



## FICHA METODOLÓGICA

### Informações Gerais sobre a Base de Dados

<b>Nome</b>	CGIL
<b>Descrição</b>	Autorizações de residência para fins laborais e de investimentos
<b>Fonte</b>	Ministério da Justiça e Segurança Pública
<b>Período Trabalhado</b>	2023 a 2025

### Variáveis Utilizadas

Variáveis Originais	Nome Variável	Descrição Variável
	Nome do Estrangeiro	Nome do trabalhador imigrante
	Data de Nascimento	Data de nascimento do trabalhador
	País de Origem	País de nascimento do trabalhador
	Gênero	Sexo Biológico
	Tipo de Visto	Classifica as autorizações em residência e residência prévia
	CBO 2002	Classificação Brasileira de Ocupação
	UF	Unidade da Federação
	Andamento	Situação do processo
	Modalidade	Indica se autorização está sendo analisada na Coordenação Geral de Imigração Laboral (CGIL) ou no Conselho Nacional de Imigração (CNIg)
	Escolaridade	Escolaridade do trabalhador
	Valor de investimento	Valores investidos em pessoas jurídicas ou em imóveis
	CNAE	Cadastro Nacional de Atividades Econômicas

Variáveis Geradas	Nome Variável Original	Nome Variável Gerada	Descrição Variável Gerada
	Amparo_legal	RN	Resolução Normativa
	Data de nascimento	Idade	Idade
	País	País_IBGE	Nome do país conforme ortografia considerada pelo IBGE
	CBO 2002	Grupo ocupacional	
	UF	Regiões	
	RN	Qualificado	Indicadora se o migrante é qualificado ou não
	Valor de investimento	Valor_ajustado	Valor de investimento deflacionado com base nos índices do IGP-M
	CNAE	Setores	Setores e subsetores de atividade

### Tratamento da Base

A limpeza aplicada aos dados da CGIL a partir de 2018 segue os seguintes passos:

**Passo 1 (Separação dos tipos de residência).** Os dados de autorização foram separados por aqueles que possuem concomitância (estes não sofrem alterações), Residência Prévia e Residência.

**Passo 2 (Residência Prévia).** No caso de migrantes com mais de uma autorização de residência prévia e o tempo entre duas delas for inferior a três meses, a mais antiga é mantida e eliminada a mais recente.

**Passo 3 (Residência).** No caso de migrantes com mais de uma autorização de residência e o tempo entre duas delas for inferior a três meses, a mais antiga é mantida e eliminada a mais recente.

### Limites e Potencialidades

#### Potencialidades

- O conjunto de variáveis possibilita estabelecer o perfil sociodemográfico das solicitações;
- Permite que se dimensione e acompanhe a evolução da demanda por mão-de-obra imigrante no mercado de trabalho formal do país;
- Diferentemente das bases de registros migratórios, a profissão declarada é exatamente aquela que o imigrante exercerá, caso efetivamente ingresse no mercado laboral brasileiro

#### Limitações

- A base de dados não mostra o quantitativo de migrantes autorizados a trabalhar no Brasil, no respectivo ano, e sim o número de autorizações. A análise por indivíduo necessita de tratamento prévio na base de modo a evitar duplicações;
- A desagregação espacial chega apenas no nível de Unidade da Federação;
- Outro ponto a ser destacado é que nem todas as autorizações concedidas são efetivadas, pois como o migrante as recebem antes de chegar ao país



## FICHA METODOLÓGICA

### Informações Gerais sobre a Base de Dados

<b>Nome</b>	Cadastro Geral de Admitidos e Demitidos – CAGED
<b>Descrição</b>	O CAGED é uma base de dados do Ministério do Trabalho e Emprego (MTE) com os registros mensais de admissões e desligamentos do imigrante no mercado formal de trabalho
<b>Fonte</b>	MTE
<b>Período Trabalhado</b>	2023 a 2025

