* Nhóm em có cập nhật những phần sau so với buổi vấn đáp
  + Slide
  + Thay đổi cách chia dữ liệu train,validation,test
    - Kết quả dự đoán trên tập test độ đo R2 của nhóm em lần trước là 0.95. Nhóm em có tìm hiểu trở lại và nhận ra cách chia dữ liệu train,validation,test của nhóm em chưa đúng.
    - Sau khi tách riêng khách sạn (train, validation và test trên những tập dữ liệu khách sạn riêng biệt) và kiểm tra thấy độ đo thấp rõ rệt, nhóm em nghĩ nguyên nhân là do dữ liệu train đã có chứa dữ liệu của những phòng của những khách sạn có trong tập validation và test.
    - Tuy tập validation và tập test được tách riêng nhưng dữ liệu lại không đáp ứng được mục tiêu chung của bài toán là đánh giá phòng của một **khách sạn mới** thêm vào hệ thống.
    - Nhóm em tổ chức lại dữ liệu train,validation,test. Chia ra như sau, dữ liệu các phần là hoàn toàn tách biệt, không có quan hệ là cùng khách sạn với nhau
      * train\_data: 99 khách sạn, 1578 dữ liệu phòng
      * validation\_data: 27 khách sạn, 487 dữ liệu phòng
      * test\_data: 25 khách sạn, 331 dữ liệu phòng
    - Nhóm em đã cập nhật lại file preprocesscingData\_and\_train.ipynb quá trình huấn luyện mô hình mới trên những tập train,validation mới. Toàn bộ phần tiền xử lý đều được giữ nguyên
    - Kết quả đạt được tuy thấp hơn rất nhiều nhưng em nghĩ hiệu quả hơn so với độ đô R2 cao ban đầu. Nhóm em đạt được kết quả cuối cùng trên tập test là R2: 0.76 (kết quả khi em chạy trên Linux và Windows là khác nhau, trên Win OS chỉ đạt được 0.72)
  + Trong quá trình tổ chức chia lại dữ liệu và huấn luyện mô hình mới. Nhóm em nhận ra phần lớn dữ liệu crawl thời gian qua (chỉ một mục **price**) chưa đúng. Nhóm đã tiến hành crawl lại dữ liệu chính xác. Dự liệu mới gồm 2397 dòng.