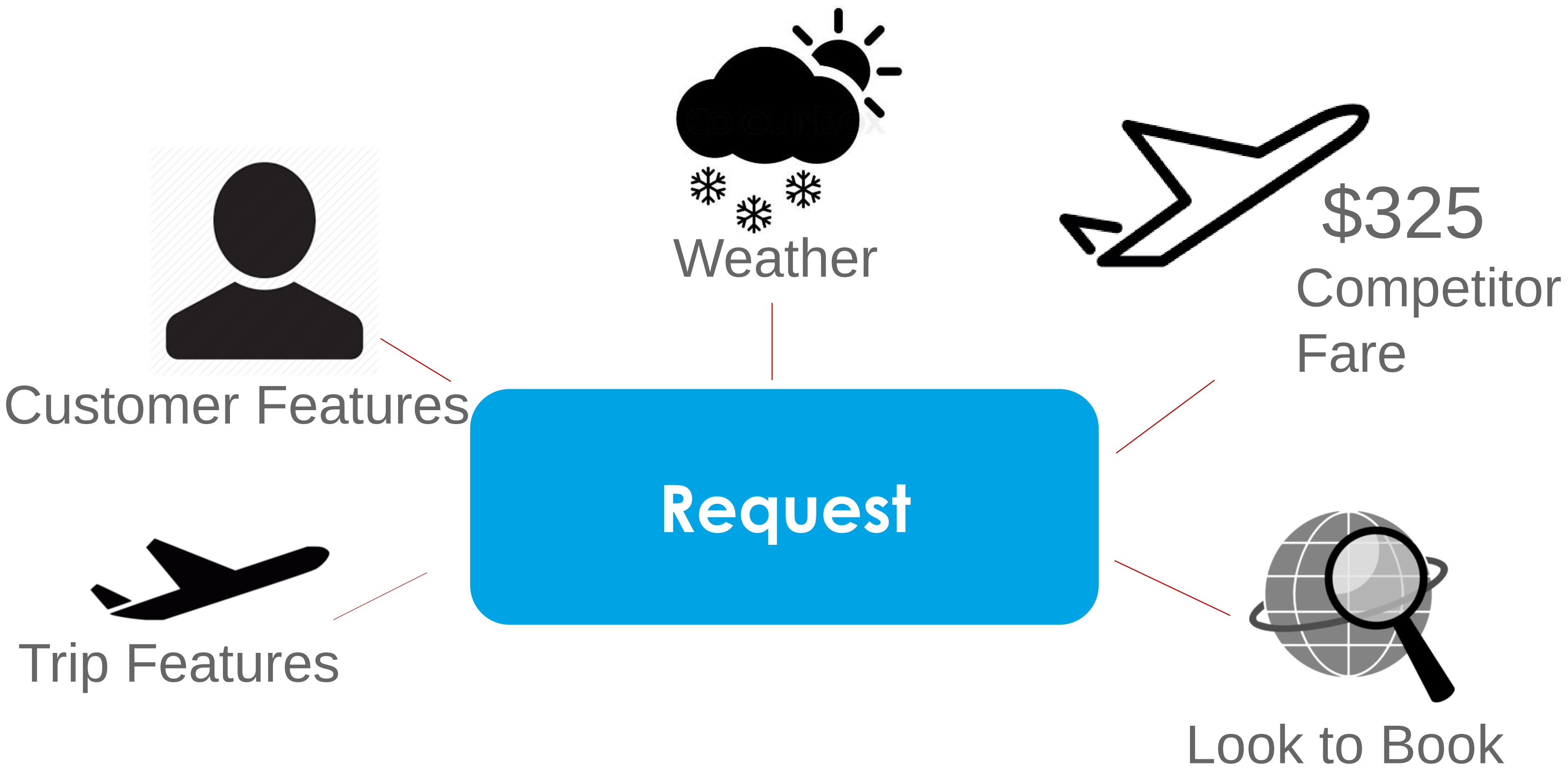
Ensemble methods in dynamic pricing

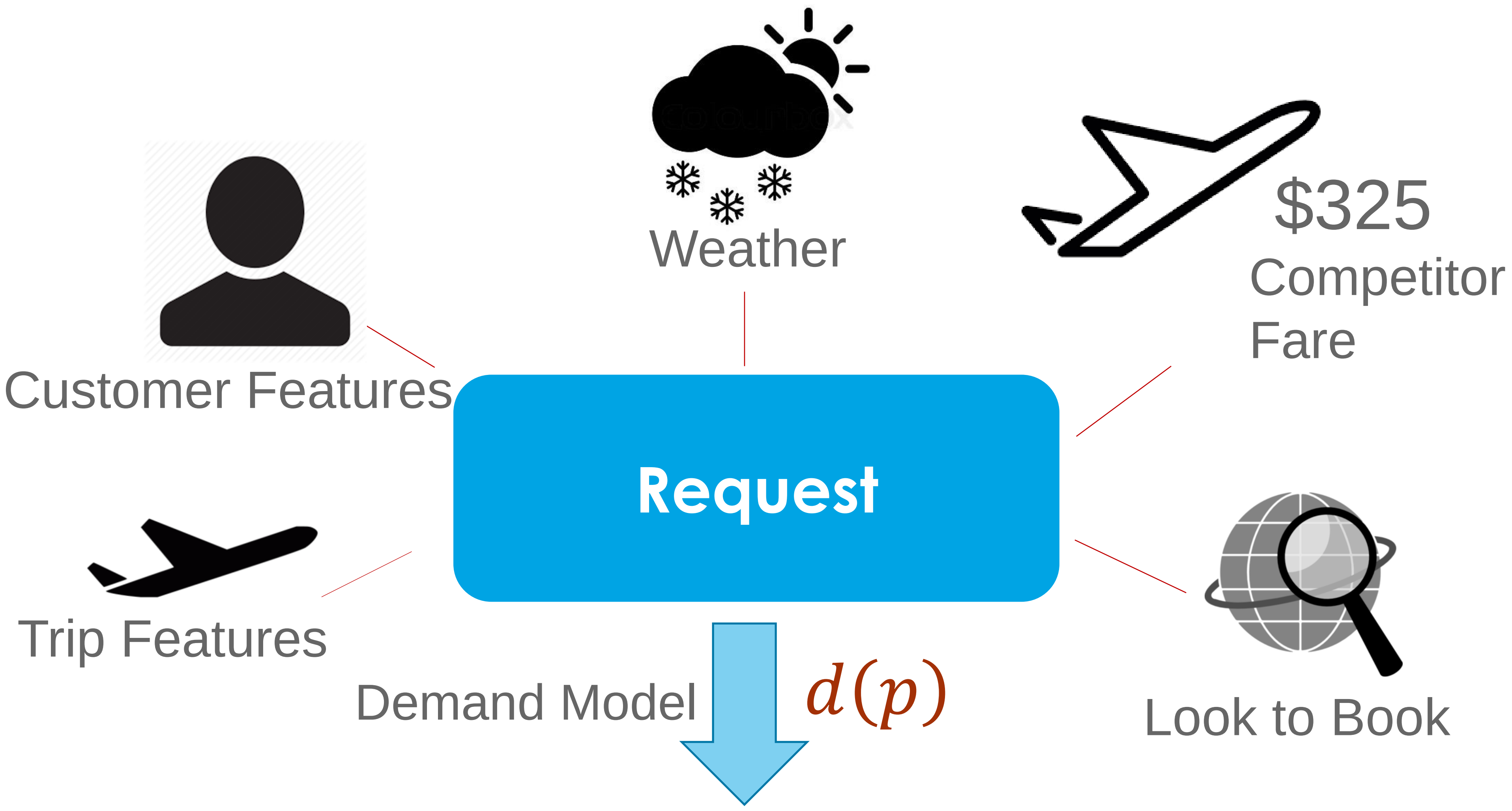
Dirk Sierag,
Data Scientist
Collaboration with Ravi Kumar

