**Estácio de Sá de Goiás**

**Campus Estação**

**Trabalho Big Data PlayStore**

**João Pedro Lino Bontempo (202103023002)**

**Gustavo Henrique Martins Seabra (202102076854)**

**Daniel Nogueira Candido (202002766565)**

**Nathan de Sousa Silva (202104298651)**

**kayky firmino lima (202202615171)**

**Fabio Lopes Vieira**

**2023**

**Goiania/Goias**

Sumário

[**1.** DIAGNÓSTICO E TEORIZAÇÃO 3](#_Toc136240353)

[1.1. Identificação das partes interessadas e parceiros 3](#_Toc136240354)

[1.2. Problemática e/ou problemas identificados 3](#_Toc136240355)

[1.3. Justificativa 3](#_Toc136240356)

[1.4. Objetivos/resultados/efeitos a serem alcançados 3](#_Toc136240357)

[1.5. Referencial teórico 3](#_Toc136240358)

[2. PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO 4](#_Toc136240359)

[2.1. Plano de trabalho 4](#_Toc136240360)

[2.2. Descrição da forma de envolvimento do público participante na formulação do projeto, seu desenvolvimento e avaliação, bem como as estratégias pelo grupo para mobilizá-los. 4](#_Toc136240361)

[2.3. Grupo de trabalho 4](#_Toc136240362)

[2.4. Metas, critérios ou indicadores de avaliação do projeto 4](#_Toc136240363)

[2.5. Recursos previstos 5](#_Toc136240364)

[2.6. Detalhamento técnico do projeto 5](#_Toc136240365)

[3. ENCERRAMENTO DO PROJETO 10](#_Toc136240366)

[3.1. Relato Coletivo: 10](#_Toc136240367)

[3.2. Relato de Experiência Individual 10](#_Toc136240368)

[3.2.1. CONTEXTUALIZAÇÃO 10](#_Toc136240369)

[3.2.2. METODOLOGIA 10](#_Toc136240370)

[3.2.3. RESULTADOS E DISCUSSÃO: 10](#_Toc136240371)

[3.2.4. REFLEXÃO APROFUNDADA 10](#_Toc136240372)

[3.2.5. CONSIDERAÇÕES FINAIS 10](#_Toc136240373)

[3.3. Relato de Experiência Individual 10](#_Toc136240374)

[3.3.1. CONTEXTUALIZAÇÃO 10](#_Toc136240375)

[3.3.2. METODOLOGIA 10](#_Toc136240376)

[3.3.3. RESULTADOS E DISCUSSÃO: 10](#_Toc136240377)

[3.3.4. REFLEXÃO APROFUNDADA 10](#_Toc136240378)

[3.3.5. CONSIDERAÇÕES FINAIS 10](#_Toc136240379)

[3.4. Relato de Experiência Individual 10](#_Toc136240380)

[3.4.1. CONTEXTUALIZAÇÃO 10](#_Toc136240381)

[3.4.2. METODOLOGIA 10](#_Toc136240382)

[3.4.3. RESULTADOS E DISCUSSÃO: 10](#_Toc136240383)

[3.4.4. REFLEXÃO APROFUNDADA 10](#_Toc136240384)

[3.4.5. CONSIDERAÇÕES FINAIS 10](#_Toc136240385)

[3.5. Relato de Experiência Individual 10](#_Toc136240386)

[3.5.1. CONTEXTUALIZAÇÃO 10](#_Toc136240387)

[3.5.2. METODOLOGIA 10](#_Toc136240388)

[3.5.3. RESULTADOS E DISCUSSÃO: 10](#_Toc136240389)

[3.5.4. REFLEXÃO APROFUNDADA 10](#_Toc136240390)

[3.5.5. CONSIDERAÇÕES FINAIS 10](#_Toc136240391)

[3.6. Relato de Experiência Individual 10](#_Toc136240392)

[3.6.1. CONTEXTUALIZAÇÃO 10](#_Toc136240393)

[3.6.2. METODOLOGIA 10](#_Toc136240394)

[3.6.3. RESULTADOS E DISCUSSÃO: 10](#_Toc136240395)

