

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA
Pós-graduação *Lato Sensu* em Ciência de Dados e Big Data

Debora Joana T. Corsino

**REAJUSTE DO SEGURO DE AUTOMÓVEIS BASEADO NA SINISTRALIDADE
POR GÊNERO**

Belo Horizonte
2020

Debora Joana T. Corsino

**REAJUSTE DO SEGURO DE AUTOMÓVEIS BASEADO NA SINISTRALIDADE
POR GÊNERO**

Trabalho de Conclusão de Curso apresentado
ao Curso de Especialização em Ciência de
Dados e Big Data como requisito parcial à
obtenção do título de especialista.

Belo Horizonte

2020

SUMÁRIO

Sumário

1. Introdução.....	4
1.1. Contextualização.....	4
1.2. O problema proposto.....	6
2. Coleta de Dados	7
3. Processamento/Tratamento de Dados	8
4. Análise e Exploração dos Dados	10
5. Criação de Modelos de Machine Learning	10
6. Apresentação dos Resultados	11
7. Links	17
REFERÊNCIAS.....	18

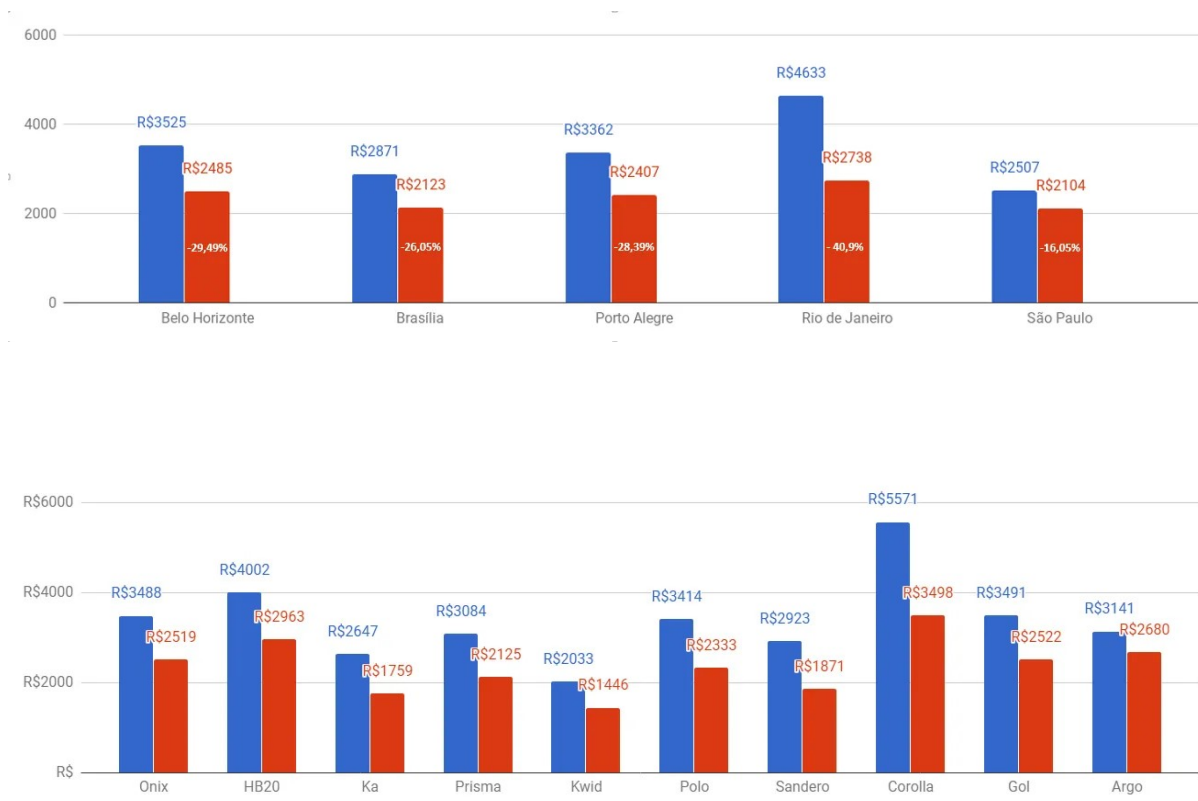
1. Introdução

1.1. Contextualização

As seguradoras constantemente ajustam o valor do seguro de automóveis baseadas em algumas premissas. Uma delas é a diferença do valor pelo gênero do segurado. O seguro de automóveis para mulheres em determinada idade, supostamente têm um custo menor do que para homens, na mesma condição.

A revista [Apólice](#) publicou o relatório de levantamento com os dez veículos mais vendidos (de acordo com a Federação Nacional da Distribuição de Veículos Automotores – Fenabrave – Maio de 2018), onde analisa o preço médio mensal do seguro em cinco capitais brasileiras: Belo Horizonte, Brasília, Porto Alegre, Rio de Janeiro e São Paulo, comparando o perfil de homens e mulheres, aos 35 anos, contratando o seguro pela primeira vez.

No comparativo entre todas as cidades e modelos do estudo, o perfil feminino apresenta a menor média de preço do seguro, representando uma economia de quase 30% para as mulheres:



