

## RICARDO TOMPSON FIORAVANTE

### Analisando Dados com a Linguagem de Programação Python

Maringá-PR

09/2021

#### Resumo

Os avanços tecnológicos proporcionaram à humanidade a expansão da utilização de dados. Atualmente, eles são provenientes de diversas fontes distintas. Uma das consequências desse fato, é que um grande volume de dados é armazenado e isso traz a necessidade de desenvolvimento/utilização de ferramentas capazes de analisá-los e transformá-los em informações sintetizadas que possuam valor para pessoas e/ou empresas. Este trabalho tem como objetivo estudar e desenvolver um produto de software capaz de analisar uma base de dados armazenada em um arquivo no formato '.CSV', por meio da linguagem de programação Python, bem como algumas de suas notórias bibliotecas, como Pandas, e extrair insights relevantes sobre os dados analisados, de maneira automatizada. Os requisitos para desenvolvimento do software foram levantados conforme preconiza a metodologia de desenvolvimento ágil Scrum, ou seja, por meio de histórias de usuários. Posteriormente, elas foram organizadas e priorizadas através de artefatos Scrum, como o Backlog do Produto, o Backlog da Sprint e o Gráfico de Burndown. Após a elicitação dos requisitos, foi desenvolvido o software utilizando a linguagem de programação Python, que possibilitou a análise e a formulação de conclusões a respeito da base de dados, como a obtenção de algumas medidas estatísticas, tais como correlação, além da plotagem de gráficos que viabilizaram a visualização de informações muitas vezes ocultas no conjunto de dados.

**Palavras-chaves:** Python. Pandas. Matplotlib. Seaborn. Scrum.

## 1. INTRODUÇÃO

Softwares estão presentes em quase todas as rotinas da nossa vida. Eles são elementos indispensáveis para o progresso humano, o que implica no aumento do interesse por eles, em seus recursos, em suas soluções. Esse interesse tem crescido de maneira significativa recentemente.

Como a humanidade está em constante desenvolvimento, novas demandas surgem a cada dia e um tema atual é a análise de dados. Segundo Faceli (p. 11), o volume de dados cresce exponencialmente e estima-se que a quantidade de dados produzidos no mundo dobra a cada dois anos, sendo grande parte deles armazenada em banco de dados relacionais.

Para utilizar essas informações de forma positiva, é preciso ter conhecimento inerente para lidar efetivamente com elas. Manipular dados é uma atividade essencial e conhecer suas características e suas implicações é de suma importância na sociedade atual. Serviços como analytics estão sendo cada vez mais inseridos no portfólio das empresas.

Porém, transformar dados brutos em informações úteis tem sido um grande desafio. Visando contornar esses obstáculos, novas formas, novas ferramentas para manipulação de dados e elaboração de relatórios surgiram ao longo dos anos, como algumas bibliotecas da linguagem de programação Python, que buscam auxiliar o ser humano nessas tarefas. Com elas, informações ocultas presentes nos dados podem se revelar. No entanto, muitos desses recursos ainda são inexplorados. Assim, surgem algumas questões: Como Python pode auxiliar na exploração de um determinado conjunto de dados armazenado em um arquivo ‘CSV’? Quais bibliotecas do Python podem auxiliar nessas tarefas? Quais tipos de gráficos são utilizados para representar visualmente os dados? Quais medidas estatísticas são comumente empregadas nas análises?

Com o intuito de abordar algumas dessas ferramentas e analisar suas contribuições para a análise de dados, objetivos foram definidos para convergir a esse propósito, sendo: Objetivo geral e objetivos específicos.

## 1.2 OBJETIVOS

A seguir, são expostos o objetivo geral e os objetivos específicos, formulados para abordar o problema levantado neste trabalho.

- **1.2.1 OBJETIVO GERAL:** O objetivo geral deste TCC foi desenvolver um software capaz de manipular dados armazenados em um arquivo ‘CSV’ (mencionado no apêndice D) e extrair computacionalmente insights importantes (como algumas medidas estatísticas) sobre a base de dados, de maneira prática, automática e com agilidade.
- **1.2.2 OBJETIVOS ESPECÍFICOS:** Revisar a literatura pertinente para obter embasamento sobre alguns conceitos, tipos de gráficos e definições sobre algumas medidas estatísticas, através de pesquisa bibliográfica realizada em livros e artigos científicos; Levantamento de alguns requisitos de usuários para desenvolvimento do software; Organização e priorização dos requisitos; Escrita e interpretação do código-fonte capaz de satisfazer os requisitos.

## 2. JUSTIFICATIVA

A ferramenta mais utilizada para análise de dados ainda é o Microsoft Excel, mas aprender corretamente muitos de seus recursos demanda tempo e esforço. Geralmente, a elaboração de fórmulas lógicas/matemáticas/estatísticas no Excel e uso de seus recursos mais avançados requer conhecimentos que muitos não possuem, dificultando sua utilização. Além do mais, pode ocorrer dos usuários o

utilizarem incorretamente, como por exemplo, inserindo fórmulas incorretas para efetuar determinados cálculos.

Muitas ferramentas de análise de dados possuem algumas licenças grátis (como versões para estudantes, do Microsoft Excel), mas em geral são recursos pagos, que geram custos para os adquirentes do produto. Além disso, as licenças grátis não disponibilizam todos os recursos das licenças pagas. Já Python é uma linguagem de programação disponibilizada completa e totalmente gratuita.

A linguagem de programação Python e algumas de suas bibliotecas são ferramentas práticas que automatizam a tarefa de análise de dados, sem a necessidade de dedução de fórmulas por parte dos usuários, que contribuem para reduzir o tempo gasto nas análises e evitam erros nos cálculos efetuados.

Além disso, disponibilizam informações de forma instantânea (quando os usuários precisam). Apesar de ser uma linguagem interpretada, o tempo de execução do software escrito com a linguagem Python é satisfatório.

Algumas bibliotecas para Python são capazes de plotar gráficos visuais que proporcionam aos usuários compreensão mais direta dos dados quando comparada com conjunto de números, revelando informações muitas vezes ocultas.

Dessa forma, essas ferramentas contribuem para o aumento da performance das pessoas (aspecto muito discutido no âmbito corporativo), além de influenciar diretamente na tomada de decisões/ações fundamentadas a partir da análise dos dados.

Exposto a relevância, inicia-se os demais itens relacionados a este trabalho, com detalhamento das demais etapas.

### 3. ESPECIFICAÇÕES INICIAIS DO SOFTWARE

Dados do tipo planilha (tabulares) representam boa parte das informações armazenadas em banco de dados relacionais. Além disso, uma grande quantidade de conjunto de dados não-estruturados pode ser convertida para o formato estruturado (McKinney, p. 20).

Os arquivos ‘CSV’ são sequências de caracteres. Eles armazenam diversos registros dissociados por quebras de linhas, em que cada registro ocupa uma linha. Os registros possuem campos (um ou mais deles), separados por um delimitador (Character-separated), que podem ser a vírgula, o ponto-vírgula e o tab. Os arquivos separados por vírgula ou ponto-vírgula geralmente têm a extensão ‘CSV’, enquanto os separados por tab, a extensão ‘TSV’.

Python surgiu em 1991, e atualmente é uma das linguagens de programação mais populares do mundo, contando com vastas e ativas comunidades. É uma linguagem de programação considerada simples, sobretudo pela sua sintaxe clara. Apesar dessa característica (simples), ela é uma linguagem com recursos extraordinários. É clara, é objetiva. (Menezes, p. 26)

Ao longo dos anos, as comunidades desenvolveram bibliotecas poderosas para Python, como Pandas e Matplotlib, que tornou Python numa rica alternativa para análise de dados, inclusive para tratamento de arquivos ‘CSV’.

O software utiliza algumas das bibliotecas do/para Python, como Pandas, Matplotlib e Seaborn, para facilitar a manipulação de dados e a transformação de dados brutos do arquivo ‘CSV’ em informações compiladas. Com essas poderosas bibliotecas, extrai-se medidas estatísticas e plotam-se gráficos visuais que fornecem insights importantes sobre os dados.

Gráficos são recursos visuais que facilitam a análise e são muito utilizados principalmente na elaboração de dashboards.

Além dos gráficos, as bibliotecas possibilitam extração de medidas estatísticas, como medidas de frequência, medidas de espalhamento e de localidade, sendo uma

forma muito eficaz para exploração de dados e extração de informações, bem como resumir de maneira quantitativa as características relevantes dos dados.

### **3.1 Escopo do Produto**

Software desenvolvido no trabalho de conclusão do curso de Engenharia de Software, com o intuito de possibilitar que os usuários possam efetuar consultas, visualizar informações e obter insights com facilidade, tendo como base um arquivo com extensão ‘CSV’. O programa foi desenvolvido para atender essa finalidade específica.

Levou-se em consideração a dificuldade dos usuários na utilização de outras ferramentas similares, custos de aquisição de aplicações semelhantes e adversidades enfrentadas por usuários na elaboração de fórmulas lógicas/matemáticas/estatísticas. Os detalhes do escopo estão descritos no apêndice A.

### **3.2 PÚBLICO-ALVO**

Usuários que almejam manipular os dados do arquivo ‘CSV’ utilizado como base de dados e extrair computacionalmente algumas informações.

### **3.3 Definições, Acrônimos e Abreviações**

No decorrer do documento, algumas abreviações serão utilizadas, explicando os conceitos de termos importantes que serão citados no decorrer do documento de desenvolvimento de software, conforme a seguir:

Sigla	Significado	Definição
HU	História de Usuário	Descrição suscinta de uma funcionalidade, sob a perspectiva do usuário
[XXX]	<i>Identificador da história de usuário</i>	Identificador único de cada história de usuário

**Quadro 1:** Abreviações e Definições. **Fonte:** autor

### 3.4 Convenções

Os requisitos foram levantados considerando as práticas da metodologia de desenvolvimento ágil Scrum. Assim, foram elicitados por meio de Histórias de Usuários, com a seguinte convenção:

Identificador	Nome	Descrição
[HU[<<identificador>>]]	<<nome>>	SENDO<usuário> POSSO<funcionalidade> POIS ASSIM<vantagem>

**Quadro 2:** Histórias de usuários. **Fonte:** autor

Exemplo:

Identificador	Nome	Descrição
[HU001]	Consultar estatísticas	SENDO um usuário com intenção de extrair informações da base de dados (arquivo 'CSV') abordado  POSSO consultar as estatísticas de alguns atributos  POIS ASSIM posso visualizar informações sobre os atributos.

**Quadro 3:** Exemplo de histórias de usuários. **Fonte:** autor

Na história de usuário acima utilizada como exemplo, [HU001] denota que ela está localizada numa seção denominada ‘Histórias de Usuários’, cujo identificador é [HU001]. HU é uma abreviação para ‘História de Usuário’; 001 significa que é a primeira levantada e o nome da história de usuário é “Consultar estatísticas”.

Para evitar repetição desnecessária em requisitos de um mesmo grupo de usuários, foi utilizado um contexto, conforme descrito a seguir:

Contexto
SENDO<ator que se repete em algumas histórias de usuário>

**Quadro 4:** Definição Contexto. **Fonte:** autor

Exemplo:

Contexto
SENDO Usuário com intenção de extrair informações sobre a base de dados (arquivo ‘CSV’) abordado

Identificador	Nome	Descrição
[HU001]	Interface/Menu	POSSO acessar os dados por meio de uma interface

		POIS ASSIM é possível visualizar informações a respeito dos clientes.
--	--	---

**Quadro 5:** Exemplo de Contexto. **Fonte:** autor

### 3.5. Requisitos de Produto de Software

Funcionalidades levantadas segundo preconiza a metodologia de desenvolvimento ágil Scrum, ou seja, através de histórias de usuário. Elas foram dispostas no apêndice B. Além disso, elaborou-se o diagrama de casos de uso (apêndice G) para representar graficamente os requisitos e seus relacionamentos.

### 3.6 Perspectiva do Produto

Com o software, a perspectiva é que as consultas sejam intuitivas, instantâneas, sem a necessidade de dedução de expressões lógicas por parte dos usuários, transformando dados brutos em informações úteis, com auxílio de gráficos visuais.

### 3.7 Funcionalidade do Produto

As funcionalidades são detalhadas no apêndice B e visam satisfazer os requisitos elicitados e convergir para a solução do problema abordado neste trabalho, observando as práticas recomendadas pela Engenharia de Software.

### 3.8 Usuários

Usuários que almejam manipular os dados do arquivo ‘CSV’ utilizado como base de dados e extrair computacionalmente algumas informações.

### **3.9 Ambiente Operacional**

Sistema operacional Windows

### **3.10 Restrições de Projeto e Implementação**

Linguagem Python;

Sistema operacional Windows;

Memória: 1 GB;

Espaço em Disco: 1 GB;

Bibliotecas Python: Tkinter, Pandas, Statsmodels, Numpy, Datetime,

Matplotlib, Sklearn.preprocessing, Pandastable e Seaborn.

### **3.11 Documentação do Usuário**

Documentação do usuário disponibilizada através de um helper de ajuda.

### **3.12 Suposições e Dependências**

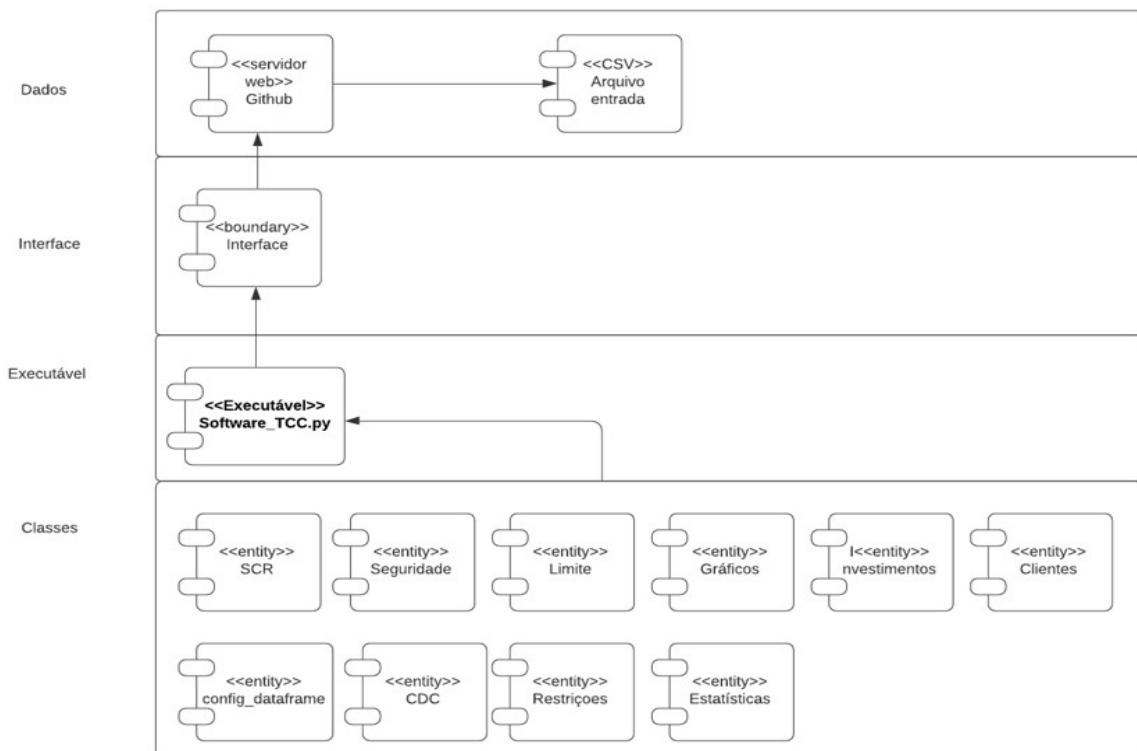
Supõe-se que os dados foram exportados para um arquivo ‘CSV’ por um banco de dados robusto, experimentado e confiável, que adota os aspectos ACID (atomicidade, consistência, isolamento e durabilidade). Portanto, não foram considerados aspectos de pré-processamento dos dados, como dados incompletos, inconsistentes, redundantes ou com ruídos, que extrapola os objetivos do trabalho.

