Brand Tracking with Bayesian Models and Metaflow

Corrie Bartelheimer
Senior Data Scientist @ Latana

Berlin Bayesian, 10/11/2021

How many people have heard of our brand?

TIER

headspace

monday.com

Blinkist

duolingo

N26

EPIC GAMES

Latana

How well known is the brand in target group?

How do people perceive the brand?

Have there been changes?

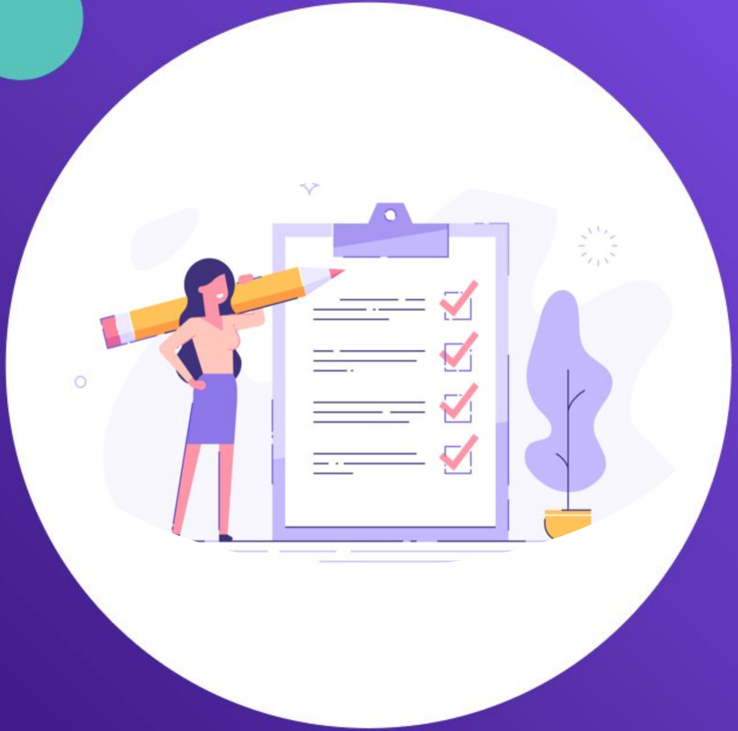
Latana

What makes Brand Tracking difficult?

