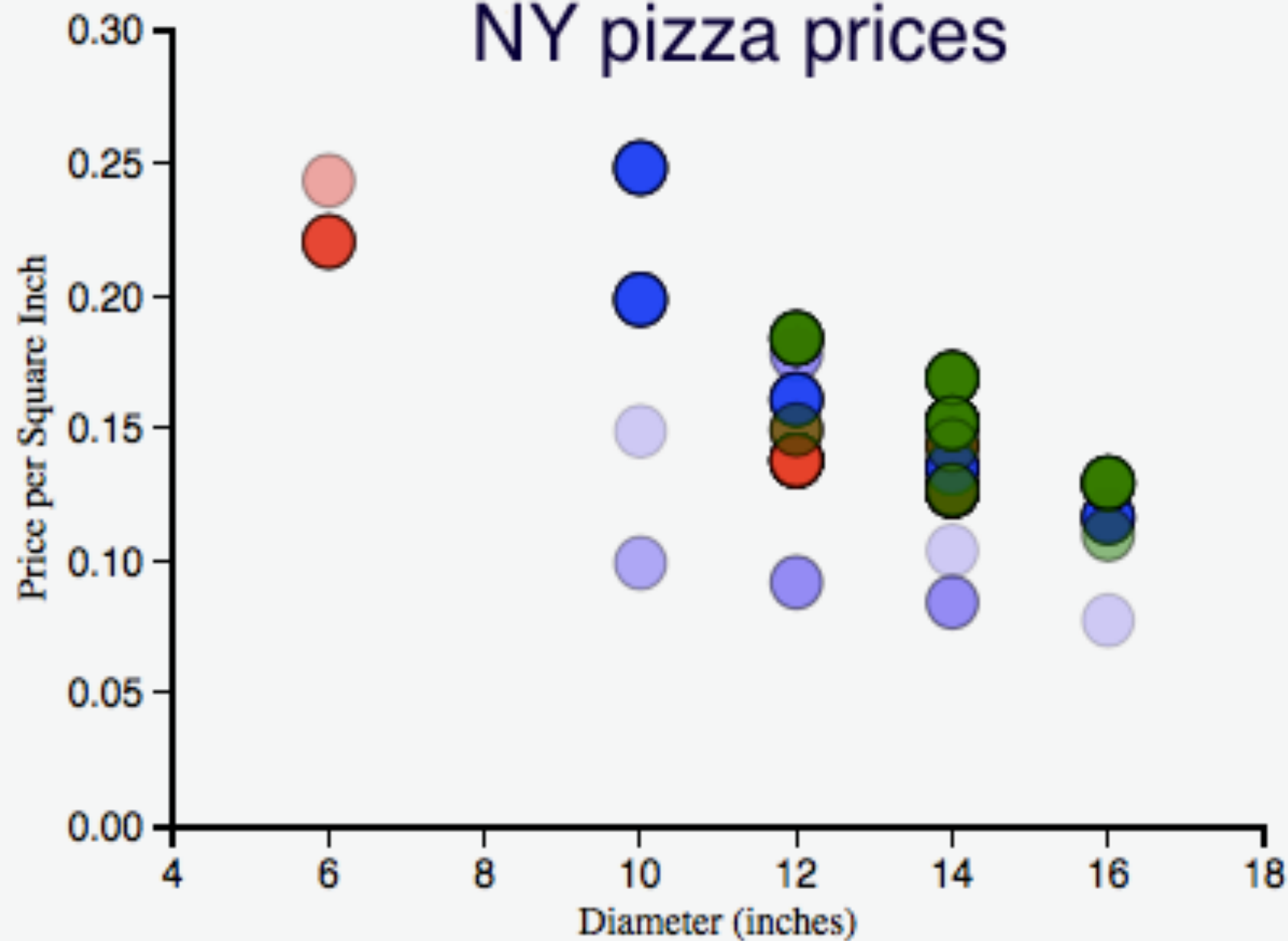
- 11th hour idea: fastfoodmenuprices.com
- Selenium, you sob. Bring it on.
- Entering the rabbit hole: outsourcing insights, pros and cons
- Learning new packages and modules on the fly