### Variáveis Utilizadas

Variáveis Originais	Nome Variável	Descrição Variável
	competenciamov	competencia
sexo	sexo	Sexo do trabalhador
idade	idade	Idade do trabalhador
graudeinstrucao	instrucao	Grau de instrução do trabalhador
uf	uf	Unidade da Federação de acordo com o código do IBGE
municipio	municipio	Município, sem tratamento de confidencialidade
cbo2002ocupacao	cbo2002ocupacao	Código da ocupação do trabalhador de acordo com a Classificação Brasileira de Ocupações (CBO 2002)
subclasse	subclasse	Código da subclasse da classificação nacional de atividade econômica (CNAE 2.0)

  

Variáveis Geradas	Nome Variável Original	Nome Variável Gerada	Descrição Variável Gerada
	paistenacionalidade	paistenacionalidade	pais_ibge
paistenacionalidade	paistenacionalidade	continente	Continente do país de nacionalidade do trabalhador, em formato padronizado. Previamente o dado era obtido através do pareamento com CTPS e RAIS
competenciamov	competenciamov	ano	Ano do movimento
competenciamov	competenciamov	quadrimestre	Quadrimestre do movimento
uf	uf	gr	Grande Região: primeiro dígito da UF
movimento	movimento	admitidos_desligados, saldo_mov, movimento_num	Tipo de movimento
idade	idade	idade_tab, idade_tab2	Categorizações da idade do trabalhador
graudeinstrucao	graudeinstrucao	instrucao_tab, instrucao_tab2	Grupos de instrução
graudeinstrucao	graudeinstrucao	qualificado	Foram definidos como migrantes qualificados aqueles com ao menos ensino superior completo
graudeinstrucao, cbo2002ocupacao	graudeinstrucao, cbo2002ocupacao	altqualificado	Foram definidos como migrantes altamente qualificados aqueles que, além de qualificados, estavam ocupados nos grupos ocupacionais 0, 1 ou 2
salário	salário	salario_real	Salário deflacionado pelo INPC e com remoção dos outliers
cbo2002ocupacao	cbo2002ocupacao	grupos_ocup	Grupo ocupacional
cbo2002ocupacao	cbo2002ocupacao	subgrupos_ocup	Subgrupo ocupacional
cbo2002ocupacao	cbo2002ocupacao	subgrupos_ocup3	Subgrupo ocupacional 3 dígitos
subclasse	subclasse	divisao_ativ	Divisão de atividade

## Tratamento da Base

O processo de integração e tratamento conjunto das bases administrativas da Carteira de Trabalho e Previdência Social (CTPS), da Relação Anual de Informações Sociais (RAIS) e do Cadastro Geral de Admitidos e Demitidos (CAGED), com o objetivo de ampliar a identificação e a caracterização dos trabalhadores migrantes no mercado formal de trabalho brasileiro, gera, como resultado, o que se denominou como bases harmonizadas. A utilização integrada dessas bases permite explorar informações sobre estoque de vínculos, movimentações de admissões e desligamentos e características sociodemográficas e ocupacionais dos trabalhadores.

Inicialmente – até 2019 – a base do CAGED não possuía a informação do país de nacionalidade ou de nascimento de cada indivíduo, de forma que este dado precisou ser agregado a partir de outra fonte. Contudo, em atualização posterior, tais variáveis passaram a ser incluídas no extrato do CAGED disponibilizado. Para os anos mais recentes, especialmente a partir de 2023, a maior desagregação das informações do CAGED, aliada à descontinuidade da base da CTPS e à crescente convergência das informações, levou à decisão metodológica de não realizar mais o pareamento com as bases RAIS e CTPS, permanecendo-se com os registros originários do CAGED para os estudos de movimentação no mercado de trabalho formal.

