

TÓPICOS DE BIG DATA EM PYTHON
Universidade Estácio de Sá campus Nova América

Relatório de projeto, Tópicos de Big Data em Python: Padaria e restaurante
Grano & Farina

Geovanna Braga de Azevedo - 202303540485
Giovanna Ximenes Maggessi Sousa - 2023033006711

Orientado por: Lucas Antunes Floriano

2024.1
Rio de Janeiro/ RJ

Sumário

1. DIAGNÓSTICO E TEORIZAÇÃO	3
1.1. Identificação das partes interessadas e parceiros	3
1.2. Problemática e/ou problemas identificados	3
1.3. Justificativa	3
1.4. Objetivos/resultados/efeitos a serem alcançados (em relação ao problema identificado e sob a perspectiva dos públicos envolvidos)	3
1.5. Referencial teórico (subsídio teórico para propositura de ações da extensão)	3
2. PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO	4
2.1. Plano de trabalho (usando ferramenta acordada com o docente)	4
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
2.3. Grupo de trabalho (descrição da responsabilidade de cada membro)	4
2.4. Metas, critérios ou indicadores de avaliação do projeto	4
2.5. Recursos previstos	5
2.6. Detalhamento técnico do projeto	5
3. ENCERRAMENTO DO PROJETO	5
3.1. Relatório Coletivo (podendo ser oral e escrita ou apenas escrita)	5
3.2. Avaliação de reação da parte interessada	5
3.3. Relato de Experiência Individual	5
3.1. CONTEXTUALIZAÇÃO	5
3.2. METODOLOGIA	6
3.3. RESULTADOS E DISCUSSÃO:	6
3.4. REFLEXÃO APROFUNDADA	6
3.5. CONSIDERAÇÕES FINAIS	6

1. DIAGNÓSTICO E TEORIZAÇÃO

1.1. Identificação das partes interessadas e parceiros

Grano e Farina, a parte interessada, é uma empresa que faz parte do ramo alimentício, sendo uma padaria e restaurante. A empresa possui um perfil socioeconômico de médio para alto, com alta demanda em sua área. Segundo reuniões com o gerente geral, os funcionários possuem conhecimento do sistema utilizado pela empresa (Colibri), e setor administrativo utiliza de tecnologias de escritório (pacote office). Gerente geral da empresa demonstrou interesse pelo projeto, se manteve disponível para reuniões com o representante do grupo.

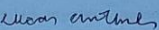


• Poderá ser solicitado à parte interessada o compartilhamento de dados não-sensíveis a serem usados durante o desenvolvimento do projeto em BIG DATA, e nunca fora deste escopo.


Aproveitamos a oportunidade e solicitamos que, em caso de aceite, seja formalizado, mediante assinatura da Carta de Autorização, as atividades e informações que o(s) aluno(s) poderá(ão) ter acesso.

Desde já nos colocamos à sua disposição para quaisquer esclarecimentos. Professor Lucas Antunes Floriano – 98820-4968 e/ou lucas.floriano@estacio.br e aluno(a) Giovanna Ximenes Maggessi Sousa – 979553402 e/ou 202303300611@alunos.estacio.br

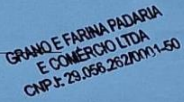
Grupo de Alunos
GIOVANNA XIMENES MAGGESSI SOUSA Matrícula: 202303300611
GEOVANNA BRAGA DE AZEVEDO Matrícula: 202303540485

Atenciosamente, 

Lucas Antunes Floriano
Docente da disciplina: TÓPICOS DE BIG DATA EM PYTHON
Semestre: 2024.1
Matrícula: 10643417


<<parte interessada>>

Rio de Janeiro, ____ de ____ de 2024.



1.2. Problemática e/ou problemas identificados

A empresa possui uma grande planilha de compras, tendo nela: mercadoria, fornecedor, valores gastos e economizados nos meses de janeiro a abril de 2024. Com mais de 8 mil linhas de dados, é impossível analisar tudo de forma prática e eficiente. Controle de estoque, gerenciamento de gastos

1.3. Justificativa

O Big Data agrega grandes volumes de informações e torna o relatório mais rápido, trabalha com gráficos e KPIs, que ajudam na compreensão das informações e medem o nível de desempenho da empresa.

