

vivo

Relatório de Análise de Fluxo Turístico

Erechim

Período 2022



GOVERNO DO ESTADO
RIO GRANDE DO SUL
SECRETARIA DE TURISMO

Índice

| | |
|---|-----------|
| 1. Contexto e motivação do estudo | 6 |
| 1.1 Introdução a plataforma LUCA..... | 6 |
| 2. Origem e precisão geográfica dos dados | 7 |
| 2.1 A rede celular | 7 |
| 2.2 Estação rádio base (ERB)..... | 7 |
| 2.3 Os eventos de rede | 8 |
| 2.4 Precisão geográfica dos eventos..... | 8 |
| 3. Amostra e limpeza de dados | 11 |
| 3.1 Seleção de amostra confiável | 11 |
| 3.2 Limpeza de dados..... | 11 |
| 3.3 Tratamento dos dados | 12 |
| 3.4 Estadias | 12 |
| 3.5 Viagens..... | 12 |
| 3.6 Pontos intermediários..... | 12 |
| 4. Representatividade estatística da amostra | 14 |
| 4.2 Construção da extrapolação: | 18 |
| 5 Framework de trabalho..... | 19 |
| 5.1 Escopo de trabalho | 20 |
| 6. Abrangência | 20 |
| 7. Resultados do ano de 2022..... | 20 |
| 7.1 Estados Emissores | 20 |
| 7.2 Cidades emissoras do Rio Grande do Sul..... | 21 |
| 7.3 Cidades emissoras de outros Estados | 22 |
| 7.4 Tempo médio de permanência | 24 |
| 7.5 Volume do fluxo turístico por dados sociodemográficos | 25 |
| 7.6 Área de captação | 28 |

7.7 Restrição de uso 28

ÍNDICE DE TABELAS E ILUSTRAÇÕES

| | |
|--|----|
| Tabela 1: Representatividade da amostra Telefônica - set/22 e população ago/2021 | 16 |
| Tabela 2: Exemplificando o processo de expansão dos dados | 18 |
| Tabela 3: Exemplificando o processo de expansão dos dados | 18 |
| Tabela 4: Exemplificando o processo de expansão dos dados | 19 |
| Tabela 5 Volume de Turistas por UF | 21 |
| Tabela 6 - TOP 10 Volume de Turistas Cidades emissoras..... | 22 |
| Tabela 7 – Lista Top 10 Cidades Emissoras fora RS – 2022 | 23 |
| Tabela 8 Área de captação dos turistas em 2022. | 28 |
| | |
| Figura 1: Diferentes formatos da cobertura de rede | 8 |
| Figura 2: Parâmetros de uma célula de telefonia celular. | 9 |
| Figura 3: Histograma da distância (em metros) da torre ao evento GPS..... | 10 |
| Figura 4: Histograma da distância (em metros) da torre ao evento GPS..... | 11 |
| Figura 5: Ilustração gráfica do processo de tratamento de dados | 14 |
| Figura 6: Conceito de estadia, viagem e ponto intermediário a nível temporal | 14 |
| Figura 7 -Fluxo de trabalho | 19 |
| Figura 8 - TOP10 Cidades Emissoras RS..... | 22 |
| Figura 9 TOP 10 Cidades Emissoras Fora do RS..... | 23 |
| Figura 10 - Tempo médio de permanência – Brasil, Rio Grande do Sul (RS) e Fora do RS | 24 |
| Figura 11- Distribuição por Gênero..... | 25 |
| Figura 12 Distribuição por Classe Social..... | 26 |
| Figura 13 Distribuição por Faixa de Idade..... | 27 |

1. Contexto e motivação do estudo

Em um país de proporções continentais, com importantes polos de produção distribuídos por todo território, o sistema de transportes é fundamental para fazer com que toda essa riqueza chegue de forma homogênea as diversas regiões do país. A boa integração da rede de transportes promove o desenvolvimento econômico, o fortalecimento do turismo, a melhora do ambiente de negócios e uma maior integração social das regiões do país.

As estatísticas de mobilidade populacional são o que permitem ao poder público melhorar a assertividade no planejamento e execução das políticas públicas que atendem o setor de transportes. Tradicionalmente a obtenção destes dados requer a realização de pesquisas de campo, que demandam altos recursos financeiros e de tempo para a sua execução, limitando tanto a frequência quanto a amplitude geográfica que elas podem abarcar.

Neste contexto apresenta-se a possibilidade de se usar fontes de dados alternativas as coletadas via pesquisa presencial. A utilização de recursos tecnológicos de telefonia móvel para obtenção de dados mobilidade vem sendo alvo de numerosos estudos ao redor do mundo (FORTUNATI, 2001; ASAKURA, 2003; AHAS, 2007 e 2008; GIRARDIN, 2009; NUNO, 2010; KHAN, 2012), sendo que esta tecnologia permite que sejam coletados dados de deslocamento e fluxo populacional baseados no uso que os usuários fazem da rede de telefonia móvel.

Visando atender esta demanda de dados de mobilidade, serão utilizadas neste estudo informações coletadas da rede de telefonia móvel para se desenvolver uma matriz origem-destino de passageiros em âmbito nacional.

1.1 Introdução a plataforma LUCA

LUCA Tourism é uma ferramenta estatística, suportada por uma infraestrutura de Big Data, que utiliza as informações de operação das redes de telefonia móvel da Telefônica, anônimas e agrupadas, para extrapolar o fluxo de pessoas por todo o território nacional. Para fornecer um serviço eficiente, a rede e os equipamentos telefônicos devem estar em comunicação frequente. A compreensão íntima desses eventos de rede permite que se construa uma compreensão contextual do movimento dos celulares no espaço e no tempo, permitindo a criação de registros que se interpretam como estadias e viagens.

O uso dos celulares pela rede de antenas é processada através de uma tripla camada, que garante a proteção absoluta à privacidade dos usuários e a relevância estatística da informação apresentada.

1. Anonimização: quaisquer registros pessoais são substituídos por identificadores encriptados na análise da movimentação de cada dispositivo.

2. **Agregação:** os dados sobre múltiplos dispositivos são agregados no espaço e no tempo, de forma que se obtenha relevância estatística em quantidade de informação, e de modo que não seja possível rastrear comportamentos individuais, ainda que anônimos.
3. **Extrapolação:** os dados são extrapolados para toda a base censitária brasileira, de forma que a base de clientes seja apenas uma inferência do comportamento de toda a população.

2. Origem e precisão geográfica dos dados

2.1 A rede celular

A rede celular é uma rede sem fio distribuída em áreas terrestres chamadas células, cada uma servida por pelo menos um transceptor de localização fixa, conhecido como um site celular ou estação rádio base (ERB). Em uma rede celular, cada célula usa um conjunto diferente de frequências de células vizinhas, para evitar interferências e fornecer largura de banda garantida dentro de cada célula. Quando juntas, essas células fornecem cobertura de rádio em uma ampla área geográfica. Isso permite que uma grande quantidade de telefones celulares se comunique entre si e com transceptores fixos e telefones em qualquer lugar da rede, através de estações base, mesmo que alguns telefones celulares estejam passando por mais de uma célula durante a transmissão.

2.2 Estação rádio base (ERB)

Sempre que um telefone móvel se conecta à rede, o sinal é encaminhado através de uma ERB. Isso se refere a um site de telefone celular onde antenas e equipamentos de comunicação eletrônica são colocados, normalmente em uma torre, mastro ou outro lugar alto, para criar uma célula dentro de uma rede celular. Uma célula é específica da tecnologia - ou seja, existem células separadas para 2G, 3G e 4G (eventos variando por geografia de implantação). O alcance de uma célula depende do tipo de tecnologia, pois as células 2G tendem a ser capazes de cobrir áreas maiores que as células 3G e 4G.

No contexto de como a Telefônica usa os dados, o termo célula não se refere ao local onde o equipamento é colocado (site da célula), mas à área coberta por este equipamento (ou seja, a área a partir da qual o sinal de um celular provavelmente será encaminhado para o equipamento neste site celular). As células podem ter formas diferentes direcionalidades e tamanhos, como mostrado na figura 1.

