# sRecruit Restaurant Visitors

2157 competitors

## Entities:

**Train**

Store ()

**Test**

Store ()

## Data:

Date (DoW)

Holiday flag

Store

Store info (genre, area, latitude, longitude) not always available

Reservations (store, time, reserve\_time, visitors)

**Target:** Visitors by store/date

Benchmarks:

Median visits (0.62392)

## Takeaways:

* Holiday handling important
  + Small trick of treating holidays as Saturday
  + Days before/after as Friday/Monday
  + Gives pretty good perf boost

## Notable write-ups

1st Place Team PPPP

* Plantsgo approach:
  + 3 models with same features
    - Predict 14 days using 14\*30 samples
    - Predict 28 days using 28\*15 samples
    - Predict 42 days using 42\*10 samples
  + Features:
    - Aggregations
      * Funcs: Sum, KernelMedian, Count, Mean, Min, Max, Std
      * Group: Store/Holiday, Store/Weekday, Store/Date, Area/Genre/Holiday, Area/Holiday, Genre/Weekday, Genre/Holiday
      * Targets: Visitors, Reservations
  + 5-fold CV for training
  + Last 42 days as validation
  + LGBM
* Fakeplastic trees:
  + LGBM / XGB
  + 39 models one for each pred day.
  + Validated on final 80 days of train
  + Features:
    - Lagged values up to 80 days of visitors
    - Lagged values up to 80 days of reservations
    - # reservations
    - Holiday dummies for prev 30 days up to +2 days
    - Lat, longitude
    - Region, area, genre
    - DoY, DoM, DoW
* Feiyang Pan
  + NN FC 3 layer
  + Relu, 128, 64, 32 hidden size
  + Features similar to plantsgo
* Piupiu
  + LGBM
  + Features similar to others:
    - Aggregations of statistics by groups, including exp. Smoothing

5th Place Danijel Kivaranovic

* LGB and Keras
* No holiday adjustments
* Similar features to Shize Zu
* Lag 20 mean weekly visitors
* Lag 8 mean monthly visitors

7th Place Michael Maguire

* Time series CV -> 6 week blocks
* Also used 5-fold CV for some of models
* Features
  + Aggregations
    - Mean, std reservations and “conf. interval” by store/city
    - Weekday rolling averages on reservations
    - Weekday rolling average on visitors instead in some of ensemble models
  + Competitor counts (comp. within 800 m)
  + Holiday flag
* Ensemble: NN to combine models
  + LGBM, XGBoost

8th Place Max Halford

* Features
  + Holiday lag/lead
  + Weather: Avg temp & precip.
  + Temporal
    - DoM, Weekend
  + Exp smoothing
    - Optimized alpha
    - Done by store/weekday, store/weekend
    - Done on both logarithm and raw
  + Aggregations
    - Funcs: Mean, median, std, count, max, min, exp. Smooth(0.1, 0.25, 0.3, 0.5, 0.75 alpha)
    - Group: Store/Dow, Store/Weekend, Store
    - Target: Visitors, Log1p Visitors
* Predicted log transform
* Removed outliers
  + Assumed normal dist per restaurant, used 2.4 as cut-off and removed with max of non-outliers
* 6-Fold CV
* Important features:
  + DoM
  + Visitor count by store/weekend
  + Optimized exp. Smooth by store/dow
  + ….

10th Place Jerome Vallet

* Ensemble of XGB, LGBM and CatBoost
* 4-fold CV
* Test set 5 weeks in end (2017-03-15 to 2017-04-22)
* Used CV + Test set perf as guide for model improvements
* Features:
  + Window aggregations:
    - Exp. Weighted averages
    - Means
    - Sum
    - Std
    - Max
    - Count
  + By: Genre, Visit, Dow,
  + Weather
    - Avg. temp
    - Sunlight

11th Place Xiuqi

* Ensemble of LGBM, XGB, RF, Prophet
* Time based split, 2 testsets:
  + Last 33 days
  + 33 days over new year to capture holiday performance
* Features
  + Surpise me kernel +
  + Min, Max, Median, Mean, Count for Visitors
  + Rolling weighted mean
  + Holiday counters
  + Weather: Rain and avg. temp

12th Place pocket

* LGBM
* CV:
  + 4-fold TSCV
    - 7 day prediction window
* Features
  + Holiday adjustments
    - Set dow of holidays to Sat/Sun
    - Next day is holiday Boolean
    - Before/after holiday
  + Temporal
    - Dow
    - Month
    - Year
    - Holidays in next/prev week
  + Lagged reserve data
  + Aggregations (1, 5, 15, 55 weeks)
    - Mean, median, max, skewness
    - Group: storeId, storeId/dow
    - Rolling medians ratios
    - Exponential weighted mean
  + Regression for trend
    - One for each month
    - Reg by store id
    - DoW, Time-trend as features
    - Ridge/Lasso ensemble
  + Weather data
    - Precip
    - Avg. temp
* 5 models, one for each week
* Removed outliers