PROS

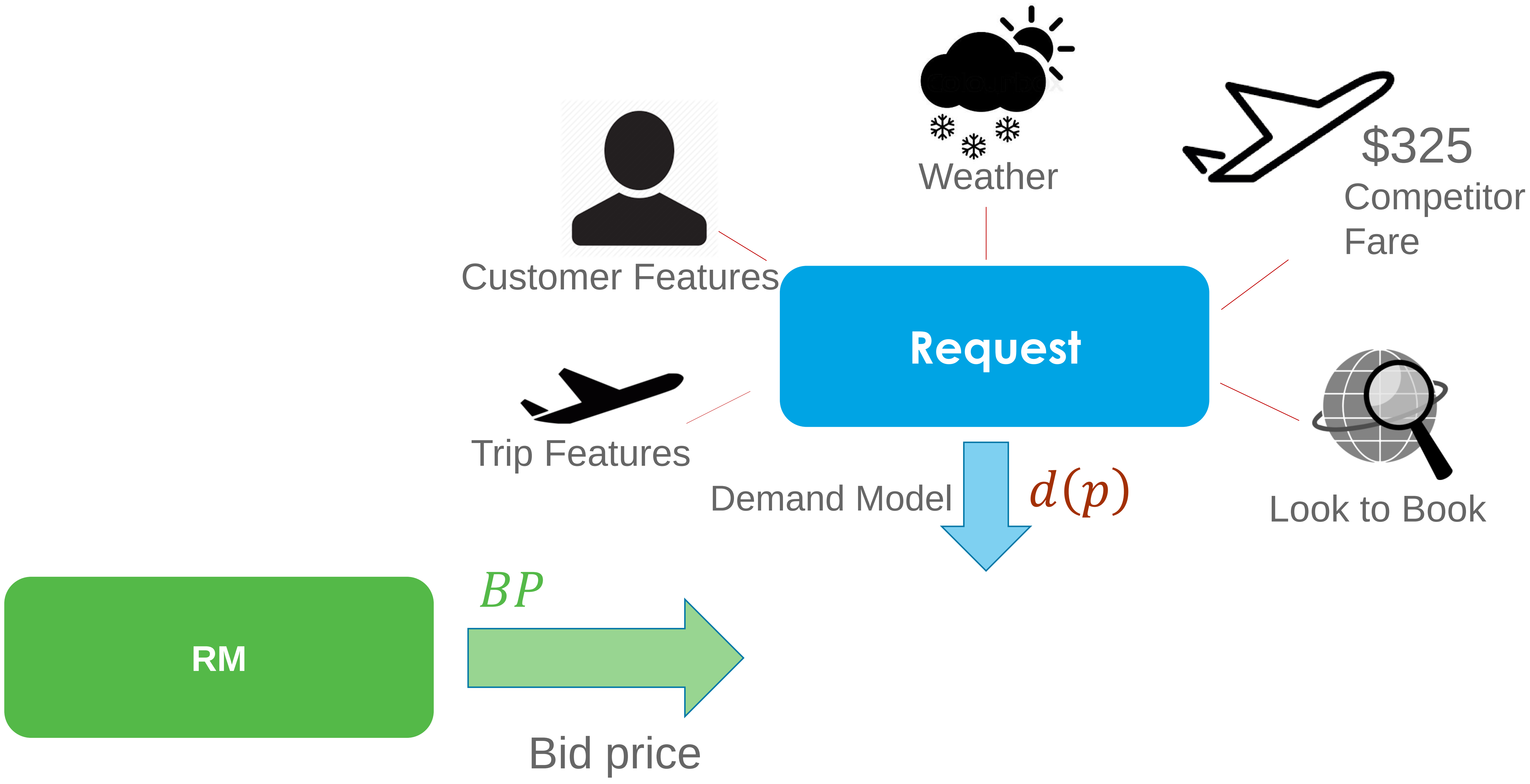
Dynamic/request specific pricing for airlines



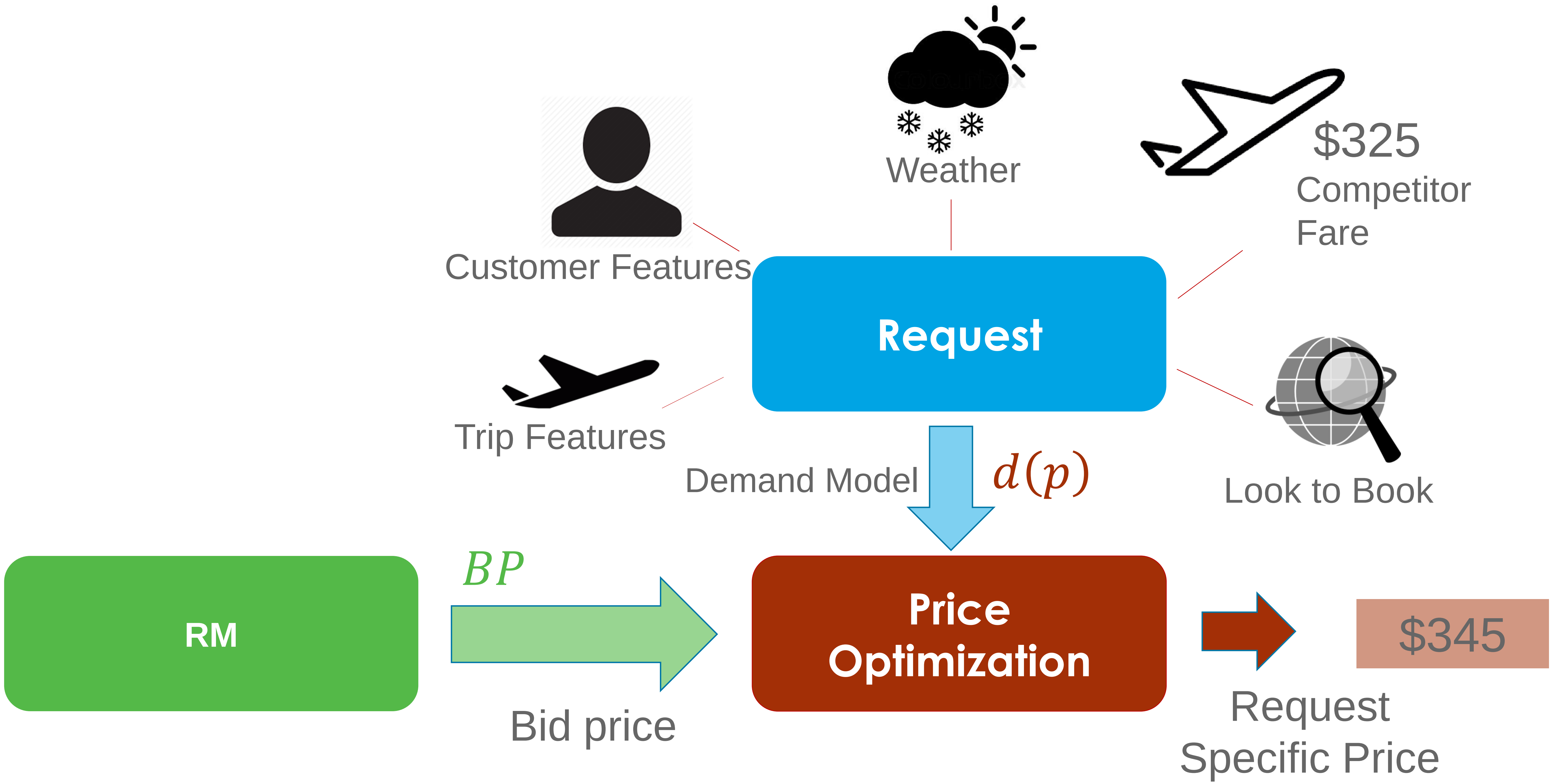
Dynamic/request specific pricing for airlines