Para a construção da base de dados são removidos os registros removendo os registros de brasileiros ou que o dado de nacionalidade era ausente. Para as análises, são desconsiderados os registros de naturalizados. É importante mencionar que o MTE organiza estas bases em três arquivos, cada um contemplando um conjunto distinto de informações:

- a) Movimentações declaradas dentro do prazo: registros que ocorreram e foram declarados no próprio mês, ou dentro do prazo estabelecido;
- b) Movimentações declaradas fora do prazo: registros que ocorreram em um mês mas foram declarados posteriormente, devem ser considerados nas estatísticas e ocasionam pequenas variações nos agregados; e c) Movimentações excluídas: referentes a registros declarados em momento passado mas que foram excluídos por alguma necessidade. Estes são contabilizados de maneira a desconsiderar o movimento, ou seja, uma admissão na base de excluídos representa uma admissão a menos.

## Limites e Potencialidades

### Potencialidades

- A base do CAGED possui periodicidade mensal, de forma que permite acompanhar a evolução da inserção dos trabalhadores no mercado de trabalho formal;
- Desagregação geográfica e por classificações de ocupação ou atividade permitem um detalhamento da informação;
- A partir da implantação do eSocial, a RAIS e o CAGED passaram a compartilhar uma origem administrativa comum, baseada em imputação de dados de diferentes fontes, o que impactou na natureza das próprias bases de dados, derivados de uma mesma infraestrutura informacional

### Limitações

- Cobertura restrita a empregos formais;
- Ao longo da série histórica (desde 2011), ocorreram ajustes e inovações metodológicas que trazem a necessidade de cautela na interpretação dos resultados, pela possibilidade de descontinuidade na série histórica



## FICHA METODOLÓGICA

### Informações Gerais sobre a Base de Dados

<b>Nome</b>	Relação Anual de Informações Sociais – RAIS
<b>Descrição</b>	A RAIS é um registro administrativo anual do MTE, que fornece informações sobre o mercado de trabalho formal no Brasil, com informações amplas sobre vínculos empregatícios
<b>Fonte</b>	MTE
<b>Período Trabalhado</b>	2010 a 2024

