AMERICANAS
DESAFIO TÉCNICO

Felipe Lourenço Angelim Vieira

https://www.linkedin.com/in/felipe-angelim/

THE DATA

1. Transaction data for 9 different products
2. Data of competitor's prices

Can we create a model to predict the sales given a price?

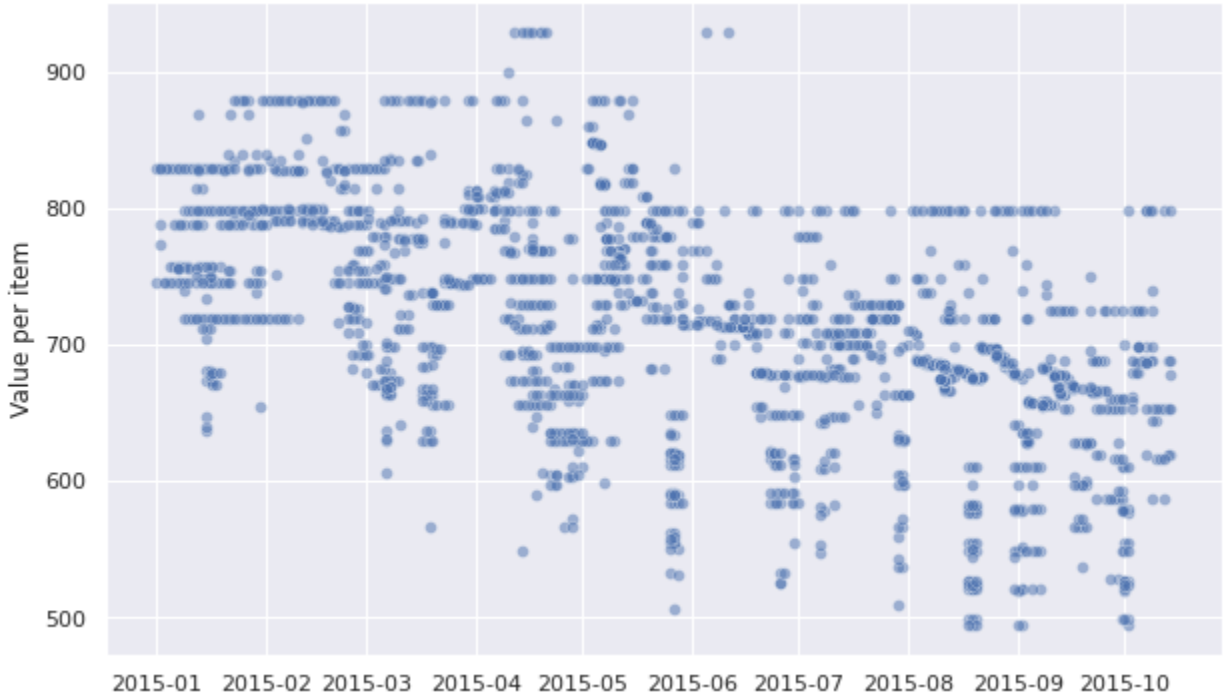
THE DATA

- We may aggregate the number of orders by average value per item
- Cheap offers are, often, not the most sold one.
- That may be related to out-of-stock products, that we may treat as noise.

	date_order	prod_id	value_per_item	qty_order
9882	2015-10-14	P9	341.10	8.0
9883	2015-10-14	P9	344.89	1.0
9884	2015-10-14	P9	356.56	1.0
9885	2015-10-14	P9	360.05	16.0
9886	2015-10-14	P9	379.00	557.0
9887	2015-10-14	P9	388.65	3.0
9888	2015-10-14	P9	399.99	1.0
9889	2015-10-14	P9	449.00	3.0

