지난 발표 이후 우리는 당장 집중해야 할 과제에 대해 알게 되었습니다.

먼저 데이터를 취득하고 이 데이터를 읽어서 필요한 데이터를 구분해줄 필요가 있다

저희가 시뮬레이션 프로그램을 통해서 받아들이는 데이터는 여러 종류가 있지만, 그 중에는 저희의 실험에 불필요한 데이터도 포함되어 있습니다. 예를 들어, 데이터 중 Fuel\_level과 Average\_fuel\_consumption은 둘 다 연료에 관련된 것으로 저희가 목표로 하는 주행 상황 파악에는 필요가 없는 데이터들입니다. 그래서 저희는 먼저 시뮬레이션을 통해 얻어낸 데이터를 필요한 부분과 필요 없는 부분으로 나누어서 정제하는 것부터 시작하였습니다.

**예를 들에 우리 실험에서 당장은 fuel level에 대한 데이터는 필요 없다**

**이런 필요없는 데이터를 구분하고 필요한 데이터만을 읽어올 필요가 있다**

그렇게 필요한 데이터를 가려낸 후에는 Google colab에서 데이터 시각화 도구를 사용하여 각각의 데이터를 분석하기 쉽게 시각화 하였습니다. 이를 통해서 각각의 필요한 데이터에 대해서 여러가지 주행 case 학습을 위한 data를 만들어 내야 합니다.

우리는 우선 이전에 얻은 데이터를 사용하여 colab에서 표를 통해 출력하고, 필요하지 않은 데이터는 구분해내는 작업을 수행했다

또한 우리가 얻은 데이터에서 각각의 데이터가 정확하게 어떤 의미를 가지는지 파악을 해 줄 필요가 있다.

동시에 이 데이터들 중에서 변화가 없는 데이터들의 경우에 matplotlib을 통해서 그래프로 출력하여 확인하고 이 데이터는 사용하지 않거나 굳이 사용할 필요가 없는 데이터로 판단하고 제외해줄 필요가 있다

이러한 데이터들을 종합하여 해당 자료가 어떤 case의 운전 유형인지 구분해줄것이다.

그를 위해선 더 많은 양의 동일한 case data가 필요하다. 다시말해 직진에 대한 여러 개의 datat와 커브에 대한 여러 개의 data, 차선변경에 대한 여러 개의 data 등

다음과 같이 동일한 data를 여럿 수집해서 일정하게 나타내는 패턴을 분석해 나가고 임의의 data를 넣었을 때 해당 data가 어느 case에 해당하는지 구분해 낼 수 있는 것이 우리의 목표이다

우리는 한동안 몇가지 경우에 대한 data를 계속 얻을 필요가 있다.

계속해서 데이터를 얻고 이 데이터를 분석하는 작업을 지속할것이다

Domain Randomization

시뮬레이터 상에서 얻은 데이터는 실제 데이터에 적용허가앤 차이가 존재한다

몇가지 방법론이 존재하는데 system identification, domain adaptation, domain randomization

이중에서 DR의 경우 랜덤한 특성을 가진 여러가지 simulation 환경을 만들 수 있고 이 환경을 기반으로 model을 학습

* Domain Randomization에 대한 학습 및 자료조사​
* 시뮬레이션 공간 상에서 얻어지는 데이터의 경우는 실제 데이터와 차이가 존재하는데 이 Gap을 줄이기 위한 여러가지 방법론이 존재한다. 그 중 하나인 DR(Domain Randomization)은 여러가지 data에 대한 random