Latana

Survey Problems

- Small target groups
- Signal or Noise?
- Representativeness of respondents



Latana

Traditional Approaches

Latana

Weighting

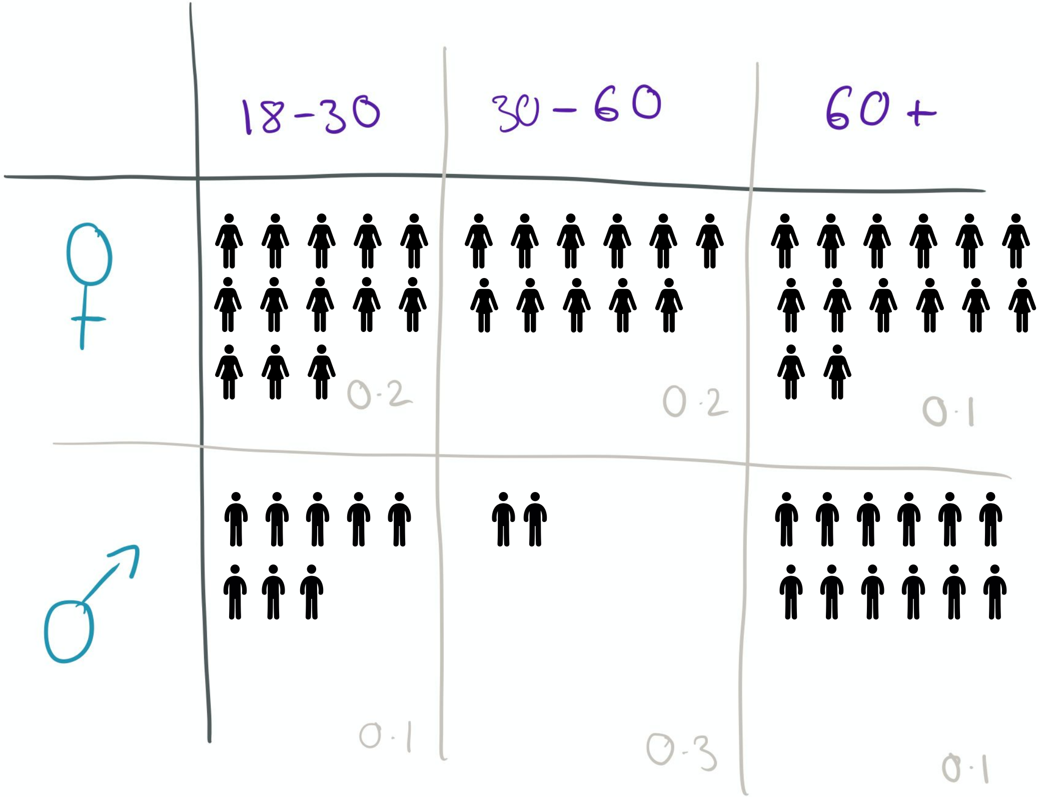
	18-30	30-60	60+
♀			
♂			

Latana

Weighting

	18-30	30-60	60+
♀	0.2	0.2	0.1
♂	0.1	0.3	0.1

Weighting



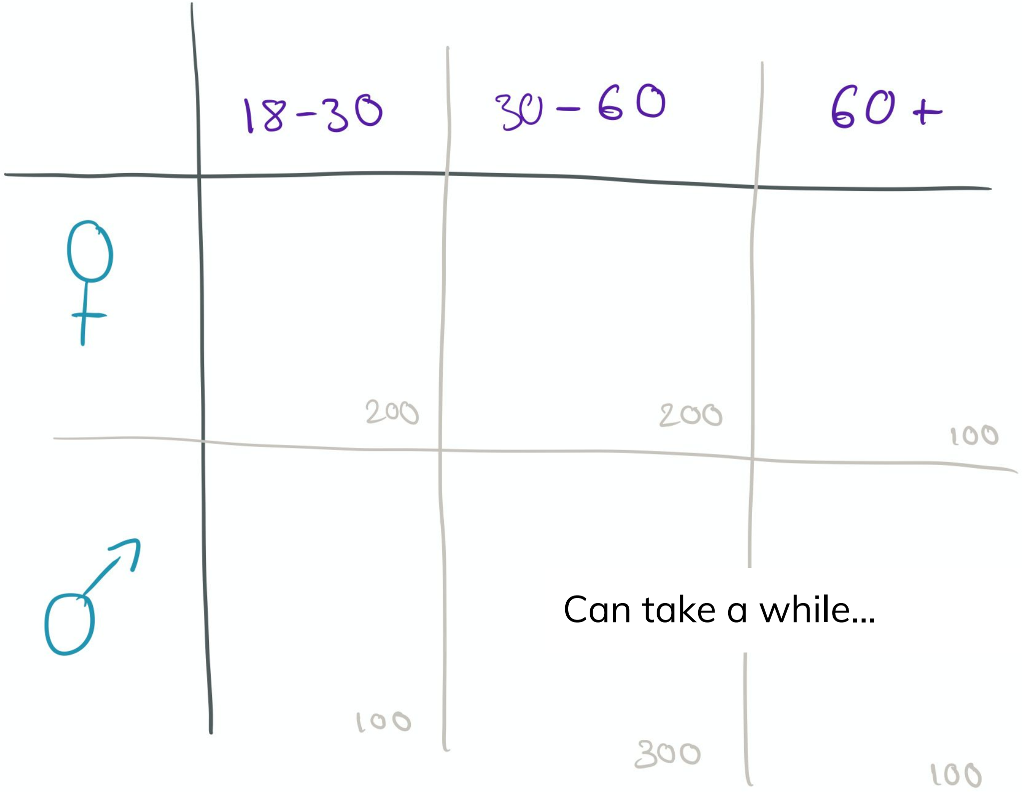
Latana

Quota Sampling

	18-30	30-60	60+
♀	200	200	100
♂	100	300	100

Latana

Quota Sampling



Can take a while...