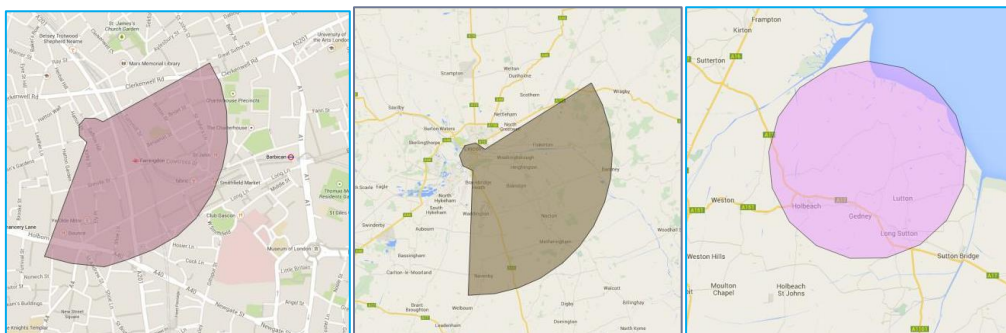


Figura 1: Diferentes formatos da cobertura de rede

2.3 Os eventos de rede

Os telefones celulares conectados à rede móvel geram eventos de rede. A natureza desses eventos (tipo, frequência no tempo) depende de uma série de fatores, incluindo o comportamento do usuário (fazer uma chamada, receber um SMS etc.), e das características da rede móvel. Para este estudo consideram-se os seguintes registros como evento de rede:

- Fazer ou receber uma ligação
- Enviar ou receber pacotes de dados

2.4 Precisão geográfica dos eventos

As células têm as seguintes características, todas disponíveis no catálogo de células, que é uma fonte confiável e robusta dentro das entidades de rede, visto que são dados importantes para a manutenção da rede.

- Tecnologia: é a geração de tecnologia da comunicação móvel na qual a célula opera. Pode ser 2G, 3G ou 4G.
- Tipo: está relacionado a amplitude da célula. Pode ser Macro, Micro, Pico ou Femto.
- Azimute: é a orientação da antena. 0 graus é orientação norte, 90 graus é orientação leste etc.
- Largura do feixe: é o ângulo de amplitude do feixe da célula. As mais comuns são as células de 65 graus de amplitude de feixe.
- Altura: é a altura em que a antena está localizada em relação ao solo.
- Inclinação: O ângulo de inclinação da antena em direção ao solo.
- Banda: é a frequência básica na qual a célula opera. Pode ser de 700 MHz, 900 MHz etc.

Metodologia

- Latitude: é a latitude exata do mastro da antena.
- Longitude: é a longitude exata do mastro da antena.
- Comprimento: é o comprimento exato do mastro da antena.
- Potência máxima: é a potência nominal máxima da antena.
- Alcance: É a distância máxima em que a célula pode oferecer na direção de seu azimute.

Essas características são representadas graficamente na Figura 2, abaixo:

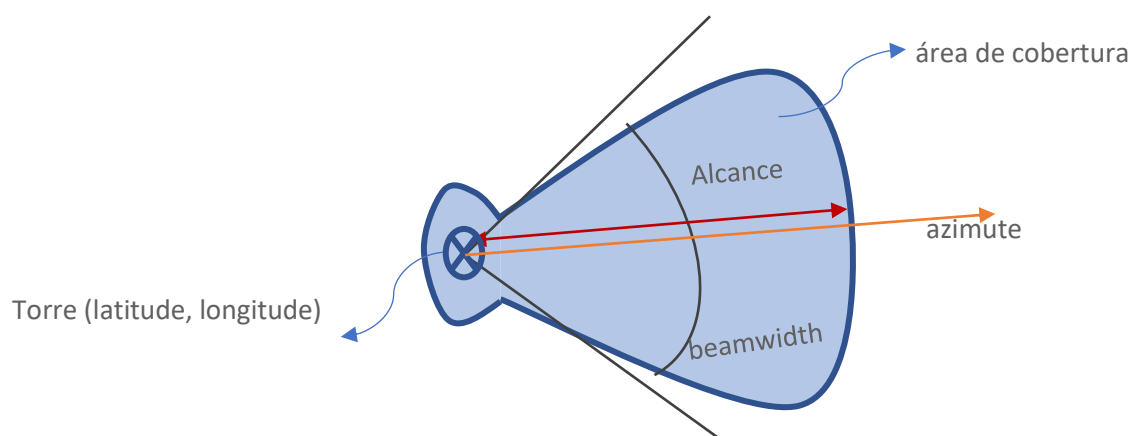


Figura 2: Parâmetros de uma célula de telefonia celular.

Com todos esses parâmetros físicos, consegue-se obter de forma muito aproximada a área teórica de cobertura das células de telefonia móvel, de modo que os eventos sejam localizados na área de cobertura da célula em que foram produzidos.

As células podem variar de alcance, cerca de 200m no caso da tecnologia 4G (mais comum em áreas urbanas), ou até mesmo vários quilômetros (macro células de baixa frequência podem chegar a ter um alcance de 12 Km). Por esta razão, a granularidade geográfica dos eventos de rede varia dependendo do intervalo de células em que ocorre e, conseqüentemente, se ocorre em ambientes urbanos ou rurais.

Por outro lado, é conhecido que a cobertura real das células é afetada diariamente por agentes externos às características físicas da antena, como a saturação de usuários na célula ou o ambiente celular. Portanto, a cobertura das células é ajustada, especificamente com um recálculo do intervalo através de um processo empírico baseado na observação dos eventos que realmente ocorreram. Para isso, é utilizado o fenômeno de *flickering*, que ocorre quando um usuário se conecta a duas ou mais células em um curto espaço de tempo em que o usuário

difícilmente consegue se mover. Cada vez que esse fenômeno é observado, pode-se afirmar que as coberturas das células envolvidas se sobrepõem. Este conceito é representado na Figura 3:

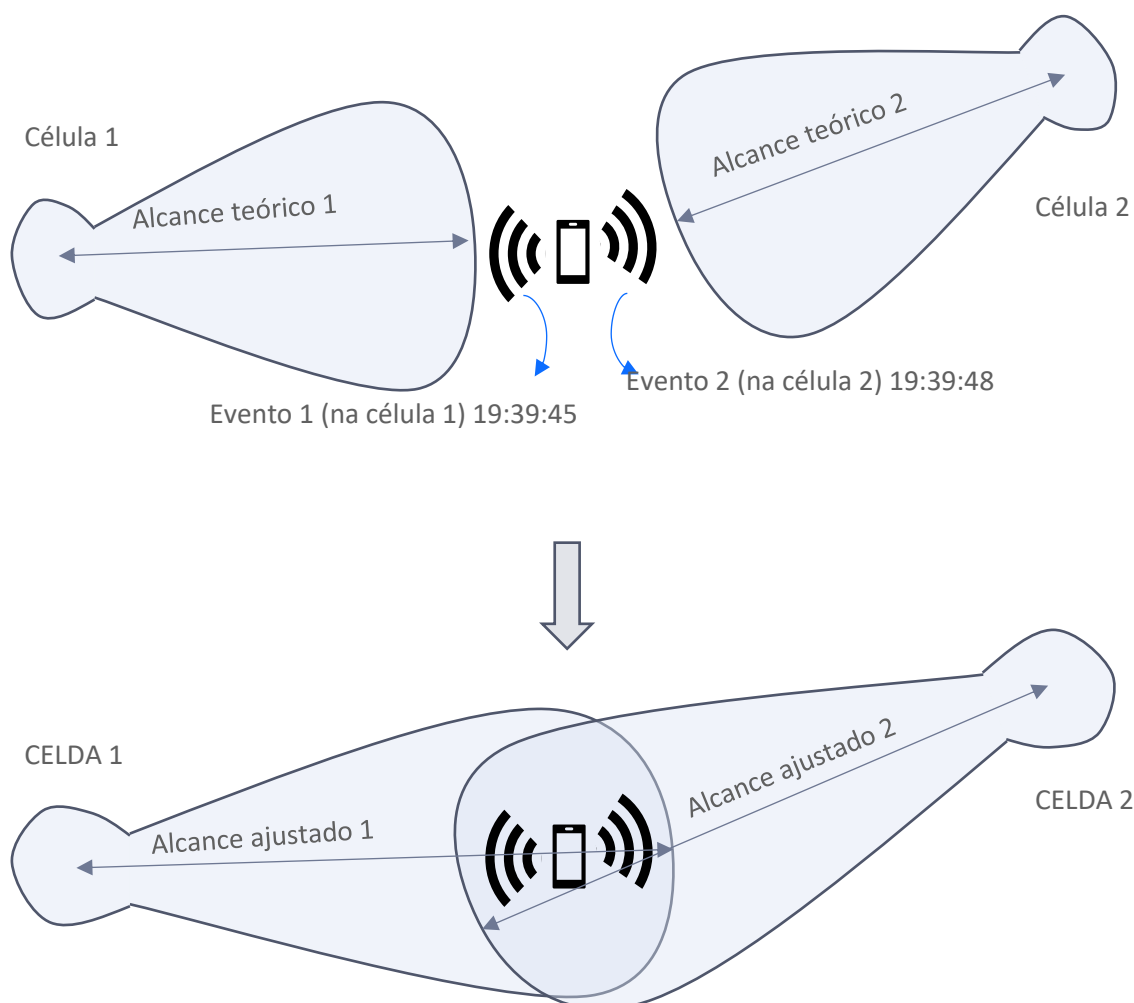


Figura 3: Histograma da distância (em metros) da torre ao evento GPS.

