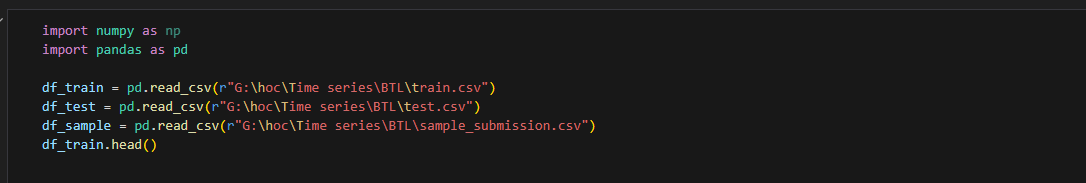
Báo cáo bài thực hành số 1

Đề 7 : Sunday

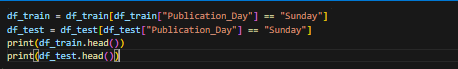
Link code github: <https://github.com/b2004x/BTH1_NGUYENGIABACH_64TTNT1_2251262579.git>

1. Xử lý dữ liệu

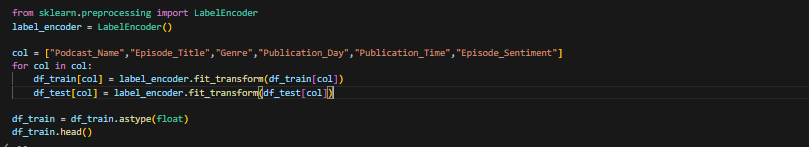
- Truyền tập dữ liệu vào sử dụng thư viện pandas



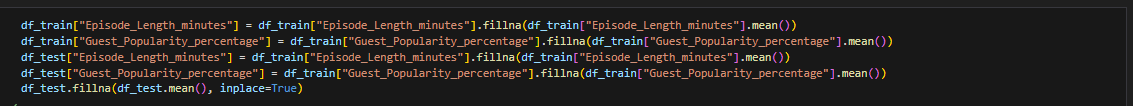
- Lọc các dữ liệu khác ngoài Publication\_day = Sunday:



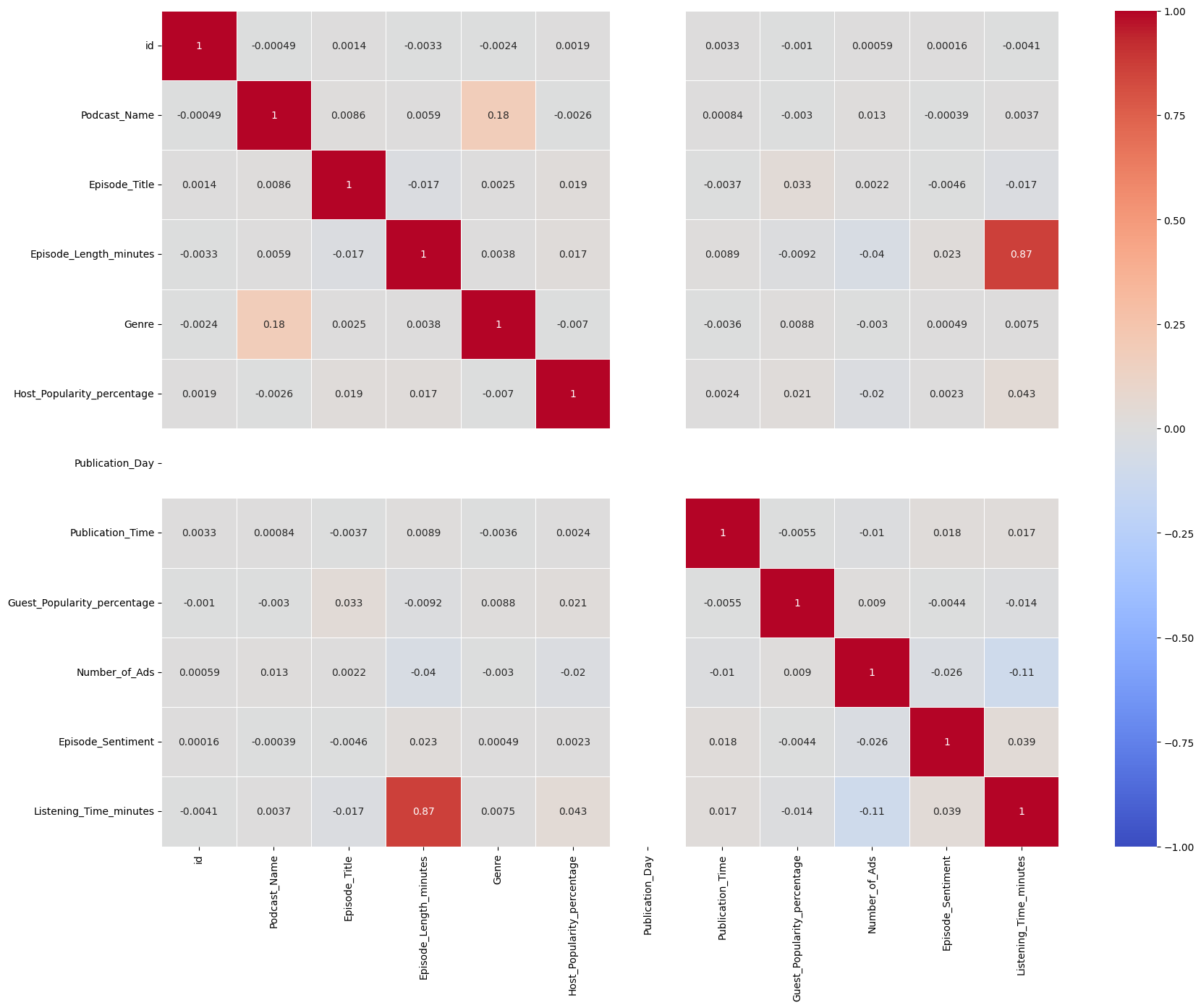
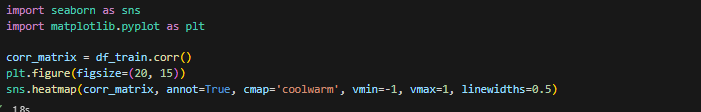
- Sử dụng label Encoding để mã hoá các cột không phải kiểu dữ liệu float và ép kiểu float cho các giá trị



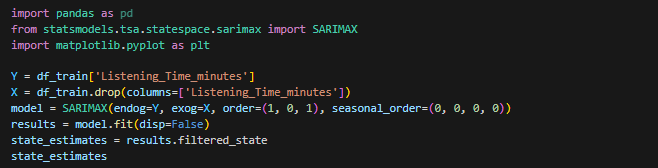
- Xử lý dữ liệu thiếu bằng mean(giá trị trung bình)



- Vẽ heatmap tương quan:



1. Sử dụng mô hình ARIMA với Kalman Filter (SARIMAX trong thư viện có sẵn)



- Y : Cột nhãn là Listening\_Time\_minutes

- X: Cột dữ liệu ngoại sinh (mọi cột trừ Listening\_Time\_minutes)

- endog = Y biến nội sinh hay biến cần dự đoán

- endog = X biến ngoại sinh hay các đặc trưng khác cần thiết để dự đoán Y

- Khởi tạo mô hình ARIMA (1,0,1):

- order = (p,d,q) gồm có:

+ p (AR): Thành phần tự hồi quy

+ d (I): Thành phần sai phân (để làm chuỗi dừng)

+ q (MA) Thành phần trung (Thành phần trung bình trượt sai số)

- seasonal\_order = (0,0,0,0):

+ P = 0 : Không có thành phần tự hồi quy mùa vụ  
+ D = 0 : Không có sự khác biệt mùa vụ