Para nortear o leitor menos acostumado com esses termos, Faceli (2011, p. 34) dá uma definição sobre eles: Dados incompletos são “a ausência de valores para alguns atributos de alguns objetos.”; inconsistentes seriam “aqueles que possuem valores conflitantes em seus atributos”. Dados redundantes são quando um objeto “é muito semelhante a um outro objeto do mesmo conjunto de dados”. “Dados com ruídos são dados que contêm objetos que, aparentemente, não pertencem à distribuição que gerou os dados analisados”.

### 3.13. Diagrama de Caso de Uso

Elaborou-se o diagrama de casos de uso que ilustra a forma como os usuários podem interagir com o software, fornecendo uma visão geral sobre os relacionamentos entre os usuários, o software e os casos de uso. O diagrama é descrito com detalhes no apêndice G.

### 3.14. Diagrama de Componentes



**Figura 3.14:** Diagrama de componentes. **Fonte:** autor

#### 4. METODOLOGIA

Em se tratando de classificação de pesquisa, este trabalho é aplicado, pois seu objetivo geral é construir um software.

Quanto aos procedimentos metodológicos, adotou-se uma abordagem quantitativa, já que utiliza métodos e medidas estatísticas com o intuito de explicar os resultados através da quantificação dos dados (ALMEIDA, 2014).

Efetuou-se uma pesquisa bibliográfica sobre as principais medidas da estatística descritiva, sobre a linguagem de programação Python e algumas de suas bibliotecas, bem como sobre os principais gráficos utilizados em análise de dados. Essa pesquisa baseou-se em materiais especializados no tema, principalmente em: MCKINNEY (2018), MENEZES (2010) e FACELI (2011).

Extraiu-se algumas informações da base de dados (descrita no apêndice D) com o Microsoft Excel (ferramenta mais popular para esta tarefa) e foram analisadas algumas dificuldades em obter informações analíticas e como contorná-las por meio da linguagem de programação Python.

Na sequência, levantou-se alguns requisitos através de conceitos da metodologia de desenvolvimento ágil Scrum, ou seja, por meio de histórias de usuários. Após o levantamento inicial, realizou-se a organização e priorização dos requisitos com auxílio de artefatos e ferramentas utilizadas no Scrum, como o Backlog do Produto, o Backlog das Sprints, o Scrum Board e o Gráfico de Burndown.

Posteriormente, desenvolveu-se um software capaz de realizar manipulação dos dados e plotagens de gráficos sem a necessidade de trabalho manual e/ou elaboração de fórmulas pelo(s) usuário(s), por meio da escrita e interpretação do código-fonte na linguagem de programação Python.

O Quadro 1 descreve as atividades realizadas no desenvolvimento do software:

Quadro 1 - Atividades para Construção de Projeto de Software

<b>Atividade</b>	<b>Artefatos resultantes</b>
Engenharia de Requisitos	<ul style="list-style-type: none"> <li>• Delimitação do escopo</li> <li>• Definição dos Stakeholders</li> <li>• Levantamento de requisitos (funcionais e não-funcionais)</li> <li>• Planejamento de teste de aceitação</li> </ul>
Elaboração do Projeto de Software	<ul style="list-style-type: none"> <li>• Diagrama de caso de uso</li> <li>• Diagrama de componentes.</li> </ul>
Tecnologias utilizadas e os requisitos de implantação	<ul style="list-style-type: none"> <li>• VS Code, versão 1.60.2: Editor desenvolvido pela Microsoft para depuração de código-fonte. Pode ser utilizado com o Microsoft Windows, MacOS e Linux. É apto a depurar código-fonte de algumas linguagens de programação, como Python.</li> <li>• Microsoft Excel 365, versão 2108: Popular editor de planilhas, elaborado pela Microsoft. Possui funções para cálculos matemáticos/estáticos e para construção de gráficos.</li> <li>• Trello: Ferramenta online gratuita para gerenciamento de projetos. É</li> </ul>

	<p>uma alternativa versátil, que possui interface intuitiva e com vários recursos úteis, como checklists. Organiza os projetos por meio de quadros, similar ao Kanban.</p> <ul style="list-style-type: none"><li>• Linguagem de programação Python, versão 3.9.5: Uma das linguagens de programação mais populares do mundo. É uma linguagem de programação considerada simples, sobretudo pela sua sintaxe clara. Ao longo dos anos, as comunidades desenvolveram bibliotecas poderosas para Python, como Pandas e Matplotlib, que tornaram Python uma excelente alternativa para análise de dados.</li></ul>
--	--

Fonte: o autor.

## RESULTADOS

Um dos requisitos levantado nas etapas anteriores demandava a exibição das profissões cadastradas na base de dados. Isso foi obtido através do software e as profissões totalizaram quatorze distintas, sendo elas: auxiliar\_servicos\_gerais, técnica\_enfermagem, radiotecnico, corretor, trabalhador\_informal, motorista, auxiliar\_escritorio, pensionista, aposentado, professor\_ensino\_fundamental, servidor\_publico\_estadual, vendedor, empresário e professor\_ensino\_superior. A figura R.1 expõe elas:

Profissões	
	Profissão
1	auxiliar_servicos_gerais
2	tecnica_enfermagem
3	radiotecnico
4	corretor
5	trabalhador_informal
6	motorista
7	auxiliar_escritorio
8	pensionista
9	aposentado
10	professor_ensino_fundamental
11	servidor_publico_estadual
12	vendedor
13	empresario
14	professor_ensino_superior

**Figura R.1:** Profissões registradas na base de dados. **Fonte:** autor

Outro requisito exigido solicitava a busca de um registro específico por meio do ID (único) e exibição dos atributos de tal registro. A figura R.2 demonstra as informações retornadas por meio da busca por meio do ID 360000036. Nela, constam informações sobre dados básicos (ID, nome, ano de nascimento e sexo); dados profissionais (profissão, status de aposentadoria e renda); dados de segurança (status de seguro auto, residencial, vida, capitalização, previdência privada e consórcios); informações sobre o endividamento no sistema de crédito (SCR); demais

informações (histórico de contratação de CDC, status de uso do limite da conta, histórico de restrições, status de pagamento parcial de cartão e telefone); dados sobre valores investidos (total de investimentos).

Dados Básicos		Seguridade		Demais Informações	
ID:	360000036	Seguro Auto:	nao	Tomou CDC Antes:	nao
Nome:	cliente35	Seguro Residencial:	nao	Usando Limite:	nao
Ano Nasc.:	1984	Seguro Vida:	sim	Hist. Restrição:	nao
Sexo:	m	Capitalização:	nao	Pgto. Parcial Cartão:	nao
		Previdência Privada:	nao	Telefone:	333-333
		Consórcios:	nao		
Dados Profissionais		SCR		Investimentos	
Profissão:	auxiliar_servicos_gerais	SCR Noso:	0	Investimentos:	6792.01
Aposentado:	nao	SCR Outros:	0		
Renda:	1822.13	SCR Total:	0		

**Figura R.2:** Dados cadastrais de um registro específico filtrado por ID. **Fonte:** autor

A figura R.3 demonstra o resultado de uma consulta geral dos registros da base de dados (trinta registros, para efeito de ilustração). Por meio dessa funcionalidade, exibiu-se todos os registros da base, bem como todos os seus atributos (alguns foram omitidos na figura, por questões de formatação).

Consulta Geral

	Id	Nome	CDC	Investimento	SCR Nossa	SCR Outros	SCR Total	Usando Limite	Restrição	Pgto. Parcial Fatura	Ano Nasc.
1	360000036	cliente35	nao	6792.01	0	0	0	nao	nao	nao	1984
2	400000004	cliente3	nao	6792.00	0	0	0	nao	nao	nao	1984
3	410000041	cliente40	nao	0	0	0	0	nao	nao	nao	1979
4	680000068	cliente67	nao	6792.00	0	0	0	nao	nao	nao	1984
5	730000073	cliente72	nao	0	0	0	0	nao	nao	nao	1979
6	900000009	cliente8	nao	0	0	0	0	nao	nao	nao	1979
7	220000022	cliente21	sim	0	204	0	204	sim	nao	nao	1997
8	250000025	cliente24	sim	0	204	0	204	sim	nao	nao	1997
9	540000054	cliente53	sim	0	204	0	204	sim	nao	nao	1997
10	570000057	cliente56	sim	0	204	0	204	sim	nao	nao	1997
11	860000086	cliente85	sim	0	204	0	204	sim	nao	nao	1997
12	890000089	cliente88	sim	0	204	0	204	sim	nao	nao	1997
13	390000039	cliente38	sim	0	0	2495	2495	nao	nao	nao	1984
14	700000007	cliente6	sim	0	0	2495	2495	nao	nao	nao	1984
15	710000071	cliente70	sim	0	0	2495	2495	nao	nao	nao	1984
16	240000024	cliente23	sim	0	1900	1902	3802	sim	sim	nao	1979
17	560000056	cliente55	sim	0	1900	1902	3802	sim	sim	nao	1979
18	880000088	cliente87	sim	0	1900	1902	3802	sim	sim	nao	1979
19	100000010	cliente9	nao	0	3300	1600	4900	nao	sim	nao	1983
20	420000042	cliente41	nao	0	3300	1600	4900	nao	sim	nao	1983
21	740000074	cliente73	nao	0	3300	1600	4900	nao	sim	nao	1983
22	270000027	cliente26	sim	0	0	5100	5100	nao	nao	np	1975
23	590000059	cliente58	sim	0	0	5100	5100	nao	nao	np	1975
24	910000091	cliente90	sim	0	0	5100	5100	nao	nao	np	1975
25	170000017	cliente16	sim	0	4000	1700	5700	nao	nao	nao	1996
26	490000049	cliente48	sim	0	4000	1700	5700	nao	nao	nao	1996
27	810000081	cliente80	sim	0	4000	1700	5700	nao	nao	nao	1996
28	380000038	cliente37	sim	0	3900	2300	6200	nao	nao	nao	1996
29	600000006	cliente5	sim	0	3900	2300	6200	nao	nao	nao	1996
30	700000070	cliente69	sim	0	3900	2300	6200	nao	nao	nao	1996

**Figura R.3:** Consulta geral de dados cadastrais. **Fonte:** autor

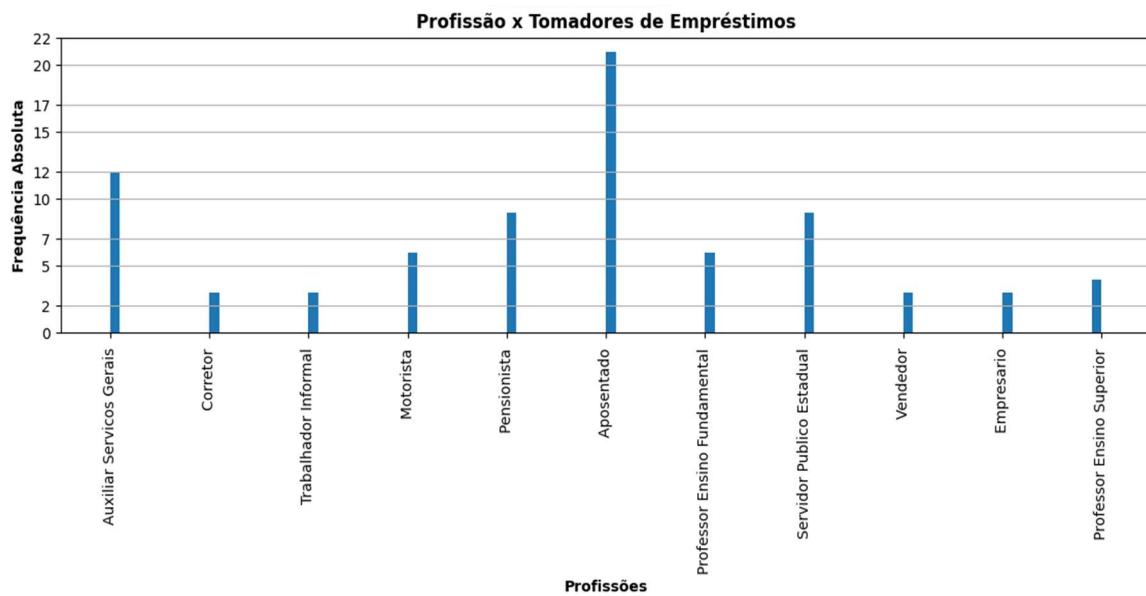
Já a figura R.4 expõe o resultado da execução de uma funcionalidade que filtra apenas clientes tomadores de empréstimos (CDC=sim). Por meio do software, foi possível exibir somente os clientes que já contrataram CDC (CDC = sim), bem como dar ênfase na coluna contendo o atributo “CDC” (destacada na cor amarela pelo software).

Posse CDC: SIM

	Id	Nome	Investimento	CDC	SCR Nossa	SCR Outros	SCR Total	Usando Limite	Restrição	Pgto.Parcial	Ano Nasc.
1	220000022	cliente21	0	sim	204	0	204	sim	nao	nao	1997
2	250000025	cliente24	0	sim	204	0	204	sim	nao	nao	1997
3	540000054	cliente53	0	sim	204	0	204	sim	nao	nao	1997
4	570000057	cliente56	0	sim	204	0	204	sim	nao	nao	1997
5	860000086	cliente85	0	sim	204	0	204	sim	nao	nao	1997
6	890000089	cliente88	0	sim	204	0	204	sim	nao	nao	1997
7	390000039	cliente38	0	sim	0	2495	2495	nao	nao	nao	1984
8	700000007	cliente6	0	sim	0	2495	2495	nao	nao	nao	1984
9	710000071	cliente70	0	sim	0	2495	2495	nao	nao	nao	1984
10	240000024	cliente23	0	sim	1900	1902	3802	sim	sim	nao	1979
11	560000056	cliente55	0	sim	1900	1902	3802	sim	sim	nao	1979
12	880000088	cliente87	0	sim	1900	1902	3802	sim	sim	nao	1979
13	270000027	cliente26	0	sim	0	5100	5100	nao	nao	np	1975
14	590000059	cliente58	0	sim	0	5100	5100	nao	nao	np	1975
15	910000091	cliente90	0	sim	0	5100	5100	nao	nao	np	1975
16	170000017	cliente16	0	sim	4000	1700	5700	nao	nao	nao	1996
17	490000049	cliente48	0	sim	4000	1700	5700	nao	nao	nao	1996
18	810000081	cliente80	0	sim	4000	1700	5700	nao	nao	nao	1996
19	380000038	cliente37	0	sim	3900	2300	6200	nao	nao	nao	1996
20	600000006	cliente5	0	sim	3900	2300	6200	nao	nao	nao	1996
21	700000070	cliente69	0	sim	3900	2300	6200	nao	nao	nao	1996
22	370000037	cliente36	11584.00	sim	4300	5900	10200	nao	nao	nao	1954
23	500000005	cliente4	11584.00	sim	4300	5900	10200	nao	nao	nao	1954
24	690000069	cliente68	11584.00	sim	4300	5900	10200	nao	nao	nao	1954
25	200000020	cliente19	0	sim	3780	11900	15680	sim	nao	nao	1956
26	520000052	cliente51	0	sim	3780	11900	15680	sim	nao	nao	1956
27	840000084	cliente83	0	sim	3780	11900	15680	sim	nao	nao	1956