Portanto, levando em conta todos os eventos de cada dia e as relações entre todas as células, o alcance de todas as células é ajustado. Após esse ajuste de intervalo existe uma grande segurança de que um usuário que produziu um evento esteja nesta área de cobertura ajustada, uma vez que um processo empírico foi introduzido.

Para medir a precisão geográfica real Telefónica comparou os eventos de rede com uma fonte de eventos GPS de um subconjunto de usuários da sua base de clientes. A distribuição da distância entre o mastro da célula do evento a localização GPS produzida pelo mesmo usuário ao mesmo tempo é mostrada na Figura 4:

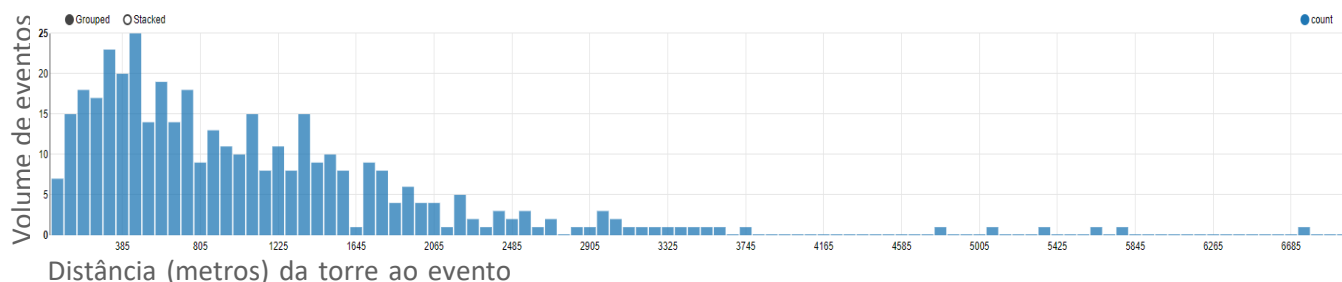


Figura 4: Histograma da distância (em metros) da torre ao evento GPS.

3. Amostra e limpeza de dados

3.1 Seleção de amostra confiável

Um dos primeiros passos na execução de cada projeto é a definição do que pode ser considerado como um usuário válido. Alguns usuários podem se tornar clientes da Telefônica durante o período de estudo ou podem deixar de ser clientes durante esse período. Alguns outros clientes fazem uso muito esporádico dos serviços, o que limita as conclusões que se pode obter do seu comportamento. Dessa forma, é importante que esses usuários sejam removidos do banco de dados do estudo. Por definição, um usuário inválido é um usuário que não gera eventos suficientes para permitir que se capture sua mobilidade de forma confiável. O mesmo usuário pode ser válido para um período de estudo, mas não é válido para outro. Portanto, a determinação de uma base de usuários válida é um processo específico do projeto.

3.2 Limpeza de dados

A limpeza dos dados ocorre no final de cada fase de processamento das informações, de modo que os erros possam ser detectados não apenas nos dados de entrada, mas também nos resultados dos algoritmos de extração dos insights. Dentro dessas "verificações de qualidade", o mais importante é, sem dúvida, a limpeza de dados de eventos de rede e das tabelas do catálogo de células. Seguem abaixo alguns exemplos de verificações realizadas:

- Se o formato da data e hora dos eventos não for o esperado, os dados desse dia serão descartados.
- Se houver muitos eventos sem célula associada, os dados desse dia serão descartados.
- Se o volume de eventos estiver abaixo de um limite determinado os dados desse dia serão descartados.
- Se o volume de células estiver abaixo de um limite determinado os dados desse dia serão descartados.

- Se os volumes variarem drasticamente em relação à média dos volumes dos dias anteriores os dados desse dia serão descartados.

3.3 Tratamento dos dados

Após a limpeza da base de dados os eventos de rede são tratados para que se transformem em informação estruturada útil para a análise de mobilidade.

3.4 Estadias

Uma estadia é criada quando um usuário fica em determinado local por um período igual ou superior a 30 minutos. Para determinar se o usuário permaneceu de fato neste local aplica-se um algoritmo que leva em consideração a localização da célula que o usuário gerou os eventos, seu alcance, a propensão a trocar entre células próximas e o tipo de evento gerado. O cálculo das estadias permite que se diferencie as origens e destinos das viagens dos pontos de passagem.

3.5 Viagens

Correspondem a deslocamentos entre duas estadias que aconteçam em locais distintos

3.6 Pontos intermediários

Correspondem a pontos intermediários da viagem detectados entre as origens e destinos. Abaixo ilustra-se graficamente este processo de tratamento de dados:

Tratamento de dados

Estadias

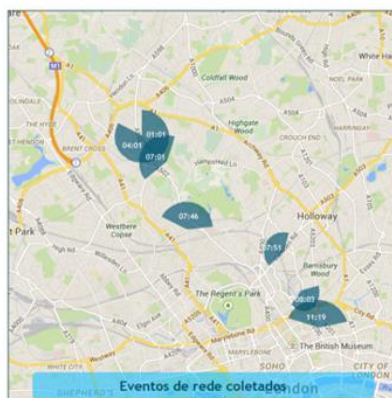
Local onde o usuário permanece um tempo sem movimentar-se

Viagens

Deslocamento entre duas estadias

Pontos intermediários

Eventos observados entre a origem e o destino da viagem



Tratamento de dados

Estadias

Local onde o usuário permanece um tempo sem movimentar-se

Viagens

Deslocamento entre duas estadias

Pontos intermediários

Eventos observados entre a origem e o destino da viagem



Tratamento de dados

Estadias

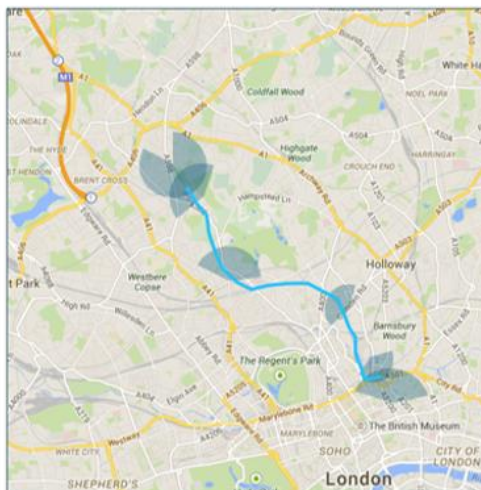
Local onde o usuário permanece um tempo sem movimentar-se

Viagens

Deslocamento entre duas estadias

Pontos intermediários

Eventos observados entre a origem e o destino da viagem



Tratamento de dados

Estadias

Local onde o usuário permanece um tempo sem movimentar-se

Viagens

Deslocamento entre duas estadias

Pontos intermediários

Eventos observados entre a origem e o destino da viagem

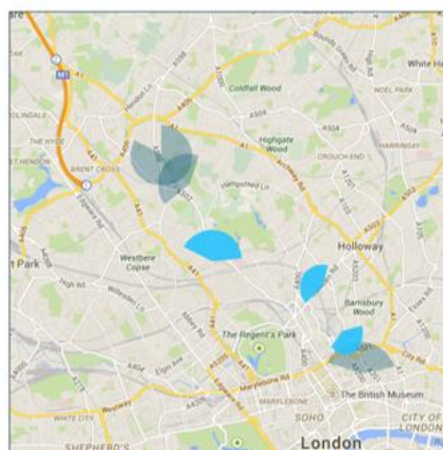


Figura 5: Ilustração gráfica do processo de tratamento de dados

Ilustra-se na figura 6 como um usuário ao longo do dia poderia ter seus eventos de rede transformados em estadias (Dwell), Viagens (Journeys) e pontos intermediários.