1.4. Objetivos/resultados/efeitos a serem alcançados (em relação ao problema identificado e sob a perspectiva dos públicos envolvidos)

Desejamos facilitar a empresa em sua gestão de gastos, sendo possível então analisar os gastos de cada mês, quais os fornecedores mais utilizados, e quais itens a empresa vem comprando em maior quantidade. Para isso propomos esse projeto que visa simplificar a atuação da gestão sobre a empresa. Utilizando do Power BI é possível mostrar de forma ainda mais simples e de fácil entendimento.

1.5. Referencial teórico (subsídio teórico para propositura de ações da extensão)

SAGIROGLU [1] fala que a era da Big Data é uma consequência da inevitável capacidade do ser humano de gerar, coletar e armazenar dados digitais em níveis de alta demanda, e em decorrência, vem a necessidade de analisar e extrair os valores desses dados. Sendo assim, Big Data veio para analisar e avaliar as melhores decisões de acordo com as necessidades de cada empresa.

SUKAWIRASA [2] diz que o Power BI tem a capacidade de ajudar a analisar dados e tomadas de decisões diárias referentes a gestão da empresa, pois o aplicativo da Microsoft foi criado para Business Intelligence, que permite exibir visualizações, consultas e conexões de dados. E sendo possível aprender a utilizá-lo de forma rápida e eficaz, já que a sua interface é semelhante a seus aplicativos irmãos, Excel, Word etc., que grande parte dos usuários já têm certa familiaridade.

LABRINIDS [3] diz que o primeiro passo, para utilizar bem o big data, é conseguir as aquisições dos dados e filtra-los por ordem de grandeza, e gerar também os dados mais importantes como: Nome da parte interessada; Gastos excessivos do mês. Os dados do big data são uma previsão, com base nos acontecimentos da empresa nos últimos meses, por isso pode ocorrer determinadas vezes que algumas notícias sejam imprecisas.

[1] Sagioglu, Seref, and Duygu Sinanc. "Big data: A review." *2013 international conference on collaboration technologies and systems (CTS)*. IEEE, 2013.

[2] SUKAWIRASA, I. Kadek Anom et al. Implementasi Data Warehouse Dan Penerapannya Pada PHI-Minimart Dengan Menggunakan Tools Pentaho dan Power BI. **Jurnal Elektronik Ilmu Komputer Udayana P-ISSN**, v. 2301, p. 5373, 2008.

[3] LABRINIDIS, Alexandros; JAGADISH, Hosagrahar V. Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, v. 5, n. 12, p. 2032-2033, 2012.

2. PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO

2.1. Plano de trabalho (usando ferramenta acordada com o docente)

DATA	ATIVIDADE	DESCRIÇÃO
02/03	Big Data	O primeiro contato sobre o que é e para que serve o "Big Data".
09/03	Regex	Aprendemos os processos do big data.
16/03	Google Colab	Aprendemos a utilizar os códigos, pandas e csv.
22/03	Primeiro contato com a parte interessada.	Levamos a carta a parte interessada, que a assinou.

19/04	Atividade “Detetive”	Realizamos uma atividade que foi de grande importância.
25/04	Recebimento dos dados.	A parte interessada fez o envio da planilha de compras da empresa.
03/05	Gráfico Power BI.	O gráfico em Power BI, referente a planilha foi feito.
16/05	Relatório.	Começamos a preencher o relatório.
27/05	Desenvolvimento.	Grupo começou o desenvolvimento dos dados simulados.
27/05	Utilização do Git.	Grupo começou a utilizar o Git.
31/05	Término do projeto.	Finalizamos o projeto.
07/06	Apresentação.	Apresentamos o projeto ao professor orientador e a turma.

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.

Re: Gráfico de finanças, referente ao projeto Universitário da faculdade Estácio de Sá

17 de maio de 2024 18:06

Marcos Sousa para 2 destinatários

Detalhes

Boa tarde Srta. Giovanna,

Estarei analisando todas as informações enviadas, qualquer dúvida entrarei em contato.

Desde já agradecemos.

Em sex., 17 de mai. de 2024 10:21, giovannaxms <giovannaxms@gmail.com> escreveu:
Bom dia senhor Marcos, tudo bem?

Segue em anexo o gráfico de finanças que nosso grupo criou a partir da planilha de vendas de janeiro a abril de 2024 que o senhor disponibilizou.

