

Trabalho Prático II

Miguel Antonio Ribeiro e Silva - 4680

João Victor Graciano Belfort de Andrade - 4694

Mateus Henrique Vieira Figueiredo - 4707

Alan Gabriel Martins Silva - 4663

1. Introdução.....	1
2. Resumo.....	1
3. Preparação do ambiente.....	2
4. Modelo estrela.....	3
5. Data Warehouse.....	4
6. Aplicação OLAP.....	5
7. Drill-Down e Roll-Up.....	8
8. Relatórios.....	9
9. Power BI – Outra aplicação OLAP.....	15
10. Data Mining.....	19
11. Comparação.....	19
12. Análise do negócio.....	20
13. Conclusão.....	21
14. Bibliografia.....	21

1. Introdução

A presente documentação aborda as atividades executadas em conformidade com as diretrizes estabelecidas. Cada seção subsequente delinea uma tarefa específica realizada.

O escopo deste trabalho consiste na análise de uma seguradora de veículos, com foco na modelagem do seu banco de dados relacional, implementação de um data warehouse e condução de análises pertinentes. As ferramentas empregadas para a realização dessas atividades incluíram o SQL Server 22, Visual Studio 2019, Excel e Power BI.

Esta documentação refere-se ao projeto desenvolvido no âmbito do Trabalho Prático II da disciplina Gestão Recuperação e Análise de Informações, identificada pelo código CCF 424. No decorrer do trabalho, propôs-se a análise de um banco de dados transacional de uma seguradora de veículos, com o intuito de conceber e implementar um data warehouse. O detalhamento do projeto apresentado neste documento encontra-se disponível na página do GitHub sob o título [SQL-Server-Data-Analysis](#).

2. Resumo

A análise começou com a compreensão aprofundada do banco de dados transacional da seguradora de veículos. Esta fase envolveu a revisão minuciosa das tabelas, relações entre os dados e os requisitos específicos do projeto. Com base nessa compreensão, foi possível realizar a modelagem do banco de dados relacional. Esse passo é essencial, pois estabelece a estrutura sobre a

qual o data warehouse será construído, garantindo que as informações sejam organizadas de maneira lógica e acessível.

A implementação do data warehouse foi conduzida com a ajuda do SQL Server 22 e do Visual Studio 2019. Durante esse processo, os dados foram extraídos, transformados e carregados (ETL) no data warehouse, garantindo consistência e integridade das informações. A criação de dimensões e fatos adequados foi cuidadosamente realizada para oferecer uma base sólida para as análises subsequentes.

Além disso, a utilização do Excel e do Power BI permitiu a criação de visualizações interativas e relatórios dinâmicos. Essas ferramentas facilitaram a interpretação dos dados e a comunicação eficaz das descobertas. As análises realizadas abrangeram uma variedade de métricas e indicadores relevantes para a seguradora de veículos, oferecendo insights valiosos sobre padrões de sinistros, comportamentos dos clientes e tendências de mercado.

As ferramentas mencionadas desempenharam um papel fundamental, proporcionando as funcionalidades necessárias para transformar dados brutos em insights acionáveis. O acesso ao projeto completo no GitHub permite uma revisão mais aprofundada do trabalho realizado, oferecendo uma visão completa e transparente do processo e dos resultados obtidos.

3. Preparação do ambiente

No processo de tratamento dos dados recebidos pelo grupo, foi adotada uma abordagem estruturada para criar um modelo relacional utilizando o **MySQL Workbench**, conforme indicado nas especificações do projeto. O resultado desse esforço pode ser visualizado na Figura [1], que representa o Modelo Relacional desenvolvido.

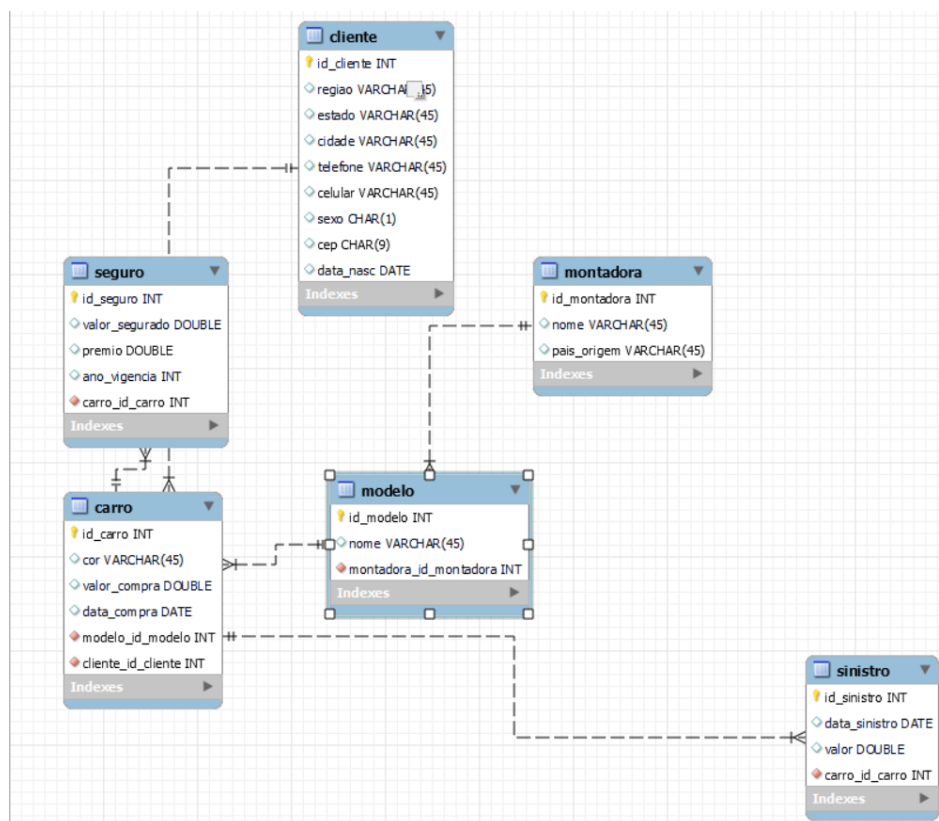


Figura 1 – Modelo Relacional

Para transferir dados brutos para um modelo relacional, foi usado um método eficiente. Os dados foram exportados para arquivos **.csv** para garantir portabilidade e facilidade de manipulação, automatizado por um script Python. O script extrai e converteu os dados para **.csv**, mantendo integridade e precisão. O arquivo **.csv** serviu como ponto de entrada para o banco de dados no SQL Server, onde os dados foram inseridos de maneira eficiente e mapeados para tabelas correspondentes. Esse processo permitiu consultas complexas e análises detalhadas, facilitando a exploração significativa dos dados.

4. Modelo estrela

Toda a **seção 4** deste trabalho se refere à **Tarefa 1**. No processo de transformação dos dados, foi decidido criar um modelo estrela com base nas informações já inseridas no modelo relacional. Este modelo estrela foi projetado para otimizar a análise dos dados, permitindo uma compreensão mais **profunda dos sinistros** na seguradora de veículos. As tabelas que compõem o modelo estrela são as seguintes:

Tabelas Dimensão: (dim_carro, dim_cliente, dim_modelo, dim_montadora) As tabelas de dimensões escolhidas são aquelas referentes e correlacionadas com a tabela fato_sinistro, portanto guardam as informações referentes às demais informações necessárias para a análise dos dados

Tabela fato: (fato_sinistro) A tabela de fatos, fato_sinistro, é o núcleo do modelo estrela. Ela contém métricas e medidas relacionadas aos sinistros, representando o centro das análises. As informações cruciais sobre os sinistros.