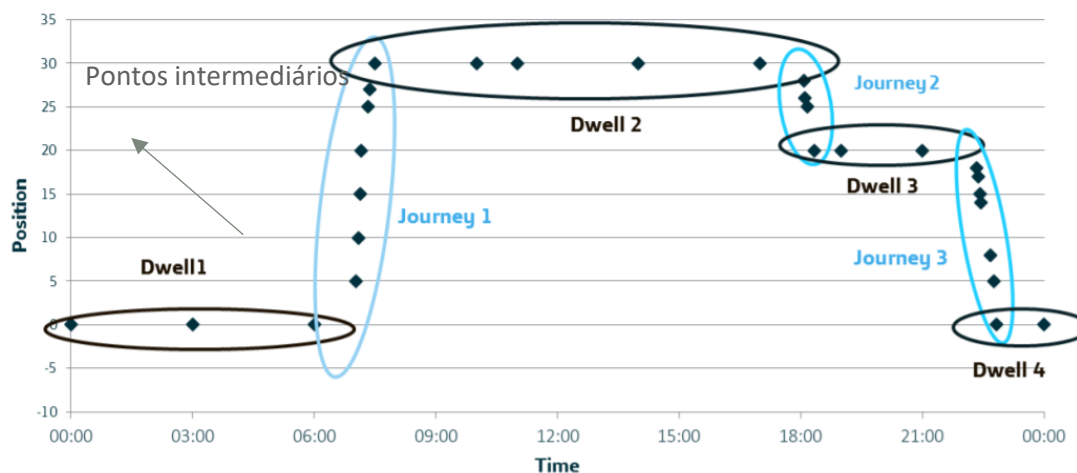


Figura 6: Conceito de estadia, viagem e ponto intermediário a nível temporal

4. Representatividade estatística da amostra

Uma das grandes vantagens de se trabalhar com dados da rede móvel é que a amostra da população é muito maior do que em qualquer pesquisa tradicional, o que aumenta muito a representatividade e diminui o "viés" das conclusões obtidas. Para calcular a representatividade de uma amostra, deve-se monitorar dois parâmetros: a margem de erro e o nível de confiança.

• **Margem de Erro:** é o intervalo em que se espera encontrar os dados a serem medidos no universo. Isto é, tomando uma proporção do universo, espera-se que as conclusões tiradas daquela amostra levada ao universo estejam em um intervalo centrado no valor obtido com uma amplitude de duas vezes a margem de erro (um uma vez para baixo e uma vez para cima). Visto com um exemplo: se a margem de erro oferecida pela amostra for de 5%, isso significa que se a conclusão for de que 70% da minha amostra mora em uma determinada cidade, essa conclusão tomada para todo o universo é traduzida em que entre 66,5% (70 menos 5%) e 73,5% (70 mais 5%) do universo vive nessa cidade.

• **Nível de Confiança:** É a certeza de que os dados que se pretende estimar em relação ao universo inteiro estão no intervalo estabelecido pela margem de erro explicada acima. No exemplo anterior, se a amostra permitir um nível de confiança de 95%, significa que os dados reais de todo o universo estarão no intervalo de 66,5% -73,5% com uma probabilidade de 95%.

A relação entre o tamanho do universo, o tamanho da amostra, a margem de erro e o nível de confiança é governada pela seguinte fórmula:

$$n = \frac{N \cdot Z^2 \cdot p \cdot (1-p)}{(N-1) \cdot e^2 + Z^2 \cdot p \cdot (1-p)}$$

Onde:

N:tamanho do universo.

n:tamanho da amostra.

e: margem máxima de erro permitida.

Z: desvio do valor médio para atingir o nível de confiança desejado (é dado pelas tabelas da distribuição normal, de modo que, se o nível de confiança desejado for, por exemplo, 95%, então $Z = 1,96$).

p: proporção esperada para ser encontrada. Se não for conhecido, 50% é usado, o que representa o valor mais conservador.

Quando o tamanho do universo é muito grande (maior que 100.000 indivíduos, a fórmula pode ser simplificada da seguinte forma:

$$n = \frac{Z^2 \cdot p \cdot (1-p)}{e^2}$$

Metodologia

Por questões regulatórias e de respeito a privacidade dos usuários a Telefônica tem algumas restrições quanto aos dados que pode compartilhar:

- 1- Telefônica não pode compartilhar dados não extrapolados, o que significa que não se pode abrir o valor da amostra utilizada nos estudos. Ao fazê-lo se estaria expondo informações da base de clientes da empresa, o que não se permite.
- 2- Telefônica não pode compartilhar estatísticas ou o detalhamento do deslocamento que representem volumes inferiores a 10 casos. Ao fazê-lo se facilitaria a identificação individual de usuários o que estaria em desacordo com a Lei nº 13.709 de 14 de agosto de 2018 – Lei Geral de Proteção de Dados.

Entretanto, existem informações públicas a nível macro que permitem entender a magnitude desta amostra. Grande parte destas informações podem ser encontradas no site: <http://www.teleco.com.br>, que resume uma série de estatísticas a respeito do mercado brasileiro de telecomunicações. Na tabela a seguir oferece-se uma visão de representatividade estatística da amostra para cada unidade da federação:

Tabela 1: Representatividade da amostra Telefônica - set/22 e população ago/2021

| UF | População* | Quantidade de celulares ** | Marketing Share Vivo** | Nível de confiança | Margem de erro |
|----|------------|----------------------------|------------------------|--------------------|----------------|
| AC | 906.876 | 354.095 | 41,20% | 95% | 0,183% |
| AL | 3.365.351 | 1.176.082 | 38,80% | 95% | 0,111% |
| AM | 4.269.995 | 1.861.544 | 44,20% | 95% | 0,089% |
| AP | 877.613 | 337.321 | 40,60% | 95% | 0,193% |
| BA | 14.985.284 | 5.273.279 | 33,80% | 95% | 0,065% |
| CE | 9.240.540 | 3.787.311 | 41,20% | 95% | 0,018% |
| DF | 3.094.325 | 1.170.533 | 29,90% | 95% | 0,111% |
| ES | 4.108.508 | 3.237.663 | 70,90% | 95% | 0,054% |
| GO | 7.206.589 | 2.528.273 | 30,00% | 95% | 0,076% |
| MA | 7.153.262 | 2.165.718 | 35,70% | 95% | 0,119% |
| MG | 21.411.923 | 11.977.008 | 48,70% | 95% | 0,041% |
| MS | 2.839.188 | 1.493.324 | 48,40% | 95% | 0,098% |
| MT | 3.567.234 | 2.232.040 | 52,70% | 95% | 0,082% |
| PA | 8.777.124 | 3.202.516 | 40,00% | 95% | 0,071% |
| PB | 4.059.905 | 1.763.857 | 40,20% | 95% | 0,171% |
| PE | 9.674.673 | 3.668.292 | 37,10% | 95% | 0,109% |
| PI | 3.289.290 | 928.521 | 29,70% | 95% | 0,121% |
| PR | 11.597.484 | 3.161.644 | 22,90% | 95% | 0,079% |
| RJ | 17.463.349 | 6.696.597 | 32,20% | 95% | 0,053% |
| RN | 3.560.903 | 1.241.732 | 36,80% | 95% | 0,121% |

| | | | | | |
|----|------------|------------|--------|-----|--------|
| RO | 1.815.278 | 385.880 | 21,70% | 95% | 0,223% |
| RR | 652.713 | 358.793 | 55,10% | 95% | 0,123% |
| RS | 11.466.630 | 6.439.091 | 46,90% | 95% | 0,052% |
| SC | 7.338.473 | 2.572.555 | 29,80% | 95% | 0,092% |
| SE | 2.338.474 | 1.496.145 | 62,80% | 95% | 0,109% |
| SP | 46.649.132 | 27.572.214 | 35,20% | 95% | 0,021% |
| TO | 1.607.363 | 245.479 | 15,00% | 95% | 0,144% |

**Fonte: http://www.teleco.com.br/cel_adl.asp * <https://www.ibge.gov.br/estatisticas-novoportal/sociais/populacao/9103-estimativas-de-populacao.html?=&t=downloads>

Expansão dos dados

Todos os resultados deste estudo são expandidos, o que significa que representam o comportamento de toda população e não somente dos usuários Telefônica. Descreve-se a seguir as etapas executadas durante este processo de expansão.

4.1 Construção da população residente por zona (Telefônica):

Para que se execute o processo de expansão é necessário assignar um local de residência aos usuários Telefônica. Tal definição é feita baseando-se exclusivamente no hábito de uso do celular dos usuários. É marcado como local de residência a zona de maior frequência onde o usuário foi encontrado no período entre 22:00 horas e 06:00 horas (durante todo o mês de análise).

