

O Impacto na Mobilidade Urbana Norte-Americana em decorrência da chegada do Novo Coronavírus.

Aspectos técnicos da análise

Para começar, as duas APIs que utilizei nesse processo seletivo foram <https://rapidapi.com/Gramzivi/api/covid-19-data>, para capturar dados da Covid-19 nos EUA, e <https://rapidapi.com/pillious/api/google-mobility-data>, para capturar dados de mobilidade urbana de usuários que tenham dispositivos Google (celulares Android, por exemplo).

Para capturar os dados, utilizei Python e a biblioteca Requests. Com a Requests, eu batia nos endpoints e tratava os arquivos Json no próprio Python. Assim, armazenava os dados tratados num DataFrame, do Pandas, e jogava esse DataFrame num Excel. Basicamente, a forma com que escolhi armazenar os dados nesse processo foram alguns simples arquivos em Excel devido à praticidade.

Na pasta “data”, do projeto, estão os 3 arquivos que usei como base para montar o dashboard em PowerBI: “calendar”, que possui dados de tempo. Esta base eu já tinha pronta, não precisei capturá-la em nenhuma plataforma. O arquivo “covid_raw_data” provém dos dados obtidos através da API sobre Covid e o arquivo “mobility_raw_data” provém dos dados obtidos através da API de dados de mobilidade.

Escolhi armazenar tudo num repositório do Git porque essa ferramenta foi citada durante a entrevista.

Pegando referências de “*Storytelling With Data*”, de Cole Nussbaumer Knaflic, resolvi desenvolver o *report* em duas direções: a análise exploratória – o dashboard, quando o usuário final tem um contato visual com dados para conhecer o “campo” que irá entrar, e a análise explanatória – o documento, que buscará explicitar a relação dos acontecimentos baseado em dados.

Contextualizando o uso das APIs

Não tive grandes problemas para encontrar uma boa API para obter dados da Covid, a plataforma está cheia desse tipo de conteúdo. Optei por pegar a primeira API verificada por eles que atendia a condição de ter os dados segmentados por data. Pelo fato de o plano gratuito ser um pouco limitado, precisei me atentar para não exceder os limites da API. Apesar de ser uma API verificada pela plataforma, os dados sobre pacientes recuperados não retornaram da API de forma consistente e resolvi não usá-los na análise.

Com isso, antes de começar a desenvolver qualquer tipo de código, fui buscar outra API que me retornasse dados relevantes e que não me levaria tanto tempo para conseguir cruzá-los com os dados da doença, e esse foi um dos grandes desafios para mim. Tive ideias para cruzar desde dados de viagens, hospedagem de hotéis, compras em plataformas digitais, produção global de vacinas, até a busca por máscaras na Amazon, mas não encontrei nenhuma API que me retornasse conteúdo relevante para tal. Encontrei uma API de busca de notícias de qualquer tipo de assunto, mas eu não conseguia aplicar praticamente nenhum filtro na requisição utilizando o modo gratuito. Isso me fez ter um certo receio em usá-la em relação ao tempo que levaria extraindo os dados. Acho que vale destacar essa API:

<https://rapidapi.com/newscatcher-api/newscatcher-api-default/api/newscatcher>.

Assim, enquanto eu pesquisava APIs, vi a API sobre dados de mobilidade de usuários Google e me interessei por dois motivos: era possível filtrar os dados por país, com dados separados para cada UF, e existiam dados temporais que davam margem para uma análise de antes/depois.

Optei por escolher os EUA por ser um país que, em ambas as APIs, eu conseguiria obter dados relevantes.

Entretanto, na segunda API, tive alguns problemas: a documentação dela não é clara e os dados começam no dia 29 de Março, ou seja, no meio da crise, e vão até o dia 11 de Abril. Por isso optei por trazer dados do Covid referentes aos meses de Janeiro à Abril. Sobre a documentação não ser clara, a única medida que temos na API é o que eles chamam de “change”. Por ex.: Dia 29/03, Arizona, Parks, Change 10 (dados fictícios). Pela falta de conteúdo na documentação, eu não consigo definir se essa “change” é em cima do dia anterior, do mês anterior, da semana anterior, enfim. Para motivos analíticos, **eu assumi** que as datas mostram a oscilação referente ao período pré-pandêmico.

Com esses pontos claros, vamos a análise propriamente dita.