Grata,
Giovanna Ximenes Sousa.

2.3. Grupo de trabalho (descrição da responsabilidade de cada membro)

Integrante	Função
Geovanna Braga de Azevedo	Preenchimento do relatório.
Giovanna Ximenes Maggessi Sousa	Busca pela parte interessada, comunicação com a parte interessada, criação dos gráficos em Power BI, desenvolvimento dos dados simulados em pandas, utilização do Git, preenchimento do relatório.

2.4. Metas, critérios ou indicadores de avaliação do projeto

Estabelecer conhecimento sobre os principais fornecedores e seus custos, se são ou não benéficos para empresa. Dar a possibilidade da empresa ter um controle de gastos novo e eficiente.

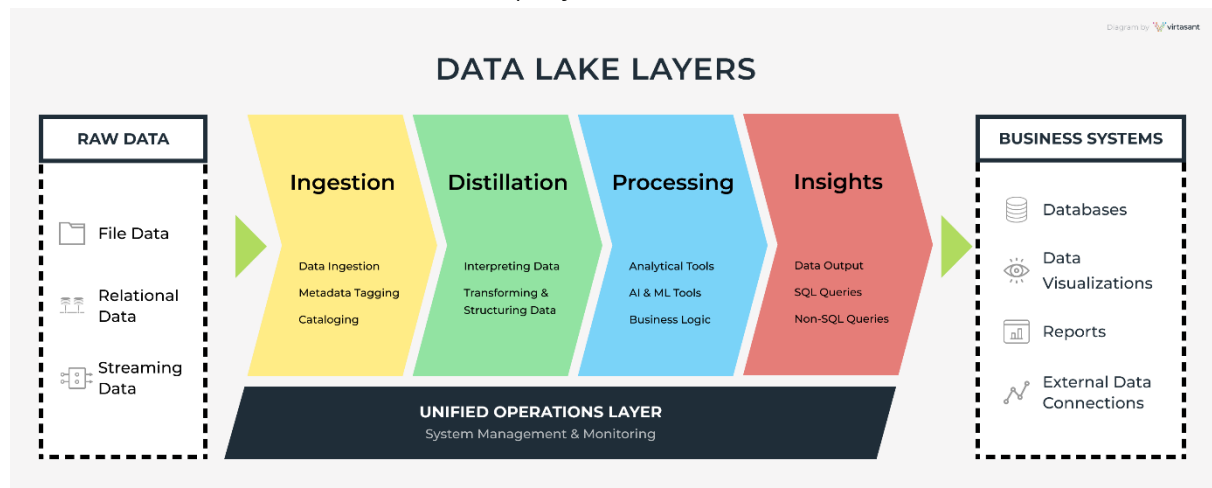
2.5. Recursos previstos

Fizemos Utilização do Excel, formato pelo qual os dados foram enviados via e-mail, Power BI para a criação dos gráficos e CSV formato que foi utilizado para extrair os dados. Google colab, Python e a biblioteca Pandas, para a elaboração dos dados simulados.

Esses são os dados essenciais que utilizamos, a planilha deve vir dessa forma.

