



Universidade Federal do Rio de Janeiro

Escola Politécnica

MBA em Big Data, Business Intelligence e Business Analytics
(MB3B)

**ESTUDO DE CORRELAÇÃO ENTRE VARIÁVEIS ECONÔMICAS E A
INDÚSTRIA DE VIDRO PLANO COM A UTILIZAÇÃO DO
SOFTWARE R**

Autor:

Gabriel Lima Souto do Vabo, B. Sc.

Orientador:

Manoel Villas Boas Junior, M. Sc.

Coorientador:

Roberto Ivo, D. Sc.

Examinador:

Flávio Luis de Mello, D. Sc.

Examinador:

José Airton Chaves Cavalcante Junior, D. Sc.

Examinador:

Vinicius Drumond Gonzaga, M. Sc.

**Rio de Janeiro
Dezembro 2021**

Declaração de Autoria e de Direitos

Eu, **Gabriel Lima Souto do Vabo**, CPF 160.723.717-28, autor da monografia *ESTUDO DE CORRELAÇÃO ENTRE VARIÁVEIS ECONÔMICAS E A INDÚSTRIA DE VIDRO PLANO COM A UTILIZAÇÃO DO SOFTWARE R*, subscrevo para os devidos fins, as seguintes informações:

1. O autor declara que o trabalho apresentado na defesa da monografia do curso de Pós-Graduação, Especialização MBA em Big Data, Business Intelligence e Business Analytics da Escola Politécnica da UFRJ é de sua autoria, sendo original em forma e conteúdo.
2. Excetuam-se do item 1 eventuais transcrições de texto, figuras, tabelas, conceitos e ideias, que identifiquem claramente a fonte original, explicitando as autorizações obtidas dos respectivos proprietários, quando necessárias.
3. O autor permite que a UFRJ, por um prazo indeterminado, efetue em qualquer mídia de divulgação, a publicação do trabalho acadêmico em sua totalidade, ou em parte. Essa autorização não envolve ônus de qualquer natureza à UFRJ, ou aos seus representantes.
4. O autor declara, ainda, ter a capacidade jurídica para a prática do presente ato, assim como ter conhecimento do teor da presente Declaração, estando ciente das sanções e punições legais, no que tange a cópia parcial, ou total, de obra intelectual, o que se configura como violação do direito autoral previsto no Código Penal Brasileiro no art.184 e art.299, bem como na Lei 9.610.
5. O autor é o único responsável pelo conteúdo apresentado nos trabalhos acadêmicos publicados, não cabendo à UFRJ, aos seus representantes, ou ao(s) orientador(es), qualquer responsabilização/ indenização nesse sentido.
6. Por ser verdade, firmo a presente declaração.

Rio de Janeiro, 04 de dezembro de 2021.

Gabriel Lima Souto do Vabo

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Av. Athos da Silveira, 149 - Centro de Tecnologia, Bloco H, sala - 212,
Cidade Universitária Rio de Janeiro – RJ - CEP 21949-900.

Este exemplar é de propriedade Escola Politécnica da Universidade Federal do Rio de Janeiro, que poderá incluí-lo em base de dados, armazenar em computador, microfilmear ou adotar qualquer forma de arquivamento.

Permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es).

DEDICATÓRIA

Dedico esse trabalho primeiramente a minha família, que sempre me apoiou em busca de todos os meus objetivos e me permitiu chegar até onde cheguei hoje, principalmente minha mãe. Foi a pessoa que sempre me incentivou, me apoiou e me proporcionou tudo o que tenho até hoje. Esse trabalho é fruto de um trabalho em conjunto de tudo o que me foi proporcionado por ela ao longo da minha trajetória.

Aproveito para dedicar esse trabalho a meu tio Claudio Magnanini (in memoriam), Engenheiro Eletricista formado pela Universidade Federal Fluminense, com especialização em Engenharia de Produção pela UFRJ e Mestrado em Arquitetura e Urbanismo pela Universidade Federal Fluminense. Foi uma das pessoas que me incentivou a escolher minha profissão e teve um papel importante na minha vida não apenas como inspiração de carreira, mas também exercendo o papel de pai por muitas vezes. Fica aqui o meu eterno agradecimento a tudo o que ele fez e contribuiu para a minha trajetória.

Também gostaria de agradecer a UFRJ, que me proporcionou a oportunidade de adquirir novos conhecimentos a respeito de um tema que está em constante crescimento e desenvolvimento, sempre com inovações que permitem que a sociedade se desenvolva de forma mais bem orientada e utilizando a razão baseada em fatos comprovados por meio de estudos.

Por fim, gostaria de agradecer a empresa que trabalho, a qual me despertou o interesse em estudar cada vez mais sobre seu mercado e me permitir realizar esse curso. Nele pude aprender onde e como conseguirei levar diversos novos conhecimentos para aplicar em projetos que trarão benefícios para empresa não apenas atualmente, mas também visando soluções futuras e a longo prazo.

RESUMO

O trabalho tem como objetivo obter uma análise detalhada do índice de correlação entre algumas variáveis econômicas (IPCA, IGP-M, SELIC, IBOVESPA, INCC, ICMS, IAIC) e a produção de vidro plano no Brasil, buscando direcionar o planejamento estratégico do setor para os próximos anos de recuperação frente ao período pós pandemia do novo coronavírus. Inicialmente o texto apresenta conceitualmente o que são os índices analisados e faz uma breve introdução ao leitor sobre o mercado vidreiro nacional. Apresenta-se também uma explicação sobre o processo de extração, tratamento e manipulação de dados, além do software que foi utilizado e suas principais características, evidenciando, portanto, as razões de sua escolha frente a outras possíveis soluções. Dando continuidade, o projeto demonstra todo o processo de trabalho com os dados dentro do programa, desde os locais de onde foram extraídos até como foram armazenados, tratados e manipulados para trazer os resultados esperados. O projeto demonstra de forma detalhada todos os cálculos realizados pelo programa a fim de evidenciar ao leitor como funciona conceitualmente o que está sendo apresentado nas telas do programa. Assim, é possível analisar com mais clareza os passos de desenvolvimento do estudo das correlações. Por fim, o trabalho traz como foi possível obter diferentes níveis de correlações entre as variáveis selecionadas, o que possibilita, por exemplo, que a empresa direcione o plano de produção, ações de marketing e até mesmo as vendas conforme os resultados. Isso também abre possibilidades para que o estudo seja utilizado para outros tipos de correlação que possam fazer sentido para a empresa.

Palavras-Chave: Pearson, Estatística, Vidro, Economia, Índices

ABSTRACT

This study was made to make a detailed analysis of the correlation coefficient between some economic indices (IPCA, IGP-M, SELIC, IBOVESPA, INCC, ICMS) and the float glass production in Brazil. As a result of this correlation, is expected to give some crucial information that can help glass industries in its strategic planning during the years after the COVID-19 pandemic. At first, the study shows the most important concepts about the indices, presents a quick review about the glass market in Brazil and about the production process. Because it is a study that involves data analysis, some explanation about data preparation is exposed to make the reader familiar about what will be made, together with de reasons why a specific software was defined to be used, instead of others. Continuing the process, the project demonstrates all the data work and preparation outside and inside the software, since the data extraction from websites to the data use at graphs and detailed analysis. The study makes a point of highlighting every single step made at the software to show clearly what is being doing not only by the machine, but also at the statistical analysis. To finish the work, it was possible to define the correlation coefficient between the indices and the glass production during a determined range of years. This type of analysis, as said before, can help industries to make their demand planning and strategic orientation for the years that will come and also can be used as a base study to many other correlation analyses that the company wants to do.

Keywords: Pearson, Statistics, Economics, Glass, Indices

SIGLAS

ANFAVEA	Associação Nacional de Fabricantes de Veículos Automotores
BACEN	Banco Central
BC	Banco Central
CBIC	Câmara Brasileira da Indústria da Construção
CNI	Confederação Nacional da Indústria
COPOM	Comitê de Política Monetária
FGV	Fundação Getúlio Vargas
IBGE	Instituto Brasileiro de Geografia e Estatística
IMECC	Instituto de Matemática, Estatística e Computação Científica
IMOB	Índice Imobiliários
INCC	Índice Nacional de Construção Civil
IPCA	Índice de Preços ao Consumidor Amplo
IPEA	Instituto de Pesquisa Econômica Aplicada
PIB	Produto Interno Bruto
IGP-M	Índice Geral de Preços do Mercado
SELIC	Sistema Especial de Liquidação e Custódia
IBOVESPA	Índice da Bolsa de Valores do Estado de São Paulo
IAIC	Índice de Atividade da Indústria da Construção
PVB	Polivinil Butiral
SNIPC	Sistema Nacional de Índices de Preços ao Consumidor
UNICAMP	Universidade Estadual de Campinas

LISTA DE FIGURAS

Figura 2.1	Gráfico 2.1: Frete x Distância (Autor)	14
Figura 4.1	Tela do Programa R – Diretório	25
Figura 4.2	Tela do Programa R – Arquivos importados	25
Figura 4.3	Tela do Programa R – Cabeçalhos das bases	26
Figura 4.4	Tela do Programa R – Últimas linhas das bases	27
Figura 4.5	Tela do Programa R – Remoção 12 primeiros meses da Base BOVESPA	28
Figura 4.6	Tela do Programa R – Base BOVESPA ajustada	28
Figura 4.7	Tela do Programa R - Remoção 12 primeiros meses da Base PIB	29
Figura 4.8	Tela do Programa R - Remoção 336 primeiros meses da Base SELIC	29
Figura 4.9	Tela do Programa R – Removendo 2 últimos meses da Base BOVESPA	30
Figura 4.10	Tela do Programa R – Removendo último mês da base IGPM	30
Figura 4.11	Tela do Programa R – Removendo último mês da base INCC	31
Figura 4.12	Tela do Programa R – Removendo último mês da base IPCA	31
Figura 4.13	Tela do Programa R – Removendo último mês da base PIB	32
Figura 4.14	Tela do Programa R – Removendo 2 últimos meses da base SELIC	32
Figura 4.15	Tela do Programa R – Data Frame Único	33
Figura 4.16	Tela do Programa R – Exclusão colunas “Mês” repetidas	33
Figura 4.17	Tela do Programa R – Ajuste nome coluna “Mês”	34
Figura 4.18	Tela do Programa R – Identificação da formatação das colunas	34
Figura 4.19	Tela do Programa R – Conferência de elementos vazios	34
Figura 4.20	Tela do Programa R – Conferência de elementos vazios por fórmula	35

Figura 4.21	Tela do Programa R – Instalação do pacote “CORPLOT”	35
Figura 4.22	Tela do Programa R – Gráfico Matriz de Correlação	36
Figura 4.23	Tela do Programa R – Tabela Matriz de Correlação	36
Figura 4.24	Tela do Programa R – Código para gráfico BOVESPA x Produção de Vidro	37
Figura 4.25	Tela do Programa R – Gráfico de Dispersão BOVESPA x Produção de Vidro	37
Figura 4.26	Tela do Programa R – Código para correlação BOVESPA x Produção de Vidro	38
Figura 4.27	Tela do Programa R – Código para gráfico IGPM x Produção de Vidro	38
Figura 4.28	Tela do Programa R – Gráfico de Dispersão IGPM x Produção de Vidro	38
Figura 4.29	Tela do Programa R – Código para correlação IGPM x Produção de Vidro	39
Figura 4.30	Tela do Programa R – Código para gráfico INCC x Produção de Vidro	39
Figura 4.31	Tela do Programa R – Gráfico de Dispersão INCC x Produção de Vidro	39
Figura 4.32	Tela do Programa R – Código para correlação INCC x Produção de Vidro	40
Figura 4.33	Tela do Programa R – Código para gráfico IPCA x Produção de Vidro	40
Figura 4.34	Tela do Programa R – Gráfico de Dispersão IPCA x Produção de Vidro	40
Figura 4.35	Tela do Programa R – Código para correlação IPCA x Produção de Vidro	41
Figura 4.36	Tela do Programa R – Código para gráfico PIB x Produção de Vidro	41
Figura 4.37	Tela do Programa R – Gráfico de Dispersão PIB x Produção de Vidro	41
Figura 4.38	Tela do Programa R – Código para correlação PIB x Produção de Vidro	42
Figura 4.39	Tela do Programa R – Código para gráfico SELIC x Produção de Vidro	42
Figura 4.40	Tela do Programa R – Gráfico de Dispersão SELIC x Produção de Vidro	42
Figura 4.41	Tela do Programa R – Código para correlação SELIC x Produção de Vidro	43

Figura 4.42	Tela do Programa R – Data Frame Produção de Veículos	44
Figura 4.43	Tela do Programa R – Ajuste coluna mês em df_carros	44
Figura 4.44	Tela do Programa R – União df_carros com o Data Frame principal	44
Figura 4.45	Tela do Programa R – Data Frame Índice de Atividade da Indústria de Construção	45
Figura 4.46	Tela do Programa R – Ajuste nos meses do Data Frame Original	45
Figura 4.47	Tela do Programa R – União do df_construcao com o Data Frame principal	45
Figura 4.48	Tela do Programa R – Ajuste coluna “Mês” no Data Frame Final	46
Figura 4.49	Tela do Programa R – Gráfico Produção Mensal de Vidro	46
Figura 4.50	Tela do Programa R – Gráfico Produção Mensal de Veículos	46
Figura 4.51	Tela do Programa R – Gráfico Índice de Atividade da Indústria de Construção	47
Figura 4.52	Tela do Programa R – Exclusão da coluna “Mês”	47
Figura 4.53	Tela do Programa R – Base para correlação ajustada	47
Figura 4.54	Tela do Programa R – Gráfico Matriz de Correlação Final	48
Figura 4.55	Tela do Programa R – Tabela Matriz de Correlação	48
Figura 4.56	Tela do Programa R – Gerar Gráfico de Produção Veículos x Produção de Vidro	49
Figura 4.57	Tela do Programa R – Gráfico de Produção de Veículos x Produção de Vidro	49
Figura 4.58	Tela do Programa R – Correlação Produção de Veículos x Produção de Vidro	49
Figura 4.59	Tela do Programa R – Gerar Gráfico de IAIC x Produção de Vidro	50
Figura 4.60	Tela do Programa R – Gráfico de IAIC x Produção de Vidro	50
Figura 4.61	Tela do Programa R – Correlação IAIC x Produção de Vidro	50

LISTA DE QUADROS

Quadro 2.1	Distância x Frete	14
Quadro 2.2	Interpretação de Correlações	18
Quadro 2.3	Interpretación Correlación	18
Quadro 2.4	Strenght of Association	18
Quadro 4.1	Resultados do estudo de correlação	43

Sumário

Capítulo 1: Introdução	1
1.1 – Tema.....	1
1.2 - Justificativa	1
1.3 – Objetivos	1
1.4 - Delimitação.....	2
1.5 – Metodologia	2
1.6 – Descrição.....	2
Capítulo 2: Embasamento Teórico.....	4
2.1 – Índices Econômicos no Brasil.....	4
2.1.1 – SELIC/Taxa de Juros	4
2.1.2 – IPCA.....	5
2.1.3 – ICMS	6
2.1.4 – PIB	6
2.1.5 – IGP-M	7
2.1.6 – INCC	8
2.1.7 – IBOVESPA	8
2.1.8 – IAIC	8
2.2 – A Indústria do Vidro Plano	9
2.3 – Software R.....	11
2.4 – Conceitos Estatísticos.....	12
2.4.1 – Tipos de Variáveis	12
2.4.2 – Séries Estatísticas	13
2.4.3 – Gráfico de Dispersão.....	13
2.4.4 – Covariância	15
2.4.5 – Correlação	16
2.5 – Análise de Dados.....	19
Capítulo 3: Metodologia	21
Capítulo 4: Estudo de Caso	24
4.1 – Análise dos Dados.....	24
4.1.1 – Definição das Bases de Dados Utilizadas	24
4.1.2 – Análise dos Dados Importados.....	25
4.1.3 – Ajustando a Base de Dados.....	28
4.1.4 – Análise dos Dados.....	35
4.1.5 – Análise dos Resultados	37
4.1.5.1 – BOVESPA x Produção de Vidros.....	37
4.1.5.2 – IGP-M x Produção de Vidros.....	38
4.1.5.3 – INCC x Produção de Vidros	39
4.1.5.4 – IPCA x Produção de Vidros.....	40
4.1.5.5 – PIB x Produção de Vidros.....	41
4.1.5.6 – SELIC x Produção de Vidros.....	42

4.1.6 – Análise de mercados clientes	44
4.1.6.1 – Produção de Veículos x Produção de Vidros.....	49
4.1.6.2 – Índice de Atividade da Indústria de Construção x Produção de Vidros	50
Capítulo 5: Conclusão e Trabalhos Futuros.....	51
5.1 – Conclusão.....	51
5.2 – Trabalhos Futuros.....	52
Referências Bibliográficas	53

CAPÍTULO 1

Introdução

1.1 Tema

O trabalho tem como tema principal um estudo estatístico de correlação simples entre determinados índices econômicos presentes no Brasil e a variação da produção de vidro plano no país, utilizando-se de uma ferramenta de mercado chamada “R”.

1.2 Justificativa

Atualmente, as empresas se baseiam em previsões de produção principalmente de acordo com demandas históricas. Com a variação constante de índices econômicos no Brasil, as empresas produtoras de vidro podem se utilizar de estudos que possibilitem identificar quais outros fatores possuem influência direta em seus clientes. Dessa forma, é possível fazer investimentos direcionados e adequar planos de produção para que elas sigam planos baseados em dados confiáveis e que gerem informações estratégicas, derivados de análises realizadas em um software que permita esses tratamentos de forma rápida e correta.

1.3 Objetivo

O principal objetivo desse trabalho é fornecer um estudo capaz de evidenciar as correlações existentes entre índices e taxas brasileiras e a produção de vidro plano no Brasil de forma assertiva. Com isso, espera-se que seja possível comprovar determinadas situações que possibilitem as empresas guiarem seus planejamentos estratégicos não apenas baseando-se nas já conhecidas previsões de demanda, que levam em consideração as médias históricas, mas também novas variáveis que também possuem influência direta em seu ramo de atuação.

1.4 Delimitação

Esse estudo delimita-se apenas as variáveis econômicas IPCA, IGP-M, SELIC, IBOVESPA, INCC, ICMS que afetam diretamente o poder de compra do consumidor de forma geral, impactando, provavelmente, na produção de vidro plano, e no PIB, que identifica tudo o que é produzido no país, inclusive o vidro plano, e, portanto, deve possuir algum tipo de relação. Além desses indicadores, o estudo abre a possibilidade de utilizar mais dois indicadores, que são o IAIC e Produção de Veículos no Brasil.