É importante mencionar que a tabela "seguro" foi deliberadamente retirada durante a construção do modelo estrela. A decisão foi tomada com base na falta de associação entre a tabela seguro e a tabela de sinistro, que foi escolhida como a tabela de fatos. Além disso, as informações contidas na tabela seguro não eram pertinentes para a análise em questão e poderiam resultar em dados duplicados durante a inserção no modelo estrela.

A escolha de focar na tabela de sinistro como a principal tabela de fatos reflete a natureza crítica dos sinistros para a seguradora. Pagar pelos sinistros aos segurados é uma das atividades mais importantes da empresa e, portanto, a análise desses dados é de grande relevância para a tomada de decisões estratégicas. Os scripts .sql para a transferência do modelo relacional para o modelo estrela estão disponíveis no GitHub.

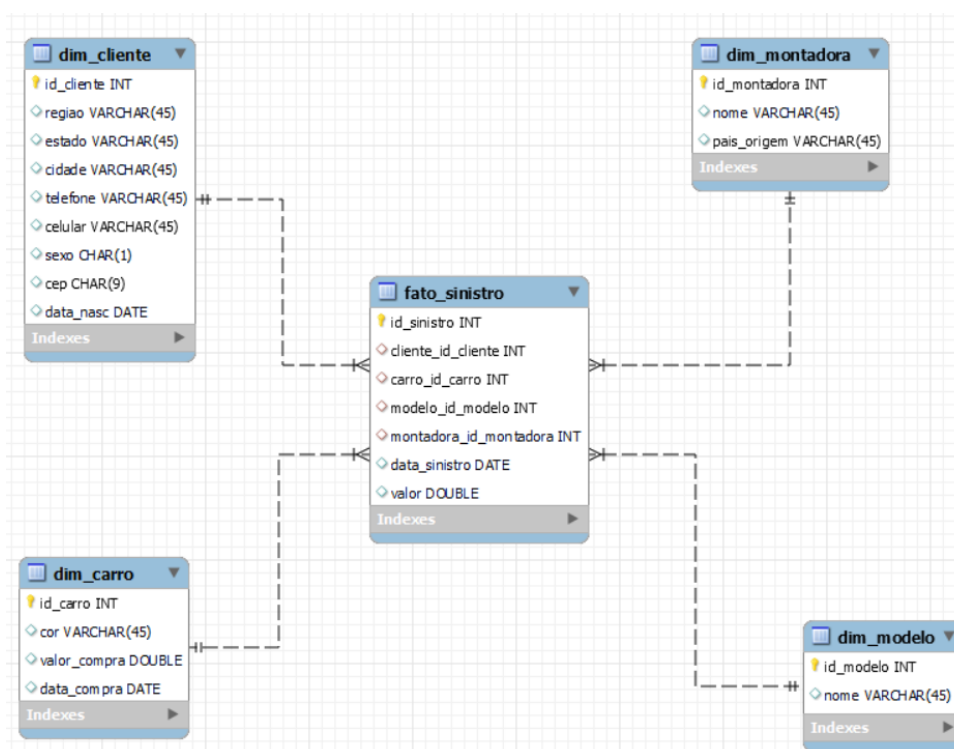


Figura 2 – Modelo Estrela

5. Data Warehouse

Toda a **seção 5** deste trabalho se refere à **Tarefa 2**. No âmbito do projeto, foi desenvolvido um Data Warehouse para a análise eficaz dos dados. Para este propósito, foi criado um cubo de dados usando o **SQL Server**. Essa criação foi facilitada por meio de uma extensão do **Analysis Services** no **Visual Studio 2019**, que proporcionou uma plataforma robusta para o design e a implementação do cubo.

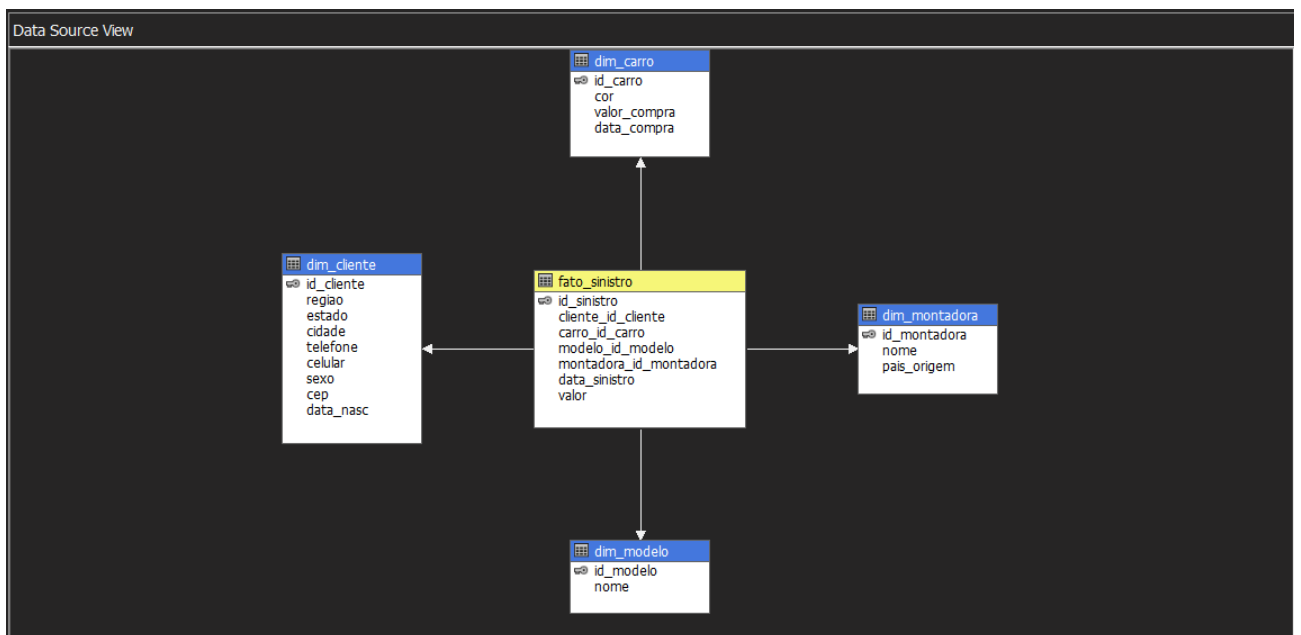


Figura 3 – Data Warehouse

A **figura 3** representa a visualização do Data Source View no Visual Studio. Esta visualização é essencial para entender a estrutura dos dados dentro do Data Warehouse. Mostra como as tabelas do modelo estrela, como `dim_carro`, `dim_cliente`, `dim_modelo`, `dim_montadora` e `fato_sinistro`, estão inter-relacionadas dentro do ambiente do Data Warehouse. No Visual Studio 2019, a extensão do Analysis Services foi instrumental para criar esse ambiente de Data Warehouse e cubo de dados de forma eficiente. A integração entre as ferramentas facilitou o processo de desenvolvimento, garantindo que o Data Warehouse estivesse pronto para análises profundas e significativas.

6. Aplicação OLAP

Toda a **seção 6** deste trabalho se refere à **Tarefa 3**. Após a criação do cubo no Visual Studio 2019, os dados foram exportados para o Excel usando o Analysis Services do SQL Server, permitindo uma análise mais detalhada e interativa. Diversas perguntas foram formuladas para explorar os dados e orientar as decisões estratégicas da seguradora. As tabelas derivadas do cubo principal oferece insights valiosos sobre os sinistros, possibilitando respostas precisas para questões específicas a serem respondidas aos executivos:

Qual é o valor médio do sinistro por cidade, estado e região?- Permite identificar padrões geográficos nos custos dos sinistros. Essa informação é vital para alocar recursos de forma eficaz e ajustar as políticas de seguro conforme as regiões.

