Restaurant Revenue Prediction

Predict annual restaurant sales based on objective measurements

$30,000 · 2,257 teams · 2 years ago

Competition Data

Edit

sampleSubmission.csv

test.csv.zip

train.csv.zip

train.csv.zip 4.54 KB

Download

Data Description

TFI has provided a dataset with 137 restaurants in the training set, and a test set of 100000 restaurants. The data columns include the open date, location, city type, and three categories of obfuscated data: Demographic data, Real estate data, and Commercial data. The revenue column indicates a (transformed) revenue of the restaurant in a given year and is the target of predictive analysis.

File descriptions

- **train.csv** - the training set. Use this dataset for training your model.
- **test.csv** - the test set. To deter manual "guess" predictions, Kaggle has supplemented the test set with additional "ignored" data. These are not counted in the scoring.
- **sampleSubmission.csv** - a sample submission file in the correct format

Data fields

- **Id** : Restaurant id.
- **Open Date** : opening date for a restaurant
- **City :** City that the restaurant is in. Note that there are unicode in the names.
- **City Group**: Type of the city. Big cities, or Other.
- **Type**: Type of the restaurant. FC: Food Court, IL: Inline, DT: Drive Thru, MB: Mobile

- **P1, P2 - P37**: There are three categories of these obfuscated data. **Demographic data** are gathered from third party providers with GIS systems. These include population in any given area, age and gender distribution, development scales. **Real estate data** mainly relate to the m2 of the location, front facade of the location, car park availability. **Commercial data** mainly include the existence of points of interest including schools, banks, other QSR operators.
- **Revenue**: The revenue column indicates a (transformed) revenue of the restaurant in a given year and is the target of predictive analysis. Please note that the values are transformed so they don't mean real dollar values.