Construção da população residente por zona:

O processo inicia-se com a alocação das informações censitárias nas zonas de estudo da matriz. Todos os setores censitários cujos centroides estejam contidos em uma zona de estudo tem sua população alocada como residente nesta zona. Assim constrói-se a população estimada de cada zona baseando-se no censo de 2010. Posteriormente um fator linear é aplicado sobre cada zona, de forma que a população de cada município atinja o valor projetado para o ano de estudo. Utiliza-se neste caso a projeção de população feita pelo IBGE para a data de primeiro de agosto de 2021.

<https://www.ibge.gov.br/estatisticas-novoportal/sociais/populacao/9103-estimativas-de-populacao.html?=&t=downloads>

Identificação de usuários por dia:

Para cada dia de análise, cria-se uma lista dos usuários que foram identificados naquele dia. Entenda-se por usuário identificado, o usuário, que neste dia tenha deixado um registro de permanência (tenha permanecido em um determinado local por um tempo igual ou superior a 30 min). Este registro de permanência pode ter ocorrido em qualquer região geográfica do país e independe deste usuário ter ou não realizado alguma viagem. Assim sendo, para cada dia de análise gera-se esta lista de usuários únicos identificados. Adiciona-se a esta tabela de dados as informações da zona de residência dos usuários, além da informação de gênero e idade. Estas 3 últimas informações são fixas por usuário. Nota-se que nesta etapa ainda se está trabalhando com dados individuais (cada usuário é identificado por um código encriptado). O resultado de tal tabela tem o seguinte formato:

Tabela 2: Exemplificando o processo de expansão dos dados

| Dia | Usuário | Zona Residência | Idade | Gênero |
|------------|-----------|-----------------|-------|--------|
| 01/01/2022 | AJFIEJDOE | 45 | 20-29 | m |
| 01/05/2022 | SIWWOSS | 23 | 30-39 | f |
| 02/01/2022 | IOWJSOSK | 67 | 20-29 | f |
| 03/01/2022 | SKSOSKW | 98 | 40-49 | f |
| 03/01/2022 | SRTUSKW | 25 | 20-29 | m |
| 04/01/2022 | IOWJSOSK | 12 | 40-49 | f |
| 04/01/2022 | AJFIEJDOE | 45 | 40-49 | m |

Tal tabela é em seguida agrupada por dia, zona de Residência, idade e gênero, enquanto realiza-se a contagem de usuários distintos. Ao final desta agregação pode-se compreender, por exemplo, quantos usuários homens, de 20 a 29 anos, residentes da zona 45 foram identificados no dia 03. A tabela neste momento tem o seguinte formato:

Tabela 3: Exemplificando o processo de expansão dos dados

| Dia | Quantidade | Zona Residência | Idade | Gênero |
|------------|------------|-----------------|-------|--------|
| 03/01/2022 | 240 | 45 | 20-29 | m |
| 03/01/2022 | 260 | 45 | 20-29 | f |
| 03/01/2022 | 400 | 45 | 30-39 | m |
| 03/01/2022 | 430 | 45 | 30-39 | f |

4.2 Construção da extrapolação:

A partir desta tabela acima é possível adicionar a informação populacional projetada (que pode nos dizer, por exemplo, quantos homens de 20 a 29 anos residem na zona 45). Com a utilização dos dados demográficos populacionais e das quantidades identificadas de usuários com esse

Metodologia

detalhamento pode-se criar os fatores de expansão através da fórmula: $\text{Fator} = \text{Censo}/\text{Quantidade}$.

Tabela 4: Exemplificando o processo de expansão dos dados

| Dia | Quantidade | Zona Residência | Idade | Gênero | Censo | Fator |
|------------|------------|-----------------|-------|--------|-------|-------|
| 03/01/2022 | 240 | 45 | 20-29 | m | 720 | 3 |
| 03/01/2022 | 260 | 45 | 20-29 | f | 832 | 3,2 |
| 03/01/2022 | 400 | 45 | 30-39 | m | 1800 | 4,5 |
| 03/01/2022 | 430 | 45 | 30-39 | f | 2107 | 4,9 |

Tendo estes pesos calculados, se construirá os estimadores necessários para obter as populações estimadas com base nas informações do usuário que possui dados demográficos cadastrados, sendo que o usuário precisa necessariamente ser detectado realizando uma viagem em um determinado dia. Assim, se aplica o fator calculado e se mensura a quantidade de usuários que realizariam essa viagem caso a base possuísse 100% de cobertura da população. O cálculo desse fator é construído com base na integração das variáveis como Zona de residência, idade e gênero. Para cada viagem realizada o peso é aplicado, se o usuário realiza 2 viagens em dia, aplica-se o peso para cada uma das viagens.

5 Framework de trabalho

Pode-se resumir o fluxo de trabalho explicitado até o momento de acordo com o seguinte fluxograma:

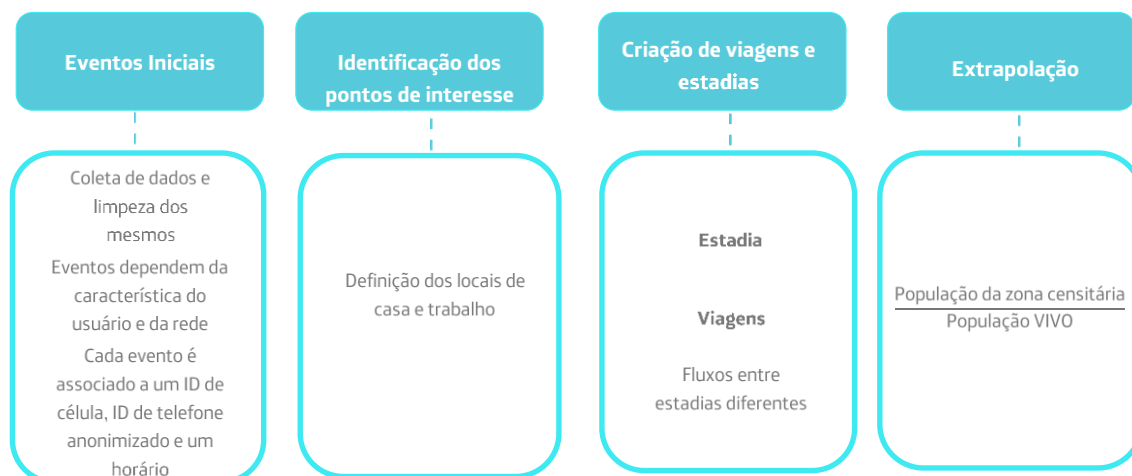


Figura 7 -Fluxo de trabalho

5.1 Escopo de trabalho

Este trabalho visa identificar o perfil do turista, em âmbito nacional, que meça o volume de turistas dos 5570 municípios do Brasil que agregados teremos os 4998 que possuem a cobertura da operadora VIVO, podendo esse número sofrer alterações devido ao número amostral de usuários nos municípios para determinar seu perfil, padrão de comportamento, permitindo-lhes realizar uma segmentação adequada e realizar campanhas dirigidas, no momento e local certos, e pode definir o preço otimize seu produto e melhore a tomada de decisões. Ser capaz de identificar o lugar certo para lançar um produto aumentando a proposição de valor em relação à concorrência e garantindo o ROI.

6. Abrangência

A base entregue nesse estudo, foi o ano completo de 2022, ou seja, buscamos realizar análises que correspondem esse processamento, estudos e investigação por meio da ferramenta **SmartSteps**.

Lembrando que Santana Livramento por não estar presente nos estudos de 2019, fica impossibilitado de fazer algum tipo de comparação com 2022.

| | |
|------------|-------------------|
| Municípios | Local |
| Erechim | Município Inteiro |

7. Resultados do ano de 2022

Durante o ano de 2022 verificou-se um volume 428.485 turistas únicos provenientes de todas as regiões do Brasil, incluindo também os moradores de outros municípios do Rio Grande do Sul. Optamos por incluir Caixas do Sul nas análises, pois representa a segunda maior população do estado.

7.1 Estados Emissores

Na Tabela 7 detalha-se o volume de turistas por Unidade da Federação (UF), onde é possível observar que os 5 principais estados emissores, não incluindo o RS são: SP, SC, PR, RJ e MG. Os turistas do RS que por razões de distância é o estado mais emissor representa 34% dos turistas,

somado a esses outros 5 estados representam aproximadamente 75% dos turistas durante o ano de 2022.

