

1. می خواهیم تابعی داشته باشیم که کاراکترهای خاص را از ستون transcript حذف کند به طوری که مجموعه داده را در batch های 1000 تایی پردازش کند. برای این کار لازم است توابع batchwise_remove_special_chars و remove_special_characters را تکمیل کنید. در نظر داشته باشید برای گرفتن نمره کامل باید تست ها را پاس کنید.

2. می خواهیم تابعی داشته باشیم که یک vocabulary از ستون transcript مجموعه داده ایجاد کند، جایی که key ها کاراکترهای منحصر به فرد و value تعداد هر کاراکتر است. برای این کار کافی است تابع generate_vocab_with_frequencies را تکمیل کنید. در نظر داشته باشید برای گرفتن نمره کامل باید تست را پاس کنید.