

PROJETO FINAL

DESCRIÇÃO GERAL DO PROJETO

- Todas as equipes deverão entregar as mesmas especificações, de acordo com o seu respectivo tema.
- Vocês deverão aplicar os conceitos vistos durante o curso para tratar, organizar e modelar os dados de no mínimo 2 datasets escolhidos por vocês seguindo o tema de sua equipe.
- Obrigatoriamente deverá conter as tecnologias Google Cloud Platform (Cloud Storage), Python, Pandas, SQL, PySpark, Looker Studio, PowerBI, Big Query e MongoDB.

APRESENTAÇÃO

A apresentação do trabalho se dará da seguinte maneira: Cada grupo deverá ser totalmente responsável pela forma pela qual vai interpretar o dataset, apresentando suposições e conclusões dos dados. Todas essas situações devem ser explicadas. Deverá iniciar pela apresentação do dataset, informando de qual local foi obtido o dataset e quais as principais informações sobre o mesmo.

- Deverá apresentar as funções e ferramentas utilizadas no código.
- Explicar o porquê do dataset escolhido.
- Todos os componentes (integrantes) deverão se apresentar.
- Deverá ser usado termos técnicos, evitando o uso de gírias ou expressões coloquiais e/ou culturais.
- Cada grupo terá até 60 minutos para se apresentar.
- A ordem da apresentação será comunicada pelos professores próximo à data de apresentação.

PRINCIPAIS HABILIDADES A SEREM AVALIADAS

- Oralidade e comunicação em público.
- Análise SWOT: ferramenta de planejamento estratégico que ajuda a avaliar as forças, fraquezas, oportunidades e ameaças de uma empresa ou projeto.
- Storytelling na apresentação dos dados: a apresentação dos dados pode ser feita de forma mais eficaz e envolvente através da técnica de storytelling. Ao contar uma história, os dados são apresentados de maneira mais clara e compreensível, e a audiência é mais facilmente motivada e engajada.
- Principais métricas e KPI's observadas: definição completa de quais métricas e/ou KPI's serão aplicadas ao projeto, suas razões e definições estratégicas.
- Estabeleça algumas sugestões de ações: construa algumas sugestões de ações frente às informações coletadas no projeto. Ações destinadas a campanhas, objetivos e segmentações, por exemplo.
- Capacidade de argumentação
- Habilidade de codificação em Python e utilização de suas bibliotecas
- Habilidade de interpretação e análise de dados.
- Capacidade de implementação de códigos utilizando as bibliotecas Pandas e PySpark.
- Capacidade de implementação de consultas utilizando a linguagem SQL.
- Capacidade Analítica e Interpretativa.
- Capacidade de organização e grupo

REQUISITOS OBRIGATÓRIOS

- Obrigatoriamente os datasets devem ter formatos diferentes (CSV / Json / Sql / NoSql/ Excel, outros) e 1 deles obrigatoriamente tem que ser em CSV.
- Operações com Pandas (limpezas , transformações e normalizações)
- Operações de análise de dados usando PySpark com a justificativa de cada uma das transformações utilizadas com o mesmo.
- Utilizar pelo menos 2 gráficos distintos (bibliotecas python) para representar os dados e encontrar possíveis inconsistências
- Os datasets utilizados podem ser em língua estrangeira , mas devem ao final terem seus dados/colunas exibidos na língua PT-BR (Utilizar a diretrizes abordadas em aula)
- Os datasets devem ser salvos e operados em armazenamento cloud obrigatoriamente dentro da plataforma GCP (não pode ser usado Google drive ou armazenamento alheio ao google)
- Os dados tratados devem ser armazenados obrigatoriamente em um datalake(Bucket) , DW(BigQuery) ou em ambos.
- Os Dataframe(s) resultante(s) deve(m) estar em uma coleção do MongoDB atlas (informar a key de acesso ao cluster)
- Deve ser feita no mínimo 4 análises dentro do Big Query utilizando a linguagem padrão SQL com a descrição das consultas feitas.
- Deve ser criado no Looker Studio ou no PowerBI um dashboard para exibição gráfica dos dados tratados trazendo insights importantes
- E deve ser demonstrado em um workflow simples (gráfico) as etapas de ETL com suas respectivas ferramentas.
- Fazer a documentação do projeto;

TEMA:

- MERCADO IMOBILIÁRIO