**TCC - Felipe Assad e Jorge Chagas - (22/05/2019)**

**Objetivo da atividade:**

Fazer o pré processamento dos dos dados:

**O que foi feito:**

Foram extraídas cinco base de dados da polícia de São Paulo:

1. furto\_veiculo
2. roubo\_veiculo
3. intervencao\_policial
4. latrocinio
5. lesao\_corporal

Em cada uma delas, foram criadas mais três colunas: mês (varchar), tipo\_crime (varchar) e sofreu (boolean).

Em seguida, descartou-se os dados em que não se têm o valor de latitude e longitude, eles são cerca de 10% do dataset total.

Por fim, foi separado de forma aleatória, para cada crime, 66% dos dados e salvo como treinamento e o resto salvo como teste, para a partir de agora usarmos sempre os mesmos dados tanto para teste quanto para treinamento, para podermos avaliar os distintos métodos de aprendizado de máquina.