O objetivo deste documento, é descrever os primeiros passos para executar o notebook principal do projeto.

Antes de executar o notebook com as análises dos dados do Enem, será necessário montar o ambiente, pois utilizamos o freamework Spark para melhorar a desempenho de carregamento dos microdados recuperados do site do INEP.

Procedimentos.

Para o rodar os Spark será necessário fazer o download do java e deverá estar na versão 8 ou superior.

- Após a instalação, configurar as variáveis de ambiente JAVA\_HOME e JRE\_HOME

Exemplo:

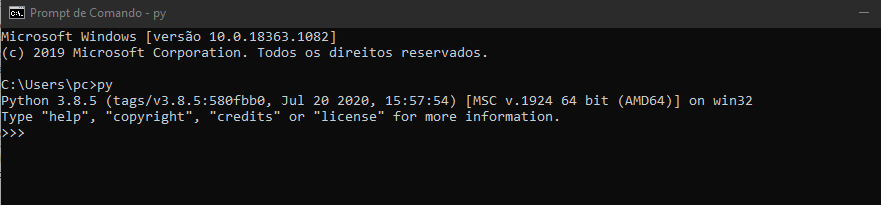
JAVA\_HOME= <local de instalação>\Java\jdk1.8.0\_251

JRE\_HOME=<local de instalação>\Java\jre1.8.0\_251

Fazer a instalação do Python, o software de instalação na versão 3.8.5 está disponível no link abaixo.

<https://www.python.org/downloads/>

Para confirmar a instalação rodar comando no prompt “PY” conforme o print abaixo:



- Fazer o download do freamework Spark na versão 3.0.1

<https://spark.apache.org/downloads.html>

Configurar as variáveis de ambiente SPARK\_HOME, HADOOP\_HOME e PYTHONPATH

SPARK\_HOME= <Local de Instalação>\spark-3.0.0-bin-hadoop2.7

HADOOP\_HOME=<Local de Instalação>\spark-3.0.0-bin-hadoop2.7

PYTHONPATH=%SPARK\_HOME%\python;%SPARK\_HOME%\python\lib\py4j-0.10.9-src.zip

Caso estja utilizando Windows incluir o arquivo winutils.exe na pasta SPARK\_HOME\bin

Rodar o comando “pyspark” no prompt e verificar se a instalação foi realizada com sucesso.