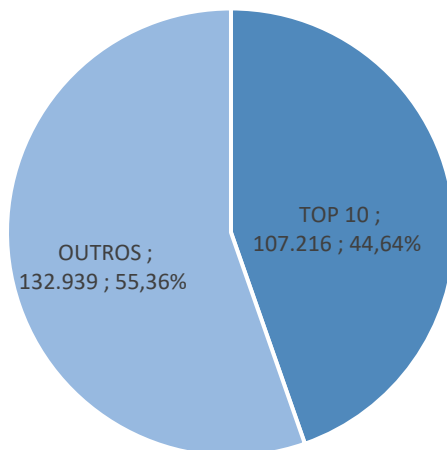
Tabela 5 Volume de Turistas por UF

| Estados Emissores | Qtd turistas únicos | % do total |
|---------------------|---------------------|---------------|
| RIO GRANDE DO SUL | 240.155 | 56,05% |
| SANTA CATARINA | 83.092 | 19,39% |
| PARANÁ | 68.519 | 15,99% |
| SÃO PAULO | 11.253 | 2,63% |
| MATO GROSSO DO SUL | 3.430 | 0,80% |
| GOIÁS | 3.194 | 0,75% |
| MATO GROSSO | 2.628 | 0,61% |
| MINAS GERAIS | 2.458 | 0,57% |
| RIO GRANDE DO NORTE | 1.985 | 0,46% |
| RIO DE JANEIRO | 1.954 | 0,46% |
| CEARÁ | 1.842 | 0,43% |
| ALAGOAS | 1.537 | 0,36% |
| PARAÍBA | 1.343 | 0,31% |
| DISTRITO FEDERAL | 1.116 | 0,26% |
| TOCANTINS | 852 | 0,20% |
| BAHIA | 754 | 0,18% |
| RONDÔNIA | 513 | 0,12% |
| MARANHÃO | 452 | 0,11% |
| PARÁ | 351 | 0,08% |
| PERNAMBUCO | 253 | 0,06% |
| ESPIRITO SANTO | 237 | 0,06% |
| AMAZONAS | 234 | 0,05% |
| SERGIPE | 199 | 0,05% |
| PIAUI | 59 | 0,01% |

7.2 Cidades emissoras do Rio Grande do Sul

Descreve-se na Figura 8 a representatividade do grupo com as 10 principais cidades emissoras de turistas para cada município da UF, bem como um somatório de todas as outras remanescentes. Observa-se que o grupo TOP 10 representa 46% do total de turistas emissores do estado do RS, conseqüentemente é nítido que existe uma concentração de turistas no TOP 10.

Figura 8 - TOP10 Cidades Emissoras RS



A Tabela 6 detalha as 10 principais cidades com seu respectivo volume de turistas, observa-se que o volume de Porto Alegre e Caxias do Sul representa quase 50% dos turistas que visitam Gramado.

Tabela 6 - TOP 10 Volume de Turistas Cidades emissoras

| Cidades Emissoras RS | Qtd turistas únicos | % do total |
|----------------------|---------------------|---------------|
| PASSO FUNDO | 33.351 | 13,89% |
| PORTO ALEGRE | 20.094 | 8,37% |
| CAXIAS DO SUL | 12.886 | 5,37% |
| MARAU | 7.398 | 3,08% |
| CANOAS | 7.369 | 3,07% |
| IJUI | 7.165 | 2,98% |
| CARAZINHO | 7.013 | 2,92% |
| BENTO GONCALVES | 4.395 | 1,83% |
| SAO LEOPOLDO | 3.914 | 1,63% |
| PELOTAS | 3.631 | 1,51% |
| OUTROS | 132.939 | 55,36% |

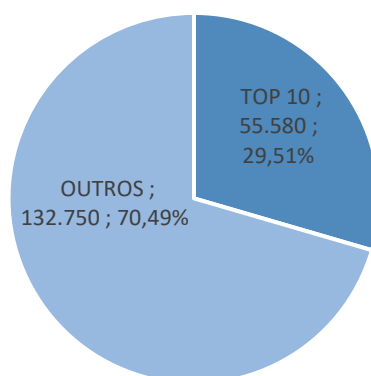
7.3 Cidades emissoras de outros Estados

Abaixo apresenta-se na Figura 9, a representatividade do grupo TOP 10 dos outros Estados. Verifica-se que o grupo TOP 10 representa pouco mais de 29% dos turistas provenientes dos

Metodologia

estados fora do Rio Grande do Sul, onde é possível observar uma fragmentação para outras cidades fora do TOP 10.

Figura 9 TOP 10 Cidades Emissoras Fora do RS



Percebe-se que Santa Catarina é o principal estado emissor com pouco mais de 83 mil turistas, e possui o maior número de turistas espalhados sendo 4 cidades do top cidades emissoras de fora do RS, juntamente com o estado do Paraná (68mil turistas) que possui 5 cidades emissoras dentro do TOP 10 de fora do RS, estados o perfil de visitantes de fora do Rio Grande do Sul, são as capitais das regiões sul, sudeste e centro-oeste, com exceção de Joinville que no ranking de cidades que mais visitaram Gramado em 2022 se encontra em 8º lugar.

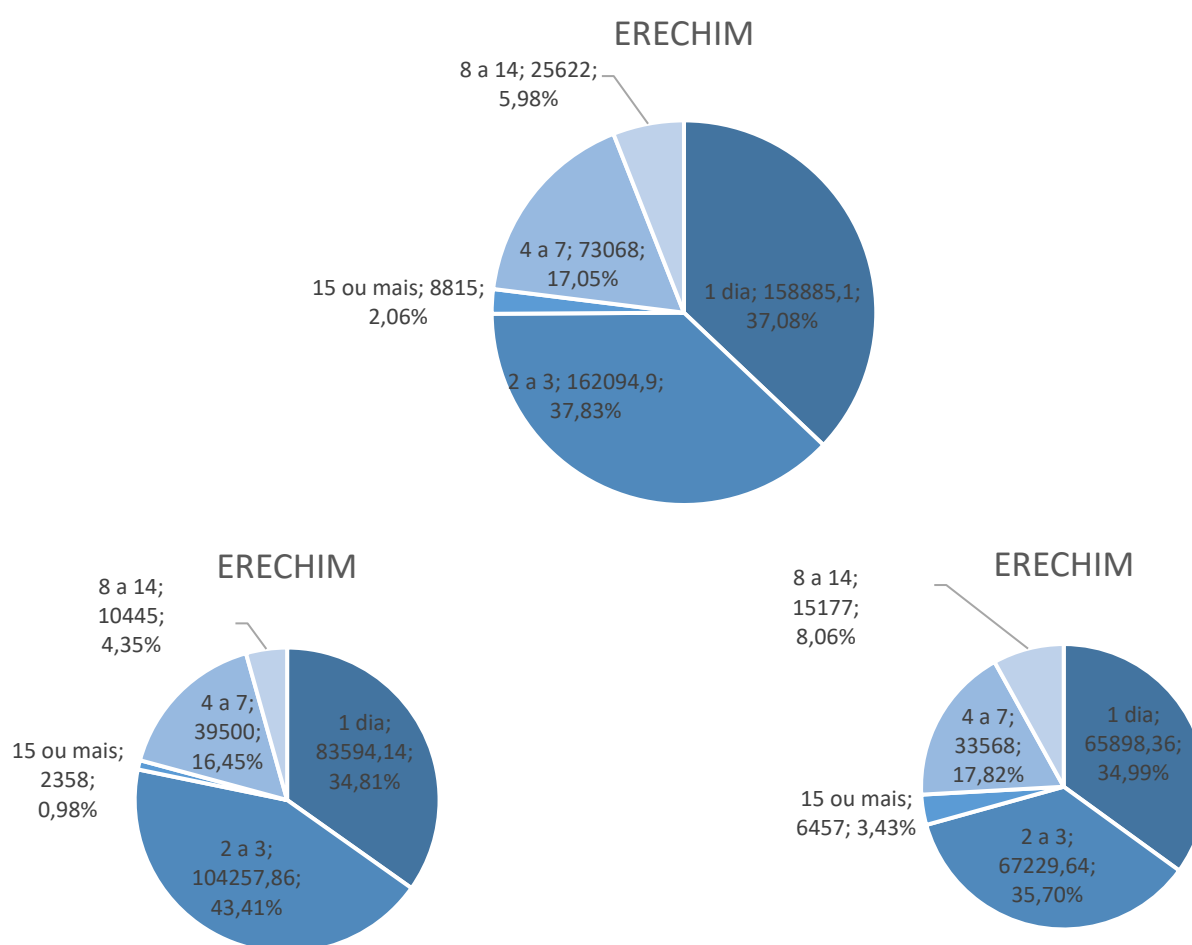
Tabela 7 – Lista Top 10 Cidades Emissoras fora RS – 2022

| Cidades Emissoras fora RS | Qtd turistas únicos | % do total |
|---------------------------|---------------------|---------------|
| FLORIANOPOLIS – SC | 11.292 | 6,00% |
| CURITIBA – PR | 7.704 | 4,09% |
| SAO MATEUS DO SUL – PR | 7.314 | 3,88% |
| PAULA FREITAS - PR | 5.129 | 2,72% |
| BALNEARIO CAMBORIU – SC | 4.786 | 2,54% |
| CAMPOS NOVOS - SC | 4.786 | 2,54% |
| CASCADEL – PR | 4.062 | 2,16% |
| ITAPEMA - SC | 3.664 | 1,95% |
| SAO PAULO – SP | 3.610 | 1,92% |
| UNIAO DA VITORIA - PR | 3.233 | 1,72% |
| OUTROS | 132.750 | 70,49% |

