Automating Inventory at Stitch Fix

Using Beta Binomial Regression for Cold Start Problems

Sally Langford - Data Scientist

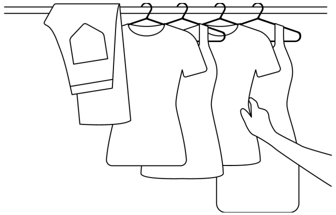
STITCH FIX™

How Stitch Fix works:

How Stitch Fix works:

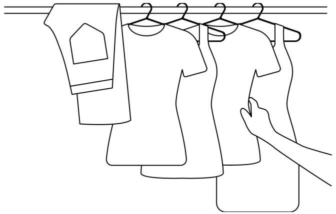
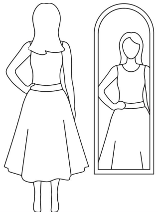
- Tell us about your style, fit and price preferences.

How Stitch Fix works:



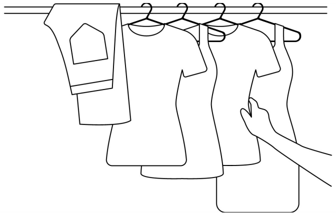
- Tell us about your style, fit and price preferences.
- A personal stylist will curate five pieces for you.

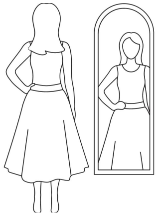
How Stitch Fix works:

- Tell us about your style, fit and price preferences.
- A personal stylist will curate five pieces for you.
- Try all the items on at home.

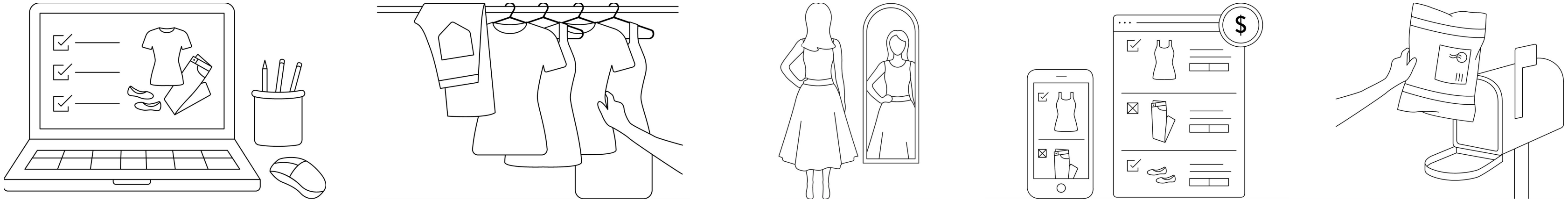
How Stitch Fix works:





- Tell us about your style, fit and price preferences.
- A personal stylist will curate five pieces for you.
- Try all the items on at home.
- Give your stylist feedback on all items, then only pay for what you keep.

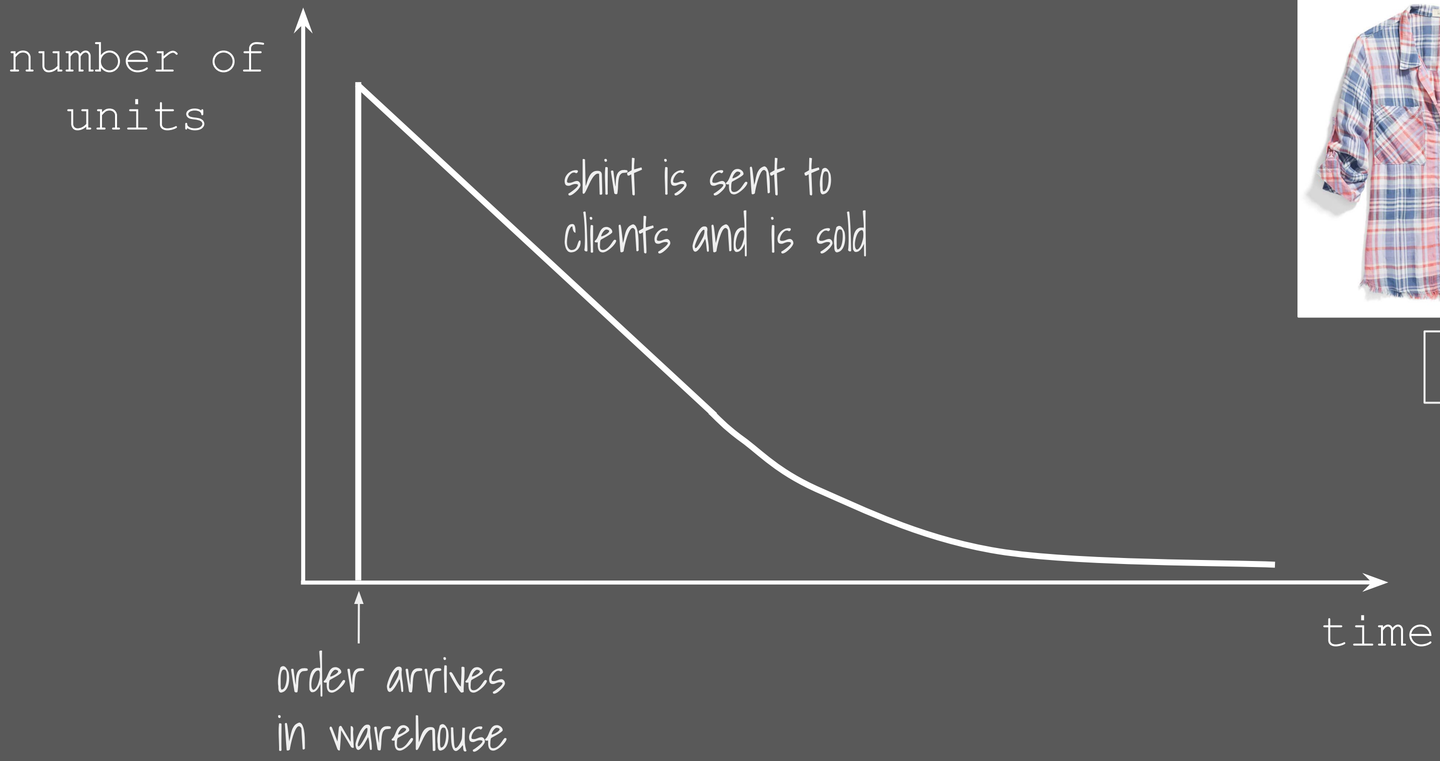
How Stitch Fix works:



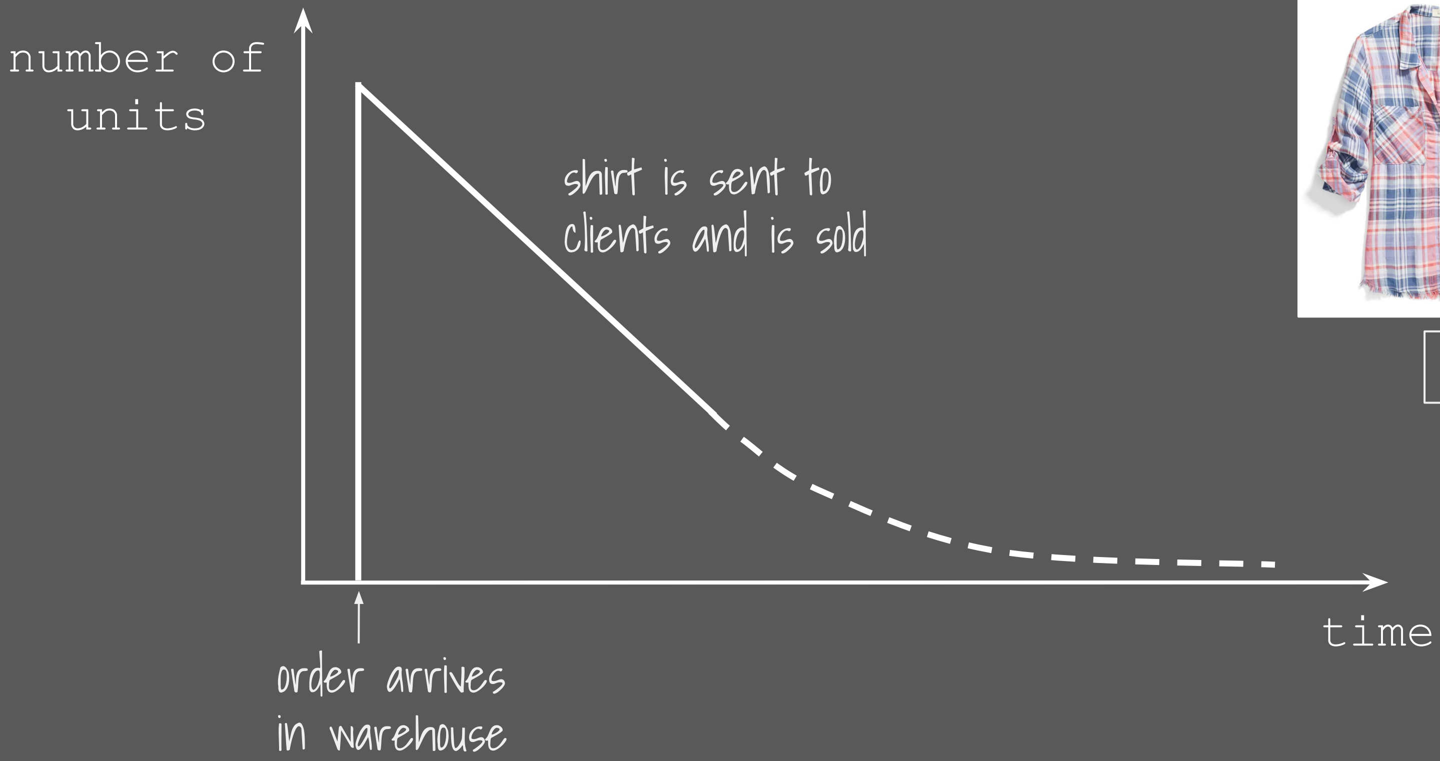
- Tell us about your style, fit and price preferences.
- A personal stylist will curate five pieces for you.
- Try all the items on at home.
- Give your stylist feedback on all items, then only pay for what you keep.
- Return the other items in envelope provided.

Benefits of Machine Learning in Inventory Management:

- Scalable with business.
- Rapid reforecasting.
- Capture nonlinear relationships.
- Cold start problems.


number of units
shirt is sent to clients and is sold
time
order arrives in warehouse

plaid shirt


number of units
shirt is sent to clients and is sold
time
order arrives in warehouse

plaid shirt

Inventory consumption of a style is proportional to;

- daily demand,
- clients for which the style is recommended,
- whether there are units in the warehouse,
- probability a stylist chooses to send the client this style,
- if the client buys the style.

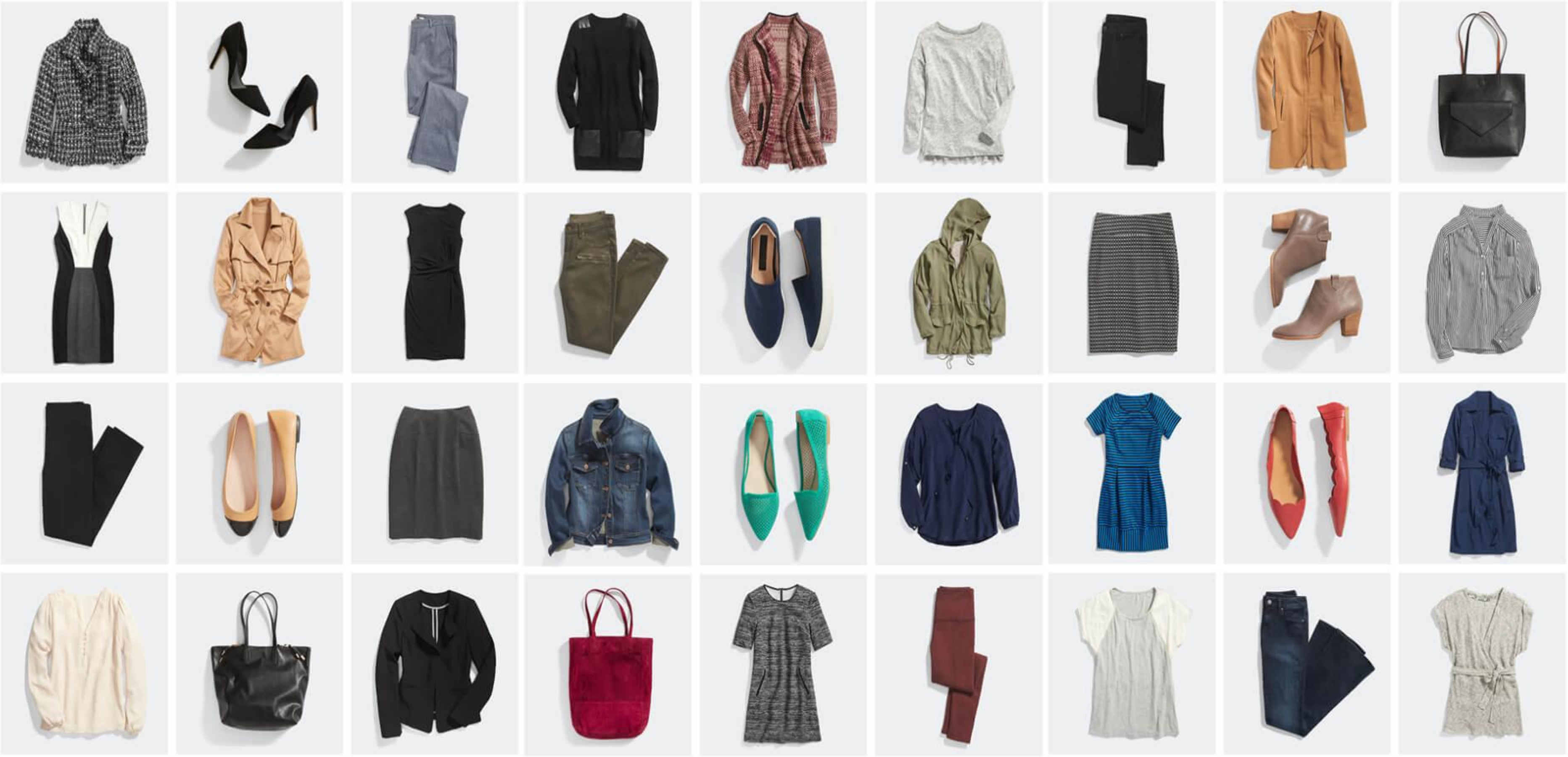
Inventory consumption of a style is proportional to;

