Analise explanatória da Criminalidade Chicago

1Alvaro Cristian da Silva Botelho

1MBA Data Science e Big Data – Universidade do Vale dos Sinos (UNISINOS)  
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

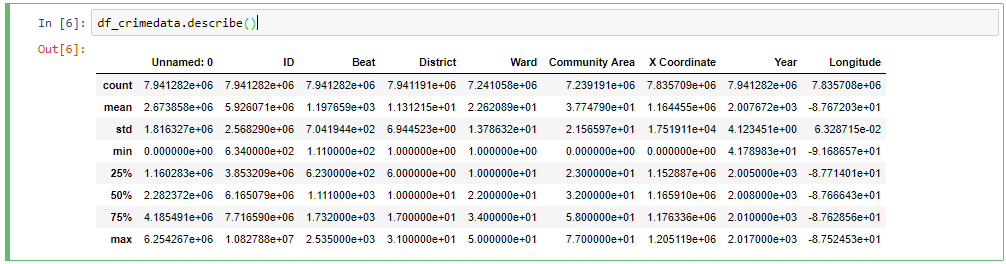
Cristian.ucpel@gmail.com

INTRODUÇÃO

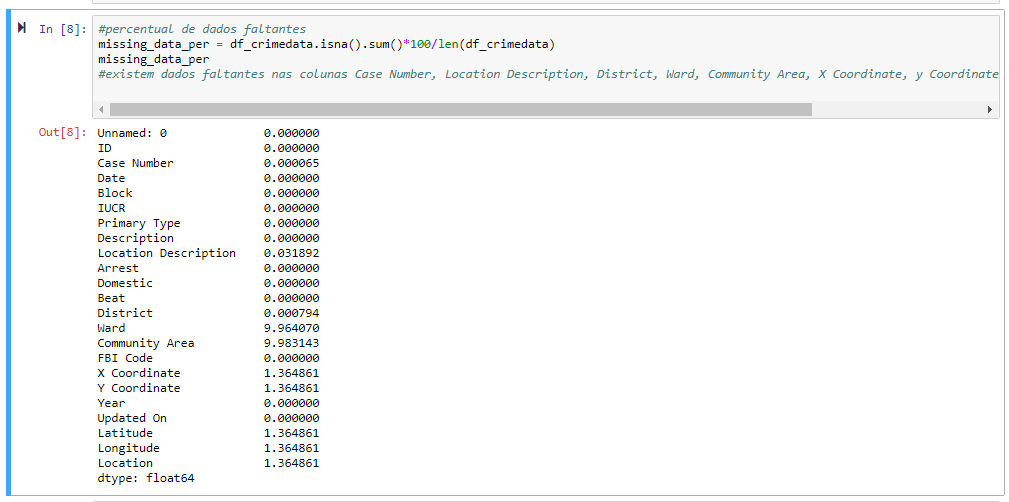
O presente documento visa relatar a analise realizada no conjunto de dados sazonais obtidos através da plataforma Kaggle <https://www.kaggle.com/currie32/crimes-in-chicago>). A base contem 7.941.282 Linhas e 23 atributos. Durante o processo foi utilizado o algoritmo de forecast FBProphet (<https://facebook.github.io/prophet/>). Prophet é um algoritmo para forecasting time series com base em modelos aditivos no qual as tendências não lineares se ajustam a sazonalidade, também funciona com dados faltantes e tem uma qualidade alta quanto a outliers. Mapas e os scripts foram anexados juntos deste documento. Mais informações estão no https://github.com/alcristian/MachineLearningTask10-23-2018

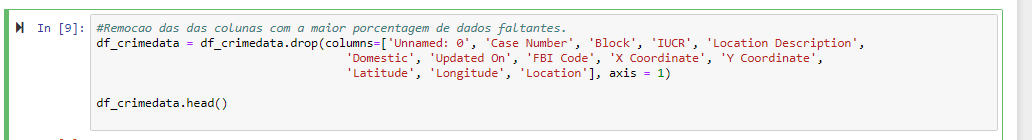
ANALISE EXPLANATORIA

Alguns dados estatísticos sobre a massa de dados.