THE DATA

- Price per item can vary a lot during the same day or week

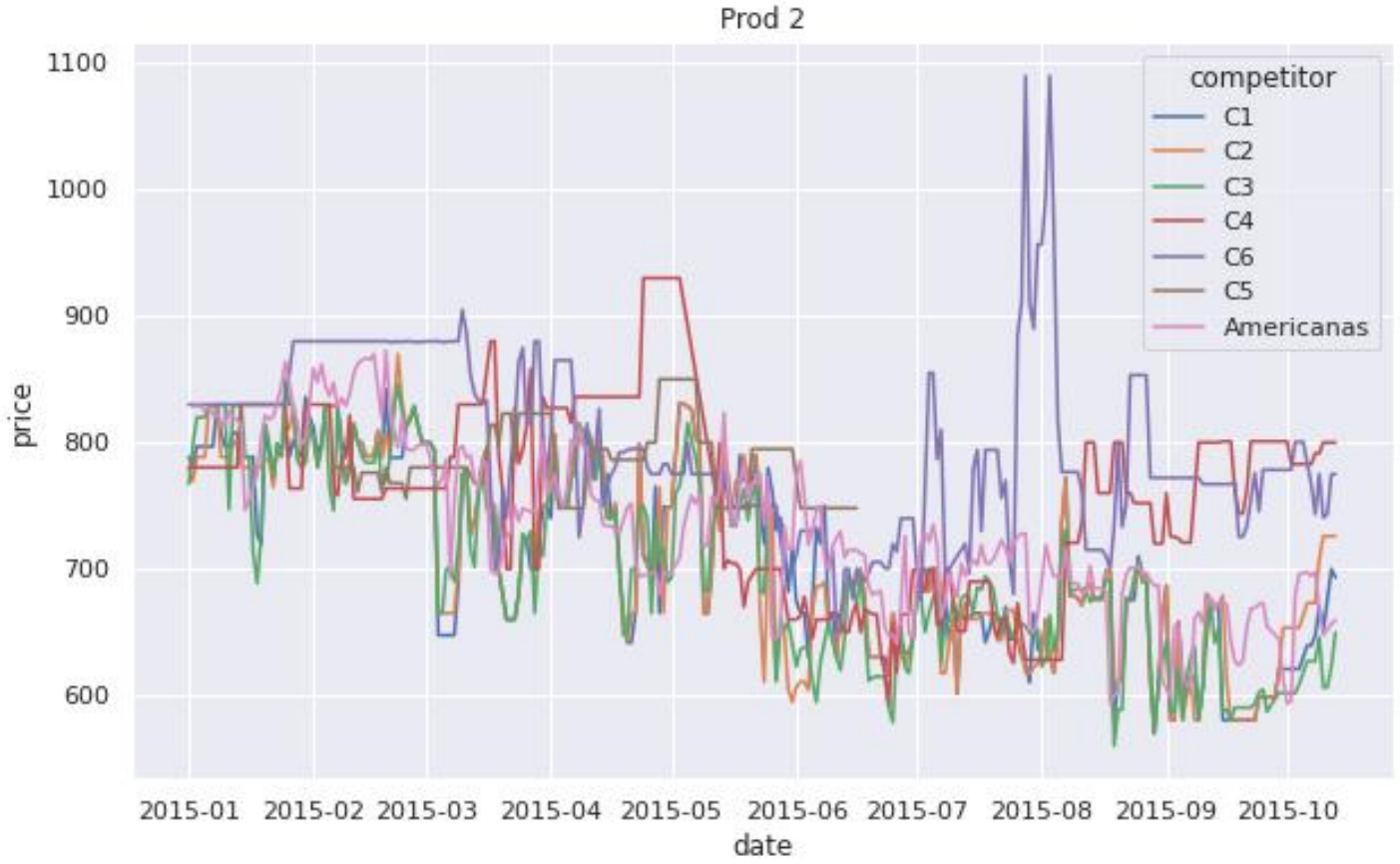


THE DATA

- We may aggregate the number of orders by average value per item
- Often, cheap offers aren’t the most sold one.
- That may be related to out-of-stock products, that we may treat as noise.

