**Assignment - Data Engineer.**

## **Diário de Bordo**

A amostra de dados em anexo (info\_transportes.csv), possui dados de um aplicativo de transporte privado, cujas colunas são:

* DATA\_INICIO (formato: "mm-dd-yyyy HH") o DATA\_FIM (formato: "mm-dd-yyyy HH")
* CATEGORIA o LOCAL\_INICIO o LOCAL\_FIM o PROPOSITO
* DISTANCIA

Uma equipe está elaborando um modelo para compreender como os clientes estão utilizando o aplicativo. Para isso, você precisa fornecer uma nova tabela “info\_corridas\_do\_dia", com dados agrupados pela data de início do transporte utilizando a formatação “yyyy-MM-dd”, contendo as seguintes colunas:

|  |  |
| --- | --- |
| **Nome da coluna** | **Descrição** |
| DT\_REFE | Data de referência. |
| QT\_CORR | Quantidade de corridas. |
| QT\_CORR\_NEG | Quantidade de corridas com a categoria “Negócio”. |
| QT\_CORR\_PESS | Quantidade de corridas com a categoria “Pessoal”. |
| VL\_MAX\_DIST | Maior distância percorrida por uma corrida. |
| VL\_MIN\_DIST | Menor distância percorrida por uma corrida. |
| VL\_AVG\_DIST | Média das distâncias percorridas. |
| QT\_CORR\_ REUNI | Quantidade de corridas com o propósito de "Reunião". |
| QT\_CORR\_NAO\_REUNI | Quantidade de corridas com o propósito declarado e diferente de "Reunião". |

Abaixo temos um exemplo de uma possível linha da tabela:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DT\_REFE** | **QT\_CORR** | **QT\_CORR\_NEG** | **QT\_CORR\_PESS** | **VL\_MAX\_DIST** | **VL\_MIN\_DIST** | **VL\_AVG\_DIST** | **QT\_CORR \_REUNI** | **QT\_CORR\_NAO\_ REUNI** |
| 2022-01-01 | 20 | 12 | 8 | 2.2 | 0.7 | 1.1 | 6 | 10 |

## **Monitoramento de Sensores IoT**

**Descrição do Problema**

Você precisa criar um sistema de monitoramento de sensores IoT que envia dados de sensores em tempo real para um tópico Kafka (producer) e consome esses dados para processamento e armazenamento (consumer).

1. **Criar o Producer**
   * Desenvolver um script em Python (ou outra linguagem de sua escolha) que gera dados falsos de sensores IoT e envia esses dados para um tópico Kafka.
   * Utilizar uma biblioteca como faker para gerar dados falsos.
2. **Criar o Consumer**
   * Desenvolver um script que consome os dados do tópico Kafka e processa esses dados.
   * Armazenar os dados consumidos em um banco de dados..