1/a = df.to\_csv(sep='\t', index=False)

Nếu bạn không chuyển tham số **index=False** cho phương thức **to\_csv**, pandas sẽ tự động tạo một chỉ mục mặc định với các giá trị từ 0 đến **n-1** (trong đó **n** là số hàng trong DataFrame).

Tuy nhiên, khi bạn đọc lại tệp CSV đó vào Spark DataFrame, Spark sẽ tự động tạo một chỉ mục mới cho DataFrame của mình và bỏ qua chỉ mục mặc định mà Pandas đã tạo ra trước đó. Do đó, việc bạn chuyển **index=False** chỉ đơn giản là không lưu chỉ mục mặc định của Pandas vào tệp CSV, nhưng không ảnh hưởng đến chỉ mục mới được tạo bởi Spark khi bạn đọc lại tệp CSV vào Spark DataFrame.

2/spark\_df.registerTempTable("time\_table")

Cho phép tạo bảng tạm thời tên là time\_table để thực hiện truy vấn sql trên spark sql.

3/list trong python có tính năng xem index của phần tử là bao nhiêu

print(['P', 'Q', 'R', 'S'].index('R'))

🡪output: 2

4/ sử dụng phương thức keys.() để lấy all giá trị key trong dct

my\_dict = {'a': 1, 'b': 2, 'c': 3}

keys = my\_dict.keys()

print(keys) # Output: dict\_keys(['a', 'b', 'c'])

🡪Nếu key không có trong dictionary, sẽ xảy ra lỗi **KeyError**

5/ sử dụng phương thức values() để lấy list giá trị values trong dct:  
my\_dict = {'a': 1, 'b': 2, 'c': 3}

values\_list = list(my\_dict.values())

print(values\_list) # Output: [1, 2, 3]