A relação entre a chegada da pandemia e a diminuição da visitação a espaços públicos nos EUA explícita com dados

É notável que, com o surgimento de um vírus altamente contagioso, medidas de isolamento social devem ser tomadas para evitar a propagação da doença. Também é notável que, caso esse controle não esteja dando bons resultados, medidas como *lockdown* são plausíveis para o cenário. Os EUA acabaram se tornando o epicentro da doença no mundo e endureceram as medidas de distanciamento social apenas no final de Março. É aí que começamos a validar os números que temos: primeiro, vamos limitar nosso cenário no dashboard para até o dia 29 de Março, data em que a API de dados de mobilidade deu a primeira resposta. Use os filtros no canto superior direito do dashboard para aplicar tais filtros. Teremos o resultado apresentado na figura 1.

Mais de 2 mil mortes no período, 141 mil casos confirmados e uma proporção de 18 mortes para cada 1000 casos. A curva de mortes apresenta crescimento exponencial e a de casos confirmados não fica para trás. Nesse ponto, os EUA já eram considerados o novo epicentro da doença no mundo. Curiosamente, nesse mesmo período, a visitação de parques subiu em torno de 7% em relação ao período pré-pandêmico, entretanto, nos demais tipos de locais, a visitação pública despencou, como é esperado com o aumento de casos. Note que as mortes subiram quase 6 vezes mais e os casos confirmados deram um salto de pouco mais de 4 vezes comparados à uma semana antes, dia 22/03.

Se analisarmos um pouco mais no detalhe, podemos filtrar ainda mais nossos dados pra ver como era a situação no período. Obs.: Note que dei duas possibilidades de filtrarmos os dados aqui: tanto clicando nos tipos de locais e países (forma que mais prefiro, mas pode não ser tão intuitivo) quanto utilizando os filtros que criei ao lado do gráfico de visitação média. No print a seguir, filtrei meus dados clicando em “Parks”, para analisarmos a anomalia no detalhe, dentro do gráfico de visitação média, o que faz com que os dados sejam **realçados** ao invés de **filtrados** (efeito caso utilize os filtros). Veja o resultado na figura 2.

Antes, a oscilação cumulativa mostrava o compilado de todos os tipos de local. Agora, mostra dados apenas do local “Parks”. Notamos que parques da Dakota do Sul tiveram um curioso aumento na procura de 126%. Talvez isso seja explicado pelo fato de que as pessoas, durante o início da pandemia, confundiam o “ficar em casa” com “estar de férias”. Outro ponto que reforça ainda mais o porquê desse tipo de local ser tão procurado, mesmo durante uma crise de saúde mundial, é de que existem mais de 10 parques estaduais concentrados ali, entre eles o icônico *Mount Rushmore*, que possui rostos esculpidos de 4 famosos presidentes dos EUA.

A visitação de áreas residenciais também cresceu no período da pandemia, mas isso é algo esperado. A imensa queda de mercados, áreas comerciais, áreas de recreação e áreas de trânsito também é esperado com um regime de isolamento social.

Agora que analisamos uma suposta anomalia em nosso conjunto, que seria uma crescente da visitação pública em algum tipo de local durante uma pandemia, vamos analisar no detalhe o resultado esperado. Olhemos a figura 3.

Nela, removi a nossa anomalia (use CTRL + click para seleção múltipla nos filtros do dashboard), que é a crescente na visitação de parques, e mantive apenas os resultados esperados. É importante notar os picos na variação média: o pico mais negativo, ou seja, onde houve menor visitação média de pessoas desde o início da pandemia, é de -46.57%, em locais de transição de pessoas, como metrô e importantes vias. Já o pico mais positivo, +11.35%, em áreas residenciais. Ambos reforçam o argumento de que, com o aumento de casos, as pessoas passam a ficar mais em áreas residenciais e transitar por ali do que ir para áreas de intenso tráfego.

Ainda com o dashboard como o da figura 3, eu encontrei um dado muito curioso: Vermont é o estado que lidera o cumulativo de diminuição de visitação em locais públicos. Vamos filtrar o dashboard apenas pelo respectivo estado, como na figura 4, para tirarmos um insight curioso: note que esse estado teve uma diminuição massiva comparado ao período posterior à pandemia na visitação de dois tipos de locais: locais de recreação e áreas de trânsito de pessoas. Com isso, consultei os dados brutos do dashboard e notei que as mortes e casos confirmados neste estado estavam baixíssimos comparados à estados vizinhos até o dia 29/03. Isso evidencia o resultado de seguir as orientações de grandes instituições globais de saúde: com uma menor aglomeração social, mais difícil é para o vírus se propagar.