[3.6.4. REFLEXÃO APROFUNDADA 10](#_Toc136240396)

[3.6.5. CONSIDERAÇÕES FINAIS 10](#_Toc136240397)

# DIAGNÓSTICO E TEORIZAÇÃO

## Identificação das partes interessadas e parceiros

Tendo em vista que o banco de dados é focado nos App do Google playstore, nosso público alvo são grupos que buscam saber dados básicos sobre tais Apps, como por exemplo número de download, data de lançamento, se o download é pago ou gratuito, categoria e etc. com tais dados podemos ter uma base da categoria que mais foi instalado, do número de apps que possui nomes parecidos, o que pode deixar os consumidores confusos.

## Problemática e/ou problemas identificados

Identificar características sobre apps, com o intuito de identificar o melhor mercado para se investir, criando um App que está em alta no mercado atual. Tendo em vista que o mercado está repleto de concorrência, ter dados que nos mostram a área que possui uma grande dinâmica, é um grande diferencial para se ter sucesso no projeto.

## Justificativa

Tendo em vista o projeto, aprendemos a fazer a coleta de dados de um banco, vendo que a várias formas de extrair informações de acordo com a necessidade do cliente. Também podemos identificar o aprendizado ao criar a interface onde criamos uma dinâmica para melhor se ler os dados.

## Objetivos/resultados/efeitos a serem alcançados

Nesse projeto, os objetivos a serem alcançados são:

Auxiliar pessoas e empresas que buscam uma coleta de dados para fins de comparação, sendo eles a construção de projetos próprios relacionados a criação de aplicativos semelhantes.

Saber a demanda para aplicativos específicos, ordenando pela maior quantidade de pesquisas.

Verificar o crescimento de downloads durante a o período da pandemia devido ao crescimento da demanda digital.

## Referencial teórico

Como base do projeto, foram utilizados algumas referências e sites para se obter as informações necessárias como:

Site utilizado para obter o banco de dados: <https://www.kaggle.com/datasets/gauthamp10/google-playstore-apps>

Vídeos para criação do programa base: https://www.youtube.com/watch?v=vmEHCJofslg / <https://www.youtube.com/playlist?list=PLCQT7jmSF-LrwYppkB3Xdbe6QC81-ozmT>

No projeto foram utilizadas algumas bibliotecas como base do programa:

Pandas: Uma data frame fácil e eficiente para manipulação de dados com indexação integrada. Sendo uma ferramenta para leitura e escrita de dados entre estruturas in-memory e diferentes formatos como: CSV e SQL. Alinhamento automático baseado em etiquetas nos cálculos e manipulação ordenada de dados. Reformulação flexível de conjunto de dados. Essa biblioteca tem a missão de ser o bloco fundamental de construção de alto nível para fazer análises de dados práticos e do mundo real em Python disponível em qualquer língua.

PathLib: Biblioteca que oferece um conjunto de classes para lidar com caminhos do sistema de arquivos, podendo ser feito por meio de operadores, acessos e atributos a chamadas de métodos. Possui uma semântica bem definida, que elimina quaisquer falhas ou ambiguidades.

- Tkinter: Interface padrão do Python para o kit de ferramentas gráficas. Útil para a criação de interfaces simples e complexas. Geralmente é uma biblioteca compartilhada, mas pode estar conectada estaticamente ao interpretador Python.

- TkInterDnD e TkInterDnD2: Fornece suporte para arrastar e soltar objetos dentro se uma única aplicação, dentro da mesma janela ou entre janelas, sendo necessário criar uma associação de eventos.

# PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO

## Plano de trabalho

Durante o projeto foi realizado a criação de um programa em python junto ao banco de dados da PlayStore obtido através do site "Kaggle", que tinha como objetivo a pesquisa e coleta de dados desse banco para auxiliar pessoas e empresas que buscam uma coleta de dados para fins de comparação, saber a demanda para aplicativos específicos, ordenando pela maior quantidade de pesquisas e verificação do crescimento de downloads durante a o período da pandemia devido ao crescimento da demanda digital, onde os membros do grupo foram designados para diferentes funções com o objetivo de cumprir o objetivo proposto.

