

Para la descarga del dataset, tras registrarse en Kaggle, se ha comprobado las diferentes competiciones que se encontraban activas.

The screenshot shows the Kaggle homepage with the search bar set to 'Search competitions'. The sidebar on the left includes links for 'Competitions' (which is selected), 'Datasets', 'Models', 'Code', 'Discussions', 'Learn', and 'More'. The main content area displays several active competitions:

- Titanic - Machine Learning from Disaster**: Knowledge, Ongoing. Start here! Predict survival on the Titanic. Getting Started. 16017 Teams.
- House Prices - Advanced Regression Techniques**: Knowledge, Ongoing. Predict sales prices and practice feature engineering, RFs, and gradient boosting. Getting Started. \$207 Teams.
- Spacexship Titanic**: Knowledge, Ongoing. Predict which passengers are transported by ship. Getting Started. 1579 Teams.
- AI Mathematical Olympiad - Progress Prize 2**: Knowledge, Ongoing. Solve national-level math challenges used in the AIMO. Featured: Code Competition. 325 Teams. \$2,117,152, 5 months to go.
- ARC Prize 2024**: Knowledge, Ongoing. Create an AI capable of solving reasoning problems. Featured: Code Competition. 1986 Teams. \$1,100,000, 7 days to go.
- Gemma 2**: Knowledge, Ongoing. Google - Unlock Global Communication with Gemma. Create Gemma model variants for a specific application. Analytics. 1027 Teams. \$150,000, 2 months to go.
- Jane Street Real-Time Market Data Forecasting**: Knowledge, Ongoing. Predict financial market responses using Jane Street's real-time market data. Featured: Code Competition. 1021 Teams. \$120,000, 2 months to go.
- Google - Gemini Long Context**: Knowledge, Ongoing. Demonstrate interesting use cases for Gemini. Analytics. 1 month to go. \$100,000.
- NFL Big Data Bowl 2025**: Knowledge, Ongoing. Help use pre-scrub behavior to predict a team's performance. Analytics. 2 months to go. \$100,000.
- Child Mind Institute - Promoting Internet Use**: Knowledge, Ongoing. Reduce Physical Activity to Problematic Internet Use. Featured: Code Competition. 1900 Teams. \$60,000, 2 months to go.
- Edai - Mining Misconceptions in...**: Knowledge, Ongoing. Predict affinity between misconceptions and age groups. Featured: Code Competition. 1024 Teams. \$55,000, 1 month to go.

Se ha buscado entre las activas, una que resultara interesante, eligiendo **House Prices - Advanced Regression Techniques**

The screenshot shows the detailed view of the 'House Prices - Advanced Regression Techniques' competition on Kaggle. The sidebar on the left is identical to the previous screenshot. The main content area includes:

- Overview**: This competition runs indefinitely with a rolling leaderboard. Learn more.
- Description**: Start here if... You have some experience with R or Python and machine learning basics. This is a perfect competition for data science students who have completed an online course in machine learning and are looking to expand their skill set before trying a featured competition. Getting Started Notebook.
- Competition Description**: Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. But this playground competition's dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket fence.
- Competition Host**: Kaggle
- Prizes & Awards**: Knowledge. Does not award Points or Medals.
- Participation**: 878,426 Participants, 5,425 Participants, 5,207 Teams, 24,132 Submissions.
- Tags**: Regression, Tabular, Root Mean Squared Logarithmic Error.
- Table of Contents**: Evaluation, Tutorials, Frequently Asked Q..., Citation.

Para acceder al dataset es necesario unirse a la competición, pulsando sobre el botón JOIN COMPETITION y aceptando las condiciones.

Join Competition

...

## Review competition terms and conditions

By clicking on the "I Understand and Accept" button below, you agree to be bound by the competition rules for [Exploring Mental Health Data](#).