Qual é o valor médio do sinistro por marca e modelo de carro?- Ajuda a entender as diferenças nos custos dos sinistros entre fabricantes e tipos de veículos. Isso é crucial para determinar prêmios de seguro e avaliar o desempenho dos modelos em termos de segurança.

Como o valor médio do sinistro varia de acordo com o sexo dos segurados e em diferentes regiões?-Permite avaliar se existem disparidades de custos dos sinistros com base no gênero dos segurados, além de fornecer uma visão geográfica dessas discrepâncias. Essa análise é essencial para garantir a equidade nas políticas de seguro.

Qual é o valor médio do sinistro por cidade, estado e região?

Rótulos de Linha	Valor	Quantidade	Média
centro-oeste			
GO			
anapolis	92555977.45	3349	27636.89981
baliza	65256374.84	2349	27780.49163
batiza	29675554.27	1068	27786.0995
caldas novas	154018898.8	5526	27871.67912
goiania	68997593.42	2440	28277.70222
trindade	108345902.5	3881	27917.00656
MT			
cuiaba	204948620.3	7420	27621.10786
jangada	278752405	10153	27455.1763
ponte branca	296582576.8	10680	27769.9042
sorriso	465731171.7	16748	27808.16645
vila rica	335554159.6	11930	28126.92034
nordeste			
BA			
brejolandia	286455775.6	10236	27985.12853
caravelas	305720576.8	11113	27510.17518
ilheus	456034769.6	16485	27663.61963
juazeiro	336098469.1	11997	28015.20956
salvador	208902065.3	7497	27864.7546
PE			
belo jardim	1254811159	44900	27946.79641
olinda	1133590826	40828	27765.03443
petrolina	1056191529	38193	27654.06041
recife	767307876.7	27568	27833.2805
vertentes	1740987379	62753	27743.49241

Figura 4

Como o valor médio do sinistro varia de acordo com o sexo dos segurados e em diferentes regiões?

Rótulos de Linha	Valor	Quantidade	Média
chevrolet			
camaro	5114020540	67568	75687.02
cruze	1011841780	44879	22546
onix	280393765.5	30503	9192.334
onix plus	181169534.9	18729	9673.209
s10	3812635349	72893	52304.55
fiat			
argo	1060195510	98145	10802.34
mobi	357039523.5	41082	8690.899
strada	605412887.2	60116	10070.74
toro	4279123949	92145	46439.02
uno	235723707.6	25671	9182.49
ford			
fiesta	46282442.44	4441	10421.63
focus	191524715.9	10805	17725.56
fusion	799353968.4	17319	46154.74
ka	62422093.12	7255	8604.01
mustang	1254934313	16133	77786.79
honda			
city	298545210.7	29169	10235.02
civic	1857240840	42324	43881.51
fit	209626005	17684	11853.99
hrv	801425556.1	69292	11565.92
wrv	720302789.1	65452	11005.05

Figura 5

Existe uma correlação entre a frequência dos sinistros e a idade de compra dos veículos segurados?

Rótulos de Linha	Valor	Quantidade	Média
F			
centro-oeste	146242766.5	6314	23161.67
nordeste	518219503.9	22297	23241.67
norte	617598684.4	23198	26622.93
sudeste	479516358.5	20243	23688.01
sul	123502770.9	5463	22607.13
M			
centro-oeste	1954176468	69230	28227.31
nordeste	7027880922	249273	28193.51
norte	8108364238	255735	31706.12
sudeste	6299297492	222695	28286.66
sul	1682981846	59607	28234.63
Total Geral	26957781051	934055	28861.02

Figura 6

7. Drill-Down e Roll-Up

Toda a seção 7 deste trabalho se refere à Tarefa 4. Buscamos utilizar técnicas avançadas como Drill-Down e Roll-Up para enriquecer a análise dos dados. Essas estratégias permitem explorar a hierarquia dos cubos já apresentados anteriormente, oferecendo uma visão mais detalhada ou mais geral dos dados, conforme necessário. Ao utilizar o Drill-Down, podemos aprofundar nossa análise, explorando categorias específicas em uma dimensão ou em vez de analisar dados em um nível muito detalhado, podemos usar o Roll-Up para ter uma visão mais geral. No Excel, isso é feito expandindo uma categoria ou colapsando conforme o necessário.

Rótulos de Linha ▾	Valor	Quantidade	Média
⊕ chevrolet	10400060969	234572	44336.33
⊕ fiat	6537495578	317159	20612.68
⊕ ford	2354517533	55953	42080.27
⊕ honda	3887140401	223921	17359.43
⊕ toyota	2422620317	46625	51959.69
⊕ volkswagen	1355946252	55825	24289.23
Total Geral	26957781051	934055	28861.02

Figura 7 – Granularidade aumentada para marca-modelo

Rótulos de Linha ▾	Valor	Quantidade	Média
⊖ centro-oeste			
⊕ GO	518850301.3	18613	27875.69447
⊕ MT	1581568933	56931	27780.45236
⊖ nordeste			
⊕ BA	1593211656	57328	27791.16063
⊕ PE	5952888770	214242	27785.8159
⊖ norte			
⊕ AM	2006925848	63919	31397.9544
⊕ PA	6719037075	215014	31249.30039
⊖ sudeste			
⊕ MG	1963302474	70453	27866.83994
⊕ SP	4815511376	172485	27918.43567
⊖ sul			
⊕ RS	476097361	17221	27646.3249
⊕ SC	1330387256	47849	27803.8675
Total Geral	26957781051	934055	28861.02109

Figura 8 – Granularidade aumentada para região-estado

Rótulos de Linha ▾	Valor	Quantidade	Média
⊕ centro-oeste	2100419235	75544	27803.91871
⊕ nordeste	7546100426	271570	27786.94416
⊕ norte	8725962922	278933	31283.36526
⊕ sudeste	6778813850	242938	27903.4727
⊕ sul	1806484617	65070	27762.17331
Total Geral	26957781051	934055	28861.02109

Figura 9 – Granularidade aumentada para estado

Rótulos de Linha ▾	Valor	Quantidade	Média
⊕ F	1885080084	77515	24318.91
⊕ M	25072700966	856540	29272.07
Total Geral	26957781051	934055	28861.02

Figura 10 – Granularidade aumentada para sexo

8. Relatórios

A **seção 8** aborda a **Tarefa 5**, na qual estamos empenhados em gerar relatórios por meio do Excel para visualizar os dados, analisar a relação de sinistros conforme diferentes parâmetros e compará-los. O objetivo é fornecer um feedback aos executivos da empresa.

Os dois gráficos a seguir representam a média de sinistros por região e a quantidade total de sinistros. Observa-se que a média é bastante homogênea entre as diversas regiões do Brasil, com uma pequena discrepância de 2% na região Norte.

No entanto, torna-se evidente que a quantidade de sinistros acionados é significativamente maior nas regiões Norte, Nordeste e Sudeste.

Partindo para o gráfico que representa a média por estado, observamos informações relevantes. A média de sinistros nos estados do Norte é significativamente mais alta do que nos demais. Essa disparidade pode ser atribuída ao fato de que os clientes desses estados possuem veículos cujos sinistros foram mais frequentemente acionados.

Entretanto, para uma análise mais completa, seria útil investigar os fatores que contribuem para essa maior incidência de sinistros. Pode ser interessante explorar aspectos como as condições de tráfego, padrões de comportamento dos motoristas ou até mesmo questões relacionadas à infraestrutura viária.

