

Proposta de Projeto

Alexsandro Carvalho, Jeffson Simões

Abstract

Este texto é um relatório das atividades desenvolvidas no trabalho previamente proposto para a cadeira de Introdução à Ciência dos Dados do Centro de Informática - UFPE.

Keywords: data science, tourism

1. Introdução

Nosso grupo decidiu analisar a influência dos diversos empreendimentos turísticos existentes no Brasil sobre a quantidade de turistas estrangeiros que visitam nosso país. Este texto vem relatar os resultados da nossa análise.

5 2. Os Dados

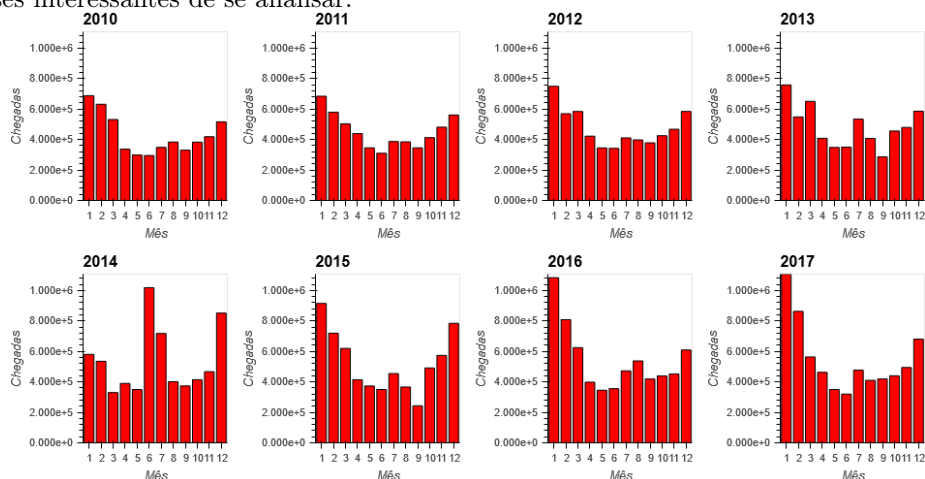
Originalmente, pretendíamos utilizar os datasets disponibilizados no Portal Brasileiro de Dados Abertos (abreviado no resto deste artigo como PBDA). Porém apenas os dados das hospedagens foram obtidos do mesmo.

10 Durante o desenvolvimento do projeto, encontramos uma ferramenta do Ministério do Turismo, o Extrator de Chegadas de Turistas Internacionais ao Brasil. Foi dada preferência a usá-lo no lugar do PBDA por 2 motivos: primeiro, porque ele estava mais atualizado e continha os dados de 2017; o segundo motivo foi porque os csvs do PBDA não mantinham um padrão nas suas colunas, o que dificultava juntar os dados.

Email addresses: `avsc@cin.ufpe.br` (Alexsandro Carvalho), `jcss3@cin.ufpe.br` (Jeffson Simões)

