

Uso de novas técnicas para obtenção de dados de inflação de passagem aérea para o SNIPC

VI Semana de Qualidade da Informação do Transporte Aéreo
SP, 27/11/18

Vladimir Miranda – IBGE/GEPLA

vladimir.miranda@ibge.gov.br

Estrutura da apresentação

- i) Rápida introdução a IPCs
- ii) Metodologia de cálculo da inflação para a passagem aérea no SNIPC
- iii) Uso de ferramentas de web scraping para a coleta de dados: benefícios e desafios
- iv) Pontos para possíveis colaborações
- v) Conclusões;

Breve introdução a IPCs

Breve introdução a IPCs

Índices de preço ao consumidor (IPC): são medidas que buscam refletir a variação nos custos de uma cesta de bens e serviços, **consumida por uma população de referência**, ao longo do tempo.

A cesta de bens e serviços é determinada pelas pesquisas de orçamento familiar (POF), que determinam, além dos constituintes da cesta, os pesos relativos de cada um deles em relação ao total de bens e serviços da cesta.

A POF também dá o peso relativo a cada estado em relação ao “consumo nacional”.

Índices de preços ao consumidor são calculados a partir de uma combinação de preços e pesos associados aos constituintes da cesta de consumo e agregados segundo um dado Sistema de classificação das despesas.

Introdução ao SNIPC

Populações de referência e cobertura para o Brasil, atualmente:

IPCA: famílias de 1 a 40 salários mínimos e residentes nas áreas urbanas do Brasil;

INPC: famílias de 1 a 5 salários mínimos e residentes nas áreas urbanas do Brasil;

Abrangência geográfica:



Introdução ao SNIPC

Sistema de classificação SNIPC:

Grupo	→	5000000	Transportes
Subgrupo	→	5100000	Transportes
Item	→	5101000	Transporte público
Subitem	}	5101001	Ônibus urbano
		5101002	Táxi
		5101004	Trem
		5101006	Ônibus intermunicipal
		5101007	Ônibus interestadual
		5101010	Passagem aérea
		5101011	Metrô
		5101022	Transporte hidroviário
		5101026	Transporte escolar
Produtos	}	Passagem aérea Rio-São Paulo, Tarifa “alfa”, com bagagem..., ida data xx, volta data yy.	
		Passagem aérea Rio-São Paulo, Tarifa “beta”, com bagagem..., ida data xx, volta data yy.	

Aumentando agregação

Introdução ao SNIPC

Aumentando a agregação para gerar o índice final, ou subíndices intermediários



		Descrição	Peso (%)
Transporte público	4,71		
Ônibus urbano	2,69	Transportes	18,10
Táxi	0,35		
Trem	0,05		
Ônibus intermunicipal	0,78		
Ônibus interestadual	0,25		
Passagem aérea	0,43		
Metrô	0,06	Transporte público	4,71
Transporte hidroviário	0,01	Veículo próprio	8,17
Transporte escolar	0,09	Combustíveis	5,22

Descrição	Peso (%)
Índice Geral	100,00
Alimentação e bebidas	24,71
Habituação	15,71
Artigos de residência	4,03
Vestuário	5,94
Transportes	18,10
Saúde e cuidados pessoais	12,04
Despesas pessoais	10,95
Educação	4,85
Comunicação	3,67

Peso subitem: em geral, nível mais desagregado com ponderação.

Áreas	Peso (%)
Brasília	2,06
Vitória	0,78
Rio de Janeiro	0,66
Curitiba	0,48
São Paulo	0,41
Belém	0,35
Salvador	0,33
Goiânia	0,26
Belo Horizonte	0,26
Campo Grande	0,23
Recife	0,21
Porto Alegre	0,18
Fortaleza	0,17
Brasil	0,43

Pesos regionais da passagem aérea.

Aspectos básicos para cálculo de um IPC

Cálculo do IPC dividido em duas etapas: cálculos para os agregados elementares (“produtos”) e agregação.

Fórmula para os agregados elementares (caso da passagem aérea) no SNIPC:

$$R_{t-1;t}^k = \left(\prod_{j=1}^m R_{t;t-1}^j \right)^{1/m}$$

$R_{t;t-1}^j$ é a média dos relativos do produto j , que fazem parte do subitem k .

Fórmula para agregação (pesos e preços):

$$L_{0;t}^b = \left\{ \sum_{k=1}^n W_0^k R_{0;t}^k \right\} / \left\{ \sum_{k=1}^n W_0^k \right\}$$

Onde W_0^k são os pesos da POF para o subitem k e $L_{0;t}^b$ é o índice para o item b entre a base de pesos e o mês de referência t .