Solution: compute the average price per item in a day weighting by the number of sales

PRICE TIME-SERIES FOR PRODUCT “P2”

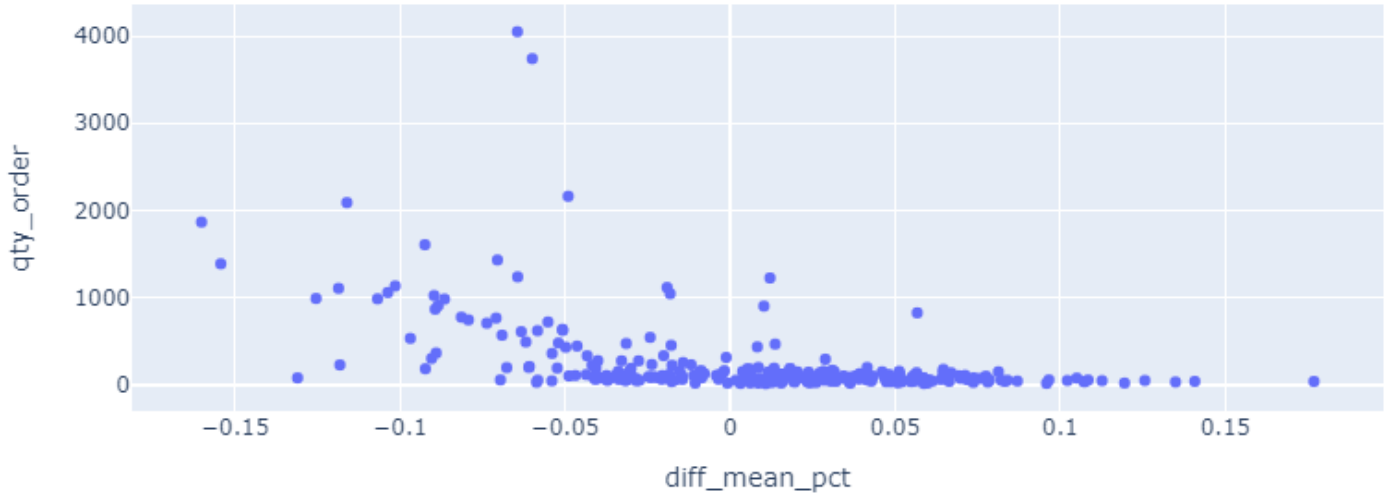


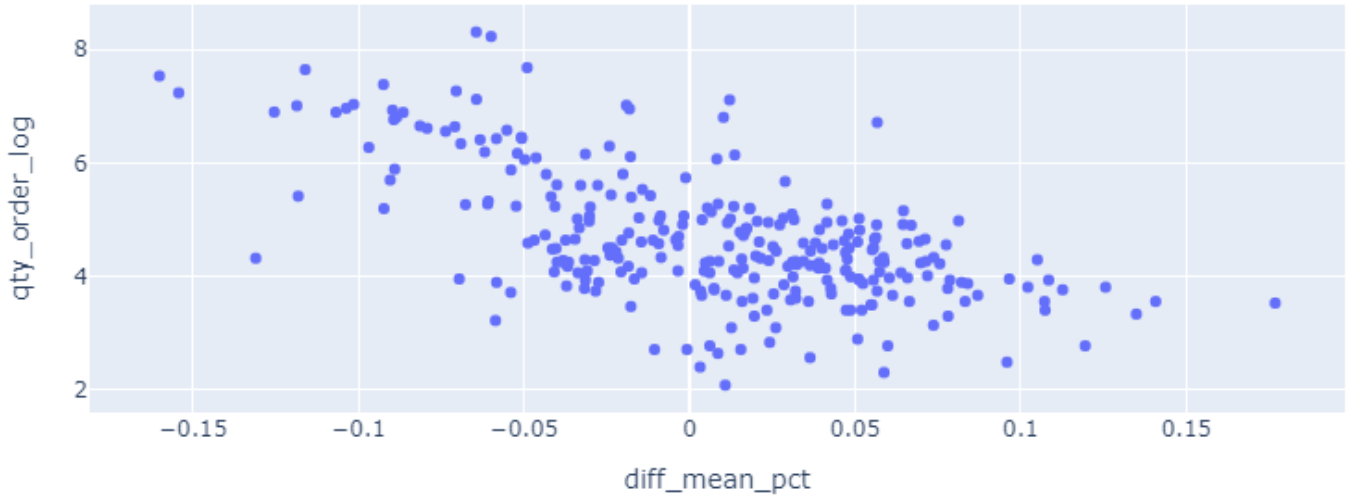
FEATURE ENGINEERING

- Features:
  - The number of total sales 7 days before (log scale)
  - The percentage difference between the company's price and the average price of the competitors

