

Universidade Estácio de Sá

Campus Nova América

Projeto de Big Data para Six Zen Acessórios

Integrantes:

Pedro Henrique(202102148201)

André Luiz Gomes de Araujo(202203753851)

Luiz Felipe Dias Cardoso Feres Lima(202002456086)

Gabriel de Paulo Mulinari(202203961004)

Renan Soares Candido(202208622577):

Professor Orientador:

Lucas Antunes Floriano

2023

Rio de Janeiro

1) Diagnóstico e Teorização

1.1 - Six Zen Acessórios

Idealizadora Kenia Gomes Bastos, pós-graduada em Pedagogia, mulher, 50 anos.

Trabalhamos com projeto de criação de acessórios exclusivos, destinados ao público feminino.

Atualmente praticamos parceria com 2 fábricas, Ourivesaria e Lapidário de Pedras Naturais.

Nossa equipe é composta por 5 pessoas na mão de obra, produzindo rigorosamente o protótipo do design, 5 vendedoras distribuídas em pontos de vendas variáveis.

Já identificamos nosso público, mulheres que buscam auxílio terapêutico dos cristais, faixa de idade 25 até 65 anos.

Temos a responsabilidade de produzir acessórios com significado e beleza, unindo o objetivo pessoal e adorno.





CARTA DE APRESENTAÇÃO

Vimos por esta apresentar o grupo de acadêmicos da Universidade Estácio de Sá – UNESA listado na tabela a final deste documento, a fim de convidá-lo a participar de uma atividade extensionista associada à disciplina TÓPICOS DE BIG DATA EM PYTHON sob responsabilidade do Prof. Lucas Antunes Floriano.

Em consonância ao Plano Nacional de Educação e demais normativas educacionais vigentes, a Universidade Estácio de Sá – UNESA desenvolve atividade extensionista que, norteados pela metodologia de aprendizagem baseada em projetos, tem por princípios fundantes o diagnóstico dos problemas/demandas/necessidades, a participação ativa dos interessados/públicos participantes, a construção dialógica, coletiva e experencial de conhecimentos, o planejamento de ações, o desenvolvimento e avaliação das ações, a sistematização dos conhecimentos, a avaliação das ações desenvolvidas.

Nesse contexto, a disciplina acima mencionada tem como principal escopo os temas relacionados à BIG DATA, com foco em aplicações em python.

Sendo assim, pedimos o apoio dessa organização/entidade/coletivo/associação/outro, que aqui chamaremos de parte interessada, para a realização das seguintes atividades: diagnósticos, análises, entrevistas, levantamentos, projetos ou qualquer outra metodologia de estudo de caso que auxilie no desenvolvimento das competências de nossos acadêmicos e ao mesmo tempo possa contribuir para a comunidade em que estamos inseridos.

Como se trata de atividade de ensino/aprendizagem de caráter extensionista, prevista no Projeto Pedagógico do Curso, salientamos que:

- não há cobrança de remuneração de qualquer natureza por parte da Universidade Estácio de Sá, seus alunos ou o docente da disciplina, à parte interessada;
- as atividades desenvolvidas no âmbito do projeto extensionista não configuram relação de trabalho entre os alunos e o docente da Universidade Estácio de Sá – UNESA disciplina TÓPICOS DE BIG DATA EM PYTHON , e a parte interessada;
- os resultados do projeto só poderão ser implantados para uso efetivo mediante Anotação de Responsabilidade Técnica de um profissional habilitado;
- os resultados do projeto podem ser implantados pela parte interessada para fins lucrativos, sem a necessidade de pagamento de quaisquer benefícios aos alunos, ao docente da disciplina e à Universidade Estácio de Sá – UNESA;
- quaisquer custos relativos à implantação e operação contínua do projeto fora do escopo das atividades do presente projeto serão arcados pela parte interessada.