C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
de Emissão	Data de Lançamento	Sigla	Fornecedor	Número	Situação	Cód. Material	Descrição Material	Cbde.	Unidade	Preço Unitário (R\$)	Acréscimo (R\$)	Desconto (R\$)	Total (R\$)	Grupo do Material	Conta Gerencial do Material
2024	02/01/2024	REC	UMEHARA ALIMENTOS LTDA	88341	normal	500054	Abobrinha Italiana	5,00	KG	4,50	0,00	0,00	22,50	Hortifrut	Despesas(Estoques)(Materia Prima)(Hc
2024	02/01/2024	REC	UMEHARA ALIMENTOS LTDA	88341	normal	500054	Abobrinha Italiana	8,00	KG	4,50	0,00	0,00	36,00	Hortifrut	Despesas(Estoques)(Materia Prima)(Hc
2024	02/01/2024	REC	UMEHARA ALIMENTOS LTDA	88341	normal	500075	Aneto / dill	2,00	UNI	2,25	0,00	0,00	4,50	Hortifrut	Despesas(Estoques)(Materia Prima)(Hc
2024	02/01/2024	REC	UMEHARA ALIMENTOS LTDA	88341	normal	500087	Manjericao	1,00	MC	2,75	0,00	0,00	2,75	Hortifrut	Despesas(Estoques)(Materia Prima)(Hc
2024	02/01/2024	REC	UMEHARA ALIMENTOS LTDA	88341	normal	500107	Ovo Branco Extra	1,00	CX	145,00	0,00	0,00	145,00	Hortifrut	Despesas(Estoques)(Materia Prima)(Hc
2024	02/01/2024	REC	UMEHARA ALIMENTOS LTDA	88341	normal	500113	Batata branca	10,00	KG	8,75	0,00	0,00	87,50	Hortifrut	Despesas(Estoques)(Materia Prima)(Hc
2024	02/01/2024	REC	UMEHARA ALIMENTOS LTDA	88341	normal	500118	Beterraba	3,00	KG	4,50	0,00	0,00	13,50	Hortifrut	Despesas(Estoques)(Materia Prima)(Hc
2024	02/01/2024	REC	UMEHARA ALIMENTOS LTDA	88341	normal	500122	Cenoura	2,00	KG	4,50	0,00	0,00	9,00	Hortifrut	Despesas(Estoques)(Materia Prima)(Hc
2024	02/01/2024	REC	UMEHARA ALIMENTOS LTDA	88341	normal	500122	Cenoura	8,00	KG	4,50	0,00	0,00	36,00	Hortifrut	Despesas(Estoques)(Materia Prima)(Hc
2024	02/01/2024	REC	UMEHARA ALIMENTOS LTDA	88341	normal	500137	Tomate salada	7,00	KG	8,75	0,00	0,00	61,25	Hortifrut	Despesas(Estoques)(Materia Prima)(Hc
2024	02/01/2024	REC	UMEHARA ALIMENTOS LTDA	88341	normal	500153	Manga	7,00	KG	7,00	0,00	0,00	49,00	Hortifrut	Despesas(Estoques)(Materia Prima)(Hc
2024	03/01/2024	REC	REZENDE E KONKOSKI COMERCIO ATACADISTA	218210	normal	500267	Lagarto redondo	25,16	KG	25,30	0,00	0,00	636,55	Proteinas	Despesas(Estoques)(Materia Prima)(Hc
2024	02/01/2024	REC	UMEHARA ALIMENTOS LTDA	88341	normal	500599	Mamao Formosa	2,00	UNI	11,00	0,00	0,00	22,00	Hortifrut	Despesas(Estoques)(Materia Prima)(Hc
2024	02/01/2024	REC	UMEHARA ALIMENTOS LTDA	88341	normal	500604	Uva verde thompson	2,00	CX	19,50	0,00	0,00	39,00	Hortifrut	Despesas(Estoques)(Materia Prima)(Hc
2024	02/01/2024	REC	UMEHARA ALIMENTOS LTDA	88341	normal	500605	Melao Amarelo	2,00	KG	15,00	0,00	0,00	30,00	Hortifrut	Despesas(Estoques)(Materia Prima)(Hc
2024	03/01/2024	REC	UMEHARA ALIMENTOS LTDA	88413	normal	500605	Melao Amarelo	2,00	KG	15,00	0,00	0,00	30,00	Hortifrut	Despesas(Estoques)(Materia Prima)(Hc
2024	02/01/2024	REC	RIO QUALITY COMERCIO DE ALIMENTOS LTDA	5526674	normal	500617	Geleia Frutas vermel	5,00	UNI	37,95	0,00	0,00	189,75	Secos	Despesas(Estoques)(Materia Prima)(Se

2.6. Detalhamento técnico do projeto



3. Ingestão – Os dados foram capturados por uma planilha Excel enviada por e-mail.
4. Storage – Recebemos uma planilha bem tratada, utilizando a biblioteca pandas para o desenvolvimento dos dados simulados.
5. Metadata (application) – Com o Power BI desenvolvemos um gráfico simples e explicativo.

6. ENCERRAMENTO DO PROJETO

6.1. Relato Coletivo:

Foi uma experiencia maravilhosa, acreditamos que fizemos algo importante para nossa parte interessada.

