**RICARDO TOMPSON FIORAVANTE**

**Analisando Dados com a Linguagem de Programação Python**

Maringá-PR

09/2021

**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 McKinney (p. 20), o volume de dados cresce exponencialmente a cada ano e estimatima-se que a quantidade de dados produzidos no mundo dobra a cada dois anos, sendo a grande maioria deles dados estruturados (como arquivos CSV) armazenados em banco de dados relacionais.

Para utilizar essas informações de forma positiva, é preciso ter conhecimento necessário para lidar efetivamente com elas. Saber operar dados é essencial para o crescimento dos negócios. Conhecer suas características e suas implicações têm grande 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 desconhecidos. 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 visulalmente os dados? Quais medidas estatísticas são comumente empregadas na análise?

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 especifícos, 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 nos apêndices C e 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 conceitos-chaves, 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 gera 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 gratuitamente, sem custos.

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 na análise 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) são a grande maioria dos dados armazenados em banco de dados relacionais ou arquivos do tipo CSV (da língua inglesa, “Character-separated values”), que são usados para armazenar dados tabulares, em que pode ocorrer de cada coluna ser de um tipo distinto (numérico, string, etc). Além disso, uma grande quantidade de conjunto de dados não-estruturados podem ser convertidos 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 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. Os mais comumente usados nas análises são: histograma, boxplots, gráficos de pizza e scatter plot.

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 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 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.

* 1. **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] | *Identificador da história de usuário* | 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 ágil Scrum. Assim, foram elicitados por meio de Histórias de Usuário, 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 extrar informações da base de dados (arquivo CSV) abordado  POSSO connsultar 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 Usuário’, 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, convencionou-se a utilização de 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 visualiza-se 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 ágil Scrum, ou seja, através de histórias de usuário. 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.

* 1. **Usuários**

Usuários que almejem 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.

* 1. **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 redudantes 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**

Diagrama

Descrição gerada automaticamente

**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 ágil de desenvolvimento 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

|  |  |
| --- | --- |
| **Atividade** | **Artefatos resultantes** |
| Engenharia de Requisitos | * Delimitação do escopo * Definição dos Stakeholders * Levantamento de requisitos (funcionais e não-funcionais) * Planejamento de teste de aceitação |
| Elaboração do Projeto de Software | * Diagrama de caso de uso * Diagrama de componentes. |
| Tecnologias utilizadas e os requisitos de implantação | * 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. * Microsoft Excel 365, versão 2108: Popular editor de planilhas, elaborado pela Microsoft. Possui funções para cálculos matemáticos/estáticos e também para construção de gráficos. * Trello: Ferramenta online gratuita para gerenciamento de projetos. É uma alternativa versátil, que possui interface intuitiva e com vários recursos úteis, como checklists. Organiza os projetos por meio de quadros, similar ao Kanban. * 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. |

Fonte: o autor.

4.1 CRONOGRAMA

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ATIVIDADES** | **Jan** | **Fev** | **Mar** | **Abr** | **Mai** | **Jun** | **Jul** | **Ago** | **Set** | **Out** | **Nov** | **Dez** |
| Pesquisa bibliográfica |  |  |  |  |  |  |  | x |  |  |  |  |
| Definição escopo |  |  |  |  |  |  |  | x |  |  |  |  |
| Levantamento, organização e priorização de requisitos |  |  |  |  |  |  |  | x |  |  |  |  |
| Escritura, depuração e teste do código-fonte | x | x | x |  |  |  |  | x | x | x | x | x |
| Conclusão do ciclo de desenvolvimento | x | x | x | x | x | x |  |  |  |  |  | x |

**REFERÊNCIAS**

ALMEIDA, Mario de Souza. **Elaboração de projeto, TCC, dissertação e tese: uma abordagem simples, pratica 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 emu ma 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.

PRESSMAN*, R. S.; MAXIM, B. R.* ***Engenharia de software: uma abordagem profissional.*** *8. ed. Porto Alegre: AMGH, 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 |

1. **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 foi desenvolvido para atender essa finalidade específica.

