

# Intro. to Kaggle

Making Your First Competition Submission

03-03-2021



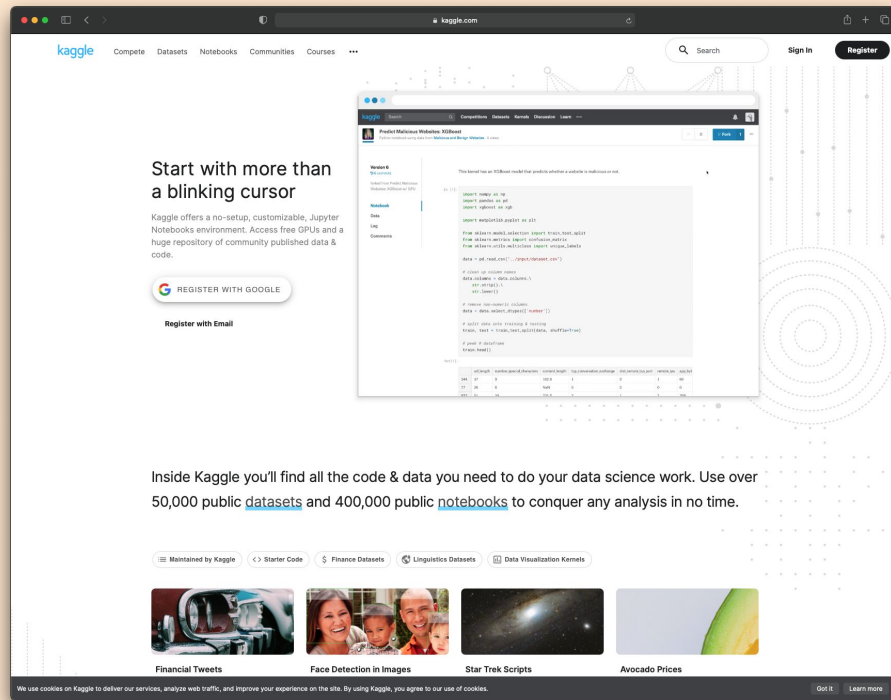
# Sign in sheet

## Important links

- Sign in here: <sign in link>
- Workshop Repository: <https://github.com/Davis-Data-Science-Club/kaggle-workshops>
- Main Website: <https://kaggle.com>
- Example Competition: <https://www.kaggle.com/c/restaurant-revenue-prediction/data>

# What is Kaggle

1. Free data science playground.
2. In browser code editor.
3. GPU support
4. A place where companies can host a dataset and problem for the public to solve.
5. Free data science courses
  - a. Data transformation
  - b. Numpy, pandas, matplotlib



# Competitions

1. Presented problem and data by company.
2. Leaderboard (Sometimes open notebook)
3. Public notebooks with voting and discussion
4. You can attempt a score even after competition is over.

The screenshot displays the Kaggle website interface for the 'Restaurant Revenue Prediction' competition. The left sidebar contains navigation links: Home, Compete, Data, Code, Communities, Courses, More, User Rankings, Tags, Blog, Documentation, Progression, Host a Competition, Support/Contact, Community Guidelines, Team, Terms, and Privacy. Below these are 'Recently Viewed' items and a 'View Active Events' link. The main content area features a search bar, a 'Featured Prediction Competition' banner for 'Restaurant Revenue Prediction' with a \$30,000 prize, and tabs for Overview, Data, Code, Discussion, Leaderboard, Rules, and Team. The 'Data Description' section explains the dataset provided by TFI. The 'File descriptions' section lists training and test files. A terminal window shows the command to download the competition data. The 'Data Explorer' section shows the 'sampleSubmission.csv' file (1.52 MB) with a table view of its contents. The 'Summary' section shows 3 files and 2 columns. A 'Download All' button is at the bottom.

**Restaurant Revenue Prediction**  
Predict annual restaurant sales based on objective measurements  
\$30,000 Prize Money  
2,257 teams · 6 years ago

Overview **Data** Code Discussion Leaderboard Rules Team My Submissions **Late Submission**

### 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 defer 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

```
> kaggle competitions download -c restaurant-revenue-prediction
```

### Data Explorer

3.97 MB

`sampleSubmission.csv`  
test.csv.zip  
train.csv.zip

### Summary

3 files  
2 columns

**Download All**

### sampleSubmission.csv (1.52 MB)

Detail	Compact	Column
41		4453532.6
42		4453532.6
43		4453532.6
44		4453532.6
45		4453532.6
46		4453532.6
47		4453532.6
48		4453532.6
49		4453532.6

# Solution Submission

1. Leaderboard (Sometimes open notebook)
2. Public notebooks with voting and discussion
3. You can attempt a score even after competition is over.

The screenshot shows the Kaggle website interface for the 'Restaurant Revenue Prediction' competition. The page is titled 'Restaurant Revenue Prediction' with a subtitle 'Predict annual restaurant sales based on objective measurements'. It indicates a prize money of '\$30,000' and that the competition was '2,257 teams · 6 years ago'. The navigation bar includes links for 'Overview', 'Data', 'Code', 'Discussion', 'Leaderboard', 'Rules', 'Team', 'My Submissions', and a 'Late Submission' button. A terminal-like input field shows the command: `kaggle competitions submit -c restaurant-revenue-prediction -f submission.csv -m "Message"`. Below this, there's a section for 'Step 1: Upload submission file' with a dashed box and an upload icon. It specifies the 'File Format' as CSV or zip/gz/rar/7z and the 'Number of Predictions' as 100,000 rows. 'Step 2: Describe submission' includes a rich text editor with a toolbar and a 'Make Submission' button at the bottom.

# Social Media

- Check us out on social media
  - Discord: <https://discord.com/invite/TvEc7ca>
  - Facebook: <https://www.facebook.com/data.ucd/>
  - Github: <https://github.com/Davis-Data-Science-Club>
  - Instagram: <https://www.instagram.com/data.ucd/>
- Next events:
  - 
  -