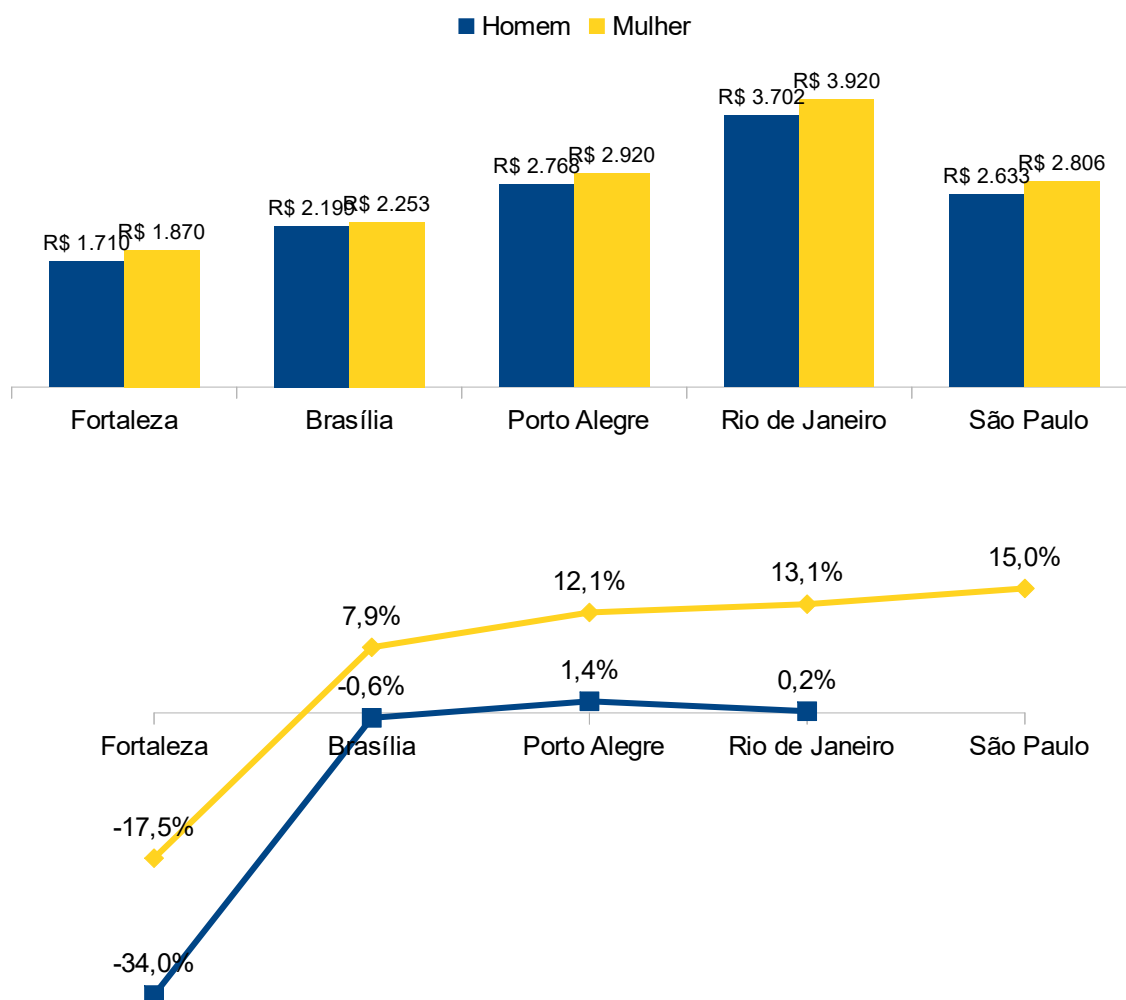
Uma das hipóteses que sustenta essa decisão é o índice de sinistralidade relacionada à acidentes provocados pelos dois conjuntos e sua severidade.

Segundo dados do Departamento Nacional de Trânsito (Denatran), dos mais de 60 milhões de motoristas no Brasil, quase 20 milhões são mulheres. Em relação a acidentes, 71% são provocados por homens e apenas 11% por mulheres.

Outro dado relevante: o Ministério da Saúde aponta que o número de homens que morreram no trânsito, em 2010, é quase quatro vezes maior do que o de mulheres. Considerando os que perderam a vida no trânsito naquele ano, 31.675 eram homens (78%) e 8.935 eram mulheres (22%).

No entanto, em 2019, o Estadão publicou uma matéria identificando que, em 04 das mesmas capitais acima relatadas, o seguro para as mulheres aumentou em relação aos homens.

A reportagem foi baseada no relatório da plataforma de seguros online Bidu – março 2019 – que identificou a quebra desta tendência.



Paulo Marchetti – CEO ComparaOnline do Brasil – deu a seguinte declaração:

“Apesar de ambos possuírem experiência, as mulheres costumam conduzir com mais segurança e prudência no trânsito e, na maioria dos casos, se envolvem colisões leves que não valem o custo da franquia para acionar o seguro para o conserto do carro, homens registram acidentes mais graves, que levam abertura de sinistros para perda parcial ou total do veículo nas seguradoras”

A redução na diferença se faz porque, cada vez, mais os sexos se equiparam em diversas questões – mulheres estão estudando mais a noite, saindo mais para baladas e comprando mais carros, e passam a correr riscos semelhantes aos homens.

Se relacionarmos os homens e mulheres que se encontram na faixa dos 35 anos, a diferença é um pouco maior, chegando a pouco mais de 10% de variação:

Para homens e mulheres solteiros, na faixa dos 60 anos de idade, os valores se tornam mais igualitários ainda, com uma variação média de valores de 4 a 5%. Mesmo assim, para ambos, o valor é mais baixo do que nas demais idades, já que as seguradoras confiam em um perfil mais experiente

No entanto, pelo cada vez maior senso de igualdade de gênero, e o comportamento cada vez mais semelhante entre homens e mulheres, as seguradoras estão nivelando os valores do seguro auto, acabando com a fama de seguro de carro mais barato para mulheres.

Através de dados oficiais, iremos identificar se esta decisão deverá ser continuada ou se houve alteração no comportamento dos dois conjuntos que altere ou invalide essa premissa, desde que a idade seja equivalente para ambos.

1.2. O problema proposto

A questão proposta é importante para que as seguradoras possam oferecer um reajuste financeiro justo, baseado em fatos.

Iremos analisar dados estatísticos de acidentes com condutores de ambos os gêneros para identificar se este é, realmente, um fator que deva ser considerado

relevante no reajuste do seguro de automóveis ou não, a exemplo da “*Proposition 103*” vigorada em Julho de 2019, no estado da Califórnia, que proibiu as seguradoras de basear seus critérios para valores de apólices no gênero dos segurados, encaixando o gênero junto com a etnia e sexualidade nos fatores que não podem ser levados em consideração na apólice de seguro.

Os dados foram coletados do Departamento de Polícia Rodoviária Federal (DPRF) e da Superintendência de Seguros Privados (SUSEP).

Trata-se de um levantamento nacional, apontando cidades/estados em que haja discrepância nos dados.