1. **PATROCINADOR**

Não possui

1. **GERENTE DO PROJETO**

Ricado Tompson Fioravante

1. **EQUIPE DO PROJETO**

|  |  |
| --- | --- |
| **Nome** | **Área** |
| RTF | Gerente do Projeto |

1. **OBJETIVOS DO PROJETO**

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

1. **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.

1. **FATORES PARA O SUCESSO DO PROJETO**

* Observação do cronograma
* Correção dos apontamentos da orientadora

1. **RESTRIÇÕES DO PROJETO**

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

1. **PREMISSAS DO PROJETO**

* Equipe disponível para dar sequência no projeto
* Gerente do projeto com alçada para tomada de decisões

1. **EXCLUSÕES DO PROJETO**

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

1. **ENTREGA PRINCIPAL**

* Software em funcionamento
* Artefatos do ciclo de desenvolvimento

1. **ORÇAMENTO**

* Não definido; produto sem custos

1. **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 | 14/09/201 |
| **Execução** | Documentação concluída | 20/09/2021 |
|  | Compilação e correção | 21/09/201 |
|  | Testes finalizados | 30/09/2021 |
| **Encerramento** | Lições aprendidas | 01/10/2021 |
|  | Projeto concluído | 01/10/2021 |

**APÊNDICE B – Histórias de Usuários**

|  |
| --- |
| **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 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. |
| [HU002] | Lista Geral de Clientes | POSSO consultar todos os clientes cadastrados  POIS ASSIM visualiza-se informações gerais dos clientes. |
| [HU003] | Consultar por ID | POSSO consultar clientes pelo id.  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, ENTÂO posso fazer uma consulta aos id’s da base. |
| [HU004] | Consultar profissões | POSSO consultar todos os clientes por determinada profissão  POIS ASSIM filtra-se profissões específicas para focar em determinadas profissões de interesse. |
| [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 | POSSO consultar clientes sem posse de empréstimo (CDC)  POIS ASSIM visualiza-se os clientes que não costumam tomar empréstimos e ofertar a eles. |
| [HU007] | Gráfico profissão dos tomadores | POSSO consultar a profissão dos tomadores de empréstimo  POIS ASSIM identifica-se aquelas que têm maior aceitação do produto. E compara elas entre si. |
| [HU008] | Gráfico renda dos tomadores | POSSO consultar a faixa de renda dos tomadores de empréstimo  POIS ASSIM identifica-se as faixas que têm maior aceitação do produto E compara elas entre si. |
| [HU009] | Gráfico tomadores | POSSO consultar a quantidade de tomadores E não tomadores de empréstimo  POIS ASSIM visualiza-se uma característica essencial para o relacionamento. |
| [HU010] | Gráfico posso de empréstimos | POSSO visualizar o percentual de clientes que tomaram OU não CDC em um gráfico  POIS ASSIM compara-se a quantidade de clientes tomadores E não tomadores. |
| [HU011] | Gráfico empréstimos por ano | POSSO visualizar a quantidade de empréstimo efetuado por ano  POIS ASSIM compara-se a quantidade de empréstimos realizados por período de tempo. |
| [HU012] | Relação de empréstimos e outros atributos | POSSO visualizar se existente uma relação entre a renda E o sexo E os valores de investimento E a posse de CDC  POIS ASSIM visualiza-se se existe algum vínculo entre essas características. |
| [HU013] | Consultar SCR | POSSO fazer uma consulta sobre o saldo de SCR dos clientes  POIS ASSIM visualiza-se o endividamento deles conosco E com a concorrência, comparando-as. |
