Recap from last week

- Data warehouse (OLAP vs OLTP)
- Difference between DB vs DW
- SQL (relational database)
  - https://www.w3resource.com/sql-exercises/
- BIDM cycle
- Structured vs Unstructured Data
- ETL, Data Pipeline
- Data Engineers vs Data Scientists
- Basic Python syntax
- Make sure you can run Jupyter
- Finished chapter 1, 2, 3, and 17

SQL Query => Simple Aggregation (mean) => Simple Statistics (standard deviation) => Hypothesis Testing => Data Mining => Artificial Intelligence

A New Journey

Now we know something about Data, what's next?

Remember the Goal of Data Analytics is to find patterns

But how?

Remember in the video of Expectation vs Reality on Correlation vs Causation

Let's being our journey

Predicting your professor's mood

If I told you, this is my face
on last Thursday night

Can you explain it why?

Find important factor

Really?

How can you prove or disprove your hypothesis?

How to verify your hypothesis?

If I told you, my mood is still sad today,

what does that mean?

Find other factors

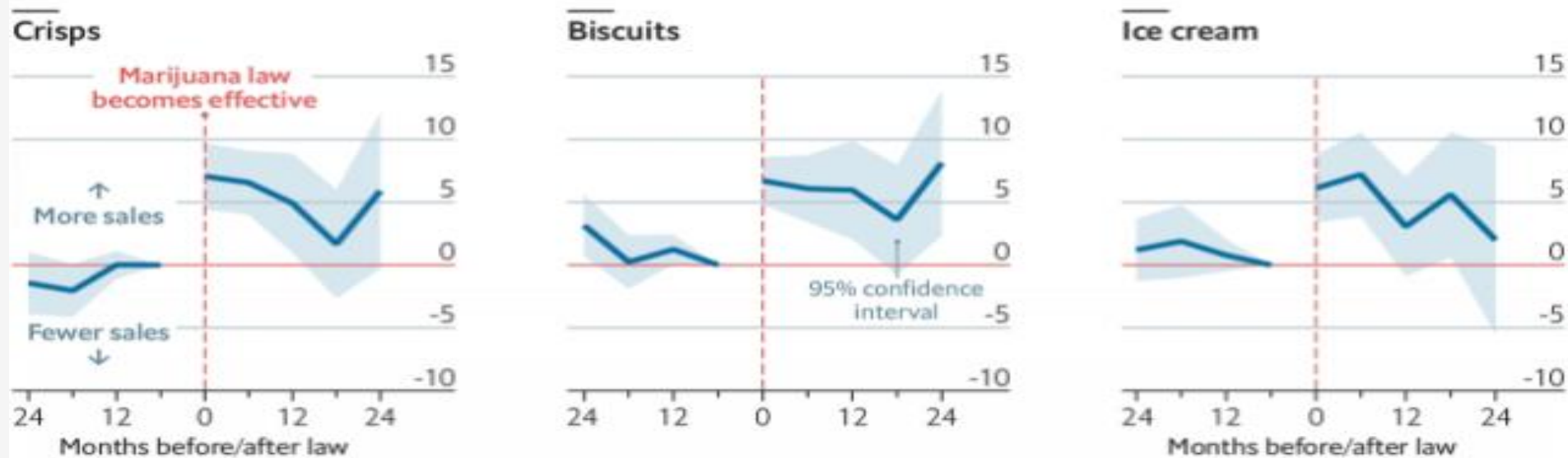Only one has responded to my real-life data analytic articles example so far.

