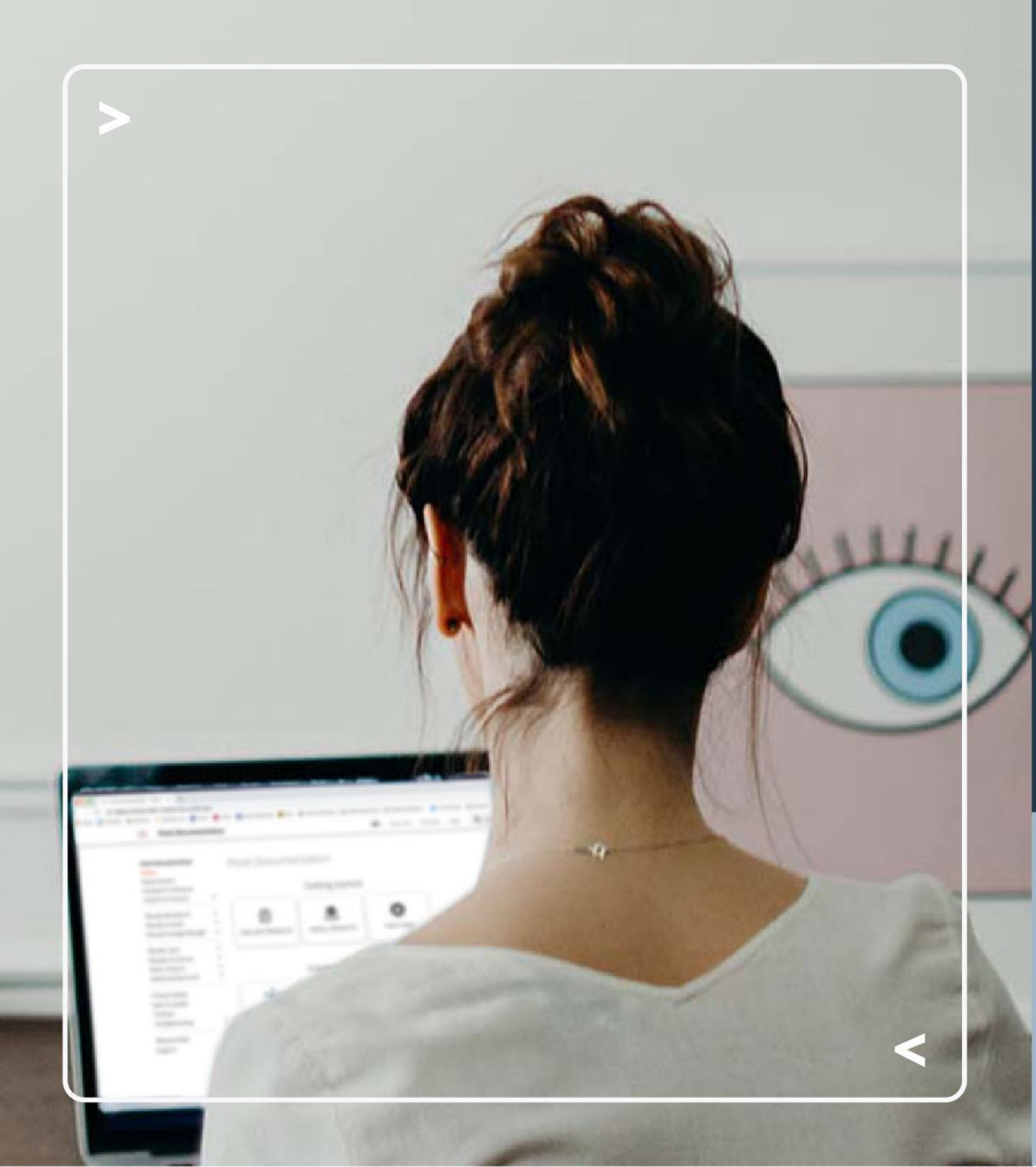
What even is “production”?