- 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 -----

Grupo de Alunos
André Luiz Gomes de Araujo Matrícula: 202203753851
Gabriel de Paulo Mulinari Matrícula: 202203961004
Luiz Felipe Dias Cardoso Feres Lima Matrícula: 202002456086
Pedro Henrique Bastos de Oliveira Matrícula: 202102148201
Renan Soares Candido Matrícula: 202208622577

Atenciosamente,

Lucas Antunes Floriano
Lucas Antunes Floriano
Docente da disciplina: TÓPICOS DE BIG DATA EM PYTHON
Semestre: 2023.2
Matrícula: 10643417

Rio de Janeiro 23 de outubro de 2023.

Xenit Gomes Bastos
<<parte interessada>>

1.2 - Problemas

Nosso atual problema identificado é a concorrência desleal, copiando a metodologia para replicar acessórios. A empresa define os problemas como:

“Trabalhamos com matéria-prima da natureza, seguimos contornos de forma orgânica das Pedras Naturais, somos dependentes do desempenho do Lapidário, que muitas vezes não entende nosso design, isso dificulta a movimentação do site, vendas e produção em geral.

Estamos buscando maior harmonia nessa parceria para acelerar e aprimorar a conclusão dos processos internos, para estabelecer maiores resultados nas vendas.

Não seguimos tendências da moda, nossas peças são atemporais e conceituais, desse modo não podemos vender em qualquer lugar, e mais, nem sempre as vendedoras têm o discurso correto para transmitir nosso propósito. Gerando um impasse, uma conotação de problema a ser combatido com muito treinamento e dedicação.”

1.3 - Justificativa

A justificativa para a realização da solução de acordo com a empresa:

“Para combater a concorrência desenvolvemos um roteiro de vendas singular característico da marca, baseado em estudos, pesquisas e identificação do ponto alvo das nossas clientes.

Estamos promovendo um treinamento de vendas para envolver nossas consumidoras, elevando sua autoestima, alegria e motivação pessoal.

Que a mesma possa captar nossa mensagem vibracional e transmitir para outras pessoas, mesmo com passar dos tempos, o meio de vendas boca a boca ainda faz diferença.

Para aprimorar o desenvolvimento da produção orgânica com pedras naturais, buscamos um segundo parceiro ourives para confeccionar as peças com perfil e melhor custo, sempre respeitando a origem da marca.

Entendemos que precisamos produzir mais, mas nossas peças são exclusivas, arrojadas e charmosas, não podemos perder essa linha de conceito.

Por isso, visamos crescer com consciência da essência da marca, exclusividade é nosso nicho.”

1.4 - Objetivo

- * Fortalecer as vendas no atacado, já trabalhamos nesse ramo, mas reconhecemos que precisamos explorar mais novos clientes em potencial.
- * Aumentar o quadro de vendedoras com devido treinamento para representar a marca e concluir a venda varejo.
- * Aumentar o número de pontos de vendas e principalmente estabelecer um ponto fixo em um shopping conceituado.

Nosso objetivo é identificar oportunidade de mercado para criar e desenvolver um projeto expansivo, buscar oportunidades e transformá-las em um negócio viável e lucrativo, por meio de criação, organização e gestão de empreendimento.

Nosso planejamento consta criação de empregos, inovação, crescimento econômico, flexibilidade e dissipar problemas.

1.5 - Referencial Teórico

De acordo com Davenport (2014)[1], Big Data é um termo genérico para dados que não podem ser contidos nos repositórios usuais; refere-se a dados volumosos demais para caber em um único servidor; não estruturados demais para se adequar a um banco de dados organizado em linhas e colunas; ou fluidos demais para serem armazenados em uma data warehouse estático.

É necessário um estudo para a empresa ter as informações necessárias sobre o sistema antes de implantá-lo, evitando ao máximo perdas e resultados negativos, conforme é dito por Braian de Souza Bulian(2017)[2].

Utilizaremos a obra "Aprenda mineração de dados: teoria e prática" escrito por Fernando Amaral[3] e o artigo “A Importância do Big Data para a compreensão dos hábitos de compra dos consumidores, aumentando as vantagens competitivas no comércio varejista” desenvolvido por Braian de Souza Bulian[2] para compreender o conceito de Big data e

suas particularidades, além de entender na teoria e prática como funciona a manipulação de dados e como extrair informações que agreguem valor ao negócio.

