Defending Your Data

FREDERICK THAYER



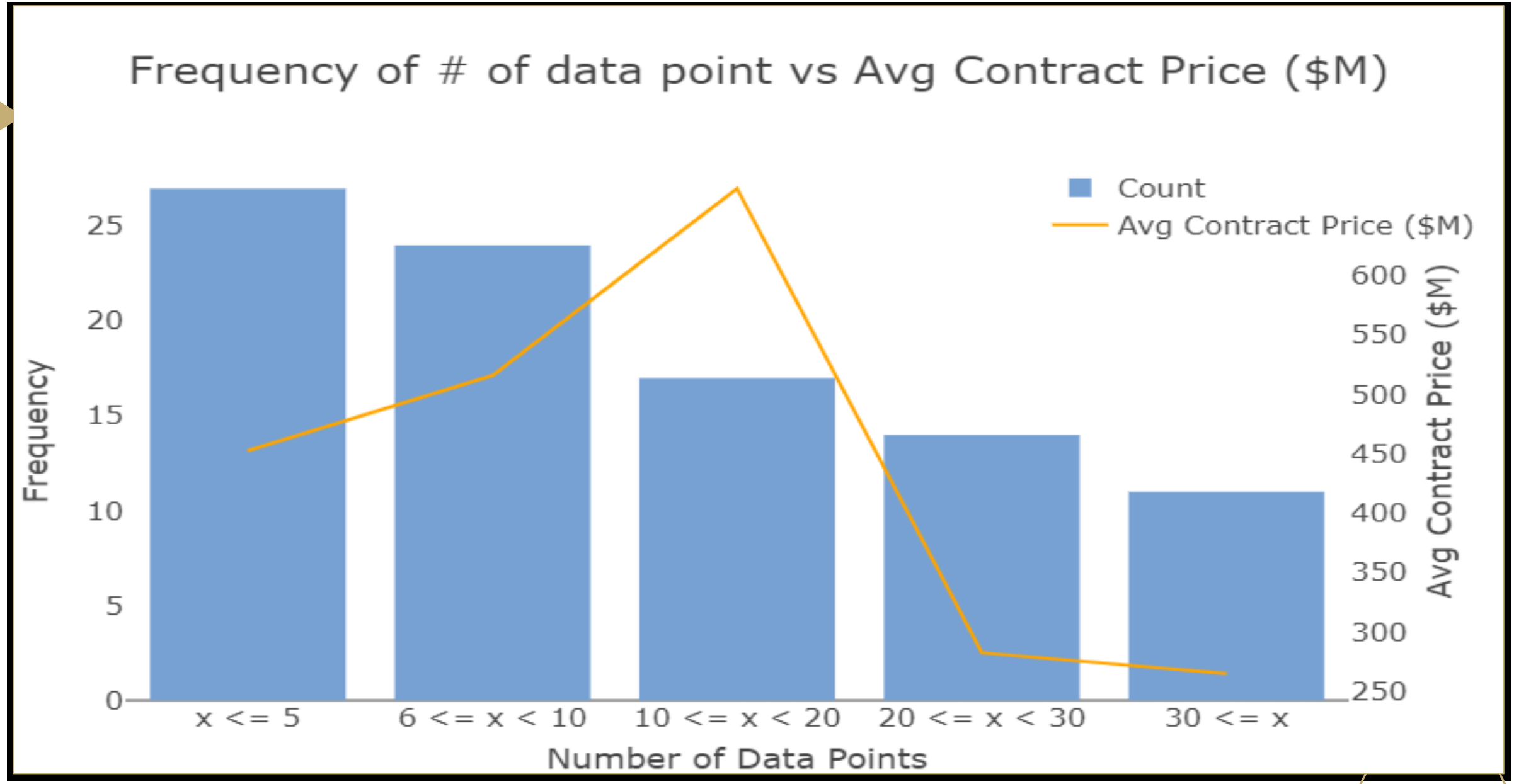


Who Am I and What Do I Do?