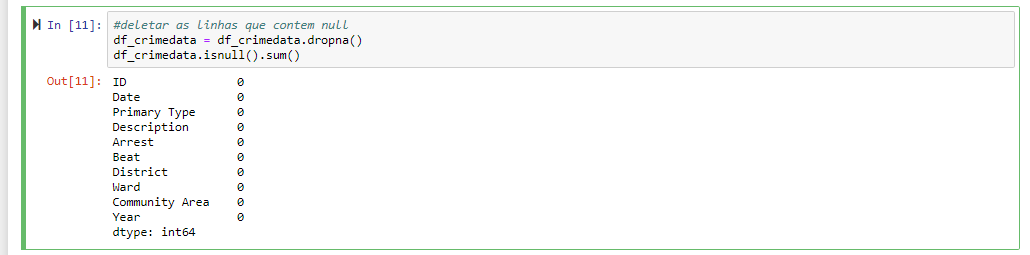


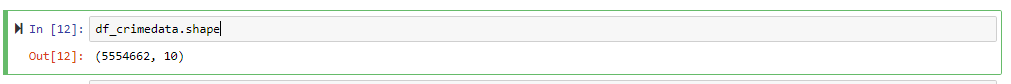
Após analise exploratória nos dados foi verificado uma grande quantidade de dados faltantes em algumas colunas, as quais foram removidas .



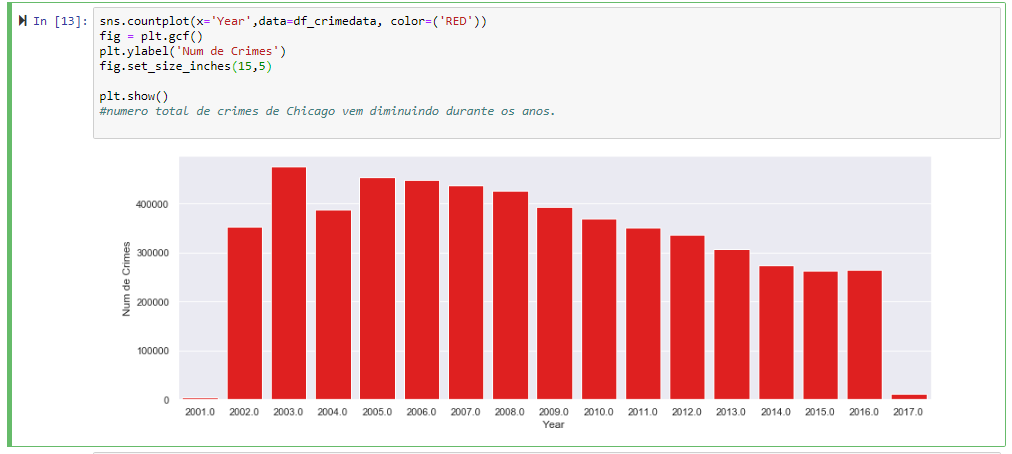


Após a remoção das linhas faltantes foi removida também as linhas com valores nulos, mesmo sabendo que o algoritmo trabalha bem com valores nulos.