Através do Livro "Big Data no Trabalho", de autoria de Davenport[1], no qual explica sobre Big data, suas tecnologias e formas de implementação em uma empresa, utilizaremos a obra como base para obter melhoria nas estratégias de negócio da Six Zen Acessórios, bem como as etapas, objetivos e tecnologias necessárias para atingir o propósito de fortalecer o número de vendas no atacado, aumentar os pontos de vendas e quadro de vendedoras.

2) Planejamento e Desenvolvimento

2.1 - Plano de Trabalho

1. Definição dos Requisitos

Descrição: Melhorar a visualização do histórico de vendas e metas anuais.

Responsável: [Pedro Henrique]

Prazo de Entrega: [02/10/2023]

2. Coleta de Dados

Descrição: Obter os comprovantes de pagamento de fontes de dados diversas.

Responsável: [Pedro Henrique]

Prazo de Entrega: [09/10/2023]

3. Pré-processamento de Dados

Descrição: Limpar e preparar os dados para análise utilizando o regex.

Responsável: [Luiz Lima, Gabriel Mulinari]

Prazo de Entrega: [16/10/2023]

4. Análise de Dados

Descrição: Realizar análises estatísticas.

Responsável: [Luiz Lima, Gabriel Mulinari]

Prazo de Entrega: [23/10/2023]

5. Visualização de Dados

Descrição: Criar visualizações e relatórios com os resultados.

Responsável: [Luiz Lima, André Luiz]

Prazo de Entrega: [1/11/2023]

6. Implantação

Descrição: Implementar o projeto na empresa para ter a melhor visualização no ambiente de produção.

Responsável: [Todos]

Prazo de Entrega: [03/11/2023]

7. Avaliação e Acompanhamento

Descrição: Avaliar o desempenho do projeto e acompanhar os resultados ao longo do tempo.

Responsável: [Todos]

Prazo de Entrega: [10/11/2023]

Recursos Necessários

- Equipe de Desenvolvimento
- Conjunto de Dados de Comprovantes de Pagamento
- Software de Big Data
- Ferramentas de Análise de Dados em Python
- Plataformas de Visualização de Dados

Formas de Acompanhamento dos Resultados

- Reuniões semanais de acompanhamento do projeto

Cronograma

- Definição dos Requisitos: [28/09/2023] - [02/10/2023]
- Coleta de Dados: [03/10/2023] - [09/10/2023]
- Pré-processamento de Dados: [10/10/2023] - [16/10/2023]
- Análise de Dados: [17/10/2023] - [23/10/2023]
- Visualização de Dados: [24/11/2023] - [01/11/2023]
- Implantação: [02/11/2023] - [03/11/2023]
- Avaliação e Acompanhamento: [04/11/2023] - [10/11/2023]

Conclusão do Projeto

Descrição: Apresentar os resultados e lições aprendidas.

Responsável: [Todos]

Prazo de Entrega: [10/11/2023]

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.

Realizamos as tarefas de criação das ideias, aplicação do conteúdo passado em sala de aula, juntamente com a elaboração do código, virtualmente via grupo de WhatsApp e Google Colab.

As entregas e comunicação direta com a parte interessada ficaram por parte de um dos membros do grupo, do qual tem proximidade e facilidade de contato com a empresa Six Zen Acessórios, que manteve o feedback de como proceder para alcançar seus objetivos de

alavancar vendas e aumentar vendedores. Seguem fotos do membro do grupo no estabelecimento Six Zen Acessórios junto com a proprietária:



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

Pedro Henrique:

Luiz Felipe Dias Cardoso Feres Lima(202002456096): Tratamento dos dados brutos utilizando o regexpr (<https://regexpr.com/>), e passar para uma planilha para poder ser utilizada no google colab ou em qualquer outra IDE de visualização de dados. Esse código foi usado

no regex para poder pegar os horários: `\d{2}:\d{2}` e esse foi usado para poder pegar os valores: `R\$(\d{1,3}(?:\.\d{3})*(?:\,\d{2})?)|\d+(?:\,\d{2})?)`.