Introducing: Mr. P
Multilevel Regression & Poststratification

Latana

Multilevel Regression

Multilevel Regression

knows brand $\sim$ Bernoulli(p)

Latana

Multilevel Regression

$$\text{knows brand} \sim \text{Bernoulli}(p)$$

$$\text{logit}(p) = \alpha_{[\text{gender}]} + \beta_{[\text{age}]}$$

Latana

Multilevel Regression

$$\text{knows brand} \sim \text{Bernoulli}(p)$$

$$\text{logit}(p) = \alpha_{[gender]} + \beta_{[age]}$$

$$\alpha_{[gender]} \sim \text{Normal}(\mu_\alpha, \sigma_\alpha)$$

$$\beta_{[age]} \sim \text{Normal}(\mu_\beta, \sigma_\beta)$$

Latana

Multilevel Regression

$$\text{knows brand} \sim \text{Bernoulli}(p)$$

$$\text{logit}(p) = \alpha_{[gender]} + \beta_{[age]}$$

$$\alpha_{[gender]} \sim \text{Normal}(\mu_\alpha, \sigma_\alpha)$$

$$\beta_{[age]} \sim \text{Normal}(\mu_\beta, \sigma_\beta)$$

$$\mu_\alpha, \mu_\beta \sim \text{Normal}(0, 1)$$

$$\sigma_\alpha, \sigma_\beta \sim \text{Normal}_+(0, 1)$$

Latana

Multilevel Regression

$$\text{knows brand} \sim \text{Bernoulli}(p)$$

$$\text{logit}(p) = \alpha_{[gender]} + \beta_{[age]}$$

Multilevel

$$\alpha_{[gender]} \sim \text{Normal}(\mu_\alpha, \sigma_\alpha)$$

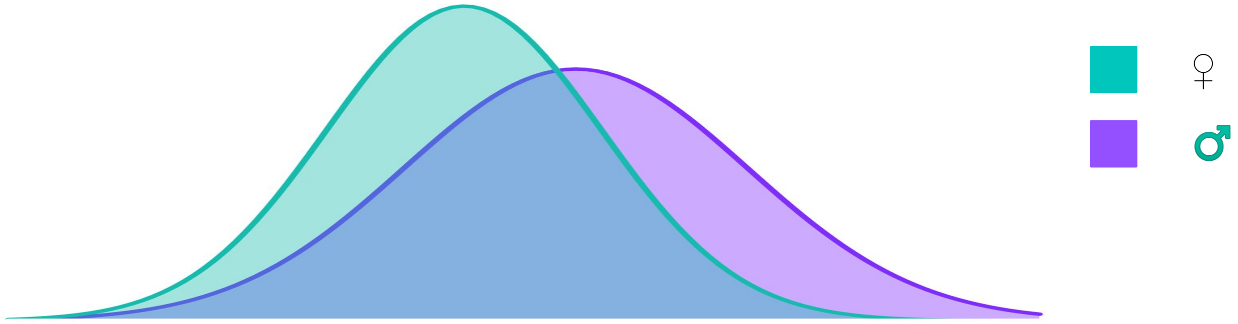
$$\beta_{[age]} \sim \text{Normal}(\mu_\beta, \sigma_\beta)$$