MLOps for the curious

Julia Silge / January 2022

MLOps is…

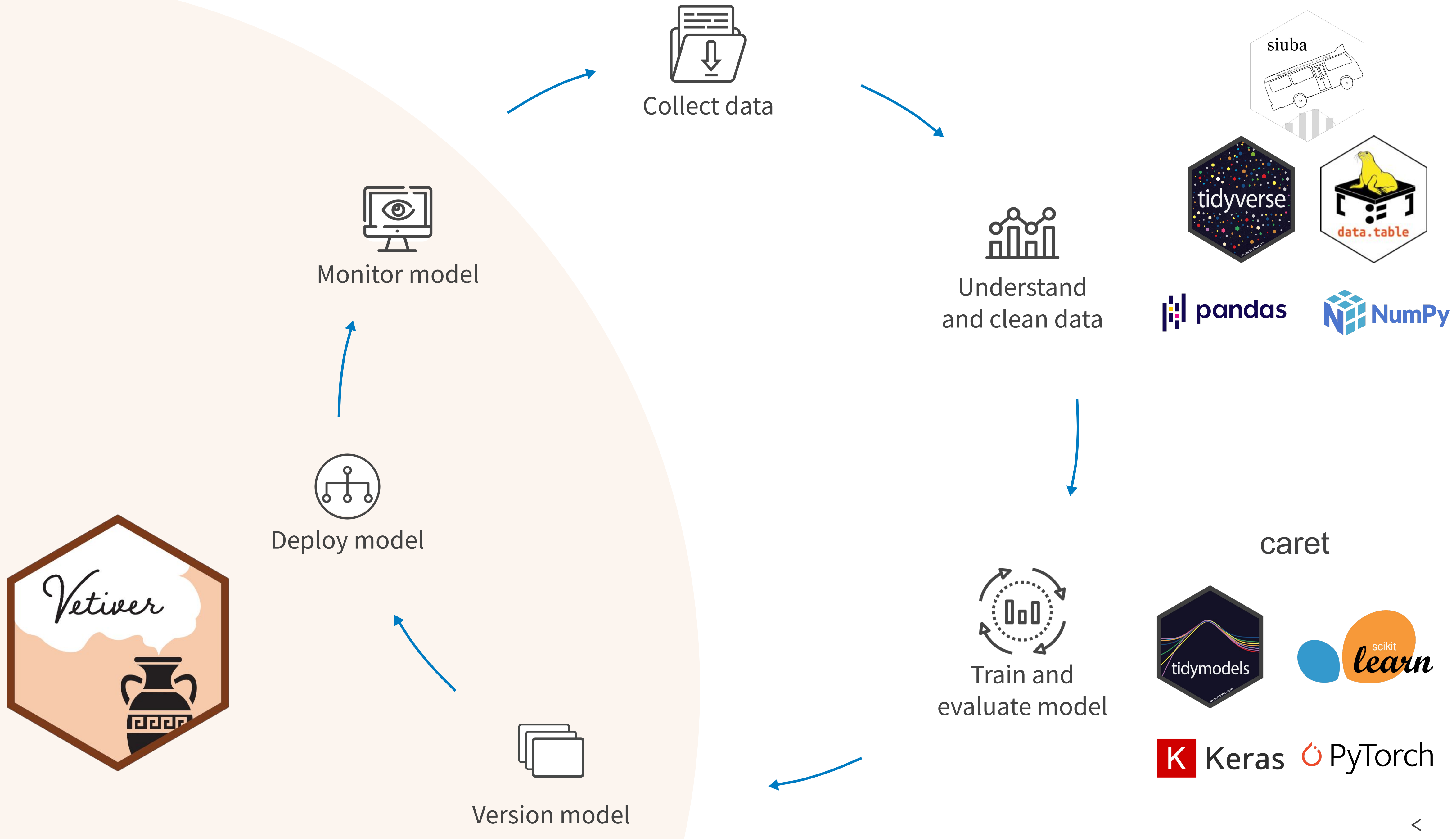
a set of practices to *deploy* and *maintain* machine learning models in production **reliably** and **efficiently**