Média de Valor de Sinistro por Região

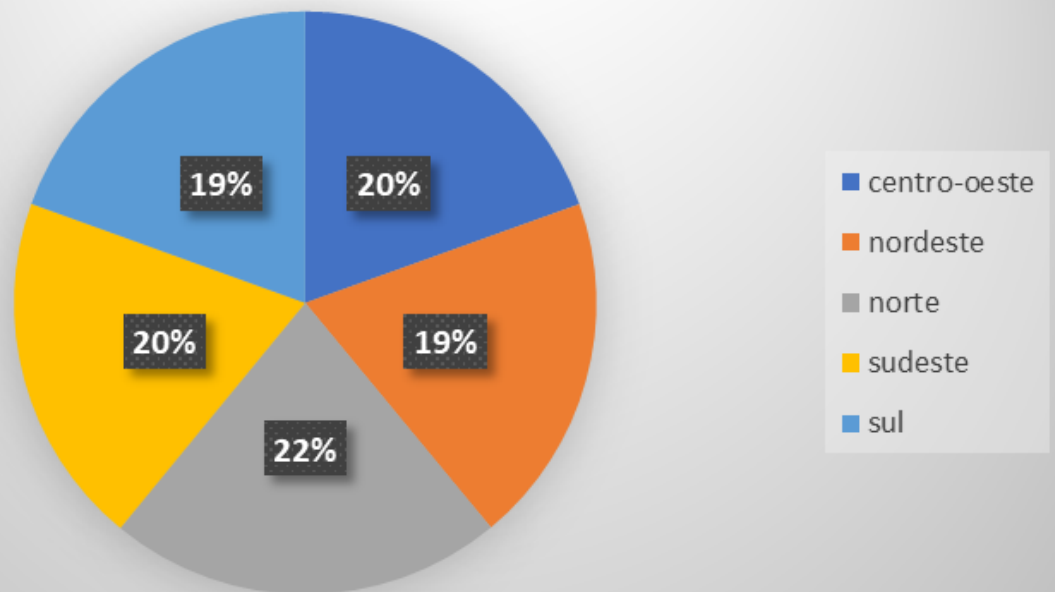


Figura 11 – Média de valor de sinistro por região

Quantidade de Sinistros Por Região

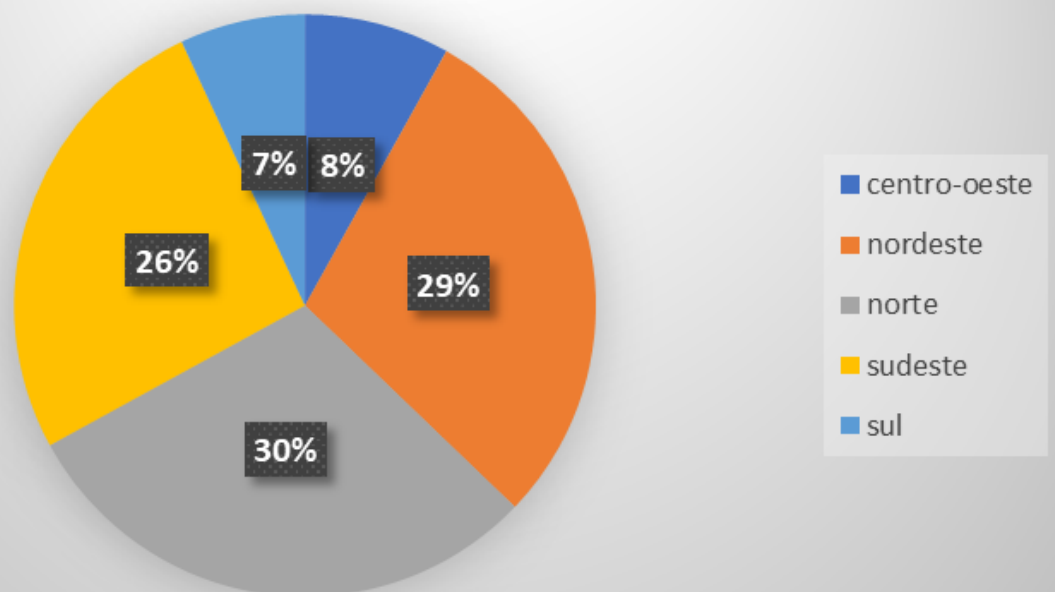


Figura 12 – Quantidade de sinistros por região

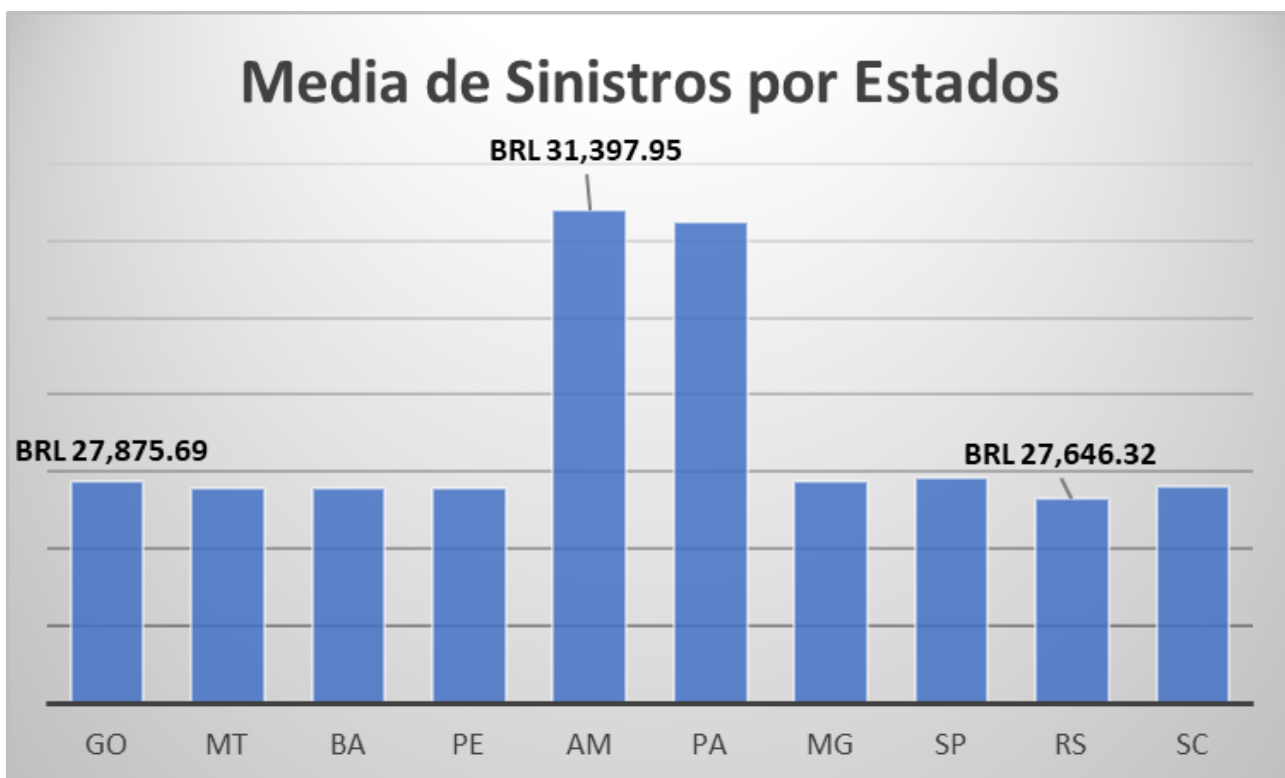


Figura 13 – Média do valor de sinistros por estado

Partindo para a próxima análise e visualizando os relatórios referentes a **modelos e marcas de carros**, examinamos o gráfico a seguir, que representa a média de sinistros por modelo de carro, utilizando um gráfico de Pareto para ilustrar as proporções. No gráfico, observa-se que os modelos cujos média de sinistros são maiores estão no início (Mustang), seguindo uma ordem decrescente, e no final (Uno).