| [HU014] | Filtrar endividamento no SCR | POSSO selecionar os clientes com saldo de SCR maior E menor que determinado valor  POIS ASSIM visualiza-se aqueles que possuem maior E menor endividamento |
| [HU015] | Gráfico SCR | POSSO visualizar o endividamento dos clientes no SCR em um gráfico  POIS ASSIM posso efetuar uma comparação entre os endividamentos em nossa base E na concorrência. |
| [HU016] | Uso de Limite | POSSO consultar clientes que não estejam OU estejam utilizando limite da conta  POIS ASSIM visualiza-se aqueles(as) clientes que estão OU não estão pagando juros pela utilização do limite. |
| [HU017] | Gráfico Uso de Limite | POSSO visualizar o percentual de clientes usando OU não o limite em um gráfico  POIS ASSIM posso efetuar uma comparação entre a quantidade de clientes utilizando E não utilizando limite |
| [HU018] | Histórico de restrições | POSSO consultar clientes com restrições vigentes OU sem restrições  POIS ASSIM verifica-se um característica importante sobre o perfil do clientes. |
| [HU019] | Sem restrições e uso de limite | POSSO consultar clientes sem restrições vigentes E que usam o limite  POIS ASSIM filtra-se cliente sem restrições E que estejam usando limite, que são públicos propensos a tomada de crédito. |
| [HU020] | Sem restrições e parcial cartão | POSSO consultar clientes sem restrições vigentes E que pagaram a fatura do cartão de crédito parcialmente  POIS ASSIM verifica-se os clientes sem restrições e que talvez precisem de crédito para pagar integralmente a fatura do cartão. |
| [HU021] | Sem restrições e aposentados | POSSO consultar clientes sem restrições vigentes E que são aposentados  POIS ASSIM verifica-se os clientes sem restrições e que talvez precisem de crédito com taxa diferenciada para esse público. |
|  |  | POSSO visualizar o percentual de clientes com restrições OU não em um gráfico  POIS ASSIM posso efetuar uma comparação entre a quantidade de clientes com restrições E sem restrições. |
| [HU022] | Gráfico restrições | POSSO visualizar o percentual de clientes com restrições OU não em um gráfico  POIS ASSIM posso efetuar uma comparação entre a quantidade de clientes com restrições E sem restrições. |
| [HU023] | Investimentos ordenados | POSSO consultar os valores investidos em ordem crescente OU decrescente  POIS ASSIM consulta-se os clientes com menores ou maiores investimentos que são perfis de interesse para os negócios. |
| [HU024] | Gráfico Investimentos | POSSO consultar os valores investidos em ordem crescente OU decrescente  POIS ASSIM consulta-se os clientes com menores ou maiores investimentos que são perfis de interesse para os negócios. |
| [HU025] | Estatísticas | POSSO consultar estatísticas sobre os dados  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 necessário consultar essas informações para várias caracteristicas dos clientes E também analisar a quantidade de posse de produtos, por sexo**.** |
| [HU027] | Correlação | POSSO consultar alguma medida estatística, algo do tipo  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). |
| [HU028] | Gráfico Seguridade | POSSO visualizar a quantidade de clientes que tenham posse OU não de alguma seguridade (seguro auto, residencial, vida, capitalização, previdência privada e consórcio)  POIS ASSIM compara-se a quantidade de pessoas que ja adquiram OU não algum desses produtos. |

