Chegou a hora de fixarmos o conhecimento adquirido neste curso e a melhor forma de fazer isso é praticando. Para isso, baixe o [notebook](https://caelum-online-public.s3.amazonaws.com/2651-databricks-conhecendo-ferramenta/aula-5/projeto_final.ipynb" \t "_blank) com as instruções sobre o mini projeto que vamos desenvolver utilizando os recursos que aprendemos até aqui.

Neste desafio, vamos usar o dataset **bikeSharing** que pode ser obtido no conjunto de datasets do Databricks. Este dataset possui os seguintes atributos:

* **instant**: índice de registro
* **dteday** : data
* **season** : estação (1: inverno, 2: primavera, 3: verão, 4: outono)
* **yr**: ano (0: 2011, 1: 2012)
* **mnth**: mês (1 a 12)
* **hr** : hora (0 a 23)
* **holiday** : se o dia é feriado ou não
* **weekday** : dia da semana
* **workingday** : se o dia não for nem fim de semana nem feriado é 1, caso contrário é 0.
* **weathersit**: condições climáticas
  + 1: Claro, Poucas nuvens, Parcialmente nublado, Parcialmente nublado
  + 2: Névoa + Nublado, Névoa + Nuvens quebradas, Névoa + Poucas nuvens, Névoa
  + 3: Neve fraca, Chuva fraca + Trovoada + Nuvens dispersas, Chuva fraca + Nuvens dispersas
  + 4: Chuva Forte + Paletes de Gelo + Tempestade + Névoa, Neve + Neblina
* **temp** : Temperatura normalizada em Celsius. Os valores são derivados via (t-t\_min)/(t\_max-t\_min), t\_min=-8, t\_max=+39 (somente na escala horária)
* **atemp**: Sensação térmica normalizada em Celsius. Os valores são derivados via (t-t\_min)/(t\_max-t\_min), t\_min=-16, t\_max=+50 (somente na escala horária)
* **hum**: umidade normalizada. Os valores são divididos por 100 (máx)
* **windspeed**: velocidade do vento normalizada. Os valores são divididos por 67 (máx)
* **casual**: contagem de usuários casuais
* **registered**: contagem de usuários registrados
* **cnt**: contagem do total de bicicletas alugadas, incluindo bicicletas casuais e registradas