Ao analisar o valor total, focando em alguns carros específicos, nota-se que a empresa desembolsa valores substanciais em sinistros para veículos como Camaro, Toro e S10. Nesse contexto, é imperativo que a empresa realize uma análise mais aprofundada para compreender as razões por trás desses elevados custos associados a sinistros. Estratégias devem ser estudadas para abordar e explicar essas discrepâncias, potencialmente incluindo a revisão de políticas de seguro, implementação de medidas de prevenção ou até mesmo negociações diretas com fabricantes. Mas para o grupo, entendemos que por serem carros mais luxuosos, o gasto é maior.

Analisando a média de sinistros por montadora, é relevante observar que as montadoras Toyota, Chevrolet e Ford apresentam médias mais elevadas de sinistros. Isso pode ser atribuído ao fato de que carros dessas marcas geralmente são mais luxuosos no nosso *dataset* e têm um custo mais elevado para a empresa em caso de sinistro. Esta observação destaca a necessidade de a empresa considerar estratégias específicas para gerenciar sinistros relacionados a veículos dessas montadoras, como implementação de programas de treinamento para motoristas, revisão de políticas de seguro ou até mesmo negociações comerciais com as montadoras para obter condições mais favoráveis.

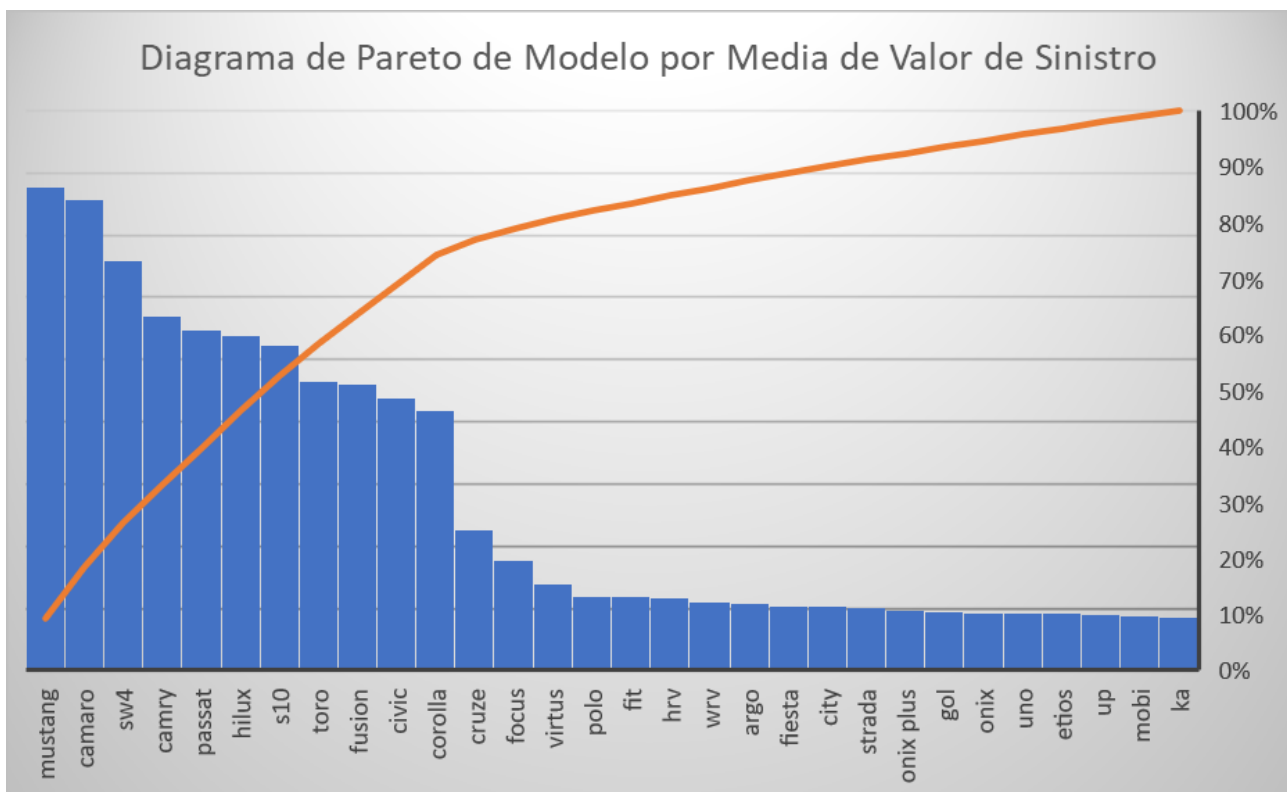


Figura 14 – Diagrama de Pareto de modelo por médio do valor do sinistro

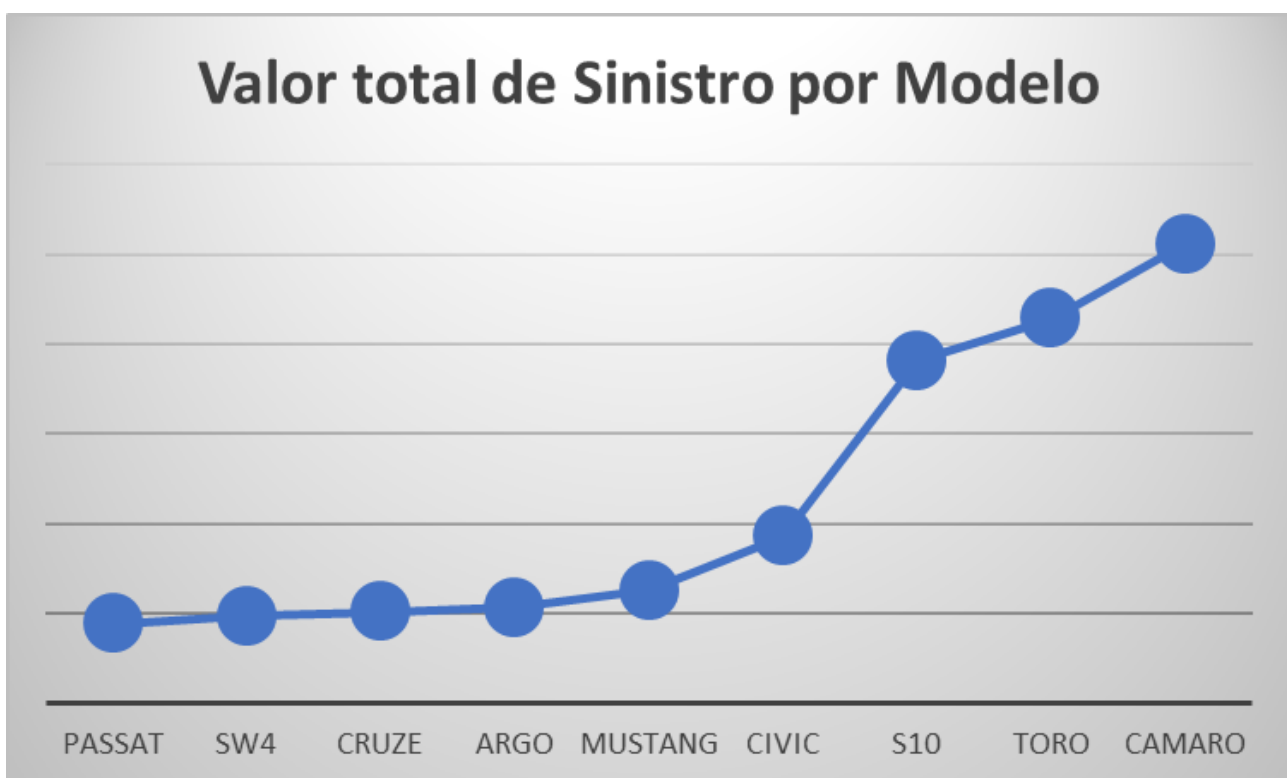


Figura 15 – Valor total de sinistro por modelo

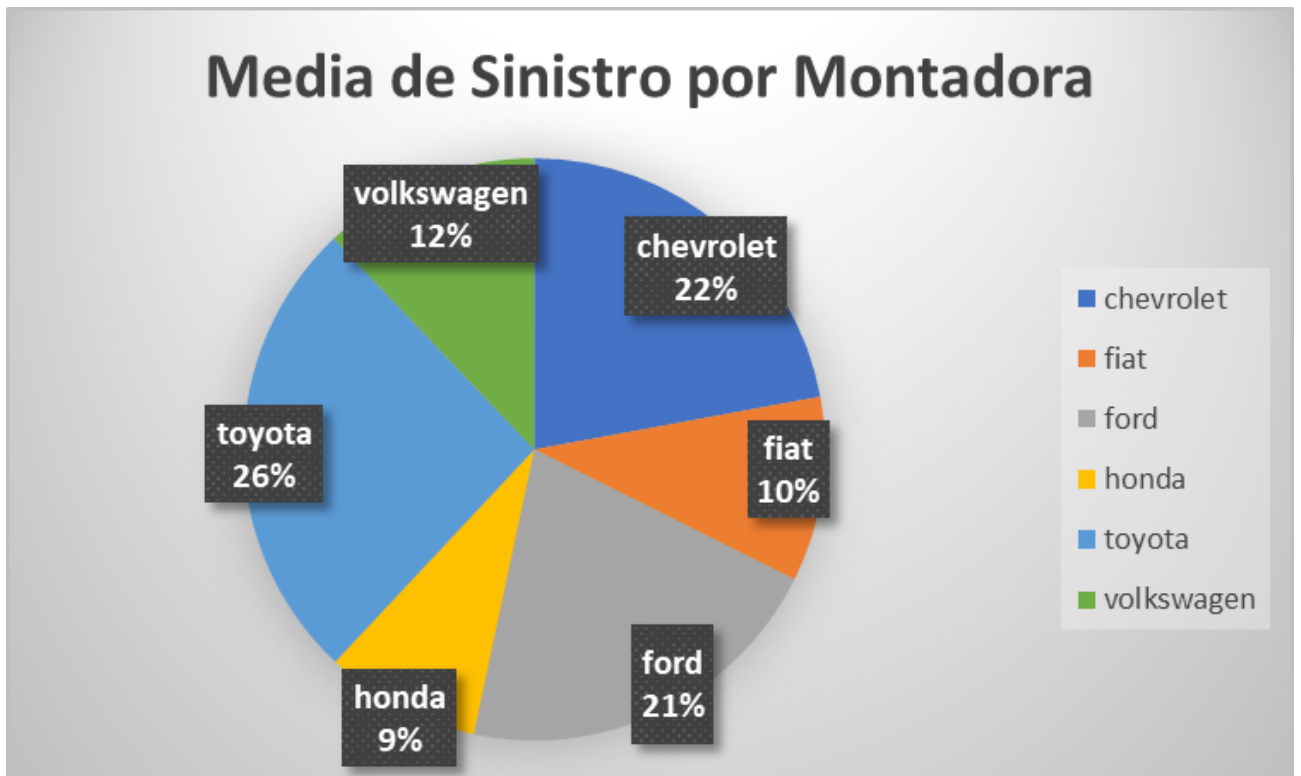


Figura 16 – Média de sinistro por montadora