Gabriel de Paulo Mulinari(202203961004): Criação do código, através do *Google Colab*, para obtenção dos dados através de um arquivo e sua adequação usando *Pandas*, fazendo manipulação através de *DataFrames*.

André Luiz Gomes de Araujo(202203753851): Gerar a visualização de dados em um gráfico, apresentar o mesmo para a parte interessada e realizar o acompanhamento do projeto.

Renan:

2.4 - Metas, critérios ou indicadores de avaliação do projeto

Definimos as seguintes etapas: Definição dos requisitos, coleta de dados, pré-processamento de dados, análise de dados, visualização dos dados, implantação e avaliação de desempenho.

Os critérios definidos, juntamente com a parte interessada, foram de realizar as entregas e executar testes e demonstrações do projeto conforme as datas estabelecidas e receber o feedback da mesma para prosseguir ou modificar o que foi realizado.

O indicador de avaliação será encontrar padrões através dos gráficos (Semelhança ou diferença evidente nos horários, datas, valores) nas vendas da empresa que consigam mostrar pontos a serem melhorados e, consequentemente, atinja os objetivos acordados com a parte interessada.

2.5 - Recursos previstos

Recursos Necessários

- Equipe de Desenvolvimento
- Conjunto de Dados de Comprovantes de Pagamento
- Software de Big Data
- Ferramentas de Análise de Dados em Python
- Plataformas de Visualização de Dados

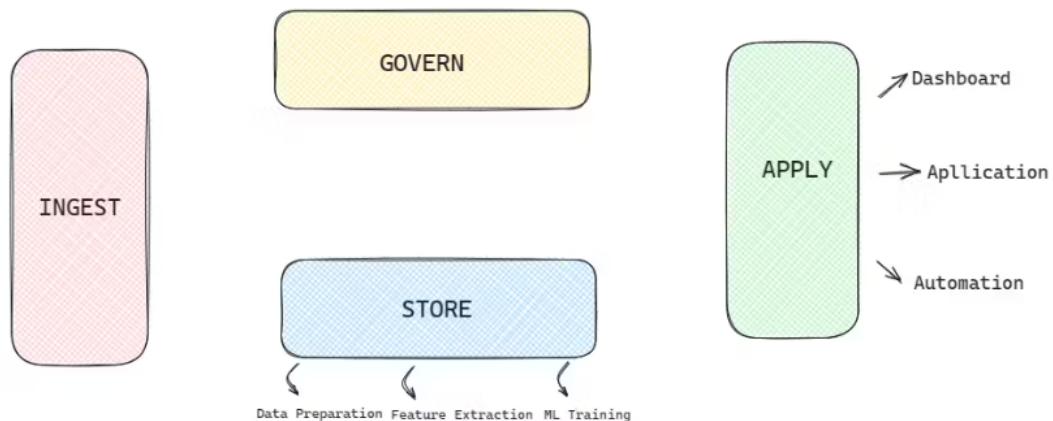
Planejamos utilizar apenas ferramentas gratuitas, e, para isso, estamos utilizando Regex (através do site [regex101.com](https://www.regex101.com)) para extração de dados, Google Colab para programação paralela de forma simples e intuitiva, biblioteca Pandas através do Python para realizar a manipulação de dados, Microsoft Excel e Power BI para visualização dos dados em tabelas e gráficos.

Os dados recebidos foram: Data, hora e valor de compras na loja Six Zen Acessórios. Esses dados foram enviados via texto por conversa do Whatsapp(txt) com a parte interessada, recebemos 88 linhas de dados, tendo como colunas essenciais a data, horário e valor da compra.

2.6 - Detalhamento Técnico do Projeto

Criamos uma solução em Big Data que realize análises estatísticas das vendas realizadas com os parâmetros estabelecidos e gere um gráfico para alcançar o objetivo acordado com a parte interessada.

Utilizamos Arquitetura de Data Lake para facilitar e deixar mais organizado o uso dos dados.



Data Lakes são utilizados para armazenar qualquer tipo de dados, podendo ser estruturados, não estruturados ou semi-estruturados. O Data Lake apresenta 4 níveis: Ingest, Store, Apply e Govern.