Dynamic/request specific pricing for airlines



Dynamic/request specific pricing for airlines

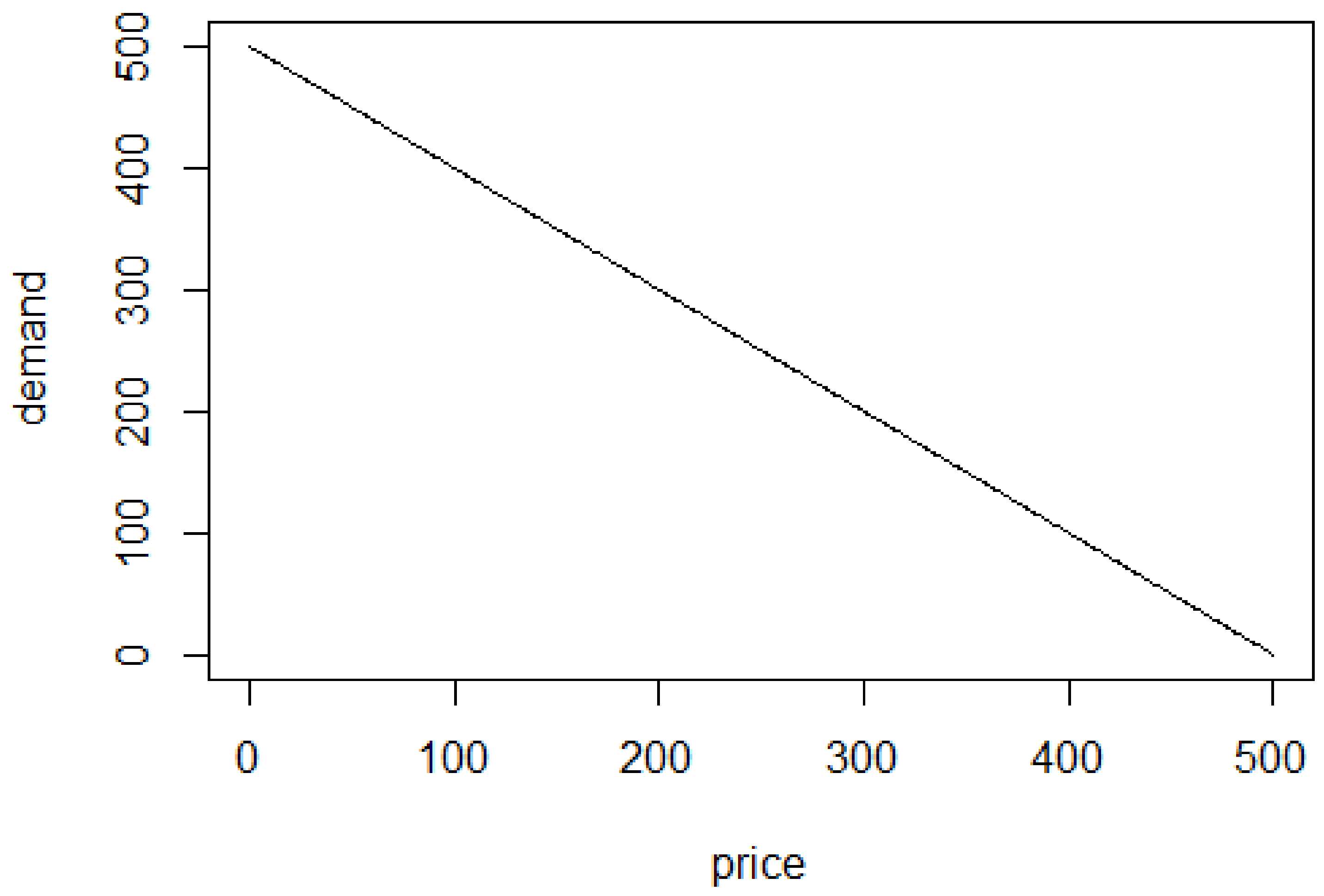


What is request specific pricing?

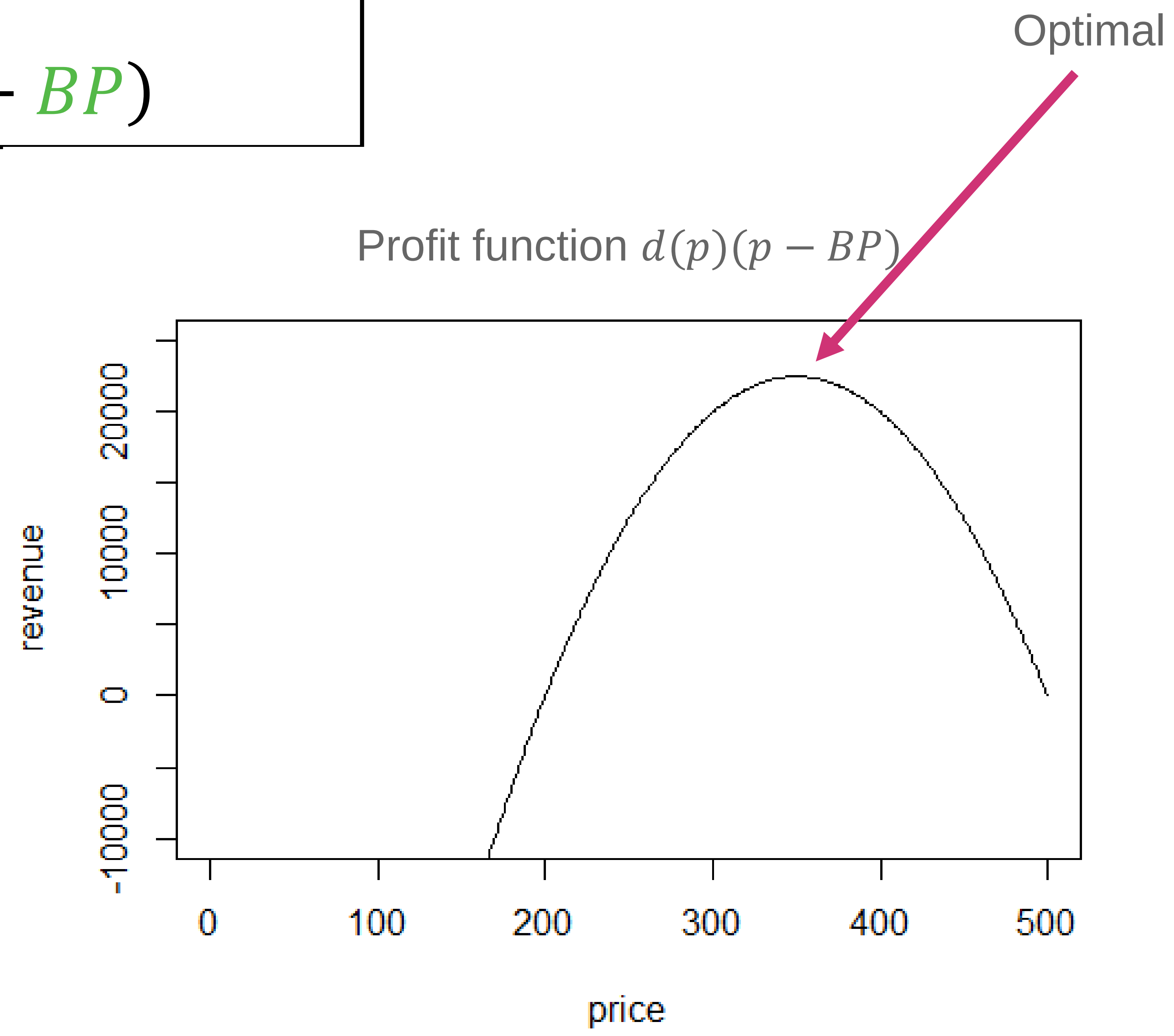
Optimal price:

$$p^* = argmax_p \ d(p) \cdot (p - BP)$$

Demand function $d(p)$



Profit function $d(p)(p - BP)$



Challenges in Practice

Data Sparsity

With so many features, data becomes very sparse

Computation Time

Fitting complex models on big data is time-consuming – potentially intractable

Flat Slopes

Often flat slopes are observed – demand is overestimated and hence prices are overestimated

Ensemble model

Idea:

Instead of fitting
one big model,
fit several smaller models
and ensemble the results
(forecast/prices).

Upside:

- Overcome sparsity issues
- Computation time

Downsides:

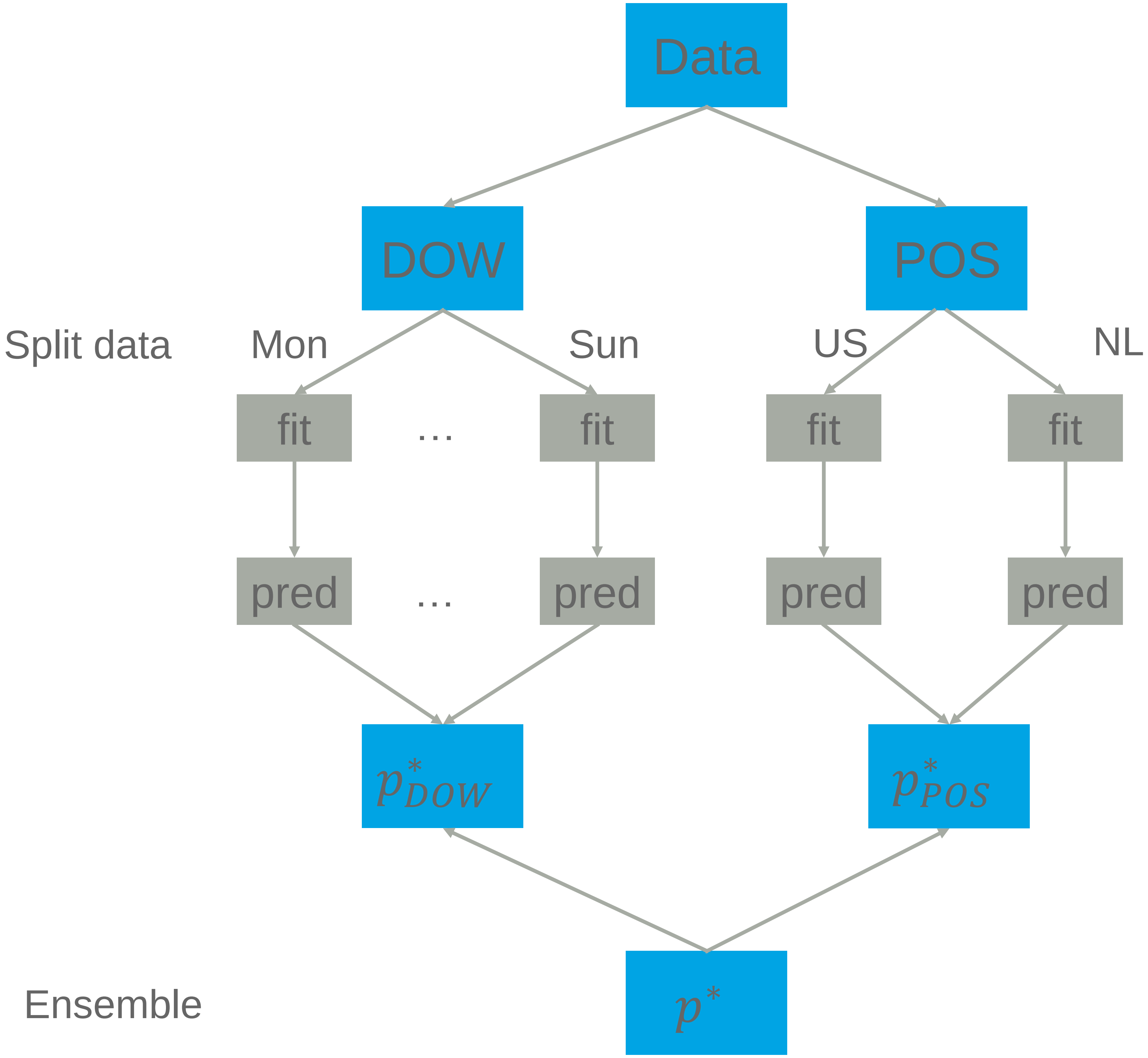
- Potential misspecified model

Example

- Data:
  - Demand D
  - Price p
  - Point of sale (POS): US & NL
  - Day of week (DOW): Mon, Tue, …, Sun

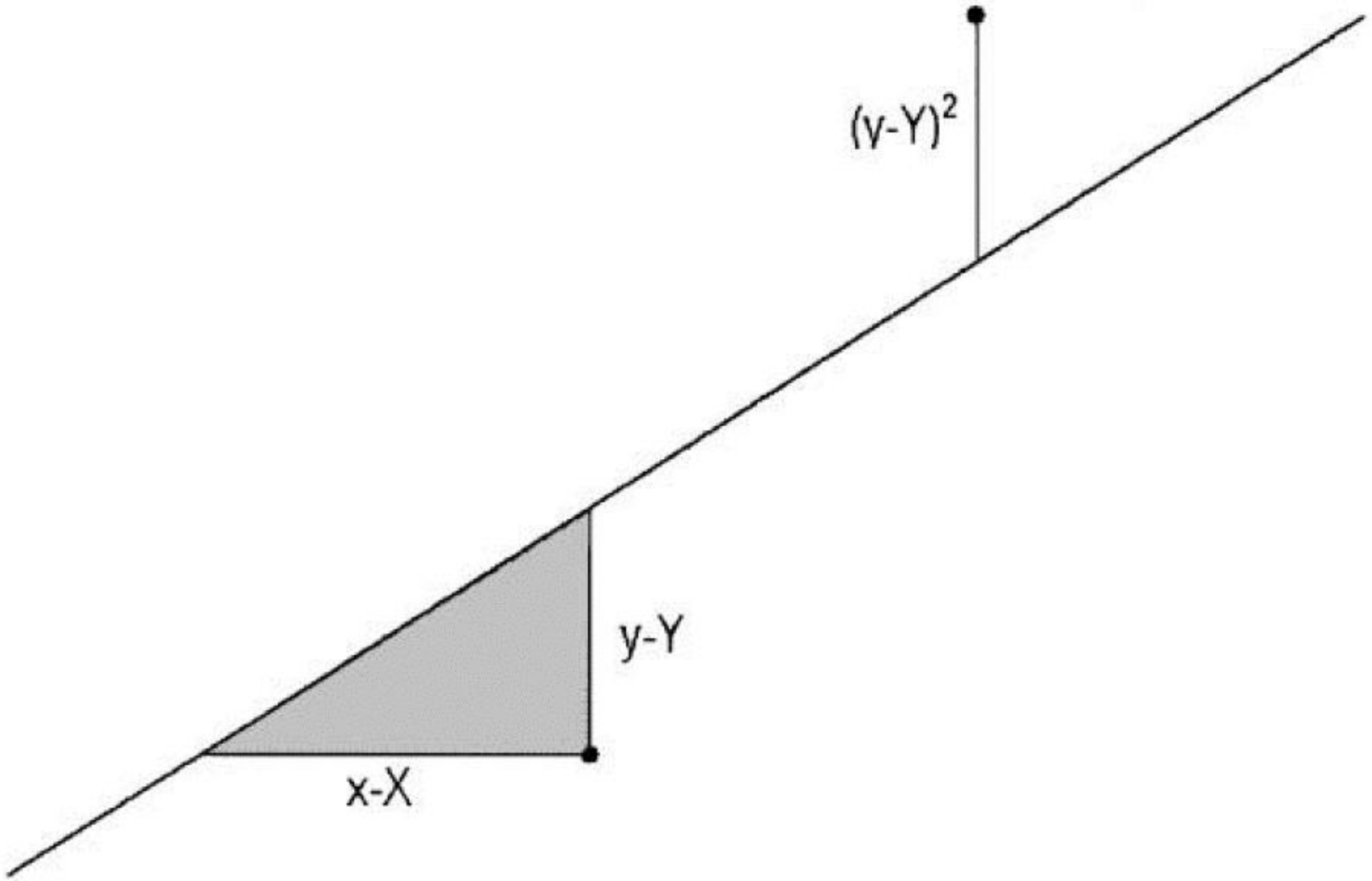
Example

- ‘One model’ on all data:
  - $D = \beta_p p + \sum_{i \in \{NL,US\}} \beta_i \cdot \mathbf{1}_{POS=i} + \sum_{j \in \{Mon,\ldots,Sun\}} \beta_j \cdot \mathbf{1}_{DOW=j} + \beta_0$
- Grand Canyon model on subset (e.g., POS=NL, or DOW=Tue):
  - $D = \beta_p^{POS=NL} p + \beta_0^{POS=NL}$ → $\boldsymbol{p^*_{POS=NL}}$
  - $D = \beta_p^{DOW=Tue} p + \beta_0^{DOW=Tue}$ → $\boldsymbol{p^*_{DOW=Tue}}$
  - $p^* = f(p^*_{POS}, p^*_{DOW})$
  - E.g., $p^* = \frac{1}{2}[p^*_{POS=NL} + p^*_{DOW=Tue}]$



Example

- Any demand model can be used:
  - Linear demand: $D = ap + b$
  - Log-linear demand: $D = \exp(ap + b)$
  - (Multinomial) logit models: $P(X = 1) = \frac{\exp(\beta^T x)}{1+\exp(\beta^T x)}$
  - Gradient boosting regression
  - ...