Agora, ao analisarmos a última pergunta sobre a média de sinistros para homens e mulheres, observamos uma diferença significativa. A média de sinistros para homens é maior do que para mulheres, provavelmente devido a algumas razões. São só suposições, para uma análise mais certa deveríamos aprofundar esse estudo.

Comportamento de condução: Estudos sugerem que, em média, as mulheres tendem a adotar comportamentos de condução mais cautelosos, como respeitar limites de velocidade e manter uma condução defensiva. Esses comportamentos podem resultar em uma menor incidência de acidentes e, portanto, menor número de sinistros.

Escolha de veículos: Mulheres podem optar por modelos de carros que historicamente estão associados a menores taxas de sinistros. Se a base de clientes femininos tiver uma preferência por veículos que, estatisticamente, estão envolvidos em menos acidentes, isso pode contribuir para uma menor média de sinistros.

Hábitos de uso do veículo: Padrões de uso do veículo também podem desempenhar um papel. Se as mulheres, em média, percorrem distâncias mais curtas, utilizam menos o veículo para deslocamentos de longa distância ou conduzem em ambientes menos propensos a acidentes, isso pode influenciar a média de sinistros.

Diferenças na exposição ao risco: Pode haver diferenças nos ambientes de condução e exposição ao risco entre homens e mulheres. Por exemplo, se os clientes masculinos estiverem mais expostos a condições de tráfego congestionado ou a áreas urbanas com maior incidência de acidentes, isso pode contribuir para uma média de sinistros mais elevada.

Ao analisar o número absoluto de sinistros, percebemos que há muito mais sinistros associados aos clientes masculinos. Essa disparidade pode ser atribuída ao fato de que a base de clientes masculinos é substancialmente maior em comparação com a base de clientes femininos. Por região, ficou bem equilibrado.