Develop an ML model

Photo by Rodion Kutsaiev on Unsplash

Maintain the model



Collect data

Understand and clean data

Train and evaluate model

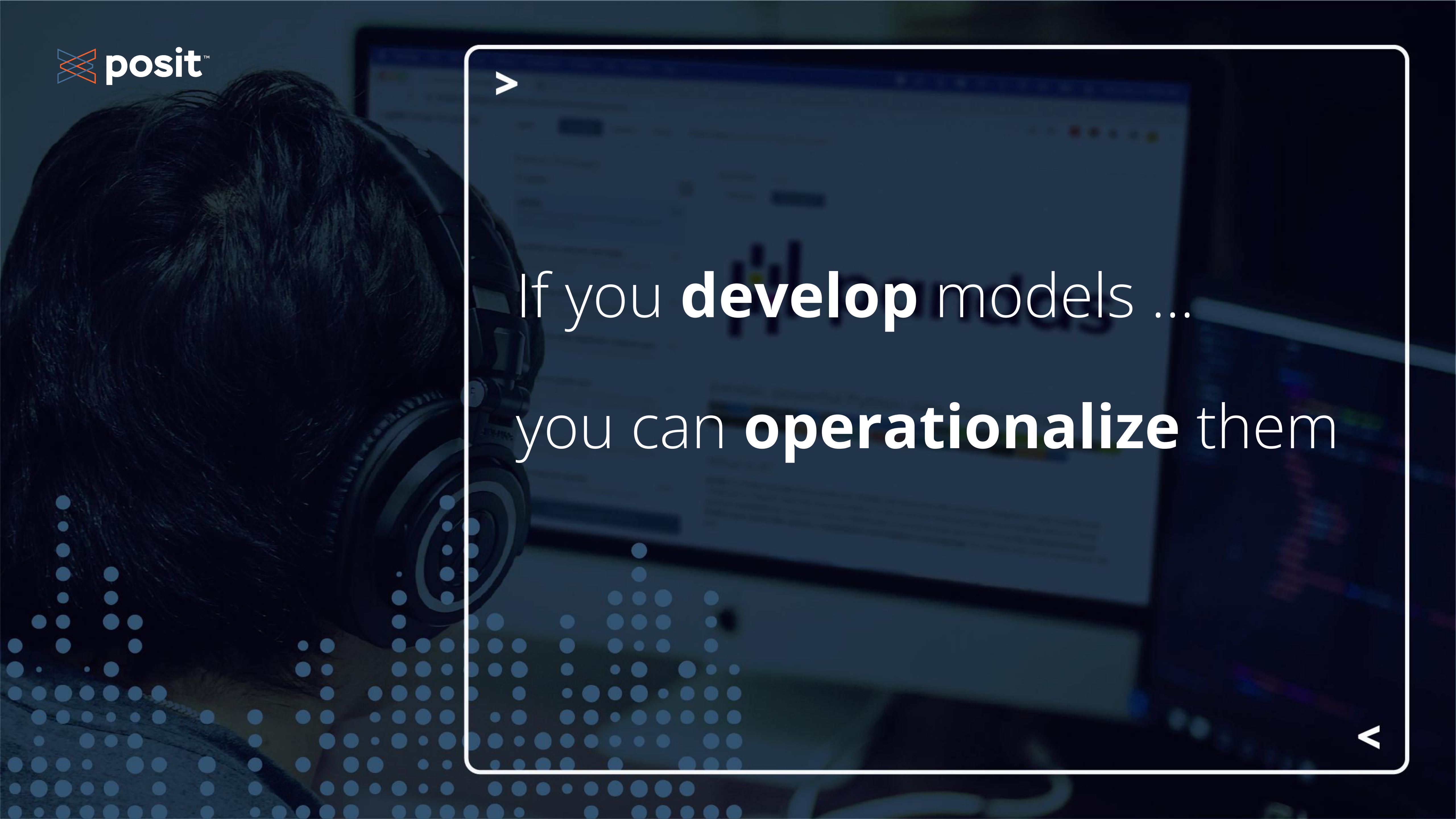
Version model

Deploy model

Monitor model

caret