Agora, vamos analisar o período até 11/04, conforme figura 5. Vejamos que os picos da variação média de visitação a locais públicos aumentou. Isto é outro indicativo da relação do aumento de casos com a limitação da mobilidade em locais públicos: agora, os picos negativos e positivos, são, respectivamente, de -47,07% em áreas de transição de pessoas e +12,07% em áreas residenciais.

Ainda na figura 5, é possível analisarmos que, neste período, quem lidera o cumulativo é a capital dos EUA, Washington D.C. Fazendo uma rápida comparação entre os períodos de 29/03 e 11/04 podemos encontrar a possível justificativa dessa diminuição da visitação de locais públicos, provavelmente impulsionada por medidas mais críticas de isolamento social: acessando os dados brutos, em 29/03, a capital norte-americana possuía um total de 342 casos confirmados e 5 mortes. Já em 11/04, o mesmo lugar possuía 1778 casos confirmados e um total de 47 mortes, um aumento exponencial em ambas as métricas que, mais uma vez, reitera a relação entre o aumento de casos da COVID-19 e a diminuição da visitação de locais públicos nos EUA.

Figuras

Figura 1

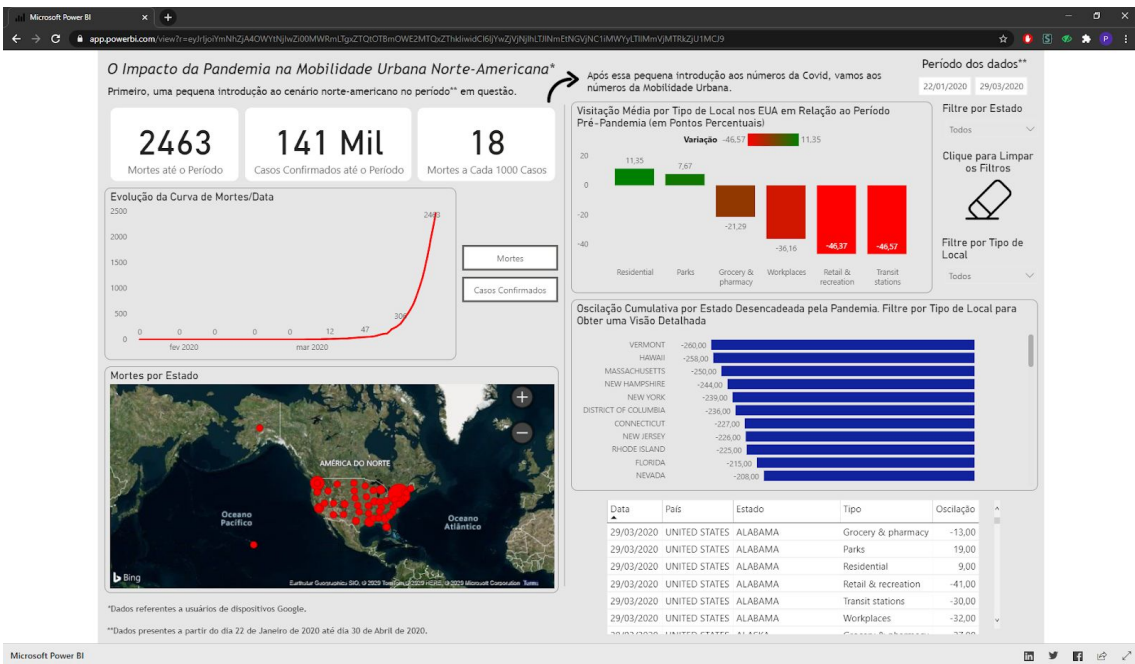


Figura 2

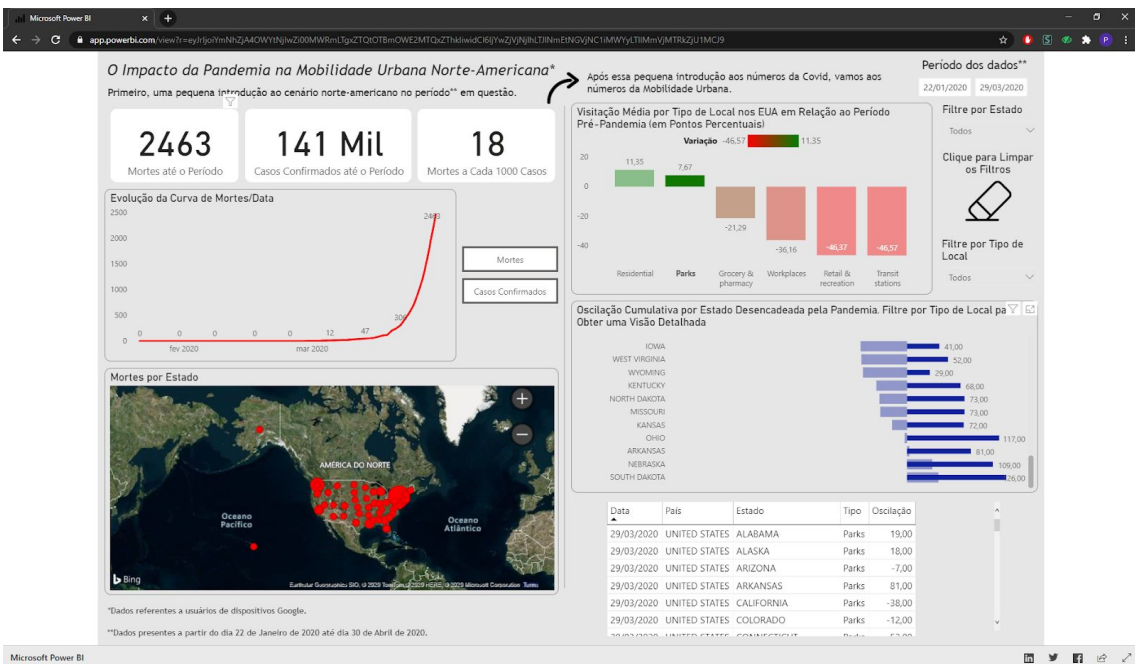


Figura 3

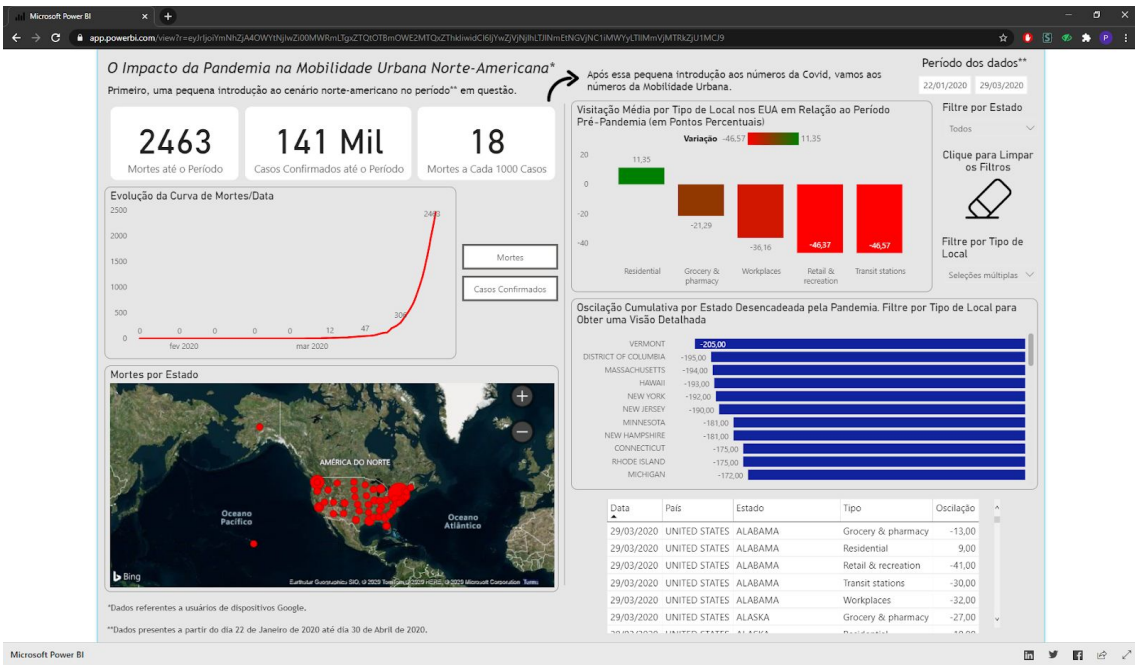
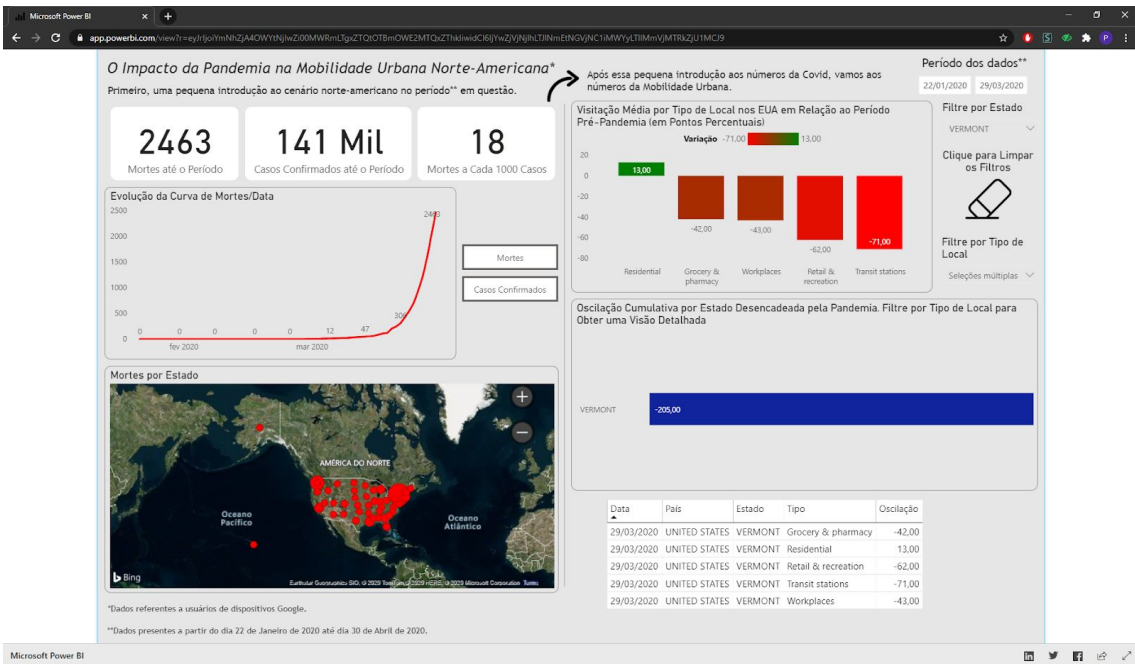