Ingest: Nesse nível, os dados são recebidos, podendo chegar em: png, json, xls, txt, etc. Posteriormente todos os dados são armazenados em um único lugar. No nosso caso, recebemos os dados via txt por Whatsapp.

Store: Aqui ocorre o tratamento de dados podendo ser utilizado o Regex(que foi utilizado neste trabalho)

Apply: Nesse nível, é realizado a visualização dos dados via dashboard ou aplicação. Utilizamos o Power Bi.

Govern: Aqui há o tratamento sobre metadados, com a origem dos dados e a relação entre os datasets.

3) Encerramento do Projeto

3.1) Relatório Coletivo

Através do nosso software de big data, conseguimos observar padrões importantes a serem melhorados para alavancar as vendas da Six Zen Acessórios. Após utilizar os softwares e métodos anteriormente citados, conseguimos compreender a importância do big data para uma empresa, como ocorre o uso de tal tecnologia na prática e também como resolver o problema de uma empresa através de tecnologia, além de como trabalhar melhor em equipe dividindo as tarefas a fim de chegar ao objetivo acordado.

3.1.1) Avaliação de reação da parte interessada

“A equipe Six Zen Acessórios considerou o trabalho muito interessante, efetivo e estimulante.

Nossa participação foi de extrema importância para reavivar nossos objetivos e classificar novos ideais.

Estamos sempre em busca de melhores resultados e amplitude da marca.

O empreendedorismo no Brasil requer resiliência constante, por isso nosso depoimento verdadeiro, além de colaborar com excelente trabalho dos alunos, nos faz acreditar que estamos no caminho certo para novas conquistas.”

3.2) Relato de Experiência Individual

Pedro Henrique: Fui responsável por fornecer informações da Six Zen Acessórios para efetuar o trabalho, transmitindo as informações necessárias da logística, planejamento e objetivos considerados extremamente importante para marca citada. Por texto passei todas as datas, horários, custo e valor de venda e todo funcionamento da empresa, não tive problemas em passar as informações necessárias, visto que, a parte interessada é familiar.

Luiz Felipe Dias Cardoso Feres Lima: Aprendi como que funciona visualização e tratamento de dados em uma empresa. Além disso, também foi de extrema importância colocar em prática o funcionamento do tratamento bruto de dados utilizando o regexpr e a visualização no powerBI.

Gabriel de Paulo Mulinari:

Neste trabalho fiquei responsável pela manipulação dos dados. Para isso utilizei a biblioteca Pandas, bastante utilizada para manipulação e análise de dados, além do uso do Google Colab para fazer os códigos.

Os dados que manipulei vieram de um arquivo CSV, que continha as colunas, data, hora e valor. A primeira coisa que fiz foi usar a função “`read_csv`” do Pandas para ler o arquivo e gerar um `DataFrame` para enfim fazer a manipulação dos dados. Após isso retirei o ponto dos valores de venda que chegavam a casa dos milhares, usando o método “`replace`” nativo do Python, pois estava dando mais trabalho de usar esses dados na visualização com o PowerBI, dessa forma ficou mais fácil de usá-los. Posteriormente fiz uma manipulação das datas para que elas ficassem no formato de Dia/Mês/Ano, porém acabou que não seria mais necessária essa formatação, porém deixei para ter certeza de que as datas ficariam no formato correto. A coluna hora não precisou de nenhum tipo de manipulação, após analisar

vi que estava tudo correto. Por último, após fazer toda a manipulação do *DataFrame*, escrevi os dados para um arquivo chamado “*dados_de_vendas.csv*”, utilizando a função “*to_csv*” do Pandas. Com isso bastaria o responsável pela visualização dos dados executar o código e ele obteria o arquivo com todos os dados manipulados para serem exibidos.

Aprendi, como analisar e manipular os dados e deixá-los da maneira que fosse necessário. A maior dificuldade foi saber quais eram as funções para fazer o que eu estava pensando, para isso fui pesquisar através de vídeos e do material de aula, acabei descobrindo as funções que faziam o que eu precisava.