Também foi definido que o estudo se restringe ao segmento de Vidros Planos, uma vez que movimenta uma parcela grande da economia e possui relação direta com vários setores como construção civil, indústria automotiva, indústria moveleira e de eletrodomésticos.

Por fim, foram utilizadas apenas aplicações do software “R” devido a sua conhecida confiabilidade em trabalhar com dados e fornecer informações relevantes em questões relacionadas à estudos estatísticos.

1.5 Metodologia

Para realização desse trabalho, foi utilizada a metodologia da pesquisa bibliográfica somada a um estudo de caso prático. Na pesquisa bibliográfica, buscou-se apresentar os principais conceitos abordados no desenvolvimento do estudo, bem como uma apresentação das ferramentas que foram utilizadas para as análises das correlações.

No estudo de caso, apresenta-se todo o processo desde a seleção dos locais de onde os dados serão extraídos, até a extração, tratamento e utilização deles para a realização das análises mais detalhadas.

1.6 Descrição

O trabalho está organizado em 5 capítulos, onde é possível acompanhar detalhadamente tudo o que foi realizado de pesquisas e análises.

- 1- No primeiro capítulo é apresentado um resumo geral de todos os passos que foram seguidos no trabalho.

- 2- Já no segundo capítulo são apresentados os conceitos e definições pesquisados na bibliografia disponível atualmente, para que o leitor possa conhecer melhor os temas que serão apresentados.
- 3- No capítulo 3 é demonstrada a proposta de solução do projeto, explicando com detalhes o que será feito e como será estruturado o estudo.
- 4- No capítulo 4 é apresentado o estudo de caso, com todo o processo de extração dos dados, utilizando a ferramenta selecionada até o tratamento final e as análises, detalhando todas as variáveis e suas respectivas correlações.
- 5- Por fim, o capítulo 5 apresenta a conclusão do trabalho, demonstrando as aplicações dos resultados obtidos tanto no capítulo 3, quanto nas análises do capítulo 4.

CAPÍTULO 2

Embasamento Teórico

2.1 - Índices Econômicos utilizados

Os índices econômicos selecionados para o estudo foram os principais índices presentes na economia brasileira, variando desde indicadores relativos a crescimento econômico, até taxas controladas por órgãos públicos no país. Foram selecionados a taxa SELIC (Sistema Especial de Liquidação e de Custódia), o IPCA (Índice de Preços ao Consumidor Amplo), ICMS (Imposto sobre Circulação de Mercadorias e Serviços), PIB (Produto Interno Bruto), IGP-M (Índice Geral de Preços do Mercado), INCC (Índice Nacional da Construção Civil), IBOVESPA (Índice da Bolsa de Valores do Estado de São Paulo) e IAIC (Índice de Atividade da Indústria da Construção), além da produção de vidro plano e de veículos.

2.1.1 - SELIC/Taxa de Juros

A SELIC é a taxa básica de juros presente na economia brasileira. Está diretamente relacionada com todas as demais taxas de juros aplicadas dentro do país e é definida pelo Banco Central com o objetivo principal de controlar a inflação ao longo do tempo.

Segundo o Banco Central do Brasil (2021): “A taxa Selic refere-se à taxa de juros apurada nas operações de empréstimos de um dia entre as instituições financeiras que utilizam títulos públicos federais como garantia.”

Aprofundando-se inicialmente no setor de construção civil, há uma característica de o mesmo se beneficiar com a baixa dessa taxa, pois o financiamento para compra de terrenos visando empreendimentos fica mais atrativa. Por consequência, surgem mais obras e o setor vidreiro pode se beneficiar, uma vez que é o seu principal mercado cliente.

A baixa na taxa de juros também facilita os financiamentos para compra de veículos automotivos em concessionárias, o que provavelmente aumenta a demanda por vidro verde (utilizado na montagem de automóveis) e a necessidade de produção do mesmo.

Por fim, uma redução da SELIC também induz bancos a subirem suas ofertas de crédito e, conseqüentemente, as pessoas passam a ter poder aquisitivo para compra de aparelhos eletrodomésticos e móveis, que costumam ser financiados a longo prazo e englobam parte dos clientes da indústria do vidro, em teoria aumentando também sua demanda.

Segundo o superintendente de investimento da Infinity Acess, Camilo Cavalcanti Jr (2020):

O crédito mais fácil também pode impactar positivamente a construção civil, uma vez que tanto a compra dos terrenos como a dos imóveis costumam ser financiadas. O setor imobiliário depende muito dessas taxas de longo prazo e com ela mais barata a margem de lucro fica maior. Por isso, as ações do setor podem subir rápido, em caso de corte de juros. No ano passado, quando a Selic saiu de 6,5% para 4,5%, as construtoras passaram por forte valorização na Bolsa.

2.1.2 - IPCA

O Índice Nacional de Preços ao Consumidor Amplo é o índice oficial de inflação do país, funcionando como parâmetro para o sistema de metas de inflação, calculado pelo IBGE. A sua alta indica aumento direto do preço dos produtos consumidos pela população brasileira, o que pode acabar levando a diminuição de gastos em certos setores e postergação de possíveis investimentos para priorizar a manutenção de itens básicos.

É um índice que mede, portanto, a variação dos preços de produtos e serviços no Brasil, influenciando de forma direta o ajuste da taxa SELIC pelo Banco Central. A forma de controle feita pelo banco consiste em analisar possíveis sinais de aumento da inflação por meio da variação do IPCA e aumentar a taxa de juros (SELIC). Essa mudança tende a reduzir as demandas de produtos e serviços e, por consequência, diminuir os preços praticados no país.

Sua fórmula de cálculo é uma média ponderada entre alguns setores como alimentos e bebidas, transportes, habitação, vestuário, despesas pessoais, entre outros. O IBGE faz o cálculo baseado na coleta de preços (entre o dia 1 e o dia 30 de cada mês) em comércios, residências, concessionárias e divulga mensalmente.

Segundo o site oficial do IBGE (2021):

Esse índice de preços tem como unidade de coleta estabelecimentos comerciais e de prestação de serviços, concessionária de serviços públicos e internet e sua coleta estende-se, em geral, do dia 01 a 30 do mês de referência. Atualmente, a população-objetivo do IPCA abrange as famílias com rendimentos de 1 a 40 salários-mínimos, qualquer que seja a fonte, residentes nas áreas urbanas das regiões de abrangência do SNIPC, as quais são: regiões metropolitanas de Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Vitória, Rio de Janeiro, São Paulo, Curitiba, Porto Alegre, além do Distrito Federal e dos municípios de Goiânia, Campo Grande, Rio Branco, São Luís e Aracaju.

2.1.3 - ICMS

O Imposto sobre Operações relativas à Circulação de Mercadorias é um tributo estadual que está diretamente ligado ao poder de compra do consumidor. Mais conhecido principalmente pela sua relação com a disparidade do preço dos combustíveis entre estados próximos no Brasil, ele afeta também outros setores como automotivo, restaurantes, transportes, moveleiro e construção civil. A alta nesse imposto acaba elevando o preço de produtos e serviços e reduzindo o poder de compra da população.

É um imposto que incide em diversas operações financeiras, desde a venda de produtos e mercadorias como bebidas e alimentos, prestação de serviços de transportes até mesmo na importação de serviços e mercadorias advindas do exterior. Acaba sendo um imposto cobrado de forma indireta no custo do produto vendido ou serviço prestado, sendo diferente de estado para estado e uma responsabilidade de cada governo determinar a porcentagem a ser praticada dentro de seu território.

Por se tratar de um imposto direto na comercialização de produtos e serviços, pode impactar diretamente nas vendas do mercado do vidro plano, tendo efeito principalmente nas variações de preços não apenas nos produtos, mas também nos fretes praticados para transportar essas mercadorias, de alto valor agregado.

Segundo o portal da Secretaria da Fazenda do Estado de São Paulo (2021), “O imposto é não-cumulativo, compensando-se o que for devido em cada operação relativa à circulação de mercadorias ou prestação de serviços de transporte interestadual e intermunicipal e de comunicação com o montante cobrado nas anteriores pelo mesmo ou por outro Estado.”

2.1.4 - PIB

O Produto Interno Bruto é a soma de tudo o que é produzido dentro do Brasil ao longo do ano. Há de se esperar que, com o aumento do PIB esteja diretamente relacionado com o aumento da produção na indústria de vidro plano, porém é necessária uma análise mais detalhada ao longo dos anos para mostrar se essa relação de fato é verdadeira ou não.

É importante ressaltar que, para seu cálculo, são considerados apenas bens e serviços finais, ou seja, não são incluídos no cálculo os bens de consumo intermediário, como por exemplo insumos que empresas comprem de outras para elaborar seus produtos finais.

O PIB é calculado pelo IBGE, que o define da seguinte forma em seu site oficial:

O PIB mede apenas os bens e serviços finais para evitar dupla contagem. Se um país produz R\$ 100 de trigo, R\$ 200 de farinha de trigo e R\$ 300 de pão, por exemplo, seu PIB será de R\$ 300, pois os valores da farinha e do trigo já estão embutidos no valor do pão. Os bens e serviços finais que compõem o PIB são medidos no preço em que chegam ao consumidor. Dessa forma, levam em consideração também os impostos sobre os produtos comercializados. O PIB não é o total da riqueza existente em um país. Esse é um equívoco muito comum, pois dá a sensação de que o PIB seria um estoque de valor que existe na economia, como uma espécie de tesouro nacional. Na realidade, o PIB é um indicador de fluxo de novos bens e serviços finais produzidos durante um período. Se um país não produzir nada em um ano, o seu PIB será nulo.

2.1.5 - IGP-M

O Índice Geral de Preços do Mercado é um dos indicadores que evidencia a variação de preços dentro da economia do país, ou seja, é uma métrica utilizada por indústrias para ajustar seus preços e também por investidores, para tomarem suas decisões de alocação de recursos, pois inclui desde produtos agrícolas e industriais (primários), produtos finais e até mesmo serviços.

É um índice que possui influência direta da variação cambial, ou seja, durante o período de pandemia, onde o câmbio teve uma variação considerável, esse índice acabou também sofrendo variações significativas. Com a desvalorização cambial no Brasil, preços de produtos que são balizados pelo dólar, como as commodities agrícolas, ferro, cobre, entre outros, tiveram seus preços elevados.

É um índice comumente utilizado também para reajustes de preços em diversas situações como tarifas de energia elétrica, aluguéis, dentre outros serviços considerados essenciais à população, sendo conhecido como “a inflação do aluguel” para inquilinos e proprietários que possuem contratos de locação.

Por se tratar de um índice que engloba em seu cálculo a união de 3 outros índices relevantes, com pesos diferentes, acabam provavelmente tendo impacto e influência no setor vidreiro. Para que seja calculado, é feita uma média ponderada de 60 % do IPA-M (Índice de Preços por Atacado), que monitora as variações do atacado, 30% do IPC-M (Índice de Preços ao Consumidor), que monitora os setores como alimentação e saúde e 10% do INCC (Índice Nacional de Custo de Construção), que monitora os custos de construções no país, ou seja, materiais e mão de obra, e será analisado na próxima seção.

Sua principal diferença em relação ao IPCA é que o IPCA considera apenas a variação de preços diretamente relacionada ao consumidor final. Já o IGP-M leva em conta uma variação de preços em todos os estágios da produção.

2.1.6 - INCC

Conforme já comentado anteriormente, o Índice Nacional de Custo de Construção é um índice que possui uma participação mais baixa no cálculo do IGP-M, porém possui uma relação direta com o mercado de construção civil, um dos principais clientes das indústrias de vidro plano. O índice é obtido por meio de orçamentos de empresas de engenharia civil e calculado considerando os valores de serviços, mão-de-obra e equipamentos, dividindo-se nos grupos de estruturais, instalações e acabamentos.

Dessa forma, o índice é utilizado por construtoras para evitar variações acima do esperado em suas matérias-primas durante as obras, que costumam durar alguns anos e estão sujeitas a variações de preço em seus custos iniciais.

2.1.7 - IBOVESPA

O Ibovespa é o principal índice da Bolsa de Valores do Brasil, que evidencia o desempenho das ações mais negociadas das empresas brasileiras com capital aberto na B3. É um cálculo reavaliado a cada 4 meses de acordo com o volume negociado de diversos setores como bancos, redes varejistas e produtoras de commodities.

É um índice antigo, criado ainda na década de 60, e considera empresas famosas como Petrobras e Vale do Rio Doce em sua base de cálculo, refletindo aproximadamente 80% do volume de ações negociadas durante um pregão. Por isso, é considerado entre os investidores e acionistas do mercado financeiro como um “termômetro” do mercado ações no Brasil, indicando se as empresas do país estão se mostrando atrativas ou não para investidores, tanto internos quanto externos.

Dessa forma, sua variação demonstra de forma mais clara o interesse tanto estrangeiro quanto nacional nas empresas do país, se tornando um indicador claro do quanto o país e suas empresas estão fortes economicamente perante o mercado como um todo.

2.1.8 - IAIC

O Índice de Atividade da Indústria da Construção, divulgado pela CNI, é um índice que evidencia a quantidade de obras sendo realizadas no país mensalmente. Esse indicador é capaz de demonstrar não apenas como está o mercado de construção civil em termos de obras, mas também é um termômetro do nível de empregabilidade do setor ao longo do tempo.

Os indicadores desse índice variam entre 0 e 100 pontos e, quando estão abaixo dos 50 pontos, indicam queda tanto nas atividades quanto do número de empregos na área. Segundo o site do Jornal “Estadão” (2021): “A Sondagem Indústria da Construção foi feita entre 3 e 12 de fevereiro, quando 478 empresas do setor foram ouvidas. A pesquisa é feita em parceria com a Câmara Brasileira da Indústria da Construção (CBIC).”

2.2 - A indústria do vidro plano

A fabricação de vidro data de mais de 7 mil anos atrás, com uma descoberta do povo fenício ao fazer uma fogueira na praia e perceber que a areia derretida produzia um material transparente e sólido. O vidro plano, por sua vez, veio apenas nos anos 500 d.C., por meio de sopro e rotação de uma esfera, ampliando-a ao máximo dentro de um forno.

Atualmente a indústria de vidro é separada em 4 segmentos, que variam conforme o produto final fabricado, são eles:

- **Embalagens de vidro:** indústrias de alimentos e farmacêutica (potes, garrafas, frascos etc.)
- **Vidro plano:** utilizado na construção civil, no setor automobilístico, indústria moveleira e linha branca (eletrodomésticos).
- **Vidro doméstico:** produzido para itens de decoração em geral (copos, vasos, objetos, xícaras etc.)
- **Vidros técnicos:** produzidos para indústrias específicas como laboratórios, televisores, celulares, lâmpadas etc.

Nesse trabalho, o foco será no estudo do ramo de vidro plano, que se trata do mercado que mais movimenta a economia entre os 4 citados e geralmente é utilizado como base para outros tipos de vidro como temperado, laminado, espelhos e pintados.

O processo de produção do vidro plano é complexo, e possui custos muito elevados, o que faz com que o produto final tenha um alto valor agregado. São diversos tipos de vidro plano

a serem produzidos, o que leva a ajustes quase que semanais nas linhas de produção, alterando desde a grossura do vidro até seus componentes e cores, sempre objetivando atender o que o mercado está demandando.

O mercado de construção civil tem diversas aplicações para o vidro plano, sendo as principais delas para construção de janelas e fachadas de casa e prédios, tanto residenciais quanto comerciais. Sua utilização permite conforto térmico e acústico por meio de tecnologias que vêm sendo desenvolvidas e capazes de se adaptar a diversos ambientes, como proteção solar, autolimpeza, controle de temperatura (e, conseqüentemente, eficiência energética) e até mesmo antibacteriano (por meio de íons de prata) e antichamas.

Dentro do ramo de decoração, o vidro também é utilizado tanto em móveis, como armários e mesas, espelhos e até paredes, como também em eletrodomésticos como geladeiras, fogões, fornos e micro-ondas.

No caso da indústria moveleira, são comumente utilizados os vidros coloridos, que costumam ser pintados e têm sua cor adquirida diretamente na linha de produção, o que proporciona um brilho e resistência maiores para suas aplicações. Outra aplicação são os vidros serigrafados, que possuem um esmalte cerâmico em sua composição, após passarem por um forno de têmpera, tornando-o muito resistente, inclusive a metais afiados.

Um dos principais clientes da indústria do vidro plano é a indústria automotiva, que utiliza de forma frequente os componentes produzidos para os mais variados tipos de veículos. Geralmente o vidro plano é produzido por uma empresa e enviado para uma processadora, que irá recortá-lo e realizar os tratamentos necessários para a função final que o produto terá.

Em um carro, por exemplo, os vidros dianteiros são feitos de vidro laminado, onde duas chapas são unidas por uma camada de PVB (polivinil butiral), oferecendo maior segurança, principalmente em casos de acidentes, onde essa folha que une as lâminas não deixa que os pedaços se soltem.

Já para os vidros laterais e traseiros, são utilizados vidros temperados, recortados de acordo com o modelo específico do veículo e, no caso dos vidros traseiros, unidos a filetes condutores de eletricidade que tem função de desembaçar.

Também são enviados a essas processadoras, os espelhos, que são modelados e unidos a materiais de plástico e borrachas e montando como produtos finais os espelhos retrovisores laterais e internos.

Dessa forma, esse ramo se torna um dos principais clientes da indústria do vidro plano, influenciando de forma direta na necessidade de demanda ao longo do ano, variando não apenas

conforme a disponibilidade de vidros, mas também demais componentes como chips e alguns materiais eletroeletrônicos imprescindíveis para a montagem dos veículos.

2.3 - Software R

O programa “R” é um software gratuito desenvolvido para, entre outras funções, possibilitar os mais variados tipos de análises de forma rápida e confiável, desde questões estatísticas por meio de cálculos diretos de fontes externas, até mesmo análises gráficas baseadas em dados. Foi construído para rodar nos mais diversos sistemas operacionais, incluindo Windows ® e MacOS ® e de forma aberta, possibilitando aos usuários entender seu funcionamento e, devido ao código fonte aberto, adaptar o programa as mais variadas necessidades.

O software utiliza como linguagem base o R, que se trata de uma linguagem de programação interpretada, ou seja, não possui um compilador que irá rodar o código inteiro, e sim será feita uma leitura linha a linha.

Atualmente, é utilizado em diversas aplicações como “Data Science”, auxiliando desde a coleta, tratamento, transformação e até mesmo exibição de dashboards¹ com os dados, o que acaba possibilitando análises estatísticas, criação de modelos de inferência e de regressão. Por ter sido criado por estatísticos, é uma linguagem fortemente utilizada em questões de estatística computacional.