Linear demand – geometric mean of two estimates

- Consider the linear demand function and it’s inverse:
  - $D(p) = ap + b$
  - $p(D) = a'D + b'$
- Use geometric mean of both predictions
- Why would the latter make sense?
  - There is uncertainty in both p and D (in practice)
- Inspired by Geometric Mean Regression or Reduced Major Axis Regression
  - Used when both dependent and independent variables have noise/measurement errors

Data description

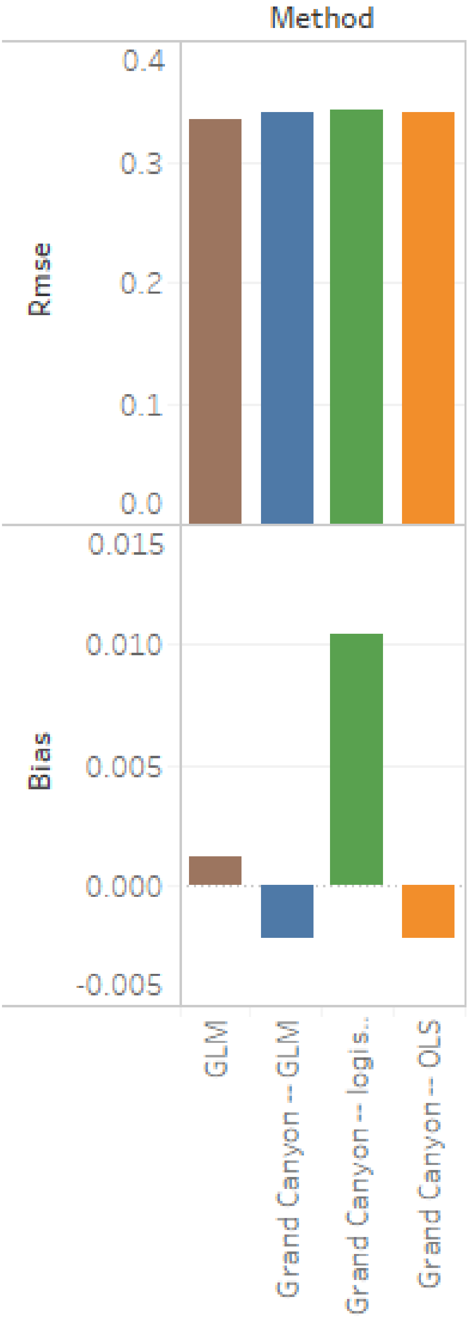
- 1 O&D
- Attributes:
  - Point of sale
  - Days booked before departure
  - Day of week
- Seasonality
- 3 years of data – 2y train/1y test

Solution methods:

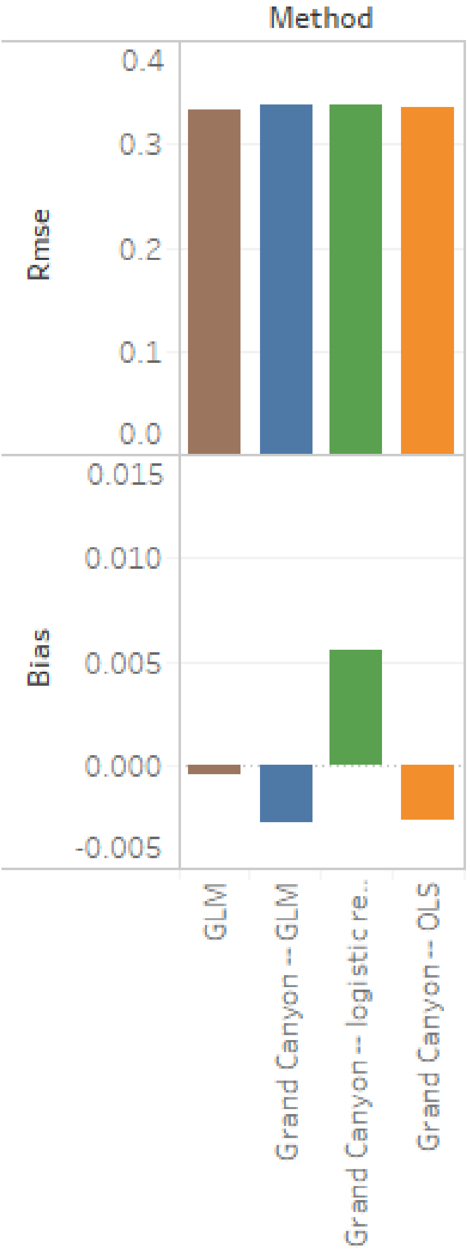
- Ensemble – log-linear
- Ensemble – logit
- Ensemble – geometric mean (linear)
- Generalized linear model

