

Data Engineering: How to Thrive

James Prescott



Hi, I'm James!

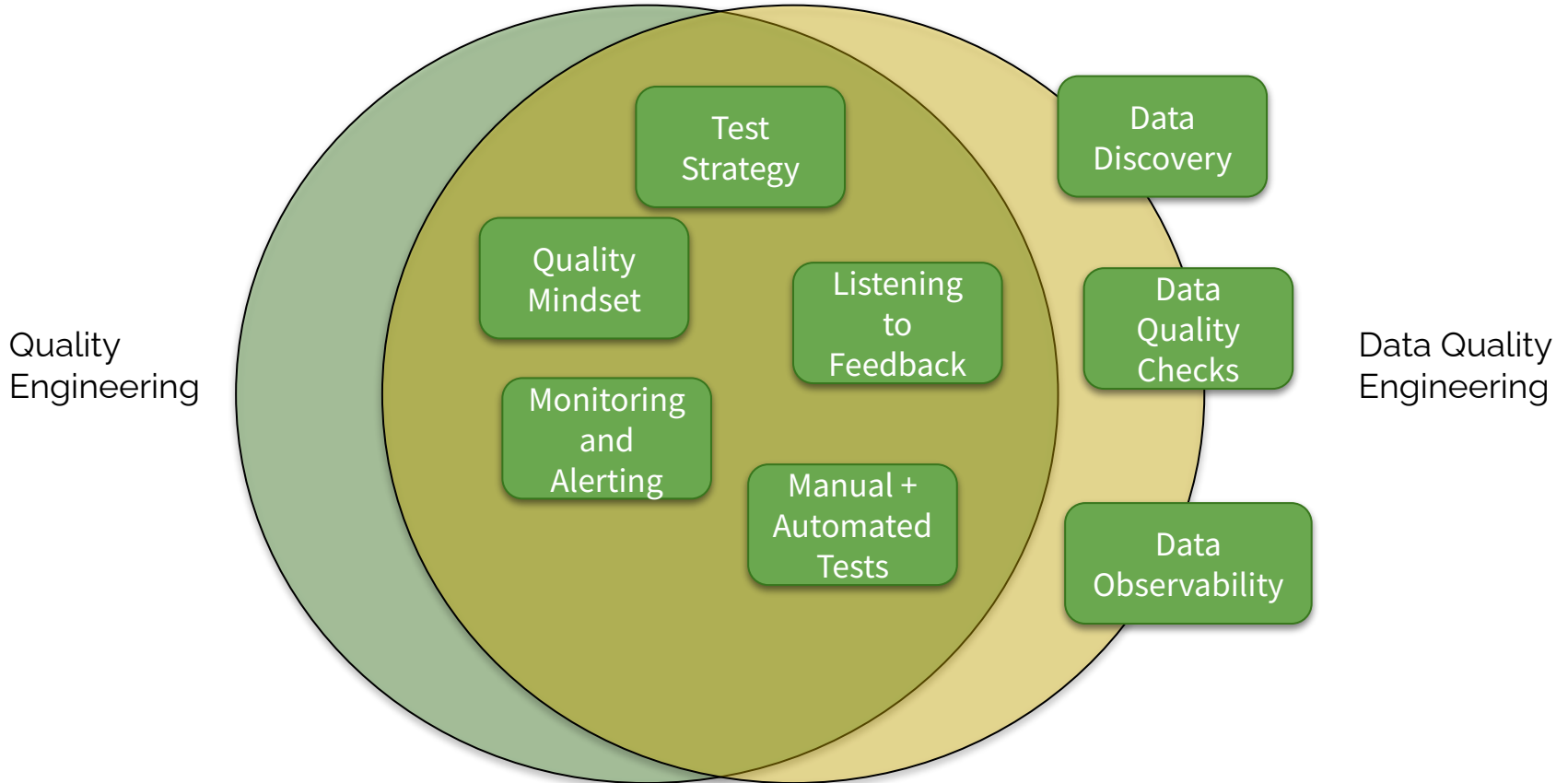


- 13 years of Quality Engineering experience
- Transitioned to Data Engineering in 2020
- Part of QE leadership since 2021
- Loves Video Games, Hiking and going to cool Conventions!

What is Data Engineering?