### Variáveis Utilizadas

<b>Variáveis Originais</b>	<table border="1"> <thead> <tr> <th>Nome Variável</th> <th>Descrição Variável</th> </tr> </thead> <tbody> <tr> <td>cod_sexo</td> <td>Sexo do trabalhador</td> </tr> <tr> <td>val_idade</td> <td>Idade do trabalhador</td> </tr> <tr> <td>cod_grau_instrucao</td> <td>Grau de instrução do trabalhador</td> </tr> <tr> <td>cod_ufmun</td> <td>Unidade da Federação de acordo com o código do IBGE</td> </tr> <tr> <td>cod_cbo</td> <td>Código da ocupação do trabalhador de acordo com a Classificação Brasileira de Ocupações (CBO 2002)</td> </tr> <tr> <td>cod_cnae20</td> <td>Código da subclasse da classificação nacional de atividade econômica (CNAE 2.0)</td> </tr> </tbody> </table>		Nome Variável	Descrição Variável	cod_sexo	Sexo do trabalhador	val_idade	Idade do trabalhador	cod_grau_instrucao	Grau de instrução do trabalhador	cod_ufmun	Unidade da Federação de acordo com o código do IBGE	cod_cbo	Código da ocupação do trabalhador de acordo com a Classificação Brasileira de Ocupações (CBO 2002)	cod_cnae20	Código da subclasse da classificação nacional de atividade econômica (CNAE 2.0)																												
	Nome Variável	Descrição Variável																																										
cod_sexo	Sexo do trabalhador																																											
val_idade	Idade do trabalhador																																											
cod_grau_instrucao	Grau de instrução do trabalhador																																											
cod_ufmun	Unidade da Federação de acordo com o código do IBGE																																											
cod_cbo	Código da ocupação do trabalhador de acordo com a Classificação Brasileira de Ocupações (CBO 2002)																																											
cod_cnae20	Código da subclasse da classificação nacional de atividade econômica (CNAE 2.0)																																											
<b>Variáveis Geradas</b>	<table border="1"> <thead> <tr> <th>Nome Variável Original</th> <th>Nome Variável Gerada</th> <th>Descrição Variável Gerada</th> </tr> </thead> <tbody> <tr> <td>cod_nacionalidade</td> <td>pais_ibge</td> <td>País de nacionalidade do trabalhador, em formato padronizado. Informação ampliada a partir do pareamento com CAGED e CTPS</td> </tr> <tr> <td>cod_nacionalidade</td> <td>continente</td> <td>País de nacionalidade do trabalhador, em formato padronizado. Informação ampliada a partir do pareamento com CAGED e CTPS</td> </tr> <tr> <td>cod_ufmun</td> <td>uf</td> <td>Unidade da Federação: 2 primeiros dígitos do município</td> </tr> <tr> <td>cod_ufmun</td> <td>gr</td> <td>Grande Região: primeiro dígito da UF</td> </tr> <tr> <td>val_idade</td> <td>idade_tab, idade_tab2</td> <td>Categorizações da idade do trabalhador</td> </tr> <tr> <td>cod_grau_instrucao</td> <td>instrucao_tab, instrucao_tab2</td> <td>Grupos de instrução</td> </tr> <tr> <td>cod_grau_instrucao</td> <td>qualificado</td> <td>Foram definidos como migrantes qualificados aqueles com ao menos ensino superior completo</td> </tr> <tr> <td>cod_grau_instrucao, cod_cbo</td> <td>altqualificado</td> <td>Foram definidos como migrantes altamente qualificados aqueles que, além de qualificados, estavam ocupados nos grupos ocupacionais 0, 1 ou 2</td> </tr> <tr> <td>val_salario_12</td> <td>salario_real_dez</td> <td>Salário em dezembro, deflacionado pelo INPC e com remoção dos outliers</td> </tr> <tr> <td>cod_cbo</td> <td>grupos_ocup</td> <td>Grupo ocupacional</td> </tr> <tr> <td>cod_cbo</td> <td>subgrupos_ocup</td> <td>Subgrupo ocupacional</td> </tr> <tr> <td>cod_cbo</td> <td>subgrupos_ocup3</td> <td>Subgrupo ocupacional 3 dígitos</td> </tr> <tr> <td>subclasse</td> <td>divisao_ativ</td> <td>Divisão de atividade</td> </tr> </tbody> </table>		Nome Variável Original	Nome Variável Gerada	Descrição Variável Gerada	cod_nacionalidade	pais_ibge	País de nacionalidade do trabalhador, em formato padronizado. Informação ampliada a partir do pareamento com CAGED e CTPS	cod_nacionalidade	continente	País de nacionalidade do trabalhador, em formato padronizado. Informação ampliada a partir do pareamento com CAGED e CTPS	cod_ufmun	uf	Unidade da Federação: 2 primeiros dígitos do município	cod_ufmun	gr	Grande Região: primeiro dígito da UF	val_idade	idade_tab, idade_tab2	Categorizações da idade do trabalhador	cod_grau_instrucao	instrucao_tab, instrucao_tab2	Grupos de instrução	cod_grau_instrucao	qualificado	Foram definidos como migrantes qualificados aqueles com ao menos ensino superior completo	cod_grau_instrucao, cod_cbo	altqualificado	Foram definidos como migrantes altamente qualificados aqueles que, além de qualificados, estavam ocupados nos grupos ocupacionais 0, 1 ou 2	val_salario_12	salario_real_dez	Salário em dezembro, deflacionado pelo INPC e com remoção dos outliers	cod_cbo	grupos_ocup	Grupo ocupacional	cod_cbo	subgrupos_ocup	Subgrupo ocupacional	cod_cbo	subgrupos_ocup3	Subgrupo ocupacional 3 dígitos	subclasse	divisao_ativ	Divisão de atividade
Nome Variável Original	Nome Variável Gerada	Descrição Variável Gerada																																										
cod_nacionalidade	pais_ibge	País de nacionalidade do trabalhador, em formato padronizado. Informação ampliada a partir do pareamento com CAGED e CTPS																																										
cod_nacionalidade	continente	País de nacionalidade do trabalhador, em formato padronizado. Informação ampliada a partir do pareamento com CAGED e CTPS																																										
cod_ufmun	uf	Unidade da Federação: 2 primeiros dígitos do município																																										
cod_ufmun	gr	Grande Região: primeiro dígito da UF																																										
val_idade	idade_tab, idade_tab2	Categorizações da idade do trabalhador																																										
cod_grau_instrucao	instrucao_tab, instrucao_tab2	Grupos de instrução																																										
cod_grau_instrucao	qualificado	Foram definidos como migrantes qualificados aqueles com ao menos ensino superior completo																																										
cod_grau_instrucao, cod_cbo	altqualificado	Foram definidos como migrantes altamente qualificados aqueles que, além de qualificados, estavam ocupados nos grupos ocupacionais 0, 1 ou 2																																										
val_salario_12	salario_real_dez	Salário em dezembro, deflacionado pelo INPC e com remoção dos outliers																																										
cod_cbo	grupos_ocup	Grupo ocupacional																																										
cod_cbo	subgrupos_ocup	Subgrupo ocupacional																																										
cod_cbo	subgrupos_ocup3	Subgrupo ocupacional 3 dígitos																																										
subclasse	divisao_ativ	Divisão de atividade																																										