$$\mu_\alpha, \mu_\beta \sim \text{Normal}(0, 1)$$

$$\sigma_\alpha, \sigma_\beta \sim \text{Normal}_+(0, 1)$$

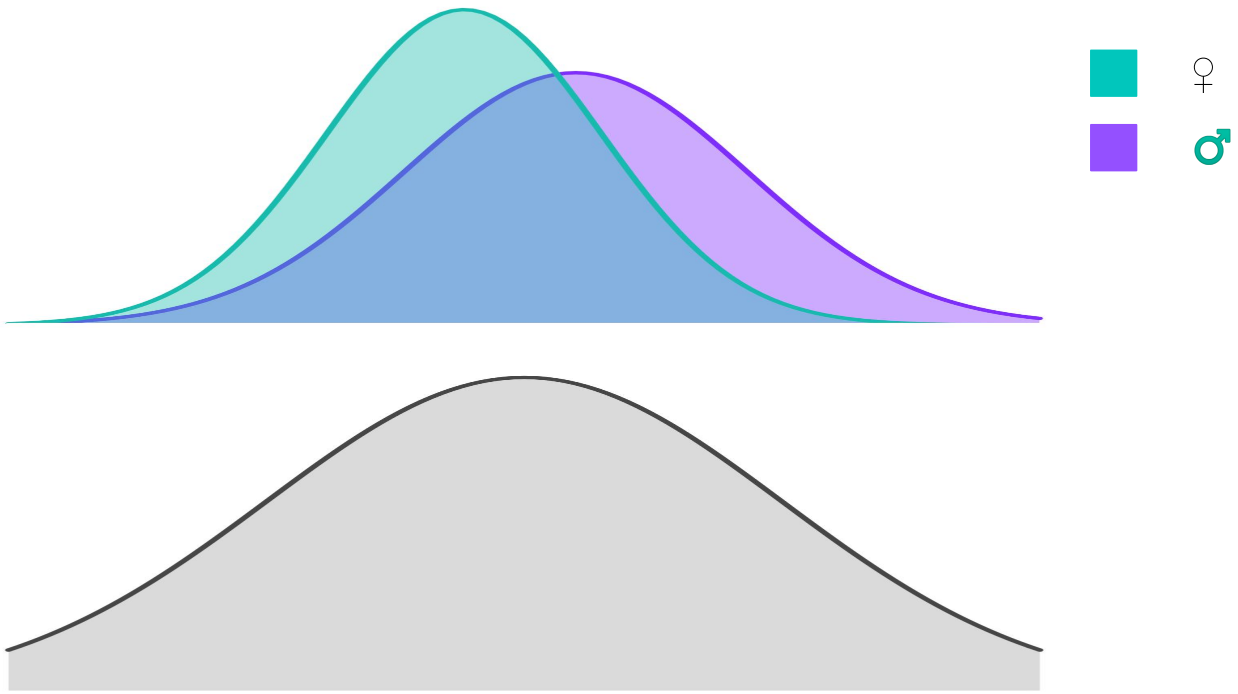
Latana

Multilevel Regression

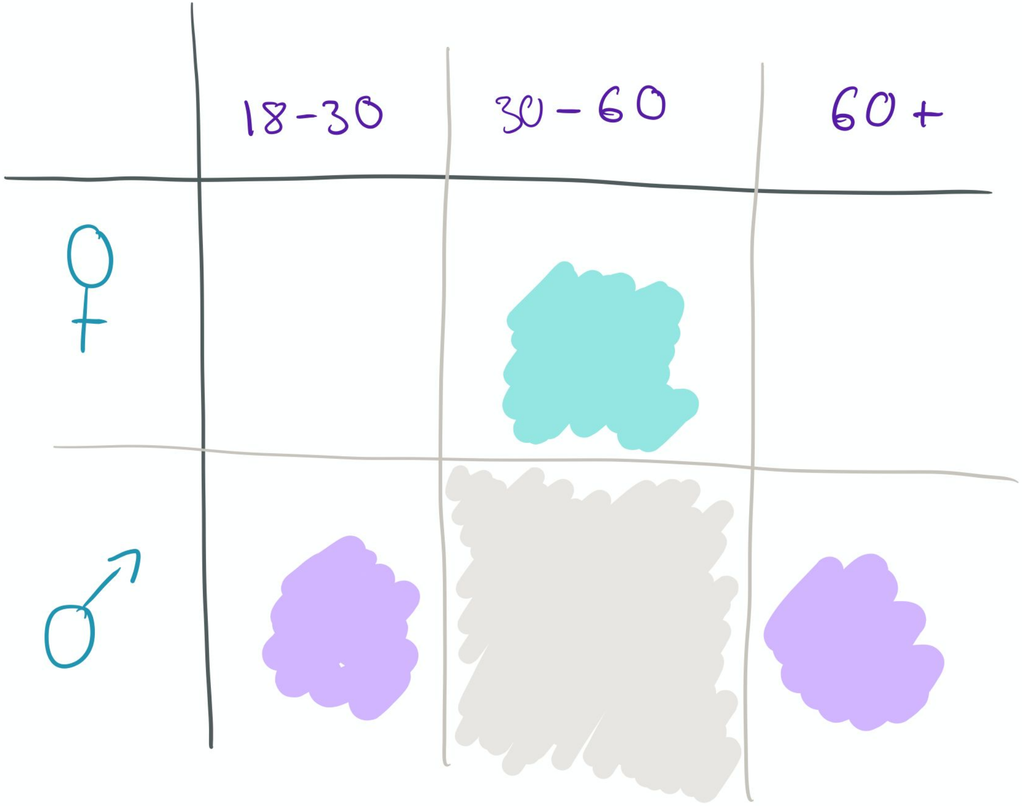


Latana

Multilevel Regression

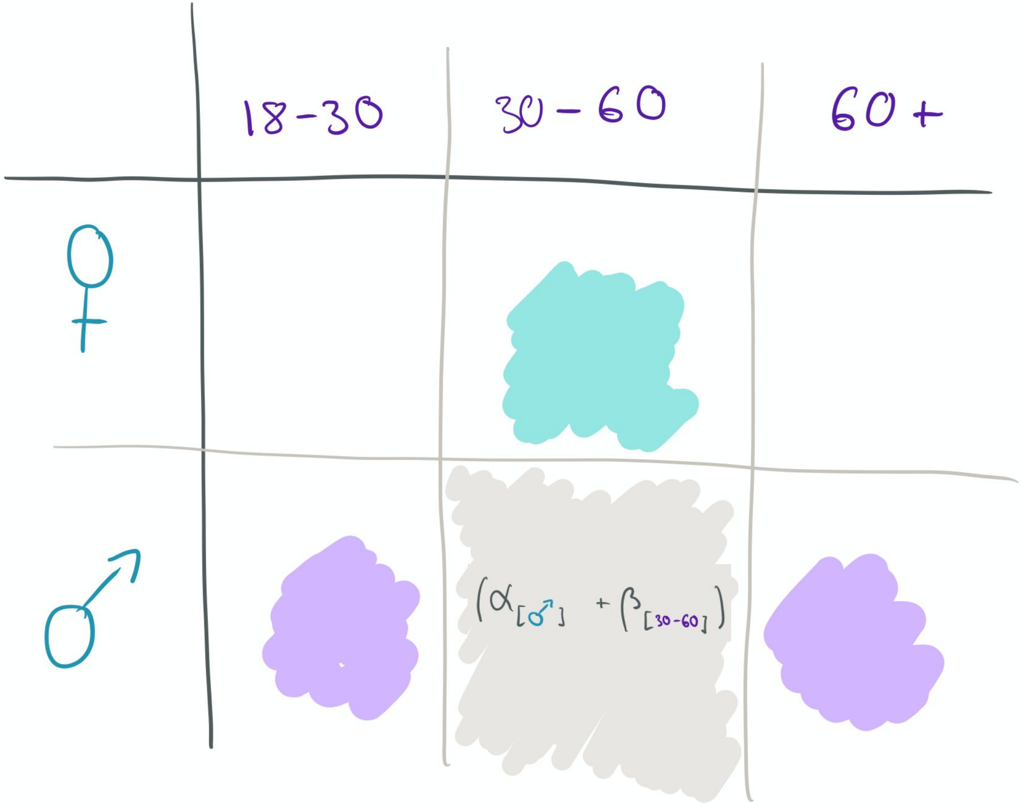


Multilevel Regression



Latana

Multilevel Regression



Latana

Poststratification

Poststratification

	18-30	30-60	60+
♀	0.2	0.2	0.1
♂	0.1	0.3	0.1

Poststratification

Total proportion of people that know brand = 0.5 × Proportion of men that know brand + 0.5 × Proportion of women that know brand