Serão analisados e comparados, dados registrados de 2018 a 2020

2. Coleta de Dados

Os dados foram obtidos das seguintes fontes:

Dados abertos da Policia Federal - <https://portal.pr.f.gov.br/dados-abertos-acidentes>
07/03/2020 - formato .csv;

ACIDENTES POR PESSOA		
Nome da coluna/campo	Descrição	Tipo
<i>id</i>	Variável com valores numéricos, representando o identificador do acidente.	INT
<i>pes_id</i>	Variável com valores numéricos, representando o identificador da pessoa envolvida.	INT
<i>data_inversa</i>	Data da ocorrência no formato dd/mm/aaaa.	DATE
<i>dia_semana</i>	Dia da semana da ocorrência. Ex.: Segunda, Terça, etc.	VARCHAR
<i>horario</i>	Horário da ocorrência no formato hh:mm:ss.	TIME
<i>uf</i>	Unidade da Federação. Ex.: MG, PE, DF, etc.	VARCHAR
<i>br</i>	Variável com valores numéricos, representando o identificador da BR do acidente.	VARCHAR
<i>km</i>	Identificação do quilômetro onde ocorreu o acidente, com valor mínimo de 0,1 km e com a casa decimal separada por ponto.	FLOAT
<i>municipio</i>	Nome do município de ocorrência do acidente	VARCHAR
<i>causa_acidente</i>	Identificação da causa principal do acidente. Neste conjunto de dados são excluídos os acidentes com avariável causada por...	VARCHAR
<i>tipo_acidente</i>	Identificação do tipo de acidente. Ex.: Colisão frontal, Saída de pista, etc. Neste conjunto de dados são excluídos os tipos...	VARCHAR
<i>classificacao_acidente</i>	Classificação quanto à gravidade do acidente: Sem Vítimas, Com Vítimas Feridas, Com Vítimas Fatais e Ignorado.	VARCHAR
<i>fase_dia</i>	Fase do dia no momento do acidente. Ex. Amanhecer, Pleno dia, etc.	VARCHAR
<i>sentido_via</i>	Sentido da via considerando o ponto de colisão: Crescente e decrescente.	VARCHAR
<i>condicao_meteorologica</i>	Condição meteorológica no momento do acidente: Céu claro, chuva, vento, etc.	VARCHAR
<i>tipo_pista</i>	Tipo da pista considerando a quantidade de faixas: Dupla, simples ou múltipla.	VARCHAR
<i>tracado_via</i>	Descrição do traçado da via.	VARCHAR
<i>uso_solo</i>	Descrição sobre as características do local do acidente: Urbano=Sim; Rural=Não.	VARCHAR
<i>id_veiculo</i>	Variável com valores numéricos, representando o identificador do veículo envolvido.	INT
<i>tipo_veiculo</i>	Tipo do veículo conforme Art. 96 do Código de Trânsito Brasileiro. Ex.: Automóvel, Caminhão, Motocicleta, etc.	VARCHAR
<i>marca</i>	Descrição da marca do veículo.	VARCHAR
<i>ano_fabricacao_veiculo</i>	Ano de fabricação do veículo, formato aaaa	INT
<i>tipo_envolvido</i>	Tipo de envolvido no acidente conforme sua participação no evento. Ex.: condutor, passageiro, pedestre, etc	VARCHAR
<i>estado_fisico</i>	Condição do envolvido conforme a gravidade das lesões. Ex.: morto, ferido leve, etc.	VARCHAR
<i>idade</i>	Idade do envolvido. O código “-1” indica que não foi possível coletar tal informação.	INT
<i>sexo</i>	Sexo do envolvido. O valor “inválido” indica que não foi possível coletar tal informação.	VARCHAR
<i>ilesos</i>	Valor binário que identifica se o envolvido foi classificado como ileso.	INT
<i>feridos_leves</i>	Valor binário que identifica se o envolvido foi classificado como ferido leve.	INT
<i>feridos_graves</i>	Valor binário que identifica se o envolvido foi classificado como ferido grave.	INT
<i>mortos</i>	Valor binário que identifica se o envolvido foi classificado como morto.	INT
<i>latitude</i>	Latitude do local do acidente em formato geodésico decimal.	FLOAT
<i>longitude</i>	Longitude do local do acidente em formato geodésico decimal.	FLOAT
<i>regional</i>	Regional responsável	VARCHAR
<i>delegacia</i>	Identificador da Delegacia de registro	VARCHAR
<i>uop</i>	Identificador UOP	VARCHAR

Acidentes por pessoa – Fonte: Dados abertos da Polícia Rodoviária Federal

Sistema de Estatísticas de Automóveis – AUTOSEG - SUSEP -
<http://www2.susep.gov.br/menuestatistica/Autoseg/respmenu1.aspx>

formato *.html/.xls*.

Estatísticas AutoSeguro - SUSEP		
Nome da coluna/campo	Descrição	Tipo
<i>Categoria</i>	Categoria do veículo segurado	Texto
<i>Região</i>	Unidade e nome da federação identificado como "pernoite do veículo" na apólice de seguro	Texto
<i>Grupo</i>	Grupo de seguro	Texto
<i>Ano Modelo</i>	Ano e modelo do veículo segurado]	Texto
<i>Sexo Condutor</i>	Gênero apontado na cláusula "principal condutor" da apólice	Texto
<i>Faixa Etária</i>	Faixa etária do "principal condutor" apontado na apólice	Texto
<i>IS Média (R\$)</i>	Indenização securitária média	Moeda
<i>Expostos</i>	Qtd de carros expostos	Numeral
<i>Prêmio Médio (R\$)</i>	Valor pago pelo seguro	Moeda
<i>Freq. Incêndio e Roubo</i>	Frequência de incêndio e roubo	Numeral
<i>Indeniz. Incêndio e Roubo (R\$)</i>	Indenização paga por incêndio e roubo	Moeda
<i>Freq. Colisão</i>	Frequência de colisão	Numeral
<i>Indeniz. Colisão (R\$)</i>	Indenização paga por colisão	Moeda
<i>Freq. Outras</i>	Outras estatísticas	Numeral
<i>Indeniz. Outras (R\$)</i>	Indenizações de outras estatísticas	Moeda

OBS.: até o momento desta consulta, o site da Susep disponibiliza dados apenas até 06/2019.