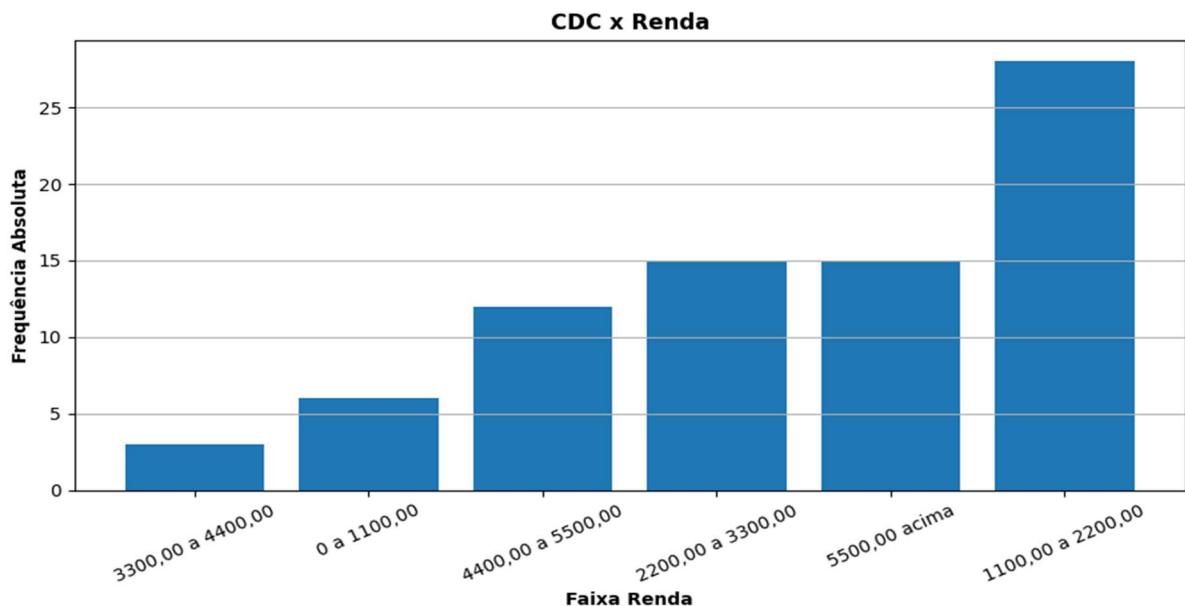
**Figura R.4:** Consulta registros com posse de CDC. **Fonte:** autor

A figura R.5 traz um histograma que fornece uma visualização sobre as profissões tomadoras de empréstimos. Por meio dela, concluiu-se que os aposentados eram os maiores tomadores de empréstimos, apresentando frequência absoluta superior a vinte; em seguida, as profissões de auxiliar de serviços gerais, com doze registros. Pensionistas e servidor público estadual apresentaram frequência absoluta maior que sete e menor do que dez. Também é possível visualizar que as profissões de empresário, corretor, trabalhador informal, vendedor e professor do ensino superior contrataram menos empréstimos (quantidade em frequência absoluta menor que sete) do que as demais profissões citadas anteriormente.



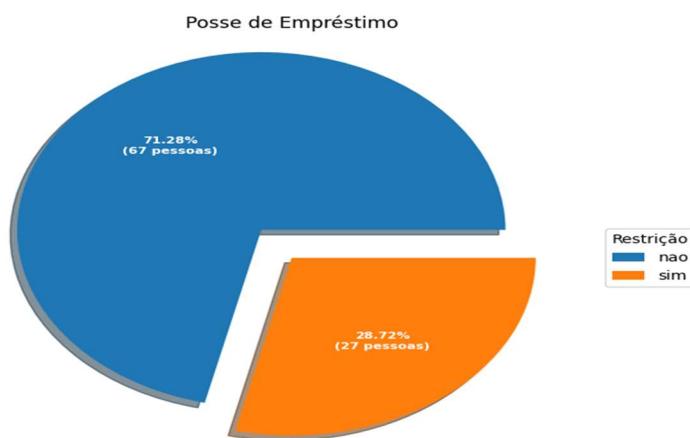
**Figura R.5:** Gráfico profissão x tomadores de empréstimos. **Fonte:** autor

Por meio da plotagem do histograma da figura R.6, verificou-se que – dentre os registros da base de dados – a faixa de renda que mais efetuou a contratação de empréstimos foi a de R\$ 1.100,00 a R\$ 2.200,00, com mais de vinte e cinco registros. Em seguida, as faixas de renda de R\$ 2.200,00 a R\$ 3.300,00 e a faixa acima de R\$ 5.500,00, ambas com quinze registros; logo na sequência, as faixas de renda de R\$ 4.400,00 a R\$ 5.500,00 e R\$ 0,00 a R\$ 1.100,00. A faixa de renda que menos contratou empréstimos foi a de R\$ 3.300,00 a R\$ 4.400,00, conforme ilustra a figura R.6 abaixo. O histograma foi ordenado pelo software para exibir os dados na ordem crescente de frequência absoluta.



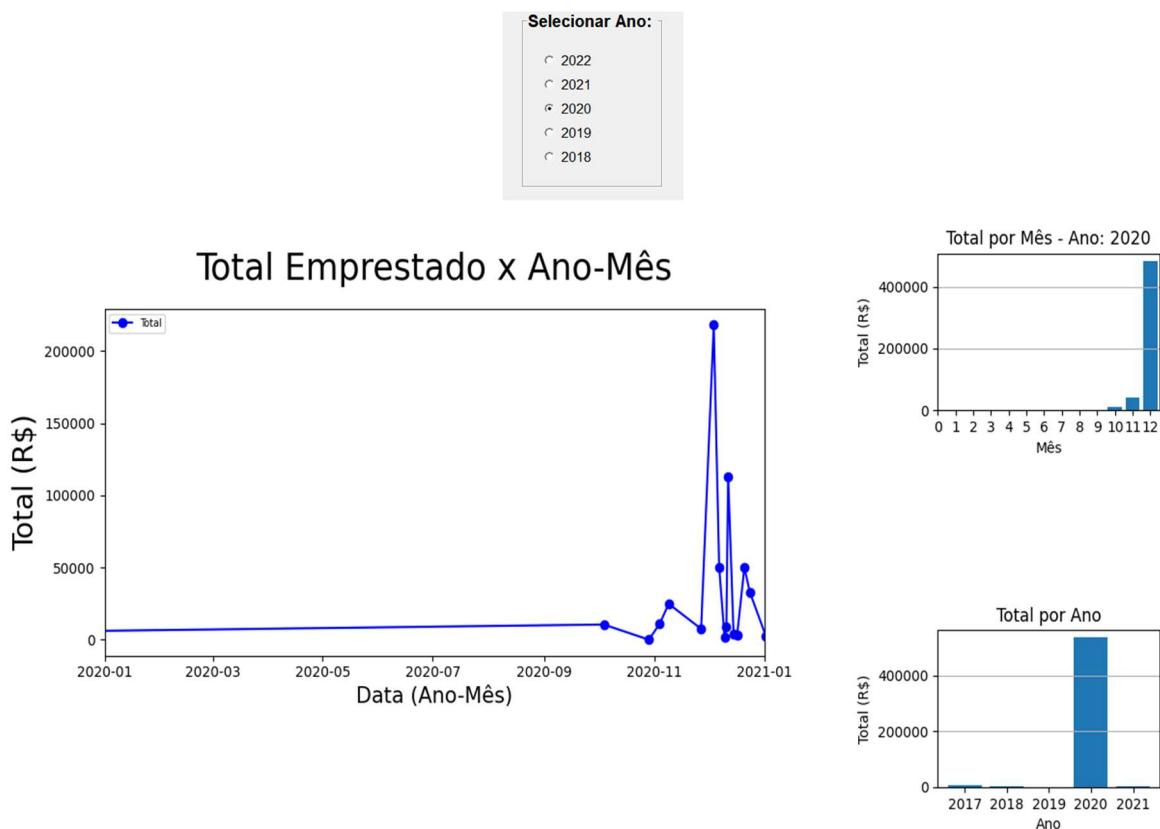
**Figura R.6:** Gráfico CDC x Renda. **Fonte:** autor

Através da figura R.7 abaixo, inferiu-se que 71,28% dos registros que contrataram empréstimos não possuíam restrições no seu histórico, enquanto 28,72% tinham e/ou já tiveram algum tipo de restrição. O setor dos registros com restrição foi destacado no gráfico de pizza (figura R.7) através do software para enfatizar esse dado.



**Figura R.7:** Gráfico pizza posse de empréstimo. **Fonte:** autor

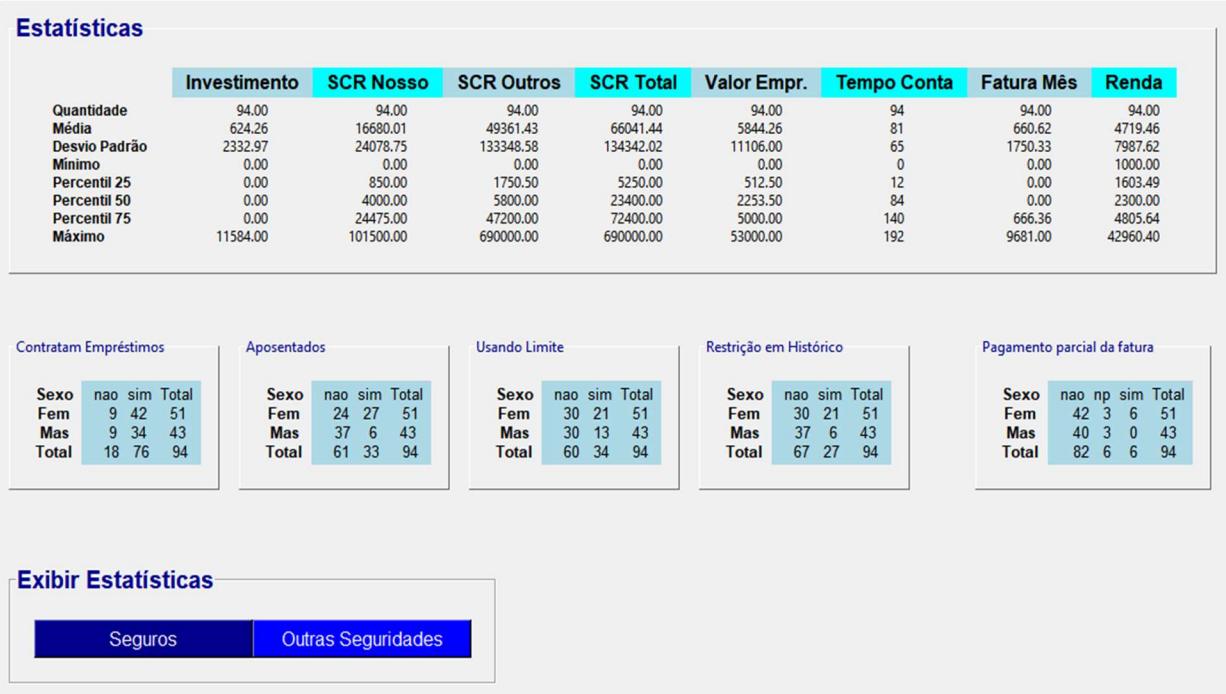
A figura R.8 ilustra o valor total emprestado por ano/mês, que permitiu deduzir que 2020 foi o ano com maior volume de empréstimos contratados quando comparando com os anos de 2017, 2018, 2019 e 2021 (comparação efetuada entre os últimos cinco anos, conforme requisitado). Nota-se também que os maiores valores de empréstimos foram efetuados nos meses de novembro e dezembro do ano de 2020.



**Figura R.8:** Gráfico empréstimo por ano. **Fonte:** autor

As figuras R.9, R.10 e R.11 demonstram algumas estatísticas sobre os dados. A figura R.9 exibe informações como quantidade de registros, média, desvio padrão, mínimo, percentis 25º, 50º e 75º, bem como o máximo de cada atributo (investimento,

scr nosso, scr outros, scr total, valor empréstimo, tempo de conta, fatura do mês e renda). Já as figuras R.10 e R.11 mencionam estatísticas referentes à quantidade de registros, quantidade de valores distintos, valores mais frequentes e frequência dos atributos: seguro auto, seguro vida, seguro residencial, consórcio, capitalização e previdência privada. Além disso, a figura R.9 traz informações sobre a quantidade de registros que contrataram empréstimos, eram aposentados, estavam usando limite da conta, histórico de restrições e a respeito do pagamento parcial da fatura do cartão de crédito, segmentados pelo sexo (Mas = masculino e Fem = feminino).



**Figura R.9:** Estatísticas I. **Fonte:** autor

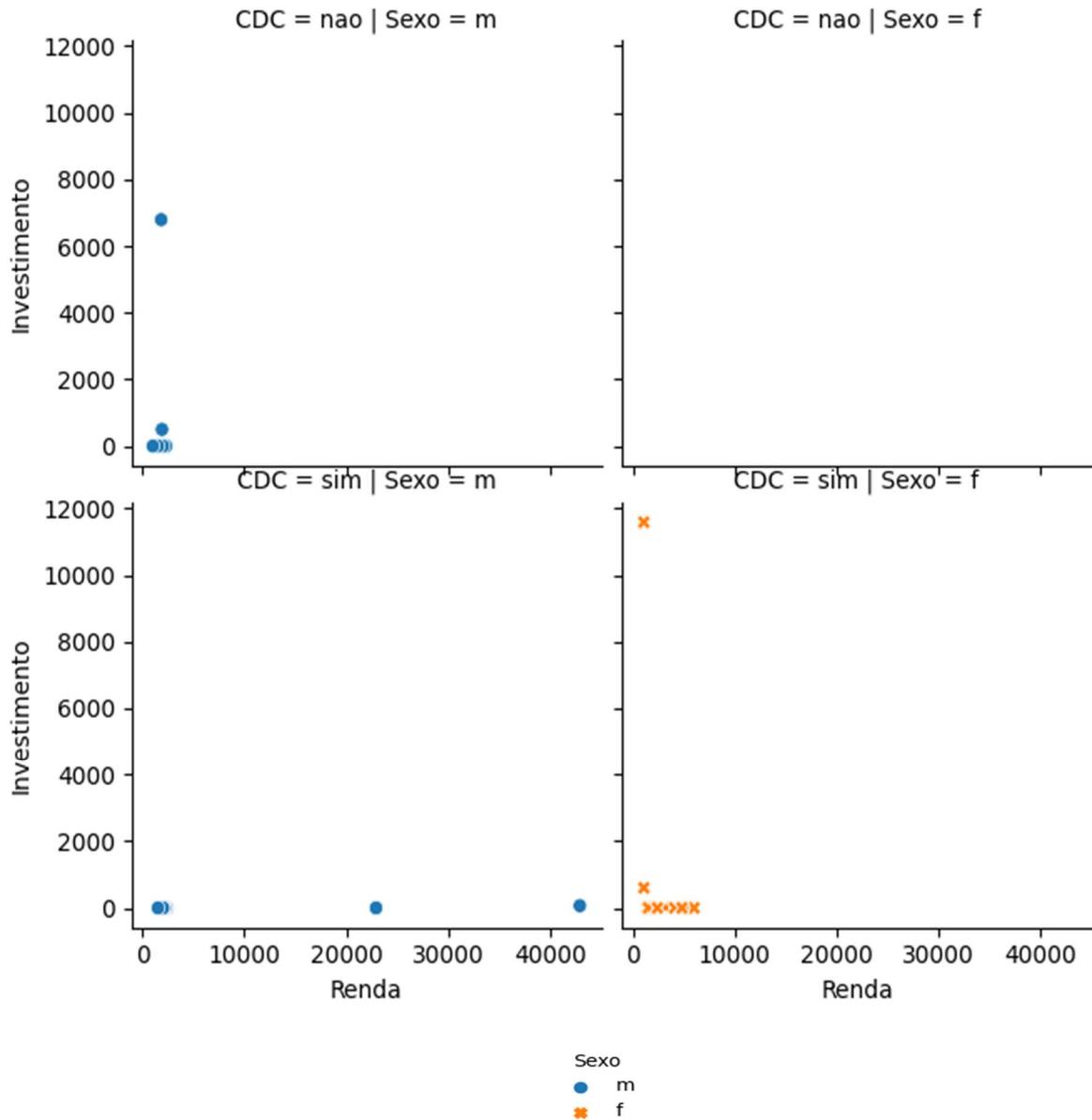
	Estatísticas	Seguro Auto	Seguro Vida	Seguro Residencial
1	Quantidade registros	94	94	94
2	Quant. de valores distintos	2	2	2
3	Mais frequente	nao	nao	nao
4	Frequência	75	83	84

