**Teste Engenheiro de Dados**

1) O [arquivo](https://drive.google.com/open?id=1XjYvVyE_q_6SnECaDTeB96MlhtfKYQeI) de entrada contém uma série histórica de preços de combustíveis do Brasil, do ano de 2019, as colunas são:

*DATA INICIAL,DATA FINAL,REGIÃO,ESTADO,MUNICÍPIO,PRODUTO,NÚMERO DE POSTOS PESQUISADOS,UNIDADE DE MEDIDA,PREÇO MÉDIO REVENDA,DESVIO PADRÃO REVENDA,PREÇO MÍNIMO REVENDA,PREÇO MÁXIMO REVENDA,MARGEM MÉDIA REVENDA,COEF DE VARIAÇÃO REVENDA,PREÇO MÉDIO DISTRIBUIÇÃO,DESVIO PADRÃO DISTRIBUIÇÃO,PREÇO MÍNIMO DISTRIBUIÇÃO,PREÇO MÁXIMO DISTRIBUIÇÃO,COEF DE VARIAÇÃO DISTRIBUIÇÃO.*

Usando o apache Spark (Pyspark ou SparkR) faça:

a) Estes valores estão distribuídos em dados semanais, agrupe eles por mês e calcule as médias de valores de cada combustível por cidade.

b) Calcule a média de valor do combustível por estado e região.

c) Calcule a variância e a variação absoluta do máximo, mínimo de cada cidade, mês a mês.

d) Quais são as 5 cidades que possuem a maior diferença entre o combustível mais barato e o mais caro.