No início do projeto foi passado o cronograma em sala pelo professor com o prazo inicial no dia 17 de abril de 2023 e prazo final para o dia 19 de junho de 2023, além do acompanhamento em sala pelo professor durante o período de aula. A partir do dia 17 foi iniciado o projeto e passado as funções para cada integrante, a primeira etapa foi realizada do dia 17 a 24 de Abril, a segunda etapa do dia 08 ao dia 15 de Maio, a criação do programa base do projeto foi feito a partir do dia 22 de Maio e finalizado aproximadamente no dia 12 de Junho, onde nessa etapa foram feitas pesquisas e coleta de informações e dados necessários para criação do programa, e a última etapa a partir do dia 12 de Junho até o dia 19 de Junho onde foi feita a apresentação e conclusão do projeto junto ao semestre letivo.

## Descrição da forma de envolvimento do público participante na formulação do projeto, seu desenvolvimento e avaliação, bem como as estratégias pelo grupo para mobilizá-los.

Para a obtenção do projeto, utilizamos por meio da plataforma da "Kaggle" para obtermos o banco de dados, no qual foi selecionado o banco da playstore, o qual possui dados como nome dos apps, data de lançamento, número médio de downloads, localização do servidor, se o app é pago e várias outras informações a respeito de aplicativos da playstore.

## Grupo de trabalho

Durante a criação do projeto não houve divisão de etapas de forma exata, pois todos os integrantes participaram de todas as partes do projeto para que todos conseguissem entender de forma coletiva tudo que foi feito.

## Metas, critérios ou indicadores de avaliação do projeto

Com a elaboração do projeto buscamos criar uma forma pratica de absorção dos dados desejados, banco de dados da playstore, com o intuito de criar ferramentas simples para o resumo dos dados.

Para alcançar o objetivo separamos as categorias e subgrupos para ordenar os dados, e por meio desta classificação criamos métodos de busca que reserva categorias selecionadas e/ou períodos pretendidos para consulta, como por exemplo, data de lançamento e media de download.

Assim, basta o usuario situar as categorias mais relevantes para sua pesquisa na área de busca, que assim será feito uma consulta no banco e consecutivamente será apresentado apenas as informações desejadas encontradas no banco.

O que torna pratica a absorção dos dados utilizados, otimizando tempo e recursos.

## Recursos previstos

No projeto foram utilizados recursos obtidos na internet através de pesquisa sobre cada tópico, usando vídeos no Youtube como base para criação do projeto, as bibliotecas: Pandas, PathLib, TkInter e TkInterDnD, fóruns de pesquisa sobre o programa, o aplicativo "PyCharm" como recurso base de programação em Python, e o site "Kaggle" para obtenção do banco de dados utilizado.

## Detalhamento técnico do projeto

**Biblioteca:**

Texto

Descrição gerada automaticamente

**TKinter:** Biblioteca responsável pela interface gráfica do usuário.

**TkinterDnD:** biblioteca usada tem a função de arrastar e soltar.

**TkinterDnD2:** É uma extensão da TkinterDnD.

**Pandas:** Biblioteca usada para ler Datas e demonstrar.

**Pathlib:** Fornece um método mais fácil para interagir com o sistema de arquivos, independentemente do sistema operacional.

**Classes:**

**Class Application(TkinterDnD.tk):**

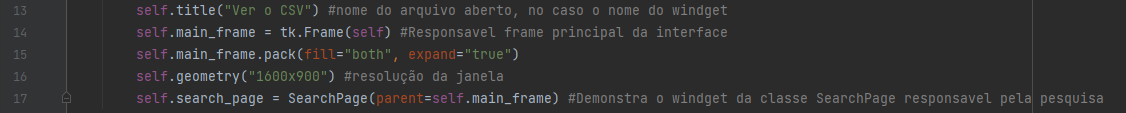
O nome da classe é Application e o (TkinterDnd.Tk) é um modulo.

Essa Classe é responsável por toda a base da interface.