- Target:
  - **The log of the number of sales in a day.**

FEATURE ENGINEERING





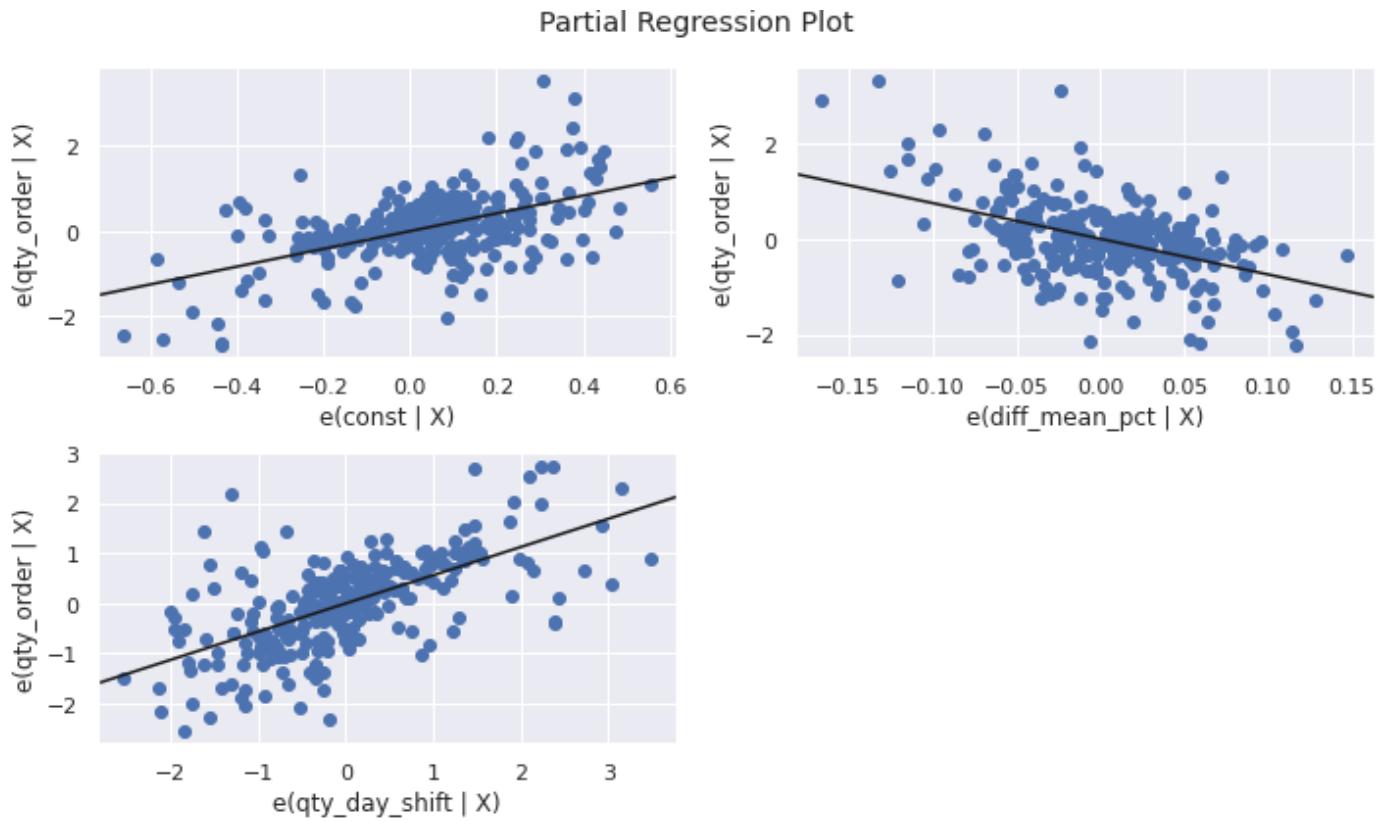
THE MODEL

We don’t have lots of data.
After cleaning, we have about 2000 observations for all 9 products.