7.4 Tempo médio de permanência

No estudo de turismo existe uma importância enorme da variável de tempo de permanência, na figura 10 apresenta-se com detalhes o tempo médio de permanência no município, onde é possível notar que 79% dos turistas têm permanência média entre 1 e 3 dias. Observa-se também que aproximadamente 90% dos turistas permanecem por no máximo até 1 semana. É possível perceber que o comportamento dos turistas do Rio Grande do Sul frente aos outros Estados é semelhante.

Figura 10 - Tempo médio de permanência – Brasil, Rio Grande do Sul (RS) e Fora do RS



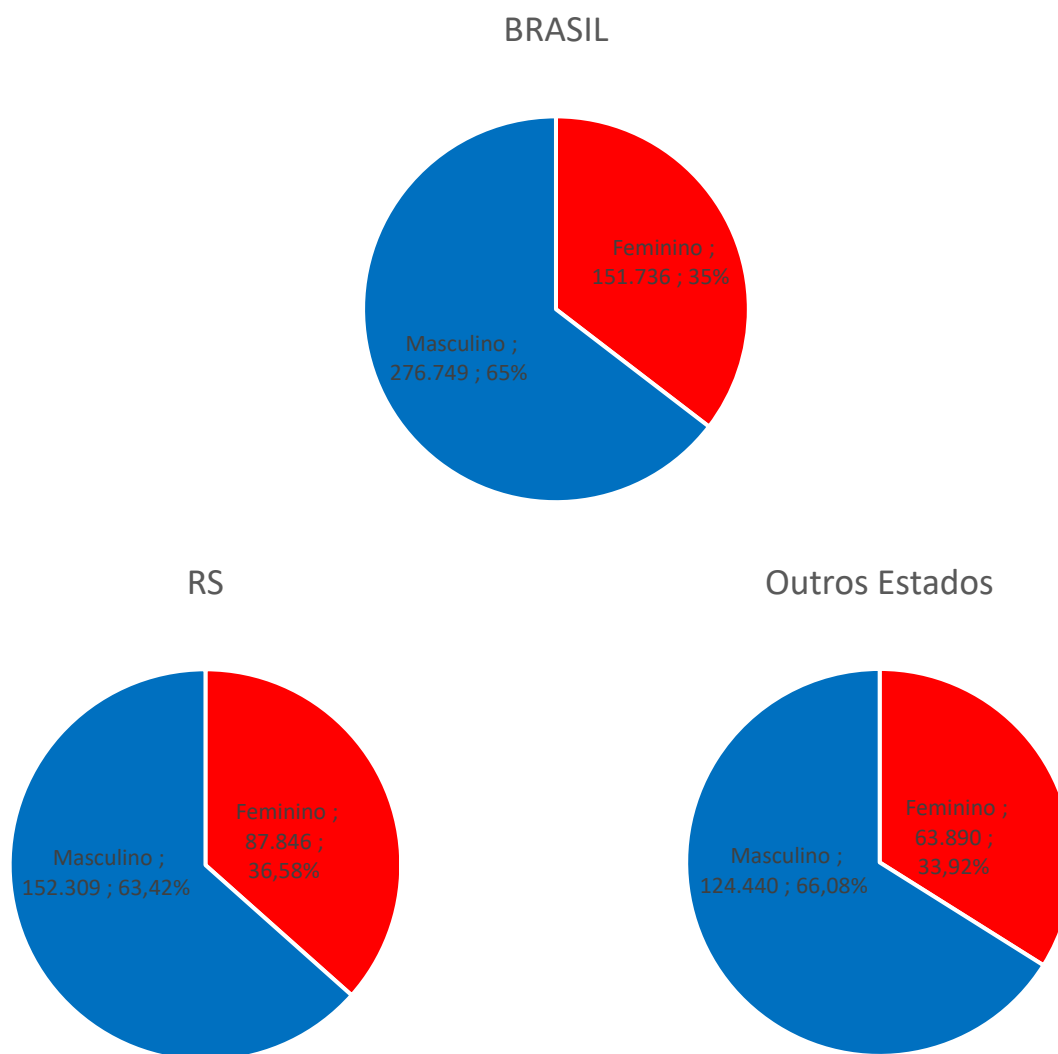
7.5 Volume do fluxo turístico por dados sociodemográficos

Nessa subseção o objetivo é ilustrar por meio de gráficos de pizza o comportamento dos turistas por dados sociodemográficos, no qual correspondem as variáveis de gênero, faixa de idade e classe social

7.5.1 Distribuição por Gênero

Observa-se na figura 11 que os turistas homens se sobressaem um pouco mais que as mulheres no turismo no Rio Grande do Sul. O comportamento se mantém o mesmo para turistas de fora do RS.