**Figura R.10:** Estatísticas II. **Fonte:** autor

	Estatísticas	Consorcio	Capitalizacao	Previdencia Privada
1	Quantidade registros	94	94	94
2	Quant. de valores distintos	2	2	2
3	Mais frequente	nao	nao	nao
4	Frequência	84	86	85

**Figura R.11:** Estatísticas III. **Fonte:** autor

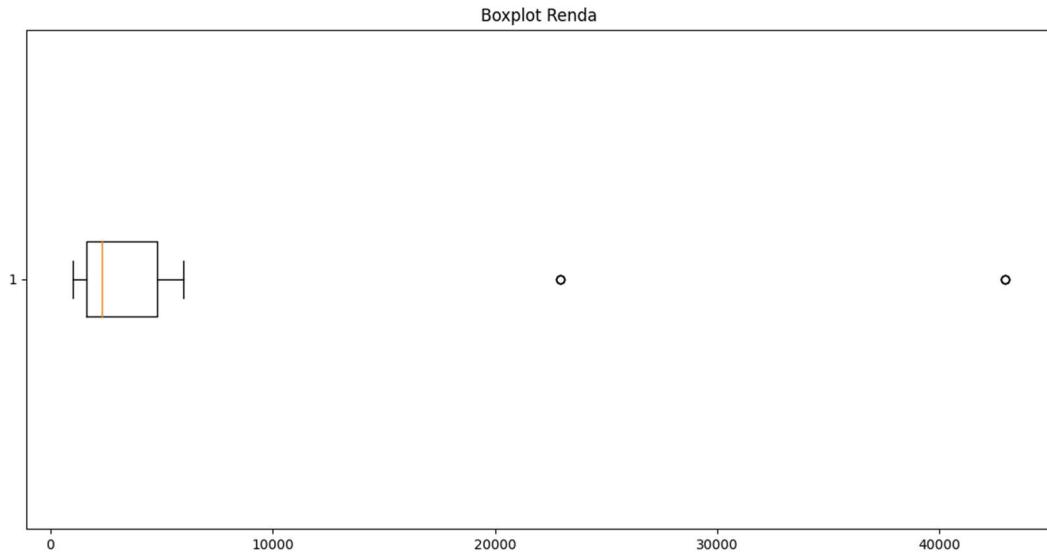
A figura R.12 se refere a um gráfico do tipo scatter e ilustra tomadores e não tomadores de empréstimos, considerando o total de investimentos (eixo y) e a faixa de renda (eixo x), segmentados pelo sexo (m – masculino e f – feminino). Por meio dele nota-se que a base de dados apresentava – em sua grande maioria – registros com baixos valores de investimentos (independentemente do sexo). Notavelmente os registros do sexo feminino que tomaram CDC (CDC = sim e sexo = feminino) também apresentaram baixa faixa de renda (concentrados em renda menores que R\$ 10 mil mensais).



**Figura R.12:** Gráfico scatter. **Fonte:** autor

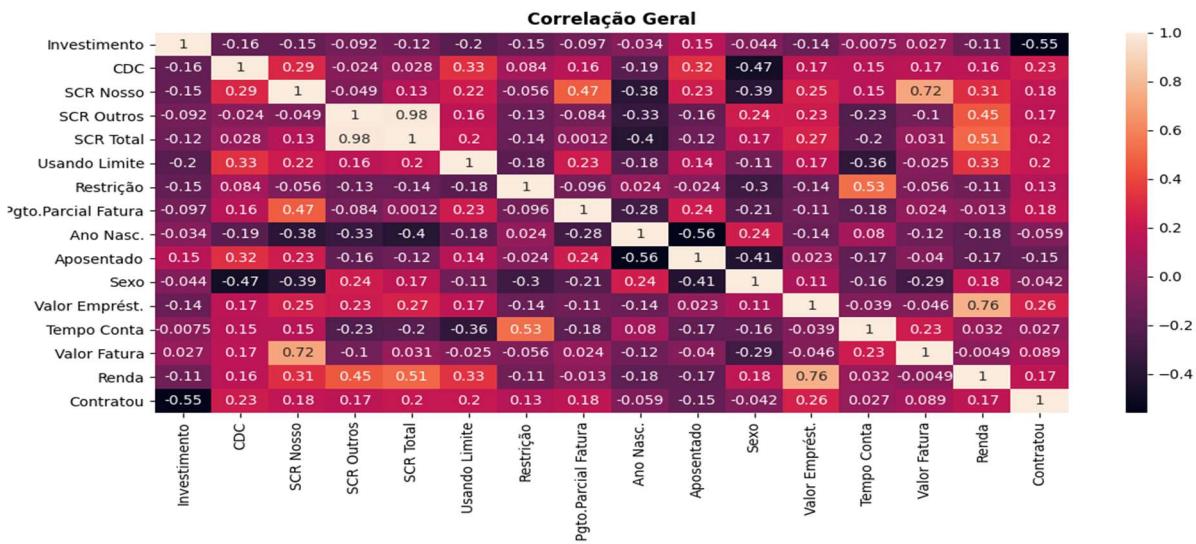
Por meio do boxplot exibido na figura R.13, foi possível visualizar a presença de dois valores considerados outliers (um valor de renda entre vinte mil e trinta mil e outro valor acima de quarenta mil). Isso permitiu uma averiguação desses valores (se realmente os registros possuam tais rendas ou se foi um erro, como erro de digitação

dos valores). Com excessão desses dois valores, nota-se que a renda máxima não ultrapassava R\$ 10 mil.



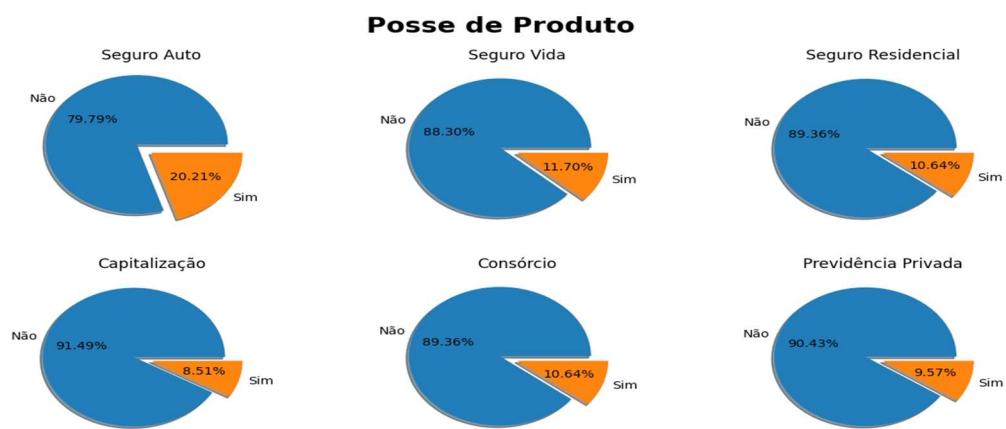
**Figura R.13:** Boxplot renda. **Fonte:** autor

A correlação entre os atributos numéricos foi obtida por meio de uma funcionalidade presente no software e desenvolvida com auxílio da biblioteca Seaborn. A figura R.14 exibe a correlação entre vários atributos, permitindo compreender a variação de um atributo com base na variação de outro atributo. Percebe-se, por exemplo, que a renda apresentava uma correlação relativamente alta com o valor emprestado (0.76).



**Figura R.14:** Correlação. **Fonte:** autor

Com relação aos produtos de seguridade (seguro auto, seguro vida, seguro residencial, consórcios, capitalização e previdência privada), a figura R.15 mostra que 79,79% contrataram seguro auto (20,21% não contrataram); 88,30% possuíam seguro de vida (11,70% não); 89,36% tinham posse de seguro residencial (10,64% não tinham); 91,49% já adquiriram capitalização (8,51% não); 89,36% possuíam consórcio (10,64% não) e 90,43% detinham previdência privada (9,57% não).



**Figura R.15:** Posse de seguridade. **Fonte:** autor

## CONCLUSÃO

Posteriormente à elaboração desse trabalho, chegou-se à conclusão de que Python é uma excelente ferramenta gratuita (inclusive para fins comerciais) para manipular dados contidos em arquivos do tipo ‘CSV’ (da língua inglesa, “Character-separated values”), que são arquivos usados para armazenar dados tabulares. Para essa finalidade específica, a biblioteca Pandas do Python trouxe inúmeros recursos que facilitaram essa tarefa, tornando-a muito simples. Além disso, como Python é uma das linguagens de programação mais populares da atualidade, existem inúmeras e ativas comunidades que buscam auxiliar nas dúvidas sobre sua utilização e sobre seus recursos, e isso é algo que facilita o trabalho de quem a utiliza, como por exemplo, programadores.

Quanto às bibliotecas empregadas na análise dos dados, as mais utilizadas foram as seguintes: Pandas, Matplotlib, Seaborn, Numpy e Tkinter. A seguir, algumas considerações sobre cada uma delas: Pandas é uma biblioteca que facilita o trabalho com arquivos do tipo ‘CSV’, tornando a leitura e manipulação dos dados rápida e eficiente por disponibilizar os dados através de tabelas facilmente manipuláveis; Matplotlib e Seaborn são bibliotecas incríveis para plotagens de dados, com a possibilidade de personalizar os gráficos de diversas maneiras(cores, tamanho de fonte, disposição das imagens, entre outras configurações); Numpy realiza processamentos numéricos e oferece um amplo rol de funções e operações que auxiliam o processamento de cálculos numéricos, sendo muito utilizada em tarefas matemáticas; Tkinter é uma biblioteca embutida na instalação padrão do Python e possui ferramentas para desenvolvimento de interfaces gráficas do usuário. Qualquer computador que disponha do interpretador Python pode gerar interfaces gráficas com Tkinter. Chama a atenção na sua utilização a facilidade em acessar seus recursos, já que permite inserir widgets (botões, caixas de textos etc.) na interface do usuário com extrema simplicidade.

Quanto os gráficos utilizados para analisar os dados, os que mais destacaram-se foram o histograma, o boxplot, o scatter plot e o gráfico de pizza. Por meio deles,

foi possível deduzir inúmeras afirmações, sobretudo a respeito de medidas estatísticas da base de dados. Os gráficos representaram recursos visuais que facilitaram a análise. Diversas figuras foram inseridas nos apêndices para ilustrar essa afirmação.

No que se refere às medidas estatísticas, abordou-se sobre: média, percentuais, frequência absoluta, mediana, desvio-padrão, mínimo, máximo, percentis e correlação entre os atributos da base de dados. Essas medidas auxiliaram na descrição da base de dados e na identificação de outliers, sendo fundamentais para a análise em questão, pois é uma forma de explorar dados e extrair informações úteis do conjunto de dados. Elas resumiram de maneira quantitativa as características relevantes de grupos de elementos.

Vislumbrando trabalhos futuros, poderia ser utilizado o arquivo ‘CSV’ abordado nesse trabalho, porém contendo mais registros. Além disso, abordar recursos mais avançados das bibliotecas citadas, visando extrair informações úteis a respeito da base analisada e dessa forma incentivar cada vez mais o aprendizado sobre os poderosos recursos da linguagem de programação Python e de algumas de suas bibliotecas empregadas na análise de dados.

## REFERÊNCIAS

ALMEIDA, Mario de Souza. **Elaboração de projeto, TCC, dissertação e tese: uma abordagem simples, prática e objetiva.** 2. Ed. São Paulo: Atlas, 2014.

CARVALHO, B. V e MELLO, C. H. P. **Aplicação do método ágil Scrum no desenvolvimento de produtos de software em uma pequena empresa de base tecnológica.** Disponível em: <https://www.scielo.br/j/gp/a/34xH953TFwLPYDB9BYdJghL/?lang=pt>. Acesso em: 7 de jan. de 2021.

COHN, M. Desenvolvimento de **software** com Scrum: aplicando métodos ágeis com sucesso. São Paulo: Bookman, 2011

FACELI, Katti et. al Inteligência Artificial: **Uma Abordagem de Aprendizagem de Máquina.** Rio de Janeiro: LTC, 2011.

FREITAS, Janaina Aparecida de. **Gerenciamento de Software.** Reimpressão, Maringá-PR: Unicesumar, 2017.

Guia para o conhecimento Scrum (Guia SBOK™) by ScrumStudy  
PERSEGUINE, V. R. **Engenharia de Requisitos.** Reimpressão, Maringá-PR.: Unicesumar, 2016.

MARTINELLI, M. **Gráficos e Mapas: Construa-os Você Mesmo.** São Paulo: Moderna, 1998.

McKINNEY W. **Python para Análise de Dados,** 2. ed, Ed. O'Reilly Media, Inc, Wes McKinney 2018

MENEZES, Nilo Ney Coutinho. **Introdução à programação com python.** 3.ed. Editora Novatec, 2010.

Sbrocco, J. H. T. C. e Macedo, P. C. 2012 **Metodologias ágeis engenharia de software sob medida.** Editora Érica Ltda. 1º ed. 2012.

SOMMERVILLE, I. **Engenharia de Software.** 9.ed. São Paulo: Pearson Prentice Hall, 2011.

SOMMERVILLE, I. **Engenharia de Software.** 8. ed. São Paulo: Pearson Addison-Wesley, 2007.

SUTHERLAND, Jeff. **Scrum - a arte de fazer o dobro de trabalho na metade do tempo.** São Paulo: LEYA BRASIL, 2016.

## APÊNDICES

## APÊNDICE A – Declaração de escopo do projeto

### DECLARAÇÃO DE ESCOPO DO PROJETO

Gerente do Projeto: Ricardo Tompson Fioravante

DATA 05/09/2021

### HISTÓRICO DE REVISÕES DO DOCUMENTO

DATA	VERSÃO	DESCRIÇÃO DA ALTERAÇÃO	AUTOR
05/09/2021	1	CRIAÇÃO DESTE DOCUMENTO	GERENTE DO PROJETO
07/09/2021	2	COMPILAÇÃO	GERENTE DO PROJETO
12/09/2021	3	REVISÃO	GERENTE DO PROJETO
15/09/2021	4	CONCLUSÃO/APRESENTAÇÃO	GERENTE DO PROJETO

## I. DESCRIÇÃO DO PROJETO

Software desenvolvido no trabalho de conclusão do curso de Engenharia de Software, com o intuito de possibilitar que os usuários possam efetuar consultas, visualizar informações e obter insights com facilidade, tendo como base um arquivo ‘CSV’. O programa destinado a atender essa finalidade específica.