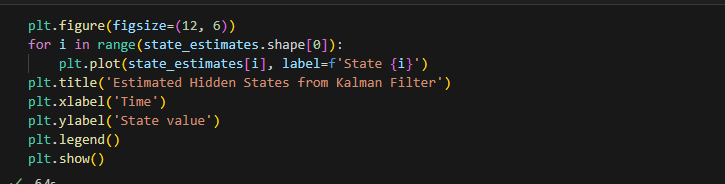
+ Q = 0: Không có thành phần trượt mùa vụ

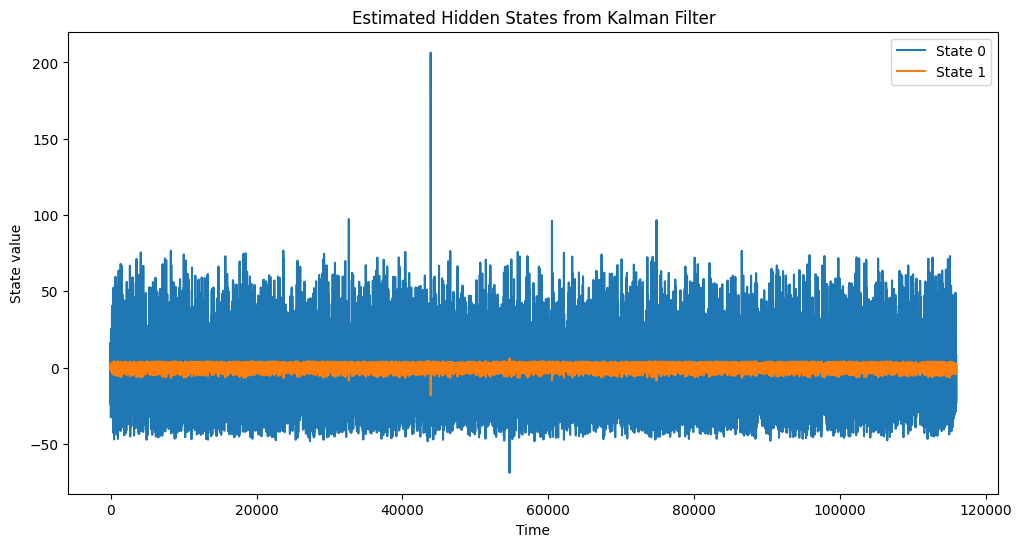
+ S = 0: Không có chu kỳ mùa vụ

- model.fit : Huấn luyện mô hình (disp = False – tắt log trong quá trình trainning)

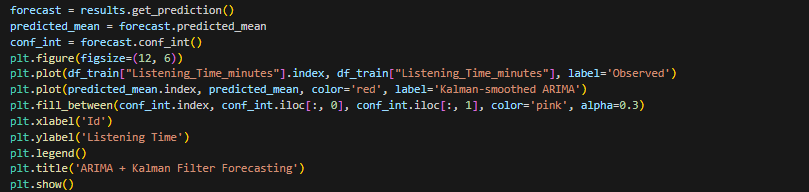
- result.filthered\_state: Ước lượng trạng thái ẩn

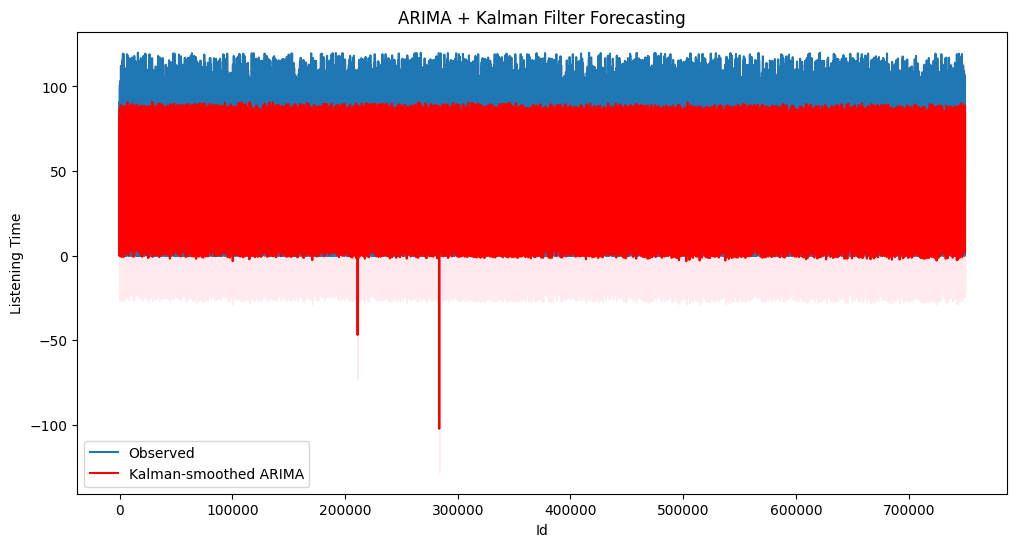
- Vẽ biểu đồ trạng thái ẩn:





- Biểu đồ so sánh giá trị quan sát và giá trị dự đoán của Arima





- forecast = result.get\_prediction(): Dùng để dự đoán giá trị bằng mô hình vừa huấn luyện

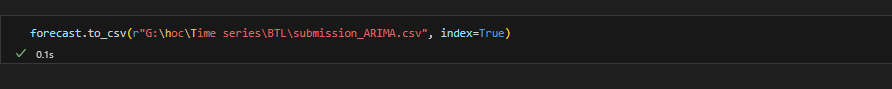
- forecast.predicted\_mean :Trả về trung bình giá trị dự đoán của forecast

- conf\_int: Trả về khoảng tin cậy cho các dự đoán

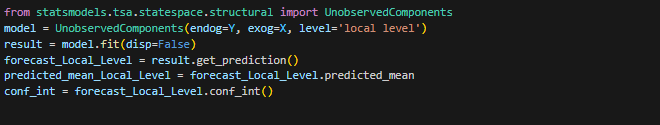
- Dự đoán với tập test và so sánh kết quả với submission\_sample:



- Chuyển dự đoán sang file submission:

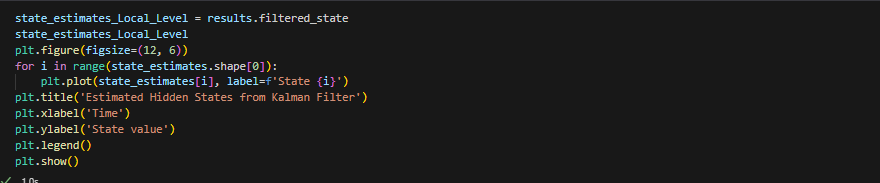


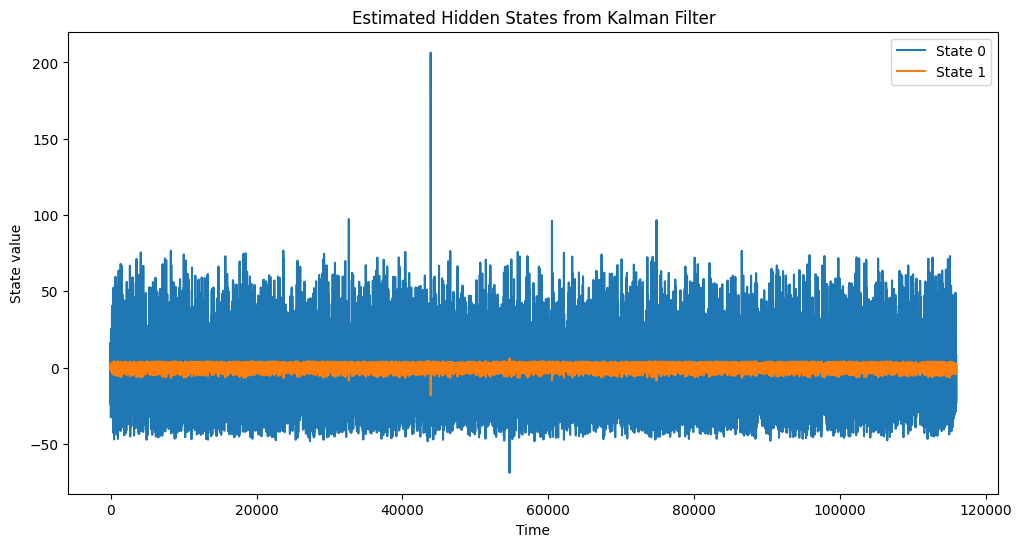
1. Mô hình Local Level(State-space)