6.1.1. Avaliação de reação da parte interessada

Link do vídeo onde nossa parte interessada relata sua experiência e opiniões sobre nosso projeto: <https://youtu.be/D6P1YyEzmZY>

Relato de Experiência Individual (Pontuação específica para o relato individual)

6.1.2. CONTEXTUALIZAÇÃO

Geovanna Braga de Azevedo.

Depois de conseguir a parte interessada e receber as planilhas de compras, usamos o Power BI, para destacar pontos importantes para serem utilizados, por exemplo: qual foi o produto mais comprado, qual fornecedor, a média de compras.

Nesta parte tive muita dificuldade para entender como usava o Power BI, como separar em blocos, colocar os arquivos, mas pesquisei bastante e consegui usar.

Com os códigos do DataFrame pandas foi um pouco mais fácil, com os slides das aulas entendi no que cada código ajudava como: csv, apply, lambda....

Foi utilizado através do Google colab, para ajudar nas buscas de itens específicos na planilha.

No geral, ajudamos a parte interessada, Padaria Gran & Fino, a ver com mais clareza os produtos que mais gastam e a ter uma organização mais eficiente, com os gráficos do PowerBI, além disso com os dados simulados eles vão ter conhecimento, e a segurança, se a quantidade de funcionários na equipe daria o suporte necessário caso algum funcionário ficasse doente por 1 semana ou se uma licença a maternidade ocorresse.

Com esse trabalho aprendi a importância de uma boa organização e a diferença que faz, principalmente se tratando de uma quantidade enorme de linhas na planilha, que não são fixas podendo ocorrer mudanças ao longo do meses como: a mudança do fornecedor principal, a abertura de franquias pelo estado, ou também a diminuição por uma crise.

São tantas as possibilidades de mudanças que estar acompanhando esses dados e comparando de 3 em 3 meses ajudará muito no futuro. fazer dados simulados nesses períodos também ajudará a diminuir grandes imprevistos.

Também fiquei responsável por fazer a maior parte do relatório.

Giovanna Ximenes Maggessi Sousa:

Minha participação no projeto foi total, fiz o possível para aprender e realizar tudo da melhor forma possível. Foi de certa forma estressante, mas também muito gratificante já que senti que realmente estava ajudando a parte interessada e acabei desenvolvendo muito apressa pela área de manipulação de dados num geral.

Minhas contribuições além do relatório:

1. Desenvolvimento dos dados simulados em pandas:

```
In [ ]: def imprimir_informacoes(compras, item):
        print('Mercadoria:', item)

        item_compras = compras[compras['Descrição Material'] == item]

        preco_unitario_por_fornecedor = {}
        valor_total_por_fornecedor = {}
        quantidade_por_fornecedor = {}

        for index, row in item_compras.iterrows():
            fornecedor = row['Fornecedor']
            preco_unitario = float(row['Preço Unitário (R$)'].replace(',', ''))
            quantidade = float(row['Qtde.'].replace(',', ''))
            valor_total = row['Total (R$)']

            if isinstance(valor_total, float):
                valor_total = str(valor_total)

            valor_total = float(valor_total.replace(',', ''))

            if fornecedor in preco_unitario_por_fornecedor:
                preco_unitario_por_fornecedor[fornecedor] += preco_unitario * quantidade
                valor_total_por_fornecedor[fornecedor] += valor_total
                quantidade_por_fornecedor[fornecedor] += quantidade
            else:
                preco_unitario_por_fornecedor[fornecedor] = preco_unitario * quantidade
                valor_total_por_fornecedor[fornecedor] = valor_total
                quantidade_por_fornecedor[fornecedor] = quantidade

        for fornecedor in preco_unitario_por_fornecedor.keys():
            preco_unitario = preco_unitario_por_fornecedor[fornecedor] / quantidade_por_fornecedor[fornecedor]
            valor_total = valor_total_por_fornecedor[fornecedor]
            quantidade = quantidade_por_fornecedor[fornecedor]
            print('\n')
            print(f'Fornecedor: {fornecedor}')
            print(f'Preço Unitário(R$): {preco_unitario:.2f}')
            print(f'Valor Total(R$): {valor_total:.2f}')
            print(f'Quantidade: {quantidade}')
        print()

        top_fornecedores = ['UMEHARA ALIMENTOS LTDA', 'Hortifruti Genesis Ltda', 'Supermercados unidos (pai da luiza)']
        itens_comuns = encontrar_itens_semelhantes(compras, top_fornecedores)

        for item in itens_comuns:
            imprimir_informacoes(compras, item)
```