- Frederick Thayer
- Data Scientist at the Naval Aviation (NAVAIR) Proposal Analysis Team (PAT)
  - Government team
  - Part of the NAVAIR Procurement Group
  - Modernizing NAVAIR Acquisition

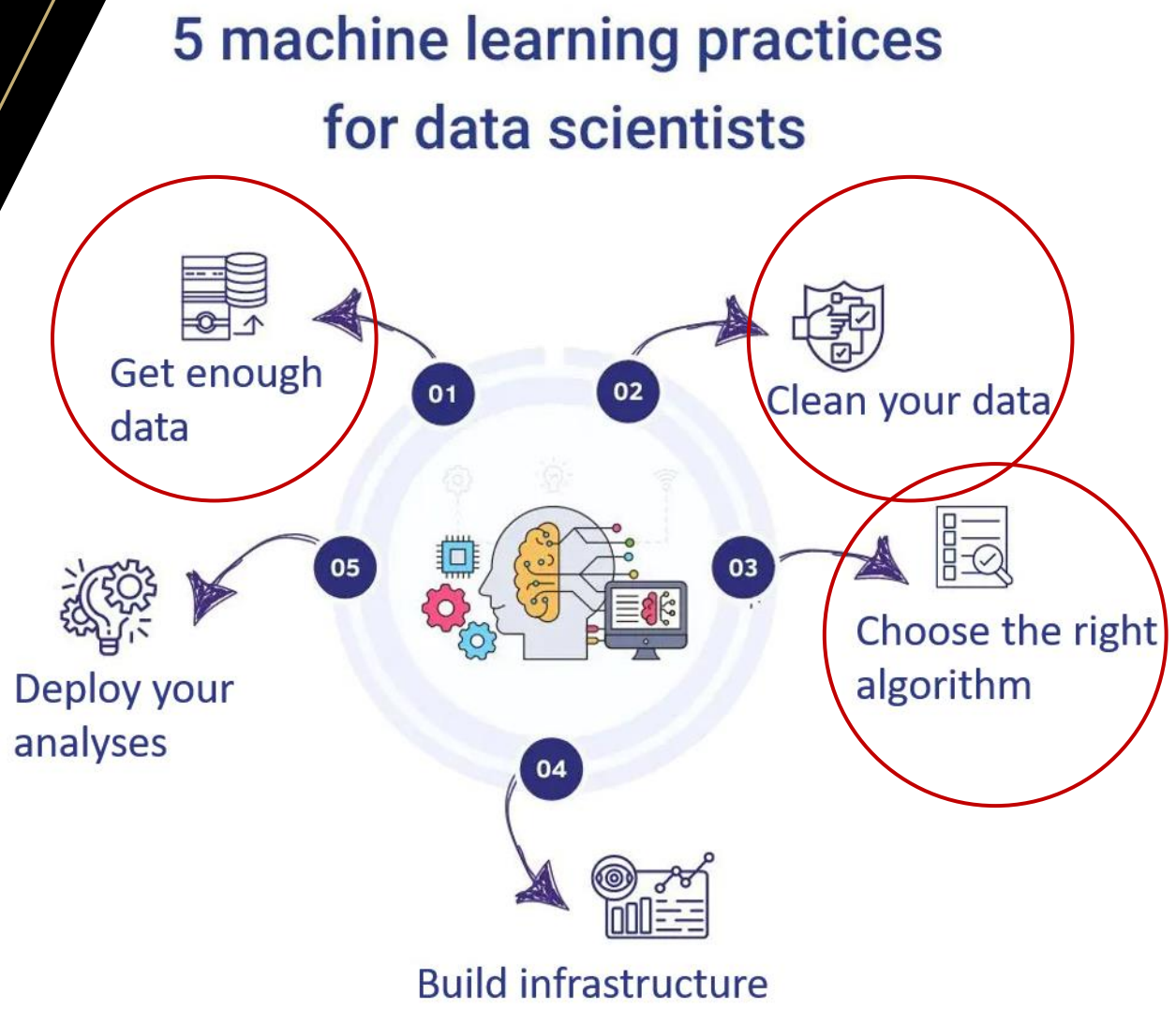
Many such examples

- $280M on 6 data points
- $930M on 7 data points
- $1.2B on 5 data points


Frequency of # of data point vs Avg Contract Price ($M)
Count
Avg Contract Price ($M)
Frequency
Avg Contract Price ($M)
25
20
15
10
5
0
600
550
500
450
400
350
300
250
x <= 5
6 <= x < 10
10 <= x < 20
20 <= x < 30
30 <= x
Number of Data Points

The world isn’t going to wait for you to get enough data.

Best Practices



Get Enough Data

- Be an expert on the data you have
  - Where does the data come from?
  - Why is it applicable?
  - What structures are in place for it?





Clean Your Data

- Know the context of your analysis
  - What other data is there?
  - Who is the intended audience?
  - When is your analysis needed by?
- Document Assumptions

Choose the Right Algorithm

- Understand the intent of your modeling
  - What real world effect are you trying to model?
  - Why are the results what they are?
  - How are they being used and interpreted?



If the data doesn't speak for itself, you need to speak for it.



Thank You!

- Frederick.Thayer@icloud.com