Forecasting

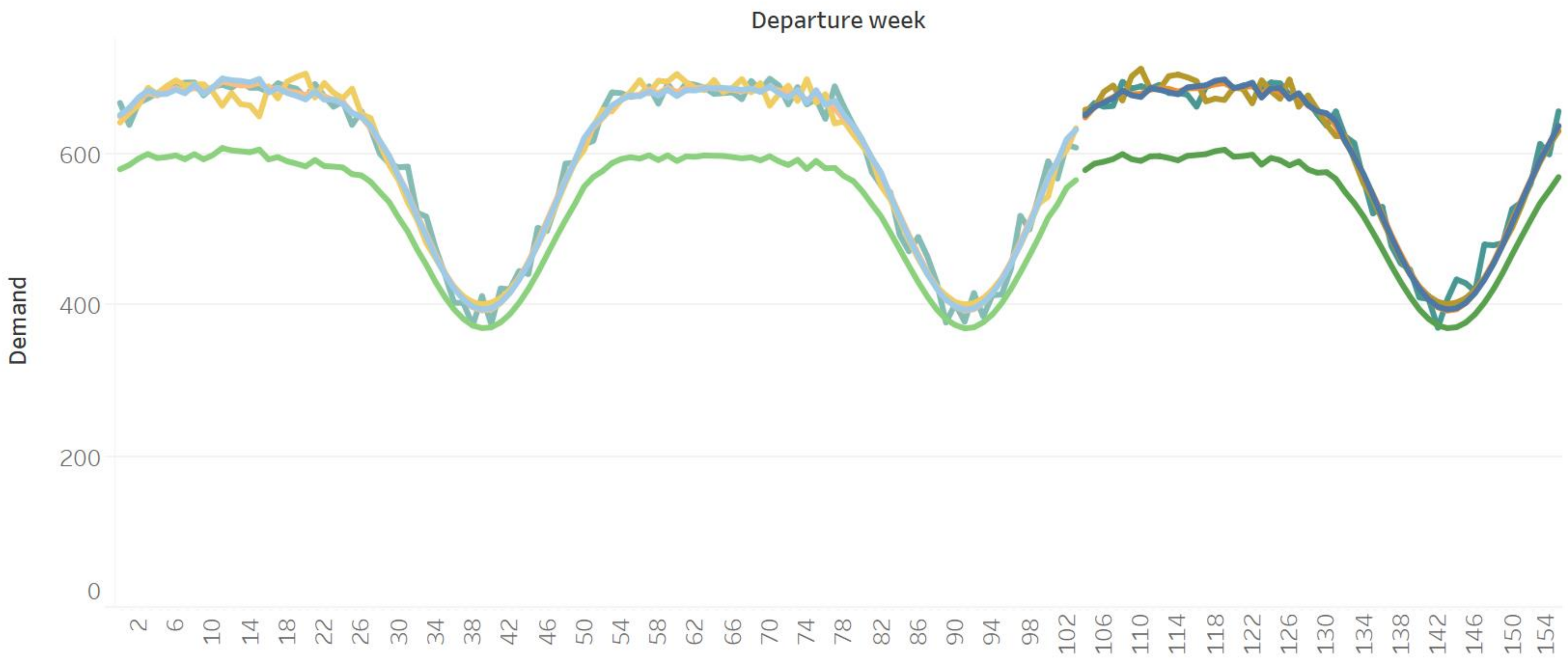
FC accuracy normal demand

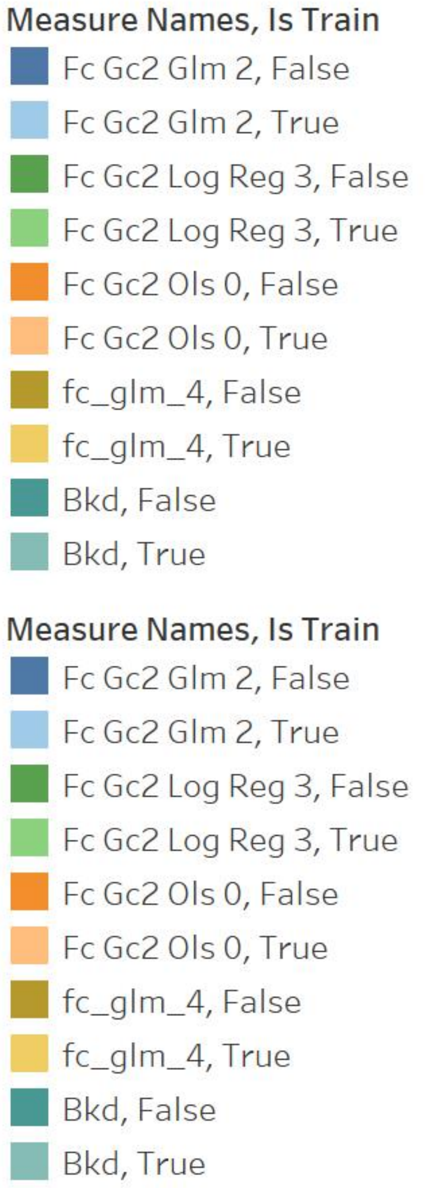


FC accuracy exponential demand

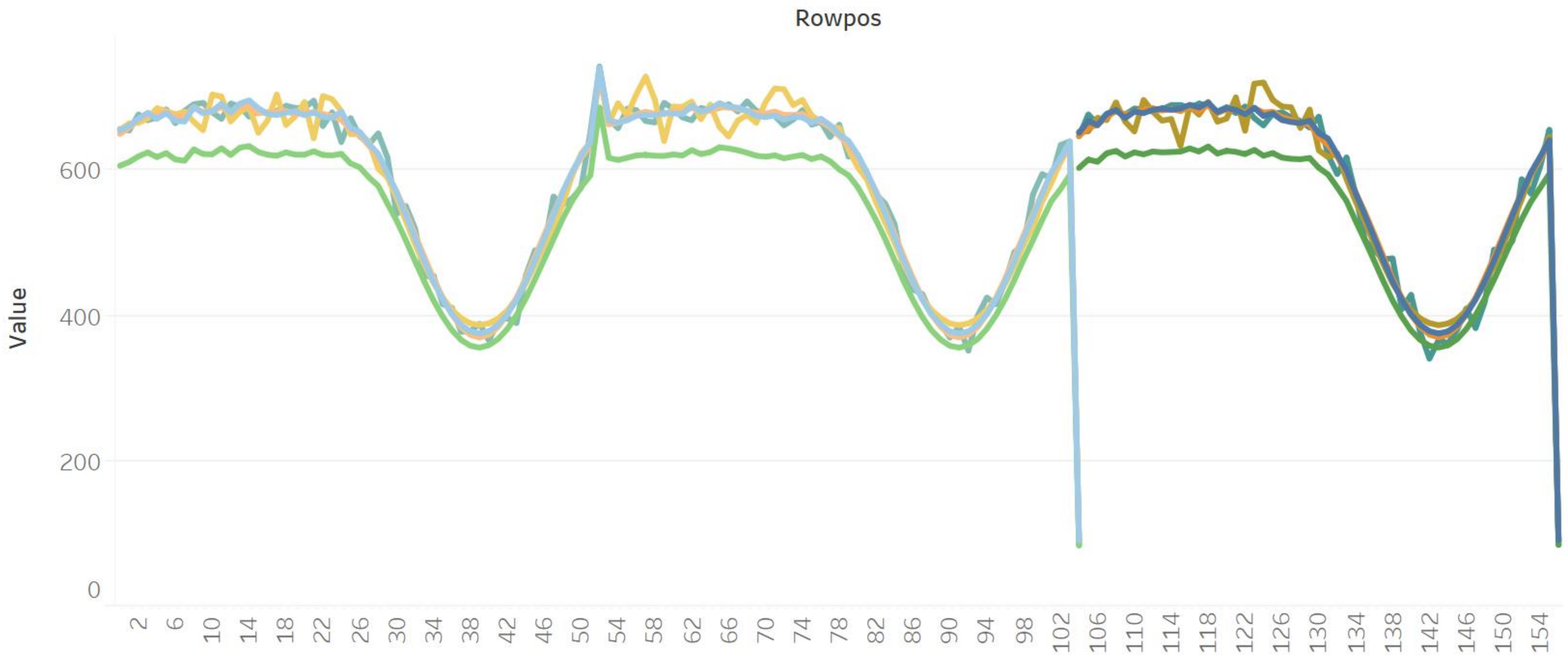


FC -- normal demand





FC -- exponential demand



Forecasting

Observations:

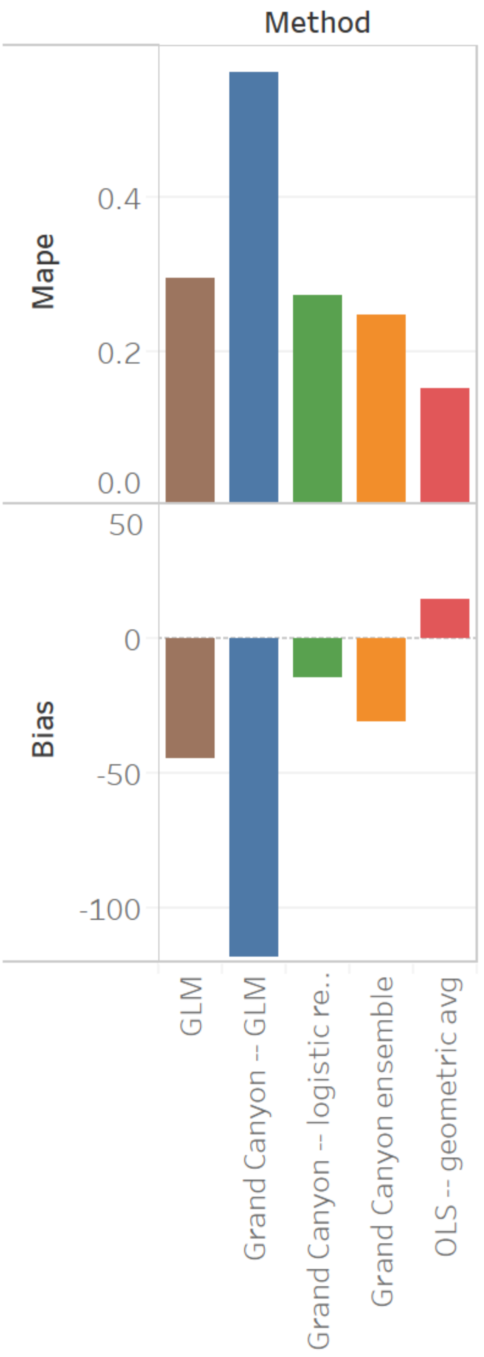
- Forecasts perform well all across the board

Conclusions:

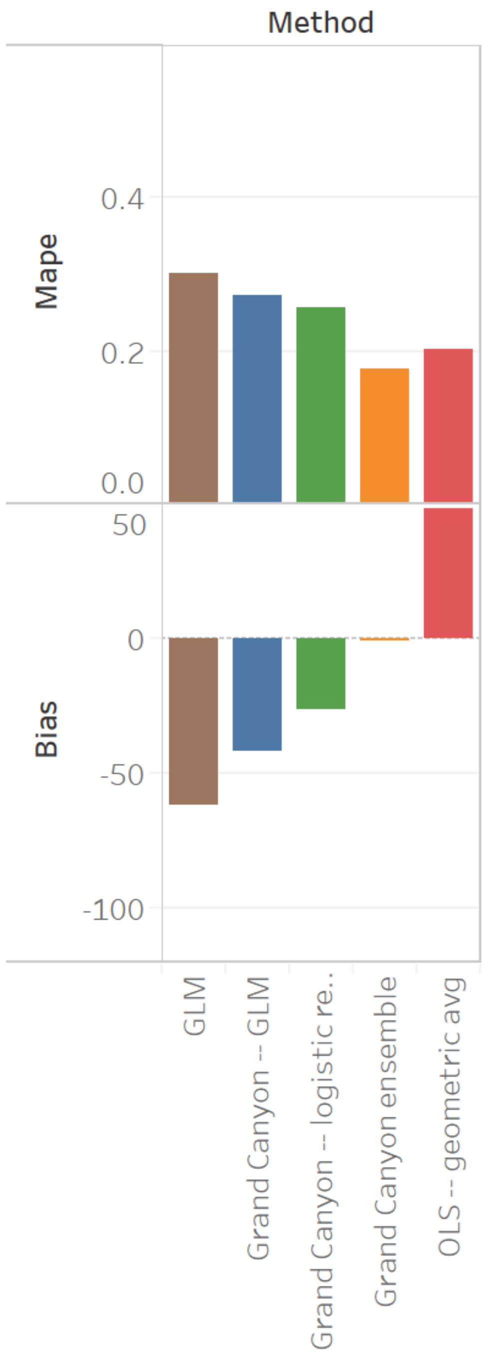
- Ensemble model suitable for forecasting
- Not clear which method (ols, glm) performs the best
- No clear indication which method to choose to do pricing