Devido a suas funcionalidades ligadas a esses conceitos também acaba sendo utilizado em diversas aplicações de *Machine Learning*, tornando-se peça fundamental para análises preditivas, por meio de seus pacotes e bibliotecas (também criadas por usuários) que permitem modelagem da dados para modelos de regressão linear, não linear, clusterização, entre outros.

Para o desenvolvimento das atividades desse trabalho, foi utilizado o “R Studio”, um programa que utiliza o R como linguagem base, porém apresenta uma interface mais amigável e de fácil entendimento tanto para o usuário, quanto para quem venha a fazer um estudo ou análise sobre o trabalho. O R Studio possui diversos tipos de versões pagas, mas para o desenvolvimento desse trabalho foi utilizado a versão gratuita, que se adequa e possui as funcionalidades necessárias ao que foi proposto.

¹Dashboard é uma ferramenta de gestão presente na análise de dados que tem a função de proporcionar uma visualização dos dados de forma gráfica, a fim de permitir uma análise de indicadores chaves para as empresas

2.4 - Conceitos Estatísticos

Para esse trabalho, será construída uma análise estatística baseada na correlação entre duas variáveis, sendo necessária, portanto, a aplicação de diversos conceitos do ramo da Inferência Estatística.

Segundo Wilton Bussab e Pedro Moretin (2017):

É uma das partes da Estatística. Esta é a parte da metodologia da Ciência que tem por objetivo a coleta, redução, análise e modelagem dos dados, a partir do que, finalmente, faz-se a inferência para uma população da qual os dados (a amostra) foram obtidos. Um aspecto importante da modelagem dos dados é fazer previsões, a partir das quais se podem tomar decisões.

2.4.1 - Tipos de Variáveis

A estatística trabalha com alguns tipos de variáveis, como por exemplo as qualitativas e as quantitativas. As variáveis qualitativas referem-se a características imensuráveis, ou seja, que não podem ser medidas. São exemplos de variáveis qualitativas características como sexo, raça, tipo físico, cores etc.

Dentro das variáveis quantitativas, por sua vez, existem as que são discretas e as contínuas. As primeiras se tratam de variáveis que sempre vão assumir valores inteiros, como por exemplo idade, número de falhas ou número de reclamações de clientes, variando também sempre em unidades inteiras. Já as variáveis contínuas, podem assumir valores fracionários e quaisquer outros tipos, tendo como exemplos peso, médias, tempos, ângulos e todas as demais que não são obrigatoriamente inteiras.

Para esse trabalho, como buscou-se analisar a correlação de variáveis de indicadores econômicos, diretamente com a produção de vidro plano ao longo do tempo, serão utilizadas variáveis quantitativas contínuas, que variam conforme o passar do tempo e possuem valores com algumas casas decimais.

Sobre os tipos de variáveis, de acordo com Wilton Bussab e Pedro Moretin (2017):

Algumas variáveis como sexo, escolaridade e estado civil, tem como possíveis respostas uma descrição ou qualidade do indivíduo, e, portanto, são chamadas de variáveis qualitativas. Já variáveis como número de filhos e salário tem como possíveis respostas um número, um valor, uma quantidade, e, portanto, são chamadas de variáveis quantitativas.

2.4.2 - Séries estatísticas

As séries estatísticas são tabelas que apresentam a distribuição de conjuntos de dados, possuindo número indefinido de colunas, sempre de acordo com a quantidade de dados obtidos para a análise a ser realizada.

São definidas por Tarciana Liberal (2019), como uma apresentação de informações (variáveis estatísticas) no formato de tabelas com o objetivo de resumir os dados observados e torná-los mais entendíveis. A autora também destaca que tanto uma tabela quanto mesmo um gráfico devem apresentar um cabeçalho, um corpo e um rodapé.

Também possuem algumas classificações, assim como os tipos de variáveis. Dentre elas estão as séries temporais ou históricas, que possuem sempre alguma informação relativa a datas, como dia, mês, ano ou período em uma coluna e outros dados em outras; geográficas, que buscam trazer informações relativas a determinados locais, sejam ruas, bairros, cidades, estados ou até mesmo países; mista ou conjugada, que são séries com diversos tipos de informações, podendo envolver locais como nas séries geográficas e datas ou períodos, como em séries temporais; categóricas, geralmente utilizadas para construção de rankings, onde os dados são ordenados seguindo uma determinada condição imposta pelo usuário que está analisando ou quem formulou a tabela.

2.4.3 - Gráfico de Dispersão

O Gráfico de Dispersão é o primeiro passo para se começar uma análise entre duas variáveis quantitativas ao longo do tempo. Colocando-se as duas variáveis em um diagrama x-y, onde cada ponto será alocado de acordo com as suas coordenadas cartesianas, pode se obter uma primeira análise visual de como se comportam os dados ao longo de uma determinada série estatística.

Segundo artigo publicado por Graça Martins e Maria Eugénia (2014):

Na presença de um conjunto de dados bivariados o primeiro passo na análise desses dados é representá-los num diagrama de dispersão. A forma da nuvem de pontos, representada no diagrama, pode mostrar uma associação linear entre as duas variáveis, que pode ser expressa numericamente pelo coeficiente de correlação amostral de Pearson ou pelo seu quadrado que se chama coeficiente de determinação.

Já segundo Áurea Sousa em seu artigo “Diagrama de dispersão, correlação e regressão linear” (Açores, 2019), a representação gráfica para uma análise estatística, permite que sejam

identificados padrões nos conjuntos de dados, assim como presença de outliers (valores muito grandes ou muito pequenos quando comparados aos demais). A representação de pares (X,Y) em um sistema cartesiano, portanto, permite a criação de um diagrama que evidencia uma possível relação existente entre as variáveis, seja ela linear ou não.

Para demonstrar um exemplo, foi criado o Quadro 2.1, que reflete dados muito utilizados na logística, envolvendo as variáveis frete e distância. Por meio da análise desses dados, é possível identificar possíveis locais onde o frete esteja muito acima do praticado pela empresa em destinos com distâncias parecidas.

Quadro 2.1 – Distância x Frete

Distância	Frete	Situação
5	3	Padrão
10	6	Padrão
14	10	Padrão
7	23	Outlier
24	12	Padrão
48	25	Padrão
98	56	Padrão
200	120	Padrão

Fonte: Autor (2021)

Para uma melhor visualização dos outliers, foi criado um gráfico, apresentado na figura 2.1:

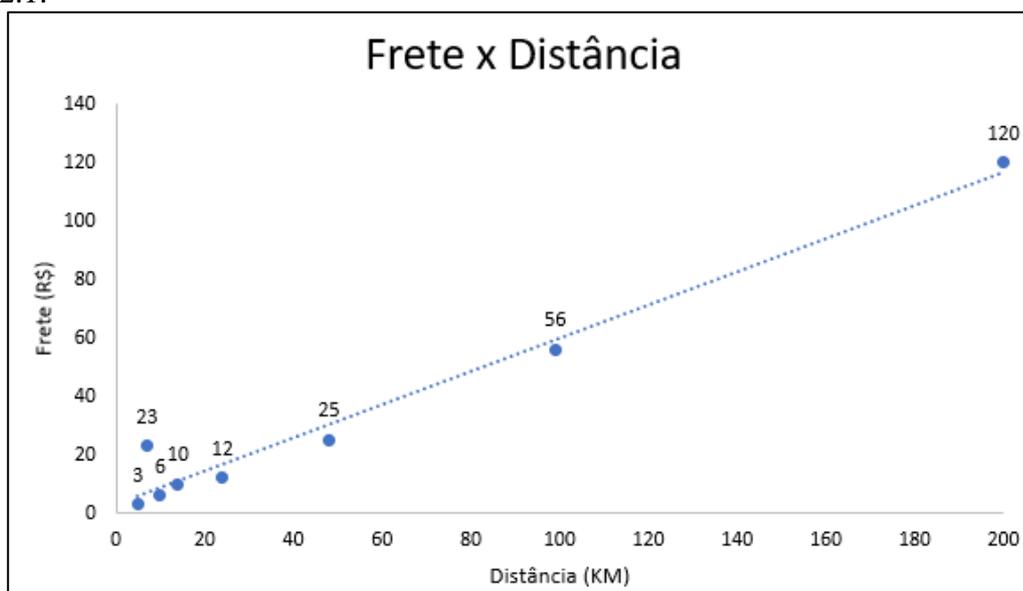


Figura 2.1 - Frete x Distância

Fonte: Autor (2021)

Desse modo, podem ser reavaliados contratos e feitas negociações de redução de custos e adequação de valores para que os pontos fiquem cada vez mais próximos de uma relação linear.

2.4.4 - Covariância

A covariância é uma estatística que permite analisar a relação entre duas variáveis aleatórias, onde em seu cálculo são usadas as respectivas médias e tamanho do conjunto de dados. Por meio dela, permite-se medir o grau de interdependência entre essas variáveis, ou seja, quando o valor da covariância é 0, diz-se que as variáveis são independentes, então não possuem nenhum tipo de relação. Quando esse valor difere de 0, existe algum tipo de relação, seja ela diretamente ou inversamente proporcional.

A equação de covariância (1) é determinada pela multiplicação da somatória entre a diferença das variáveis em cada par ordenado:

$$C(X, Y) = \frac{\sum_i^n (X_i - \bar{X}) * (Y_i - \bar{Y})}{n} \quad (1)$$

Onde $C(X, Y)$ é a Covariância entre X e Y, n é o número de pares do conjunto de dados, i é a posição i do valor a ser considerado, X_i é a variável X na posição i do conjunto de dados, Y_i é a variável Y na posição i do conjunto de dados, \bar{X} é a média do conjunto de dados X e \bar{Y} é a média do conjunto de dados Y.

O resultado dessa equação permite a indicação de uma possível relação linear entre as variáveis, onde:

- $C(X, Y) = 0$, não possuem relação, variáveis independentes.
- $C(X, Y) > 0$, possuem relação direta, quando uma aumenta a outra aumenta e quando uma diminui a outra diminui.
- $C(X, Y) < 0$, possuem uma relação oposta, ou seja, quando uma aumenta, a outra diminui.

É importante ressaltar que a covariância, apesar de indicar uma relação entre as variáveis, não é capaz de trazer a magnitude dessa relação, por não ser uma equação tratada de forma normalizada.

Segundo o professor Paulo Ricardo B. Guimarães, da Universidade Federal do Paraná (2021), “Apesar da covariância ser uma estatística adequada para medir relação linear entre duas variáveis, ela não é adequada para comparar graus de relação entre variáveis, dado que ela está influenciada pelas unidades de medida de cada variável”.

2.4.5 - Correlação

Para que seja possível medir não apenas uma relação direta ou oposta entre as variáveis, mas também a influência de seus valores na magnitude das mudanças, existe uma análise estatística denominada correlação. Essa análise utiliza o desvio padrão de cada conjunto de dados para normalizar a equação de covariância e trazer um coeficiente capaz de medir e indicar a força da relação entre os dados.

O desvio padrão, por sua vez, é uma medida que demonstra o grau de dispersão do conjunto de dados em relação à sua média central, ou seja, o quão longe da média estão os dados analisados. Quanto mais longe de 0 for o desvio-padrão, mais dispersos estão os dados, ao passo que, quanto menor o seu valor, mais uniformemente distribuídos ao redor da média estão os valores apresentados.

A equação do desvio padrão (2) é feita para cada variável de forma separada, por meio da raiz quadrada do somatório da diferença entre cada valor e a média do conjunto:

$$Dp = \sqrt{\frac{\sum_{i=1}^n (X - \bar{X})^2}{n}} \quad (2)$$

Onde Dp é o próprio desvio padrão, n é o tamanho do conjunto de dados, i é a posição do valor X a ser considerado e \bar{X} é a média dos valores do conjunto X.

A correlação, portanto, utiliza o valor do desvio padrão para normalizar a equação de covariância, tornando o cálculo um pouco mais complexo, conforme demonstrado abaixo na equação de correlação (3):

$$\text{Cor}(X,Y) = \frac{\frac{\sum_{i=1}^n (X_i - \bar{X}) * (Y_i - \bar{Y})}{n}}{\sqrt{\frac{\sum_{i=1}^n (X - \bar{X})^2}{n}} \sqrt{\frac{\sum_{i=1}^n (Y - \bar{Y})^2}{n}}} \quad (3)$$

O valor obtido por meio da equação (3) é conhecido como coeficiente de Pearson, que permite identificar o grau ou magnitude da relação entre as variáveis e não apenas se estão relacionadas diretamente, inversamente ou não estão relacionadas. O resultado dessa equação sempre irá variar entre -1 e 1 e, para se obter o seu valor, todas as variáveis devem ser quantitativas contínuas ou discretas. A partir do resultado, permite-se fazer alguns tipos de análise, como:

$Cor(X,Y) = 0$, não existe uma relação linear entre as variáveis

$Cor(X,Y) = 1$, existe uma relação linear perfeita e positiva entre as variáveis

$Cor(X,Y) = -1$, existe uma relação linear perfeita e negativa entre as variáveis

$Cor(X,Y) > 0$, existe uma correlação positiva entre as variáveis

$Cor(X,Y) < 0$, existe uma correlação negativa entre as variáveis

Alguns autores determinam tipos de correlação para medir a força da relação entre as variáveis, como Áurea Sousa (2019) demonstra no Quadro 2.2:

Quadro 2.2: Interpretação de Correlações

Cor (X,Y)	Interpretação
$ Cor(X,Y) < 0,2$	Negligenciável
$0,2 < Cor(X,Y) < 0,4$	Correlação Fraca
$0,4 < Cor(X,Y) < 0,6$	Correlação moderada
$0,6 < Cor(X,Y) < 0,8$	Correlação forte
$ Cor(X,Y) > 0,8$	Correlação muito forte

Fonte: Áurea Sousa, 2019

Já Juan Diego e Hernández Lalinda (2018), possuem uma diferenciação com menos divisões, apresentada no Quadro 2.3:

Quadro 2.3: Interpretação de Correlação

Intervalo de valores de Cor(X,Y)	Interpretação
$0.00 \leq Cor(X,Y) < 0.10$	Correlação Nula
$0.10 \leq Cor(X,Y) < 0.30$	Correlação Fraca
$0.30 \leq Cor(X,Y) < 0.50$	Correlação Moderada
$0.50 \leq Cor(X,Y) < 1.00$	Correlação Forte

Fonte: Juan Diego, et al, 2018

Por fim, o Quadro 2.4 apresenta a interpretação de Cohen (1988):

Quadro 2.4: Força de Associação

Força de Associação	Positiva	Negativa
Pequena	0.1 a 0.3	-0.1 a -0.3
Média	0.3 a 0.5	-0.3 a -0.5
Grande	0.5 a 1.0	-0.5 a -1.0

Fonte: Cohen, 1988

Para esse trabalho, será utilizada a segunda tabela, proposta por Hernandez Lalinde e Juan Diego, por segregar mais os valores obtidos na equação e permitir que seja feita uma análise mais detalhada e separada dos valores que serão obtidos e interpretados por meio do programa R.

2.5 - Análise de Dados

Uma análise de dados completa possui diversas etapas que a compõe, desde a escolha crítica dos dados que deverão ser utilizados, passando por sua extração, manipulação, tratamento e transformação até utilização final em alguma análise mais específica.

Geralmente, conjuntos de dados possuem os mais variados formatos e não tem um padrão, o que dificulta um trabalho inicial rápido em cima do que será manipulado. Dados de indicadores econômicos, por exemplo, podem vir em tabelas mensais, anuais, trimestrais ou até mesmo quinzenais. Por esse motivo, é necessária uma predefinição do formato que será utilizado para padronizar e adequar todos os conjuntos de acordo. Para a utilização desses dados, portanto, os mesmos passam por uma etapa de preparação, que engloba diversos passos até deixar os conjuntos manipuláveis de forma padronizada e adequada ao usuário que fará o trabalho em cima deles.

A primeira etapa se trata da identificação, onde são definidos quais os dados que serão utilizados para realizar a análise. Essa etapa é de grande importância, pois é responsável por definir a base que será exportada, baseada em definições passadas pela empresa ou pessoa interessada na análise que será feita, considerando objetivos financeiros, corte de custos e objetivos a serem atingidos, por exemplo.

Segundo Bruno Guerra (2020):

O processo de preparação de dados começa com a localização das informações corretas. Elas podem ser acessadas de qualquer fonte, independentemente da origem e do formato. Com a ferramenta adequada, essa extração é realizada de forma automática e inteligente, mesmo que as fontes sejam complexas e não-estruturadas – como quando a planilha está salva em PDF.

A próxima etapa trata-se de descobrir e realizar uma análise sobre os dados que serão utilizados. Nesse passo, geralmente é feito um “*Sanity Check*” do que está sendo apresentado, ou seja, uma verificação se o conjunto de dados possui algum determinado padrão, a quantidade de linhas, colunas, dados diferentes por linha ou coluna. Com esse primeiro resultado da formatação dos dados, é realizada uma união dos dados em uma tabela principal.

Feita essa análise em cima da qualidade dos dados, o próximo passo é realizar uma limpeza melhora dos dados, principalmente completando possíveis valores faltantes, remover dados inconsistentes e possíveis outliers que causem impacto significativo no resultado. É uma etapa extremamente relevante por se tratar de uma remoção de dados inconsistentes e o preenchimento de possíveis lacunas.

Após serem analisados e limpos, os dados passam por um processo de transformação, onde são manipulados de forma direta para que possam ser compreendidos pelo maior número de pessoas possível que terá acesso a base. É uma etapa relevante, pois direciona as mudanças para as áreas da empresa ou pessoas que de fato serão responsáveis por sua utilização. Com essa etapa realizada, permite-se uma construção de uma análise mais coerente com a realidade da área ou negócio a ser analisada, economizando, portando, tempo e possíveis esforços desnecessários.

Por fim é realizada a exportação dos dados e sua utilização pelas mais variadas ferramentas de Inteligência de Negócio presentes no mercado. Com dados limpos, tratados e padronizados, é possível realizar análise gráficas, estatísticas, entre diversas outras, além de dashboards dos negócios. Dessa forma analistas, gestores e até mesmo diretores tem a possibilidade de tomar decisões estratégicas baseadas em dados confiáveis e que traduzem a realidade da empresa ao longo do tempo.