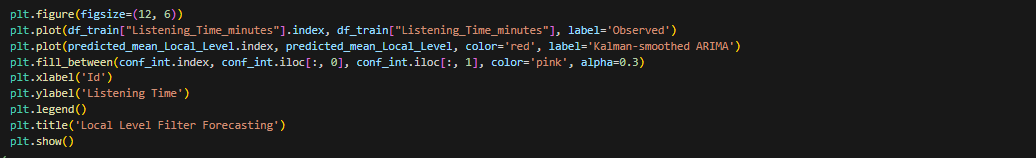
- UnobservedComponets trong thư viện statsmodels

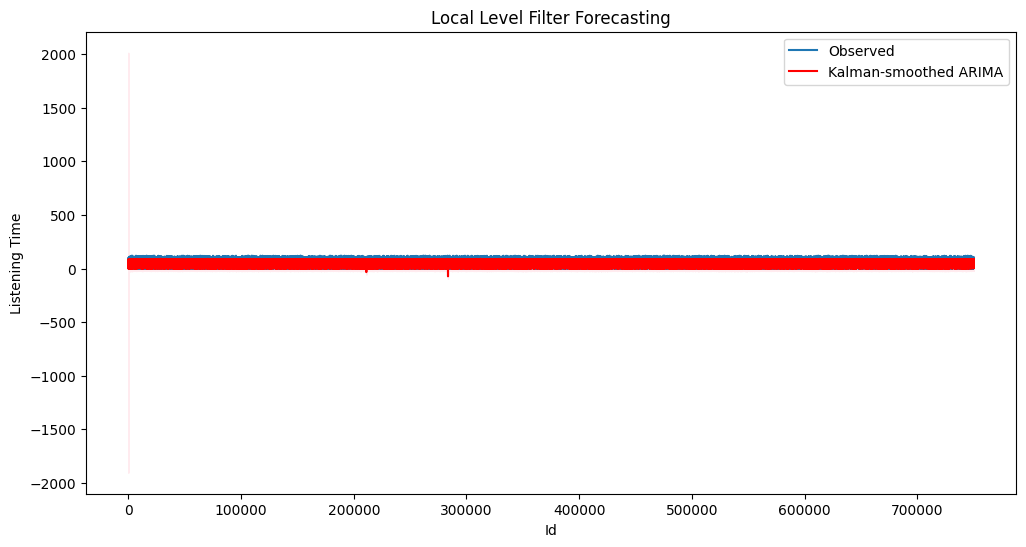
- Vẽ biểu đồ trạng thái ẩn:





- Vẽ biểu đồ so sánh giá trị thực tế và dự đoán của mô hình

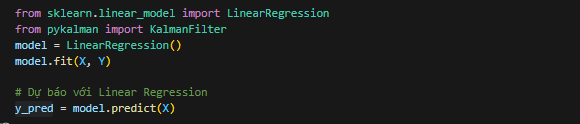




- In kết quả dự đoán ra file csv:



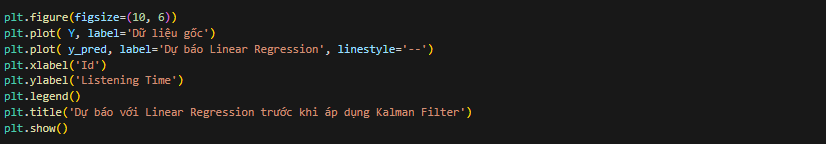
1. Mô hình Linear Regression với Kalman Filter