Weight

Prediction from our Model

Poststratification

Total proportion of people that know brand $= 0.5 \times$ Proportion of men that know brand $+ 0.5 \times$ Proportion of women that know brand

P(gender)

P(knows brand | gender)

Poststratification

$$\begin{pmatrix} 0.75 \\ 0.71 \\ \vdots \\ 0.68 \end{pmatrix} = 0.5 \times \begin{pmatrix} 0.63 \\ 0.68 \\ \vdots \\ 0.71 \end{pmatrix} + 0.5 \times \begin{pmatrix} 0.88 \\ 0.75 \\ \vdots \\ 0.64 \end{pmatrix}$$

On scale: Metaflow

Latana

What do you associate with this brand?

duolingo

- [x] Affordable
- [ ] Effective
- [x] Fun
- [ ] Innovative
- [ ] User-friendly
- [ ] None of these

What do you associate with this brand?

duolingo

- [x] Affordable
- [ ] Effective
- [x] Fun
- [ ] Innovative
- [ ] User-friendly
- [ ] None of these

One model per answer option

What do you associate with this brand?

duolingo

- [x] Affordable
- [ ] Effective
- [x] Fun
- [ ] Innovative
- [ ] User-friendly
- [ ] None of these

One model per answer option

One question per brand, different competitor brands per market

~500 - 1500 models per project

Latana

What do you associate with this brand?

duolingo

- [x] Affordable
- [ ] Effective
- [x] Fun
- [ ] Innovative
- [ ] User-friendly
- [ ] None of these

One model per answer option

One question per brand, different competitor brands per market

~500 - 1500 models per project

~20min per model

= ~10 days compute time

METAFLOW

- Integrates with AWS Batch
- Easy to use for Data Scientists
- Supports reproducibility

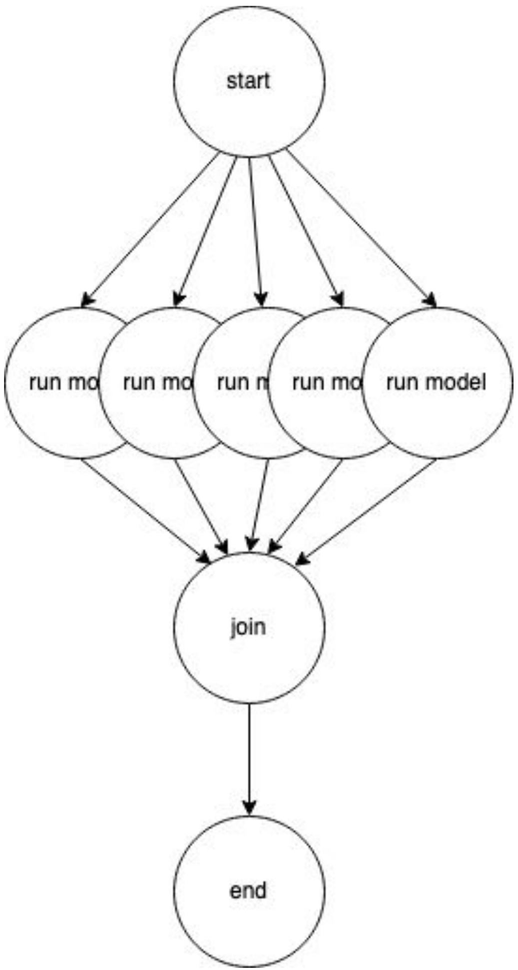
MRP as Metaflow

```
from metaflow import FlowSpec,step

class MRPFlow(FlowSpec):

    @step
    def start(self):
        Self.data, self.questions = load_data()
        self.next(self.run_model, foreach="questions")

    @step
    def run_model(self):
        question = self.input
        self.result = run_mrp(question, self.data)
        self.next

    @step
    def join(self, inputs):
        for result in inputs:
            save(result)

        self.next(self.end)

    @step
    def end(self):
        pass
```