Aspectos básicos para cálculo de um IPC

Na prática:

$$L_{t-1;t}^b = \left\{ \sum_{k=1}^n W_{t-1}^k R_{t-1;t}^k \right\} / \left\{ \sum_{k=1}^n W_{t-1}^k \right\}$$

W_{t-1}^k é peso atualizado no instante t-1.

Aspectos básicos para cálculo de um IPC

O cálculo de IPCs é baseado no **método do modelo pareado**: variações de preços dos mesmos produtos, nos mesmos locais, são seguidas ao longo do tempo.

Comparabilidade é fundamental.

Mês:

t-1



P_{t-1}

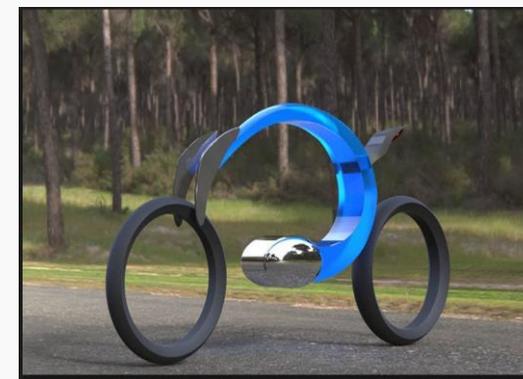
(comparável)



P_t

t

(Não comparável)



Principal personagem: relativo de preços

$$R_{t;t-1} = P_t / P_{t-1}$$

Metodologia de Cálculo da inflação para a passagem aérea no SNIPC

Alguns desafios para obter a inflação de passagem aérea

- I. Existem inúmeros tipos de passagens aéreas diferentes. Quais devem ser utilizadas para o cálculo da inflação?
- II. Há um limite de recursos e acesso para a coleta das passagens.
- III. Como acomodar o método do modelo pareado aqui? Tipos de produtos oferecidos nem sempre se repetem.
- IV. O objetivo do IPC é pegar apenas passagens consumidas pelas famílias de referência do IPC. Como separar este alvo de compras comerciais?
- V. Quais as rotas a serem consideradas?
- VI. Qual o peso das companhias em cada rota?
- VII. Qual o peso dos diferentes tipos de passagens?

Método atualmente usado

- I. Tipo de passageiro e quantidade de bilhetes: adulto, um bilhete.
- II. Origem dos voos: capital de cada um dos estados que integram o SNIPC.
- III. Destino: capitais dos estados onde estão localizadas as cidades mais procuradas do país por motivo de lazer.
- IV. Aeroportos considerados: o principal aeroporto da capital
- V. Horários: todos os horários dos voos de interesse anunciados no momento da coleta.
- VI. Modalidades de tarifa: tarifas mais comumente oferecidas, com inclusão de uma bagagem de 23kg, no máximo. Também são consideradas tarifas que já incluem o direito a despachar uma bagagem.

Método atualmente usado

- VII. Pesquisa das tarifas: coleta semanal, em um dia específico.
- VIII. Tipo de produto: passagem de ida e volta, comprada com 60 dias de antecedência, para voos com duração de 8 dias, com partida no sábado de uma semana e retorno no domingo da semana seguinte.
- IX. A inflação do mês t é a dada pelos preços coletados em $t-2$.
- X. Cálculo: o valor total médio da passagem aérea é calculado, a cada mês, através da média aritmética simples, incluindo ida e volta acrescentando a taxa de embarque, de cada modalidade de tarifa pesquisada, por empresa.

Este valor médio comparado ao do mês anterior, resulta no relativo de preços de cada modalidade de tarifa

O resultado final é obtido através da média geométrica simples dos relativos de preços das modalidades de tarifas pesquisadas.