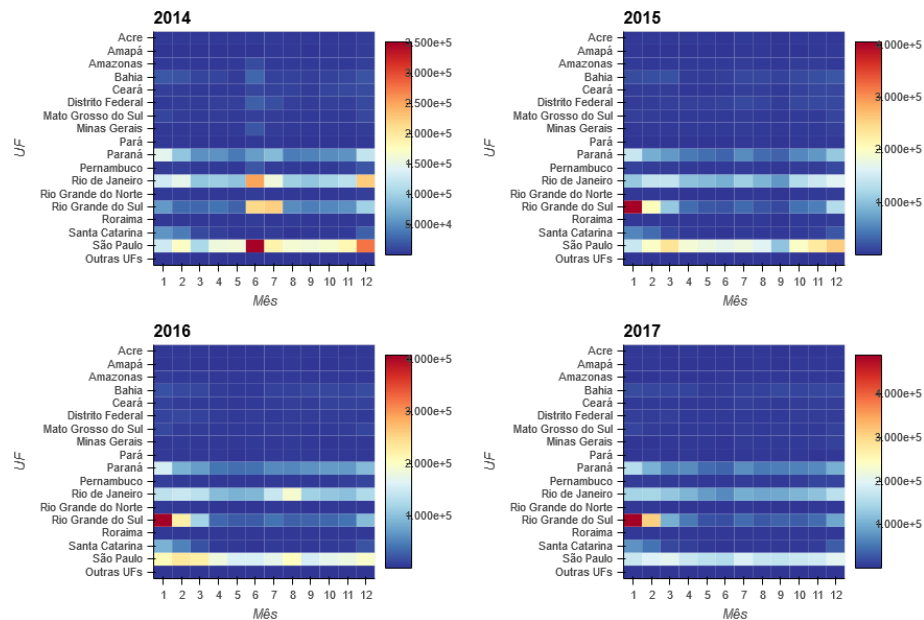
15 3. Análise Exploratória

Para entender melhor os dados a serem trabalhados, foram feitos diversos gráficos com o objetivo de revelar fenômenos dos datasets que levassem a hipóteses interessantes de se analisar.

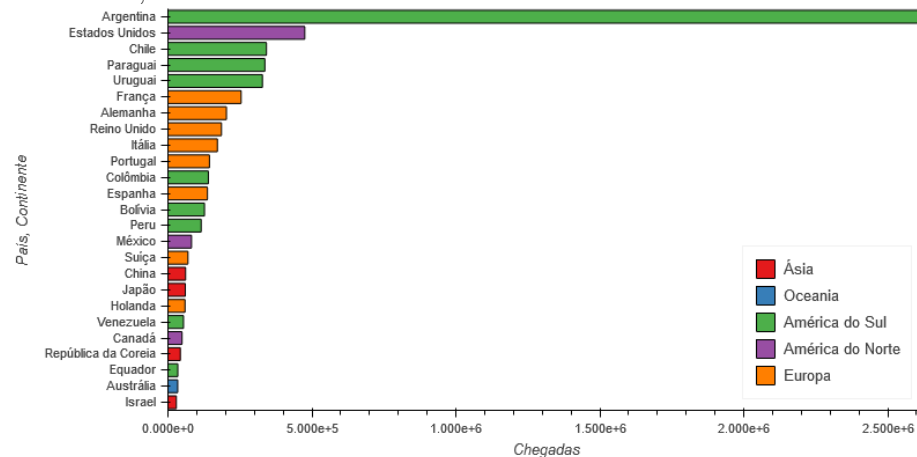


20 Inicialmente, foram feitos gráficos de barras do total de cada mês nos 8 últimos anos. Nesses gráficos, geralmente houve um padrão de o número de visitantes ser mais alto nos primeiros e últimos meses. Porém uma notável exceção foi o ano de 2014, onde os meses com maior número de turistas foram junho e julho, os meses durante os quais estava ocorrendo a Copa do Mundo.

25 Isso serviu de motivação para analisar a hipótese do efeito de eventos esportivos no turismo.



Uma nova visualização, agora exibindo as chegadas mensais por estado. Com isso, foi observado que 4 estados concentram quase todos os turistas: Paraná, Rio de Janeiro, Rio Grande do Sul e São Paulo.



Uma plotagem dos 25 países com mais turistas visitando o Brasil. Nela é possível ver que o turistas argentinos são a grande maioria dos turistas estrangeiros. Além da Argentina, os países que mais visitam o Brasil são Sulamericanos ou europeus, com os EUA sendo a única exceção.

4. Hipóteses testadas

4.1. Eventos esportivos afetam a chegada de turistas

Para avaliar a hipótese de que os eventos esportivos Copa do Mundo e Olimpíadas afetaram a chegada de turistas no Brasil, foi feita uma regressão nos
40 anos e meses, usando o histórico de 2008 a 2017. Para incluir os eventos na regressão, foram adicionadas uma variável para cada evento, indicando se ele estava ocorrendo no momento ou não.