- daily demand,
- clients for which the style is recommended,
- whether there are units in the warehouse,
- **probability a stylist chooses to send the client this style,**
- **if the client buys the style.**

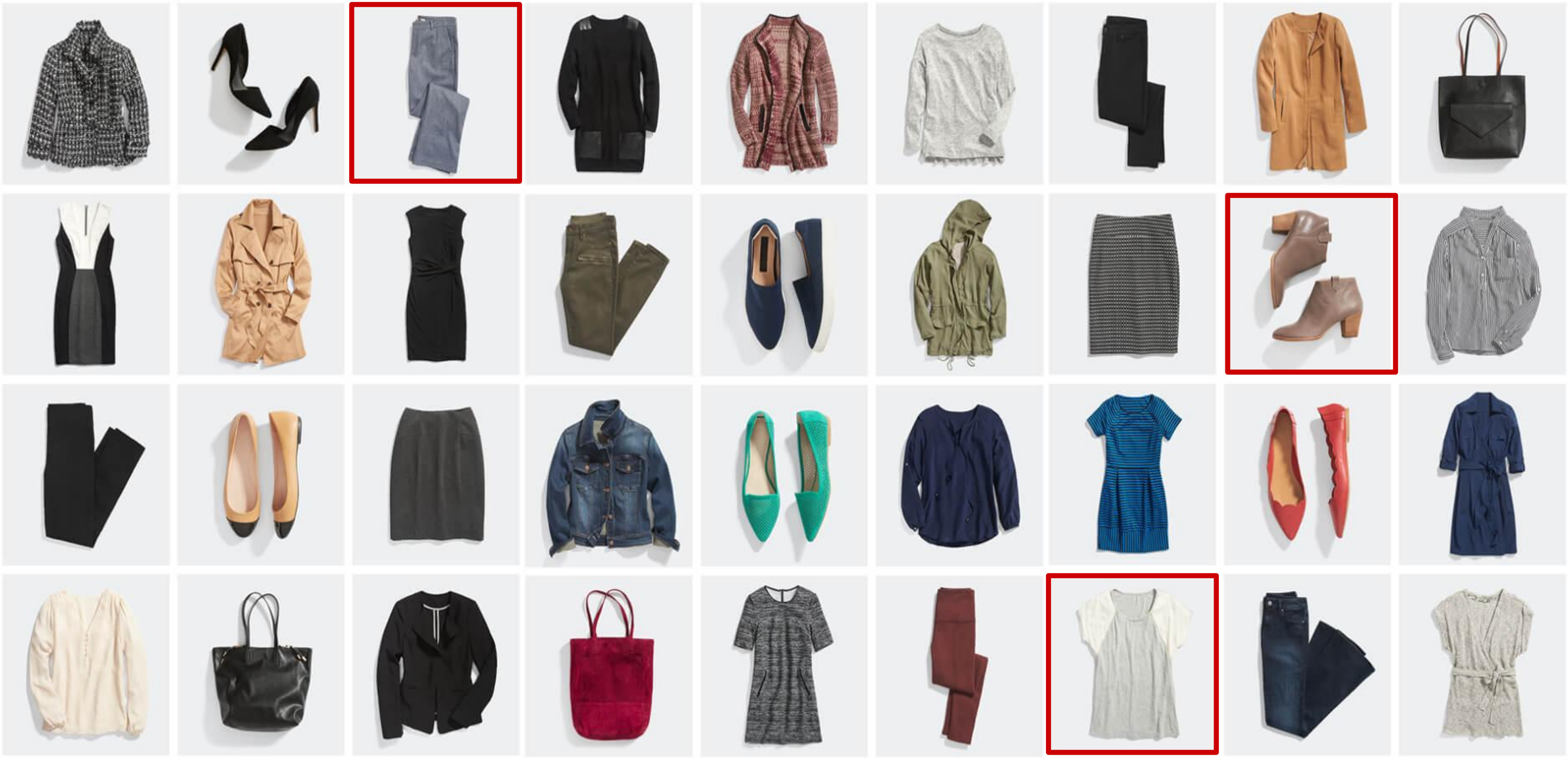
Inventory consumption of a style is proportional to;

- daily demand,
- clients for which the style is recommended,
- whether there are units in the warehouse,
- **probability a stylist chooses to send the client this style,**
- if the client buys the style.

Ranked styles recommended for client - which will the stylist choose to send?