3. Processamento/Tratamento de Dados

Ferramenta utilizada para ETL e visualização de dados: Power BI

Download e descompactação dos arquivos da PRF (.csv) na pasta definida a saber: “*Acidentes agrupados por pessoa*”;

Transformação de dados da PRF:

Importar / Folder / acidentes agrupados por pessoa / Combine & Transform Data.

Invocada função “Transform file” usando como parâmetro a coluna “Content”.

Utilizado recurso de primeira linha como título das colunas.

Usada função de expansão de colunas, para visualizar as colunas contidas em cada tabela.

Aplicação dos tipos corretos para cada coluna (string / int / float / date / time)

Criada coluna “Mês” referente ao nome mês da data do registro do acidente.

Criada coluna “Ano” referente ao nome ano da data do registro do acidente.

Criada coluna de faixa etária para corrigir desvios de preenchimento contidos na base, na coluna “idade”. Para isso, foi criada uma coluna de idade definida no tipo ‘Int’, convertendo o valor texto “NA” em 0, para que toda coluna pudesse ser transformada em numeral.

Criada uma coluna de agrupamento das causas de colisão encontradas no dataset para possibilitar uma boa visualização destes dados no card de métricas.

Registros obtidos da PRF: **221.161** registros totais, sob as seguintes definições:

Os registros identificados com classificações diferentes de “Condutor” (passageiro, testemunha, pedestre etc.), foram desconsiderados pois não são objetos deste estudo, bem como

Também foram desconsiderados registros de acidentes cujo bem não há classificação de seguro (Bicicleta, charrete, carroça), bem como gênero(f/m) classificados como “Não informado” e “Ignorado”.

Nesta análise foram considerados a quantidade de condutores envolvidos no mesmo acidente, por exemplo: um caminhão colide com um automóvel: dois condutores.

Download dos arquivos da Susep (.html) e conversão (.xlsx) na pasta definida “Sinistros Susep”.

Importar / arquivo Excel / susep 2018 e susep 2019

Criação de coluna adicional de ano, em cada arquivo

Append dos arquivos, tornando-os base única.

Criação de uma coluna agrupando os veículos de mesmo tipo, para ser similar aos dados de estudo, a saber dados da PFR.

Criação de uma coluna separando o nome do estado da UF, pois eles estão na mesma coluna.

A classificação de 'Nacional/Importado' não foi levado em conta nas estatísticas aqui apresentadas pela ausência de detalhes em ambas as fontes.

4. Análise e Exploração dos Dados

Durante a análise foi possível identificar a maior distinção de comportamentos em determinada faixa etária: quanto mais avançada a idade, maior diferença de números entre os dois grupos (Homens X Mulheres), tanto nos dados enviados pelas seguradoras (Susep) quanto na base da PRF.

Também foi possível identificar uma curva totalmente contrária entre os dados da PRF e da Susep, esta última mostrando um número maior de acidentes para as mulheres.

Esse desvio poderia ser explicado justamente pela diferença nos preços de seguro para mulheres, levando ao que as seguradoras identificam como fraude no quesito “principal condutor”. Ou seja, sabendo que o valor do seguro é mais barato para a mulher, o principal condutor faz a apólice de seguro no nome da esposa, mãe, tia. Quando aquela apólice é acionada, entra para contagem o “principal condutor” apontado na apólice, não o real condutor no momento do acidente.

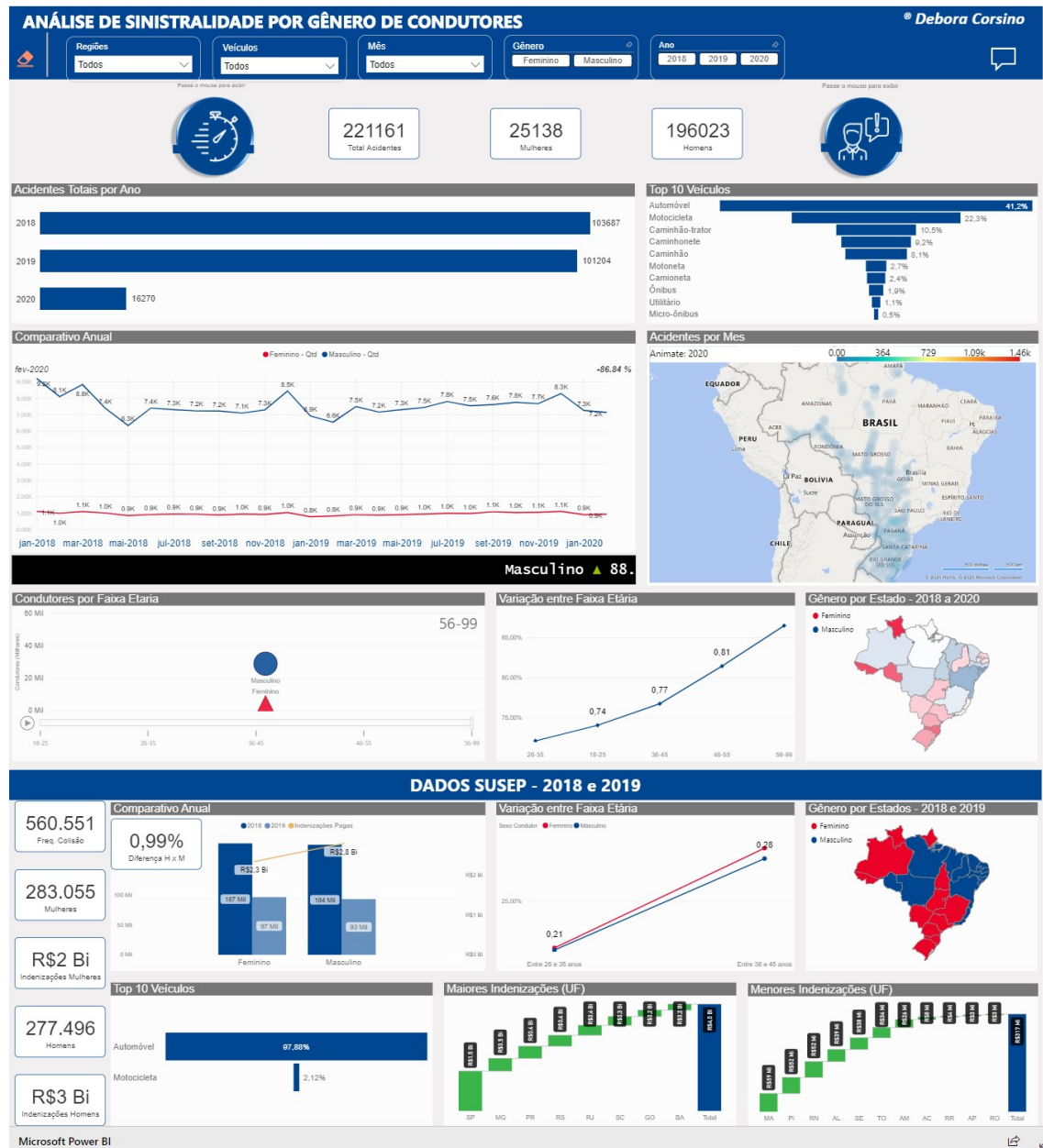
Já os boletins de ocorrência registrados pela PRF apontam o real condutor no exato momento do acidente tornando a medida mais precisa, mostrando um valor 86% maior para os homens.