Meaning my extra-credits is cheap ☹

Legal weed is linked to higher junk-food sales

Research suggests marijuana really does give you the munchies

Hey hey we're the munchies
United States, effect of recreational marijuana laws on junk-food sales, %

Source: "Recreational Marijuana Laws and Junk Food Consumption: Evidence Using Border Analysis and Retail Sales Data", by Michele Baggio and Alberto Chong, working paper (2019)

Queens neighborhoods saw drop in home sale prices in August

https://qns.com/story/2019/09/06/the-price-is-wrong-these-queens-neighborhoods-saw-drop-in-home-sale-prices-in-august/
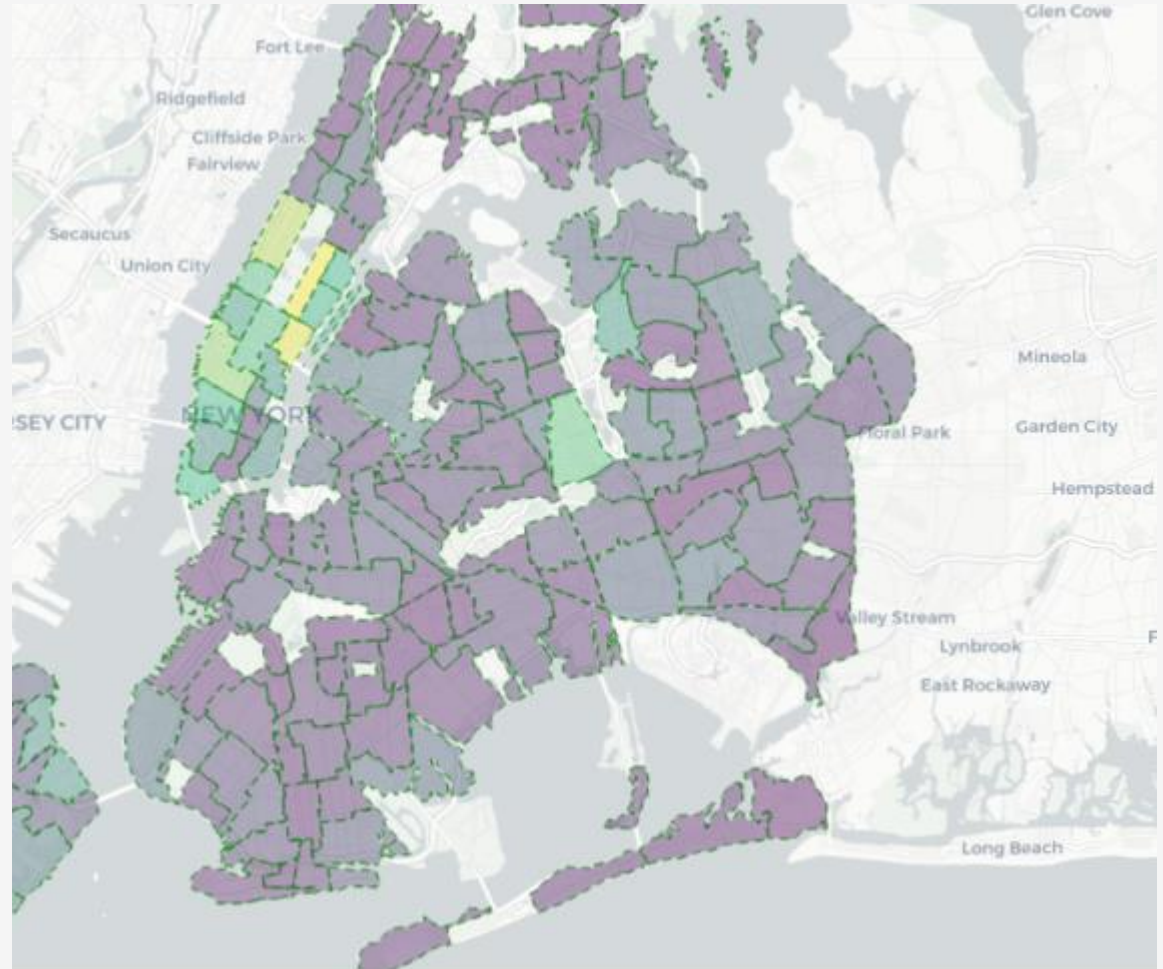
Test your hypothesis

Test the hypothesis by posting interesting news articles
And ask how do I feel next week

Data Analytics Research Process

Define your question or problem you want to solve

=> Make Observations, Collect Data

⇒ Identify possible important factors (Features, Attributes)

⇒ Test whether the factors are important or not

⇒ Continue to find important factors

⇒ Until you feel you think you got it!