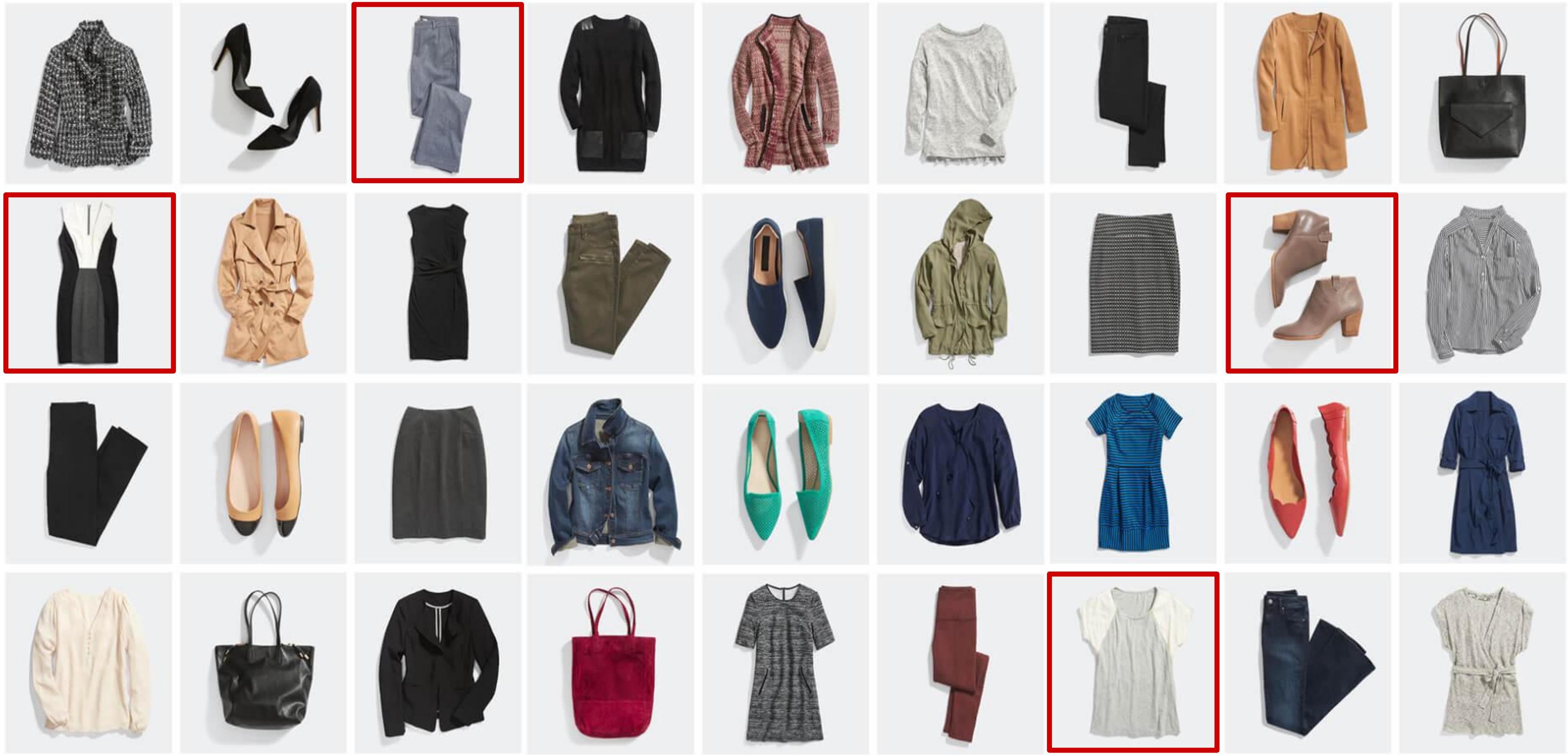


Ranked styles recommended for client - which will the stylist choose to send?

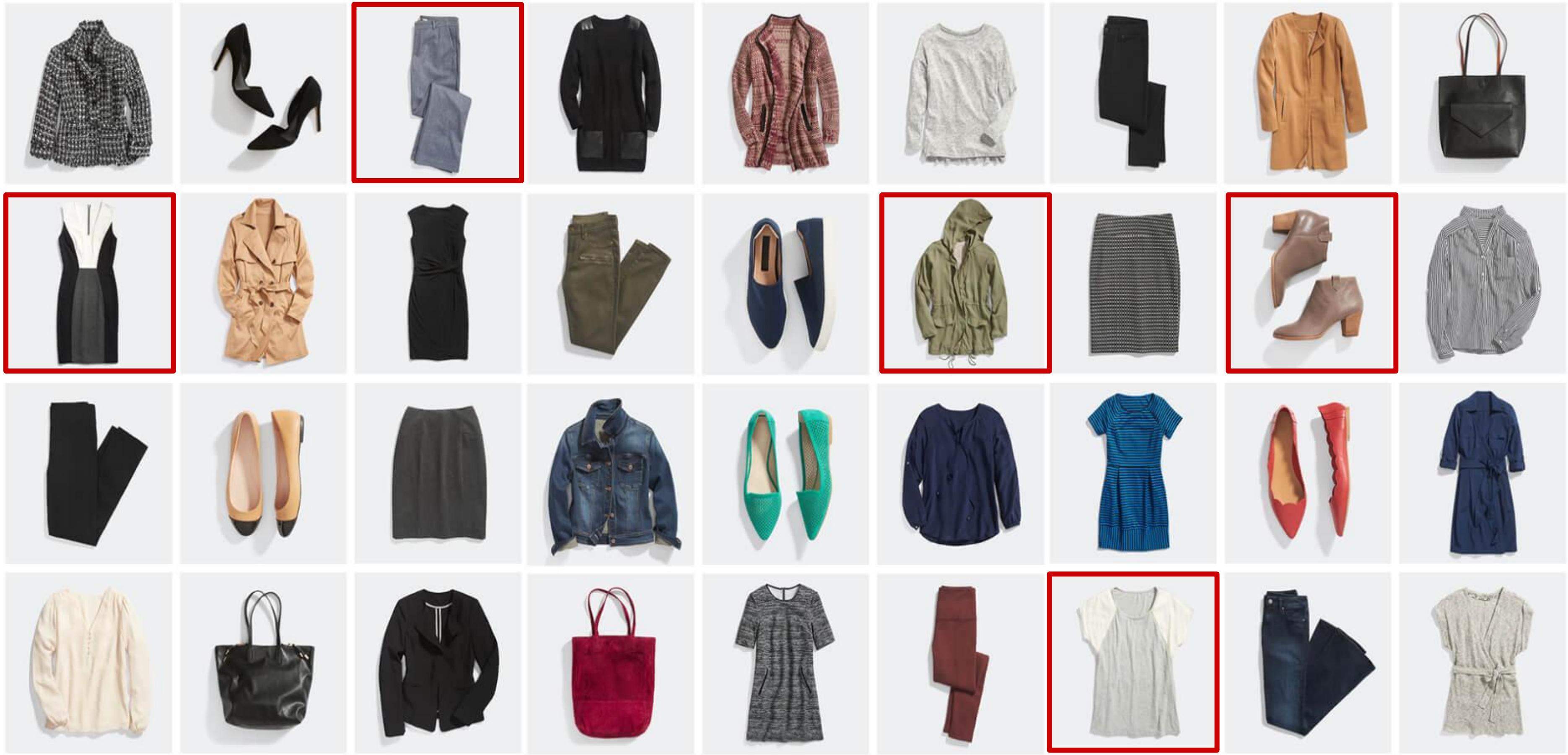


STITCH FIX™

Ranked styles recommended for client - which will the stylist choose to send?



Ranked styles recommended for client - which will the stylist choose to send?



STITCH FIX™

Ranked styles recommended for client - which will the stylist choose to send?

plaid long-sleeve shirt

plaid long-sleeve shirt

plaid long-sleeve shirt

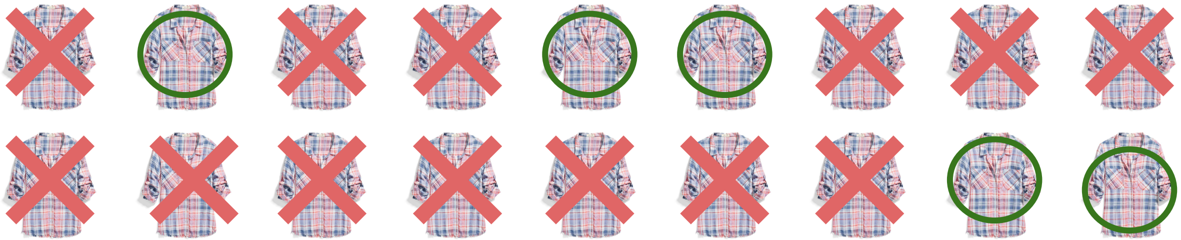
STITCH FIX™

plaid long-sleeve shirt

plaid long-sleeve shirt

...

