**Team:** Where’s my data dude

**Members:** Ricky Kharawala, Paul Rizzuto & Craig Bauer

**Overarching framework to solve the problem:**

Celebrity books are among the most difficult to forecast in Consumer Trade Book Publishing. Beyond consumers purchasing a forecastable mix of digital books and print books, the demand volume and life cycle is less predictable leading to lost sales when out of stock and excess inventory when the forecast bias was stronger than actual market demand.

The hypothesis is to determine if social media sentiment and volume can be an indicator of demand and does it lead to visible results in point-of-sale (POS) results.

Thinking through the hypothesis we discussed, we can evaluate SOURCE (celebrity originated activity) but focus on NON-SOURCE (everyone but the celebrity) twitter activity to illustrate:

1. Does the volume of social media activity influence POS demand?
2. Does social media sentiment influence POS demand?
3. Is there a visible lag period between twitter traffic and POS results?
4. Does a sustained campaign create a longer period of demand?
5. Do New York Times Book Reviews influence demand <=> social sentiment?

**Links to initial datasets you’ve found (with brief descriptions on the data structure):**

Nielson BookScan POS Data

Twitter API

NY Times API for book review data

Python 2.0 and Python 3.7; Pandas

Vader for sentiment analysis

Importing CSV files for POS, accumulating Twitter data based on Title and aligning periods to weekly POS.