Cálculo realizado

Passo 1: média nos produtos (definido por cia aérea, rota e tipo de tarifa)

Produto 104			
Tipo	Preço	Tipo	Preço
Ida	289,31	Volta	189,33
Ida	159,31	Volta	1710,9
Ida	179,31	Volta	1710,9
Ida	529,31	Volta	278,33
Ida	189,21	Volta	289,8
Ida	209,21	Volta	209,33
Ida	179,31	Volta	209,33
Ida	219,31	Volta	159,8
Ida	159,31	Volta	1710,9
Ida	208,84	Volta	219,8
Ida	1710,41	Volta	1710,9
Ida	179,31	Volta	179,8
Ida	319,21	Volta	1710,9
Ida	209,21	Volta	219,8
Ida	289,31	Volta	248,8
Ida	208,84	Volta	159,8
Ida	159,31	Volta	179,8
Ida	159,31	Volta	159,8
Ida	189,21	Volta	159,8
		Volta	119,8

Remessa: 7 Local: 074880

Produto: 05101010000104

IDA

Data: 01/10/2017 Tx. Bag.: 30,00 Média:  332,45

VOLTA

Data: 08/10/2017 Tx. Bag.: 30,00 Média:  606,88

Método atualmente usado

Cada remessa fornece um valor médio para cada produto.

Remessa: 7 Local: 074880

Produto: 05101010000104

Nome Produto: Empresa aérea - Cidade origem - destino TARIFA abc (mais bagagem até 23kg) – IDA E VOLTA COM TAXAS DE EMBARQUE (PREÇO MÉDIO)

IDA				VOLTA				Data coleta
Data	Tx. Bag.	N. voos	Média	Data	Tx. Bag.	N. voos	Média	
01/10/2017	30	19	332,45	08/10/2017	30	20	606,88	01/08/2017
07/10/2017	30	19	328,77	15/10/2017	30	20	703,38	09/08/2017
14/10/2017	30	20	621,38	22/10/2017	30	20	758,58	16/08/2017
21/10/2017	30	20	449,94	29/10/2017	30	20	584,29	23/08/2017
28/10/2017	30	20	452,05	05/11/2017	30	20	578,46	30/08/2017
Média			439,09				646,32	<u>1085,41</u>

Produto: 05101010000105

Nome Produto: Empresa aérea - Cidade de origem - destino TARIFA bcd – IDA E VOLTA COM TAXAS DE EMBARQUE (PREÇO MÉDIO)

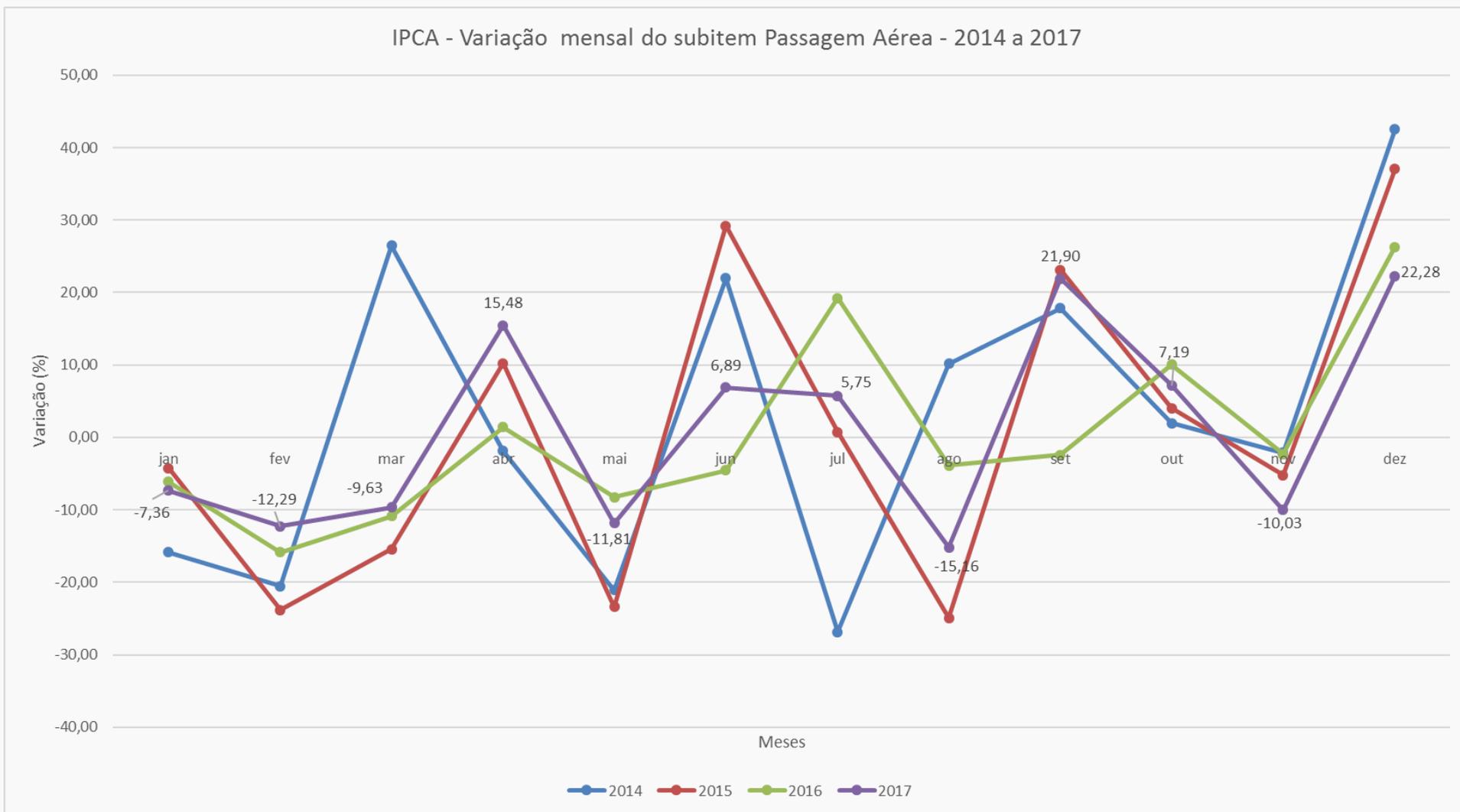
IDA				VOLTA				Data coleta
Data	Tx. Bag.	N. voos	Média	Data	Tx. Bag.	N. voos	Média	
01/10/2017	0	19	374,54	08/10/2017	0	20	647,53	01/08/2017
07/10/2017	0	19	388,46	15/10/2017	0	20	742,52	09/08/2017
14/10/2017	0	20	665,18	22/10/2017	0	20	803,26	16/08/2017
21/10/2017	0	20	507,87	29/10/2017	0	20	642,36	23/08/2017
28/10/2017	0	20	486,62	05/11/2017	0	20	628,81	30/08/2017
Média			486,64				692,90	<u>1179,53</u>