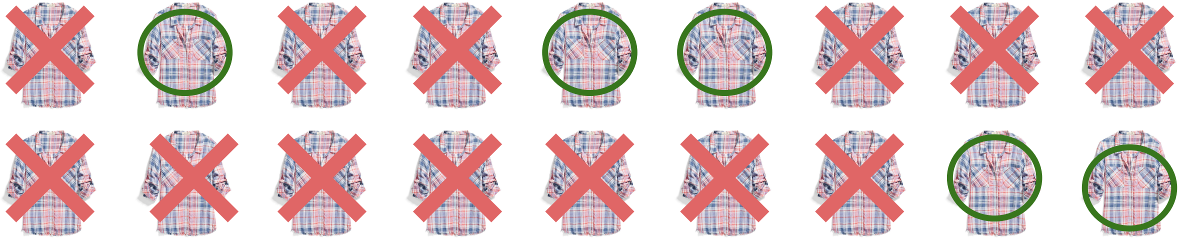
plaid long-sleeve shirt



...

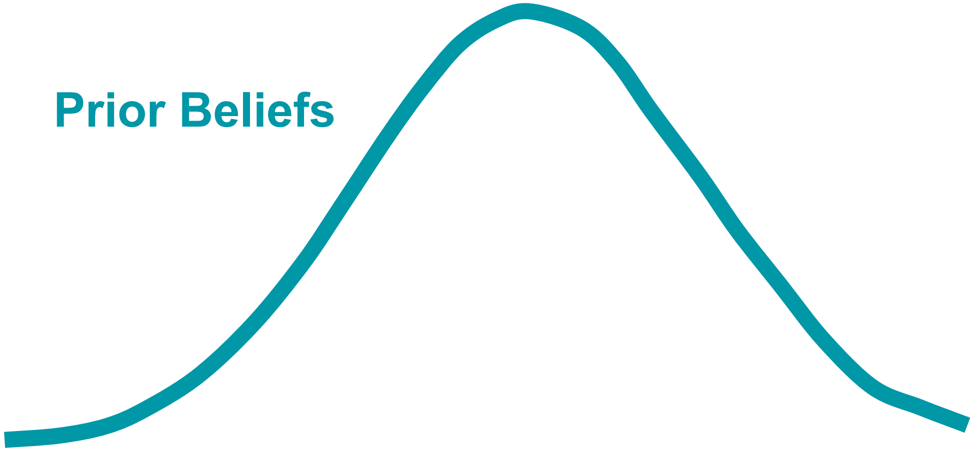
$$P(\text{chosen}) + P(\text{not chosen}) = 1$$