5. Criação de Modelos de Machine Learning

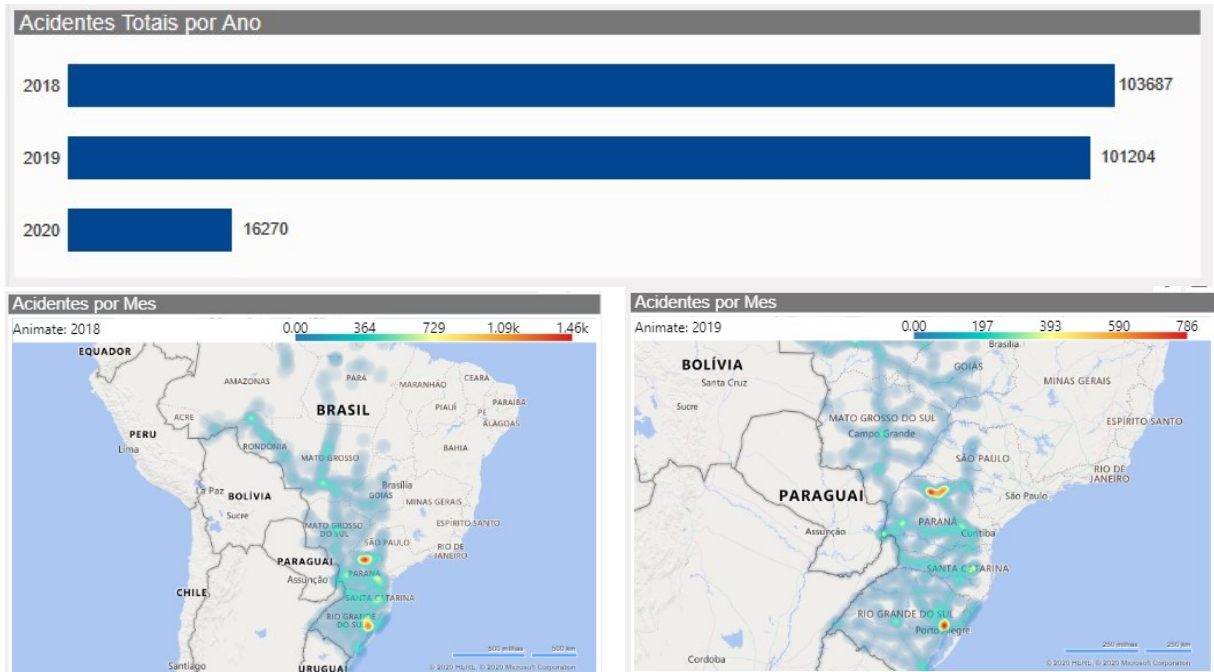
Não utilizado.

6. Apresentação dos Resultados

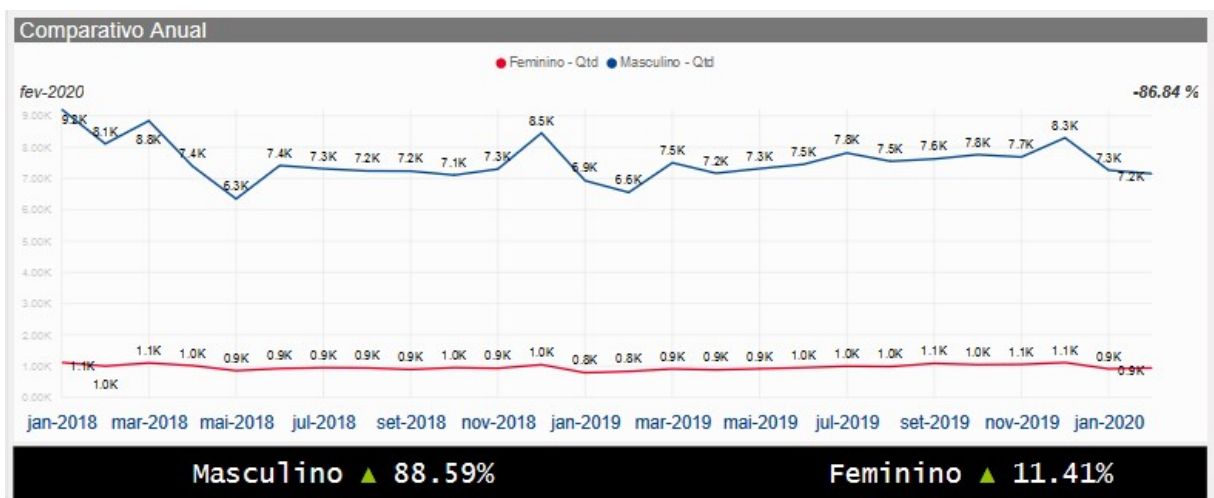
A seguir, apresento os resultados desta análise no dashboard realizado no Power BI ([disponível também neste link](#)):