Pricing

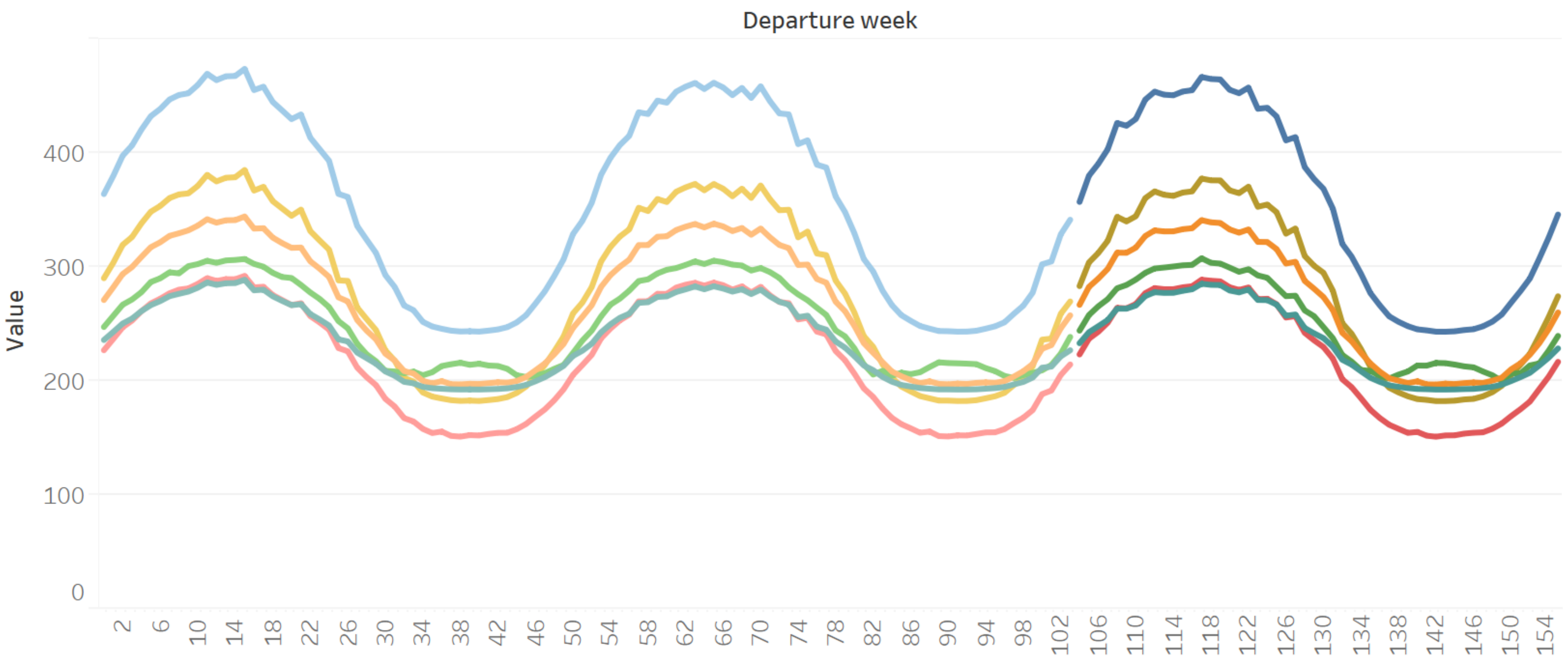
Pricing accuracy normal

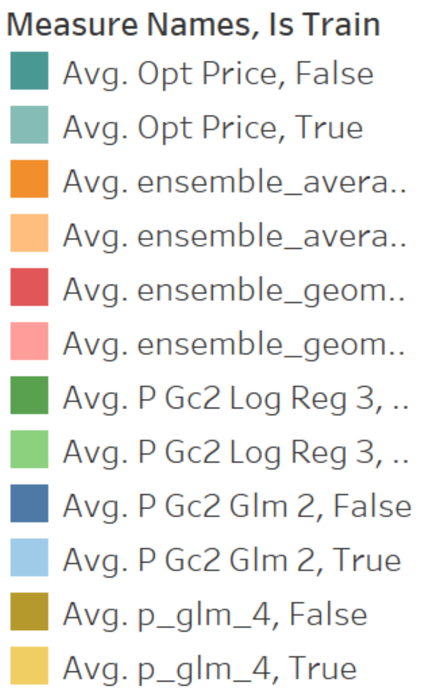


Pricing accuracy exponential

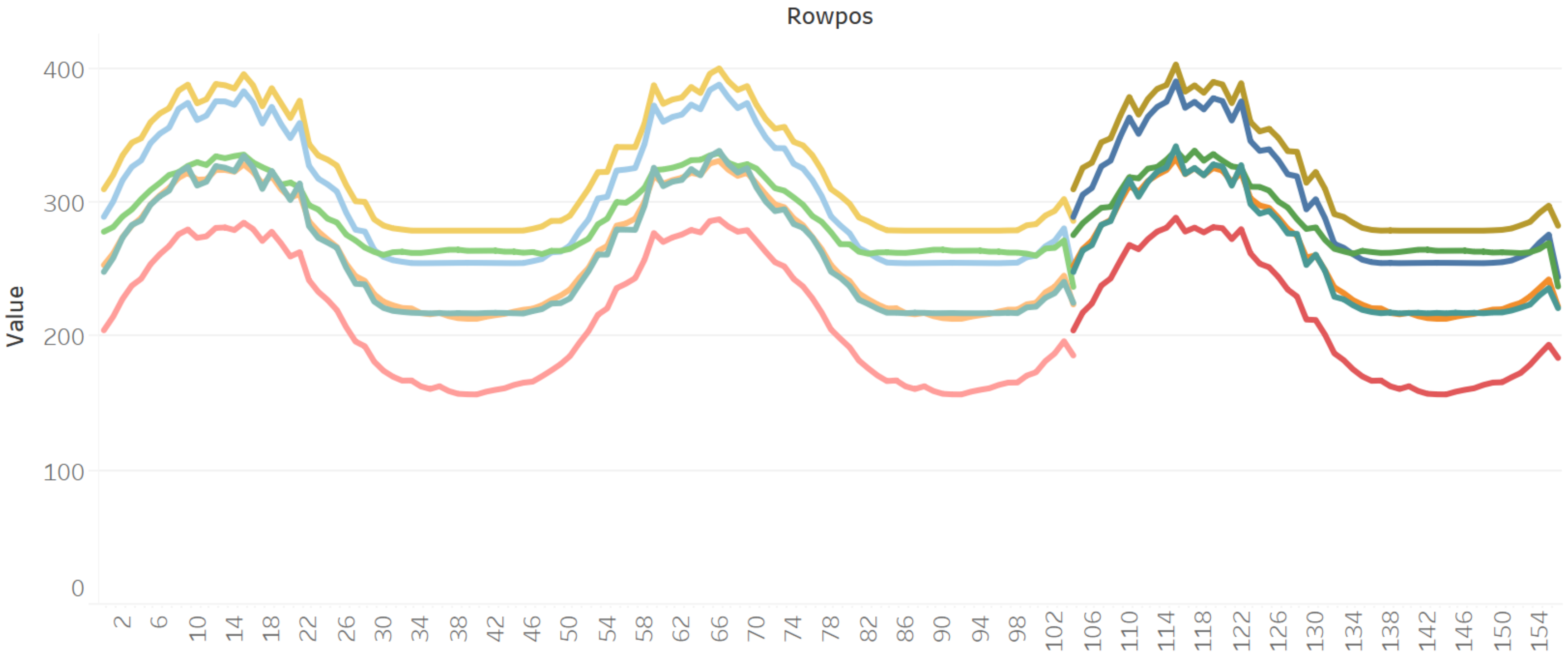


Pricing -- normal demand



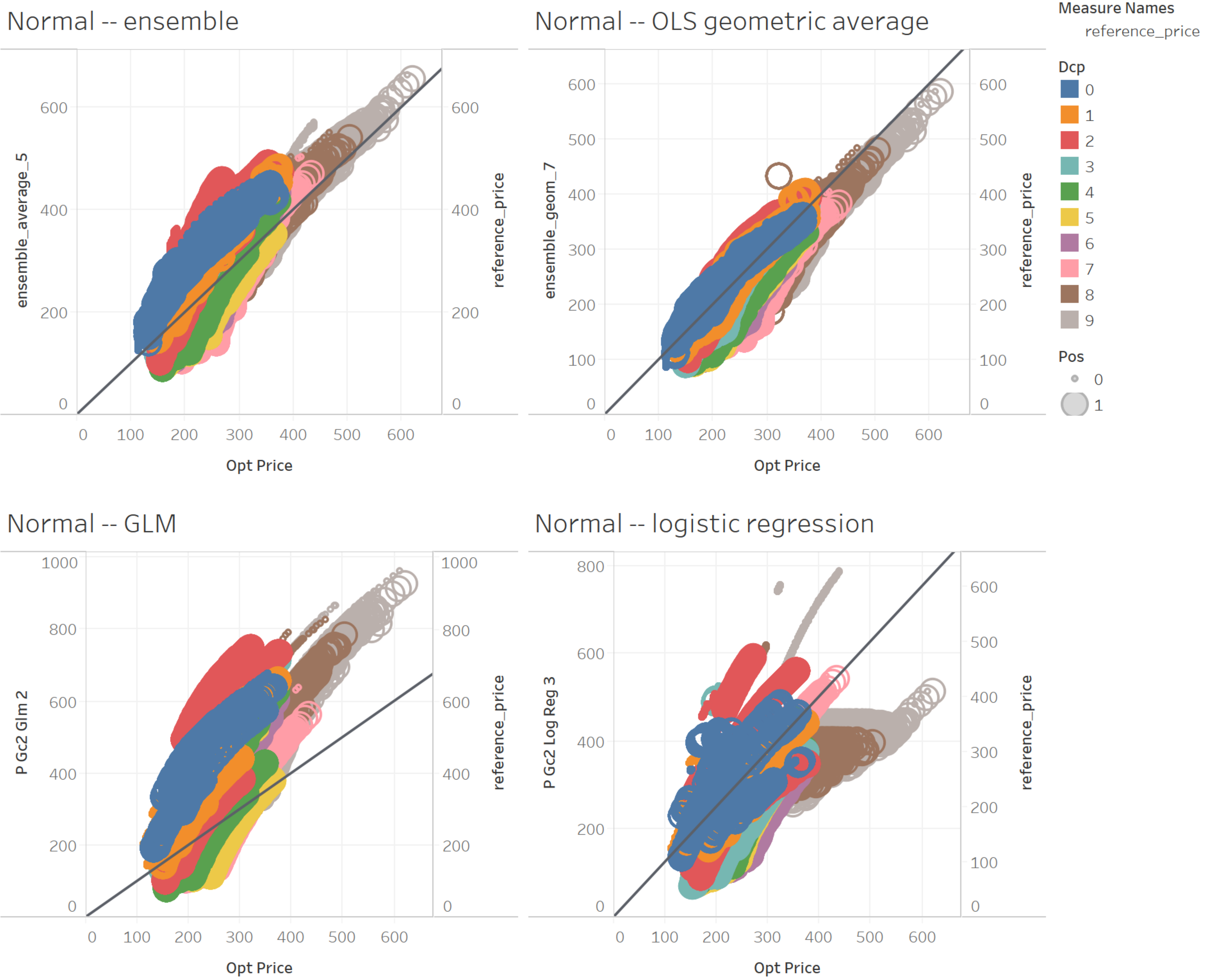


Pricing -- exponential demand



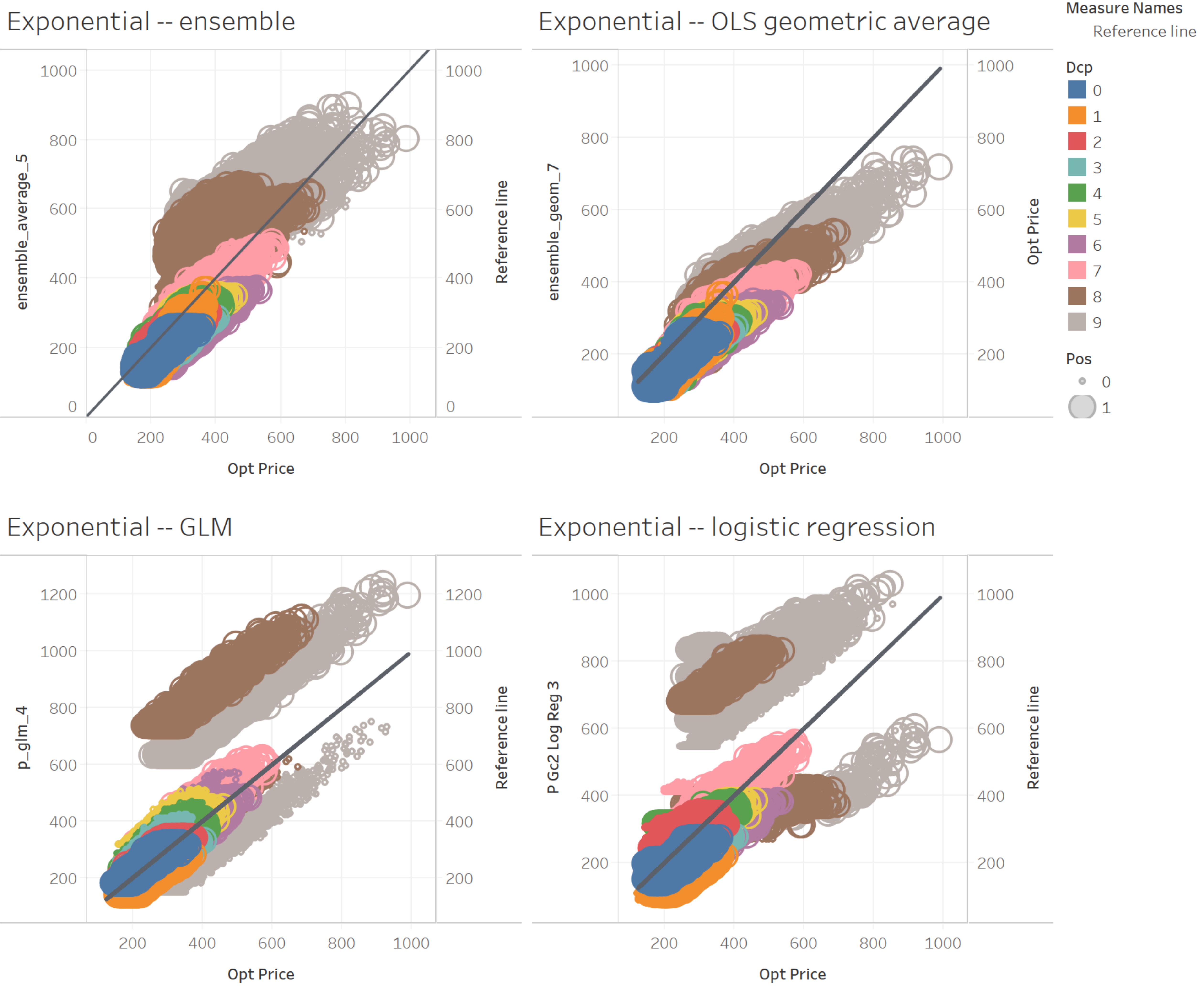
PROS

Pricing



Normal -- ensemble
ensemble_average_5
reference_price
Opt Price
Normal -- OLS geometric average
ensemble_geom_7
reference_price
Opt Price
Measure Names
reference_price
Dcp
0
1
2
3
4
5
6
7
8
9
Pos
0
1
Normal -- GLM
P Gc2 Glm 2
reference_price
Opt Price
Normal -- logistic regression
P Gc2 Log Reg 3
reference_price
Opt Price

Pricing


Exponential -- ensemble
Exponential -- OLS geometric average
Exponential -- GLM
Exponential -- logistic regression
ensemble_average_5
ensemble_geom_7
p_glm_4
P Gc2 Log Reg 3
Reference line
Opt Price
Measure Names
Reference line
Dcp
0
1
2
3
4
5
6
7
8
9
Pos
0
1

Pricing

Observations

- Pricing performance varies a lot across the board
- Ensemble methods seem to do well across all accuracy measures

Conclusions/future research:

- Good forecasts do not imply good prices
- Overpricing is an issue across the board (future research)

Thank You

PROS Headquarters | Houston, Texas | 3100 Main Street, Suite 900 | Houston, TX 77002, USA | +1-713-335-5105 +1-800-555-3548