Mercadoria: Abobrinha italiana

Fornecedor: UMEHARA ALIMENTOS LTDA
Preço Unitário(R\$): 4.41
Valor Total(R\$): 528.82
Quantidade: 120.0

Fornecedor: Hortifruti Genesis Ltda
Preço Unitário(R\$): 5.58
Valor Total(R\$): 942.48
Quantidade: 169.0

2. Utilização do Git e GitHub, mantendo a organização do projeto:

ProjetoBigData
Public

Unpin
Unwatch 1
Fork 0
Star 1

main
5 Branches
0 Tags

Add file
Code

ginaxim

Rename roteiro semi-final.docx to Relatório.docx

98a180f · now

60 Commits

Dados-Simulados.py	Update and rename projeto_bigdata_final.py to Dados-Simu...	10 hours ago
Dados-Simulados_Colab.ipynb.ipynb	Rename Cópia de Projeto BigData Final.ipynb to Dados-Sim...	9 hours ago
Gráficos Grano & Farina.pbix	Rename RelatórioCompras - BI.pbix to Gráficos Grano & Fari...	8 hours ago
README.md	Update README.md	yesterday
Relatório.docx	Rename roteiro semi-final.docx to Relatório.docx	now
VendasGrano_csv.csv	Add files via upload	yesterday

Projeto de Tópicos de Big Data em Python, disciplina de extensão cursada no 3º período do curso de Análise e Desenvolvimento de Sistemas na UNESA.

• O objetivo do projeto é lidar com a manipulação de dados em grandes volumes. Com uma planilha disponibilizada pela parte interessada, deve-se manipular os dados, criar dados simulados, utilizar o git para upload de códigos e a criação de gráficos Power BI.

• Nossa parte interessada, uma padaria e restaurante, Grano & Farina, enviou sua planilha de compras referente aos meses de janeiro a abril de 2024. Devido à planilha já estar devidamente tratada, realizamos dados simulados baseados nos fornecedores, descrição do material, valor unitário e valor total. Dessa forma, podemos simular qual dos dois fornecedores principais tem o maior custo-benefício para a empresa.

About

Projeto de extensão da matéria tópicos de BigData da Universidade Estácio de Sá.

