historical data

- time   *hourly data*
(- duration of usage)
(- sales)
(- population of the city)

  - season
✓ - major events
  (- demographics)
  (- appreciation)
  - ~~stations usage~~
  - weather
  → traffic
  - fuel prices

ML: Loss function → for optimizing the algo

→ evaluation for model selection

- MSE · symmetric (!)

algo → MSLE

split data   ? ratio 20%

how important is the time-dependence?

10'00
- magnitude of outcome? (log might help)
- benchmark   avg   Post test result on Socrative

DATA → ML → ☁

$f(X)$

$f(x) = \bar{y}_{train}$

Linear Regression:

→ $y = \beta_0 + \varepsilon$ → $\hat{\beta}_0 = \bar{y}$

→ x: holiday → x0 : x = 0 β_0
                  x1 : x = 1 $\beta_0 + \beta_1$

$y = \beta_0 + \beta_1 x_1 + \varepsilon$