CAPÍTULO 3

Metodologia

O presente estudo busca apresentar uma análise estatística baseada em dados reais, com séries históricas, da produção de vidro plano ao longo dos últimos 20 anos e diversos índices econômicos. As empresas produtoras de vidro plano presentes no Brasil não divulgam seus resultados de produção em dados abertos e públicos, por isso foi necessário buscar os dados presentes na base de dados do IBGE, que apresenta números gerais da produção de vidro plano e de segurança mensalmente. Já os dados dos indicadores econômicos foram obtidos em fontes do governo como o Banco Central e o mesmo IBGE por meio do IPEA (Instituto de Pesquisa Econômica Aplicada).

Diversos estudos de correlação são realizados por empresas para definir seus planejamentos estratégicos ao longo dos anos. Análises de regressão para calcular, por exemplo, previsões de demanda, são utilizadas pelas mais variadas organizações, a fim de buscar ajustar linhas de produção e seus respectivos planejamento de vendas, visando não apenas vender mais, mas também manter estoques que suportem as variações de demanda e evitem custos desnecessários.

A análise de correlação, por sua vez, traz uma estimativa estatística, que é feita por meio de cálculos que utilizam dados históricos para verificar possíveis influências que determinadas variáveis terão sobre outras. Isso significa que é possível presumir, por exemplo, que se uma variável tiver uma correlação positiva forte com outra, é provável que a tendência de aumento de uma irá impactar no aumento da outra, da mesma forma que a queda de uma deverá impactar na queda da outra.

A proposta desse estudo, portanto, é apresentar uma análise de correlação entre algumas variáveis econômicas que supostamente impactariam diretamente produção de vidro plano e de segurança. Os impactos, como demonstrado no capítulo anterior, poderiam ser na diminuição de juros que facilitam parcelamentos e financiamentos para indústria de construção e de veículos, ou até mesmo na possibilidade de comprar matéria-prima mais barata, o que aumentaria a produção ou número de obras, por exemplo.

Para fazer esse estudo de correlação, foi selecionado o software R, que possui um histórico de utilização para cálculos estatísticos e está presente em diversos estudos e

congressos a respeito dessa área. Como mencionado no capítulo anterior, por se tratar de um software de código aberto, diversos pesquisadores e estudiosos criaram bibliotecas que facilitam a utilização do programa e a análise dos cálculos e resultados.

O programa permite que sejam utilizados e criados “Data Frames” para se trabalhar com os dados, o que facilita na manipulação e visualização em formato de tabelas, auxiliando nos ajustes que serão necessários. Como os dados presentes no estudo, conforme mencionado anteriormente, foram extraídos de fontes oficiais do governo e estão em planilhas de Excel, todos serão colocados em *Data Frames*¹ separados, analisados individualmente e padronizados para que possam ser utilizados nos cálculos e construção dos gráficos.

Para criação dos “Data Frames” o programa permite que o usuário trabalhe sem a necessidade de instalação de pacotes externos, mas nesse estudo, por exemplo, serão utilizadas bibliotecas como a “GGPLOT” e a “CORPLOT” para facilitar a auxiliar no tratamento e manipulação dos dados, além da criação de elementos visuais essenciais para a análise. O primeiro pacote (“GGPLOT”), permite a criação dos mais variados tipos de gráficos, como gráficos de variações mensais, gráficos dispersão, que serão utilizados como uma primeira análise visual para verificar possíveis correlações entre os dados. Já o segundo pacote (“CORPLOT”) será utilizado por se tratar de um pacote focado em análises estatísticas, que permite a criação de uma matriz geral que correlaciona todas as variáveis presentes no estudo, já demonstrando visualmente por meio de cores e tamanhos se há algum tipo de correlação mais evidente entre os dados.

Por fim, será utilizada uma função já presente no próprio R que faz o cálculo da correlação entre as variáveis, trazendo diretamente o resultado, que será analisado juntamente com o gráfico de dispersão para que se chegue nos resultados e, conseqüentemente, na conclusão. Como se trata de um estudo ainda não realizado, as correlações que serão calculadas serão justificadas de acordo com o Quadro 2.3, encontrado na literatura e definida como padrão para os resultados.

O estudo também buscará uma análise de correlação entre a produção de vidro e a produção de veículos e o indicador de construção civil, divulgados pelas respectivas organizações responsáveis. Essa análise busca comprovar e reforçar a tese que é um estudo eficaz para ser utilizado em outras situações.

¹Data Frames são uma forma de se organizar dados semelhante a uma matriz, onde cada linha corresponde a um registro e cada coluna a um campo relativo à propriedade.

Vale destacar que, como não foi possível obter os indicadores relativos as vendas de vidro plano, é possível que os resultados demonstrem alguma disparidade com a realidade das empresas. A produção de vidro plano e de segurança (única fonte de dados disponível), varia não apenas com as vendas, mas também com o planejamento de estoques e projetos que as empresas tenham visando o futuro e não necessariamente com o seu faturamento naquele mês.

Isso significa que não necessariamente uma variação na taxa de juros que facilite o financiamento de um veículo no mês atual, terá impacto na produção do mesmo mês, uma vez que a produção daquele mês já estava prevista anteriormente de acordo com outro momento de mercado.

CAPÍTULO 4

Estudo de Caso

Como mencionado no capítulo 3, o presente trabalho fará um estudo de caso a fim de obter uma análise que demonstre a correlação entre determinadas variáveis econômicas brasileiras com a produção de vidro plano ao longo dos meses dos últimos 20 anos, tendo como início o mês de janeiro de 2002.

4.1 - Análise dos Dados

4.1.1 - Definição das Bases de Dados Utilizadas

IPCA – Para o IPCA, foi utilizada a base histórica presente no site do IBGE (Instituto Brasileiro de Geografia e Estatística), que possui a série histórica oficial.

SELIC – Para obtenção dos dados relativos à taxa de juros (SELIC), também foi utilizada uma fonte oficial do governo brasileiro, porém dessa vez foi o Instituto de Pesquisa Econômica Aplicada.

IGPM – O IGPM é calculado pela Fundação Getúlio Vargas (FGV), que é a instituição não governamental responsável por realizar o acompanhamento e cálculo desse índice ao longo do tempo. Mas os dados foram extraídos do Instituto de Pesquisa Econômica Aplicada, que possui uma lista mensal com a série histórica desde o começo de sua publicação.

PIB – O Produto Interno Bruto (PIB) é um dado também calculado pelo Instituto Brasileiro de Geografia e Estatística, ou seja, é um dado oficial oriundo de fontes governamentais e aberto ao público em geral que queira acompanhar tudo o que o país produziu em determinado ano. Também foi extraído do Instituto de Pesquisa Econômica Aplicada.

IBOVESPA – O Índice da Bolsa de Valores é obtido diretamente do site da sua empresa controladora, que atualmente é a B3. Como se trata de um índice encontrado em diversas plataformas, para não haver nenhuma divergência, foi selecionado a fonte oficial para o conjunto de dados.

INCC – o Índice Nacional do Custo de Construção também é calculado pela Fundação Getúlio Vargas (FGV), portanto, também foi retirada a série histórica diretamente do site do Instituto.

Produção de Vidro Plano - Obtido por meio do site da SIDRA, que traz a “Pesquisa Industrial Mensal - Produção Física”, com a variação mensal da produção desde janeiro de 2002.

4.1.2 - Análise dos dados importados

Após identificadas as fontes de dados que serão utilizadas, todas foram colocadas em uma pasta específica do projeto, para que posteriormente fossem extraídas para o programa R como data frames, onde seriam identificadas possíveis inconsistências nos dados e ajustes para que ficassem padronizados. Primeiramente, foi identificado no programa o diretório onde estavam armazenadas as bases, como demonstrado na Figura 4.1:



Figura 4.1 – Tela do Programa R – Diretório
Fonte: Autor (2021)

Após especificado o diretório, a Figura 4.2 apresenta a importação de todas as bases como Data Frames, pois tratavam-se variáveis de classes diferentes que fariam parte do mesmo banco, sendo cada coluna uma variável e cada linha uma observação:

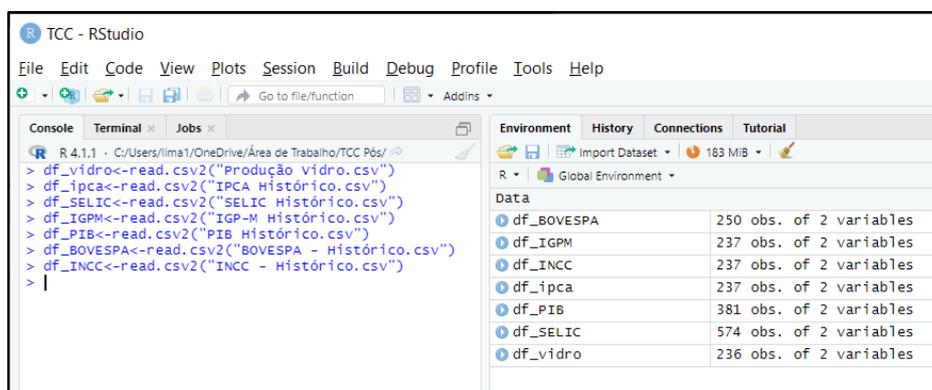


Figura 4.2 – Tela do Programa R – Arquivos importados
Fonte: Autor (2021)

Após a importação de todas as bases, foi possível identificar que todas possuíam o mesmo número de colunas, uma identificando o mês e uma identificando o valor do índice

estudado. Porém a quantidade de linhas, ou seja, a quantidade de dados disponíveis, era diferente em algumas situações, pelo fato de algumas buscarem um range maior de datas que outras. A Figura 4.3 mostra que, para identificar as colunas presentes em cada uma, foi utilizada a função `head()`, que mostra os primeiros dados e títulos das respectivas colunas:

```

R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> head(df_BOVESPA)
  i..Mes BOVESPA
1 jan/01 17672.77
2 fev/01 15891.41
3 mar/01 14438.45
4 abr/01 14917.54
5 mai/01 14649.98
6 jun/01 14559.79
> head(df_IGPM)
  i..Mes IGPM
1 jan/02 0.003594510
2 fev/02 0.000599244
3 mar/02 0.000967430
4 abr/02 0.005568851
5 mai/02 0.008238363
6 jun/02 0.015434200
> head(df_INCC)
  i..Mes INCC
1 jan/02 0.003603680
2 fev/02 0.005776001
3 mar/02 0.005468920
4 abr/02 0.003282897
5 mai/02 0.025335039
6 jun/02 0.005659962
> head(df_ipca)
  i..Mes IPCA
1 jan/02 0.0642
2 fev/02 0.0441
3 mar/02 0.0744
4 abr/02 0.1004
5 mai/02 0.0255
6 jun/02 0.0516
> head(df_PIB)
  i..Mes PIB
1 jan/01 102530.7
2 fev/01 101635.3
3 mar/01 108303.8
4 abr/01 107572.0
5 mai/01 111202.2
6 jun/01 104949.4
> head(df_SELIC)
  i..Mes SELIC
1 jan/74 0.19
2 fev/74 0.15
3 mar/74 0.15
4 abr/74 0.16
5 mai/74 0.16
6 jun/74 0.20
> head(df_vidro)
  i..Mes Producao.Vidro
1 jan/02 73.8
2 fev/02 72.0
3 mar/02 80.7
4 abr/02 79.5
5 mai/02 82.1
6 jun/02 77.3
> |

```

Figura 4.3 – Tela do Programa R – Cabeçalhos das bases
 Fonte: Autor (2021)

Observou-se, portanto, que todas as bases possuíam uma coluna chamada “Mês” e uma de acordo com o dado importado, tornando possível a junção de todos em uma tabela conjunta, com uma coluna fixa indicando mês e ano.

Foi possível identificar também os primeiros meses de cada base, observando-se que todas continham dados a partir dos meses de janeiro de 2002. Já para verificar quais os últimos meses presentes nas bases, foi utilizada a função `tail()`, que permite a visualização dos últimos 6 dados presentes, conforme demonstrado na figura 4.4:

```
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> tail(df_BOVESPA)
  i..Mes  BOVESPA
245 mai/21 126215.7
246 jun/21 126801.7
247 jul/21 121800.8
248 ago/21 118781.0
249 set/21 110979.1
250 out/21 103500.7
> tail(df_IGPM)
  i..Mes      IGP.M
232 abr/21 0.015079796
233 mai/21 0.040965333
234 jun/21 0.006022688
235 jul/21 0.007780763
236 ago/21 0.006632229
237 set/21 -0.006396100
> tail(df_INCC)
  i..Mes      INCC
232 abr/21 0.009004107
233 mai/21 0.022188921
234 jun/21 0.021602623
235 jul/21 0.008460268
236 ago/21 0.004639930
237 set/21 0.005130366
> tail(df_ipca)
  i..Mes      IPCA
232 abr/21 0.0378
233 mai/21 0.1043
234 jun/21 0.0655
235 jul/21 0.1215
236 ago/21 0.1095
237 set/21 0.1484
> tail(df_PIB)
  i..Mes      PIB
244 abr/21 717796.8
245 mai/21 714313.1
246 jun/21 711254.2
247 jul/21 727346.1
248 ago/21 727121.7
249 set/21 716186.5
> tail(df_SELIC)
  i..Mes      SELIC
569 mai/21 0.03
570 jun/21 0.04
571 jul/21 0.04
572 ago/21 0.05
573 set/21 0.05
574 out/21 0.06
> tail(df_vidro)
  i..Mes Producao.Vidro
231 mar/21 102.0
232 abr/21 101.9
233 mai/21 107.1
234 jun/21 98.9
235 jul/21 99.9
236 ago/21 99.7
> |
```

Figura 4.4 – Tela do Programa R – Últimas linhas das bases
Fonte: Autor (2021)

Por definição, portanto, foram considerados todos os meses a partir de janeiro de 2002, que estava presente em todas as bases, até agosto de 2021, que também estava presente em todas as bases e englobava o último mês antes da finalização dos dados desse estudo.

Foi identificado que todas as variáveis econômicas estavam com suas unidades em porcentagem, enquanto o IBOVESPA possuía sua chamada “pontuação”, já o “PIB” veio em

milhões de reais e a produção de vidro baseada com um índice de base fixa sem ajuste sazonal, considerando o ano de 2012 como base média (número índice 2012 = 100).

4.1.3 - Ajustando a base

O primeiro passo foi ajustar a base e eliminar dados das bases identificadas com início anterior a janeiro de 2002, que são: IBOVESPA, PIB e SELIC.

Para o IBOVESPA, conforme demonstrado com a função `head()`, a base em questão começava em Janeiro de 2001. Dessa forma, a Figura 4.5 mostra que foram retirados os 12 primeiros meses utilizando-se novamente a função `tail()`, porém para excluir as 12 primeiras linhas, como se fossem consideradas todas as linhas abaixo.

```
> head(df_BOVESPA)
  ĩ..Mes      BOVESPA
1 jan/01 17.672,77
2 fev/01 15.891,41
3 mar/01 14.438,45
4 abr/01 14.917,54
5 mai/01 14.649,98
6 jun/01 14.559,79
> df_BOVESPA<-tail(df_BOVESPA,-12)
```

Figura 4.5 – Tela do Programa R – Remoção 12 primeiros meses da Base BOVESPA
Fonte: Autor (2021)

Para verificar se agora a base ficou correta, ou seja, começando do mês de Janeiro de 2002, novamente foi utilizada a função `head()`, como demonstrado na figura 4.6:

```
> head(df_BOVESPA)
  ĩ..Mes      BOVESPA
13 jan/02 12.721,45
14 fev/02 14.033,29
15 mar/02 13.254,55
16 abr/02 13.085,12
17 mai/02 12.861,43
18 jun/02 11.139,16
> |
```

Figura 4.6 – Tela do Programa R – Base BOVESPA ajustada
Fonte: Autor (2021)

Para o PIB, da mesma forma que a BOVESPA, os dados estavam começando em janeiro de 2001, portanto, foi feito o mesmo procedimento, apresentado na Figura 4.7:

```

i..Mes      PIB
1 jan/01 102.530,70
2 fev/01 101.635,30
3 mar/01 108.303,80
4 abr/01 107.572,00
5 mai/01 111.202,20
6 jun/01 104.949,40
> df_PIB<-tail(df_PIB,-12)
> head(df_PIB)
i..Mes      PIB
13 jan/02 112.374,80
14 fev/02 111.477,10
15 mar/02 118.444,70
16 abr/02 120.385,90
17 mai/02 123.552,50
18 jun/02 123.424,40
> |

```

Figura 4.7 – Tela do Programa R - Remoção 12 primeiros meses da Base PIB
Fonte: Autor (2021)

Já para a base SELIC, como demonstrado na Figura 4.8, os dados se iniciavam em janeiro de 1974, sendo necessário, portanto, excluir $2002-1974 = 28 \text{ anos} \times 12 \text{ meses} = 336$ linhas, seguindo basicamente o mesmo procedimento:

```

> head(df_SELIC)
i..Mes SELIC
1 jan/74 0.19
2 fev/74 0.15
3 mar/74 0.15
4 abr/74 0.16
5 mai/74 0.16
6 jun/74 0.20
> df_SELIC<-tail(df_SELIC,-336)
> head(df_SELIC)
i..Mes SELIC
337 jan/02 0.20
338 fev/02 0.16
339 mar/02 0.18
340 abr/02 0.19
341 mai/02 0.18
342 jun/02 0.17
> |

```

Figura 4.8 – Tela do Programa R - Remoção 336 primeiros meses da Base SELIC
Fonte: Autor (2021)

Após ajustadas essas 3 bases, todas ficaram com o início padronizado em janeiro de 2002. O passo seguinte foi ajustar para que todas as bases ficassem com o último mês de dados como agosto de 2021. Após a verificação feita com a fórmula `tail()`, foi verificado que a base de produção de vidro está com esse padrão, então foram ajustadas todas as demais.