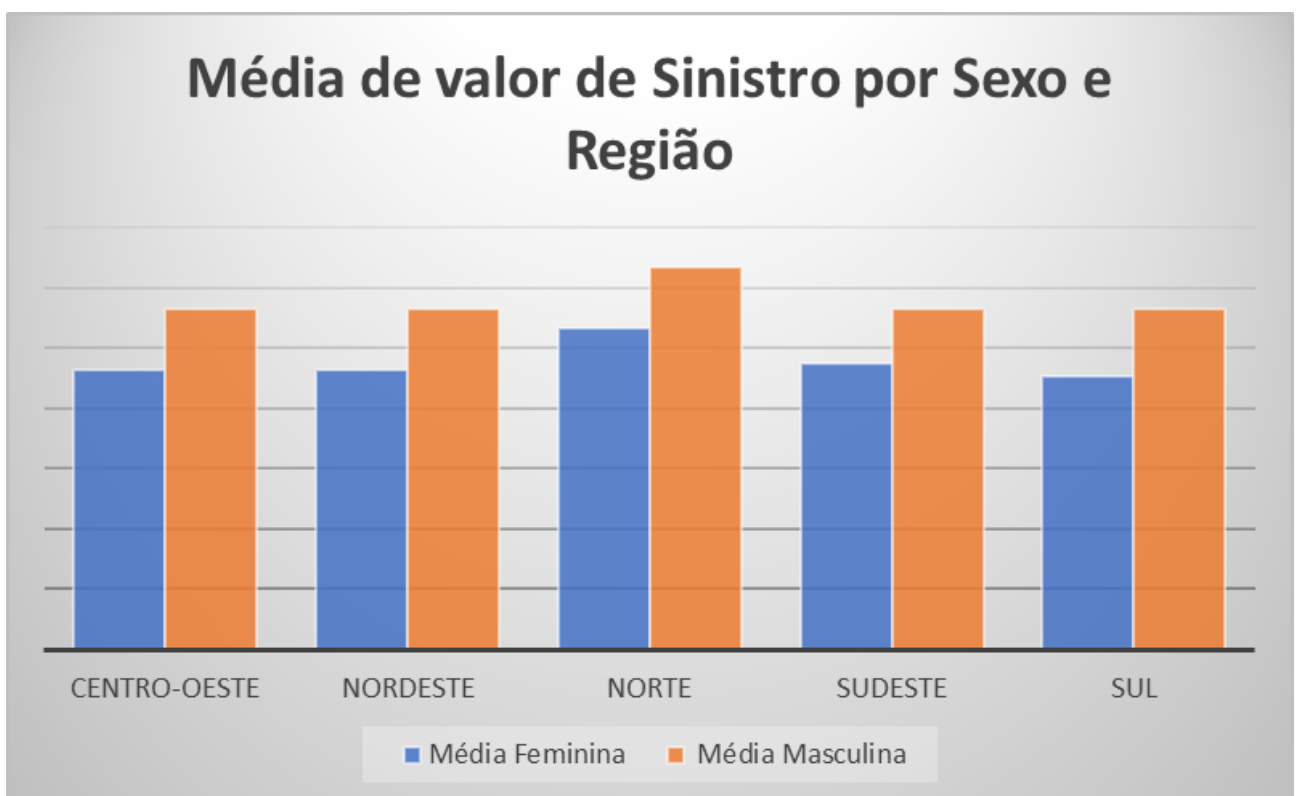


Figura 17 – Média de sinistro por sexo e região

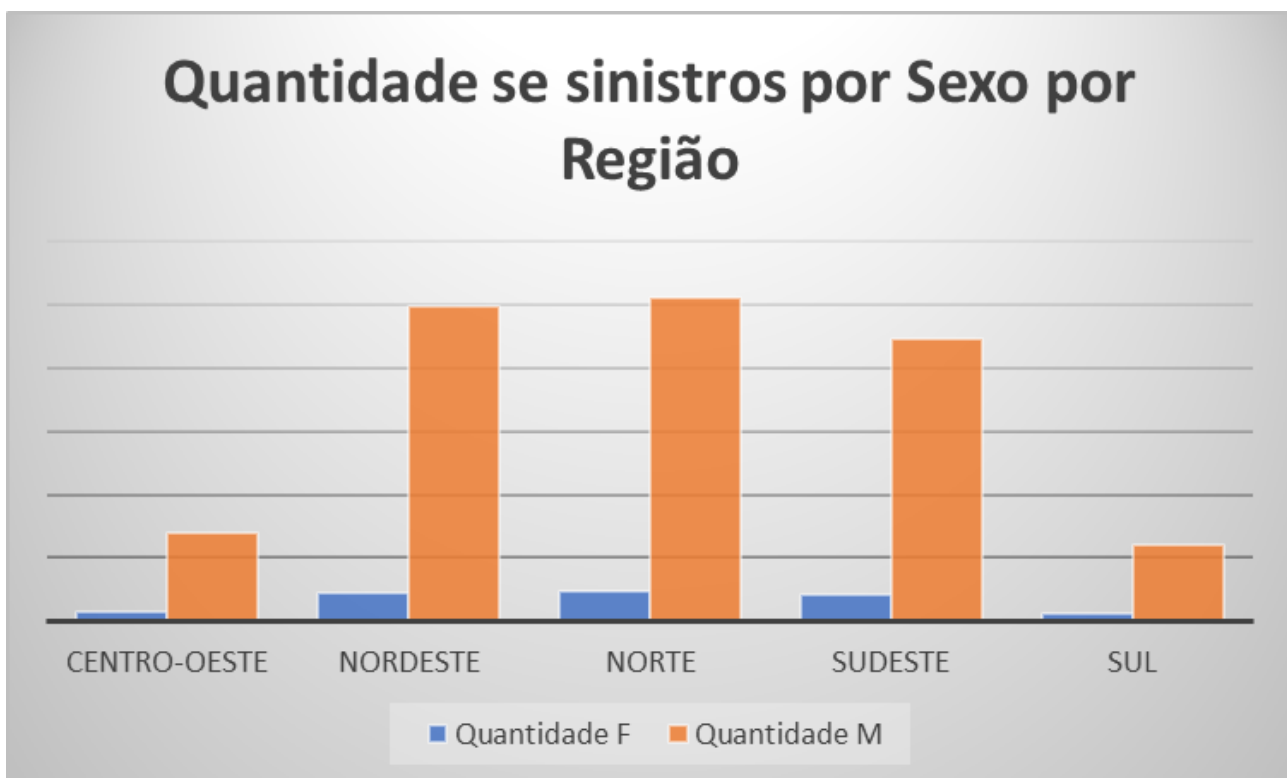


Figura 18 – Quantidade de sinistros por sexo e região

9. Power BI – Outra aplicação OLAP

O Power BI desempenhou um papel fundamental ao permitir análises complementares em conjunto com outras ferramentas, proporcionando uma abordagem mais abrangente na avaliação dos dados. Ao contrário das médias, que fornecem uma visão geral, o Power BI foi utilizado para analisar valores brutos totais, oferecendo uma compreensão detalhada e granular dos dados.

A capacidade do Power BI de integrar dados de diversas fontes e transformá-los em visualizações dinâmicas foi essencial. Ao analisar valores brutos totais, a equipe pôde identificar não apenas tendências, mas também padrões específicos e discrepâncias nos dados. Essa análise detalhada permitiu uma compreensão mais profunda das operações da seguradora. É relevante ressaltar que focamos em relatórios apenas pelo Excel, mas observamos que pelo Power BI, é possível a criação de relatórios bem completos.

Veja a seguir algumas tabelas e granularidades que importamos para nossa aplicação OLAP no Power BI:

Regiao	Estado	Cidade	Valor
centro-oeste	GO	anapolis	92,555,977.45
centro-oeste	GO	baliza	65,256,374.84
centro-oeste	GO	batiza	29,675,554.27
centro-oeste	GO	caldas novas	154,018,898.81
centro-oeste	GO	goiania	68,997,593.42
centro-oeste	GO	trindade	108,345,902.46
centro-oeste	MT	cuiaba	204,948,620.34
centro-oeste	MT	jangada	278,752,404.95
centro-oeste	MT	ponte branca	296,582,576.84
centro-oeste	MT	sorriso	465,731,171.71
centro-oeste	MT	vila rica	335,554,159.64
nordeste	BA	brejolandia	286,455,775.64
nordeste	BA	caravelas	305,720,576.76
nordeste	BA	ilheus	456,034,769.63
nordeste	BA	juazeiro	336,098,469.12
nordeste	BA	salvador	208,902,065.26
nordeste	PE	belo jardim	1,254,811,158.80
nordeste	PE	olinda	1,133,590,825.76
nordeste	PE	petrolina	1,056,191,529.26
nordeste	PE	recife	767,307,876.73
nordeste	PE	vertentes	1,740,987,379.21
norte	AM	barcelos	365,660,861.42
norte	AM	itamarati	377,135,157.40
norte	AM	manaus	262,710,843.12
norte	AM	parintins	422,521,839.63
norte	AM	tabatinga	578,897,145.93
norte	PA	belem	880,833,622.50
norte	PA	benevides	1,197,924,086.33
norte	PA	castanhal	1,274,266,695.98
norte	PA	marapanim	1,953,628,013.92
Total			26,957,781,050.55

Figura 19

Regiao	Valor
norte	8,725,962,922.28
nordeste	7,546,100,426.17
sudeste	6,778,813,850.11
centro-oeste	2,100,419,234.73
sul	1,806,484,617.26
Total	26,957,781,050.55

Figura 20

Estado	Valor
AM	2,006,925,847.50
BA	1,593,211,656.41
GO	518,850,301.25
MG	1,963,302,474.19
MT	1,581,568,933.48
PA	6,719,037,074.78
PE	5,952,888,769.76
RS	476,097,361.04
SC	1,330,387,256.22
SP	4,815,511,375.92
Total	26,957,781,050.55