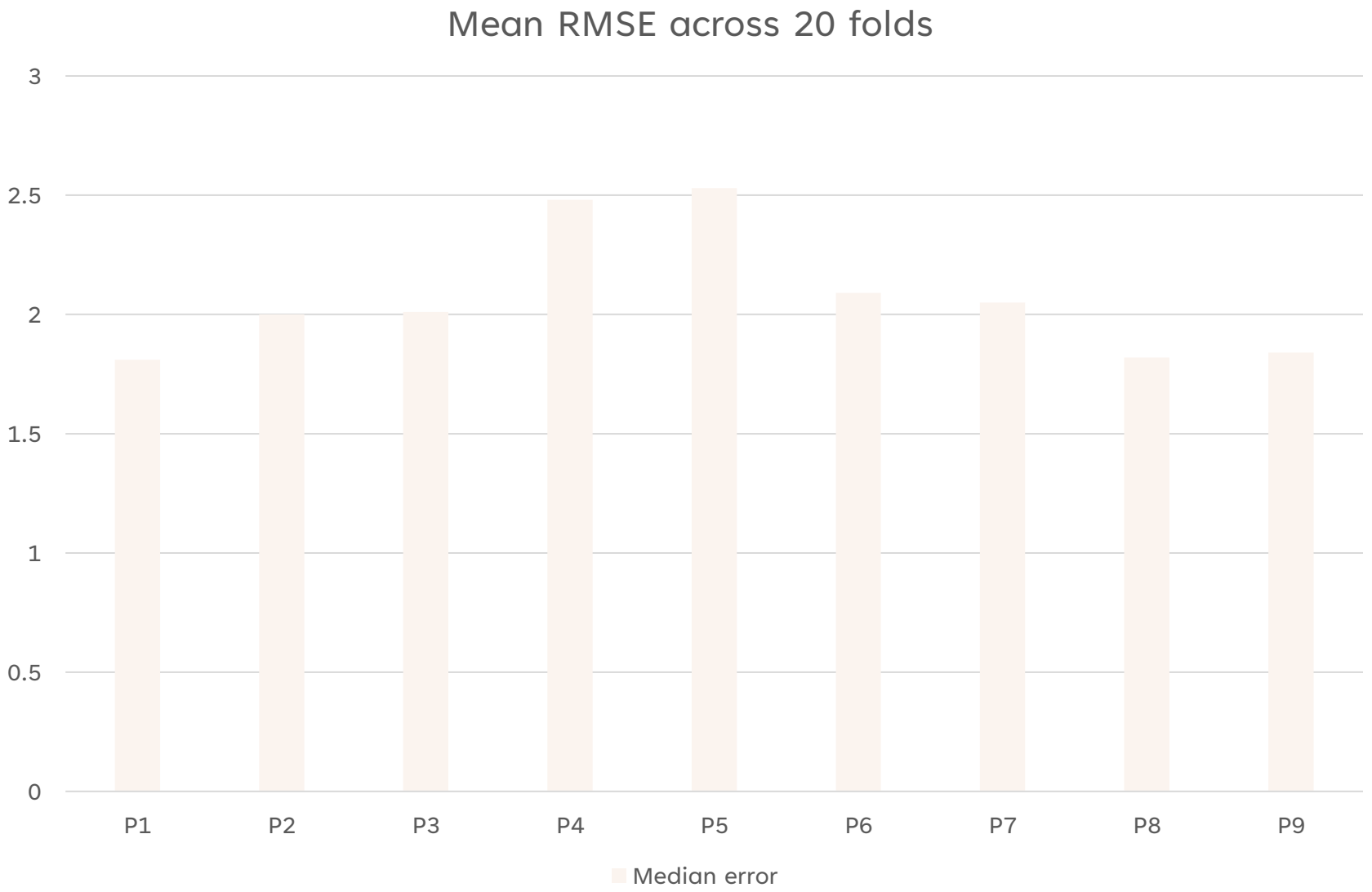
We’ll fit a linear regression model.

- Easy to understand and simulate the impacts of price policies
- No overfitting

THE MODEL


Partial Regression Plot
e(qty_order | X)
e(const | X)
e(diff_mean_pct | X)
e(qty_day_shift | X)

CROSS-VALIDATION MEAN ERROR



THANK YOU