Na Figura 4.9, é possível identificar que o último mês da base da BOVESPA era outubro de 2021, portanto, foi necessário excluir os dois últimos dados, referentes aos meses de setembro e outubro de 2021. Como haviam sido retiradas as 12 primeiras linhas, utilizou-se a função `head()` para remodelar o Data Frame e considerar apenas as 236 primeiras linhas, a partir da linha 13, que se tornou a primeira após o ajuste de início:

```
> tail(df_BOVESPA)
  i..Mes      BOVESPA
245 mai/21 126.215,73
246 jun/21 126.801,66
247 jul/21 121.800,79
248 ago/21 118.781,03
249 set/21 110.979,10
250 out/21 103.500,71
> df_BOVESPA<-head(df_BOVESPA, 236)
> tail(df_BOVESPA)
  i..Mes      BOVESPA
243 mar/21 116.633,72
244 abr/21 118.893,84
245 mai/21 126.215,73
246 jun/21 126.801,66
247 jul/21 121.800,79
248 ago/21 118.781,03
> |
```

Figura 4.9 – Tela do Programa R – Removendo 2 últimos meses da Base BOVESPA
Fonte: Autor (2021)

Já para o IGP-M, foi identificado apenas um mês a mais (Setembro de 2021), que foi eliminado utilizando-se a mesma função `head()`, porém redefinindo o Data Frame para que fossem consideradas as 236 primeiras linhas, começando da linha 1, conforme apresenta a Figura 4.10:

```
> tail(df_IGPM)
  i..Mes      IGP.M
232 abr/21 0.015079796
233 mai/21 0.040965333
234 jun/21 0.006022688
235 jul/21 0.007780763
236 ago/21 0.006632229
237 set/21 -0.006396100
> df_IGPM<-head(df_IGPM, 236)
> tail(df_IGPM)
  i..Mes      IGP.M
231 mar/21 0.029387830
232 abr/21 0.015079796
233 mai/21 0.040965333
234 jun/21 0.006022688
235 jul/21 0.007780763
236 ago/21 0.006632229
> |
```

Figura 4.10 – Tela do Programa R – Removendo último mês da base IGP.M
Fonte: Autor (2021)

Como apresentado na Figura 4.11, o INCC segue exatamente o mesmo padrão do tratamento do IGP-M, portanto seguiu-se o mesmo procedimento de excluir a última linha, ajustando o Data Frame para que fossem considerados a partir da primeira linha:

```
> tail(df_INCC)
  i..Mes      INCC
232 abr/21 0.009004107
233 mai/21 0.022188921
234 jun/21 0.021602623
235 jul/21 0.008460268
236 ago/21 0.004639930
237 set/21 0.005130366
> df_INCC<-head(df_INCC,236)
> tail(df_INCC)
  i..Mes      INCC
231 mar/21 0.013045945
232 abr/21 0.009004107
233 mai/21 0.022188921
234 jun/21 0.021602623
235 jul/21 0.008460268
236 ago/21 0.004639930
> |
```

Figura 4.11 – Tela do Programa R – Removendo último mês da base INCC
Fonte: Autor (2021)

Para o IPCA, na Figura 4.12 é possível verificar que foi feito o mesmo procedimento dos dois anteriores, por se tratar de uma base começando no período correto e com um mês a mais do que o definido.

```
> tail(df_ipca)
  i..Mes      IPCA
232 abr/21 0.0378
233 mai/21 0.1043
234 jun/21 0.0655
235 jul/21 0.1215
236 ago/21 0.1095
237 set/21 0.1484
> df_ipca<-head(df_ipca,236)
> tail(df_ipca)
  i..Mes      IPCA
231 mar/21 0.1175
232 abr/21 0.0378
233 mai/21 0.1043
234 jun/21 0.0655
235 jul/21 0.1215
236 ago/21 0.1095
> |
```

Figura 4.12 – Tela do Programa R – Removendo último mês da base IPCA
Fonte: Autor (2021)

Já o PIB, trata-se de um Data Frame que foi ajustado, tendo seu início alterado para a linha 13. Portanto, ao se excluir as últimas linhas, serão consideradas todas entre a linhas 13 e a linha 248, que se refere ao mês de agosto de 2021, conforme procedimento demonstrado na Figura 4.13.

```

> tail(df_PIB)
  i..Mes      PIB
244 abr/21 717.796,80
245 mai/21 714.313,10
246 jun/21 711.254,20
247 jul/21 727.346,10
248 ago/21 727.121,70
249 set/21 716.186,50
> df_PIB<-head(df_PIB,236)
> tail(df_PIB)
  i..Mes      PIB
243 mar/21 730.633,70
244 abr/21 717.796,80
245 mai/21 714.313,10
246 jun/21 711.254,20
247 jul/21 727.346,10
248 ago/21 727.121,70
> |

```

Figura 4.13 – Tela do Programa R – Removendo último mês da base PIB
Fonte: Autor (2021)

O Data Frame da SELIC teve seu início alterado, pois começava em janeiro de 1974. Seu início passou a ser na linha 337, terminando no mês de agosto de 2021, que ficava na linha 572. Na Figura 4.14, é possível observar que a base terminava na linha 574, então foi necessário excluir as últimas duas linhas, referentes aos meses de setembro e outubro de 2021, respectivamente.

```

> tail(df_SELIC)
  i..Mes SELIC
569 mai/21 0.03
570 jun/21 0.04
571 jul/21 0.04
572 ago/21 0.05
573 set/21 0.05
574 out/21 0.06
> df_SELIC<-head(df_SELIC,236)
> tail(df_SELIC)
  i..Mes SELIC
567 mar/21 0.02
568 abr/21 0.03
569 mai/21 0.03
570 jun/21 0.04
571 jul/21 0.04
572 ago/21 0.05
> |

```

Figura 4.14 – Tela do Programa R – Removendo 2 últimos meses da base SELIC
Fonte: Autor (2021)

A base final, portanto, será formada por 8 colunas (Mês, BOVESPA, IGPM, INCC, IPCA, PIB, SELIC e Produção Vidro) e 236 linhas, representando todos os meses entre janeiro de 2002 e agosto de 2021.

Foi escolhida uma quantidade de dados separadas por mês, pois quase todos os índices possuem medições e variações mensais, com exceção da SELIC, que é ajustada pelo Banco Central de forma não periódica, mas possui o seu valor mensalmente em todos os anos.

O período de 236 meses foi definido pois abrange diversas mudanças políticas e econômicas enfrentadas pelo Brasil e pelo mundo desde 2002. Passando, por exemplo, pela crise de 2008, Impeachment de Presidente em 2016 e Pandemia em 2020, eventos que influenciam diretamente na economia de qualquer país.

Após a padronização dos data frames, todos foram unidos em um mesmo Data Frame chamado “df_BD”, mostrado na Figura 4.15 abaixo:

```
> df_BD = cbind.data.frame(df_BOVESPA, df_IGPM, df_INCC, df_ipca, df_PIB, df_SELIC, df
_vidro)
> head(df_BD)
  i..Mes BOVESPA i..Mes      IGP.M i..Mes      INCC i..Mes      IPCA i..Mes
13 jan/02 12721.45 jan/02 0.003594510 jan/02 0.003603680 jan/02 0.0642 jan/02
14 fev/02 14033.29 fev/02 0.000599244 fev/02 0.005776001 fev/02 0.0441 fev/02
15 mar/02 13254.55 mar/02 0.000967430 mar/02 0.005468920 mar/02 0.0744 mar/02
16 abr/02 13085.12 abr/02 0.005568851 abr/02 0.003282897 abr/02 0.1004 abr/02
17 mai/02 12861.43 mai/02 0.008238363 mai/02 0.025335039 mai/02 0.0255 mai/02
18 jun/02 11139.16 jun/02 0.015434200 jun/02 0.005659962 jun/02 0.0516 jun/02
      PIB i..Mes SELIC i..Mes Producao.Vidro
13 112374.8 jan/02 0.20 jan/02      73.8
14 111477.1 fev/02 0.16 fev/02      72.0
15 118444.7 mar/02 0.18 mar/02      80.7
16 120385.9 abr/02 0.19 abr/02      79.5
17 123552.5 mai/02 0.18 mai/02      82.1
18 123424.4 jun/02 0.17 jun/02      77.3
> |
```

Figura 4.15 – Tela do Programa R – Data Frame Único
Fonte: Autor (2021)

Como é possível observar na Figura 4.16, esse comando uniu todos os data frames, tendo a coluna de “Mês” repetida, portanto, essa coluna foi excluída:

```
> df_BD<-df_BD[,-c(3,5,7,9,11,13)]
> head(df_BD)
  i..Mes BOVESPA      IGP.M      INCC      IPCA      PIB SELIC Producao.Vidro
13 jan/02 12721.45 0.003594510 0.003603680 0.0642 112374.8 0.20      73.8
14 fev/02 14033.29 0.000599244 0.005776001 0.0441 111477.1 0.16      72.0
15 mar/02 13254.55 0.000967430 0.005468920 0.0744 118444.7 0.18      80.7
16 abr/02 13085.12 0.005568851 0.003282897 0.1004 120385.9 0.19      79.5
17 mai/02 12861.43 0.008238363 0.025335039 0.0255 123552.5 0.18      82.1
18 jun/02 11139.16 0.015434200 0.005659962 0.0516 123424.4 0.17      77.3
> |
```

Figura 4.16 – Tela do Programa R – Exclusão colunas “Mês” repetidas
Fonte: Autor (2021)

Por fim, a coluna “Mês” estava com o nome fora do padrão, devido ao acento presente nas bases, pois o R não identifica o acento e acaba alterando o nome da coluna. A Figura 4.17 mostra que, para corrigir isso, foi alterado o nome da coluna para a palavra “Mês”, sem a presença do acento.

```

> colnames(df_BD)[which(colnames(df_BD) == 'i..Mes')] <- 'Mes'
> head(df_BD)
  Mes BOVESPA      IGP.M      INCC      IPCA      PIB SELIC Producao.Vidro
13 jan/02 12721.45 0.003594510 0.003603680 0.0642 112374.8 0.20      73.8
14 fev/02 14033.29 0.000599244 0.005776001 0.0441 111477.1 0.16      72.0
15 mar/02 13254.55 0.000967430 0.005468920 0.0744 118444.7 0.18      80.7
16 abr/02 13085.12 0.005568851 0.003282897 0.1004 120385.9 0.19      79.5
17 mai/02 12861.43 0.008238363 0.025335039 0.0255 123552.5 0.18      82.1
18 jun/02 11139.16 0.015434200 0.005659962 0.0516 123424.4 0.17      77.3
> |

```

Figura 4.17 – Tela do Programa R – Ajuste nome coluna “Mês”
 Fonte: Autor (2021)

Após o ajuste de colunas e tamanho da base, conforme demonstrado na Figura 4.18, foi identificado que todas as colunas que serão utilizadas na correlação estavam definidas como “num”, então não precisaram ser convertidas para “chr” para que as contas estatísticas pudessem ser realizadas. A coluna “Mês”, que foi utilizada para construção dos gráficos, foi mantida.

Data	
df_BD	236 obs. of 8 variables
\$ Mes	: chr "jan/02" "fev/02" "mar/02" "abr/02" ...
\$ BOVESPA	: num 12721 14033 13255 13085 12861 ...
\$ IGP.M	: num 0.003595 0.000599 0.000967 0.005569 0.008238 ...
\$ INCC	: num 0.0036 0.00578 0.00547 0.00328 0.02534 ...
\$ IPCA	: num 0.0642 0.0441 0.0744 0.1004 0.0255 ...
\$ PIB	: num 112375 111477 118445 120386 123553 ...
\$ SELIC	: num 0.2 0.16 0.18 0.19 0.18 0.17 0.2 0.19 0.18 0.22 ...
\$ Producao.Vidro	: num 73.8 72 80.7 79.5 82.1 77.3 72.3 75 67.8 77.4 ...

Figura 4.18 – Tela do Programa R – Identificação da formatação das colunas
 Fonte: Autor (2021)

Na Figura 4.19, verifica-se que, para conferir se a tabela estava com todos os dados preenchidos, foi utilizada a função “is.na” junto à função “sum”, onde “is.na” mostra “FALSE” caso tenha dado preenchido e TRUE caso esteja vazio:

```

Console Terminal x Jobs x
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> is.na(df_BD)
  Mes BOVESPA IGP.M INCC IPCA  PIB SELIC Producao.Vidro
13 FALSE  FALSE FALSE FALSE FALSE FALSE FALSE      FALSE
14 FALSE  FALSE FALSE FALSE FALSE FALSE FALSE      FALSE
15 FALSE  FALSE FALSE FALSE FALSE FALSE FALSE      FALSE
16 FALSE  FALSE FALSE FALSE FALSE FALSE FALSE      FALSE
17 FALSE  FALSE FALSE FALSE FALSE FALSE FALSE      FALSE
18 FALSE  FALSE FALSE FALSE FALSE FALSE FALSE      FALSE
19 FALSE  FALSE FALSE FALSE FALSE FALSE FALSE      FALSE
20 FALSE  FALSE FALSE FALSE FALSE FALSE FALSE      FALSE

```

Figura 4.19 – Tela do Programa R – Conferência de elementos vazios
 Fonte: Autor (2021)

A função “sum(is.na())” mostra a soma de todos os “TRUE” encontrados. Nesse caso, como representado na Figura 4.20, a soma deu zero, o que significa que todos os dados estavam preenchidos.

```
> sum(is.na(df_BD))
[1] 0
> |
```

Figura 4.20 – Tela do Programa R – Conferência de elementos vazios por fórmula
Fonte: Autor (2021)

4.1.4 - Análise dos Dados

Após ajustada a base e conferido que todos os valores estão preenchidos, foi utilizado o pacote “Corrplot”, que facilita a manipulação de dados para análises de correlação.

Para instalação desse pacote, conforme demonstrado na Figura 4.21, foi utilizado o comando padrão para instalação de pacotes do R.

```
> install.packages("corrplot")
WARNING: Rtools is required to build R packages but is not currently installed. Please
download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/lima1/OneDrive/Documentos/R/win-library/4.1'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/corrplot_0.90.zip'
Content type 'application/zip' length 2894402 bytes (2.8 MB)
downloaded 2.8 MB

package 'corrplot' successfully unpacked and MD5 sums checked
```

Figura 4.21 – Tela do Programa R – Instalação do pacote “CORPLOT”
Fonte: Autor (2021)

Esse comando pode ser utilizado para instalação de diversos outros pacotes e permite realizar o download dentro do próprio programa. Isso permite que as novas funcionalidades fiquem disponíveis para o usuário de maneira rápida.

Após instalado o pacote, foi utilizada construída uma matriz gráfica de correlação disponível no pacote, por meio da função corrplot(cor(df_BF), method = "circle"). Onde são correlacionadas todas as colunas da base e utilizado o método “Circle”, que traz visualmente o nível de correlação entre eles.

Essa matriz, demonstrada na Figura 4.22, permitiu a identificação de forma rápida e visual da relação entre todos os valores da tabela, conforme abaixo:

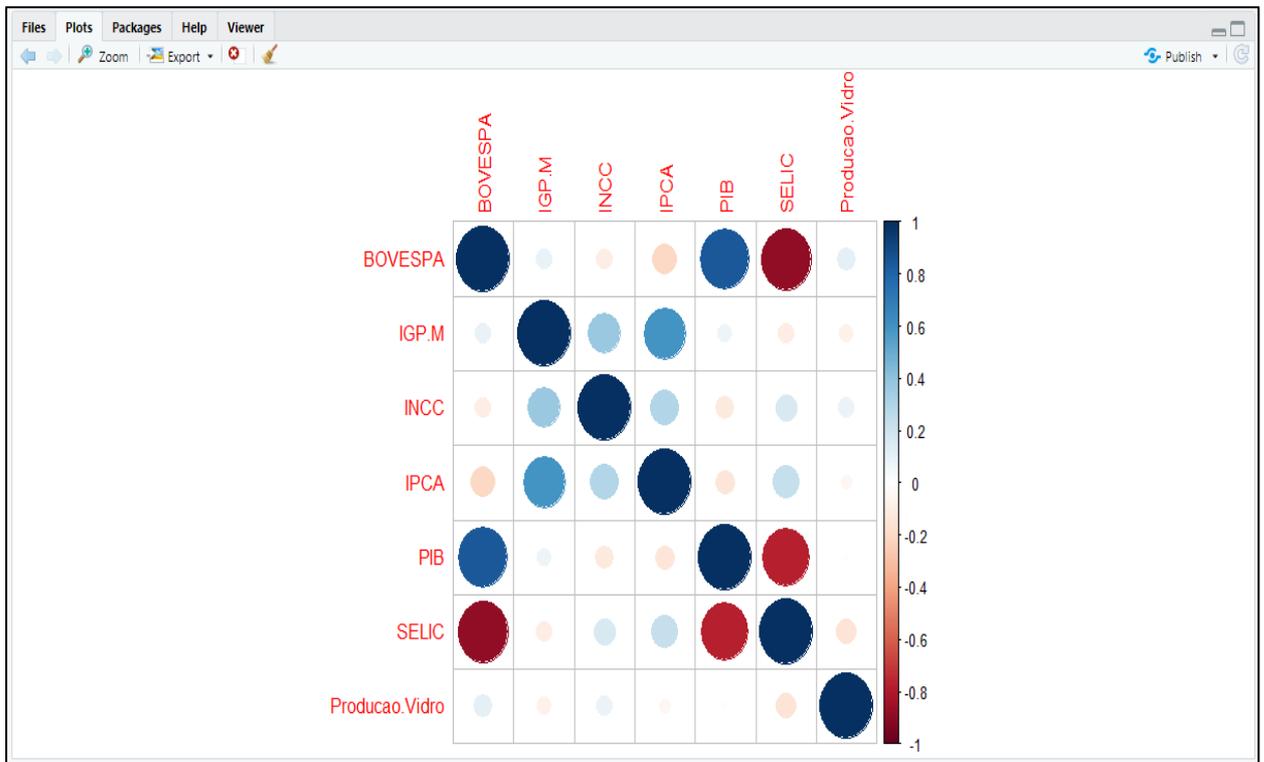


Figura 4.22 – Tela do Programa R – Gráfico Matriz de Correlação
Fonte: Autor (2021)

Como o estudo foca na Produção de Vidro, serão destacadas apenas a última linha/coluna, que já permite uma identificação de uma correlação fraca entre a produção de vidro plano e as variáveis selecionadas no estudo.

Para demonstrar de forma detalhada os resultados da matriz, foi gerada uma tabela dentro do programa, apresentada na Figura 4.23:

	BOVESPA	IGP.M	INCC	IPCA	PIB	SELIC	Producao.Vidro
BOVESPA	1.00	0.09	-0.10	-0.21	0.85	-0.88	0.11
IGP.M	0.09	1.00	0.38	0.60	0.08	-0.09	-0.07
INCC	-0.10	0.38	1.00	0.29	-0.12	0.16	0.09
IPCA	-0.21	0.60	0.29	1.00	-0.13	0.24	-0.05
PIB	0.85	0.08	-0.12	-0.13	1.00	-0.78	0.01
SELIC	-0.88	-0.09	0.16	0.24	-0.78	1.00	-0.15
Producao.Vidro	0.11	-0.07	0.09	-0.05	0.01	-0.15	1.00

Figura 4.23 – Tela do Programa R – Tabela Matriz de Correlação
Fonte: Autor (2021)

4.1.5 - Análise dos Resultados

Para evidenciar os resultados obtidos de forma gráfica, foram feitas as respectivas correlações da produção de vidro plano com cada índice. Dessa forma, é possível confrontar os dados apresentados na matriz com o gráfico de séries temporais.