**APÊNDICE C – Telas do Software em funcionamento**

Interface gráfica do usuário

Descrição gerada automaticamente com confiança média

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

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

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

Texto

Descrição gerada automaticamente com confiança média

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

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

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

Gráfico, Gráfico de cascata

Descrição gerada automaticamente

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

Gráfico, Gráfico de pizza

Descrição gerada automaticamente

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

Gráfico

Descrição gerada automaticamente

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

Gráfico, Gráfico de pizza

Descrição gerada automaticamente

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

Interface gráfica do usuário

Descrição gerada automaticamente

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

Gráfico, Gráfico de barras

Descrição gerada automaticamente

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

Interface gráfica do usuário, Texto, Tabela

Descrição gerada automaticamente

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

Gráfico, Gráfico de barras

Descrição gerada automaticamente

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

Figura A13 – Profissão dos tomadoresGráfico, Histograma

Descrição gerada automaticamente

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

**Gráfico

Descrição gerada automaticamente**

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

Gráfico, Gráfico de caixa estreita

Descrição gerada automaticamente

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

**Diagrama

Descrição gerada automaticamente com confiança baixa**

**Uma imagem contendo Tabela

Descrição gerada automaticamente**

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

**APÊNDICE D – Base de Dados**

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





**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** |
| **id** | Numérico(int) | Ponto e vírgula | Identificador único de cada cliente |
| **nome** | Nominal | Ponto e vírgula | Nome do cliente |
| **investimento** | Numérico(float) | Ponto e vírgula | Investimento total do cliente |
| **cdc\_antes** | Nominal | Ponto e vírgula | Informa se o cliente já tomou empréstimo anteriormente |
| **scr\_nosso** | Numérico(float) | Ponto e vírgula | Indica o valor do endividamento total do cliente na própria instituicao |
| **scr\_outros** | Numérico(float) | Ponto e vírgula | Indica o endividamento total do cliente em concorrentes |
| **scr\_total** | Numérico(float) | Ponto e vírgula | Valor que reflete o endividamento total do cliente (somando-se o scr\_nosso e scr\_outros) |
| **usando\_limite** | Nominal | Ponto e vírgula | Indica se o cliente está usando o limite da conta (sim) ou não |
| **restricao\_hist** | Nominal | Ponto e vírgula | Indica se o cliente tem algum tipo de restrição no histórico de cadastro. |
| **pgto\_parcial\_fatura** | Nominal | Ponto e vírgula | Indica se a pessoa pagou parcialmente a fatura do cartão de crédito nos últimos 03 meses |
| **ano\_nasc** | Nominal | Ponto e vírgula | Ano de Nascimento do cliente |
| **profissao** | Nominal | Ponto e vírgula | Indica a profissão |
| **aposentado** | Nominal | Ponto e vírgula | Indica se a pessoa é aposentado (sim) ou não |
| **sexo** | Nominal | Ponto e vírgula | Indica o sexo |
| **valor\_empr** | Numérico(float) | Ponto e vírgula | Valor dos total dos empréstimos |
| **tempo\_conta** | Numérico(int) | Ponto e vírgula | Indica o tempo de conta , desde a abertura, até a data de levantamento, em meses |
| **linha** | Nominal | Ponto e vírgula | Diz sobre a linha de crédito tomada na ligação |
| **data** | Nominal | Ponto e vírgula | A data da contração do empréstimo |
| **contatofone** | Nominal | Ponto e vírgula | Indica se houve contato telefônico junto ao cliente nos três meses anteriores |
| **valor\_fatura\_mes** | Numérico(float) | Ponto e vírgula | Indica o valor total da fatura do cartão |
| **renda** | Numérico(float) | Ponto e vírgula | Renda total bruta |
| **telefone** | Nominal | Ponto e vírgula | Número de telefone |
| **seguro\_auto** | Nominal | Ponto e vírgula | Indica se possui (sim) ou não seguro auto |
| **seguro\_residencial** | Nominal | Ponto e vírgula | Indica se possui (sim) ou não seguro residencial |
| **seguro\_vida** | Nominal | Ponto e vírgula | Indica se possui (sim) ou não seguro vida |
| **capitalizacao** | Nominal | Ponto e vírgula | Indica se possui (sim) ou não capitalização |
| **previdencia\_privada** | Nominal | Ponto e vírgula | Indica se possui (sim) ou não previdência privada |
| **consorcio** | Nominal | Ponto e vírgula | Indica se possui (sim) ou não consórcio |
| **contratou** | 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

**E.2. Código Fonte:**

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

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





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

****

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

****

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

****

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

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

**Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamenteInterface gráfica do usuário, Aplicativo, Tabela

Descrição gerada automaticamente**

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



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



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



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

**APÊNDICE G – Diagrama Casos de Uso**

Diagrama

Descrição gerada automaticamente Diagrama

Descrição gerada automaticamente

Diagrama

Descrição gerada automaticamente Diagrama

Descrição gerada automaticamente

Diagrama

Descrição gerada automaticamenteDiagrama

Descrição gerada automaticamente

Diagrama

Descrição gerada automaticamente