**Tópicos em Big Data com Python**

Faculdade Estácio do Recife Professor: Thiago Mendes

Alunos: Maria Júlia Alves da Rocha e Iverson Luna

**Visão Geral do Projeto:**

* **Objetivo:** Projeto da disciplina de Tópicos de Big Data com Python da faculdade Estácio, o objetivo do projeto é realizar o tratamento e análise de dados no Power BI, utilizar bibliotecas Python para exibir estatísticas, dataframes, importar arquivos, unir e separar tabelas com o merge do tipo Left Join.
* **Escopo:** Utilizei uma planilha de vendas no Power BI e através do Power Query exclui a ultima coluna que ficou “null” após fazer essa importação, removi linhas em branco, uni duas colunas (nome e sobrenome), dividi a coluna “Loja” em duas, transformei em “Estado” e “Cidade”, também transformei a coluna “Gênero” que estava abreviada com “M” e “F” para “Masculino” e “Feminino”. Após o tratamento da tabela, demonstrei os dados por meio dos filtros de visual gráfico de colunas, gráfico de pizza, cartão, mapa e gráfico de barras.

A próxima parte do projeto foi feita em Python, através de IDE VScode, a princípio importei as bibliotecas “Random” para gerar números aleatórios e “Pandas” para a manipulação de tabelas, criei dois dataframe: “Gols” e “Faltas”, uni ambos em um terceiro dataframe, também exibi a maior quantidade de gols, a menor e a média de gols, exibindo assim as estatísticas, também realizei o mesmo processo com a tabela de “Faltas”.

Nesta parte do projeto, criei uma planilha sobre alguns jogadores, seus clubes, gols, faltas e suas respectivas idades, também utilizando Python e a biblioteca Pandas, importei essa planilha em Excel para o Vscode, exibi um dataframe para a planilha completa, após isso, utilizei a função “iloc” para pegar colunas da planilha e armazenar em duas tabelas distindas, assim quebrando a tabela inicial. Por fim, utilizei o “Left Join” para unir as tabelas da esquerda para direita.

**Design, apresentação e template:**

* Utilizei Power Point para o template do Power BI
* Ícones retirados do site “FlatCoin”
* Utilizei Canva para o slide

**Configuração do Ambiente:**

* **Requisitos de Hardware/Software:** IDE VScode, Power BI, Excel
* Bibliotecas: Pandas e Random

**Problemas Conhecidos e Soluções:**

* A princípio, não consegui dar import em Pandas pela IDE, o problema foi resolvido pelo próprio terminal do windows.
* Outro problema encontrado foi ao importar arquivos para o python, poblema resolvido através da instalação de uma extenção para suportar alguns formatos de arquivos.