Passo 2: média ponderada dos produtos.

Passo 3: Cálculo dos relativos das médias de produtos.

Passo 4: média geométrica destes relativos.

Alguns resultados



Práticas em acordo com metodologias internacionais. **Permitem comparabilidade.**

Prática internacional



“Another version of matching methodology is to select some elements of a tariff as “representative items” and re-price them in subsequent periods. ...
...For instance, for air fares this could be for each airline carrier, a non-refundable and non-changeable airline fare from one pre-specified location to another, with pre-determined outbound and inbound dates chosen by time of day and day of week, including all surcharges.”

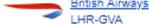
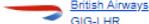
“Once chosen in the base period, the route, the departure and arrival times and the ticket type and class of travel should remain the same throughout the year”

“Transport ticket prices should feed into the index at the time of travel, not at the time they were booked. So, a ticket price for a December flight should feed into the December index even if the price relates to a purchase in October.” **CPI manual, Chapter 11.**

Uso de ferramentas de web scraping para a coleta de dados: benefícios e desafios

Coleta tradicional

- I. Coleta tradicional: cada uma das 16 unidades locais coleta manualmente os preços das rotas selecionadas para o SNIPC nos sites das companhias aéreas.

Previous Week <input checked="" type="checkbox"/> Also change my inbound dates Next Week						
Sat 08 Sep from \$595	Sun 09 Sep from \$723	Mon 10 Sep from \$510	Tue 11 Sep from \$550	Wed 12 Sep from \$550	Thu 13 Sep from \$530	Fri 14 Sep from \$550
Travel Classes						
Departs	Arrives	Flight Operator	Economy	Premium Economy	Business Class	
Sorry, there are no direct flights for this route, flights with connections are below.						
Outbound connecting flight options						
21:50 11 Sep GIG	18:00 12 Sep GVA	 British Airways GIG-LHR  British Airways LHR-GVA	<input type="radio"/> \$550 <input checked="" type="radio"/> Lowest	<input type="radio"/> \$1450 <input checked="" type="radio"/> Lowest	<input type="radio"/> \$1425 <input checked="" type="radio"/> Lowest	1 Connection Show journey details + Total journey time: 15 hours 10 minutes
21:50 11 Sep GIG	20:15 12 Sep GVA	 British Airways GIG-LHR  BA CityFlyer LCY-GVA	<input type="radio"/> \$550 <input checked="" type="radio"/> Lowest	<input type="radio"/> \$1450 <input checked="" type="radio"/> Lowest	<input type="radio"/> \$1425 <input checked="" type="radio"/> Lowest	1 Connection Show journey details + Total journey time: 17 hours 25 minutes
21:50 11 Sep GIG	20:25 12 Sep GVA	 British Airways GIG-LHR  British Airways LHR-GVA	<input type="radio"/> \$550 <input checked="" type="radio"/> Lowest	<input type="radio"/> \$1450 <input checked="" type="radio"/> Lowest	<input type="radio"/> \$1425 <input checked="" type="radio"/> Lowest	1 Connection Show journey details + Total journey time: 17 hours 35 minutes
21:50 11 Sep GIG	22:25 12 Sep GVA	 British Airways GIG-LHR  British Airways LHR-GVA	<input type="radio"/> \$550 <input checked="" type="radio"/> Lowest	<input type="radio"/> \$1450 <input checked="" type="radio"/> Lowest	<input type="radio"/> \$1425 <input checked="" type="radio"/> Lowest	1 Connection Show journey details + Total journey time: 19 hours 35 minutes

