PYPI – Big Query Data Dictionary

[Link to Dataset](https://console.cloud.google.com/bigquery?p=bigquery-public-data&d=pypi&page=dataset&project=ds-5001-final-project)

# Table 1 – Distribution Metadata

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Field name** | **Type** | **Mode** | **Description** | **Potential Use** |
| metadata\_version | STRING | NULLABLE |  |  |
| name | STRING | REQUIRED | Name of package |  |
| version | STRING | REQUIRED | Version of package |  |
| summary | STRING | NULLABLE | Short description of package |  |
| description | STRING | NULLABLE | Long description of package |  |
| description\_content\_type | STRING | NULLABLE | Not useful |  |
| author | STRING | NULLABLE |  | Contributor network |
| author\_email | STRING | NULLABLE |  | Contributor network |
| maintainer | STRING | NULLABLE |  | Contributor network |
| maintainer\_email | STRING | NULLABLE |  | Contributor network |
| license | STRING | NULLABLE |  | Select OSI licenses |
| keywords | STRING | NULLABLE |  | Categories |
| classifiers | STRING | REPEATED |  | Categories |
| platform | STRING | REPEATED | Operating systems? |  |
| home\_page | STRING | NULLABLE | Location of code repository | Getting more contributor information |
| download\_url | STRING | NULLABLE |  |  |
| requires\_python | STRING | NULLABLE | Indicates which version of python required |  |
| requires | STRING | REPEATED | Dependencies | Dependency network |
| provides | STRING | REPEATED |  |  |
| obsoletes | STRING | REPEATED |  |  |
| requires\_dist | STRING | REPEATED | Dependencies? |  |
| provides\_dist | STRING | REPEATED |  |  |
| obsoletes\_dist | STRING | REPEATED |  |  |
| requires\_external | STRING | REPEATED |  |  |
| project\_urls | STRING | REPEATED |  |  |
| uploaded\_via | STRING | NULLABLE |  |  |
| upload\_time | TIMESTAMP | NULLABLE |  |  |
| filename | STRING | NULLABLE |  |  |
| size | INTEGER | NULLABLE |  |  |
| path | STRING | NULLABLE |  |  |
| python\_version | STRING | NULLABLE |  |  |
| packagetype | STRING | NULLABLE | Options [bdist\_egg, bdist\_dumb, bdist\_wininst, bdist\_msi, sdist, bdist\_wheel, bdist\_rpm, bdist\_dmg] |  |
| comment\_text | STRING | NULLABLE | Mostly null |  |
| has\_signature | BOOLEAN | NULLABLE | Don’t know what this represents |  |
| md5\_digest | STRING | REQUIRED | Hash |  |
| sha256\_digest | STRING | NULLABLE | Hash |  |
| blake2\_256\_digest | STRING | NULLABLE | Hash |  |

# Table 2 – File Downloads

| **name** | **Type** | **Mode** | **Description** | **Potential Use** |
| --- | --- | --- | --- | --- |
| timestamp | TIMESTAMP | REQUIRED |  |  |
| country\_code | STRING | NULLABLE |  |  |
| url | STRING | REQUIRED |  |  |
| project | STRING | REQUIRED |  |  |
| file | RECORD | REQUIRED |  |  |
| filename | STRING | NULLABLE |  |  |
| project | STRING | NULLABLE |  |  |
| version | STRING | NULLABLE |  |  |
| type | STRING | NULLABLE |  |  |
| details | RECORD | NULLABLE |  |  |
| installer | RECORD | NULLABLE |  |  |
| name | STRING | NULLABLE |  |  |
| version | STRING | NULLABLE |  |  |
| python | STRING | NULLABLE |  |  |
| implementation | RECORD | NULLABLE |  |  |
| name | STRING | NULLABLE |  |  |
| version | STRING | NULLABLE |  |  |
| distro | RECORD | NULLABLE |  |  |
| name | STRING | NULLABLE |  |  |
| version | STRING | NULLABLE |  |  |
| id | STRING | NULLABLE |  |  |
| libc | RECORD | NULLABLE |  |  |
| lib | STRING | NULLABLE |  |  |
| version | STRING | NULLABLE |  |  |
| system | RECORD | NULLABLE |  |  |
| name | STRING | NULLABLE |  |  |
| release | STRING | NULLABLE |  |  |
| cpu | STRING | NULLABLE |  |  |
| openssl\_version | STRING | NULLABLE |  |  |
| setuptools\_version | STRING | NULLABLE |  |  |
| tls\_protocol | STRING | NULLABLE |  |  |
| tls\_cipher | STRING | NULLAB |  |  |

# Table 3 – Simple Requests

Looks like similar data to table 2.