Texto

Descrição gerada automaticamente

Aqui nós estamos herdando a habilidade de drag and drop na inicialização do GUI (Graphical user interface), devido ao fato de termos colocado TkinterDnD.Tk o que ao utilizar **self** acessamos os atributos e métodos da classe logo o uso da biblioteca no modulo.



**Class DataTable(ttk.Treeview):**



Um widget Treeview permite exibir dados em estruturas tabulares e hierárquicas. Nessa Classe terá todos os aspectos relacionados a parte da interface na qual mostra os dados em ordem e suas devidas colunas.

Texto

Descrição gerada automaticamente

Nessa def é feito a base como o tamanho, cor do background, as barras de scroll e a demonstração do Data Frame.

Texto

Descrição gerada automaticamente

Nesse def estamos provendo para esse data table com data frame o que fara com que seja armazenado e depois mostrar na interface.

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Nessa def será responsável por fazer as tabelas e colunas na interface

Texto

Descrição gerada automaticamente

Basicamente nessa def pairs será como um dicionário para demonstrar e organizar as colunas.



Isso fara com que a table volte para seu estado original.

**Class SearchPage(tk.Frame):**



Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Basicamente muitas funções básicas da base do código estão nessa def então recomendo ler as anotações no código

Interface gráfica do usuário, Texto

Descrição gerada automaticamente

Def para a função de arrastar e soltar basicamente aqui é utilizado a pathlib para identificar o caminho do arquivo e ler.

Texto

Descrição gerada automaticamente

Def na qual configura a função do clique duplo no qual quando pressionado ira demonstrar o arquivo csv.

Texto

Descrição gerada automaticamente

Essa é uma parte mais técnica, basicamente quando alcançarmos open embrace nos vamos criar uma variável chamada **j** e ela será o index (idx) +1 e enquanto ele n for igual ao closed embrace nos iremos acrescentar o nome com todos os caracteres que vemos no diretório do arquivo arrastado.

Texto

Descrição gerada automaticamente

Aqui nos convertemos os sistemas de colunas para um sistema de dicionários e depois passar esse dicionário na data table no caso a def find\_value na class datatable, ira executar a query\_string ate passar por todas as colunas.

# ENCERRAMENTO DO PROJETO

## Relato Coletivo:

O projeto de big data desenvolvido pelo grupo inicialmente foi feito com o objetivo de utilizar um arquivo CSV com dados da playstore para a busca e coleta de dados específicos de aplicativos, com o fim de auxiliar empresas e pessoas na coleta de informações que fossem necessárias para seus objetivos, e por fim também podendo ser utilizado como base para ler e buscar dados de outros arquivos CSV, não somente da playstore como foi apresentado inicialmente. As etapas desenvolvidas no projeto inicialmente foi a criação de um programa em Python com o auxílio de bibliotecas usadas como base, onde cada integrante do grupo participou de forma coletiva em todas as etapas de criação e desenvolvimento do código, com a ajuda de pesquisas e fontes para conclusão do programa, assim como na parte teórica do projeto como a de pesquisa, coleta de informação das partes que poderiam estar interessadas, identificação dos problemas que ocorreram durante o desenvolvimento do projeto, discussões sobre os objetivos que poderiam ser alcançados no final, auxilio na obtenção dos recursos necessários para complementar e criar o código, detalhamento técnico e por fim na apresentação em grupo.

## Relato de Experiência Individual

João Pedro Lino Bontempo

202103023002

### CONTEXTUALIZAÇÃO

O projeto consistiu de um trabalho em equipe a gente inicialmente tinha planejado fazer um código que conseguisse ler apenas o arquivo da playstore mas quanto mais a gente ia pesquisando a gente percebeu que seria mais interessante fazer um código no qual conseguisse ler qualquer tipo de csv, cada um participou do código como teve diversas mudanças toda vez alguém mexia em algo diferente.

### METODOLOGIA

Todo o projeto foi feito na sala de aula e em nossas casas, discutindo pelo whatsapp e testando em nossas maquinas.