Coletas para todos os voos para rotas pré-determinadas, para bilhetes comprados com 2 meses de antecedência à data da viagem e dias de ida e volta e tempo de permanência pré-estabelecidos.

Dados coletados para diferentes tipos de bilhetes.

Coletas para diferentes companhias.

Dados coletados uma vez por semana.

Coleta tradicional

- I. O processo é demorado: approx. 4h para cada coleta. **Cerca de 16h por mês, por área.**
- II. Mais sujeito a erros. Demanda de tempo extra de análise da equipe de escritório. 1-2h por mês.
- III. Desde o último trimestre de 2017, coleta extra necessária devido à implementação do programa de comparação de preços internacional da CEPAL:
 - a) Novas rotas e companhias adicionadas.
 - b) Coleta 3 semanas por mês para mais 2 áreas.
 - c) Implicação de mais 6h de coleta manual por mês.

Pergunta inicial: é possível melhorar este processo usando técnicas de web scraping?

Desafios para implementação da coleta por robô

Necessário um programa para extrair os dados da página.

Html por trás do web site de uma companhia aérea: necessário “organizar” os dados num formato estruturado.

1 opções de voos de VOLTA

Filtro +

Organizar por decolagem mais ▼	MAX	PLUS	LIGHT
	1ª e 2ª bagagens gratuitas ✓ R\$ 1 = 4 milhas Smiles ✓ Assento GOL+ Conforto gratuito ✓ Antecipação gratuita e mais vantagens +	1ª bagagem gratuita ✓ R\$ 1 = 3 milhas Smiles ✓ Marcação de assento gratuita ✓ Antecipação gratuita e mais vantagens +	Sem bagagem gratuita ✓ R\$ 1 = 2 milhas Smiles ✓ Marcação de assento gratuita no período de check-in e mais vantagens +
	R\$ 733,17	R\$ 643,17	MENOR PREÇO DO DIA R\$ 598,17

Reexibir todos os voos

```

</td>
<td class="taxa taxaExecutiva">
  <div>
    <div class="lessPriceBox"></div>
    <div class="taxaSelected">
      <div class="checkTaxaSelected"></div>
    </div><span class="smilesAndMoneyValue"></span><label class="textIdentFareValue" for="
      R$ 643,17</span><span>tarifa Plus</span></label><input id="ControlGro
    </div>
  </td>
<td class="taxa taxaPromocional">
  <div>
    <div class="lessPriceBox">
      <div class="lessPrice">Menor Preço do Dia</div>
    </div>
    <div class="taxaSelected">
      <div class="checkTaxaSelected"></div>
    </div><span class="smilesAndMoneyValue"></span><label class="textIdentFareValue" for="
      R$ 598,17</span><span>tarifa Light</span></label><input id="ControlGr
    </div>
  </td>
</tr>
<td>
  <div id="market2_journey1" class="infoGrid bgGrid popupANAC"><span class="anacInformatio
    </span></div>
  </td>
</tr>
</table>
  
```

Desafios para implementação da coleta por robô

Demanda extra: necessidade de emular um usuário navegando na página.