## II. PATROCINADOR

Não possui

## III. GERENTE DO PROJETO

Ricardo Tompson Fioravante

## IV. EQUIPE DO PROJETO

Nome	Área
RTF	Gerente do Projeto

## V. OBJETIVOS DO PROJETO

- Facilitar consultas
- Extrair informações visuais por meio de gráficos
- Sintetizar resultados

## VI. JUSTIFICATIVA DO PROJETO

- Dificuldade dos usuários na utilização de outras ferramentas similares
- Custos de aquisição de aplicações semelhantes
- Adversidades enfrentadas por usuários na elaboração de fórmulas lógicas/matemáticas/estatísticas.

## VII. FATORES PARA O SUCESSO DO PROJETO

- Observação do cronograma
- Correção dos apontamentos da orientadora

## VIII. RESTRIÇÕES DO PROJETO

- Projeto deverá estar concluído até o final da disciplina do TCC II

## IX. PREMISSAS DO PROJETO

- Equipe disponível para dar sequência no projeto
- Gerente do projeto com alcada para tomada de decisões

## X. EXCLUSÕES DO PROJETO

- Efetuar pré-processamento dos dados base
- Integrar o produto com demais ferramentas existentes.

## XI. ENTREGA PRINCIPAL

- Software em funcionamento
- Artefatos do ciclo de desenvolvimento

## XII. ORÇAMENTO

- Não definido; produto sem custos

## XIII. FASES PRINCIPAIS

Fase	Discriminação	Data de término
Iniciação	Definição do Gerente do Projeto	10/09/2021
Planejamento	Escopo definido	11/09/2021
	Cronograma estabelecido	12/09/2021
	Plano de Projeto estabelecido	14/09/2021
Execução	Documentação concluída	20/09/2021
	Compilação e correção	21/09/2021
	Testes finalizados	30/09/2021
Encerramento	Lições aprendidas	01/10/2021
	Projeto concluído	10/04/2022

## APÊNDICE B – Histórias de Usuários

Contexto
<p>SENDO Usuário com intenção de extrair informações sobre a base de dados (arquivo ‘CSV’) abordado</p>

Identificador	Nome	Descrição
[HU001]	Interface/Menu	<p>POSSO acessar os dados por meio de uma interface</p> <p>POIS ASSIM visualiza-se informações a respeito dos clientes com facilidade por meio de menus e opções de consultas, ao invés de trabalhar com planilhas menos acessíveis e intuitivas.</p>
[HU002]	Lista Geral de Clientes	POSSO consultar todos os clientes cadastrados

		POIS ASSIM visualiza-se informações gerais dos clientes.
[HU003]	Consultar por ID	<p>POSSO consultar clientes pelo ID.</p> <p>POIS ASSIM seleciona-se um(a) determinado(a) cliente E obtem-se informações cadastrais especificamente sobre ele(a) E SE não lembrar o ID de interesse, ENTÃO posso fazer uma consulta aos ID's da base.</p>
[HU004]	Consultar profissões	<p>POSSO consultar todos os clientes por determinada profissão</p> <p>POIS ASSIM filtra-se profissões específicas para focar em determinadas profissões de interesse.</p>
[HU005]	Empréstimo Posse	POSSO consultar clientes com posse de empréstimo (CDC)

		POIS ASSIM visualiza-se clientes que costumam tomar empréstimos rotineiramente.
[HU006]	Empréstimo Sem Posse	<p>POSSO consultar clientes sem posse de empréstimo (CDC)</p> <p>POIS ASSIM visualiza-se os clientes que não costumam tomar empréstimos e ofertar a eles.</p>
[HU007]	Gráfico profissão dos tomadores	<p>POSSO consultar a profissão dos tomadores de empréstimo</p> <p>POIS ASSIM identifica-se aquelas que têm maior aceitação do produto. E compara elas entre si.</p>
[HU008]	Gráfico renda dos tomadores	<p>POSSO consultar a faixa de renda dos tomadores de empréstimo</p> <p>POIS ASSIM identifica-se as faixas que têm maior aceitação do produto E compara elas entre si.</p>

[HU009]	Gráfico tomadores	<p>POSSO consultar a quantidade de tomadores E não tomadores de empréstimo</p> <p>POIS ASSIM visualiza-se uma característica essencial para o relacionamento.</p>
[HU010]	Gráfico posso de empréstimos	<p>POSSO visualizar o percentual de clientes que tomaram OU não CDC em um gráfico</p> <p>POIS ASSIM é possível comparar a quantidade de clientes tomadores E não tomadores.</p>
[HU011]	Gráfico empréstimos por ano	<p>POSSO visualizar a quantidade de empréstimo efetuado por ano</p> <p>POIS ASSIM é possível comparar a quantidade de empréstimos realizados por período de tempo.</p>

[HU012]	Relação de empréstimos e outros atributos	<p>POSSO visualizar se existente uma relação entre a renda E o sexo E os valores de investimento E a posse de CDC</p> <p>POIS ASSIM visualiza-se se existe algum vínculo entre essas características.</p>
[HU013]	Consultar SCR	<p>POSSO fazer uma consulta sobre o saldo de SCR dos clientes</p> <p>POIS ASSIM visualiza-se o endividamento deles conosco E com a concorrência, comparando-as.</p>
[HU014]	Filtrar endividamento no SCR	<p>POSSO selecionar os clientes com saldo de SCR maior E menor que determinado valor</p> <p>POIS ASSIM visualiza-se aqueles que possuem maior E menor endividamento</p>

[HU015]	Gráfico SCR	<p>POSSO visualizar o endividamento dos clientes no SCR em um gráfico</p> <p>POIS ASSIM posso efetuar uma comparação entre os endividamentos em nossa base E na concorrência.</p>
[HU016]	Uso de Limite	<p>POSSO consultar clientes que não estejam OU estejam utilizando limite da conta</p> <p>POIS ASSIM visualiza-se aqueles(as) clientes que estão OU não estão pagando juros pela utilização do limite.</p>
[HU017]	Gráfico Uso de Limite	<p>POSSO visualizar o percentual de clientes usando OU não o limite em um gráfico</p> <p>POIS ASSIM posso efetuar uma comparação entre a quantidade de clientes utilizando E não utilizando limite</p>

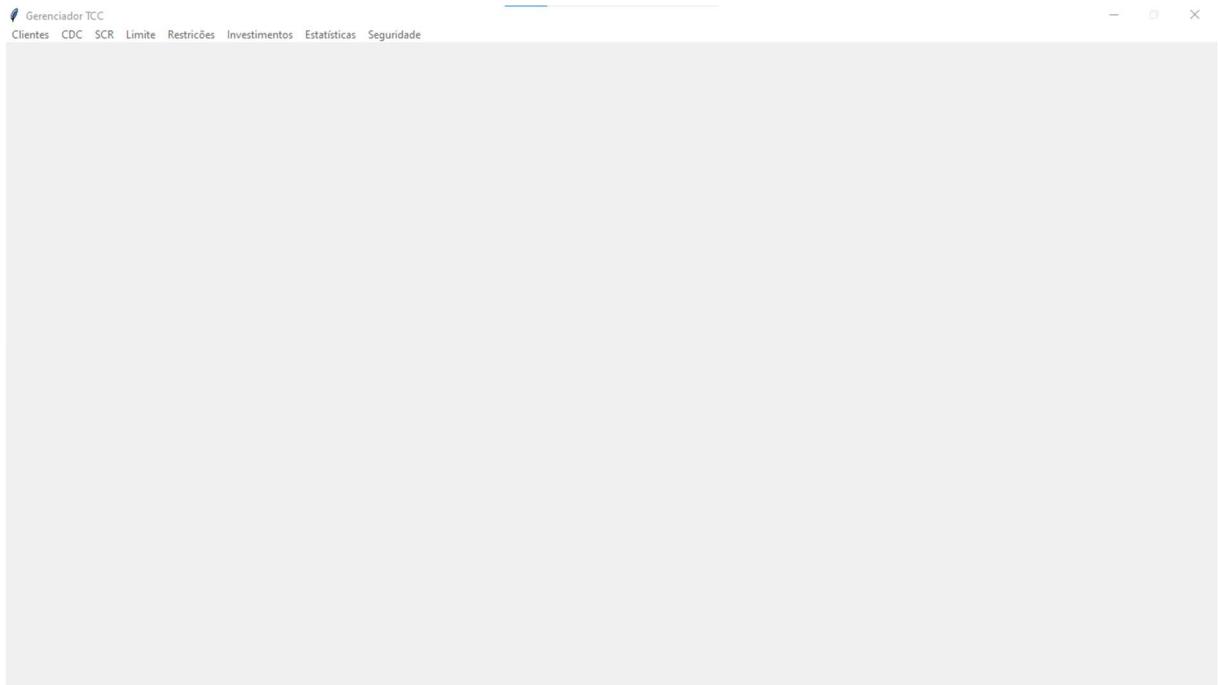
[HU018]	Histórico de restrições	<p>POSSO consultar clientes com restrições vigentes OU sem restrições</p> <p>POIS ASSIM verifica-se um característica importante sobre o perfil dos clientes.</p>
[HU019]	Sem restrições e uso de limite	<p>POSSO consultar clientes sem restrições vigentes E que usam o limite</p> <p>POIS ASSIM filtra-se cliente sem restrições E que estejam usando limite, que são públicos propensos a tomada de crédito.</p>
[HU020]	Sem restrições e parcial cartão	<p>POSSO consultar clientes sem restrições vigentes E que pagaram a fatura do cartão de crédito parcialmente</p> <p>POIS ASSIM verifica-se os clientes sem restrições e que talvez precisem de crédito para pagar integralmente a fatura do cartão.</p>

[HU021]	Sem restrições e aposentados	<p>POSSO consultar clientes sem restrições vigentes E que são aposentados</p> <p>POIS ASSIM verifica-se os clientes sem restrições e que talvez precisem de crédito com taxa diferenciada para esse público.</p>
		<p>POSSO visualizar o percentual de clientes com restrições OU não em um gráfico</p> <p>POIS ASSIM posso efetuar uma comparação entre a quantidade de clientes com restrições E sem restrições.</p>
[HU022]	Gráfico restrições	<p>POSSO visualizar o percentual de clientes com restrições OU não em um gráfico</p> <p>POIS ASSIM posso efetuar uma comparação entre a quantidade de clientes com restrições E sem restrições.</p>

[HU023]	Investimentos ordenados	<p>POSSO consultar os valores investidos em ordem crescente OU decrescente</p> <p>POIS ASSIM consulta-se os clientes com menores ou maiores investimentos que são perfis de interesse para os negócios.</p>
[HU024]	Gráfico Investimentos	<p>POSSO consultar os valores investidos em ordem crescente OU decrescente</p> <p>POIS ASSIM consulta-se os clientes com menores ou maiores investimentos que são perfis de interesse para os negócios.</p>
[HU025]	Estatísticas	<p>POSSO consultar estatísticas sobre os dados</p> <p>POIS ASSIM tem-se uma percepção maior sobre a base de dados, como: qual é o maior valor de investimento? Qual é o menor? A média? É</p>

		necessário consultar essas informações para várias características dos clientes E também analisar a quantidade de posse de produtos, por sexo.
[HU027]	Correlação	<p>POSSO consultar alguma medida estatística, algo do tipo</p> <p>POIS ASSIM verifica-se se algum acréscimo OU decréscimo nos valores de uma característica (como a renda, o sexo, etc.) influenciam nos valores do(s) outro(s).</p>
[HU028]	Gráfico Seguridade	<p>POSSO visualizar a quantidade de clientes que tenham posse OU não de segurança (seguro auto, residencial, entre outros)</p> <p>POIS ASSIM é possível comparar a quantidade de pessoas que ja adquiriram OU não algum desses produtos.</p>

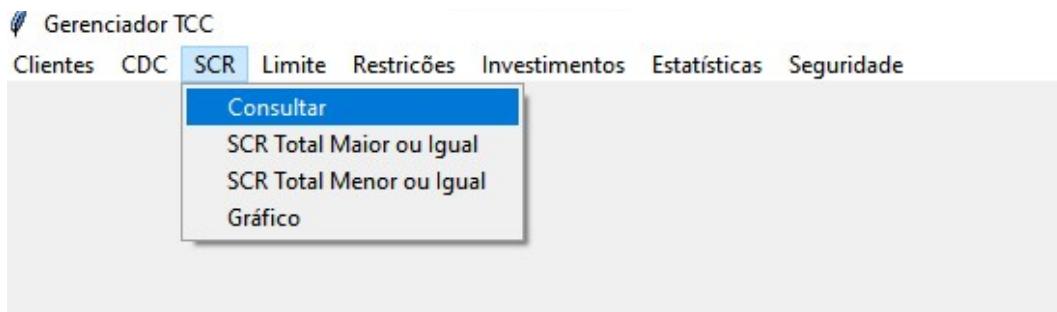
## APÊNDICE C – Telas do Software em funcionamento



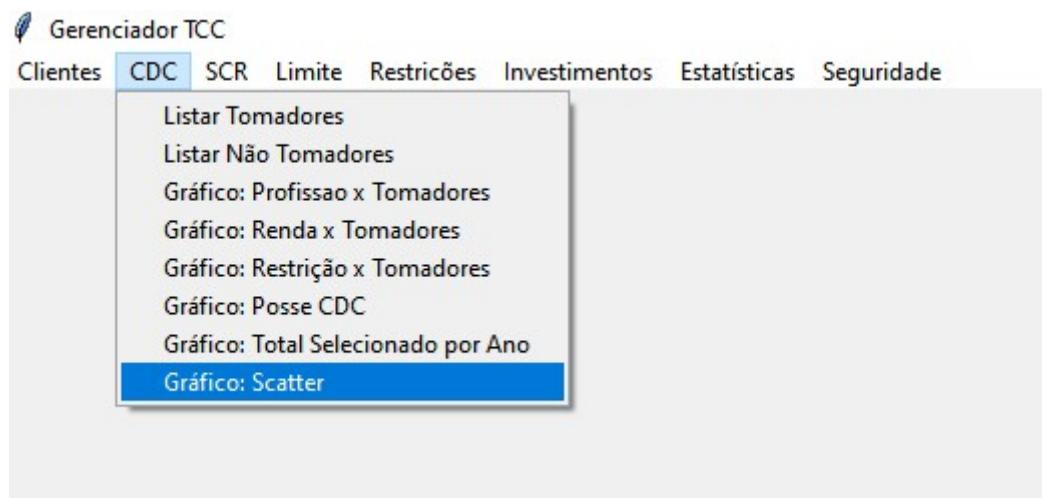
**Figura C.1:** Tela interface. **Fonte:** autor