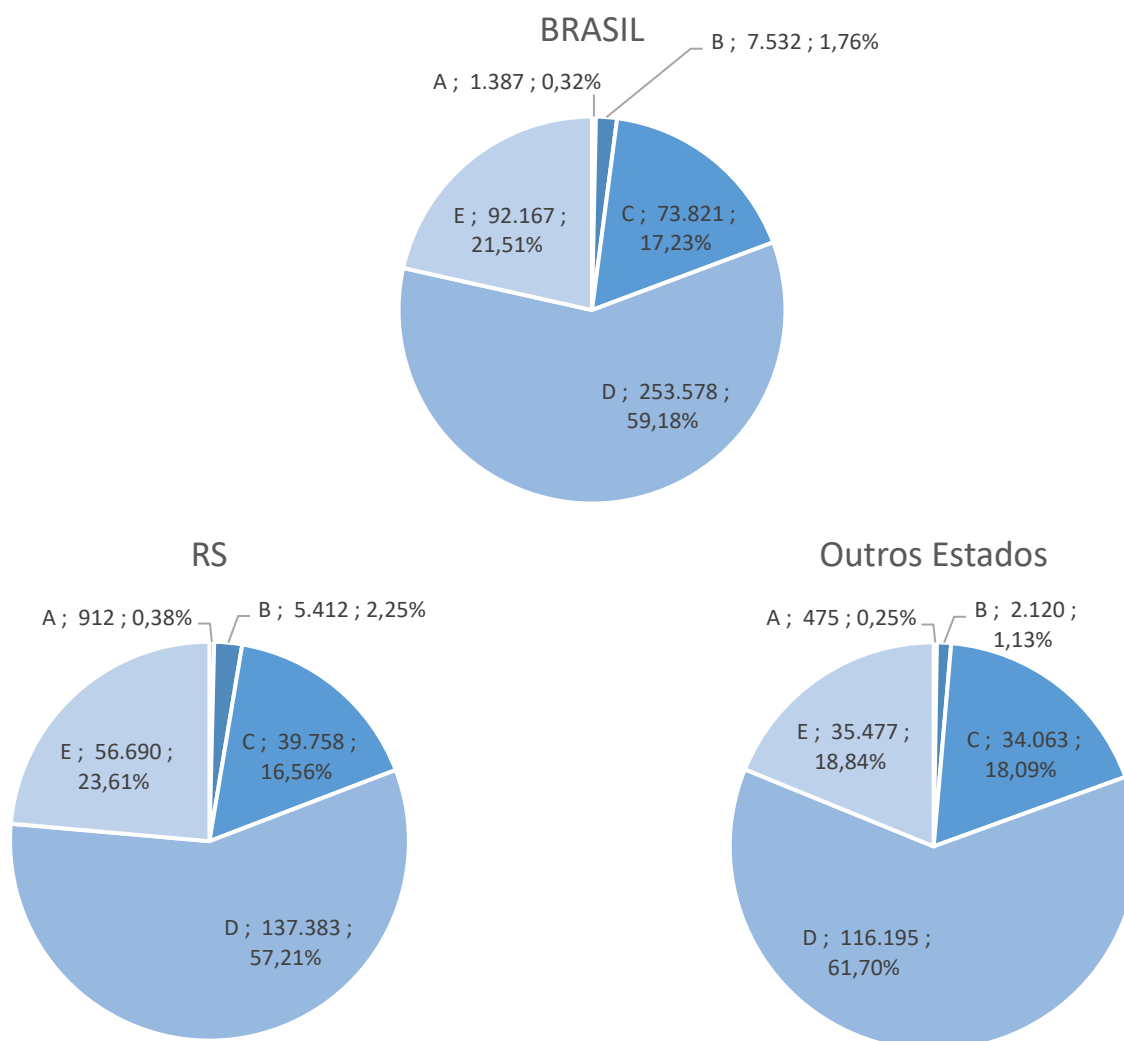
Figura 11- Distribuição por Gênero



7.5.2 Distribuição por Classe Social

Por meio da figura 12, é possível observar um domínio da visitação da Classe D, tanto em âmbito nacional, como turistas do Rio Grande do Sul e de fora dele. Portanto, é um comportamento que nos chama a atenção e bem interessante no período de recuperação pós pandemia, pois pelo confinamento obrigatório em anos anteriores, as pessoas assim que podem já iniciam um processo de viagem.

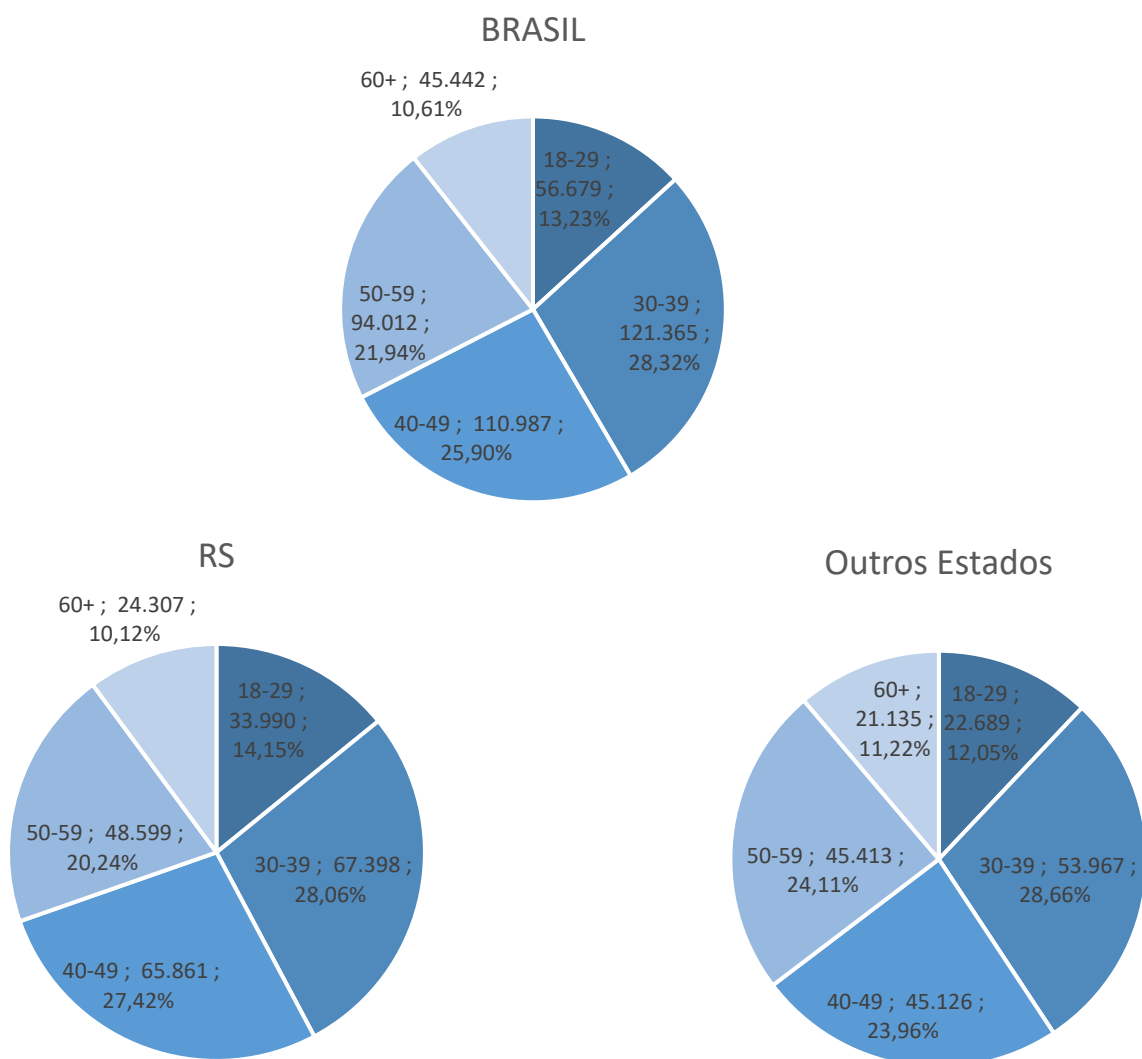
Figura 12 Distribuição por Classe Social



7.5.3 Distribuição por faixa de idade

Ao analisar a distribuição por faixa de idade na figura 13 observa-se uma predominância dos grupos de 30-39 e 40-49 anos, que é uma alta probabilidade de famílias que buscam viajar, devido a alguns motivos. Possivelmente são os mais estabelecidos familiarmente e busca viajar mais, bem como dependendo da faixa de renda, são as faixas etárias que podem gastar mais no turismo. E esse comportamento repete-se tanto para nível Brasil bem como nível Rio Grande do Sul e fora dele.

Figura 13 Distribuição por Faixa de Idade



7.6 Área de captação

Abaixo detalha-se o raio de captação dos turistas, ou seja, a quantidade de turistas que residem a uma distância dentro dos intervalos definidos. É importante ressaltar que no cálculo leva em consideração a distância ponto a ponto de acordo com os pontos visitados. Gramado é uma cidade em que é bem distribuído o volume de turistas brasileiros, podendo vir de qualquer parte do país, principalmente de capitais.

Tabela 8 Área de captação dos turistas em 2022.

| Raio captação Turistas | Qtd turistas únicos | % do total |
|------------------------|---------------------|---------------|
| 0 a 250km | 212.989 | 49,71% |
| 1000 a 2000km | 11.830 | 2,76% |
| 250 a 500km | 173.886 | 40,58% |
| 500 a 1000km | 20.424 | 4,77% |
| Mais de 2000km | 9.356 | 2,18% |

7.7 Restrição de uso

- Os bancos de dados fornecidos são considerados informação estratégica e confidencial da Telefônica, ficando proibida sua reprodução, distribuição e disponibilização pública ou qualquer utilização distinta da detalhada neste presente acordo por parte do Cliente.
- O Cliente se comprometa a adotar e documentar os mecanismos de segurança necessários para proteger a confidencialidade dos dados frente a qualquer perda, destruição, alteração, revelação ou acesso não autorizado. O nível de segurança que o Cliente deverá implementar deverá ser adequado ao risco que apresenta o acesso aos dados e a natureza deles.
- O Cliente se compromete a não duplicar, copiar ou revelar os Dados recebidos da Telefônica em todo ou em parte, a terceiros, salvo que se houver autorização prévia no presente acordo ou autorização por escrito por parte da Telefônica.

Metodologia

- O Cliente restringirá o uso da informação exclusivamente para os empregados que tenham necessidade de realização de análise e forma que se detalha no presente acordo.
- O Cliente se compromete a não realizar sobre os dados nenhum ato ilegal ou fraudulento.
- O Cliente se compromete a não tentar manipular, decompilar, aplicar engenharia inversa sobre os dados da Telefônica ou combinar os dados próprios ou de terceiros de uma forma que tal combinação possa supor a pré-identificação dos indivíduos ou de alguma maneira facilitá-la.
- Todos os direitos de propriedade intelectual sobre os dados seguirão sendo de titularidade exclusiva de Telefônica.



www.telefonica.com