4.1.5.1 - BOVESPA x Produção de Vidros

Foi criado o gráfico de dispersão por meio do comando demonstrado na Figura 4.25:

```
Console Terminal Jobs
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> ggplot(df_BD, aes(x=BOVESPA,y=Producao.Vidro))+geom_point()+ggtitle("BOVESPA x Produ
ção de vidro")
> |
```

Figura 4.24 – Tela do Programa R – Código para gráfico BOVESPA x Produção de Vidro

Fonte: Autor (2021)

Após o comando, o programa gera o gráfico apresentado na Figura 4.25:

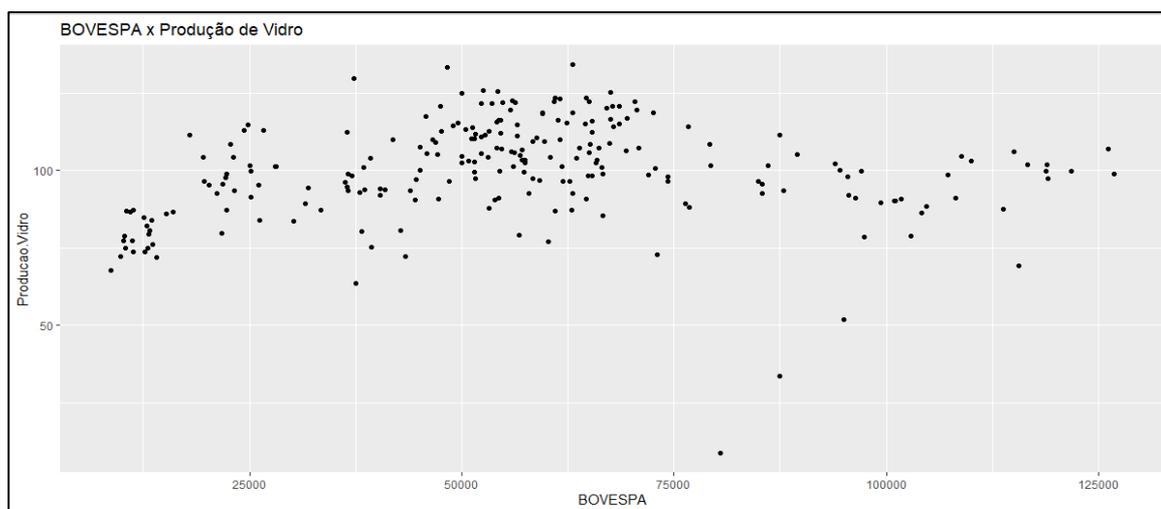


Figura 4.25 – Tela do Programa R – Gráfico de Dispersão BOVESPA x Produção de Vidro

Fonte: Autor (2021)

Pelo gráfico construído, não foi possível se obter nenhuma evidência de influência entre a variação de pontuação da Bolsa de Valores e a Produção de Vidro Plano.

Para confirmar o que foi visto de forma gráfica, a Figura 4.26 apresenta o que foi feito no cálculo de correlação dentro do R, utilizando a função “cor(x,y)”, onde x indica cada item indica uma variável que se quer relacionar:

```
Console Terminal x Jobs x
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> cor(df_BD$BOVESPA, df_BD$Producao.Vidro)
[1] 0.1138536
> |
```

Figura 4.26 – Tela do Programa R – Código para correlação BOVESPA x Produção de Vidro
Fonte: Autor (2021)

O resultado final, portanto, foi 0,11 e, segundo a tabela definida como padrão para a análise, valores compreendidos entre 0.1 e 0.3 indicam que existe uma correlação, porém fraca entre as variáveis.

Segundo Guilherme Parreira (2015): “Pelo fato de a IBOVESPA ser um índice com alta volatilidade (alta variabilidade), e com um pouco de background da área de negócios, sabe-se que é muito difícil alguma outra variável estar correlacionada com a série da IBOVESPA (caso contrário, a maioria apostaria na bolsa).”

4.1.5.2 - IGPM x Produção de Vidros

Foi criado o gráfico de dispersão por meio do comando demonstrado na Figura 4.27:

```
Console Terminal x Jobs x
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> ggplot(df_BD, aes(x=IGP.M,y=Producao.Vidro))+geom_point()+ggtitle("IGPM x Produção d
e Vidro")
> |
```

Figura 4.27 – Tela do Programa R - Código para gráfico IGPM x Produção de Vidro
Fonte: Autor (2021)

Após o comando, o programa gerou o gráfico apresentado na Figura 4.28:

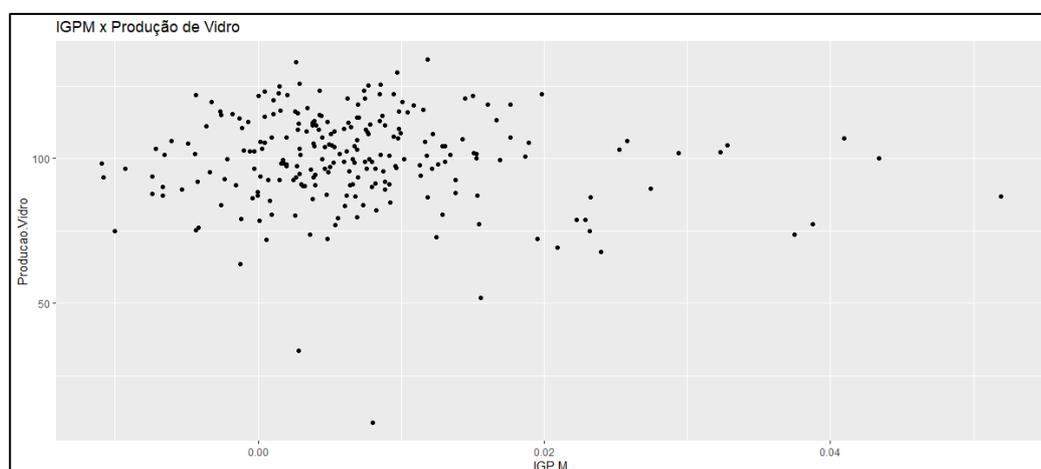


Figura 4.28 – Tela do Programa R – Gráfico de Dispersão IGPM x Produção de Vidro
Fonte: Autor (2021)

Da mesma forma que acontece com a BOVESPA, não foi possível obter nenhuma correlação analisando visualmente o gráfico. Como pode se observar, os pontos estão espalhados quase que de forma aleatória, não sendo possível identificar nenhum padrão entre os dados.

Na análise estatística feita no R, a correlação se mostrou também fraca, conforme apresentado na Figura 4.29:

```
Console Terminal x Jobs x
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> cor(df_BD$IGP.M,df_BD$Producao.Vidro)
[1] -0.07358022
> |
```

Figura 4.29 – Tela do Programa R – Código para correlação IGPM x Produção de Vidro
Fonte: Autor (2021)

Essa correlação, por sua vez, é negativa, o que indicaria que a Produção de Vidro possui um sentido contrário a variação do IGPM, porém como é um valor muito próximo de 0, a grande maioria dos autores consideram como desprezível.

4.1.5.3 - INCC x Produção de Vidros

Foi criado o gráfico de dispersão por meio do comando demonstrado na Figura 4.30:

```
Console Terminal x Jobs x
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> ggplot(df_BD, aes(x=INCC,y=Producao.Vidro))+geom_point()+ggtitle("INCC x Produção de vidro")
> |
```

Figura 4.30 – Tela do Programa R – Código para gráfico INCC x Produção de Vidro
Fonte: Autor (2021)

Após o comando, o programa gerou o gráfico apresentado na Figura 4.31:

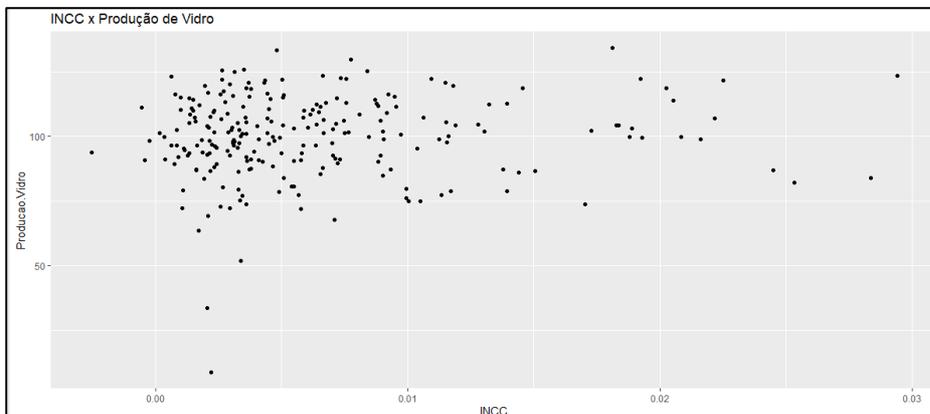


Figura 4.31 – Tela do Programa R – Gráfico de Dispersão INCC x Produção de Vidro
Fonte: Autor (2021)

O Gráfico de Dispersão mais uma vez não apresentou nenhuma relação visual capaz de identificar alguma correlação presente entre a variável, nesse caso o INCC, e a produção de vidro plano. Estatisticamente, a Figura 4.32 mostra que obteve-se a correlação:

```
Console Terminal x Jobs x
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> cor(df_BD$INCC,df_BD$Producao.Vidro)
[1] 0.0896337
> |
```

Figura 4.32 – Tela do Programa R – Gráfico de Dispersão INCC x Produção de Vidro

Fonte: Autor (2021)

Há, portanto, uma correlação positiva, indicando que possivelmente há uma mesma tendência de crescimento da produção de acordo com o INCC, porém mais uma vez está abaixo de 0,1, indicando uma correlação considerada nula pela tabela definida como base para o estudo.

4.1.5.4 - IPCA x Produção de Vidros

Foi criado o gráfico de dispersão por meio do comando demonstrado na Figura 4.33:

```
Console Terminal x Jobs x
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> ggplot(df_BD, aes(x=IPCA,y=Producao.Vidro))+geom_point()+ggtitle("IPCA x Produção de Vidro")
> |
```

Figura 4.33 – Tela do Programa R – Código para gráfico IPCA x Produção de Vidro

Fonte: Autor (2021)

Após o comando, o programa gerou o gráfico apresentado na Figura 4.34:

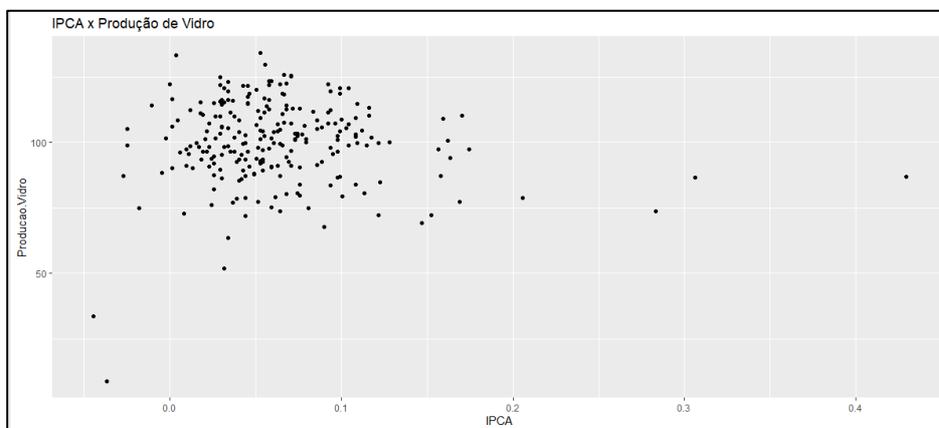


Figura 4.34 – Tela do Programa R – Gráfico de Dispersão IPCA x Produção de Vidro

Fonte: Autor (2021)

Pelo gráfico de dispersão não foi possível obter uma análise visual da correlação entre os dados, que ficam concentrados com o índice de IPCA mais baixo, variando pouco. Na análise estatística apresentada na Figura 4.35, essa correlação quase inexistente se confirma:

```
Console Terminal x Jobs x
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> cor(df_BD$IPCA,df_BD$Producao.Vidro)
[1] -0.04807197
> |
```

Figura 4.35 – Tela do Programa R – Código para correlação IPCA x Produção de Vidro
Fonte: Autor (2021)

A análise estatística comprova que praticamente não há ligação entre o IPCA e a produção de vidro plano, apenas uma tendência contrária de crescimento de um em relação a outro.

4.1.5.5 - PIB x Produção de Vidros

Foi criado o gráfico de dispersão por meio do comando demonstrado na Figura 4.36:

```
Console Terminal x Jobs x
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> ggplot(df_BD, aes(x=PIB,y=Producao.Vidro))+geom_point()+ggtitle("PIB x Produção de V
idro")
> |
```

Figura 4.36 – Tela do Programa R – Código para gráfico PIB x Produção de Vidro
Fonte: Autor (2021)

Após o comando, o programa gerou o gráfico apresentado na Figura 4.37:

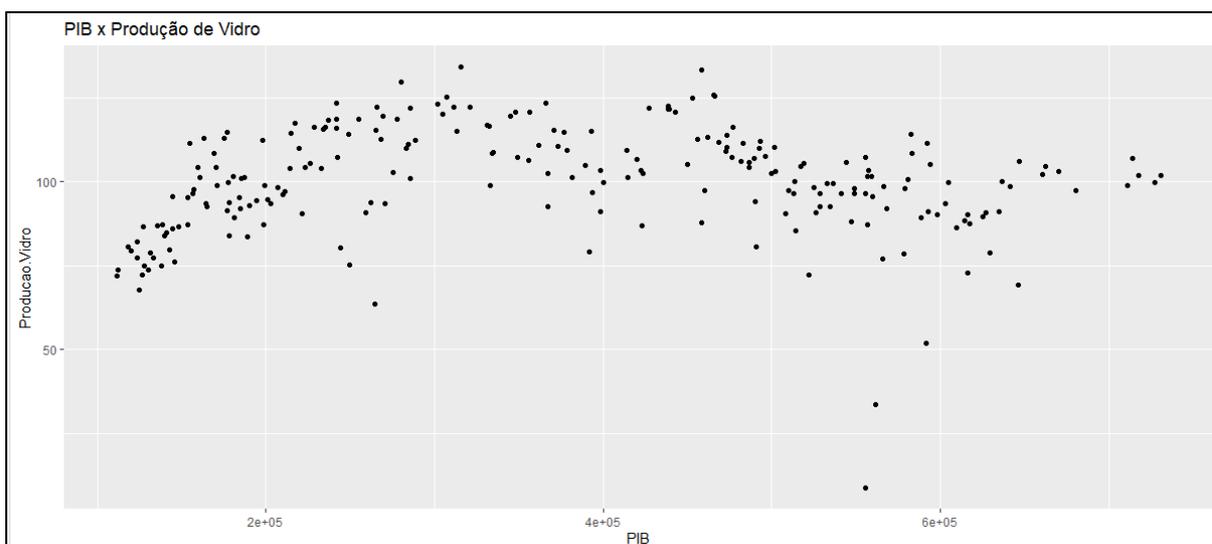


Figura 4.37 – Tela do Programa R – Gráfico de Dispersão PIB x Produção de Vidro
Fonte: Autor (2021)

O gráfico dessa vez teve pontos mais espaçados, o diferenciando dos anteriores, porém ainda assim não foi possível se obter uma relação direta entre as variáveis de forma visual.

Analisando estatisticamente, a Figura 4.38 apresenta o resultado do programa:

```
Console Terminal x Jobs x
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> cor(df_BD$PIB,df_BD$Producao.Vidro)
[1] 0.01384073
```

Figura 4.38 – Tela do Programa R – Código para correlação PIB x Produção de Vidro
Fonte: Autor (2021)

A correlação positiva indica que ambas as variáveis caminham juntas em relação ao crescimento, o que era de se esperar, uma vez que a produção de vidro faz parte da construção do valor do PIB. Por outro lado, a correlação é praticamente nula, de apenas 0,014, indicando que essa parte da indústria não influencia de forma significativa o PIB brasileiro.

4.1.5.6 - SELIC x Produção de Vidros

Foi criado o gráfico de dispersão por meio do comando demonstrado na Figura 4.39:

```
Console Terminal x Jobs x
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> ggplot(df_BD, aes(x=SELIC,y=Producao.Vidro))+geom_point()+ggtitle("SELIC x Produção de vidro")
> |
```

Figura 4.39 – Tela do Programa R – Código para gráfico SELIC x Produção de Vidro
Fonte: Autor (2021)

Após o comando, o programa gerou o gráfico apresentado na Figura 4.24:

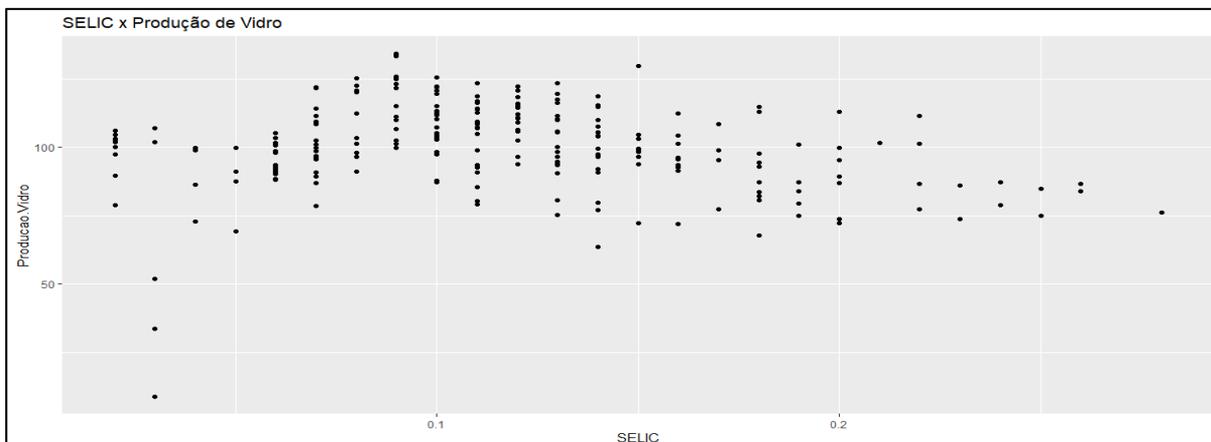


Figura 4.40 – Tela do Programa R – Gráfico de Dispersão SELIC x Produção de Vidro
Fonte: Autor (2021)

O gráfico dessa vez tinha diversos pontos em comum pois, como mencionado anteriormente, a taxa SELIC não possui uma variação mensal e sim de acordo com divulgação do Banco Central. Analisando visualmente, também não foi possível obter nenhuma relação visível entre as duas variáveis. Na análise estatística colocada na Figura 4.41, essa correlação se mostrou a mais alta entre todas as especificadas:

```

R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> cor(df_BD$SELIC, df_BD$Producao.Vidro)
[1] -0.1454942
  
```

Figura 4.41 – Tela do Programa R – Código para correlação SELIC x Produção de Vidro
Fonte: Autor (2021)

Trata-se de uma correlação mais forte que as demais, porém negativa, indicando que a Produção de Vidro e a Taxa SELIC caminham de forma oposta, porém sem uma grande influência de uma sobre a outra, sendo considerada uma correlação fraca na tabela selecionada pelo estudo.