Link do Colab com o código:

https://colab.research.google.com/drive/1vO7aB4CSmqEnNHMUy9tvU8kaTZuaFFcV?authuser=1#scrollTo=o3GsOIEIE_VR

```
[16] df_vendas = pd.read_csv("/content/sample_data/vendas.csv", delimiter=";")

    df_vendas["Valor"] = df_vendas["Valor"].apply(lambda x: x.replace(".", ""))
    #df_vendas["Valor"] = df_vendas["Valor"].apply(lambda x: x.replace(",", "."))
    #é melhor não substituir a vírgula por ponto para passar para o csv e modelar os

    #Passa os valores na coluna "Data" para o tipo "datetime" e formata a data com "d
    df_vendas["Data"] = pd.to_datetime(df_vendas["Data"], dayfirst=True)
    df_vendas["Data"] = df_vendas["Data"].dt.strftime("%d/%m/%Y") #CONVERTE A COLUNA

    #Passa os valores na coluna "Valor" para o tipo float
    #df_vendas["Valor"] = df_vendas["Valor"].astype(float)

    print(df_vendas)

    #Arquivo gerado para ser usado no PowerBi
    df_vendas.to_csv("dados_de_vendas.csv", index=False)
```

https://liveestacio-my.sharepoint.com/:x/g/personal/202203753851_alunos_estacio_br/EQI_CxHGV6tEkf2iY7INpCYBIZd8ngp24ryFw0fUWe7HVA?e=fGjIEL

https://liveestacio-my.sharepoint.com/:t/g/personal/202203753851_alunos_estacio_br/EfG0TZbg23xGmNj-PLfk1bMBi3jFSOGfqK-iopTIZI2muQ?e=52Jbg3

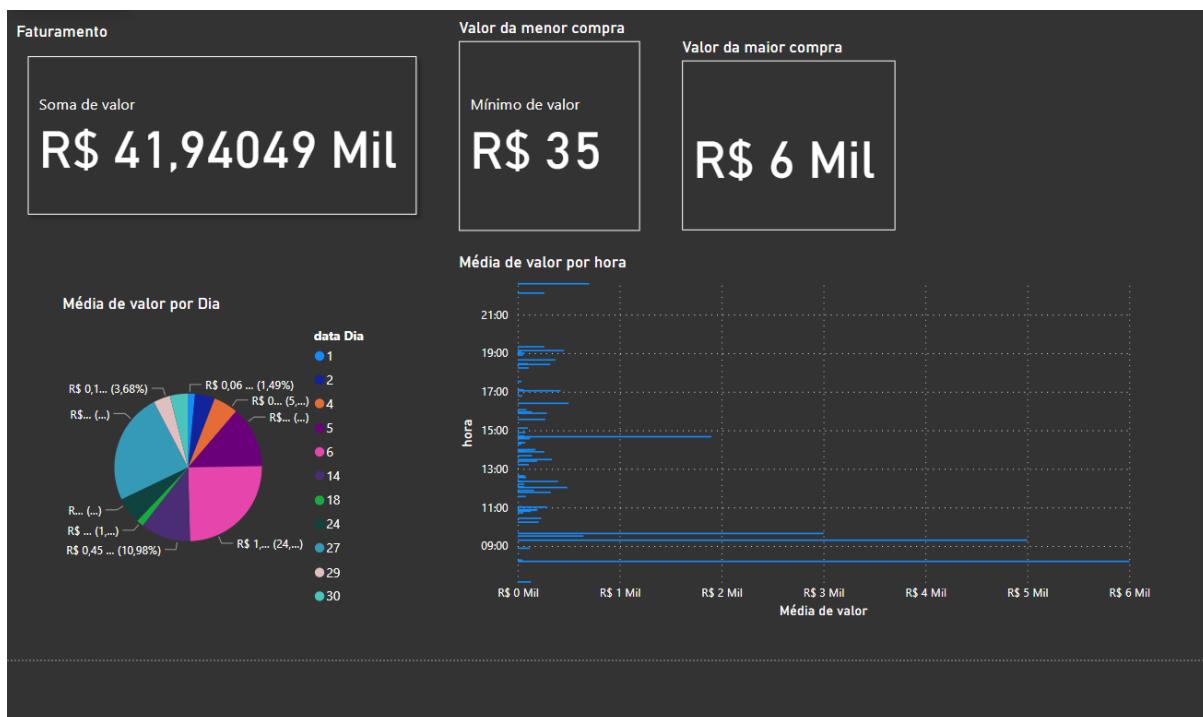
Renan: Com a conclusão do trabalho, percebi a importância do Big Data e seu impacto na visualização e tratamento de dados. Fornecendo diversas facilidades para empresas resolverem problemas e dificuldades, ele torna-se algo imprescindível nesse meio. Também me surpreendi com o Power Bi, que possibilita a criação de gráficos dinâmicos e interativos, permitindo que mesmo uma pessoa sem conhecimento técnico consiga extrair informações devido a clareza das informações disponibilizadas.