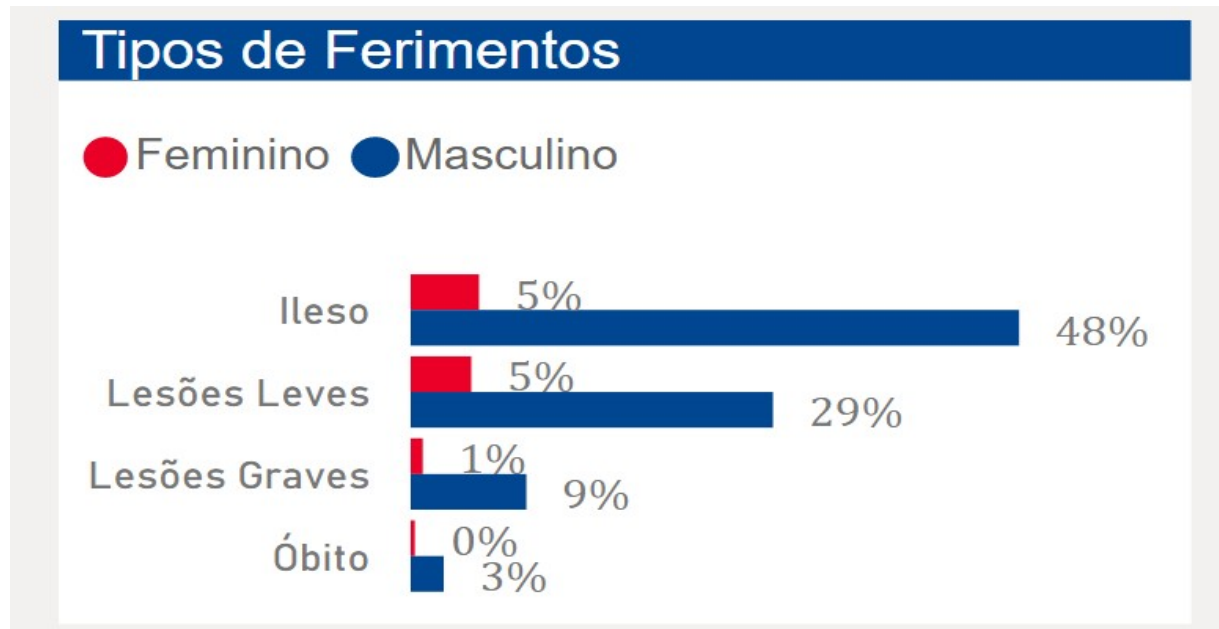
Notamos uma certa diminuição dos acidentes entre 2018 e 2019, em todo o território nacional, saindo de 103.687 para 101.204 acidentes, provavelmente graças a políticas de conscientização no trânsito:



Conseguimos comprovar a informação do Denatran quanto a quantidade de acidentes causadas por mulheres e homens, sendo esta diferença, gritante, com 87,63% em dezembro de 2018 e chegando a 86,84% a mais para os homens, em fevereiro de 2020.



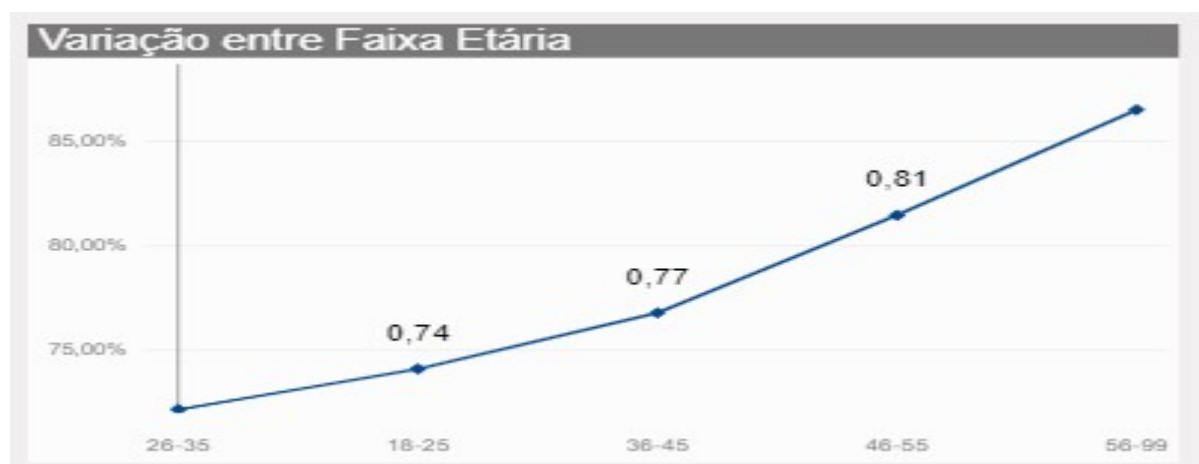
Ainda em relação às informações do governo, através do ministério da saúde, como citado anteriormente, o número de mortos nestes acidentes é maior para os homens, trazendo uma diferença de 3%.



No quesito idade, notamos que, entre 18 e 35 anos, há um aumento gradativo de acidentes tanto por motoristas homens quanto mulheres.

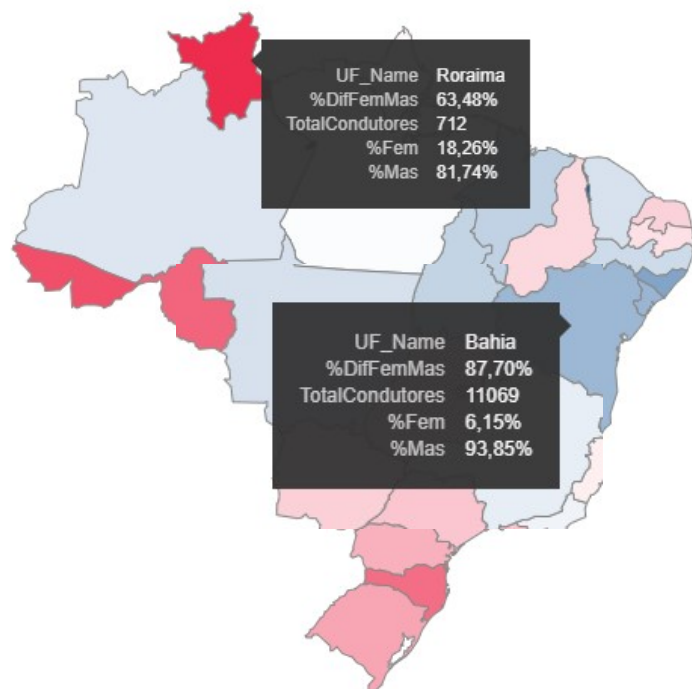


No entanto, a partir dos 36 anos, a distância entre eles vai ficando cada vez maior.