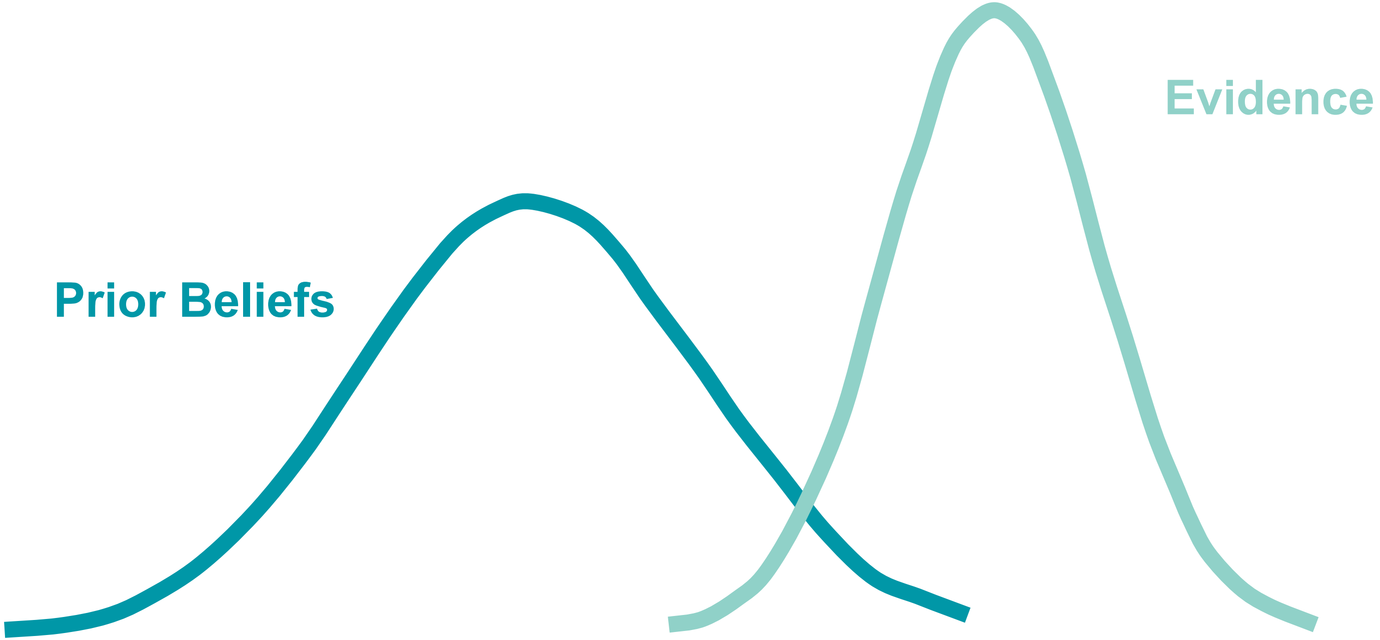
plaid long-sleeve shirt

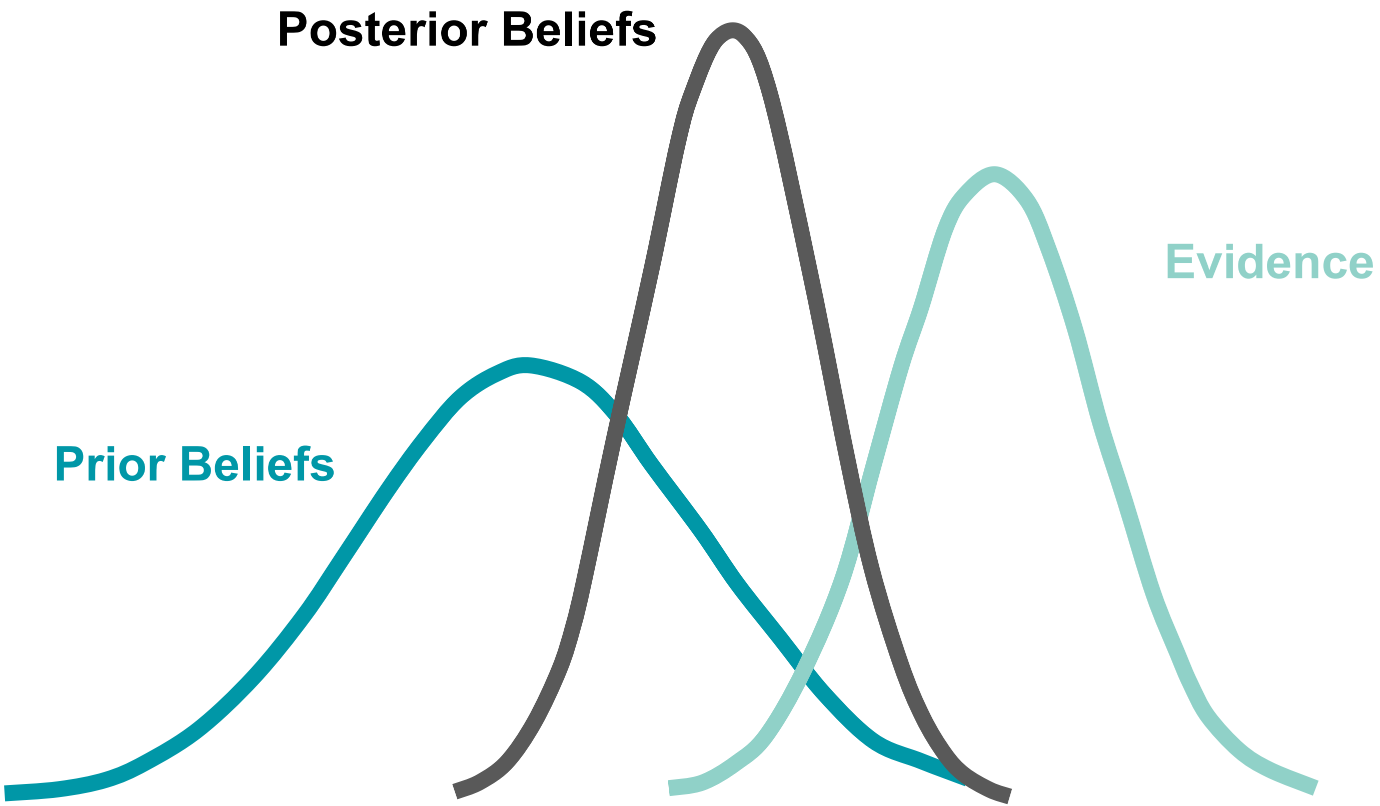


...

blue long-sleeve shirt


Prior Beliefs


Prior Beliefs
Evidence



STITCH FIX™

$$N \sim \mathrm{Binom}(N_{av}, p)$$

$$N \sim \text{Binom}(N_{av}, p)$$

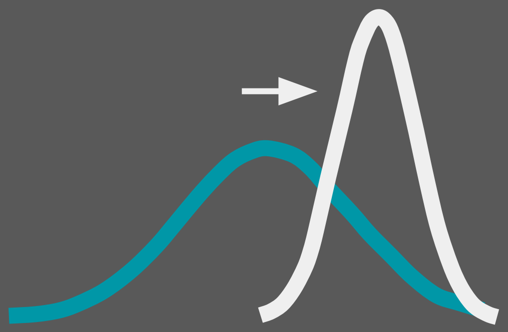
$$N \sim \mathrm{Binom}(N_{av}, p)$$

$$p \sim B(\alpha, \beta)$$

STITCH FIX™

$$B(\alpha', \beta') = B(\alpha_0 + k, \beta_0 + n - k)$$

$$B(\alpha', \beta') = B(\alpha_0 + k, \beta_0 + n - k)$$



STITCH FIX™

Step 1: Use maximum likelihood to calculate $\boldsymbol{\alpha_0}$ and $\boldsymbol{\beta_0}$ for the distribution of **p** in groups of similar styles.

Step 2: After a period of time, update this prior for the number of times the new style has been recommended for a client (**n**), and chosen to be sent (**k**).

Step 3: Calculate the mean and confidence interval of **p** from the resulting distribution. This is used as the probability that the new style will be chosen to be sent to a client.

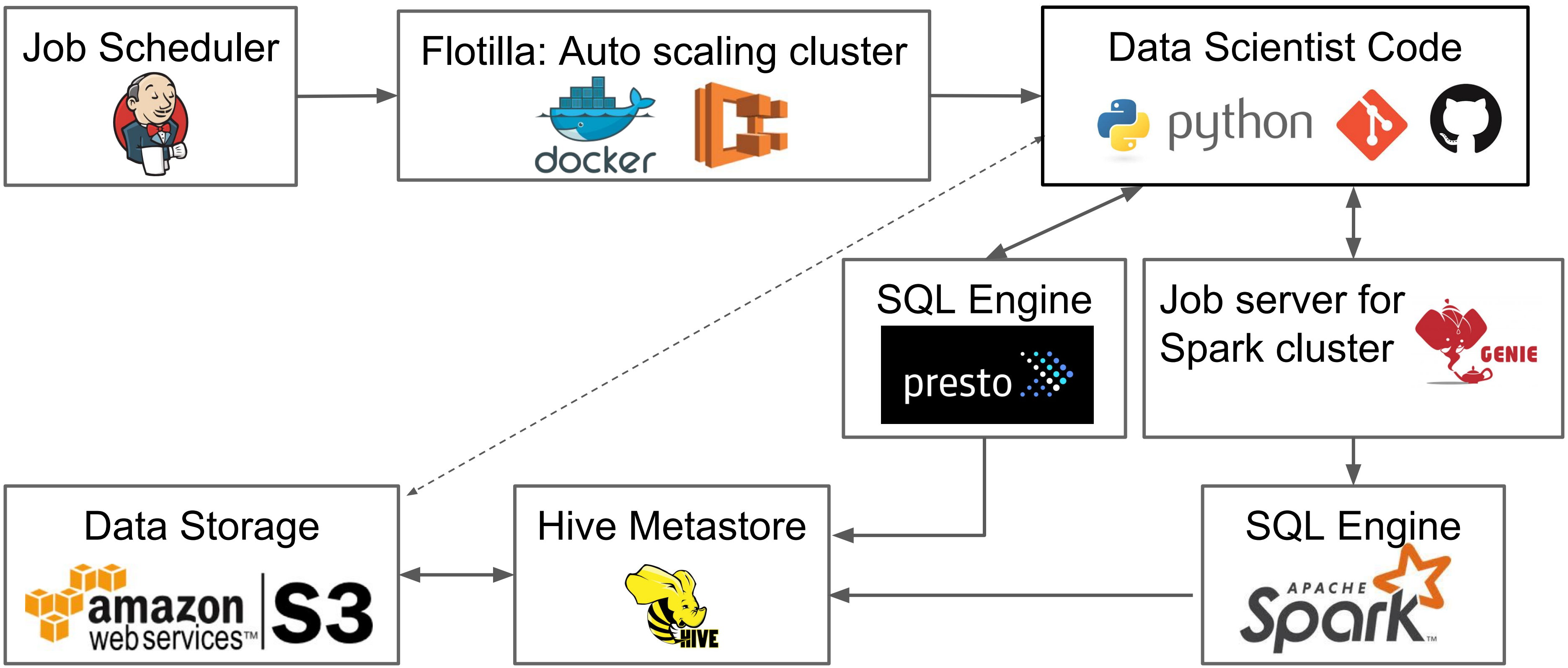
Step 4: Repeat steps 2-3.