Figura 4



O Impacto da Pandemia na Mobilidade Urbana Norte-Americana*

Primeiro, uma pequena introdução ao cenário norte-americano no período** em questão.

Após essa pequena introdução aos números da Covid, vamos aos números da Mobilidade Urbana.

Métricas Principais:

- 20 Mil Mortes até o Período
- 525 Mil Casos Confirmados até o Período
- 39 Mortes a Cada 1000 Casos

Evolução da Curva de Mortes/Data

Gráfico de linha mostrando a evolução das mortes e casos confirmados ao longo do tempo (fev 2020 a abr 2020).

Visitação Média por Tipo de Local nos EUA em Relação ao Período Pré-Pandemia (em Pontos Percentuais)

Tipo de Local	Varição (%)
Residencial	+13,07
Parks	+10,03
Grocery & pharmacy	-13,14
Workplaces	-36,63
Retail & recreation	-46,12
Transit stations	-47,07

Oscilação Cumulativa por Estado Desencadeada pela Pandemia. Filtre por Tipo de Local para Obter uma Visão Detalhada

District/State	Oscillation
DISTRICT OF COLUMBIA	-262,00
HAWAII	-253,00
FLORIDA	-213,00
Nevada	-193,00
MONTANA	-187,00
TEXAS	-184,00
CALIFORNIA	-183,00
VERMONT	-174,00
NEW YORK	-173,00
NEW JERSEY	-170,00
ARIZONA	-166,00

Mortes por Estado

Mapa mostrando a distribuição geográfica das mortes por estado nos Estados Unidos.

Tabela de Dados Detalhados (Filtro: ALABAMA):

Data	País	Estado	Tipo	Oscilação
29/03/2020	UNITED STATES	ALABAMA	Grocery & pharmacy	-13,00
29/03/2020	UNITED STATES	ALABAMA	Parks	19,00
29/03/2020	UNITED STATES	ALABAMA	Residential	9,00
29/03/2020	UNITED STATES	ALABAMA	Retail & recreation	-41,00
29/03/2020	UNITED STATES	ALABAMA	Transit stations	-30,00
29/03/2020	UNITED STATES	ALABAMA	Workplaces	-32,00

*Dados referentes a usuários de dispositivos Google.
 **Dados presentes a partir do dia 22 de Janeiro de 2020 até dia 30 de Abril de 2020.