< >

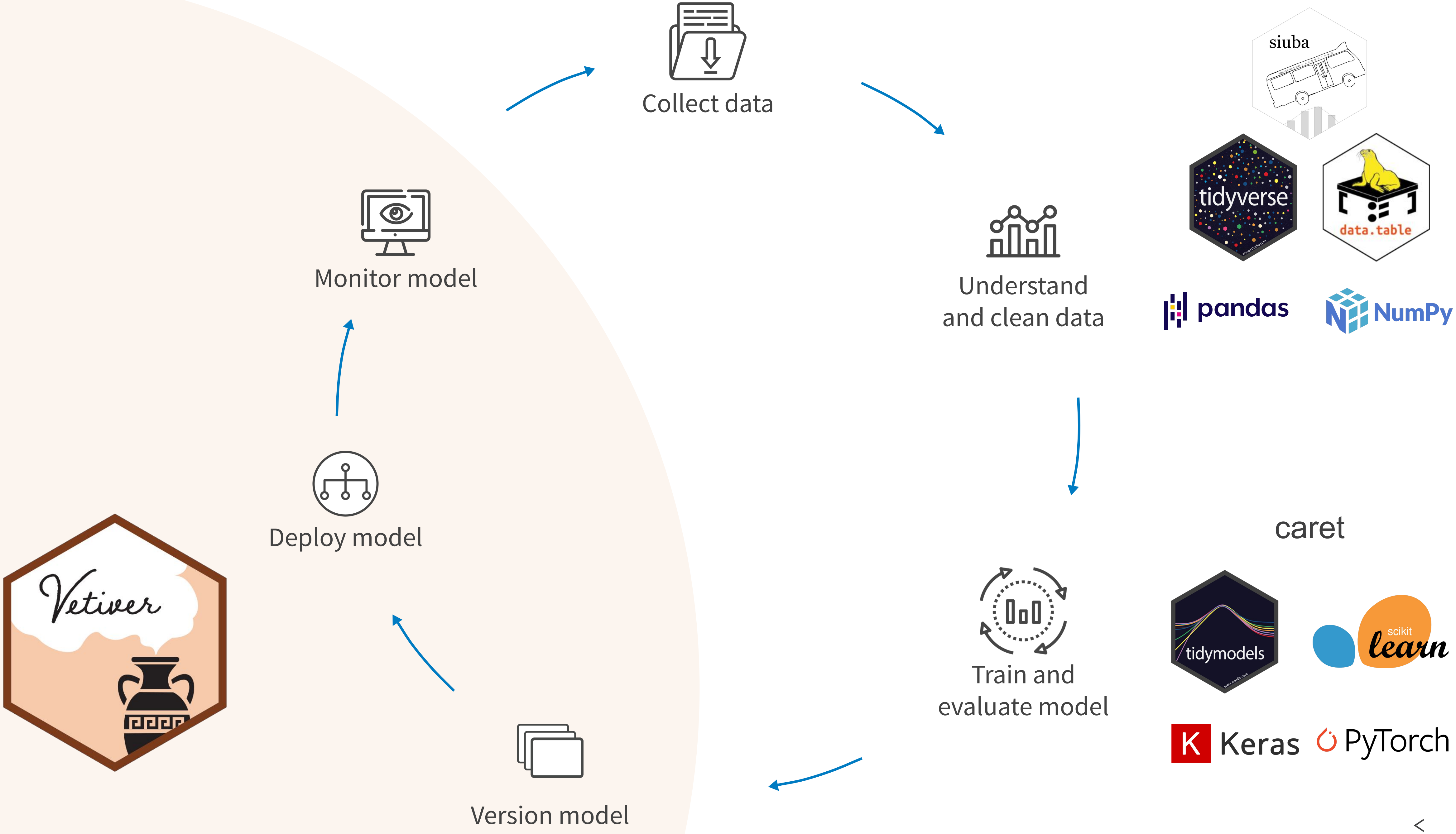


If you **develop** models ...

you can **operationalize** them

Using vetiver…

- allows those new to MLOps to get started **quickly**
- supports **scaling** safely as an org matures