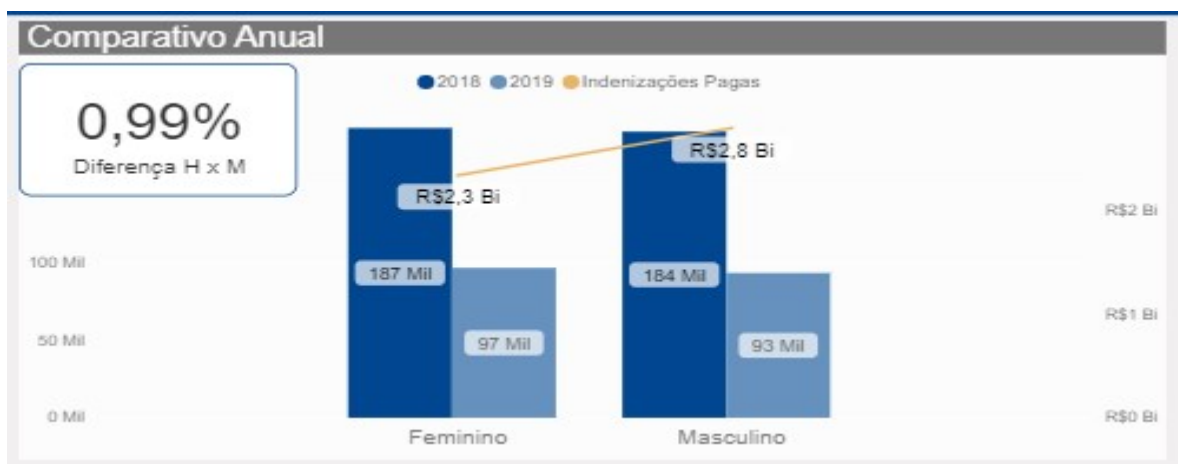


Identificamos também que a variação percentual entre homens e mulheres aumenta ou diminui em determinadas regiões. O mapa abaixo tem o colorido mostrando a diferença entre estes dois grupos de forma que, quanto mais azul, maior a diferença entre eles (maior porcentagem de acidentes de motoristas homens) e, quanto mais próximo do vermelho, menor a diferença entre eles (diferença entre eles abaixo de 80%)

GÊNERO POR ESTADO - 2018 A 2020

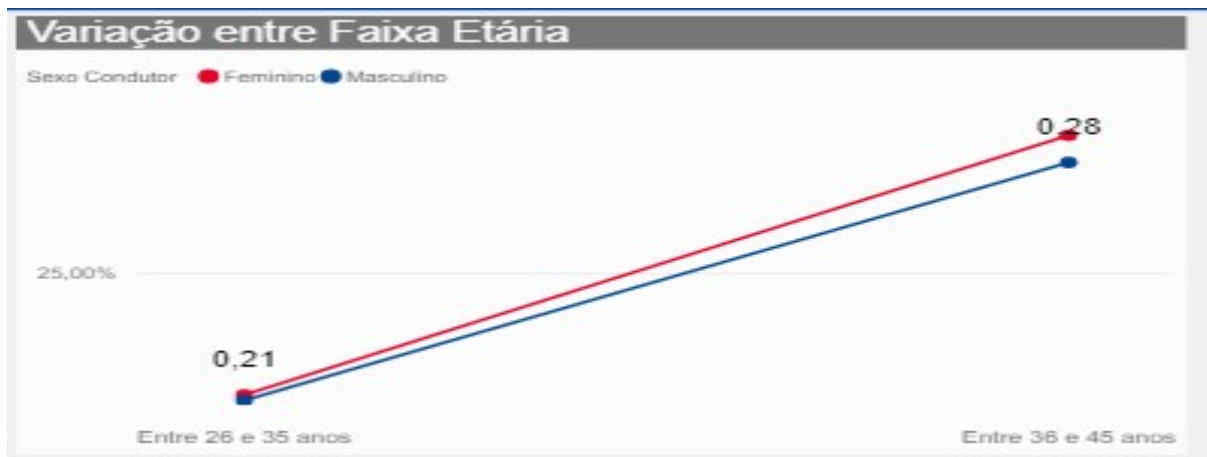


Nos dados apresentados pela Susep, o percentual maior fica para as mulheres, tendo como hipótese para tal fenômeno, o menor custo do seguro. Porém a diferença não chega a 1%:

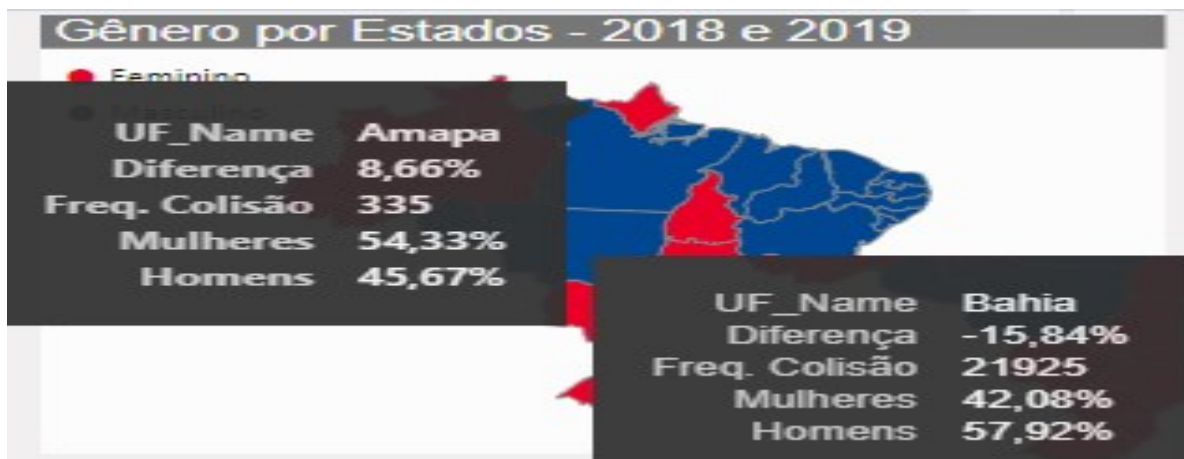


E, mesmo com essa diferença a maior para as mulheres, os valores pagos em indenizações para as mulheres é menor do que para os homens, como mostrado no gráfico acima: R\$2,3 bilhões em 2018 para apólices com mulheres como principal condutor (contratualmente) versus R\$2,8 bilhões no mesmo período, para homens.

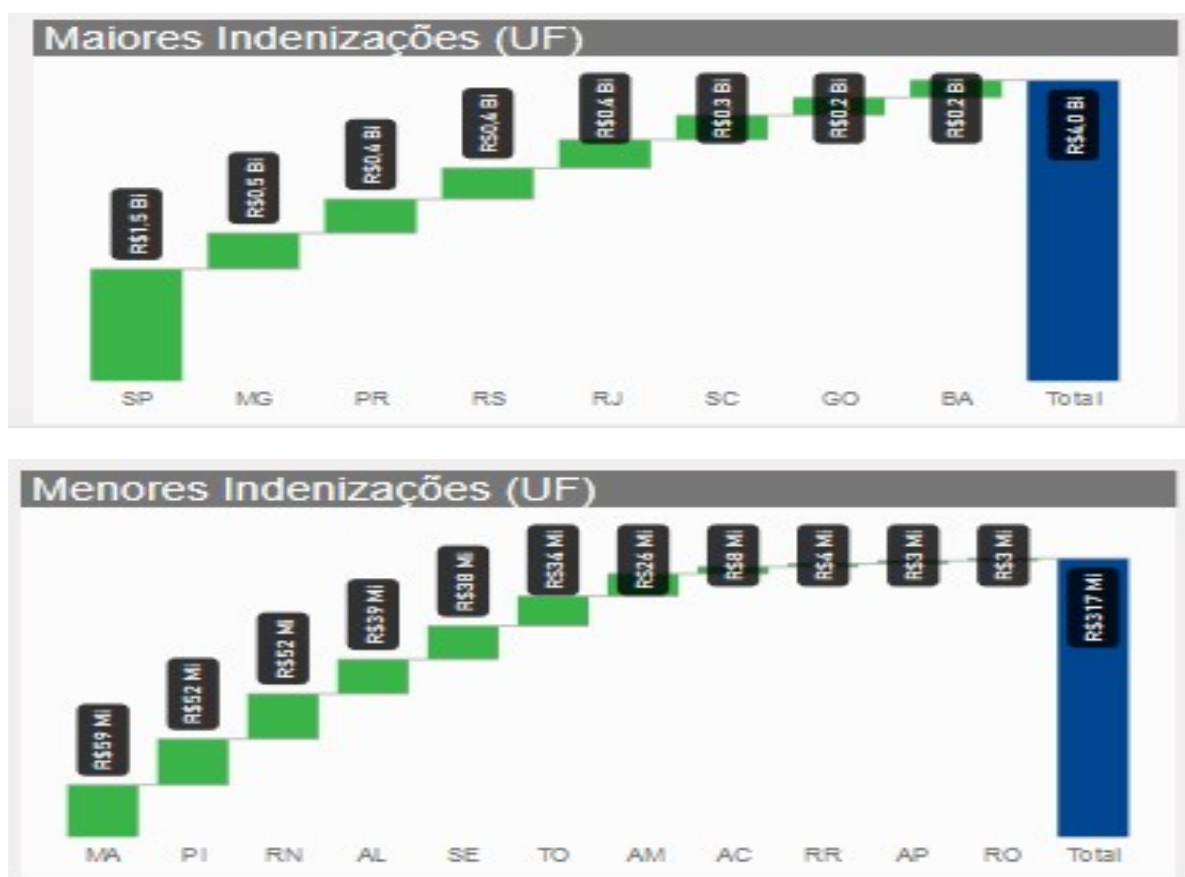
A diferença entre idade, segue o padrão apresentado pelos dados da PRF: Quanto maior a idade do condutor, mais aumenta a diferença entre eles:



Fizemos o mesmo raciocínio para apresentar o resultado pelos estados federativos. Azul representa maior quantidade de sinistros causados por homens enquanto o vermelho, maior sinistralidade para as mulheres.



Por fim, identificamos os estados federativos onde há maior gasto com indenizações, fazendo com que o preço das apólices de seguro sejam mais altas (SP R\$ 1,5Bilhões, MG R\$ 0,5Bilhões, PR e RS R\$0,4Bilhoes), bem como os que teriam menor valor do seguro, devido ao menor gasto com indenizações (MA R\$59Milhões, PI R\$52Milhões, RN R\$52Milhões) – valores pagos em 2018:



Com base no exposto, a análise mostra números que corroboram para que o seguro da mulher tenha um diferencial no preço. Contudo, é necessária aplicação de mais políticas antifraudes no momento da aquisição do seguro para que o desconto seja usufruído somente por quem tem o direito e não haver prejuízo para as seguradoras que praticarem esta política.

7. Links

Segue os links de todo material deste trabalho.

Dashboard - <https://bit.ly/3eoa8OE>

Vídeo - <https://bit.ly/2XCdE1R>

Repositório - <https://bit.ly/3bnwIVm>

REFERÊNCIAS

<https://www.revistaapolicy.com.br/2018/05/seguero-auto-para-as-mulheres-e-em-media-28-mais-barato/>

<https://blog.suhaiseguradora.com/seguero-de-carro-para-mulheres-mais-barato/>

<https://jornaldocarro.estadao.com.br/carros/seguero-para-mulheres-mais-carro/>