Devido ao fato de os meses terem demonstrado uma influência não linear na chegada de turistas, também foi incluída na fórmula o quadrado do mês, pois
45 assim a fórmula se tornaria uma equação do 2º grau, que possui o comportamento de crescer nas extremidades para cima ou para baixo, dependendo da concavidade.

	coef	SE	t	p-value
Intercept	2.911358e+09	4.304096e+09	0.676416	5.001589e-01
Copa[T.True]	5.134350e+05	6.105101e+04	8.409934	1.409461e-13
Olimpíadas[T.True]	1.465364e+05	8.539026e+04	1.716079	8.888788e-02
Mês	-1.685253e+05	9.851360e+03	-17.106805	3.825668e-33
I(Mês ** 2)	1.166118e+04	7.376492e+02	15.808564	2.108632e-30
Ano	-2.907907e+06	4.277363e+06	-0.679837	4.979984e-01
I(Ano ** 2)	7.263264e+02	1.062697e+03	0.683474	4.957062e-01

Os valores gerados acima mostram que a Copa do Mundo tem de fato efeito
50 significativo (devido ao p-value baixo) e positivo, tendo atraído cerca de 500 mil turistas ao Brasil.

As Olimpíadas, por outro lado, tiveram um desempenho menor, e o p-value alto associado ao evento demonstra que ela não teve impacto significativo no número de turistas no Brasil. Porém, diferentemente da Copa do Mundo, as
55 Olimpíadas foram um evento mais concentrado, ocorrendo apenas no Rio de Janeiro, sendo assim, foi feita uma nova regressão, porém com os turistas do Rio de Janeiro.

Analisando as outras variáveis da regressão, é possível observar que o ano teve um p-value mais alto que o das Olimpíadas, indicando que ele é ainda
 60 menos influente que o evento. Isso pode ser um sinal de que o turismo brasileiro não possui crescimento anual, o que será investigado na próxima hipótese.

Já o mês, demonstrou o comportamento esperado. Como $I(\text{Mês}^{**2})$ possui coeficiente positivo, isso indica que a concavidade do gráfico está para cima, como observado na análise exploratória.

	coef	SE	t	p-value
Intercept	-2.267001e+09	9.689132e+08	-2.339736	2.105321e-02
Copa[T.True]	1.149246e+05	1.374345e+04	8.362137	1.810564e-13
Olimpíadas[T.True]	9.435520e+04	1.922256e+04	4.908566	3.112915e-06
Mês	-2.096289e+04	2.217681e+03	-9.452616	5.638348e-16
I(Mês ** 2)	1.549951e+03	1.660553e+02	9.333942	1.061619e-15
Ano	2.247202e+06	9.628952e+05	2.333797	2.137418e-02
I(Ano ** 2)	-5.568525e+02	2.392283e+02	-2.327704	2.170797e-02