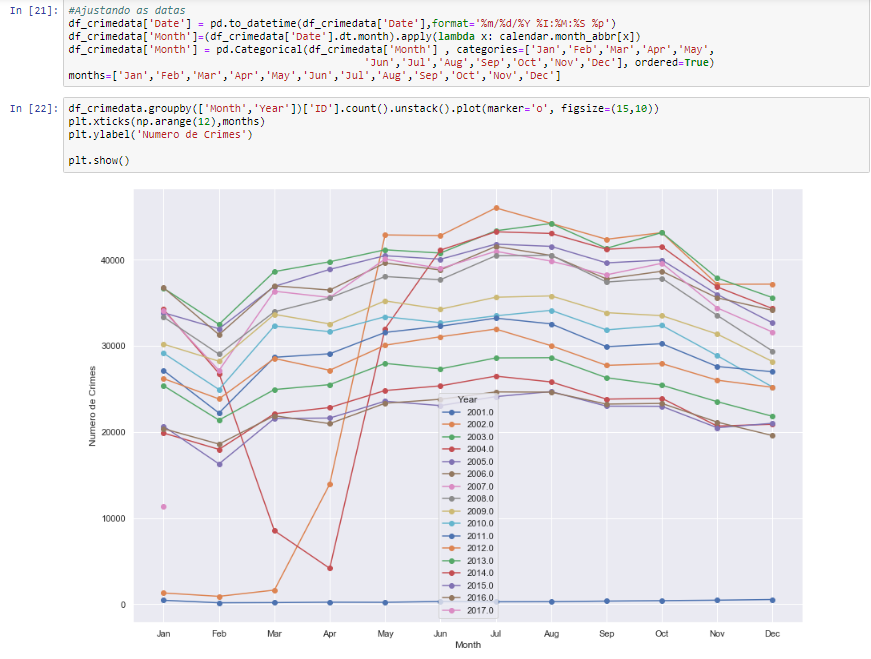


Após todas as limpezas sobraram 5.554.662 linhas e 10 atributos.

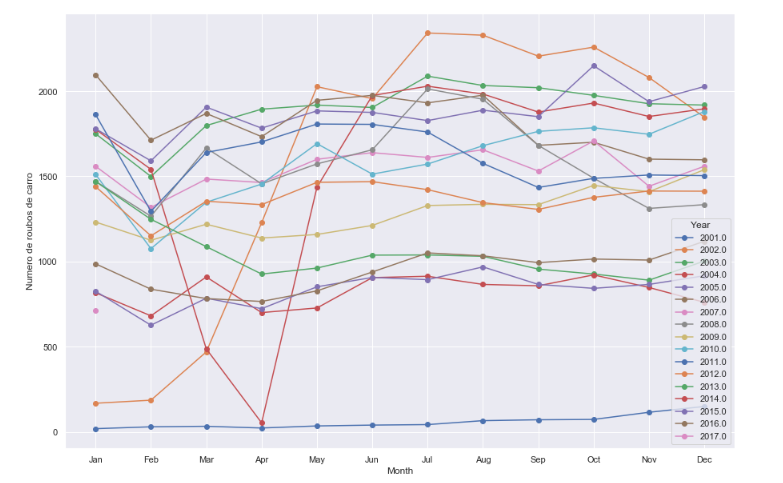


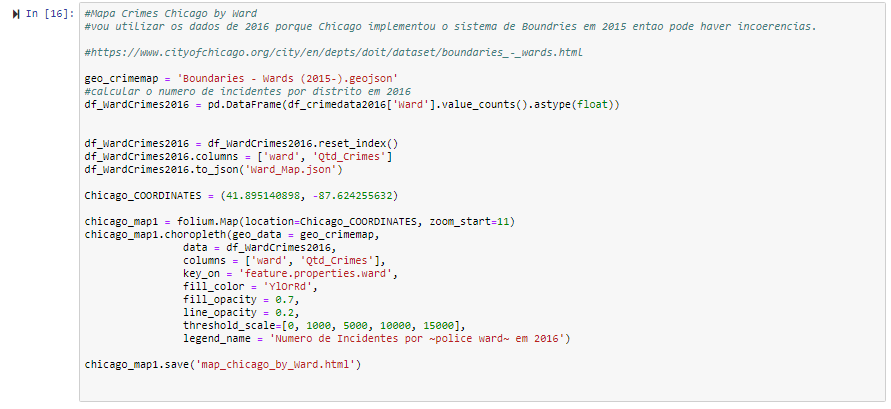
Verificando os dados históricos de Chicago a criminalidade geral vem diminuindo durante os anos. Os dados de 2017 não contemplam todo o ano por esse motivo está com um percentual muito baixo.

Separados os dados de forma mensal podemos ver que em todos os anos há um aumento da criminalidade nos meses de julho. O pico da criminalidade em Chicago em todos os anos quase sempre é no mesmo mês.

