**Governança de Dados “Projeto\_portfolio\_RosyDS.**

* Qual problema resolver para que a empresa tenha a solução na tomada de decisão?
  1. Negócio: “Financeiro”
  2. Objetivo: Entendendo cenário atual.

1. **1. Geração de caixa – valor liquido gerado em um determinado período do exercício. (usar demonstrativo de fluxo de caixa-DFC).**
2. **2. Rentabilidade – o valor de retorno do investimento (lucro Líquido/investimento \* 100 => %rentabilidade.**
3. **3. Lucratividade – lucro da empresa. Lucro líquido / receita bruta \* 100 => %de lucratividade.**
4. **4. Faturamento – soma de todas as receitas, ajudam a analisar o desempenho da empresa. Todas as receitas = faturamento bruto do período.**
   1. **Por canais de distribuição**
   2. **Por produto**
5. **5. Ticket Médio – Valor médio das vendas, melhora as estratégias de vendas, quanto maior o valor de seus produtos ou serviços ou quanto mais seus clientes comprarem, maior o ticket médio. Faturamento Bruto/Total de vendas = Ticket médio**
6. **6. Margem bruta - demonstra a porcentagem de rentabilidade nos produtos e serviços comercializados. Essa análise ajuda a identificar quais preços de vendas estão bons e quais precisam melhorar. Lucro bruto / receita total \* 100 => % de margem bruta.**
7. **7. Margem Líquida - diminuir todas as despesas envolvidas na produção da sua receita, para cada real gasto, quanto de lucro você tem? Lucro líquido após impostos / receita total \* 100 => Margem líquida**
8. **8. Margem de contribuição - representa o lucro de cada venda, não o rendimento. Valor das vendas – (custos variáveis + despesas variáveis) => margem de contribuição.**
9. **9. Ponto de Equilíbrio – Se a empresa possui receita o suficiente para pagar todas suas despesas e não recorrer a empréstimos ou ao dinheiro em caixa. Despesas / % de M.de contribuição => Ponto de equilíbrio. A partir do momento que sua empresa alcançar esse ponto, toda receita a mais que entrar será lucro.**
10. **EBITDA – demonstrativo de desempenho e produtividade da sua empresa desconsiderando influências externas. Lucro operacional antes do imposto de renda e receita / (despesas financeiras + depreciação + amortização) => EBITDA. Para uma empresa é fundamental saber qual o seu destino final para poder traçar qual caminho será percorrido, isso se chama objetivo estratégico**.

No arquivo em uma planilha tinha 3 tabelas que foi resumida em uma somente, em uma planilha. Seleciona a coluna que deseja pôr os dados-vai em Dados-obter dados-escolher o tipo de arquivo que deseja-escolhe a planilha.

**Criar Pipeline da Ciência de Dados**

**Planejamento**

* + Definir os objetivos (Junto com o setor que deseja as análises);
  + Organizar os recursos (Ferramentas, linguagens de programação e fontes);
  + Coordenar os colaboradores;
  + Fazer a programação do projeto (Scrum, Kanban, Canvas e outros).
  + Relatório com as visualizações e história.

**Preparação dos dados - ETL**

* Conseguir os dados:
  + (Excel, csv, txt, Web(scraping), BD (DW, Nosql e Hadoop), Cloud BD e DLake (IBM, AWS, Azure, Google, Oracle, Heroku, Snowflake), API, Arquivo Json, Pastas, outros;
* Limpar os dados:
  + Verificar Encoding
  + Erros
  + Anomalias
  + Escolher as variáveis ou colunas que vai precisar.
  + Transformar Colunas em linhas
  + Cabeçalho de colunas
  + Remover subtotais
  + Remover colunas, linhas em branco ou NaN
  + espaço em branco e linhas sobrando
  + Adicionar, renomear, tipo de colunas,
* Explorar os dados:
  + Estatística descritiva
    - Centralizada - média e mediana.
    - Dispersão - desvio padrão.
  + Verificar distribuição (simétrica ou assimétrica)
    - Outliers.
    - Verificar correlações entre variáveis
* Refinar os dados:
  + Escolhe o que vai incluir
    - Aqui entra definição de Variáveis, e a criação dos novos recursos, que vai gerar conteúdos no qual deseja trabalhar)

**Análise dos dados(modelagem)**

* Criar o modelo de schema ou snowflake;
  + Comparar médias e insights.
  + Teste T
  + ANOVA
* Validar o modelo;
* Avaliar o modelo;
  + Relações das variáveis
    - Quantitativas – ‘correlação’ onde as variáveis aumentam ou diminuem juntos, se ela não tem nenhuma relação entre si. ‘regressão’ traz o mesmo da correlação, mas ela eficaz e rica, ela traz a influência da variável na outra quantitativamente.
    - Qualitativas – Qui-quadrado.
* Refinar o modelo e apresentar;

**Relatório ou acompanhamento do projeto**

* Interpretar os resultados
  + Gerar insights
* Introdução – resumo do projeto, incluindo o problema, os tipos de dados usados e os insights da solução.
* Dados – detalhar os dados, variáveis e termos técnicos, incluindo observações faltantes ou anormalizados (outliers).
* Ánalise – o processo construtivo por trás da ánalise, visualização de dados, métodos, resultados interessantes ou significativos.
* Resultado e conclusão – resume os resultados e recomendação p/ tomada de decisão.
* Verificar se é necessário colocar no servidor (deploy), receber dados de novos clientes e fazer previsões, isso se estamos fazendo um modelo preditivo.

Observações:

Pode apresentar no Power Point

Usar storytelling

Colocar no GitHub

Inserir QRcode.