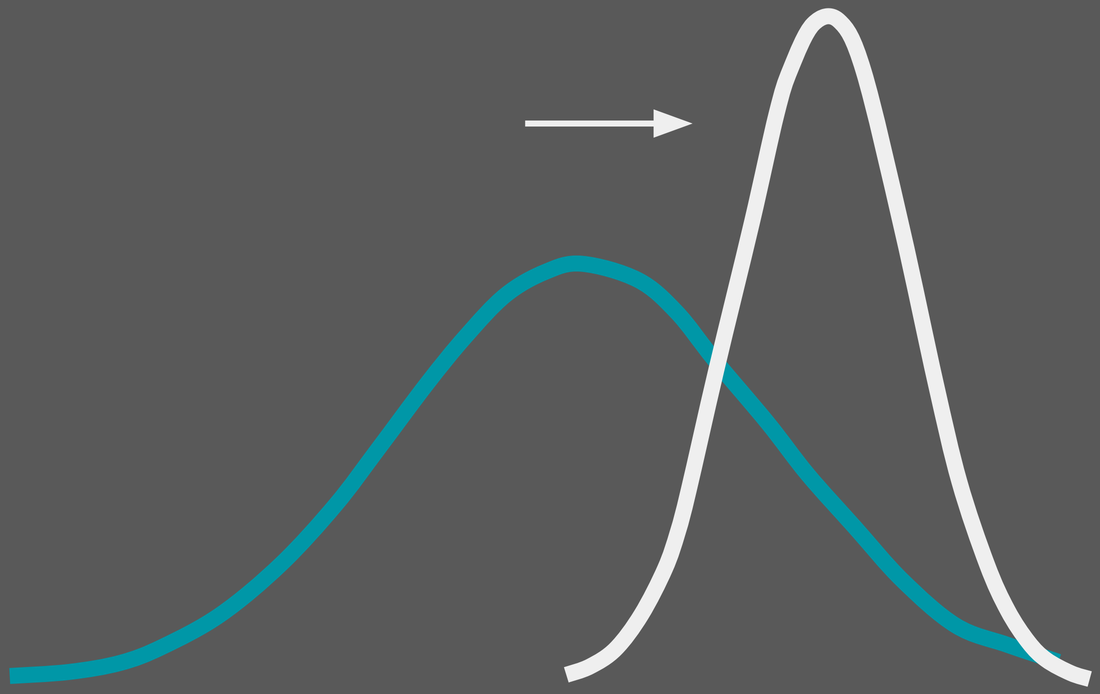
VGAM (in python):

```
import rpy2.robjects as robjects
robjects.r.library("VGAM")
robjects.r("fit = vglm(cbind(successData,trialData - successData) ~ 1,
        betabinomialff, trace=TRUE)")
alpha, beta = robjects.r("Coef(fit)")
```

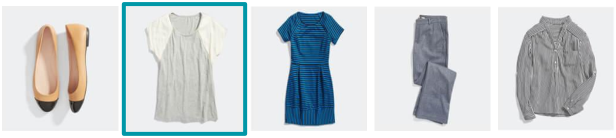
SciPy

```
import scipy
fit = scipy.stats.beta.fit(data, floc=0, fscale=1)
alpha, beta = fit[0], fit[1]

---

result = scipy.optimize.minimize(loss_function, p0, jac=True, **kwargs)
```

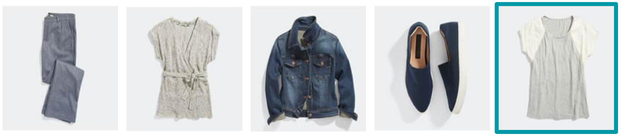

Job Scheduler
Flotilla: Auto scaling cluster
docker
Data Scientist Code
python
SQL Engine
presto
Job server for
Spark cluster
GENIE
Data Storage
amazon
web services™
S3
Hive Metastore
HIVE
SQL Engine
APACHE
Spark™