### COMPETITION-SPECIFIC RULES

*The Competition named below is a skills-based competition to promote and further the field of data science. You must register via the Competition Website to enter. To enter the Competition, you must agree to these Official Competition Rules, which incorporate by reference the provisions and content of the Competition Website and any Specific Competition Rules herein (collectively, the "Rules"). Please read these Rules carefully before entry to ensure you understand and agree. You further agree that Submission in the Competition constitutes agreement to these Rules. You may not submit to the Competition and are not eligible to receive the prizes associated with this Competition unless you agree to these Rules. These Rules form a binding legal agreement between you and the Competition Sponsor with respect to the Competition. Your competition Submissions must conform to the requirements stated on the Competition Website. Your Submissions will be scored based on the evaluation metric described on the Competition Website. Subject to compliance with the Competition Rules, Prizes, if any, will be awarded to Participants with*

### COMPETITION TERMS AND CONDITIONS



I agree to the Competition-specific rules above and Kaggle's Foundational Competition Rules. The Kaggle Competition Foundational Rules supersede the Competition-specific rules in the event of any conflict.



I understand joining this Competition creates a direct relationship between the Host and me. Kaggle has no liability for any promised prizes, commitments, content, or actions of a Host or Participant.

Decline

I Understand and Accept

Después, ya es posible descargar el dataset desde el apartado Data.

The screenshot shows the Kaggle interface with the following details:

- Left Sidebar:** Shows the user's profile (kaggle), navigation links (Create, Home, Competitions, Datasets, Models, Code, Discussions, Learn, More), and a list of recent projects (Your Work).
- Top Bar:** Includes a search bar, a "Submit Prediction" button, and a "Data Explorer" link.
- Dataset Overview:** Title: House Prices - Advanced Regression Techniques. Tabs: Overview (selected), Data (highlighted in blue), Code, Models, Discussion, Leaderboard, Rules, Team, Submissions.
- Data Description:** A file named "data\_description.txt" (13.37 kB) containing detailed descriptions of the variables:
  - MSSubClass:** Identifies the type of dwelling involved in the sale.

|     |   |
|-----|---|
| 20  | 1-STORY 1946 & NEWER ALL STYLES                       |
| 30  | 1-STORY 1946 & OLDER                                  |
| 40  | 1-1/2 STORY UNFINISHED ATTIC ALL AGES                 |
| 45  | 1-1/2 STORY - UNFINISHED ALL AGES                     |
| 50  | 1-1/2 STORY FINISHED ALL AGES                         |
| 60  | 2-STORY 1946 & OLDER                                  |
| 70  | 2-STORY 1945 & OLDER                                  |
| 75  | 2-1/2 STORY ALL AGES                                  |
| 80  | SPLIT OR MULTI-LEVEL                                  |
| 85  | SPLIT LEVEL - 1946 & NEWER                            |
| 90  | DUPLEX - ALL STYLES AND AGES                          |
| 120 | 1-STORY PUD (Planned Unit Development) - 1946 & NEWER |
| 150 | 1-1/2 STORY PUD - 1946 & NEWER                        |
| 150 | 2-STORY PUD - 1946 & NEWER                            |
| 180 | PUD - MULTILEVEL - INCL SPLIT LEVEL/FOYER             |
| 190 | 2 FAMILY CONVERSION - ALL STYLES AND AGES             |
  - MSZoning:** Identifies the general zoning classification of the sale.

|    |                              |
|----|------------------------------|
| A  | Agriculture                  |
| C  | Commercial                   |
| FV | Floating Village Residential |
| I  | Industrial                   |
| RH | Residential High Density     |
| RL | Residential Low Density      |
| RP | Residential Low Density Park |
| RM | Residential Medium Density   |
  - LotFrontage:** Linear feet of street connected to property.
  - LotArea:** Lot size in square feet.
- Data Explorer:** Shows the file structure with 4 files: data\_description.txt, sample\_submission.csv, test.csv, and train.csv.
- Summary:** Shows 4 files and 163 columns.
- Download:** A "Download All" button.