**DATASET OVERVIEW**

We have considered 2 datasets to execute this project. Considering 2 datasets gave us more scope for data-pre-processing at a good level. Following are the two datasets:

|  |  |
| --- | --- |
| **Mattermark** | **Crunchbase** |
| * Info on Companies ,Investors and 5,000+ investment events * REST API protocol used to obtain datasets * Python request library used to obtain data from Mattermark * More data available only on paid subscription * **(**<https://mattermark.com/>**)** | * 12,000 companies, 11,000 investors and 52,000+ investment events info available * API access restricted since 2016. * Last available data in mass quantity - 2013 Snapshot, officially provided by the company * **(**<https://www.crunchbase.com/>**)** |