Investigate a Dataset - Data Set Options

Choose one of the following datasets to explore in your project analysis.

Data Set	Overview and Notes	Example Questions
TMDb movie data	This data set contains information about 10,000 movies collected from The Movie Database (TMDb), including user ratings and revenue. • Certain columns, like 'cast' and 'genres', contain multiple values separated by pipe (|) characters. • There are some odd characters in the 'cast' column. Don't worry about cleaning them. You can leave them as is. • The final two columns ending with "_adj" show the budget and revenue of the associated movie in terms of 2010 dollars, accounting for inflation over time.	• Which genres are most popular from year to year? • What kinds of properties are associated with movies that have high revenues?
No-show appointments	This dataset collects information from 100k medical appointments in Brazil and is focused on the question of whether or not patients show up for their appointment. A number of characteristics about the patient are included in each row. • 'ScheduledDay' tells us on what day the patient set up their appointment. • 'Neighborhood' indicates the location of the hospital. • 'Scholarship' indicates whether or not the patient is enrolled in Brasilian welfare program Bolsa Família. • Be careful about the encoding of the last column: it says 'No' if the patient showed up to their appointment, and 'Yes' if they did not show up.	• What factors are important for us to know in order to predict if a patient will show up for their scheduled appointment?