

RELATÓRIO SOBRE OS DADOS DO CALL CENTER 311 DE NOVA YORK

Descrição Inicial

Foi disponibilizado um arquivo com 20gb com os dados sobre o call center 311 da cidade de Nova York, este call center reuni chamados de varias agencias da cidade, como reclamações de barulho, estacionamento irregular, cobrança de taxas.

O objetivo era fornecer insights sobre esses dados afim de traduzir esse grande volume de dados em algo mais palpável.

As atividades realizadas foram inserir os dados em uma base SQL, separando em tabelas menores criando relações entre essas tabelas, facilitando consultas nas tabelas individuais devido ao tamanho da base de dados sem grande, aproximadamente 36 milhões de linhas e 41 colunas, que foram separadas em 4 tabelas relacionadas através de id's:

- Complaints – contendo dados da reclamação, como data que foi criado e finalizado, status, tipo da reclamação, status e outros.
- Locations – dados da localização da reclamação, como zip do incidente reclamado, cidade, ponto de referência, bairro, endereço, latitude, longitude e outras informações.
- Agencies – informações sobre as agencias, como nome, código (abreviação), companhia de taxi.
- Additional_info – Aqui foram agrupadas informações que não se encaixavam nas outras tabelas, como tipo do veículo, como foi aberta a reclamação online ou telefone e outras informações tratadas como secundarias.

Também foi verificado a consistência do relacionamento dessas tabelas para ter certeza que a todas as tabelas estão relacionadas e tem as mesmas linhas com os mesmos id's para podermos seguir com a análise.

Análises

1. Tabelas temporárias

Após separar os dados nas 4 tabelas, foram criadas algumas tabelas temporárias para identificar alguns incidentes que poderiam ter sido reportados mais de uma vez em um curto período de tempo, indicando algum erro de preenchimento ou algum outro tipo de dado duplicado.

Foi verificado chamados no mesmo dia, onde foi levado em consideração o tipo da reclamação, o tipo de endereço, cidade, bairro (geralmente é assinalado da mesma forma), com isso foi feita uma contagem para verificar se poderia ser algum dado duplicado. Foram identificadas aproximadamente 1,6 milhões de reclamações feitas no mesmo dia, ou seja, uma indicação de duplicação, como por exemplo as mais

duplicadas, com 15 e 12 reclamações são sobre a condição da rua no Bronx e na Staten Island, como é possível ter reclamações deste tipo no mesmo dia, optou-se em manter, assim como outras reclamações que também foi considerado plausível de serem realizadas no mesmo dia.

Também foi verificado chamados na mesma semana, tentando entender se havia duplicações de dados, mas houve o mesmo entendimento que nas por dia, além de que a que mais somou reclamações em uma semana se tratava de aquecimento em uma rua do Bronx, com 19 reclamações na semana, algo que foi novamente considerado plausível, vale salientar que novamente tivemos aproximadamente 1,6 milhões de reclamações feitas na mesma semana.

Ainda olhando de forma geral os dados, notamos que algumas reclamações levam muitos dias, anos para ser mais exato, para serem concluídas, como *“Found Property”*, que seria algo como achados e perdidos, *“Adopt-a-Basket”*, adotar um lixo, que levam em média 5 e 4 anos para serem concluídos, mas não representam um grande volume, algo próximo de 3 mil em uma base de mais de 3 milhões. A partir da 9 reclamação encontramos uma média de mais razoável, junto com um volume um pouco maior para a reclamação de *“Graffiti”*, com 218 mil reclamações sendo resolvidas em um tempo médio de 291 dias.

2. Pré-Processamento

Foi ajustadas todas as nomenclaturas das 33 agencias, para que todas que tivessem o mesmo código tivessem o mesmo nome afim de facilitar em futuras analises, para esse processo foi utilizado o site da cidade de Nova York - <https://www.nyc.gov/nyc-resources/agencies.page> - onde consta todas as agencias, suas abreviações e seus nomes.

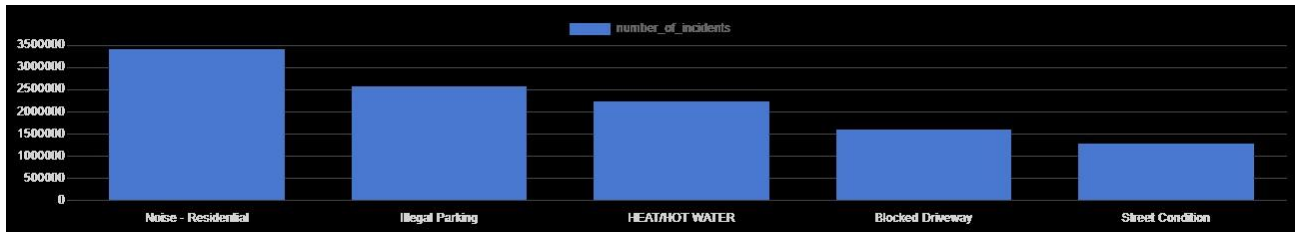
Tínhamos também alguns campos na coluna cidade sem nenhum preenchimento, apesar de ter sido verificado que neste campo, a cidade mais parecia o bairro, foi utilizado uma lista de *“zip code”* e solicitado para o chat gpt gerar de que bairro seria relacionado aquele zip, afim de preencher todos os campos. Poderia ter sido utilizado outra abordagem como utilizar o campo que aquele zip mais teria incidência ou até mesmo preencher todos os campos com a o nome da cidade de Nova York, esta opção foi pensada devido ao baixo preenchimento desse dado na coluna cidade.

Foi considerado que substituir todos os campos com o mesmo zip para uma só *“city”* não traria prejuízo para uma analise de localização pois temos outros dados de localização como o *“borough”* e até mesmo dados de geolocalização para uma análise mais aprofundada.