Collect data

Understand and clean data

Train and evaluate model

Version model

Deploy model

Monitor model

Version your model

Version your model

```
model
model_final
model_final_final
model_final_final_actually
model_final_final_actually_2
```

Version your model

```python
from vetiver import VetiverModel, vetiver_pin_write
import pins

v = VetiverModel(
     rf_pipe,
     ptype_data = X_train,
     model_name = "superbowl-ads"
)
board = pins.board_rsconnect(
     server_url = server_url, # load from an .env file
     api_key = api_key, # load from an .env file
     allow_pickle_read = True
)

vetiver_pin_write(board, v)
```

```r
library(vetiver)
library(pins)

v <- vetiver_model(rf_fit, "superbowl-ads")

model_board <- board_rsconnect() # uses env vars
vetiver_pin_write(model_board, v)
```

https://colorado.posit.co/rsc/seattle-housing-pin/

Deploy your model

Deploy your model

```python
my_api = VetiverAPI(v)
my_api.run()
```

```r
library(plumber)

pr() %>%
    vetiver_api(v, debug = TRUE) %>%
    pr_run()
```

Deploy your model

```
vetiver.deploy_rsconnect(
    connect_server = connect_server,
    board = model_board,
    pin_name = "isabel/superbowl-ads",
    version = "59869"
)
```

```
vetiver_deploy_rsconnect(
    board = model_board,
    name = "julia/superbowl-ads",
    predict_args = list(debug = TRUE)
)
```

https://colorado.posit.co/rsc/seattle-housing/

Deploy your model

```python
vetiver.prepare_docker(
    board = model_board,
    pin_name = "isabel/superbowl-ads",
    version = "59869"
)
```

```r
vetiver_prepare_docker(
    board = model_board,
    name = "julia/superbowl-ads",
    predict_args = list(debug = TRUE)
)
```

Where can you deploy a model?

posit™

AWS SageMaker

Azure ML

Google Cloud Run

Coming soon to vetiver:

AWS Lambda

GCP Vertex AI

Make it easy to do the right thing

- Robust and human-friendly checking of new data
- Track and document software dependencies of models
- Model cards for transparent, responsible reporting

Model Card Skeleton

Monitor your model

Monitor your model

```
metrics = vetiver.compute_metrics(
    new_data,
    "date",
    timedelta(weeks = 1),
    [mean_absolute_error, r2_score],
    "like_count",
    "y_pred"
)

vetiver.pin_metrics(
    model_board,
    metrics,
    "metrics_pin_name",
    overwrite = True
)

vetiver.plot_metrics(metrics)
```

```
metrics <-
    augment(v, new_data) %>%
    vetiver_compute_metrics(
        date,
        "week",
        like_count,
        .pred
    )

vetiver_pin_metrics(
    model_board
    metrics,
    "metrics_pin_name",
    overwrite = TRUE
)

vetiver_plot_metrics(new_metrics)
```

https://colorado.posit.co/rsc/seattle-housing-dashboard/

Supporting advanced use cases

```
app_file = vetiver.write_app(model_board, "cars_mpg")
```

```
from vetiver import VetiverModel
import vetiver
import pins


b = pins.pins.board_rsconnect(
    server_url = server_url,
    api_key = api_key,
    allow_pickle_read = True
)

v = VetiverModel.from_pin(
    b,
    "isabel/superbowl-ads",
    version = "81453"
)

vetiver_api = vetiver.VetiverAPI(v)
api = vetiver_api.app
```

```
vetiver_write_plumber(model_board, "cars_mpg")
```

```
# Generated by the vetiver package; edit with care

library(pins)
library(plumber)
library(rapidoc)
library(vetiver)
b <- board_rsconnect(
    "envvar",
    server = "https://colorado.posit.co/rsc"
)
v <- vetiver_pin_read(
    b,
    "julia/superbowl-ads",
    version = "61923"
)

#* @plumber
function(pr) {
     pr %>% vetiver_api(v)
}
```

Why should data practitioners be excited about MLOps?

- Connect your work to the “real world”
- Scale your impact

Thank you!