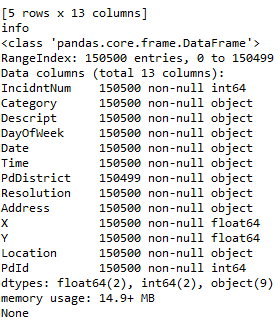
모두의 딥러닝 스터디 실습

**#1. 샌프라시스코 2016 범죄 기록을 사용하여 날짜. 시간과 범죄 종류의 관계 분석**

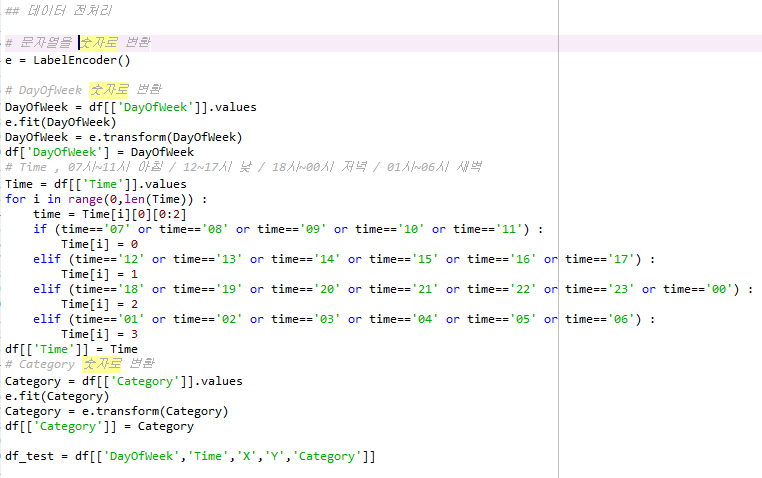
<샌프라시스코 2016 범죄 기록 사용 / 주소 : [https://www.kaggle.com/roshansharma/sanfranciso-crime-dataset)>](https://www.kaggle.com/roshansharma/sanfranciso-crime-dataset)%3E)

1. **기존 데이터 info**

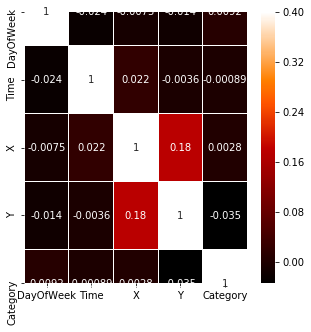


숫자로 변환할 수 있는 Category\_범죄종류, DayOfWeek\_범죄발생요일, Time\_범죄발생시간, X, Y\_범죄발생위치 칼럼 택.

1. **데이터 전처리**

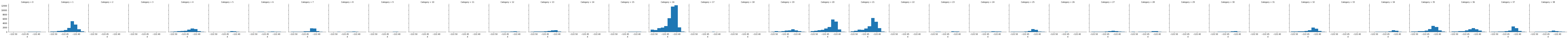


String 데이터를 숫자로 변환(월~일=>0~6/범죄카테고리=>0~38/시간=>0~3)

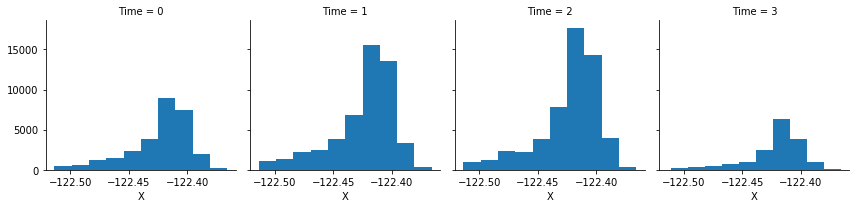


데이터간의 상관관계를 알아보기 위해 sns.heatmap() 이용.

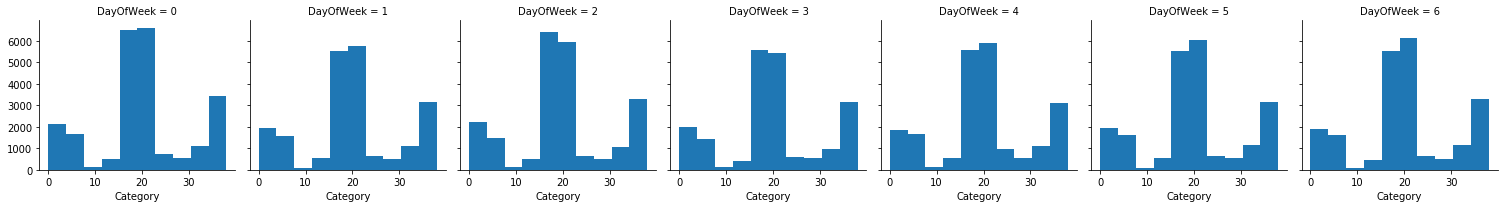
전체적으로 상관관계가 적으며, DayOfWeek는 Category와, Time은 X와, X는 Y와, Y는 X와, Category는 X와 가장 관련이 있는 것을 확인할 수 있었음



카테고리와 X의 관계. (카테고리가 총 39개라 그래프가 가로로 길어서 작게 첨부)

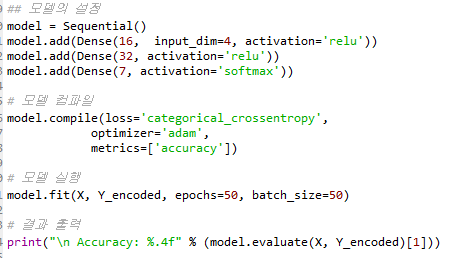


Time과 X의 관계



DayOfWeek(요일)과 카테고리 관계

1. **모델 설정 및 실행**



<발생요일,발생시간,발생장소(X,Y)=>범죄카테고리>, <범죄카테고리,발생장소,발생시간=>발생요일>, <범죄카테고리,발생장소,발생요일=>발생시간>를 위와 같은 형태의 모델에 넣어 풀어 보았음.(출력층의 Dense 뉴런 개수만 바꿈)

- <발생요일,발생시간,발생장소(X,Y)=>범죄카테고리> => 정확도 0.26

- <범죄카테고리,발생장소,발생시간=>발생요일> => 정확도 0.1553

- <범죄카테고리,발생장소,발생요일=>발생시간> => 정확도 0.37

위와 같이 매우 낮은 정확도가 나왔음. 층을 바꿔보기도 하고, 관련 없어 보이는 칼럼을 제거해보기도 하였지만 정확도는 크게 변화 없었음. **세 개의 모델 다 실패하여 다른 데이터를 찾아봄**

**#2. 버섯 분류**

<버섯 데이터 사용 / 주소: https://www.kaggle.com/uciml/mushroom-classification >