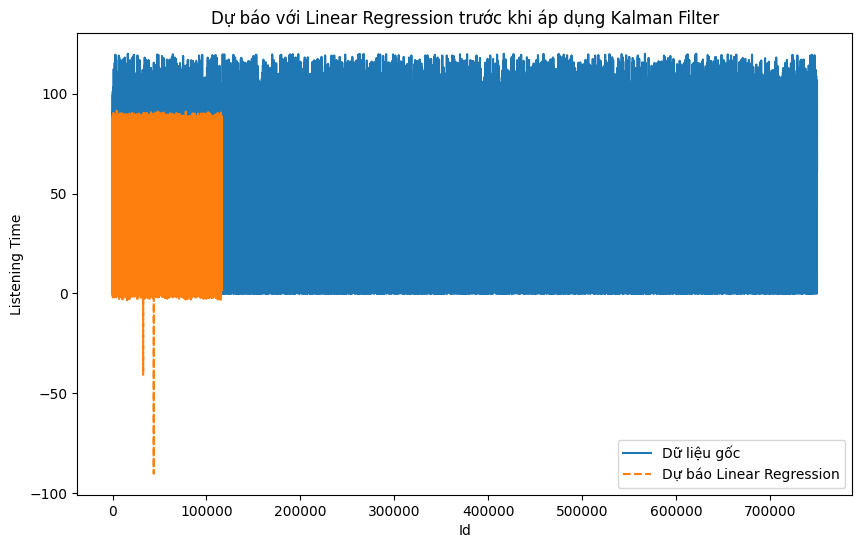


- Gọi mô hình Linear Regression từ thư viện sklearn

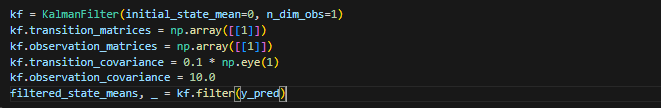
- model.fit : Trainning mô hình

- Kết quả dự đoán của mô hình trước khi Kalman Filter





- Sử dụng Kalman Filter



+ kf.transition\_matrices: Khởi tạo ma trân chuyển tiếp

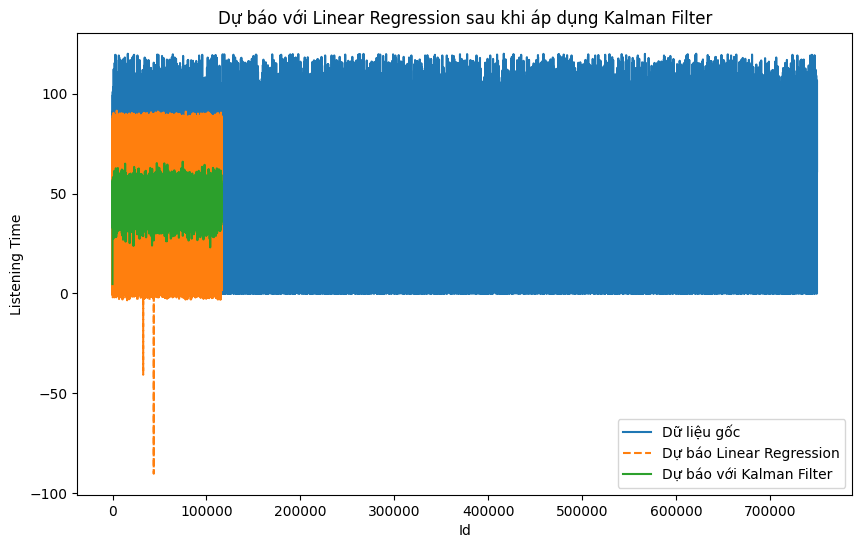
+ kf.observation\_matrices: Khởi tạo ma trận quan sát

+ kf.transition\_covarince: Ma trận hiệp phương sai chuyển tiếp

+ kf.observation\_covariance: Sai số quan sát

+ filtered\_state\_means, \_ = kf.filter(y\_pred): Truyền dự báo của linear Regression vào cho kalman filler

+ Kết quả sau khi áp dụng Kalman Filter:



- Truyền tập test vào model đã trainning :



- In kết quả ra CSV