### RESULTADOS E DISCUSSÃO:

O projeto foi bastante interessante como dito antes cada um participou um pouco de cada parte o que ajudou bastante da compreensão da minha parte em relação ao tema.

### REFLEXÃO APROFUNDADA

Inicialmente eu não tinha certeza como iriamos apresentar a parte teórica, mas com a participação dos outros membros a gente conseguiu formular tudo e fomos capazes de expandir a nossa proposta inicial.

### CONSIDERAÇÕES FINAIS

Em geral acredito que fizemos o possível, porém não fomos capazes de implementar algumas funções que queríamos no código e também não fomos capazes de otimizar o código, mas com certeza foi uma grande conquista ter feito o código ler qualquer tipo de csv.

## Relato de Experiência Individual

Nome: Gustavo Henrique Martins Seabra

Matricula 202102076854

### CONTEXTUALIZAÇÃO

No projeto em grupo de big data, primeiro pensamos em fazer um programa simples para ler um arquivos CSV, escolhemos no site kaggle um banco de dados da playstore com uma lista de todos os aplicativos e dados dos mesmos daquele respectivo ano e pesquisamos a respeito do programa para criar uma base usando algumas bibliotecas como o pandas e algumas outras, o grupo todo ajudou em todas as partes mexendo e tentando deixar o programa de uma forma que funcionasse bem e cumprisse o objetivo final, além também da parte teórica do trabalho.

### METODOLOGIA

Nosso projeto prático foi feito na maior parte em casa nas máquinas pessoais, onde cada um implementava um pouco do código e a parte teórica foi feita na maior parte durante o tempo das aulas e discutido pelo grupo do whatsapp.

### RESULTADOS E DISCUSSÃO:

A criação do projeto no geral foi bem tranquilo, além da participação de todos durante as etapas para conseguir terminar de um jeito tranquilo, e serviu para entender um pouco mais sobre a matéria de big data de uma forma prática.

### REFLEXÃO APROFUNDADA

O começo do projeto foi um pouco confuso em relação a como a gente iria começar a desenvolver e por onde começar, mas a partir de uma ideia inicial e depois de um esclarecimento das dúvidas foi bem tranquilo a criação de todas as etapas, tanto que no fim conseguimos até expandir um pouco a proposta inicial.

### CONSIDERAÇÕES FINAIS

No geral o projeto cumpriu o objetivo que esperávamos no contexto geral, porém algumas coisas ficaram faltando na implementação que não conseguimos fazer, mas nada que prejudique a proposta do programa, além de que como o tamanho do banco de dados é muito alto, em máquinas mais fracas pode ser que demore um pouco para ter o resultado, mas no geral foi uma boa experiência principalmente por conseguirmos ir além do objetivo principal.

## Relato de Experiência Individual

Daniel Nogueira Candido

202002766565

### CONTEXTUALIZAÇÃO

Tendo em vista que estamos trabalhando com Big Data, nosso projeto consiste em criar uma interface para ler e analisar dados de acordo com os filtros aplicado pelo usuário, assim, apresentando-o os dados desejados. Para tal projeto, nosso grupo, depois de algumas pesquisas e discussões, decidiu utilizar o banco de dados da Play Store extraído do site da kaggle.

### METODOLOGIA

De início foi nos apresentado um cronograma, que possui datas para concluso do projeto escrito e prático, além da apresentação. O projeto foi feito em sua grande parte na própria universidade Estácio de Sa, apenas o código foi elaborado em casa pelos nossos integrantes, Gustavo e Joao Pedro. Os quais compartilharam toda a metodologia utilizada para tal, assim, todo o grupo ficou ciente dos métodos utilizados.

Com base no cronograma, nosso projeto se saiu bem adiantado, em suma, no terceiro dia de trabalho a parte prática e escritas já estavam na fase final de conclusão. Isso foi possível graças ao trabalho conjunto, onde discutimos nos primeiros dias o que seria feito e nos dias restantes dividíamos os afazeres, relatórios escritos.