Aqui foi notado alguns dados inconsistentes em relação ao nome da cidade, alguns que pareciam estar mal preenchidos e foram mantidos, mas dois em especial com o nome de *“311 TEST SERVICE REQ”* que foram retirados.

3. Análise dos dados

Como a base trata de reclamações para a central 311, vamos analisar quais seriam as principais reclamações para a central. Temos que as 3 principais são: “*Noise–Residential*” com 9,3% do total das reclamações, “*Illegal Parking*” com 7% e “*HEAT/HOT WATER*” com 6%, as 3 juntas totalizam 22,3% de todas as reclamações.



Ao aprofundarmos um pouco essa relação de reclamações notamos que metade das reclamações de “*Noise–Residential*” estão situadas nos bairros do Bronx e do Brooklyn e mais da metade das “*Illegal Parking*” são nos bairros do Brooklyn e Queens.

Com isso já podemos notar uma concentração de reclamações, principalmente no bairro do Brooklyn, e ao verificar se a distribuição de reclamações vemos que o Brooklyn concentra 29% de todas as reclamações, seguido pelo Queens com 22,7% e Manhattan com 18,7%.

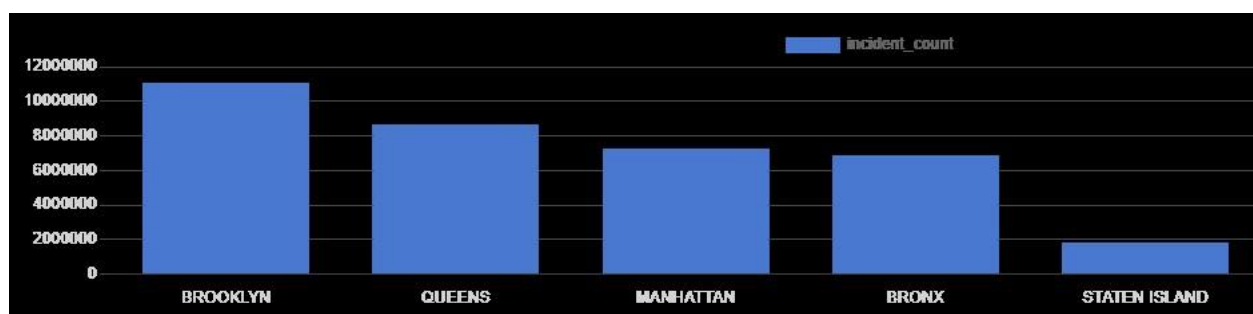
Já vimos anteriormente que algumas solicitações tem um tempo médio de atendimento bem elevado, mas ao olhar apenas as reclamações com mais de 30 dias de atendimento e separar por região, somos surpreendidos com uma grande quantidade de reclamações como “*Building/Use*” em dois bairros diferentes entre as 10 reclamações por bairro que levaram mais de 30 dias para serem finalizadas, um total de 106 mil no Queens e 59 mil no Brooklyn, também chama atenção as 74 mil reclamações de “*Graffiti*” apenas no Brooklyn, aqui vemos outra tendencia, de que o bairro Brooklyn tem mais reclamações não atendidas em menos de 30 dias junto com o Queens e Manhattan, os bairros que concentram mais reclamação. Em relação apenas as reclamações, notamos novamente o “*Graffiti*” e “*Building/Use*” como duas das que não são finalizadas antes dos 30 dias e “*General Construction/Plumbing*” que não apareceu anteriormente.

Vamos agora olhar uma relação das reclamações e soluções mais comuns, a “*Noise–Residential*” é uma reclamação bem comum e tem duas soluções comuns para ela também, a “*The Police Department responded to the complaint and with the information available observed no evidence of the violation at that time.*” ou “*The Police Department responded to the complaint and took action to fix the condition.*”, traduzindo, ou tem ação ao verificar a violação ou não tem por não verificar violação, não tomar ação corresponde a 34% resoluções do que tomar alguma ação.

4. Views e subconsultas

Ao verificar os chamados ainda abertos, temos um total de 909438 chamados em atendimento.

Como já vimos anteriormente, temos uma concentração em alguns bairros como Brooklyn, Queens e Manhattan, os três concentram 70,4% do total de chamados.



Os 10 incidentes com mais demora para ser concluído são:

ID	Bairro	Reclamação	Dias
35297016	Manhattan	Graffiti	368961
16405044	Bronx	Graffiti	69795
16774392	Brooklyn	Graffiti	69776
19640789	Bronx	Graffiti	69459
41029054	Manhattan	Graffiti	69433
37400673	Queens	Graffiti	67258
31903002	Queens	Graffiti	30739
36328706	Manhattan	Graffiti	11012
37588449	Manhattan	Graffiti	10988
15804144	Queens	Building/Use	5252

Também temos alguns endereços onde existem reclamações recorrentes, de inicio chama atenção que em sua maioria ficam em Manhattan, e são reclamações nas ruas, Brodway, 1 AV., 2 AV., 3 AV., no meio delas temos uma rua do Bronx, a 3830 bronx boulevard, todas com mais de 100 reclamações cada rua.

5. Futuras Análises

Utilizar ferramentas de geolocalização como postgis para verificar quais reclamações são feitas em pontos específicos da cidade, tendo mapas gráficos de cada reclamação e sua concentração de fácil visualização.

Outra questão que pode ser analisada é se temos algumas agencias que levam mais tempo do que outras para atender as reclamações, claro aqui teríamos que ter cuidado ao analisar devido ao volume e a dificuldade de solução para algumas reclamações que já notamos que existe.