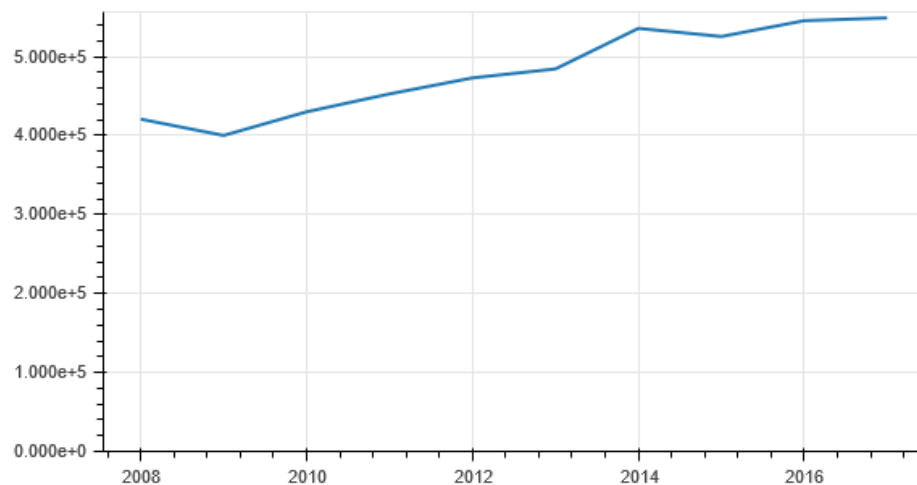
65

Na segunda regressão, restrita ao Rio de Janeiro, as Olimpíadas tiveram um p-value bem menor, e seu coeficiente se aproximou do da Copa do Mundo. Isso mostra que a localidade das Olimpíadas fez com que o efeito delas também fosse local.

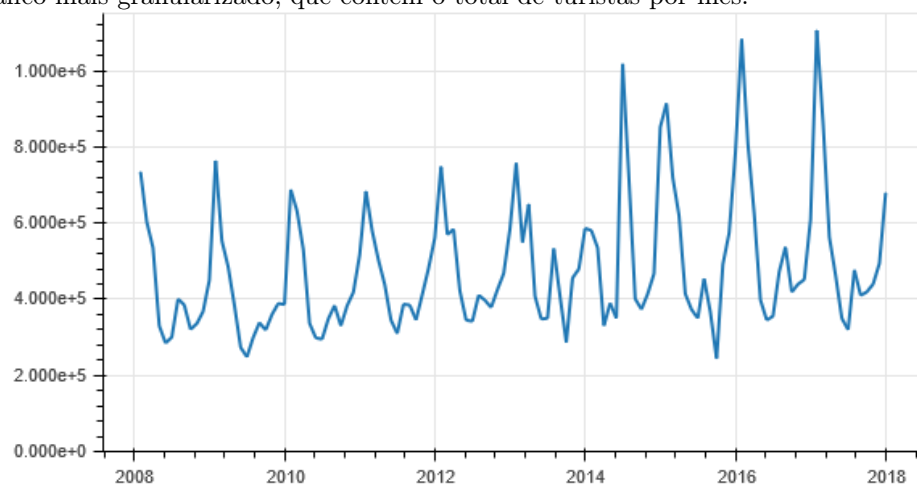
70 A conclusão geral dessa hipótese é a de que os eventos esportivos impulsionam o turismo, porém a abrangência desse impulso é limitada aos locais onde o evento está ocorrendo.

4.2. O turismo no Brasil não está crescendo

Para avaliar essa hipótese, foi plotado um gráfico contendo a média de tu-
 75 ristas pra cada ano de 2008 a 2017.