### RESULTADOS E DISCUSSÃO:

Em resumo, o trabalho foi bem tranquilo e rápido, surpreendendo minhas expectativas. O grupo foi bem colaborativos e ativos, mesmo não tendo muita afinidade entre si, salve destacar os integrantes Gustavo e Joao Pedro que tomaram a frente desde o início do projeto.

Com a elaboração do projeto, vemos que a Big Data e uma parte de supra importância para a organização dos sistemas atuais nas empresas, sendo uma grande fonte de dados logo uma fonte de informações. Como profissionais da TI temos a função de elaborar sistemas para que os usuários possam usufruir desses dados de uma forma mais direta e simplificada, e assim o fazemos.

### REFLEXÃO APROFUNDADA

Pela experiência final, foi um trabalho bem tranquilo sem muitas “dores de cabeça”, desde o início buscamos elaborar e resolver os afazeres juntos e dividir o trabalho.

### CONSIDERAÇÕES FINAIS

Tendo em vista o cronograma que nos foi passado, o qual utilizamos de base para nossos afazeres, saímos bem adiantados, finalizando o projeto em poucos dias de trabalho e ainda por cima ficando satisfeito com o trabalho elaborado, infelizmente não nos unimos com nenhuma empresa, a qual era uma das opções do projeto pra se conseguir o banco de dados, com isso não tivemos reuniões com terceiros. Entretanto, conseguimos um grande banco de dados no site da kaggle. Já que tivemos um bom entendimento, nosso grupo consegui se sair bem produtivo até o final do projeto.

## Relato de Experiência Individual

Nathan de Sousa Silva

202104298651

### CONTEXTUALIZAÇÃO

No projeto em grupo de Big Data, decidimos desenvolver um programa simples para a leitura de um arquivo CSV. Optamos por utilizar um banco de dados da Play Store, encontrado no site Kaggle, contendo uma lista de todos os aplicativos do respectivo ano. Todos os integrantes participaram tanto da criação do código como da parte teórica do projeto.

### METODOLOGIA

O projeto foi discutido e feito na sala de aula, nas nossas casas em nossas máquinas e discutido em grupo no WhatsApp, assim como a parte teórica.

### RESULTADOS E DISCUSSÃO:

A criação do projeto foi geralmente tranquila, contando com a participação de todos durante as etapas, o que nos permitiu concluí-lo de forma rápida e tranquila. Além disso, essa experiência serviu como um aprendizado prático e aprofundado do conteúdo.

### REFLEXÃO APROFUNDADA

O início do projeto foi confuso, uma vez que não tínhamos certeza de como começar. No entanto, assim que o ponto de partida foi definido, o restante do projeto transcorreu de forma tranquila, permitindo até mesmo uma expansão posterior do projeto.

### CONSIDERAÇÕES FINAIS

No geral, o projeto alcançou o objetivo que esperávamos, embora tenham ficado pendentes algumas implementações que não pudemos realizar. Nada que tenha prejudicado o programa ou algo do tipo, mas, mesmo assim, foi uma experiência muito enriquecedora, especialmente para alguém com pouca experiência em big data.

## Relato de Experiência Individual

kayky firmino lima

202202615171

### CONTEXTUALIZAÇÃO

Nesse trabalho vamos abordar leituras de um arquivo CSV, com finalidade em big data.  
utilizamos um banco aberto na play store, dosponivel dentro da kaggle, contendo uma lista de aplicativos do respectivo ano.

### METODOLOGIA

O projeto foi discutido em sala de aula, e trabalhado em casa.

### RESULTADOS E DISCUSSÃO:

A produção do projeto foi bem de boa, tendo participações de todos integrantes, e com isso o codigo foi concluido de maneia mais rapida e com eficiencia. alem da experiencia.

### REFLEXÃO APROFUNDADA

O projeto foi bem confuso, e estamos bem perdidos mas quando definimos o ponto de partida o desenvolvimento foi bem tranquilo.

### CONSIDERAÇÕES FINAIS

No entanto, o trabalho acançou o esperado o objetivo que queriamos, embora tenham ficado faltando algums questoes, que não pudemos realizar mas nada que prejudique o progama, e projeto com iniciantes pra big data, esta bem voltado e bem extruturado.