## Tratamento da Base

O processo de integração e tratamento conjunto das bases administrativas da Carteira de Trabalho e Previdência Social (CTPS), da Relação Anual de Informações Sociais (RAIS) e do Cadastro Geral de Admitidos e Demitidos (CAGED), com o objetivo de ampliar a identificação e a caracterização dos trabalhadores migrantes no mercado formal de trabalho brasileiro, gera, como resultado, o que se denominou como bases harmonizadas. A utilização integrada dessas bases permite explorar informações sobre estoque de vínculos, movimentações de admissões e desligamentos e características sociodemográficas e ocupacionais dos trabalhadores.

A base RAIS permite obter informações sobre os vínculos formais de trabalhos existentes ao longo do ano, e, de maneira mais específica para o uso que se tem no OBMigra, do estoque de vínculos ativos em 31 de dezembro de cada ano. Como ela capta todas as passagens pelo mercado de trabalho, um mesmo imigrante poderá aparecer na base mais de uma vez. Assim, o procedimento de limpeza da RAIS considera esta característica e define critérios para estabelecer o principal vínculo, em caso de mais de um registro para o mesmo trabalhador: 1) Apenas vínculos ativos em 31/12; 2) maior número de horas semanais; 3) maior salário em dezembro; 4) maior tempo de emprego; 5) maior salário contratual; 6) maior salário médio no ano; 7) se as variáveis eleitas para o filtro permanecerem iguais, o algoritmo escolhe a primeira linha empatada.

A base da RAIS possui a informação da nacionalidade de forma agregada, identificando o país em aproximadamente 40 casos, e com seis categorias do tipo "outros", em alguns casos permitindo a identificação do continente e em outros não. A partir do pareamento com a CTPS e, mais recentemente, com o CAGED, o dado final de nacionalidade pode ser definido com maior riqueza de detalhes, priorizando a informação originária da CTPS ou do CAGED.

## Limites e Potencialidades

### Potencialidades

- Cobertura muito abrangente do mercado de trabalho formal brasileiro, abrangendo trabalhadores com diversos tipos de vínculos e permitindo análises estruturais amplas;
- Desagregação geográfica e por classificações de ocupação ou atividade permitem um detalhamento da informação;
- A partir da implantação do eSocial, a RAIS e o CAGED passaram a compartilhar uma origem administrativa comum, baseada em imputação de dados de diferentes fontes, o que impactou na natureza das próprias bases de dados, derivados de uma mesma infraestrutura informacional.

### Limitações

- Cobertura restrita a empregos formais;
- A implementação do eSocial trouxe um ganho metodológico relevante, mas ao mesmo tempo traz a possibilidade de descontinuidade na série histórica;
- Defasagem temporal na disponibilização dos dados.