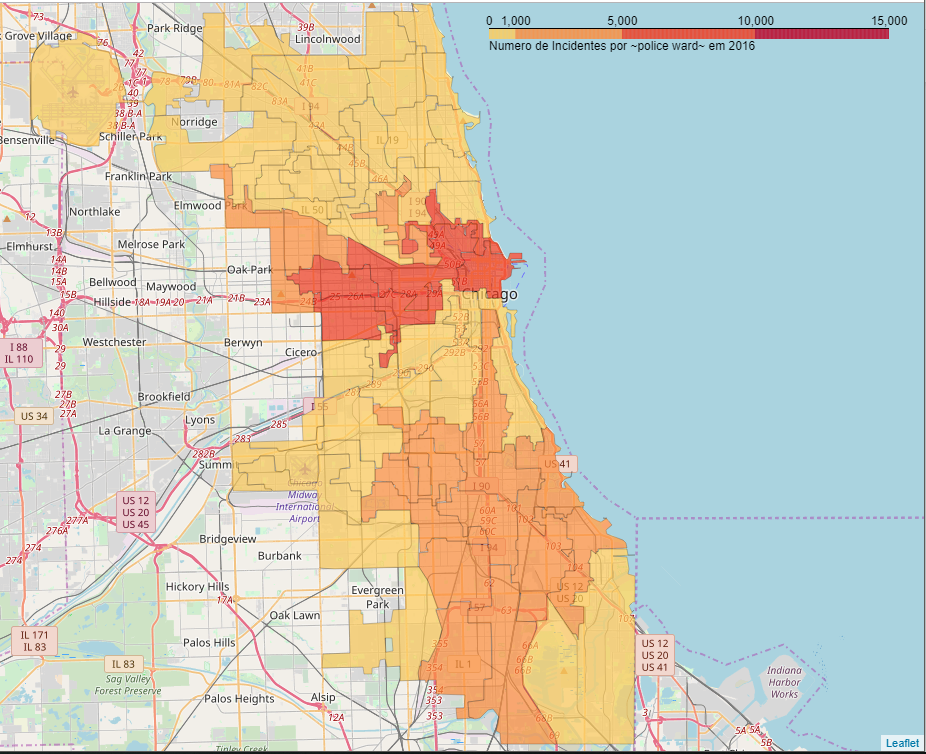


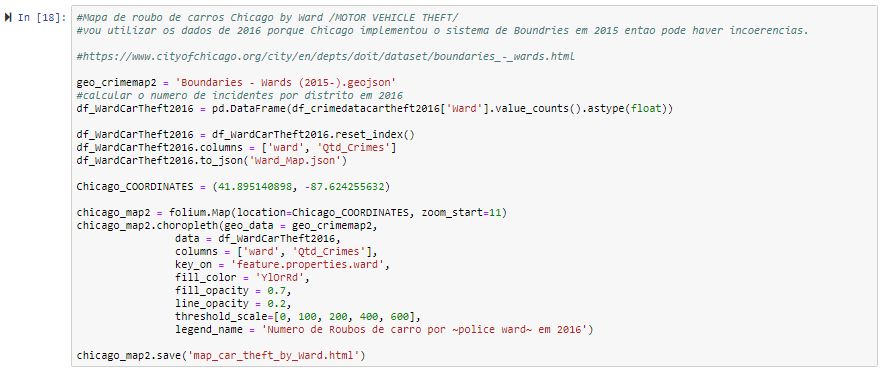
Isolando os dados referentes ao roubo de carros encontramos alguns outliers nos anos de 2012 e 2014, mas como observado no gráfico acima o pico da criminalidade não parece ser nos mesmo meses, e ela aumenta ao decorrer dos meses.





Criação de mapa para visualizar a quantidade de crimes por região policial em Chicago, foi utilizado nesse gráfico somente os dados do ano de 2016 porque a Policia de Chicago só implementou o sistema de Boundries. Existe uma região onde o crime eh concentrado com uma divergência bem maior que as outras áreas de Chicago.



O mapa dos roubos de carro mostra uma figura semelhante da distribuição de crimes geral. Porem esta concentrado em 2 areas especificas.