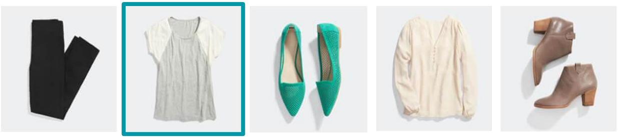
Top 5 recommendations for client A



Top 5 recommendations for client B



Top 5 recommendations for client C

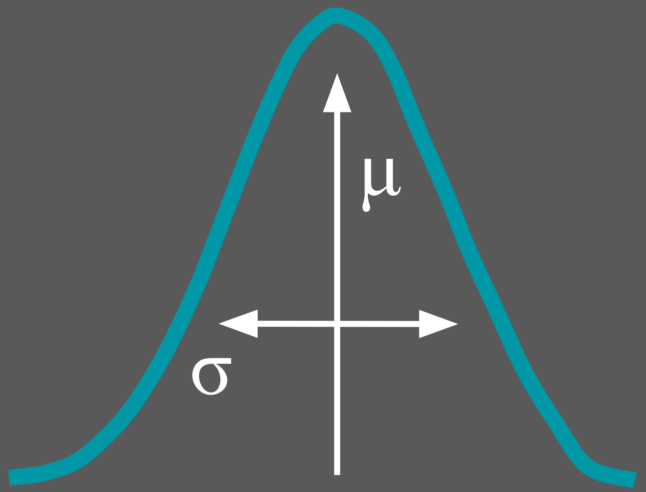


Top 5 recommendations for client D

Top 5 recommendations for client E

STITCH FIX™

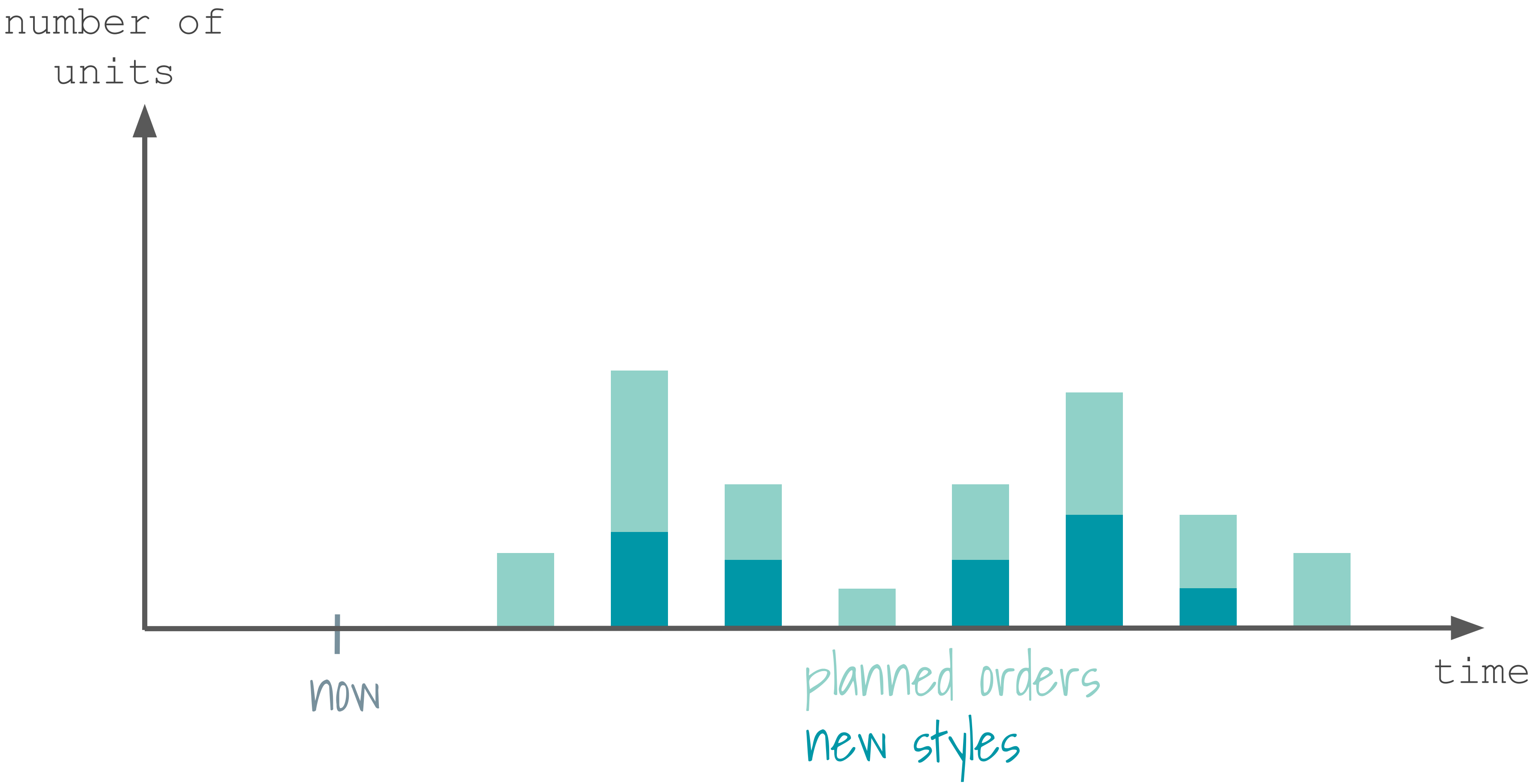
$$B(\alpha , \beta) = B(\mu/\sigma, (1 - \mu)/\sigma)$$

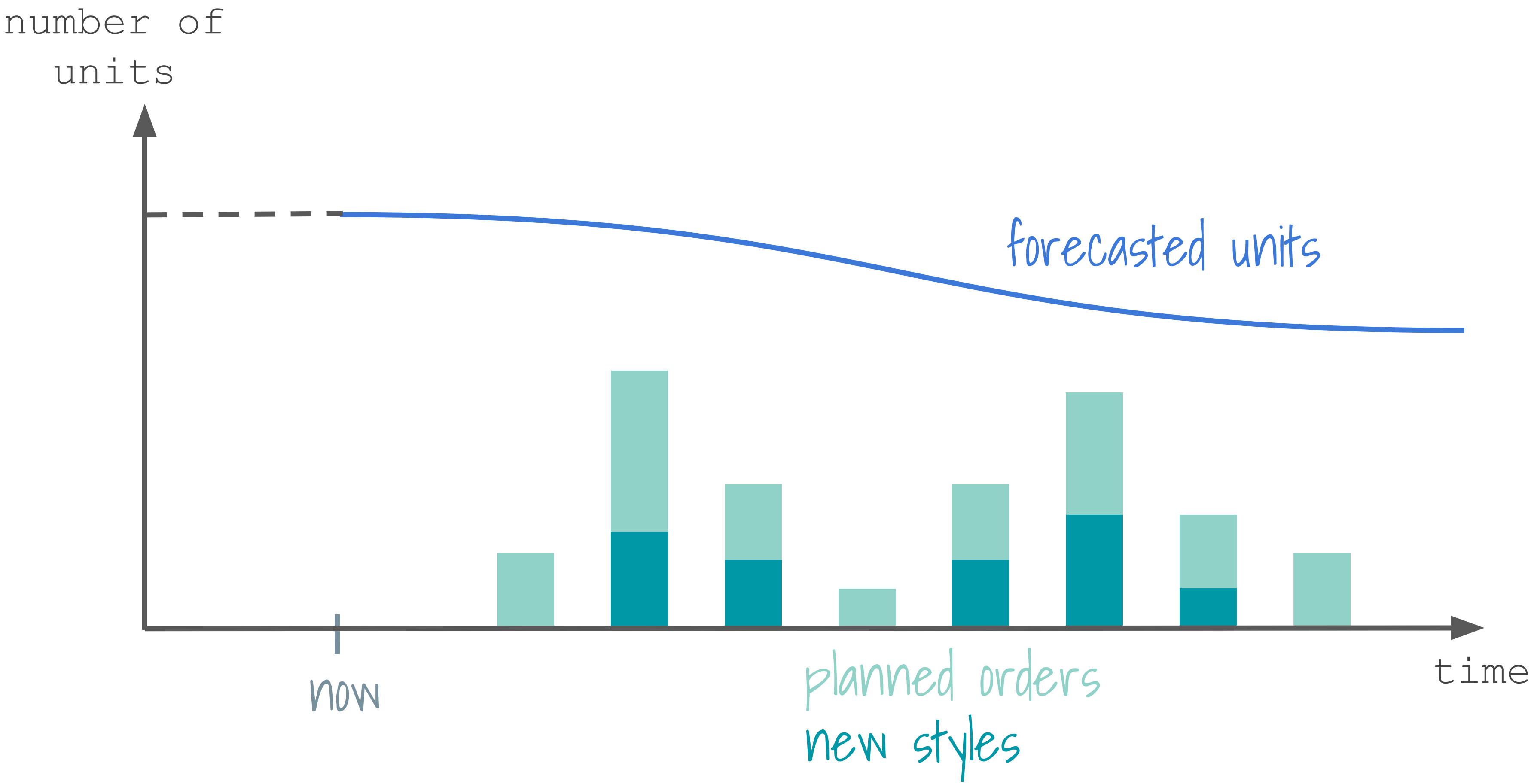


STITCH FIX™

$$\mu = \mu_0 + \mu_n \log(1 + n)$$

STITCH FIX™


number of
units
now
planned orders
new styles
time


number of
units
forecasted units
now
planned orders
new styles
time

STITCH FIX™

How do we use our inventory forecast model?

- When should we re-order inventory?
- How should we buy inventory by size?
- How should orders be separated into different warehouses?
- When should a style not be sent out anymore, in place of a new option?

Metrics of success:

- Fraction of inventory out with clients compared to in the warehouse?
- How many styles are available to send to a client?
- Δ in the beginning of month projected units.
- Cumulative units sold over time.

Do you want to calculate the probability of success in a binomial process?

Not enough data?

Use Beta Binomial Regression for your cold start problem!

slangford@stitchfix.com

[Stitch Fix]

[Algorithms Blog]

[Algorithms Tour]

@stitchfix_algo

STITCH FIX™