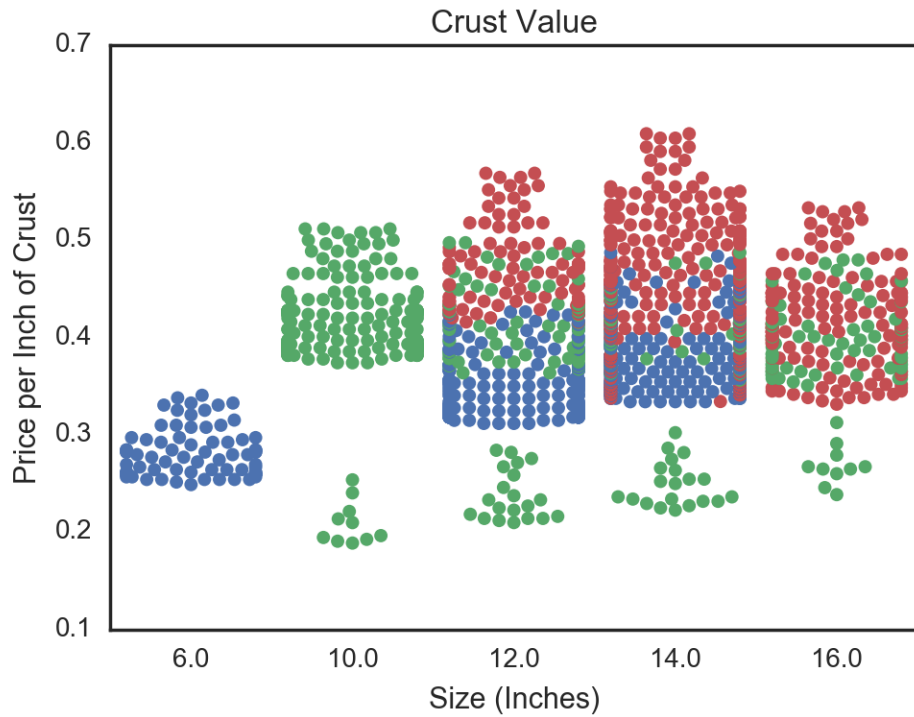
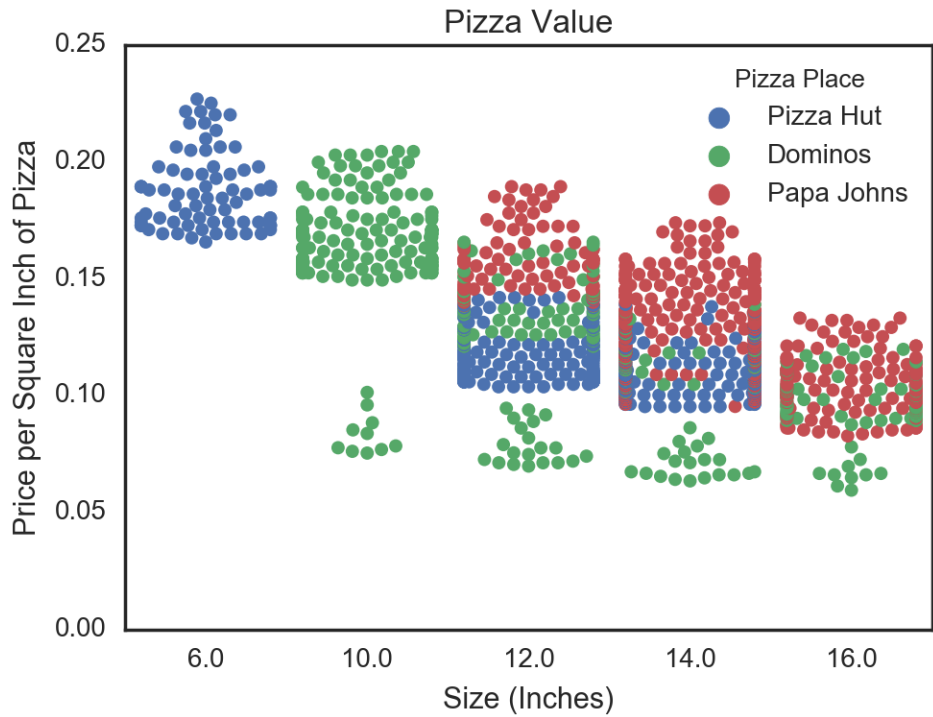


State



NY pizza prices





Key Takeaways

“If you’re not failing, you’re not trying hard enough”

Pro’s

- Successfully scraped websites using BeautifulSoup, Scrapy, Selenium, and AsyncIO
- Learned new Python modules on the fly (Pickle, Seaborn, Matplotlib)
- Discovered more effective learning style, through failures and adjustments

Cons

- Not enough time to generate substantive analysis
- Did not meet self-imposed expectations for finished product
- Budgeted time poorly(should have cut losses earlier), stuck closer to the script