The screenshot shows a flight search form with the following elements:

- Header:** "Só voo" (Only flight) and "Ferramentas de viagem" (Travel tools).
- Buttons:** "As minhas pesquisas recentes" (My recent searches).
- Origin:** "De" dropdown menu with "Brasil" selected.
- Destination:** "Para" dropdown menu with "Rio de Janeiro" selected.
- Travel Type:** Radio buttons for "viagem de ida e volta" (round trip) and "Só ida" (one-way). A link for "Várias cidades/escalas" (multiple cities/stops) is also present.
- Dates:** Two date pickers showing "11/09/18" and "28/09/18".
- Class:** "Classe de voo" dropdown menu with "Económica" selected.
- Ticket Type:** "Tipo de bilhete" dropdown menu with "Bilhete flexível" selected.
- Passenger Count:** Four columns for "Adultos (16+)", "Jovens adultos (12-15)", "Crianças (2-11)", and "Bebés (menos de 2)". Each column has a dropdown menu showing the count (1, 0, 0, 0 respectively).
- Submit:** A red button labeled "Procurar voos" (Search flights).

Blue arrows point from the text on the right to the corresponding form elements:

- From "Escolha da origem e destino." to the "De" and "Para" dropdown menus.
- From "Escolha das datas de viagem." to the date pickers.
- From "Escolha do tipo de bilhete." to the "Tipo de bilhete" dropdown menu.

Escolha da origem e destino.

Escolha das datas de viagem.

Escolha do tipo de bilhete.

Desafios para implementação da coleta por robô

Desenvolvimento de um web scraper caseiro combinando R e selenium.

Passos iniciais, foco em reproduzir o processo manual.

Resultados:

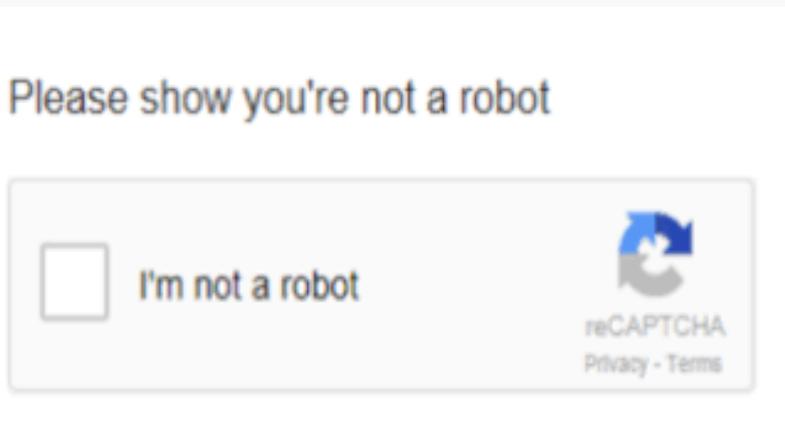
Robôs levam cerca de 30min para executar a coleta de uma área, 8 vezes mais rápido que o processo manual.

Boa concordância inicial entre dados coletados manualmente e por robô.

Scrapers também construídos para o PCI e funcionando a contento.

Desafios para a implementação (onde precisamos de vocês)

Questões legais: políticas anti-robôs



Instabilidades: sites mudam sem aviso prévio.



Cookies: de acordo com algum perfil de usuário, preços diferentes podem ser ofertados.

Cada site tem seu próprio desenho.

Pontos para colaboração: otimizar método atual e auxiliar no desenvolvimento de indicadores.

Pontos importantes

I) Para a implantação da coleta automatizada:

No processo de produção mensal, erros não são permitidos, por isso precisamos garantir acesso aos robôs e coleta da informação correta.

“Mundo ideal” seria acesso via APIs. Livre de instabilidades e bloqueios.

Necessário estabelecer termo de compromisso: não interrupção ao acesso e confiabilidade.

Processo de conversas para colaboração com as aéreas iniciado com intermédio da ANAC.

Pontos importantes

II) Possibilidades de melhoramentos metodológicos

Coleta mais intensa de dados web: mais representatividade e menor influência da precificação dinâmica.

Acesso ao “scanner data” das companhias:

Desenvolvimento de indicadores com fórmulas mais robustas e índices hedônicos;

Ponderação dos produtos vendidos;

Possibilidade de construção de índice(s) com escopo mais geral que o do SNIPC.

Conclusões

Cálculo da inflação para a passage aérea no SNIPC diferente da inflação geral do setor da aviação.

Derivação da inflação usada por IPCs baseada em alguns tipos de hipóteses e no tipo de dados disponíveis.

Avanço de estruturas de TI propiciam um novo caminho para melhoramento dos indicadores.

Melhoramentos dependem da colaboração com os especialistas do setor e de acesso às informações detidas pelas companhias e empresas do setor.

Obrigado pela atenção!

vladimir.miranda@ibge.gov.br