Para reunir os resultados de forma a facilitar a visualização, foi colocado o Quadro 4.1:

Quadro 4.1 – Resultados do estudo de correlação

Variável	Correlação	Conclusão
IBOVESPA	0,11	Fraca Positiva
IGPM	-0,07	Nula
INCC	0,09	Nula
IPCA	-0,05	Nula
PIB	0,013	Nula
SELIC	-0,14	Fraca Negativa

Fonte: Autor, 2021

Conclui-se, portanto, que apesar de alguns dos índices possuírem relação direta com os principais clientes das indústrias do vidro plano, não possuem capacidade de influência de forma significativa a produção ao longo do tempo

O trabalho abre possibilidade para utilização em diversas outras situações, como por exemplo, relacionar a produção de vidro plano e de segurança com as produções de seus principais mercados clientes, como por exemplo a construção civil e a indústria automobilística.

Para isso, o estudo buscou realizar a correlação entre essas variáveis e a produção de vidro plano para comprovar como seria uma comparação de correlação forte entre as variáveis.

4.1.6 - Análise de mercados clientes

Para a fonte de informação de produção de automóveis, foi selecionada a fonte oficial da ANFAVEA (Associação Nacional dos Fabricantes de Veículos Automotores), disponibilizada no site da Associação.

Na Figura 4.42, é demonstrado que foi criado um Data Frame já com os meses selecionados, que foi possível extrair do site com o período selecionado:

```
> df_Carros<-read.csv2("Produção veículos Histórico.csv")
> head(df_Carros)
  i..Mes Producao.Veiculos
1 jan/02          116830
2 fev/02          116000
3 mar/02          143887
4 abr/02          152334
5 mai/02          144526
6 jun/02          130458
> |
```

Figura 4.42 – Tela do Programa R – Data Frame Produção de Veículos
Fonte: Autor, 2021

Para ajustar a coluna Mês, que novamente veio com o nome errado, foi feito o mesmo procedimento anterior, apresentado na Figura 4.43:.

```
> colnames(df_Carros)[which(colnames(df_Carros) == 'i..Mes')] <- 'Mes'
> head(df_Carros)
  Mes Producao.Veiculos
1 jan/02          116830
2 fev/02          116000
3 mar/02          143887
4 abr/02          152334
5 mai/02          144526
6 jun/02          130458
> |
```

Figura 4.43 – Tela do Programa R – Ajuste coluna mês em df_carros
Fonte: Autor, 2021

O próximo passo, mostrado na Figura 4.44, foi unir o Data Frame de Carros ao da Base Criada:

```
> df_BD = cbind.data.frame(df_BD, df_Carros)
> head(df_BD)
  Mes BOVESPA      IGP.M      INCC      IPCA      PIB SELIC Producao.Vidro
13 jan/02 12721.45 0.003594510 0.003603680 0.0642 112374.8 0.20          73.8
14 fev/02 14033.29 0.000599244 0.005776001 0.0441 111477.1 0.16          72.0
15 mar/02 13254.55 0.000967430 0.005468920 0.0744 118444.7 0.18          80.7
16 abr/02 13085.12 0.005568851 0.003282897 0.1004 120385.9 0.19          79.5
17 mai/02 12861.43 0.008238363 0.025335039 0.0255 123552.5 0.18          82.1
18 jun/02 11139.16 0.015434200 0.005659962 0.0516 123424.4 0.17          77.3
  Mes Producao.Veiculos
13 jan/02          116830
14 fev/02          116000
15 mar/02          143887
16 abr/02          152334
17 mai/02          144526
18 jun/02          130458
> |
```

Figura 4.44 – Tela do Programa R – União df_carros com o Data Frame principal
Fonte: Autor, 2021

Já para obtenção desses dados do Índice de Atividade da Indústria da Construção, foi utilizada também uma fonte oficial da Confederação Nacional da Indústria. Porém os dados dessa base, como evidenciado na Figura 4.45, começavam apenas a partir de janeiro de 2010, o que fez necessário um ajuste na base original para que fosse feito o estudo de correlação.

```

Console Terminal x Jobs x
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> df_Construcao<-read.csv2("Atividade Construção - Histórico.csv")
> head(df_Construcao)
  i..Mes Índice.Ativ..Construcao
1 jan/10                    50.8
2 fev/10                    52.6
3 mar/10                    55.5
4 abr/10                    53.8
5 mai/10                    56.1
6 jun/10                    54.4
>

```

Figura 4.45 – Tela do Programa R – Data Frame Índice de Atividade da Indústria de Construção
Fonte: Autor, 2021

Como os dados começavam em janeiro de 2010, foi criado um novo Data Frame Final para realizar os gráficos apenas dessas 2 variáveis e comparação delas em um período menor de tempo. Esse ajuste é representado na tela do programa por meio da Figura 4.46.

```

> df_BD_Final<-df_BD[,-c(2,3,4,5,6,7,9)]
> df_BD_Final<-tail(df_BD_Final,140)
> head(df_BD_Final)
  Mes Producao.Vidro Producao.Veiculos
109 jan/10          112.4          233395
110 fev/10          100.9          236614
111 mar/10          122.3          318011
112 abr/10          125.4          274873
113 mai/10          134.3          297374
114 jun/10          122.2          283997
>

```

Figura 4.46 – Tela do Programa R – Ajuste nos meses do Data Frame Original
Fonte: Autor, 2021

Agora para juntar os dados, foi utilizada a função “cbind”, conforme demonstrado na Figura 4.47.

```

> df_BD_Final<-cbind.data.frame(df_BD_Final,df_Construcao)
> head(df_BD_Final)
  Mes Producao.Vidro Producao.Veiculos i..Mes Índice.Ativ..Construcao
109 jan/10          112.4          233395 jan/10                    50.8
110 fev/10          100.9          236614 fev/10                    52.6
111 mar/10          122.3          318011 mar/10                    55.5
112 abr/10          125.4          274873 abr/10                    53.8
113 mai/10          134.3          297374 mai/10                    56.1
114 jun/10          122.2          283997 jun/10                    54.4
>

```

Figura 4.47 – Tela do Programa R – União do df_construcao com o Data Frame principal
Fonte: Autor, 2021

A coluna mês se repetiu, logo, foi excluída a coluna com o nome fora do padrão (coluna 4), como mostra os comandos da Figura 4.48.

```
> df_BD_Final<-df_BD_Final[,-c(4)]
> head(df_BD_Final)
  Mes Producao.Vidro Producao.Veiculos Indice.Ativ..Construcao
109 jan/10          112.4          233395             50.8
110 fev/10          100.9          236614             52.6
111 mar/10          122.3          318011             55.5
112 abr/10          125.4          274873             53.8
113 mai/10          134.3          297374             56.1
114 jun/10          122.2          283997             54.4
> |
```

Figura 4.48 – Tela do Programa R – Ajuste coluna “Mês” no Data Frame Final
Fonte: Autor, 2021

Para uma primeira análise, foram feitos gráficos com as 3 variáveis ao longo do tempo, sendo a Produção Mensal de Vidro na Figura 4.49, a Produção Mensal de Veículos na Figura 4.50 e o Índice Nacional da Construção representado na Figura 4.51.

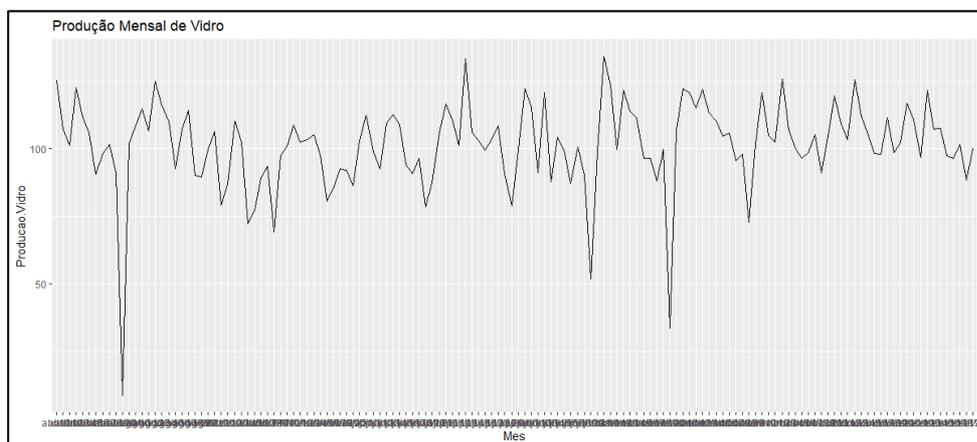


Figura 4.49 – Tela do Programa R – Gráfico Produção Mensal de Vidro
Fonte: Autor, 2021

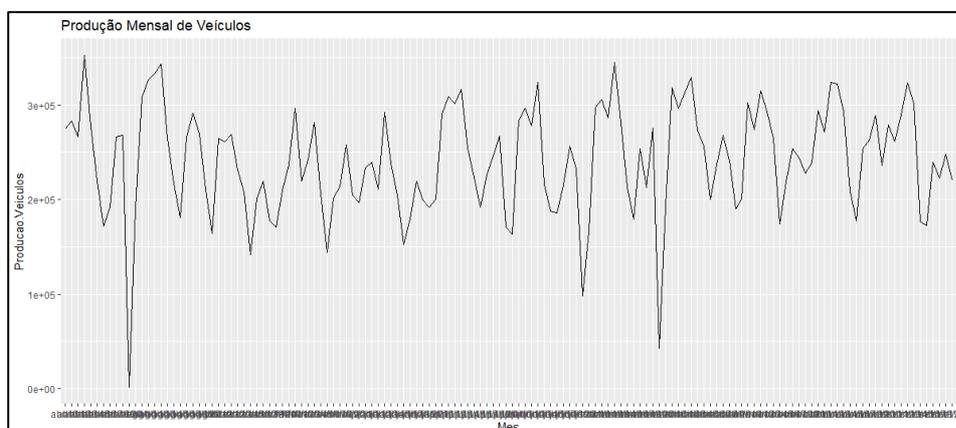


Figura 4.50 – Tela do Programa R – Gráfico Produção Mensal de Veículos
Fonte: Autor, 2021

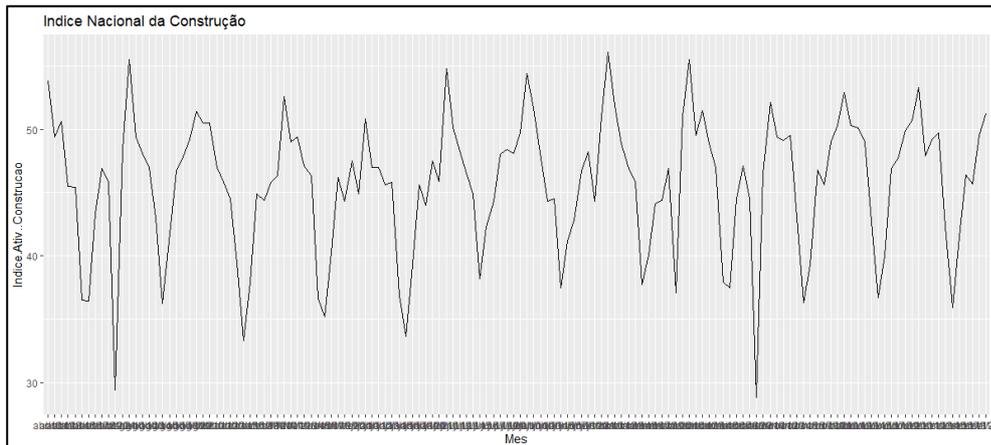


Figura 4.51 – Tela do Programa R – Gráfico Índice de Atividade da Indústria de Construção
Fonte: Autor, 2021

Já foi possível observar visualmente que os dados tinham algum tipo de correlação, o que foi verificado por meio da matriz de correlação e da fórmula estatística. Dessa forma, foi novamente montada a matriz de correlação, excluindo-se a coluna “Mês”, que não era necessária, após a base ajustada com a mesma quantidade de dados para cada variável, e começando e terminando na mesma data. A Figura 4.52 apresenta a exclusão da coluna e a Figura 4.53 a base já com o ajuste.

```
> df_cor_final<-df_BD_Final[,-c(1)]
> head(df_cor_final)
  Producao.Vidro Producao.Veiculos Indice.Ativ..Construcao
109           112.4           233395                50.8
110           100.9           236614                52.6
111           122.3           318011                55.5
112           125.4           274873                53.8
113           134.3           297374                56.1
114           122.2           283997                54.4
```

Figura 4.52 – Tela do Programa R – Exclusão da coluna “Mês”
Fonte: Autor, 2021

```
> df_cor_final<-df_BD_Final[,-c(1)]
> head(df_cor_final)
  Producao.Vidro Producao.Veiculos Indice.Ativ..Construcao
109           112.4           233395                50.8
110           100.9           236614                52.6
111           122.3           318011                55.5
112           125.4           274873                53.8
113           134.3           297374                56.1
114           122.2           283997                54.4
> |
```

Figura 4.53 – Tela do Programa R – Base para correlação ajustada
Fonte: Autor, 2021

Com a base ajustada, foi feita a Base para utilização na Matriz de Correlação, apresentada na Figura 4.54.

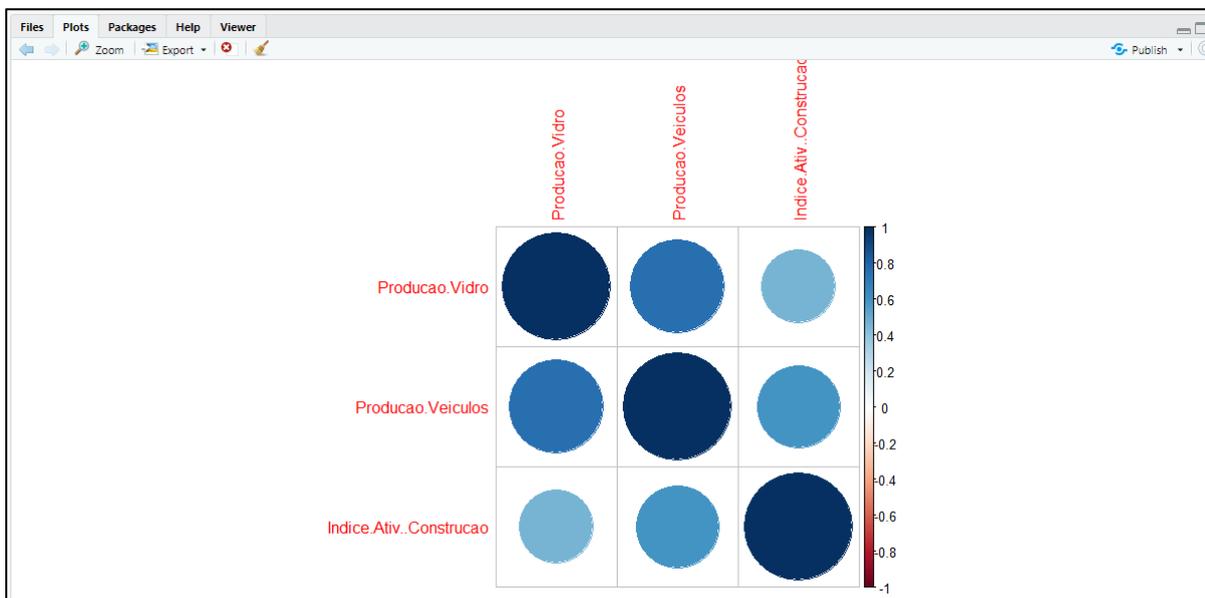


Figura 4.54 – Tela do Programa R – Gráfico Matriz de Correlação Final
Fonte: Autor, 2021

Já para se obter os dados detalhados, foi criada uma tabela, conforme feito na análise de correlação anterior, buscando trazer no detalhe os resultados obtidos, além da análise gráfica e visual, trazida na Figura 4.55.

```
> res <- cor(df_cor_final)
> round(res, 2)
              Producao.Vidro Producao.Veiculos
Producao.Vidro             1.00             0.76
Producao.Veiculos          0.76             1.00
Indice.Ativ..Construcao    0.47             0.60
              Indice.Ativ..Construcao
Producao.Vidro             0.47
Producao.Veiculos          0.60
Indice.Ativ..Construcao    1.00
> |
```

Figura 4.55 – Tela do Programa R – Tabela Matriz de Correlação
Fonte: Autor, 2021

Foi possível observar que houve uma correlação bem mais forte entre a Produção de Vidros Planos e de Segurança e a Produção de Veículos e o Índice da Atividade da Construção.