André Luiz Gomes de Araujo: Fiquei responsável pela apresentação visual do projeto, para isso, decidi usar o Power Bi para criação de gráficos e o Microsoft Excel para a transformação dos dados. Inicialmente, escolhi o Power Bi por ter maior facilidade em usar a ferramenta por ter um visual mais atrativo de ser apresentado para quem não está tão bem apresentado na área de tecnologia.

Por mais que eu tivesse um pouco mais de familiaridade com o Power Bi, consegui em pouco tempo aprender muito mais sobre o que o software tem a oferecer do que das últimas vezes que usei, principalmente quanto a inserção de dados.

Quando fui trazer os dados do Excel para o Power Bi, tive um erro em que todas as datas ficaram como: 31/12/1899, mesmo com as datas corretas no excel, por algum motivo ficou esse erro. A partir desse fato, tive que descobrir uma forma de trazer esses dados limpos e, através da aba “Colar dados em uma tabela em branco” consegui contornar a situação.

Por fim, gerei alguns cartões e gráficos que conseguiram traduzir muito bem como a parte interessada poderia atingir seu objetivo visto que:



Arquivo Power Bi:[trabalhobigdata.pbix](#)

Arquivo Power Bi com visualização para celular: [trabalhobigdatavisualizaocel.pbix](#)

Arquivo PDF: [trabalhobigdata.pdf](#)

- Com os primeiros três cartões, demonstro o valor total de vendas do período e a disparidade nos valores de compra.
- Identifiquei que há um bom volume de vendas nos primeiros e últimos dias do mês, enquanto no meio do mês existe um baixo volume de vendas, podendo ser realizado nesse período um marketing maior para captar mais clientes e manter maior constância no número de vendas mensal e consequentemente um aumento no número de vendas.
- As compras de maior valor são realizadas principalmente às 9:00, enquanto nos outros horários temos valores mais variados, com destaque para uma venda de alto valor às 15:00. A partir desses números, conseguimos entregar para a parte interessada o perfil de quanto gasta os clientes por horário, e demonstrando os períodos de maior atenção por parte dos vendedores para entregarem vendas mais caras.

Por fim, saio surpreendido com o quanto Big Data pode impactar nas vendas de uma empresa e feliz com o maior conhecimento que obtive sobre Power Bi, construção de software Big Data e maior experiência em trabalho em grupo sobre como conciliar tarefas, lidar com diferentes pessoas e dialogar com uma empresa sobre como resolver seus problemas. Como perspectiva para trabalhos futuros, saio com uma noção maior de como proceder com maior facilidade nas etapas e como extrair ainda mais o potencial do big data.

Bibliografia:

- 1-DAVENPORT, Thomas H. **Big data no trabalho.** Elsevier Brasil, 2014.
- 2-BULIAN, Braian de Souza; ALENCAR, Cícero Aparecido. A importância da big data para a compreensão dos hábitos de compra dos consumidores, aumentando as vantagens competitivas no comércio varejista. 2021.
- 3-AMARAL, Fernando. **Aprenda mineração de dados: teoria e prática.** Alta Books Editora, 2016.