**Figura C.2:** Menu Clientes. **Fonte:** autor

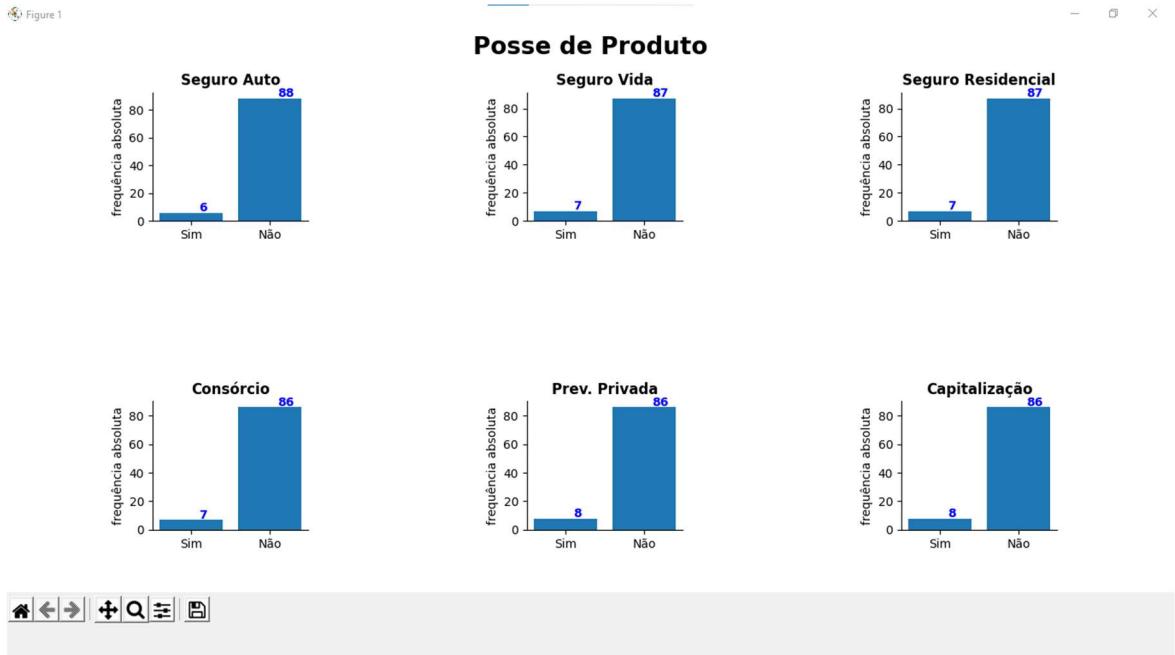


**Figura C.3:** Menu SCR. **Fonte:** autor



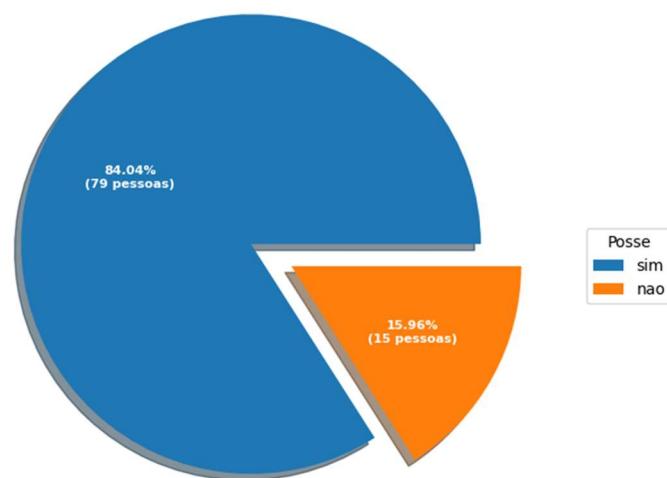
**Figura C.4:** Menu CDC. **Fonte:** autor

Figure 1



**Figura C.5:** Gráfico seguridade frequência. **Fonte:** autor

Posse de Empréstimo



**Figura C.6:** Gráfico pizza empréstimo. **Fonte:** autor

Figure 2

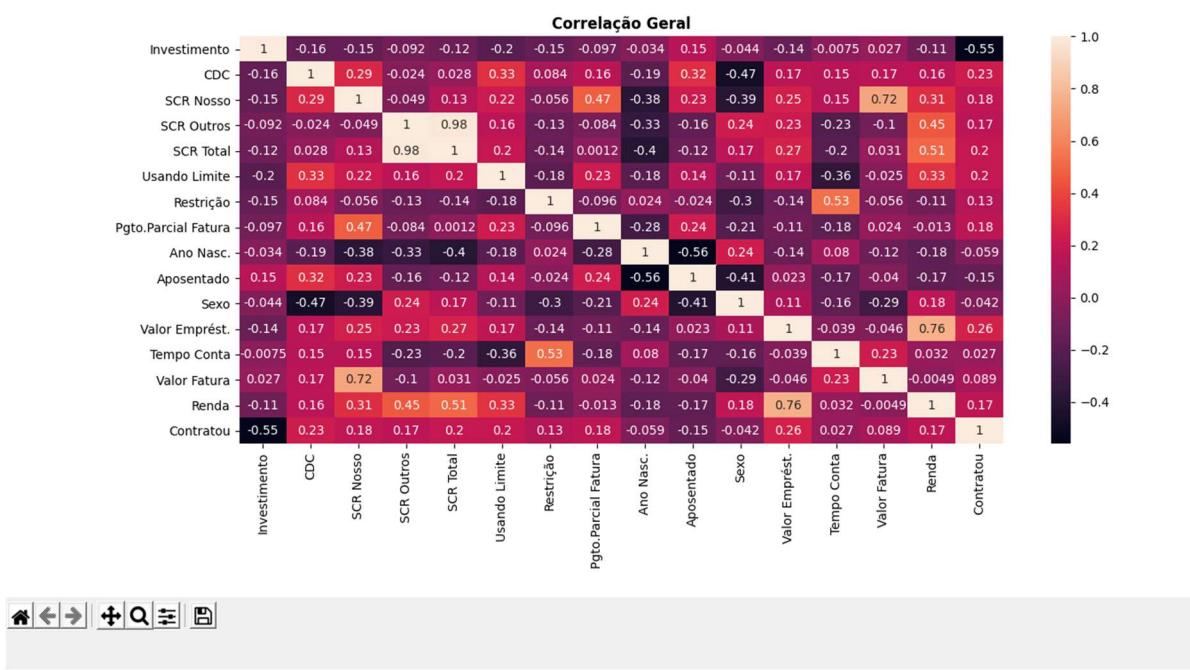


Figura C.7: Matriz de correlação Geral. Fonte: autor

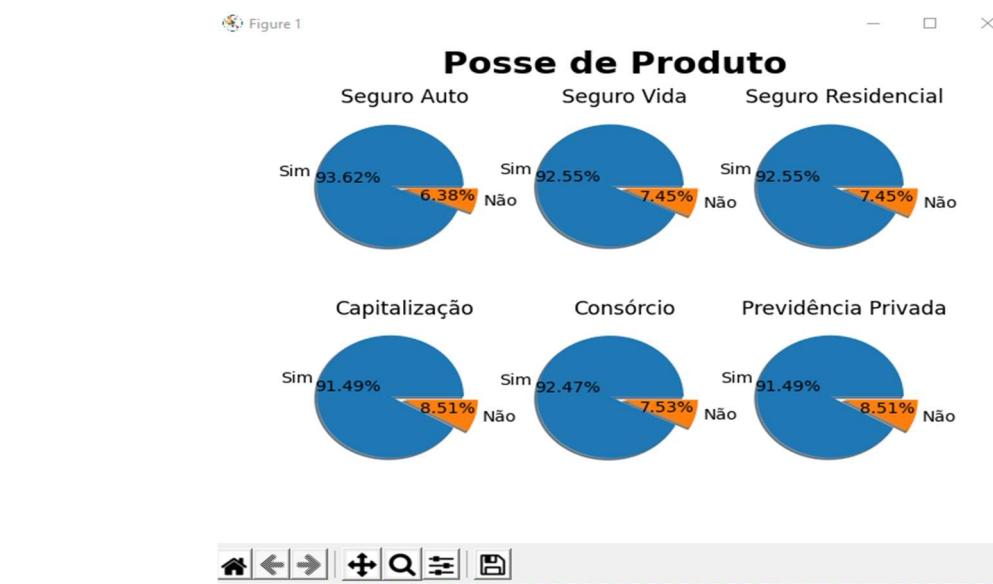
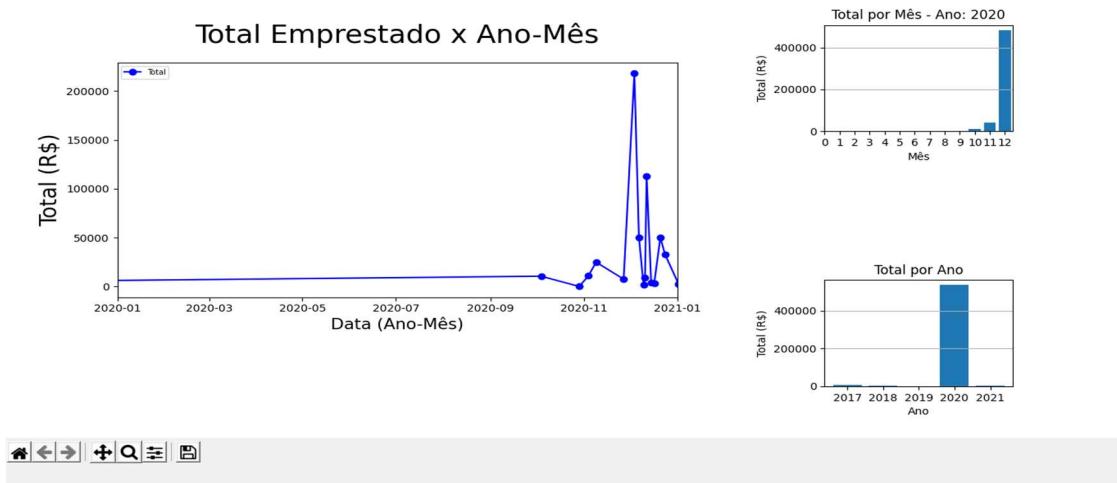


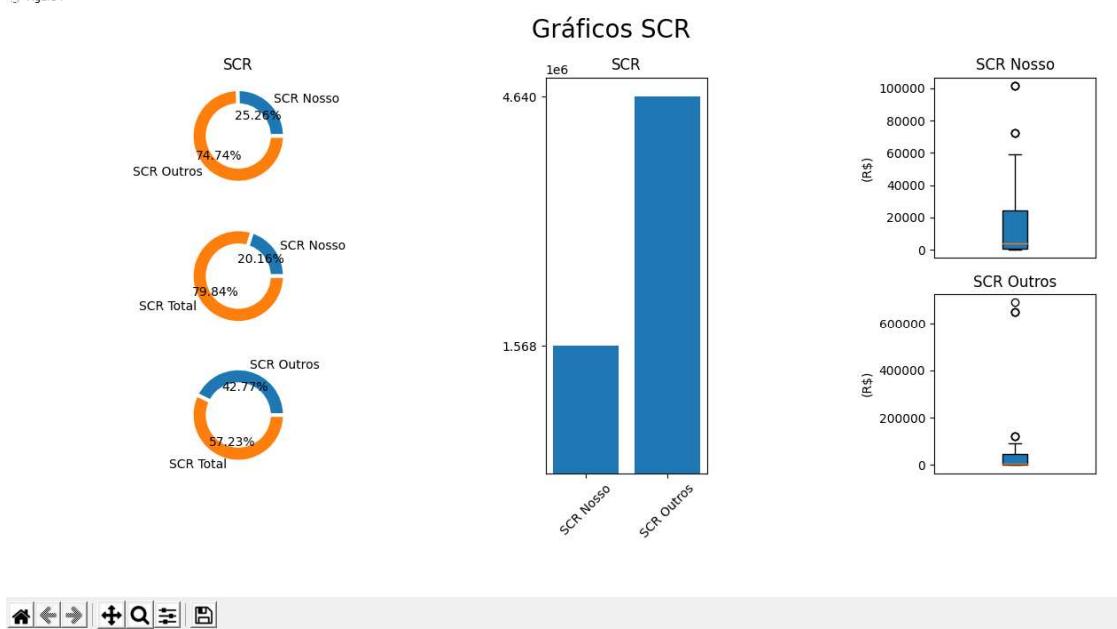
Figura C.8: Gráfico pizza posse de produto. Fonte: autor

Figure 2



**Figura C.9:** Gráfico empréstimo por ano. **Fonte:** autor

Figure 1



**Figura C.10:** Gráficos SCR. **Fonte:** autor

Estatísticas

### Estatísticas

	Investimento	SCR Nosso	SCR Outros	SCR Total	Valor Empr.	Tempo Conta	Fatura Mês	Renda
Quantidade	94,00	94,00	94,00	94,00	94,00	94	94,00	94,00
Média	624,26	16680,01	49361,43	66041,44	5844,26	81	660,62	4719,46
Desvio Padrão	2332,97	24078,75	133348,58	134342,02	11106,00	65	1750,33	7987,62
Minimo	0,00	0,00	0,00	0,00	0,00	5	0,00	1603,49
Percentil 25	0,00	850,00	1750,50	5250,00	512,50	12	0,00	2300,00
Percentil 50	0,00	4000,00	5800,00	23400,00	2253,50	84	0,00	4805,64
Percentil 75	0,00	24475,00	47200,00	72400,00	5000,00	140	666,36	42960,40
Maximo	11584,00	101500,00	690000,00	690000,00	53000,00	192	9681,00	9681,00

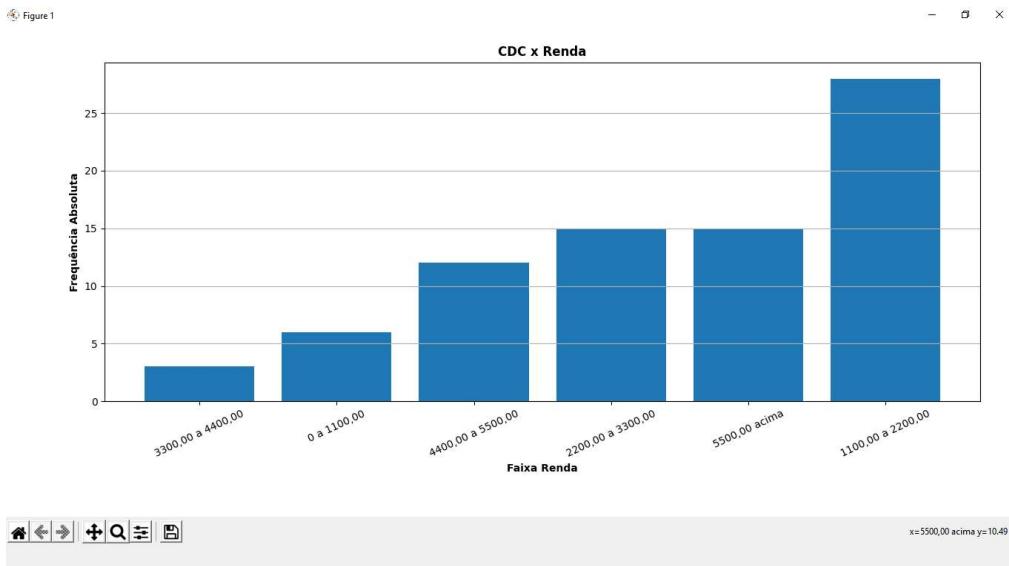
Contratam Empréstimos					Aposentados					Usando Limite					Restrição em Histórico					Pagamento parcial da fatura				
Sexo	nao	sim	Total	Sexo	nao	sim	Total	Sexo	nao	sim	Total	Sexo	nao	sim	Total	Sexo	nao	np	sim	Total				
Fem	9	42	51	Fem	24	27	51	Fem	30	21	51	Fem	42	3	6	Fem	9	42	51	Fem	9	34	43	
Mas	9	34	43	Mas	37	6	43	Mas	30	13	43	Mas	40	3	0	Mas	18	76	94	Total	61	33	94	
Total	18	76	94	Total	61	33	94	Total	60	34	94	Total	67	27	94	Total	82	6	6	Total	82	6	94	