4.1.6.1 - Produção de Veículos x Produção de Vidros

Foi criado o gráfico de dispersão por meio do comando demonstrado na Figura 4.56:

```
Console Terminal x Jobs x
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> ggplot(df_cor_final, aes(x=Producao.Veiculos,y=Producao.Vidro))+geom_point()+ggtitle
("Produção de Veículos x Produção de Vidro")
> |
```

Figura 4.56 – Tela do Programa R – Gerar Gráfico de Produção Veículos x Produção de Vidro
Fonte: Autor, 2021

Após o comando, o programa gerou o gráfico apresentado na Figura 4.57:

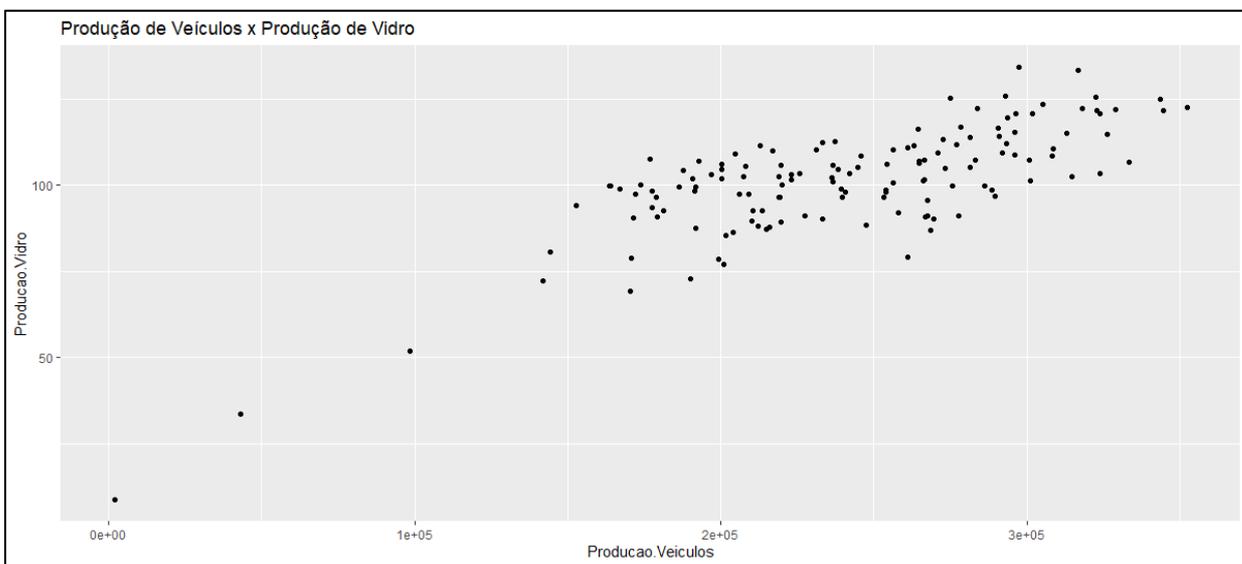


Figura 4.57 – Tela do Programa R – Gráfico de Produção de Veículos x Produção de Vidro
Fonte: Autor, 2021

No gráfico de dispersão já foi possível observar que há uma provável relação entre a produção de veículos com a produção de vidros, que foi de fato verificado por meio da análise estatística de correlação apresentada na Figura 4.58.

```
Console Terminal x Jobs x
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/
> cor(df_cor_final$Producao.Veiculos,df_cor_final$Producao.Vidro)
[1] 0.7591578
> |
```

Figura 4.58 – Tela do Programa R – Código correlação Produção de Veículos x Produção de Vidro
Fonte: Autor, 2021

Como era esperado, por se tratar de um dos principais clientes do setor vidreiro, há uma correlação forte entre a Produção de Veículos e a Produção de Vidros Planos e de Segurança.

4.1.6.2 - Índice de Atividade da Indústria de Construção x Produção de Vidros

Foi criado o gráfico de dispersão por meio do comando demonstrado na Figura 4.59:

```
Console Terminal x Jobs x  
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/  
> ggplot(df_cor_final, aes(x=Indice.Ativ..Construcao,y=Producao.vidro))+geom_point()+g  
gtitle("Atividade Cosntrução Civil x Produção de vidro")  
> |
```

Figura 4.59 – Tela do Programa R - Gerar Gráfico de IAIC x Produção de Vidro
Fonte: Autor, 2021

Após o comando, o programa gerou o gráfico apresentado na Figura 4.60:

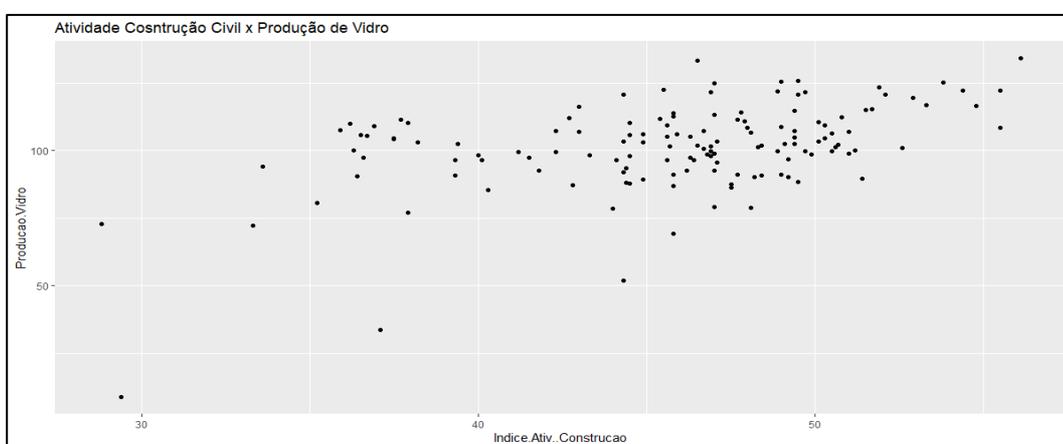


Figura 4.60 – Tela do Programa R – Gráfico de IAIC x Produção de Vidro
Fonte: Autor, 2021

Apesar de não ser tão evidente quanto a produção de veículos, também foi possível observar que havia uma relação entre o índice de construção e a produção de vidros planos e de segurança. Estatisticamente, fica formulado como demonstrado na Figura 4.61:

```
Console Terminal x Jobs x  
R 4.1.1 · C:/Users/lima1/OneDrive/Área de Trabalho/TCC Pós/Trabalho Final MB3B/  
> cor(df_cor_final$Indice.Ativ..Construcao,df_cor_final$Producao.vidro)  
[1] 0.4696031  
> |
```

Figura 4.61 – Tela do Programa R – Correlação IAIC x Produção de Vidro
Fonte: Autor, 2021

É correlação positiva, ou seja, ambos os índices caminham da mesma forma em relação a crescimento e queda, porém é uma correlação considerada moderada, ou seja, não é tão evidente quanto a produção de veículos, porém é bem mais relevante que os índices apresentados na primeira parte do estudo.

CAPÍTULO 5

Conclusão e Trabalho Futuros

5.1 - Conclusão

Após todas as análises realizadas no capítulo anterior, primeiramente foi possível verificar que de fato o programa selecionado se mostrou eficiente como uma forma detalhada, confiável e rápida para se realizar uma análise de correlação envolvendo uma base de dados relativamente grande.

Por meio do R, foi possível importar os dados de forma prática e criar data frames para tratá-los e manipulá-los de maneira rápida, para que fossem utilizados de forma eficiente. Além disso, a possibilidade de utilização de pacotes criados por usuários externos, de fato facilitou a construção do estudo, por meio das visualizações gráficas que foram utilizadas na primeira parte das análises. Para os cálculos, o programa já possui uma função direcionada a fazer a correlação entre as variáveis que foram utilizadas, fazendo esse cálculo de forma extremamente rápida e trazendo o resultado de uma correlação entre 20 anos de indicadores em poucos segundos, para que seja analisada junto a tabela definida, indicando o nível da correlação.

Em relação ao objetivo inicial do estudo, os resultados demonstraram que, diferentemente do esperado, as 7 variáveis econômicas selecionadas não possuem uma correlação forte com a produção de vidro plano e de segurança.

Um dos indicadores, o IBOVESPA, já era esperado que não tivesse nenhuma relação, uma vez que se trata da variação de certas empresas na bolsa de valores de São Paulo. Caso fosse possível obter uma correlação entre esse índice e alguma empresa, a maioria dos investidores direcionaria seus investimentos para ela ou retiraria, de acordo com seu aumento ou diminuição ao longo do tempo.

O IGP-M, apesar de possuir dentro de seu cálculo, 10% de um índice relacionado à construção civil, também não demonstrou correlação com a produção de vidro plano, uma vez que está mais relacionado aos preços que finais repassados ao consumidor. Caso fosse feita uma correlação entre o IGP-M e o número de imóveis alugados no Brasil, provavelmente a correlação seria forte.

Era esperado que o INCC demonstrasse uma correlação pelo menos moderada com a produção de vidro plano e de segurança, porém também não foi o observado nas análises. Essa ausência de uma correlação mais forte provavelmente se deve ao fato desse índice estar mais relacionado so valores e custos de materias na construção civil. É provável que um estudo de correlação entre esse índice e o os gastos mensais das empresas de construção civil resulte em uma correlação forte.

O IPCA, por sua vez, está relacionado diretamente com a variação de preços que afetam direamente o consumidor. Era esperado algum tipo de correlação devido a esse índice demonstrar o poder de compra do consumidor, que poderia impactar de alguma forma da produção de vidro, porém não foi o que aconteceu. É possível que esse índice afete de alguma forma as vendas de vidro, porém não a produção.

O Produto Interno Bruto, como já mencionado nos capítulos anteriores, é a soma de tudo o que é produzido no Brasil ao longo do tempo. A correlação entre a produção de vidro e esse índice demonstrou ser positiva, o que era esperado, uma vez que a mesma está presente dentro do PIB, porém quase nula, o que indica que não possui uma participação capaz de influenciar de forma forte a variação dele.

A taxa de juros, por sua vez, facilita a compra de veículos e a aquisição de terrenos e imóveis para a construção civil, porém na linha final não foi identificada uma correlação forte com a produção de vidro plano. Assim como mencionado com outras variáveis, é possível que uma correlação desse índice com a venda de vidro plano seja forte.

Por outro lado, nenhum dos indicadores indicou correlação 0, o que mostra que há alguma relação, mesmo que desconsiderável, entre as variáveis, porém não suficiente para que uma impacte a outra. Provavelmente isso se deve ao fato de o estudo analisar a produção de vidro plano e de segurança ao longo do tempo, o que depende de fatores mais ligados à previsão de demanda e projeção de faturamento das empresas ao longo do tempo, diferentemente das vendas desse mesmo segmento, que talvez demonstrem uma correlação maior com as variáveis econômicas.

Por fim, o trabalho buscou realizar uma análise entre duas variáveis que em teoria teriam uma correlação mais forte com a produção de vidro plano e de segurança, que são a produção de veículos e o Índice Nacional de Construção Civil, que traz o nível de atividade de obras no Brasil ao longo dos meses. Em relação a produção de veículos, o índice de fato demonstrou uma correlação forte, uma vez que a produção de veículos depende da fabricação do vidro plano, que é recortado e vendido as indústrias secundárias, que relizam os tratamentos

necessários e vendem as montadoras de forma contínua. Já em relação a construção civil, o índice de correlação se mostrou moderado, o que também faz sentido, uma vez que nem todas as obras demandam o vidro plano no momento de sua construção e a produção desse vidro é feita visando obras em momentos futuros, então há uma correlação, porém não tão forte quanto a relacionada a produção de veículos.

5.2 – Trabalhos Futuros

Como o trabalho demonstrou que é possível realizar uma análise de correlação entre quaisquer variáveis que possuam bases de dados disponíveis de forma rápida, ficam algumas sugestões de estudos futuros que podem ajudar as empresas a direcionar seus investimentos ao longo do tempo:

- **Estudo de correlação entre a venda de vidro plano e de segurança e os indicadores apresentados nesse estudo:** Como as empresas não divulgam informações relativas as vendas, seria interessante que utilizassem internamente esse estudo a fim de verificar a correlação entre suas vendas e os índices citados.
- **Estudo de Correlação entre IGP-M e a quantidade de imóveis alugados no Brasil:** Como o índice é geralmente associado e utilizado para os reajustes anuais de aluguel, provavelmente será obtido um índice de correlação forte.
- **Estudo de Correlação entre o INCC e gastos mensais de construção civil:** Como o índice apresentar a variação dos custos da construção civil ao longo do tempo, provavelmente possuirá uma correlação forte com os gastos mensais das construtoras e empreiteiras.
- **Análises de dados utilizando o software R:** Como demonstrado no estudo, o software é um programa de linguagem de fácil aprendizado, código aberto e com diversas funções que permite realizar análises de dados de forma rápida, prática e confiável. Empresas podem utilizá-lo sem custo para estudos internos, o que evita custos com programas pagos que trazem os mesmos resultados.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] LIBERAL, Tarciana. Séries, Tabelas e Gráficos. Disponível em: <<http://www.de.ufpb.br/~tarciana/CPEI/Aula2>>. Acesso em 19 set 2021.
- [2] BRITTO, Dalson *et al.* Desvendando os Mistérios do Coeficiente de Correlação de Pearson: O retorno, **Revistas USP**, N. 8, pp.66-95, 2014
- [3] HERNANDÉZ, Juan Diego Lalinde *et al.* Sobre el uso adecuado del coeficiente de correlación de Pearson: definición, propiedades y suposiciones, **Archivos Venezolanos de Farmacología y Terapéutica**, N.5, vol. 37, 2018
- [4] SOUSA, Áurea. Diagrama de dispersão, correlação e regressão linear. **Correio dos Açores**, pp16, 2019
- [5] MARTINS, Maria. Coeficiente de correlação amostral, **Rev. Ciência Elem.**, V2(02):042, 2014
- [6] BUSSAB, Wilton; MORETTIN, Pedro. Estatística Básica. São Paulo: Editora Saraiva, 2017.
- [7] Taxa SELIC. Disponível em: <<https://www.bcb.gov.br/controleinflacao/taxaselic>>. Acesso em 18 ago. 2021.
- [8] GUILHERME, Guilherme. Ações de varejo e construção serão beneficiadas com redução da Selic. Disponível em: <<https://invest.exame.com/me/varejo-e-construcao-serao-beneficiadas-com-reducao-da-selic>>._Acesso em 18 ago 2021.
- [9] Índice Nacional de Preços ao Consumidor. Disponível em: <<https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplio.html?=&t=o-que-e>>. Acesso em 19 ago 2021.

[10] ICMS. Disponível em: <<https://portal.fazenda.sp.gov.br/acessoinformacao/Paginas/ICMS.aspx>>. Acesso em 19 ago 2021.

[11] ICMS. Disponível em: <<http://www.portaltributario.com.br/tributos/icms.html>>. Acesso em 19 ago 2021.

[12] Produto Interno Bruto – PIB. Disponível em: <<https://www.ibge.gov.br/explica/pib.php>>. Acesso em 22 ago 2021.

[13] O que é e como funciona o índice geral de preços do mercado. Disponível em: <<https://www.moneytimes.com.br/igpm-o-que-e-e-como-funciona-o-indice-geral-de-precos-do-mercado/>>. Acesso em 22 ago 2021.

[14] IGPM Setembro 2021. Disponível em: <<https://portal.fgv.br/noticias/igpm-setembro-2021>>. Acesso em 22 ago 2021.

[15] Entenda o que é INCC e como é calculado. Disponível em: <<https://www.tarjab.com.br/blog/mercado-imobiliario/entenda-o-que-e-incc-e-como-e-calculado/>>. Acesso em 23 ago 2021.

[16] Ibovespa. Disponível em: <https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm>. Acesso em 23 ago 2021.

[17] O que é o ibovespa, entenda o índice da bolsa brasileira. Disponível em: <<https://invest.ename.com/guia/o-que-e-o-ibovespa-entenda-o-indice-da-bolsa-brasileira>>. Acesso em 23 ago 2021.

[18] What does effect size tells you? Disponível em: <<https://www.simplypsychology.org/effect-size.html>>. Acesso em 09 set 2021.

[19] Correlação e Regressão. Disponível em: <https://rpubs.com/rcleoni/cor_reg>. Acesso em 10 out 2021.

- [20] ROCHA, David. Sobre Correlações e visualizações de matrizes de correlação no R. Disponível em: https://rstudio-pubs-static.s3.amazonaws.com/437792_df39a5ff0a55491fb71f0f4a0f5cd0bf.html>. Acesso em 10 out 2021.
- [21] Introdução ao Software R. Disponível em: <https://operdata.com.br/blog/introducao-ao-software-r/>>. Acesso em 27 set 2021.
- [22] Tipos de Vidros e suas aplicações. Disponível em: <https://vidrado.com/noticias/mercado-e-negocios/tipos-de-vidros-e-suas-aplicacoes/>>. Acesso em 14 set 2021.
- [23] Gráfico de Dispersão. Disponível em: <https://www.fm2s.com.br/grafico-de-dispersao/>>. Acesso em 16 set 2021.
- [24] Silva. Tipos de Variáveis. Disponível em: <http://leg.ufpr.br/~silvia/CE055/node8.html>>. Acesso em 12 set 2021.
- [25] A história do vidro: saiba como o material foi descoberto. Disponível em: <https://oglobo.globo.com/economia/imoveis/a-historia-do-vidro-saiba-como-material-foi-descoberto-21591312>>. Acesso em 16 set 2021.
- [26] Você sabe como surgiu o vidro? Disponível em: <https://vidroimpresso.com.br/noticia-setor-vidreiro/voce-sabe-como-surgiu-o-vidro>>. Acesso em 16 set 2021.
- [27] Sondagem da construção consolida tendência de recuperação do setor. Disponível em: https://cbic.org.br/es_ES/sondagem-da-construcao-consolida-tendencia-de-recuperacao-do-setor/>. Acesso em 19 set 2021.
- [28] GUERRA, Bruno. O que é preparação de dados e qual sua importância? Disponível em: <https://blog.in1.com.br/o-que-e-preparacao-de-dados-e-qual-sua-importancia>> Acesso em 27 set 2021

- [29] PARREIRA, Guilherme. Coeficiente de Correlação de Pearson. Disponível em: <<https://gpestatistica.netlify.app/blog/correlacao/>>. Acesso em 16 set 2021
- [30] GUIMARÃES, Paulo Ricardo. Análise de Correlação e medidas de associação. Disponível em: <<https://docs.ufpr.br/~jomarc/correlacao.pdf>>. Acesso em 29 set 2021
- [31] CAMPOS, Geraldo. Estatística Prática para Docentes e Pós-Graduandos. Disponível em: <https://www.forp.usp.br/restauradora/gmc/gmc_livro/gmc_livro_cap02.html>. Acesso em 29 set 2021
- [32] “Série Histórica IPCA”, Disponível em: <<http://www.ipeadata.gov.br/ExibeSerie.aspx?stub=1&serid37796=37796&serid364%2082=36482>>. Acesso em 12 out 2021
- [33] “Série Histórica SELIC”, Disponível em: <<http://www.ipeadata.gov.br/exibeserie.aspx?serid=38402>>. Acesso em 12 out 2021
- [34] “Série Histórica PIB”, Disponível em: <<http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=521274780&module=M>>. Acesso em 12 out 2021
- [35] “Série Histórica IBOVESPA”, Disponível em: <https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-ibovespa-ibovespa-estatisticas-historicas.htm>
- [36] “Série Histórica INCC”, Disponível em: <<https://portalibre.fgv.br/incc>>. Acesso em 12 out 2021
- [37] “Série Histórica Produção de Vidro”, Disponível em: <<https://sidra.ibge.gov.br/tabela/3650>>. Acesso em 20 out 2021.