



# Making Agile Work for Data Science

Erika Salomon

Agile?! Yuck!

Agile?! Yuck!

*Endless, pointless meetings*

# Agile?! Yuck!

*Endless, pointless meetings*

*Micromanagement*

# Agile?! Yuck!

*Endless, pointless meetings*

Chaotic, unsupportive environment

*Micromanagement*

# Agile?! Yuck!

Endless, pointless meetings

Chaotic, unsupportive environment

Micromanagement

Process over product

We are uncovering better ways of developing software by doing it and helping others do it.

Through this work we have come to value:

**Individuals and interactions** over processes and tools

**Working software** over comprehensive documentation

**Customer collaboration** over contract negotiation

**Responding to change** over following a plan

That is, while there is value in the items on the right, we value the items on the left more.

We are uncovering better ways of **helping humans make data-informed decisions** by doing it and helping others do it.

Through this work we have come to value:

Individuals and interactions over processes and tools

**Useful information** over comprehensive information

Customer collaboration over contract negotiation

Responding to change over following a plan

That is, while there is value in the items on the right, we value the items on the left more.



# Principles for Agile Data Science

- Focus on delivery
- Psychological safety
- Making work visible
- Ruthless prioritization
- Self-organizing teams

# Principles for Agile Data Science

- **Focus on delivery**
- Psychological safety
- Making work visible
- Ruthless prioritization
- Self-organizing teams

# Principles for Agile Data Science

- Focus on delivery
- **Psychological safety**
- Making work visible
- Ruthless prioritization
- Self-organizing teams

# Two Kinds of Discomfort at Work

- Harassment
- Discrimination
- Bullying
- Taking risks
- Working outside your expertise
- Trying new things

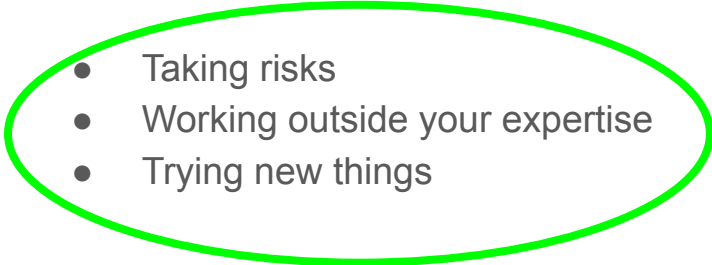
# Two Kinds of Discomfort at Work

- Harassment
  - Discrimination
  - Bullying
- 

- Taking risks
- Working outside your expertise
- Trying new things

# Two Kinds of Discomfort at Work

- Harassment
  - Discrimination
  - Bullying
- 

- Taking risks
  - Working outside your expertise
  - Trying new things
- 

# Two Kinds of Discomfort at Work

- Harassment
- Discrimination
- Bullying

- Taking risks
- Working outside your expertise
- Trying new things

1. Provide direct, private feedback

# Two Kinds of Discomfort at Work

- Harassment
- Discrimination
- Bullying

- Taking risks
- Working outside your expertise
- Trying new things

1. Provide direct, private feedback
2. When people refuse to grow, fire them



# Two Kinds of Discomfort at Work

- Harassment
- Discrimination
- Bullying

1. Provide direct, private feedback
2. When people refuse to grow, fire them

- Taking risks
- Working outside your expertise
- Trying new things

1. Talk about your own fears, uncertainties, mistakes, and lack of knowledge

# Two Kinds of Discomfort at Work

- Harassment
- Discrimination
- Bullying

1. Provide direct, private feedback
2. When people refuse to grow, fire them

- Taking risks
- Working outside your expertise
- Trying new things

1. Talk about your own fears, uncertainties, mistakes, and lack of knowledge
2. Reward people for taking risks

# Two Kinds of Discomfort at Work

- Harassment
- Discrimination
- Bullying

1. Provide direct, private feedback
2. When people refuse to grow, fire them

- Taking risks
- Working outside your expertise
- Trying new things

1. Talk about your own fears, uncertainties, mistakes, and lack of knowledge
2. Reward people for taking risks

# Two Kinds of Discomfort at Work

- Harassment
- Discrimination
- Bullying

1. Provide direct, private feedback
2. When people refuse to grow, fire them

- Taking risks
- Working outside your expertise
- Trying new things

1. Talk about your own fears, uncertainties, mistakes, and lack of knowledge
2. Reward people for taking risks
3. Encourage respectful critical feedback

# Principles for Agile Data Science

- Focus on delivery
- Psychological safety
- **Making work visible**
- Ruthless prioritization
- Self-organizing teams