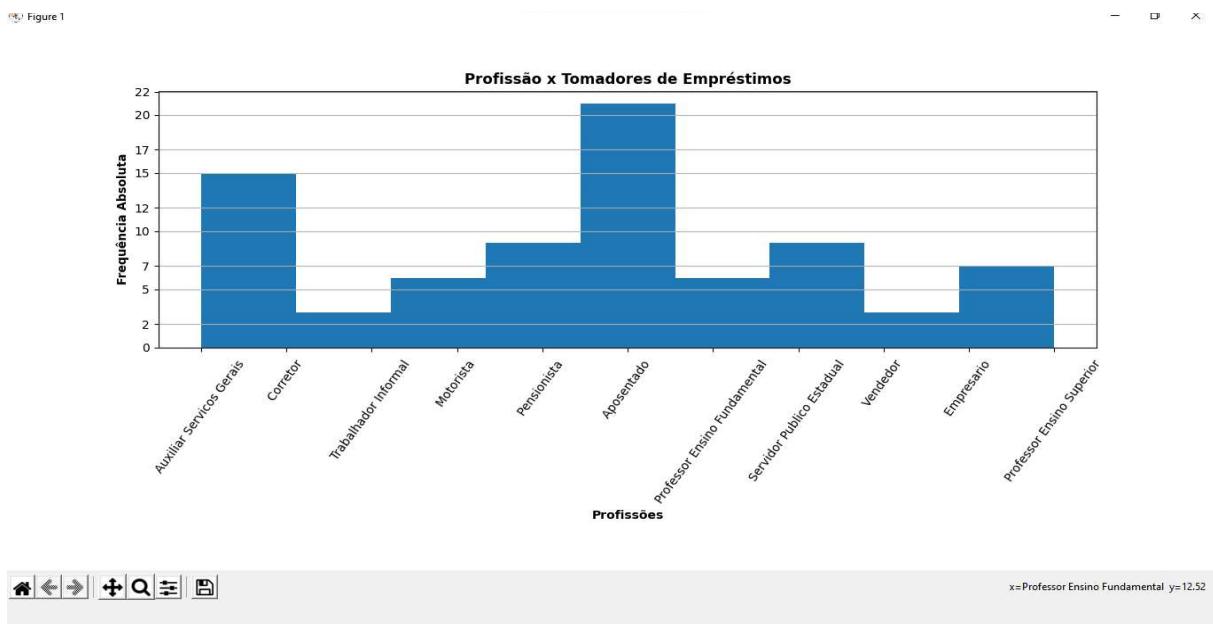
Exibir Estatísticas

Seguros	Outras Seguridades
---------	--------------------

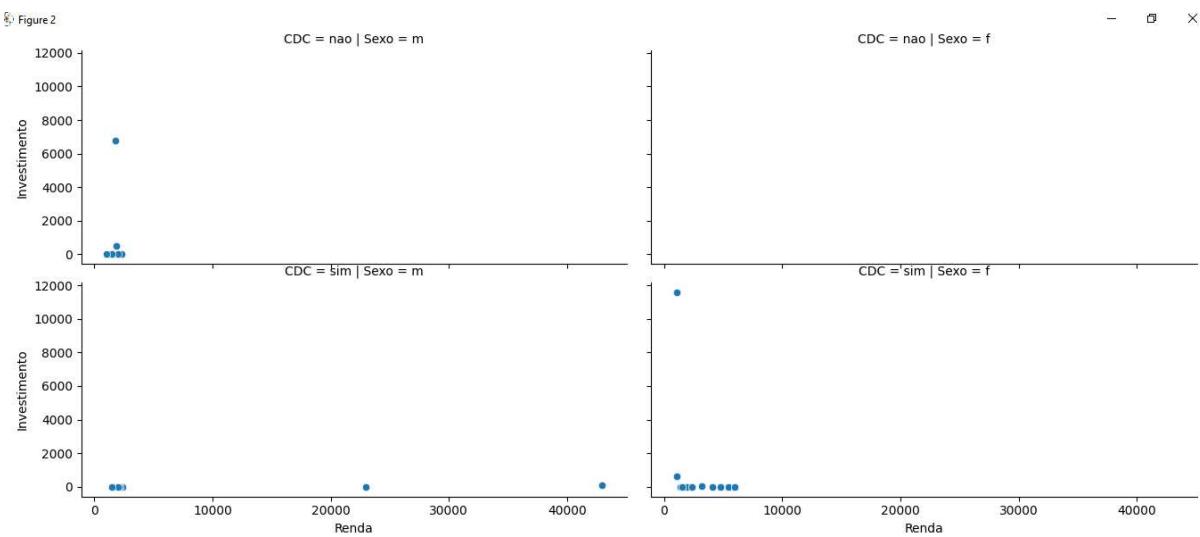
**Figura C.11:** Estatísticas. **Fonte:** autor



**Figura C.12:** Gráfico renda tomadores. **Fonte:** autor

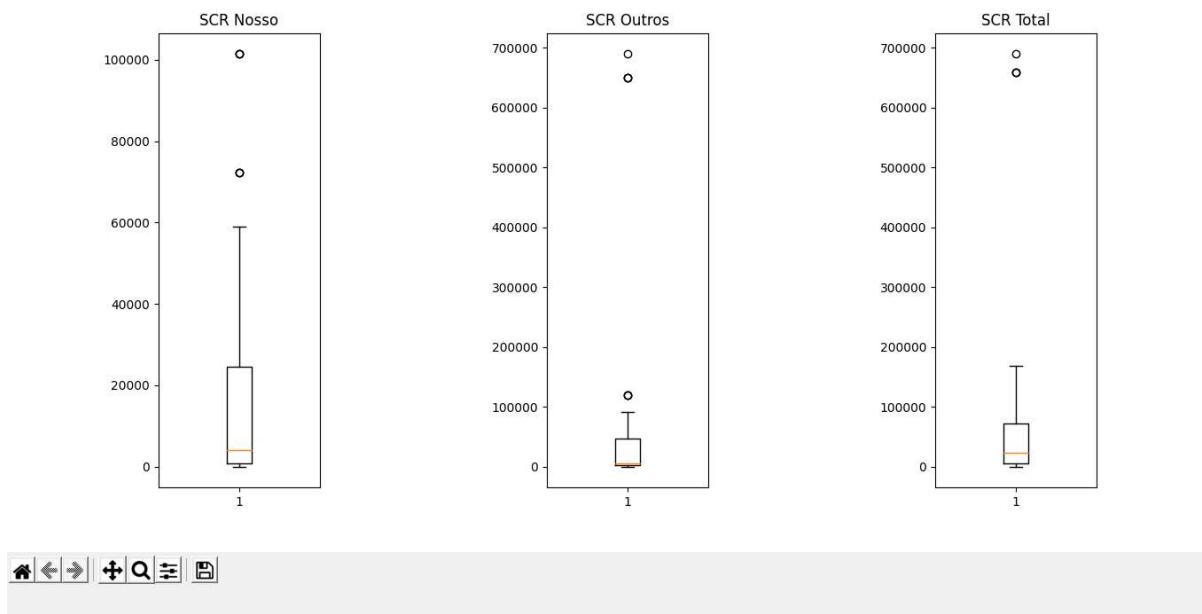


**Figura C.13:** Gráfico profissão tomadores. **Fonte:** autor



**Figura C.14:** Gráfico scatter CDC. **Fonte:** autor

Figure 1



**Figura C.15:** Gráfico Boxplot SCR. **Fonte:** autor

Seguridade

<b>Seguridade</b>	<b>Posse</b>
<input checked="" type="radio"/> Seguro Auto <input type="radio"/> Seguro Vida <input type="radio"/> Seguro Residencial <input type="radio"/> Capitalização <input type="radio"/> Consórcio <input type="radio"/> Previdência Privada	<input type="radio"/> Não <input checked="" type="radio"/> Sim <input type="radio"/> Geral

SEGUR\_O\_AUTO

	Id	Nome	Seguro Auto	Telefone
1	410000041	cliente40	sim	888-888
2	220000022	cliente21	sim	2121-2121
3	250000025	cliente24	sim	2424-2424
4	540000054	cliente53	sim	2121-2121
5	570000057	cliente56	sim	2424-2424
6	890000089	cliente88	sim	2424-2424
7	240000024	cliente23	sim	2323-2323
8	560000056	cliente55	sim	2323-2323
9	880000088	cliente87	sim	2323-2323
10	100000010	cliente9	sim	999-999
11	420000042	cliente41	sim	999-999
12	740000074	cliente73	sim	999-999
13	270000027	cliente26	sim	2626-2626
14	590000059	cliente58	sim	2626-2626
15	910000091	cliente90	sim	2626-2626
16	170000017	cliente16	sim	1616-1616
17	490000049	cliente48	sim	1616-1616
18	750000075	cliente74	sim	1010-1010
19	850000085	cliente84	sim	2020-2020

**Figura C.16:** Posse de seguridade. **Fonte:** autor

## APÊNDICE D – Base de Dados

A seguir, segue um registro da base, a título de ilustração:

id	nome	investimento	cdc_antes	scr_nosso	scr_outros	scr_total	usando_limite	restricao_hist	pgto_parcial_fatura	ano_nasc	profissao	aposentado	sexo	valor_empr
360000036	cliente35	6792.01	nao	0	0	0	nao	nao	nao	1984	auxiliar_senao	m		0
tempo_conta	linha	data	contato	fone	valor_fatura	renda	telefone	contratou	seguro_auto	seguro_residencial	seguro_vida	capitalizacao	previdencia_privada	consorcio
72	nenhuma	10/29/20	nao		0	1822,13	333-333	nao	nao	nao	nao	nao	nao	nao

**Figura D.1:** Exemplo de registro da base de dados. **Fonte:** autor

A base de dados utilizada trata-se de um arquivo com extensão .csv, contendo 29 colunas, em que cada uma representa um atributo do registro. A seguir, são discriminados os atributos de cada registro da base:

Nome	Tipo	Delimitador	Significado
<b>id</b>	Numérico(int)	Ponto e vírgula	Identificador único de cada cliente
<b>nome</b>	Nominal	Ponto e vírgula	Nome do cliente
<b>investimento</b>	Numérico(float)	Ponto e vírgula	Investimento total do cliente
<b>cdc_antes</b>	Nominal	Ponto e vírgula	Informa se o cliente já tomou empréstimo anteriormente
<b>scr_nosso</b>	Numérico(float)	Ponto e vírgula	Indica o valor do endividamento total do cliente na própria instituição
<b>scr_outros</b>	Numérico(float)	Ponto e vírgula	Indica o endividamento total do cliente em concorrentes
<b>scr_total</b>	Numérico(float)	Ponto e vírgula	Valor que reflete o endividamento total do cliente

			(somando-se o scr_nosso e scr_outros)
<b>usando_limite</b>	Nominal	Ponto e vírgula	Indica se o cliente está usando o limite da conta (sim) ou não
<b>restricao_hist</b>	Nominal	Ponto e vírgula	Indica se o cliente tem algum tipo de restrição no histórico de cadastro.
<b>pgto_parcial_fatura</b>	Nominal	Ponto e vírgula	Indica se a pessoa pagou parcialmente a fatura do cartão de crédito nos últimos 03 meses
<b>ano_nasc</b>	Nominal	Ponto e vírgula	Ano de Nascimento do cliente
<b>profissao</b>	Nominal	Ponto e vírgula	Indica a profissão
<b>aposentado</b>	Nominal	Ponto e vírgula	Indica se a pessoa é aposentada (sim) ou não
<b>sexo</b>	Nominal	Ponto e vírgula	Indica o sexo
<b>valor_empr</b>	Numérico(float)	Ponto e vírgula	Valor dos totais dos empréstimos
<b>tempo_conta</b>	Numérico(int)	Ponto e vírgula	Indica o tempo de conta, desde a abertura, até a data de levantamento, em meses
<b>linha</b>	Nominal	Ponto e vírgula	Diz sobre a linha de crédito tomada na ligação
<b>data</b>	Nominal	Ponto e vírgula	A data da contração do empréstimo
<b>contatofone</b>	Nominal	Ponto e vírgula	Indica se houve contato telefônico junto ao cliente nos três meses anteriores
<b>valor_fatura_mes</b>	Numérico(float)	Ponto e vírgula	Indica o valor total da fatura do cartão

<b>renda</b>	Numérico(float)	Ponto e vírgula	Renda total bruta
<b>telefone</b>	Nominal	Ponto e vírgula	Número de telefone
<b>seguro_auto</b>	Nominal	Ponto e vírgula	Indica se possui (sim) ou não seguro auto
<b>seguro_residencial</b>	Nominal	Ponto e vírgula	Indica se possui (sim) ou não seguro residencial
<b>seguro_vida</b>	Nominal	Ponto e vírgula	Indica se possui (sim) ou não seguro vida
<b>capitalizacao</b>	Nominal	Ponto e vírgula	Indica se possui (sim) ou não capitalização
<b>previdencia_pri-vada</b>	Nominal	Ponto e vírgula	Indica se possui (sim) ou não previdência privada
<b>consorcio</b>	Nominal	Ponto e vírgula	Indica se possui (sim) ou não consórcio
<b>contratou</b>	Nominal	Ponto e vírgula	Indica se o cliente contratou(sim) algum produto no momento da oferta.

**Quadro D.2:** Descrição da base de dados. **Fonte:** autor

---

## APÊNDICE E

### DISPONIBILIDADE DOS ARQUIVOS

#### E.1. Base de Dados:

[https://github.com/RichardTompson/EngenhariaSoftware-/blob/main/planilha\\_tcc.csv](https://github.com/RichardTompson/EngenhariaSoftware-/blob/main/planilha_tcc.csv)

#### E.2. Código Fonte:

[https://github.com/RichardTompson/EngenhariaSoftware-/blob/main/software\\_tcc.py](https://github.com/RichardTompson/EngenhariaSoftware-/blob/main/software_tcc.py)

#### E.3 Trabalho de conclusão de curso:

[https://github.com/RichardTompson/EngenhariaSoftware-/blob/main/meu\\_pre\\_projeto\\_ad.docx](https://github.com/RichardTompson/EngenhariaSoftware-/blob/main/meu_pre_projeto_ad.docx)

## APÊNDICE F – Artefatos Scrum utilizados no gerenciamento do ciclo de desenvolvimento do software

Visão do Produto						
Para	Usuários em Geral					
Que	transforma dados brutos em informações úteis e facilita a manipulação dos dados					
O	Gerenciador TCC					
É um	programa em linguagem python					
Que	permite consultas rápidas de informações dos registros das planilhas					
Diferentemente	do procedimento atual para consultar planilhas, mais demorado e menos intuitivo, dependente de filtros, seleções e conhecimento sobre funções da Excel/Libreoffice, que muitos não possuem					
O Nossa Produto	permite que qualquer usuário possa realizar consultas diretamente, de maneira mais simples, além de otimizar o tempo despendido.					