Figura 21

Estado	Quantidade
PA	215014
PE	214242
SP	172485
MG	70453
AM	63919
BA	57328
MT	56931
SC	47849
GO	18613
RS	17221
Total	934055

Figura 22

Regiao	Quantidade
centro-oeste	75544
nordeste	271570
norte	278933
sudeste	242938
sul	65070
Total	934055

Figura 23

Cidade	Quantidade
anapolis	3349
baliza	2349
barcelos	11659
batiza	1068
belem	28214
belo horizonte	8907
belo jardim	44900
benevides	38324
blumenau	8623
brejolandia	10236
cachoeirinha	3203
caldas novas	5526
campinas	31309
canela	3126
caravelas	11113
castanhal	40753
colorado	5002
contagem	14823
cuiaba	7420
divinopolis	20463
florestal	12897
florianopolis	2180
florianopolis	4003
franca	36275
goiania	2440
gramado	3543
ilheus	16485
Total	934055

Figura 24

10. Data Mining

Infelizmente, enfrentamos desafios na realização das operações de Data Mining devido à complexidade da estrutura e aos tipos de dados presentes no conjunto de dados que escolhemos analisar baseando-se no sinistro. A tentativa foi feita utilizando a ferramenta Visual Studio com a extensão Analysis Services, no entanto, encontramos um obstáculo ao processar a estrutura de mineração.

Entire Solution						1 Error	7 Warnings	0 Messages	Build + IntelliSense	Search Error List
	Code	Description	Project	File	Line	Suppression State				
✖		MeasureGroup [Gradi].[Fato Sinistro] : Semi-additive measure requires a time dimension.			0					
⚠		Dimension [Dim Carro] : Create hierarchies in non-parent child dimensions.			0					
⚠		Dimension [Dim Modelo] : Create hierarchies in non-parent child dimensions.			0					
⚠		Dimension [Dim Cliente] : Create hierarchies in non-parent child dimensions.			0					
⚠		Database [Multidimensional-CUBE] : The database has no Time dimension. Consider creating one.			0					
⚠		Dimension [Dim Montadora] : Create hierarchies in non-parent child dimensions.			0					
⚠		RelationalDataSource [MATT-DatSource] : Do not use unsupported OLE DB providers in a DataSource.			0					
⚠		Dimension [Fato Sinistro] : Create hierarchies in non-parent child dimensions.			0					

Figura 27 - Erro encontrado

Para uma análise mais profunda e potencialmente mais eficaz, seria prudente reconsiderarmos a abordagem escolhida. Uma alternativa interessante poderia ser a exploração de técnicas de previsão e classificação, utilizando algoritmos bem estabelecidos na literatura, por meio de bibliotecas em Python. Entendemos que o grupo enfrentou dificuldades com as ferramentas utilizadas, como SQL Server, Visual Studio, entre outras, o que, de certa forma, impactou o progresso do trabalho e dificultou a execução ideal das tarefas. Contudo, como mencionado anteriormente, tentamos usar recursos disponíveis no Visual Studio.

Seria benéfico considerar uma abordagem mais acessível, como o uso de bibliotecas populares de Python para Data Mining, como scikit-learn ou TensorFlow. Essas ferramentas oferecem uma ampla gama de algoritmos de machine learning e são mais acessíveis para muitos usuários, proporcionando uma curva de aprendizado mais suave em comparação com as ferramentas anteriormente escolhidas.

11. Comparação

Ao conduzir análises através do Excel, notamos descobertas interessantes ao utilizar as funções avançadas desta ferramenta. Por meio dessas funcionalidades avançadas, conseguimos criar colunas derivadas dos valores brutos presentes no banco de dados. Isso significa que muitas análises puderam ser expandidas sem a necessidade de modificar diretamente o banco de dados.

Ao explorar as capacidades avançadas do Excel, pudemos realizar transformações e manipulações nos dados de maneira eficiente, proporcionando uma maior flexibilidade na geração de relatórios. A possibilidade de criar colunas adicionais com base nos valores existentes abriu novas perspectivas para análises mais aprofundadas sem a necessidade de alterar a estrutura original do banco de dados.

Em contrapartida, ao utilizarmos o Power BI, enfrentamos algumas dificuldades na geração desses relatórios. Isso se deve, em grande parte, à falta de familiaridade do grupo com esse software específico. No entanto, é importante ressaltar que o Power BI demonstrou ser uma ferramenta avançada e poderosa, oferecendo recursos robustos para análise de dados. A complexidade da ferramenta pode ter contribuído para os desafios iniciais, mas, conforme a equipe se familiarizou mais, ficou claro que o Power BI também oferece uma plataforma robusta para análises avançadas.

Em resumo, ambas as ferramentas, Excel e Power BI, desempenharam papéis importantes em nosso processo de análise. O Excel destacou-se pela sua familiaridade e capacidade de manipulação de dados, enquanto o Power BI, apesar das dificuldades iniciais, revelou-se como uma ferramenta avançada e poderosa, oferecendo recursos mais complexos para análises mais aprofundadas.

12. Análise do negócio

A análise detalhada dos sinistros na seguradora de veículos revela informações cruciais que orientam as decisões estratégicas da empresa. Ao examinar a distribuição geográfica dos sinistros, fica evidente uma uniformidade nas médias por região, mas uma disparidade notável na quantidade total de sinistros, concentrando-se principalmente nas regiões Norte, Nordeste e Sudeste do país. A análise por estado revela uma média significativamente mais alta no Norte, o que aponta para a necessidade de investigar os fatores que contribuem para essa discrepância específica.

Ao analisar os dados por modelo e montadora de veículos, surgem modelos como Camaro, Toro e S10 como principais causadores de sinistros elevados. Essa observação exige estratégias específicas para gerenciar esses tipos de veículos e minimizar riscos. Além disso, a diferenciação na média de sinistros entre homens e mulheres destaca nuances comportamentais e preferenciais, indicando a importância de políticas de seguro personalizadas para atender às diferentes demandas dos clientes.

A utilização de ferramentas como Excel e Power BI desempenhou um papel crucial nessa análise. O Excel se mostrou valioso para a manipulação detalhada de dados, permitindo uma investigação minuciosa das informações disponíveis. Por outro lado, o Power BI destacou-se na visualização e interpretação dos valores brutos totais, oferecendo uma compreensão mais dinâmica e interativa dos dados analisados.

Em resumo, os resultados obtidos indicam a necessidade de ajustes nas políticas de seguro, a implementação de medidas preventivas e estratégias específicas para modelos de veículos e perfis de clientes. Essas ações são essenciais para otimizar a gestão de sinistros, promovendo uma abordagem mais eficaz e personalizada. Ao adotar uma estratégia informada pelos dados, a seguradora estará melhor posicionada para atender às necessidades dos clientes, minimizar riscos e fortalecer sua posição no mercado de seguros de veículos.

13. Conclusão

No decorrer do projeto, procuramos aderir às orientações fornecidas pelo professor de CCF-424. Os resultados obtidos alinharam-se adequadamente com as expectativas estabelecidas pelas diretrizes, buscando uma abordagem prática e aplicada. A condução do trabalho foi pautada pela busca de uma compreensão mais aprofundada do tema, sendo abordado com criatividade e seriedade. Dessa maneira, buscamos obter uma visão abrangente do funcionamento, aplicação e execução do tema em questão.

14. Bibliografia

[SUSEP — SUSEP - Superintendência de Seguros Privados \(www.gov.br\)](http://www.gov.br)
[Dados Estatísticos — SUSEP - Superintendência de Seguros Privados \(www.gov.br\)](http://www.gov.br)
[Fraudes continuam a afetar o setor de seguros — Ministério da Economia \(www.gov.br\)](http://www.gov.br)