|  |  |
| --- | --- |
| **Notebook** | **url** |
| Spark DataFrames.ipynb | <https://colab.research.google.com/drive/1-5pLlO9VqNJmQJpWcIs1MKXu8DUP6-e3?usp=sharing> |
| Spark Functions.ipynb | <https://colab.research.google.com/drive/1NYrkmmkX5gYNpp1WHyOEfmjaXPjclvNu?usp=sharing> |
| Tokenize Data.ipynb | <https://colab.research.google.com/drive/173_05PaICM4Pg1p3MtmXaEf_AwCGyxaT?usp=sharing> |
| Stop Words.ipynb | <https://colab.research.google.com/drive/1r8GKs31kdLVyWsRQf52ZBFDITel9NBpd?usp=sharing> |
| TF-IDF.ipynb | <https://colab.research.google.com/drive/1McJZzM5YO2fsrox6gfj61W47SShTL4A6?usp=sharing> |
| Pipeline.ipynb | <https://colab.research.google.com/drive/1bTm43-8Y5u4HDGslDhVlinmQpnGq8NXD?usp=sharing> |
| PySpark and S3 Stored Data.ipynb | <https://colab.research.google.com/drive/1X9IkB8OnSO5UsWTioP_yNCMLyJ-l4bTt?usp=sharing> |
| PySparkETL.ipynb | <https://colab.research.google.com/drive/1fILy1R3GhBSyZ516oblwCWBBPzVdeNuy?usp=sharing> |
| ChallengeETL.ipynb | <https://colab.research.google.com/drive/1CoG0mUGe9gw8o_OCDTH3os4cXfs7JuNw?usp=sharing> |
| Vine\_table\_anlaysis.ipynb | <https://colab.research.google.com/drive/1QAD7VLvs8F3Aky51bovap38OkSvI2R73?usp=sharing> |