Latana

MRP as Metaflow

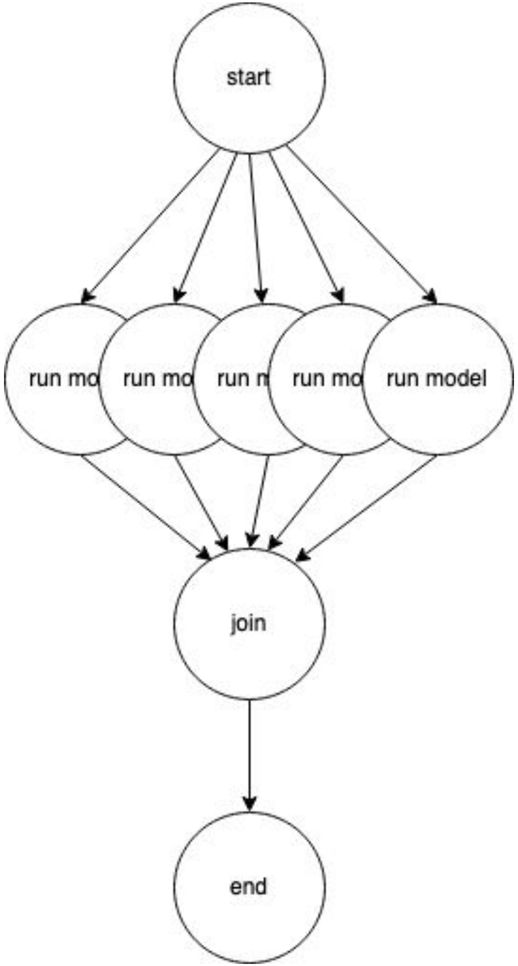
```python
from metaflow import FlowSpec,step

class MRPFlow(FlowSpec):

    @step
    def start(self):
        Self.data, self.questions = load_data()
        self.next(self.run_model, foreach="questions")

    @step
    def run_model(self):
        question = self.input
        self.result = run_mrp(question, self.data)
        self.next

    @step
    def join(self, inputs):
        for result in inputs:
            save(result)

        self.next(self.end)

    @step
    def end(self):
        pass
```



Latana

MRP as Metaflow

```
from metaflow import FlowSpec,step

class MRPFlow(FlowSpec):

    @step
    def start(self):
        Self.data, self.questions = load_data()
        self.next(self.run_model, foreach="questions")

    @step
    def run_model(self):
        question = self.input
        self.result = run_mrp(question, self.data)
        self.next

    @step
    def join(self, inputs):
        for result in inputs:
            save(result)

        self.next(self.end)

    @step
    def end(self):
        pass
```



Latana

Parallelizing models

```
@step
def start(self):
    Self.data, self.questions = load_data()
    self.next(self.run_model, foreach="questions")


@step
def run_model(self):
    question = self.input
    self.result = run_mrp(question, self.data)
    self.next
```