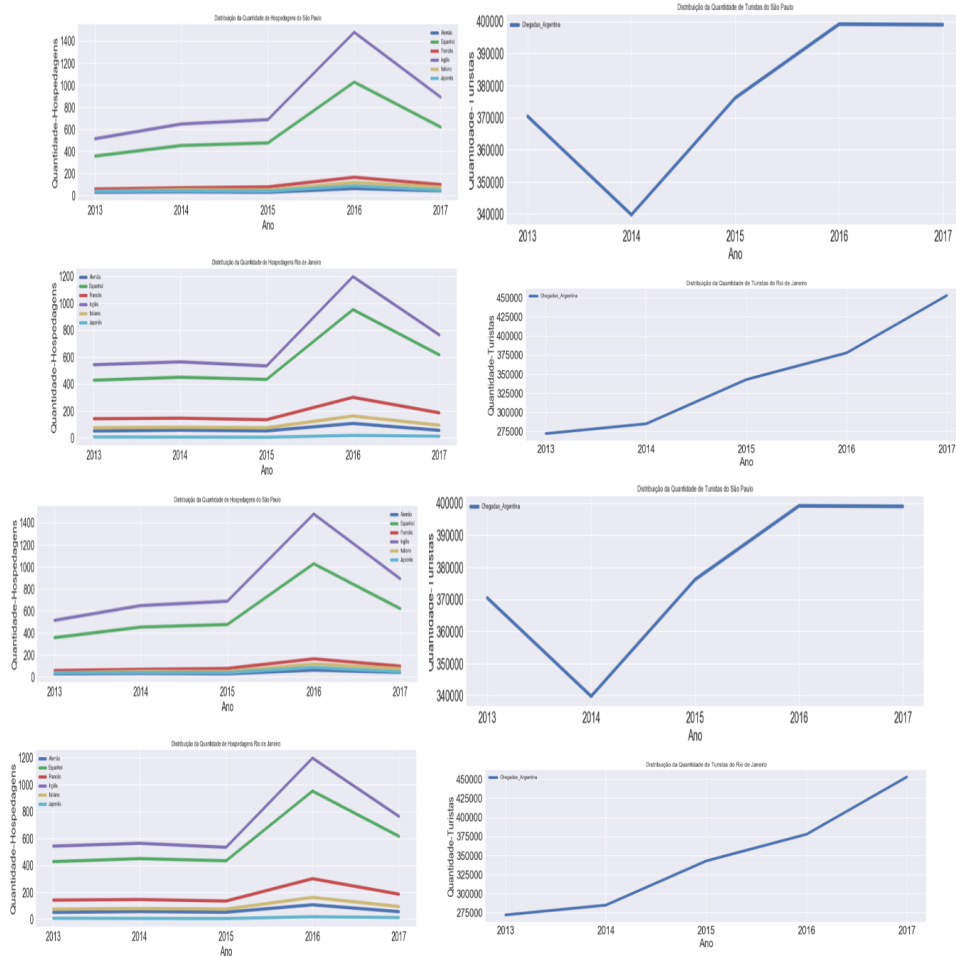
O gráfico demonstrou que houve aumento, o que a princípio, parece contradizer a hipótese. Para entender melhor a situação, foi plotado um segundo gráfico mais granularizado, que contém o total de turistas por mês.



80

O segundo gráfico mostra o porquê do ano não demonstrar influência na chegada pela regressão. As oscilações causadas pelo efeito dos meses são tão grandes que elas mascaram o crescimento anual do número de turistas. A conclusão aqui é a de que o turismo está de fato crescendo com os anos, mas pouco.

85 4.3. Causalidade entre quantidade de hospedagens que oferecem X idioma no atendimento e quantidade de turistas estrangeiros



Para avaliar as hipóteses “As Hospedagens, que oferecem X idioma no atendimento a Turistas estrangeiros, influência no fluxo de Turistas?” , “A quantidade de Turistas influencia na quantidade de hospedagens daquela região” e a última “As Olimpíadas influenciaram no aumento de hospedagens nas regiões do Rio e alguns estados próximos ?” tentamos usar um modelo de predição para séries temporais (modelo Arima) mas não tivemos muito sucesso ao tentar implementá-lo.

Com isso partimos para uma abordagem de menor acurácia que seria a de

criar gráficos das séries temporais obtidas - Chegadas de Turistas e Quantidade de Hospedagens - e comparar as variáveis nos dois gráficos de acordo com o intervalo de tempo, no caso, comparamos a quantidade de hospedagens que
100 ofereciam atendimento em espanhol com a série temporal da quantidade de turistas argentinos que chegavam. Devido a encontrarmos um pico na quantidade de hospedagens no ano de 2016 pudemos ver que as Olimpíadas influenciaram na quantidade de hospedagens.

Devido a encontrarmos um pico na quantidade de hospedagens no ano de
105 2016 pudemos ver que as Olimpíadas influenciaram na quantidade de hospedagens.

5. Conclusão

Analisando os dados pudemos entender um pouco sobre a situação do turismo brasileiro, quem nos visita, quando somos visitados, para onde esses turistas vão, e saber o que os atrai e o que não os atrai para cá.
110