Readme

Activity

1 star

1 watching

0 forks

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

Jupyter Notebook 89.5%

Python 10.5%

Suggested workflows

Based on your tech stack

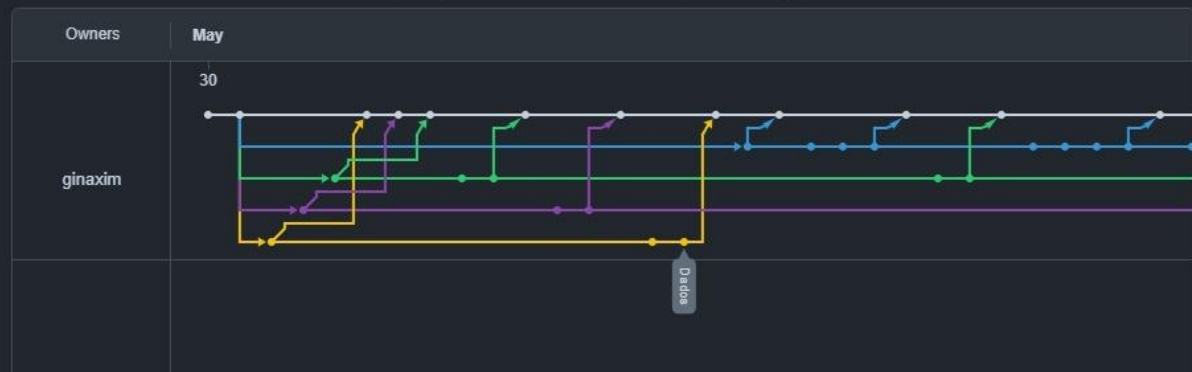
Django

Build and Test a Django Project

Configure

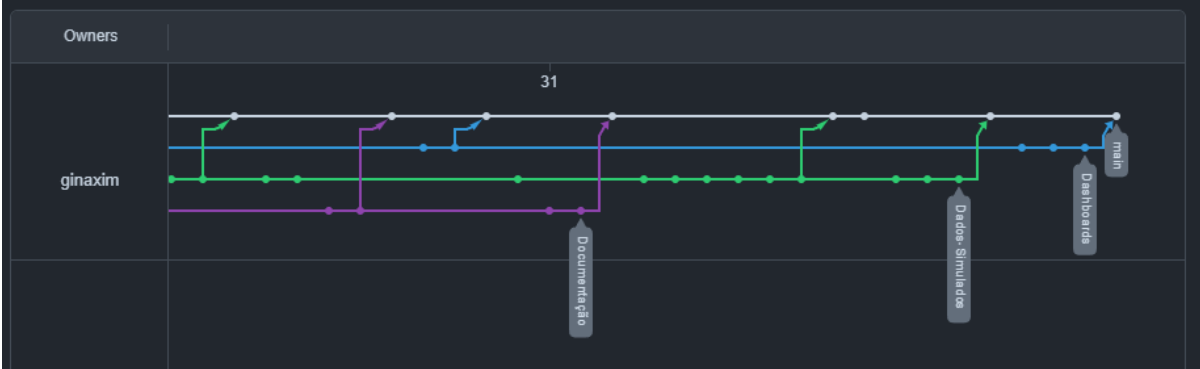
Network graph

Timeline of the most recent commits to this repository and its network ordered by most recently pushed to.