Backlog do Produto Priorizado								
HU	Item Backlog	Importância	Story Point	Sprint	Como Demonstrar	Quem Solicitou	Ready?	Obs.
003	Interface/menu	10	8	1	Criar Interface e Menu.	Usuário	Sim	
019	Sem Restrições e Uso Limite	10	3	2	Consultar clientes que não possuem restrições vigentes E estejam usando o limite da conta: QUANDO solicitar que o sistema que exiba os clientes sem restrições E usando o limite ENTÃO o sistema exibirá uma tabela com os campos: id, nome, se está usando o limite e telefone, dos clientes sem restrições E que estejam usando o limite da conta, numa tabela.	Usuário	Sim	
020	Sem Restrições e Parcial Cartão	10	3	1	Consultar clientes que tenham posse OU não de alguma seguridade	Usuário	Sim	
027	Posse de Seguridade	10	8	1	consultar cliente que já temparam cdc antes: QUANDO solicito que o sistema exiba os clientes sem restrições E que pagaram parcialmente a fatura do cartão ENTÃO o sistema exibirá uma tabela com os campos: id, nome, se pagou parcialmente a fatura, se tem restrição e telefone desse público, numa tabela.	Usuário	Sim	
028	Gráfico Seguridade	10	3	2	visualizar a quantidade de clientes que temham posse OU não de alguma seguridade	Usuário	Sim	
002	Lista Geral de Clientes	7	1	2	Exibir lista geral de clientes: QUANDO solicito visualizar todos os clientes da base ENTÃO o sistema retorna todos os clientes cadastrados em uma tabela.	Usuário	Sim	
005	Empréstimo Posse	10	1	1	Consultar clientes que já tomaram cdc antes: QUANDO solicito que o sistema exiba os clientes que já tomaram empréstimos antes ENTÃO o sistema exibirá uma tabela com os campos: id, nome, se já tomou cdc e telefone. Será exibidas informações gerais de clientes que já tomaram empréstimo antes.	Usuário	Sim	
021	Sem Restrições e Aposentados	10	3	1	Consultar clientes que não possuem restrições vigentes E são aposentados: QUANDO solicito que o sistema exiba os clientes sem restrições E que já são aposentados ENTÃO o sistema exibirá uma tabela com os campos: id, nome, se é aposentado, se tem restrição e telefone desse público, numa tabela.	Usuário	Sim	
025	Estatísticas	7	7	2	consultar estatísticas sobre os dados	Usuário	Sim	
006	Empréstimo Sem Posse	5	1	2	consultar clientes que não tomaram cdc antes: QUANDO solicito que o sistema exiba os clientes que não tomaram empréstimos anteriormente ENTÃO o sistema exibirá uma tabela com os campos: id, nome, se já tomou cdc e telefone. Será exibido apenas informações de clientes que não tomaram empréstimo.	Usuário	Sim	
013	Consulta SCR	5	2	2	Efectuar pesquisa de endividamento no SCR, total, nosso e outros: QUANDO seleciono a opção de consultar o endividamento ENTÃO o sistema exibirá uma tela para o usuário selecionar qual endividamento ele deseja consultar: total, nosso ou outros. A partir da seleção, o sistema exibirá os(s)registro(s), com os campos: id, nome, endividamento (total, nosso ou outros, de acordo com o selecionado pelo usuário) e o telefone, numa tabela.	Usuário	Sim	
026	Correlação	5	5	2	consultar alguma medida estatística, algo tipo	Usuário	Sim	
003	Consulta por ID	2	3	3	Consultar clientes pelo id	Usuário	Sim	
023	Investimentos Ordenados	2	1	3	consultar os valores investidos em ordem crescente OU decrescente	Usuário	Sim	
004	Consultar Profissões	1	3	3	filtrar clientes pela profissão	Usuário	Sim	
007	Gráfico profissão dos Tomadores	1	1	3	consultar a profissão dos tomadores de empréstimo	Usuário	Sim	
014	Filtrar Endividamento no SCR	1	1	3	selecionar os clientes com saldo de SCR maior E menor que determinado valor	Usuário	Sim	
009	Gráfico Tomadores	3	1	3	consultar a quantidade de tomadores. E não tomadores de empréstimo	Usuário	Sim	
012	Gráfico relação empréstimo e outros atributos	3	3	3	visualizar se existe uma relação entre a renda E o sexo E os valores de investimento E a posse de CDC	Usuário	Sim	
008	Gráfico Renda dos Tomadores	1	1	3	consultar a faixa de renda dos tomadores de empréstimo	Usuário	Sim	
010	Gráfico Posse de Emprestimos	1	1	4	Exibir gráfico comparativo entre a quantidade de clientes que já tomaram CDC antes e não tomaram	Usuário	Sim	
011	Gráfico Emprestimos por Ano	1	1	4	visualizar a quantidade de empréstimo efetuado por ano	Usuário	Sim	
015	Gráfico SCR	1	1	4	Exibir gráfico comparativo entre os endividamentos 'nossa e' 'outros'	Usuário	Sim	
016	Uso de Limite	1	1	4	consultar clientes que não estão no limite da conta	Usuário	Sim	
017	Gráfico Uso de Limite	1	1	4	Exibir gráfico comparativo entre os cliente que estão ou não usando o limite da conta	Usuário	Sim	
018	Histórico de Restrições	1	1	4	consultar clientes com restrições vigentes OU sem restrições	Usuário	Sim	
022	Gráfico Restrições	1	1	4	visualizar o percentual de clientes com restrições E não em um gráfico	Usuário	Sim	
024	Gráfico Investimentos	1	1	4	consultar os valores investidos em ordem crescente OU decrescente	Usuário	Sim	

Total de Pontos: 67

Figura F.1: Backlog do Produto. Fonte: autor

Definição de Preparado do Item	
Item compreendido por todos	
Com estimativas (story points)	
Com critérios de aceitação definidos	

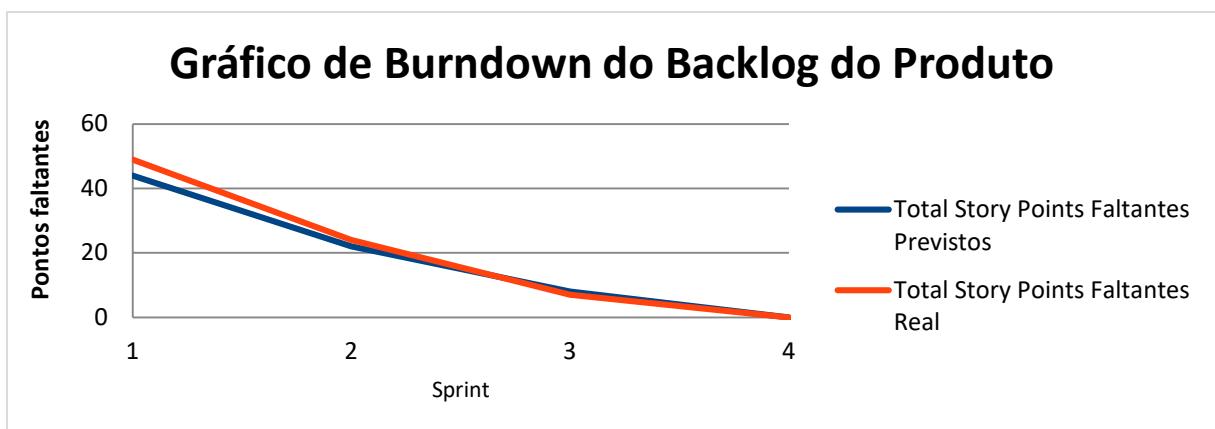
Figura F.2: Definição de Preparado. Fonte: autor

Definição de Pronto do Item				
Testado (testes unitários de caixa preta".)				
Testado (testado manualmente)				

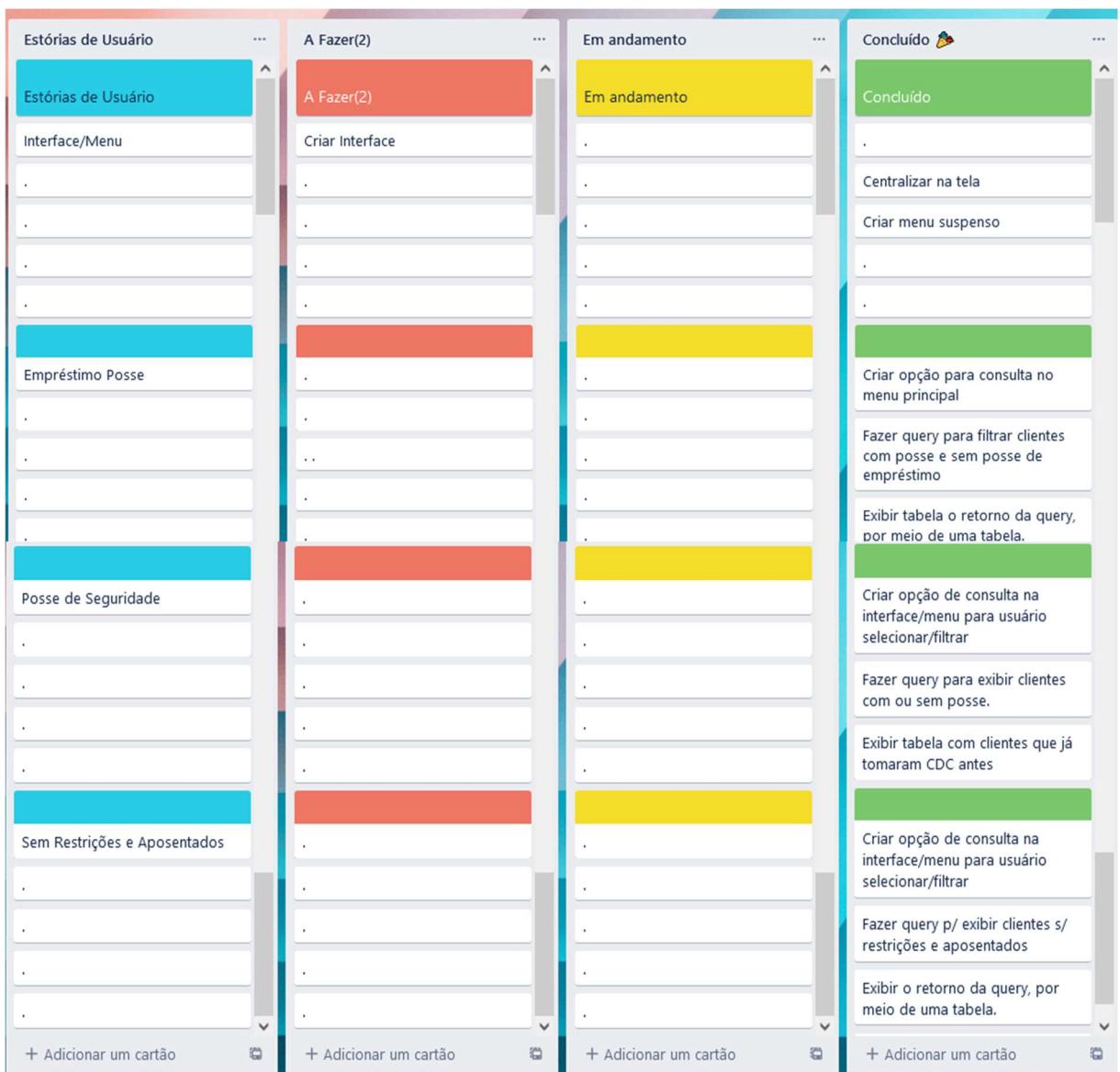
**Figura F.3:** Descrição de Pronto. **Fonte:** autor

Acompanhamento do Projeto				
Sprint	1	2	3	4
Inicio	01/08/2021	08/08/2021	15/08/2021	22/08/2021
Story Points Previstos	23	22	14	8
Story Points Realizados	18	25	17	7
Story Points Realizados Acumulados	18	43	60	67
Previsão Quant. De Sprints restantes	3	1	0	0
Horas Gastas no Sprint	15	15	15	15
Produtividade(H/S/P)	0,83	0,6	0,88	2,14
Total Story Points Faltantes Previstos	44	22	8	0
Total Story Points Faltantes Real	49	24	7	0
Situação	Em dia	Em dia	Atrasado	Em dia

**Figura F.4:** Acompanhamento do projeto. **Fonte:** autor



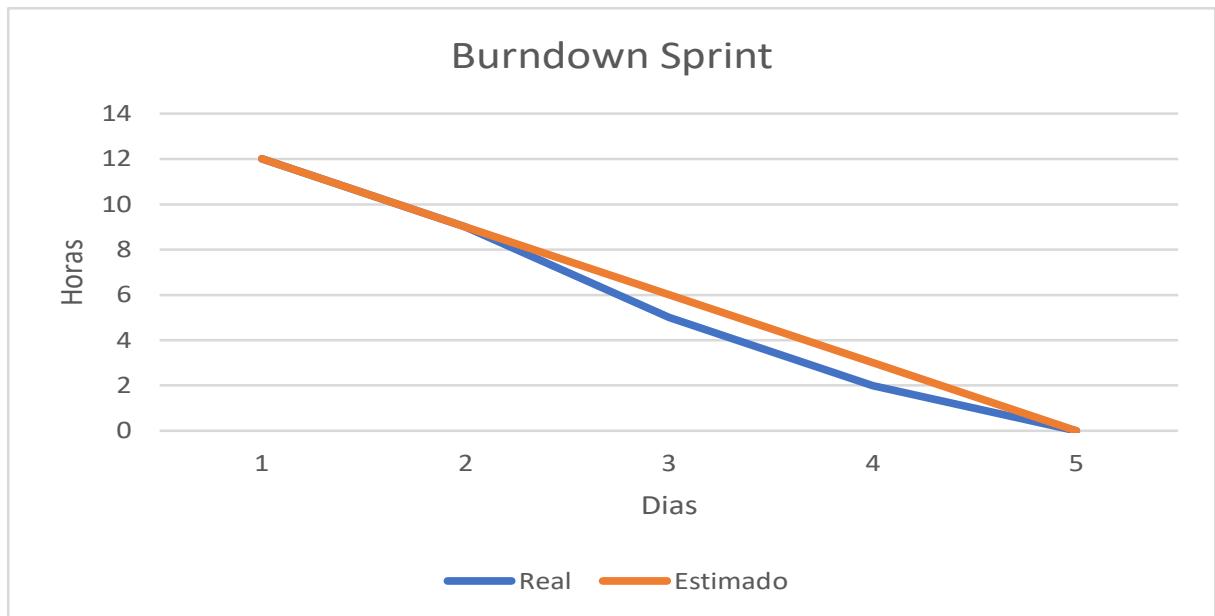
**Figura F.5:** Gráfico Burndown do Produto. **Fonte:** autor



**Figura F.7:** Scrum Board. **Fonte:** autor

Acompanhamento - Sprint (1 semana)						
	Dia	Dia 1	Dia 2	Dia 3	Dia 4	Dia 5
	Total Horas	Seg	Ter	Qua	Qui	Sex
Resolvidas	0	3	3	4	3	2
Restante Real	15	12	9	5	2	0
Restante Estimado	15	12	9	6	3	0

**Figura F.8:** Acompanhamento Sprint. **Fonte:** autor



**Figura F.9:** Burndown da Sprint. **Fonte:** autor

	Lições Aprendidas	
#	O que foi bem?	O que pode melhorar?
<b>Sprint 1</b>	Cronograma atendido	Conhecimento sobre biblioteca Pandas da linguagem de programação Python
	Dificuldade com métodos das bibliotecas	Conhecimento sobre biblioteca Tkinter da linguagem de programação Python
		Noções sobre estatísticas dos dados
<b>Sprint 2</b>	Maior facilidade no uso da linguagem Python	Conhecimento sobre biblioteca Tkinter da linguagem de programação Python
<b>Sprint 3</b>	Capacitação na linguagem de programação Python mostrou bons resultados	
	Capacitação sobre a biblioteca Tkinter do Python mostrou bons resultados	
<b>Sprint 4</b>	Cronograma atendido	Sem considerações

Plano de Ação			
#	O quê	Quem?	Quando?
<b>Sprint 1</b>	Capacitação sobre linguagem de programação Python	Programador(es)	Antes de iniciar segundo Sprint
	Capacitação sobre biblioteca Tkinter Python	Programador(es)	Antes de iniciar segundo Sprint
	Capacitação sobre biblioteca Matplotlib		
	Capacitação sobre noções de estatísticas		
<b>Sprint 2</b>	Capacitação sobre linguagem de programação Python	Programador(es)	Antes de iniciar terceiro Sprint
	Capacitação sobre biblioteca Tkinter Python	Programador(es)	Antes de iniciar terceiro Sprint
<b>Sprint 3</b>	Inserir comentários com mais frequência no código	Programador(es)/Testador(es)	Antes de iniciar quarto Sprint
<b>Sprint 4</b>	Sem considerações		Sem considerações

**Figura F.10:** Sprint Retrospective. **Fonte:** autor

## APÊNDICE G – Diagrama Casos de Uso