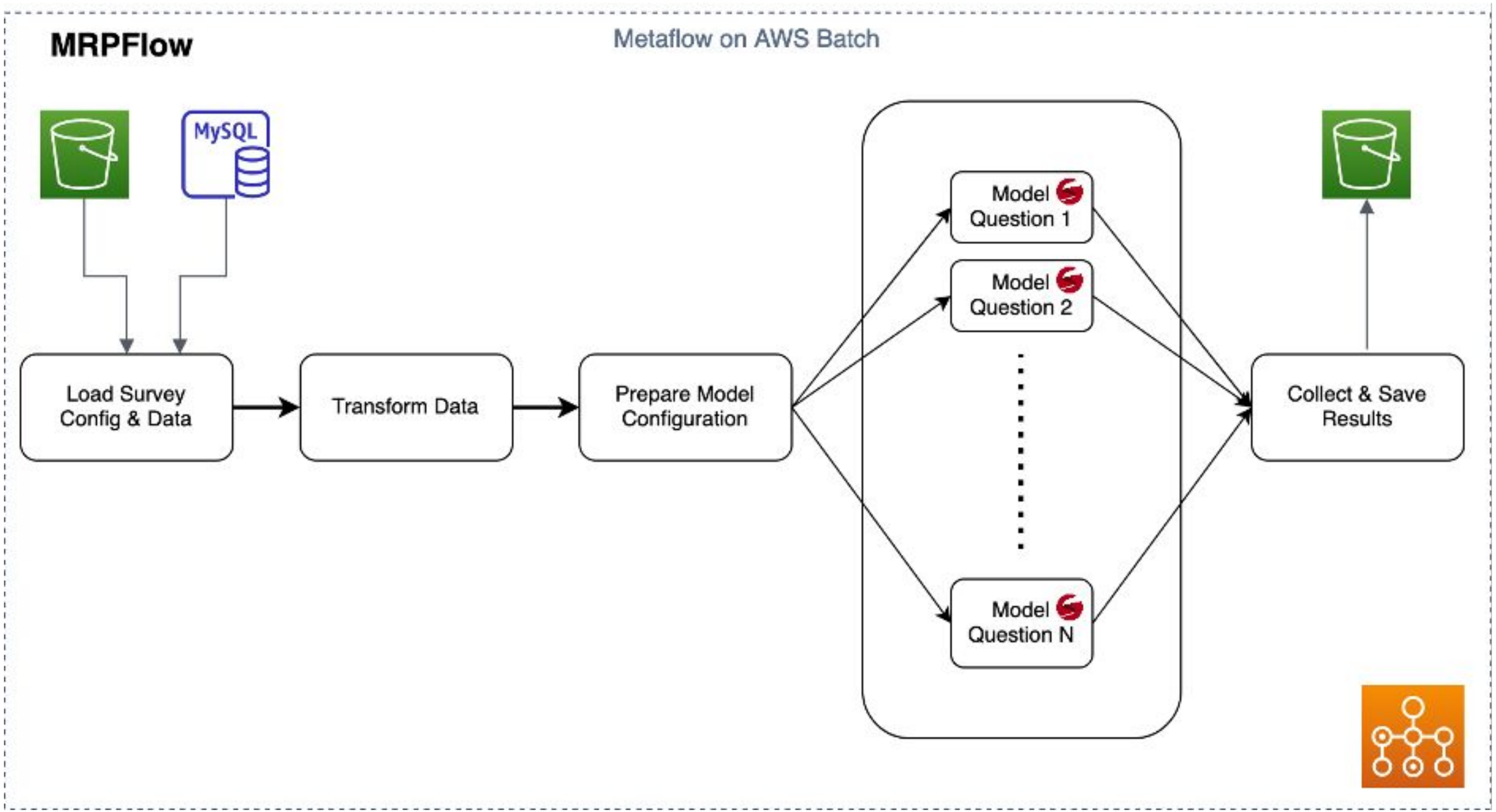
Latana

Increasing resources

```
@step
def start(self):
    Self.data, self.questions = load_data()
    self.next(self.run_model, foreach="questions")

@resources(cpu=8, memory=32000)
@step
def run_model(self):
    question = self.input
    self.result = run_mrp(question, self.data)
    self.next
```

Latana

Mr.P on AWS



Latana

Challenges

Convergence

How to monitor convergence of 1000+ models?

More predictor variables

Full joint distribution needed of all predictor variables.

Latana

Summary

- **Multilevel-regression improves errors by using grouped structure**

- **Propagation of uncertainty improves weighting**

Corrie Bartelheimer
Senior Data Scientist
corrie.bartelheimer@latana.com

Resources and Links

Introductory book on Bayesian Statistics: https://xcelab.net/rm/statistical-rethinking/

Stan: https://mc-stan.org/

Stan interface brms (R): https://paul-buerkner.github.io/brms/

MRP: Forecasting elections with non-representative polls https://www.sciencedirect.com/science/article/abs/pii/S0169207014000879

Metaflow https://metaflow.org/

MRP at Latana:

- https://latana.com/whitepapers/mrp-vs-traditional-quota-sampling-brand-tracking/
- https://aws.amazon.com/blogs/startups/brand-tracking-with-bayesian-statistics-and-aws-batch/