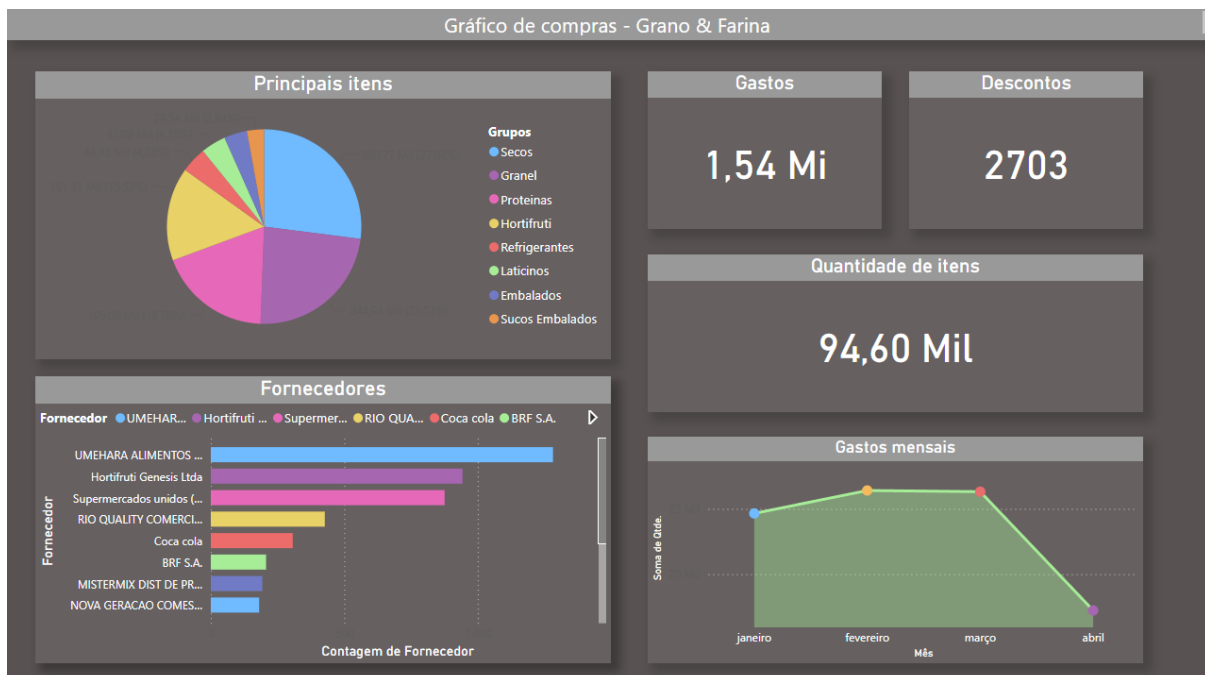


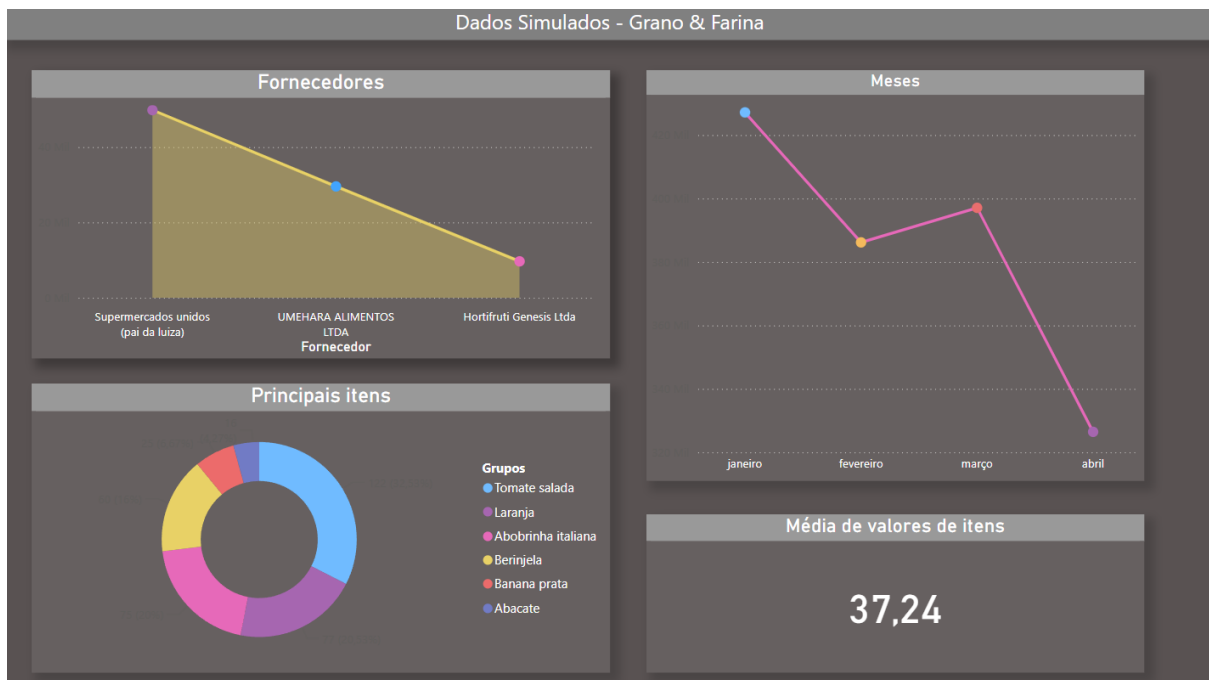
Network graph

Timeline of the most recent commits to this repository and its network ordered by most recently pushed to.



3. Criação dos gráficos em BI, de compra e dados simulados da parte interessada:





Link do repositório GitHub: <https://github.com/ginaxim/ProjetoBigData>

6.1.3. METODOLOGIA

Giovanna Ximenes Maggessi Sousa:

As aulas com o professor orientador foram muito intuitivas e tranquilas, com exercícios práticos que foram de suma importância para o desenrolar do projeto além das infinitas dúvidas que possuía que eram devidamente respondidas com muita calma.

6.1.4. RESULTADOS E DISCUSSÃO:

Giovanna Ximenes Maggessi Sousa:

Como aluna me sinto realizada em finalizar esse projeto sabendo que dei meu melhor, é claro que tenho muito a aprender, mas já com esse pequeno conhecimento me sinto realmente bem. Descobri um possível interesse em uma futura carreira, descobri que as vezes precisamos ser firmes como líderes de projeto e que nem todos possuem o mesmo senso de responsabilidade que você. Apesar de todo o estresse e dificuldades que um projeto traz, de forma geral, este foi realmente bom.

6.1.5. REFLEXÃO APROFUNDADA

Giovanna Ximenes Maggessi Sousa:

Como dito anteriormente, um projeto é algo bem estressante para os alunos, mas quando tudo sai como esperado, é extremamente gratificante.

6.1.6. CONSIDERAÇÕES FINAIS

Giovanna Ximenes Maggessi Sousa:

Como apresentado no vídeo de entrevista, a parte interessada demonstra grande interesse em utilizar nossos gráficos e fazer um estudo real utilizando-os como base.