## Do

- Add feature for number of inspections in the last 18 months
- Add inspection score to inspections fact table
- Try 1 month prediction window

## Doing

- Compare performance of ranking baseline and decision tree models
- Create an inspections fact table

## Done

- Build decision tree models with 3 features
- Build features for the number of safety violations and the number of health violations
- Add violation type to violations table
- Rank buildings by number of violations in the last 18 months
- Build feature of number of violations every 18 months
- Create a violations fact table
- Create a buildings dimension table

# Clean Ambulance Data

Create a clean version of the  
ambulance\_runs table

# Clean Ambulance Data

Create a new version of  
ambulance table





## Clean Ambulance Data

Create a new version of  
ambulance table



## Create Ambulance Runs Fact Table

### Why?

We believe that the number of ambulance runs someone has had will be an important feature. We need a convenient place to calculate that feature from

### What?

Create a table with one row per person per ambulance run with:

- person id
- run id
- run timestamp

# Principles for Agile Data Science

- Focus on delivery
- Psychological safety
- Making work visible
- **Ruthless prioritization**
- Self-organizing teams

## Do

- Try 1 month prediction window
- Add feature for number of inspections in the last 18 months
- Add inspection score to inspections fact table
- Try oversampling
- Try undersampling
- Create a weather fact table
- Add feature for rain within 2 days of inspection
- Create permits fact table
- Add feature for number of permits in the last 5 years

## Doing

- Compare performance of ranking baseline and decision tree models
- Create an inspections fact table

## Done

- Build decision tree models with 3 features
- Build features for the number of safety violations and the number of health violations
- Add violation type to violations table
- Rank buildings by number of violations in the last 18 months
- Build feature of number of violations every 18 months
- Create a violations fact table
- Create a buildings dimension table

## Do

- Try 1 month prediction window
- Add feature for number of inspections in the last 18 months
- Add inspection score to inspections fact table
- ~~Try oversampling~~
- ~~Try undersampling~~
- Create a weather fact table
- Add feature for rain within 2 days of inspection
- Create permits fact table
- Add feature for number of permits in the last 5 years

## Doing

- Compare performance of ranking baseline and decision tree models
- Create an inspections fact table

## Done

- Build decision tree models with 3 features
- Build features for the number of safety violations and the number of health violations
- Add violation type to violations table
- Rank buildings by number of violations in the last 18 months
- Build feature of number of violations every 18 months
- Create a violations fact table
- Create a buildings dimension table

## Do

- Try 1 month prediction window
- Add feature for number of inspections in the last 18 months
- Add inspection score to inspections fact table
- Create a weather fact table
- Add feature for rain within 2 days of inspection
- Create permits fact table
- Add feature for number of permits in the last 5 years
- ~~Try oversampling~~
- ~~Try undersampling~~

## Doing

- Compare performance of ranking baseline and decision tree models
- Create an inspections fact table

## Done

- Build decision tree models with 3 features
- Build features for the number of safety violations and the number of health violations
- Add violation type to violations table
- Rank buildings by number of violations in the last 18 months
- Build feature of number of violations every 18 months
- Create a violations fact table
- Create a buildings dimension table

## Do

- Add feature for number of inspections in the last 18 months
- Add inspection score to inspections fact table
- Create a weather fact table
- Add feature for rain within 2 days of inspection
- Create permits fact table
- Add feature for number of permits in the last 5 years
- Try 1 month prediction window
- ~~Try oversampling~~
- ~~Try undersampling~~

## Doing

- Compare performance of ranking baseline and decision tree models
- Create an inspections fact table

## Done

- Build decision tree models with 3 features
- Build features for the number of safety violations and the number of health violations
- Add violation type to violations table
- Rank buildings by number of violations in the last 18 months
- Build feature of number of violations every 18 months
- Create a violations fact table
- Create a buildings dimension table

# Principles for Agile Data Science

- Focus on delivery
- Psychological safety
- Making work visible
- Ruthless prioritization
- **Self-organizing teams**

- How big should a ticket be?
- Do we need standup? Does it have to be daily? Does it have to be synchronous?
- How many columns should our Kanban have?
- Which Kanban tool should we use?



We are uncovering better ways of **helping humans make data-informed decisions** by doing it and helping others do it.

Through this work we have come to value:

**Individuals and interactions** over processes and tools

**Useful information** over comprehensive information

**Customer collaboration** over contract negotiation

**Responding to change** over following a plan

That is, while there is value in the items on the right, we value the items on the left more.