- NOT DATA SCIENCE
 - Is a foundation used to gather insights and conduct analysis
 - Makes sense of normally messy data given by the business
-

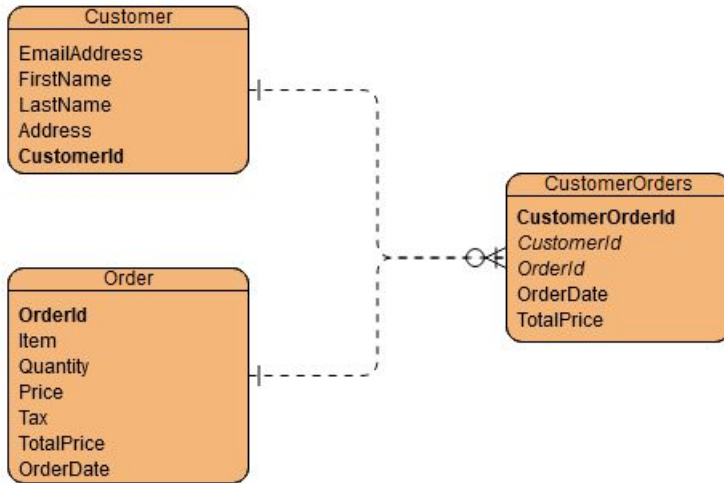
How Does Quality Fit In Here?



What is Data Quality?

How much does your data represent the world you want the business to understand?

Consider This...

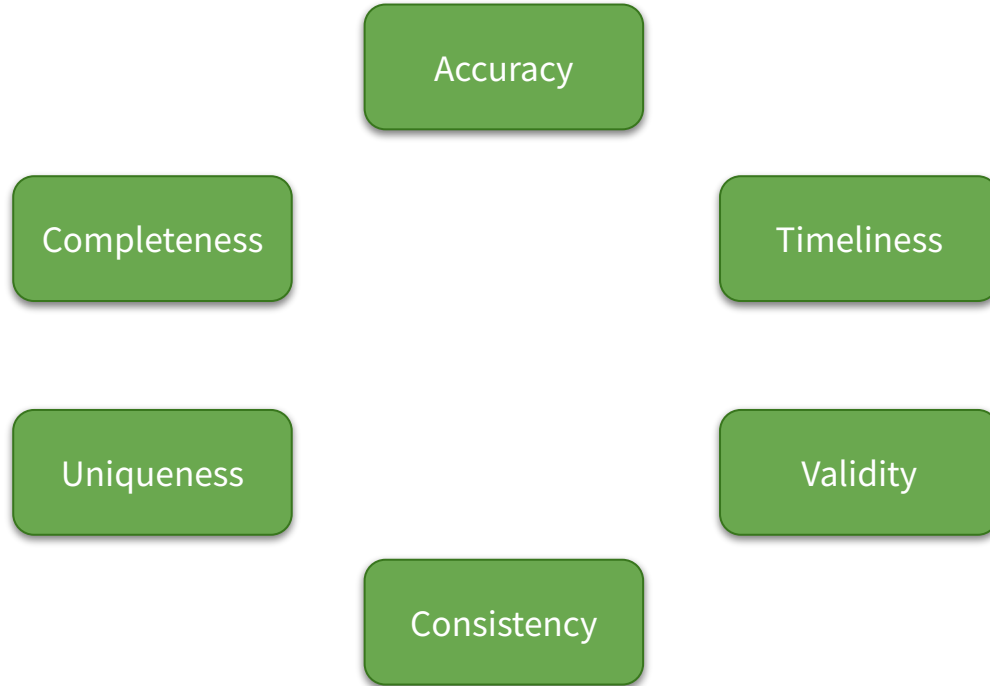


We have a very simple set of data to the left of us. There are MANY questions we can ask of this from a Data Quality perspective:

- Is the email address valid?
- Are the IDs truly unique?
- Is the OrderDate in the past? Should it be?
- What format does the Address come in?
- What currency are we talking about here?

Etc etc etc...

How Do We Measure Data Quality?



How Do We Measure Data Quality?

Tests + Checks

Data Discovery

Manual Testing

Automated DQ Checks

Observability

Monitoring
and Alerting

Data Quality
Dashboarding

Where do you even Start?



The Challenges

You're not in control!

You need to know the business

It's a niche area - Data/BI Engineers might not get it

Tips

- Shifting efforts both left and right is really important
- Don't assume software testing methods will work here
- Data is diverse and messy, so what ifs are very real!
- Pair not just with engineers, but also the business - understand their processes

Questions?