



*Big Data Real-Time Analytics Com Python e Spark 3.0*

# Big Data Real-Time Analytics Com Python e Spark Versão 3.0

## Projeto com Feedback 3

### Análise de Risco no Transporte Público



## Formação Cientista de Dados 3.0

### Big Data Real-Time Analytics com Python e Spark

#### Projeto com Feedback 3

#### Análise de Risco no Transporte Público



Quais incidentes de trânsito ocorrem com mais frequência? Qual a faixa etária que mais se envolve em incidentes de trânsito? Qual o evento mais comum nos incidentes? Passageiros ou pedestres são as maiores vítimas dos incidentes?

Essas e outras perguntas devem ser respondidas através da análise de dados reais disponíveis publicamente. Este projeto não requer Machine Learning e seu trabalho é aplicar suas habilidades de análise e responder diversas perguntas de negócio através de gráficos e storytelling.

**Para a construção desse projeto, recomendamos a utilização da Linguagem Python e Linguagem SQL e o dataset disponível para download no link abaixo:**

<https://data.world/makeovermonday/2018w51>

O conjunto de dados lista incidentes de trânsito ocorridos na cidade de Londres.

Seu trabalho é analisar os dados e construir gráficos que respondam a essas 10 perguntas abaixo:

- 1- Qual a quantidade de incidentes por gênero?
- 2- Qual faixa etária esteve mais envolvida nos incidentes?
- 3- Qual o percentual de incidentes por tipo de evento (Incident Event Type)?
- 4- Como foi a evolução de incidentes por mês ao longo do tempo?
- 5- Quando o incidente foi “Collision Incident” em qual mês houve o maior número de incidentes envolvendo pessoas do sexo feminino?
- 6- Qual foi a média de incidentes por mês envolvendo crianças (Child)?
- 7- Considerando a descrição de incidente como “Injuries treated on scene” (coluna Injury Result Description), qual o total de incidentes de pessoas do sexo masculino e sexo feminino?
- 8- No ano de 2017 em qual mês houve mais incidentes com idosos (Elderly)?
- 9- Considerando o Operador qual a distribuição de incidentes ao longo do tempo?
- 10- Qual o tipo de incidente mais comum com ciclistas?

**Quando concluir o projeto, envie o(s) script(s) que você criar para o seguinte e-mail: [suporte@datascienceacademy.com.br](mailto:suporte@datascienceacademy.com.br).**

**Caso você tenha criado datasets auxiliares e esses sejam muito grandes, armazene em um diretório virtual (existem vários na internet, como Google Drive ou Dropbox) e envie o link para que nossa equipe possa baixar os datasets. Se os arquivos forem pequenos (uma amostra do dataset original), envie no anexo junto com o script. Documente seu script tanto quanto possível.**

**Caso prefira, disponibilize seu projeto no Github e envie o link do seu repositório para nossa equipe no e-mail [suporte@datascienceacademy.com.br](mailto:suporte@datascienceacademy.com.br). Nesse caso, o Readme do repositório deve constar que este trata-se de um**

**projeto da Formação Cientista de Dados da Data Science Academy, caso contrário não será avaliado.**

**Em até 72 horas, daremos o feedback respondendo seu e-mail. Caso não receba a resposta em até 72 horas entre em contato com a nossa equipe para verificar se recebemos seu projeto.**

**Bom trabalho!**