Install Required Libraries

```
1  !pip install pypdf
2
```

```
Collecting pypdf
  Downloading pypdf-5.9.0-py3-none-any.whl.metadata (7.1 kB)
  Downloading pypdf-5.9.0-py3-none-any.whl (313 kB)
  ──────────────────────────────── 313.2/313.2 kB 1.6 MB/s eta 0:00:00
Installing collected packages: pypdf
Successfully installed pypdf-5.9.0
```

```
1  !pip install -U langchain
2  !pip install -U langchain-community
3  !pip install PyPDF2
```

```
Requirement already satisfied: typing-inspection>0.4.0 in /usr/local/lib/python3.11/dist-packages (from pydantic<3.0/0,>0.1.1->lang
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain) (3
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain) (2.5.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain) (2025.8.
Requirement already satisfied: greenlet>=1 in /usr/local/lib/python3.11/dist-packages (from SQLAlchemy<3,>=1.4->langchain) (3.2.3)
Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages (from httpx<1,>=0.23.0->langsmith>=0.1.17->langchain)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-packages (from httpx<1,>=0.23.0->langsmith>=0.1.17->lan
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.11/dist-packages (from httpcore==1.*->httpx<1,>=0.23.0->langsmith>=
Requirement already satisfied: jsonpointer>=1.9 in /usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33->langchain-core<
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-packages (from anyio->httpx<1,>=0.23.0->langsmith>=0.1.1
Requirement already satisfied: langchain-community in /usr/local/lib/python3.11/dist-packages (0.3.27)
Requirement already satisfied: langchain-core<1.0.0,>=0.3.66 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (0.
Requirement already satisfied: langchain<1.0.0,>=0.3.26 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (0.3.27)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (2.0.42)
Requirement already satisfied: requests<3,>=2 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (2.32.3)
Requirement already satisfied: PyYAML>=5.3 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (6.0.2)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (3.12.15)
Requirement already satisfied: tenacity!=8.4.0,<10,>=8.1.0 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (8.5.
Requirement already satisfied: dataclasses-json<0.7,>=0.5.7 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (0.6
Requirement already satisfied: pydantic-settings<3.0.0,>=2.4.0 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (
Requirement already satisfied: langsmith>=0.1.125 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (0.4.12)
Requirement already satisfied: httpx-sse<1.0.0,>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (0.4.1)
Requirement already satisfied: numpy>=1.26.2 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (2.0.2)
Requirement already satisfied: aiohappyeyeballs>=2.5.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langcha
Requirement already satisfied: aiosignal>=1.4.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-comm
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-communi
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-com
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-c
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-comm
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-com
Requirement already satisfied: marshmallow<4.0.0,>=3.18.0 in /usr/local/lib/python3.11/dist-packages (from dataclasses-json<0.7,>=0.5.
Requirement already satisfied: typing-inspect<1,>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from dataclasses-json<0.7,>=0.5.7-
Requirement already satisfied: langchain-text-splitters<1.0.0,>=0.3.9 in /usr/local/lib/python3.11/dist-packages (from langchain<1.0.0
Requirement already satisfied: pydantic<3.0.0,>=2.7.4 in /usr/local/lib/python3.11/dist-packages (from langchain<1.0.0,>=0.3.26->langc
Requirement already satisfied: jsonpatch<2.0,>=1.33 in /usr/local/lib/python3.11/dist-packages (from langchain-core<1.0.0,>=0.3.66->la
Requirement already satisfied: typing-extensions>=4.7 in /usr/local/lib/python3.11/dist-packages (from langchain-core<1.0.0,>=0.3.66->
Requirement already satisfied: packaging>=23.2 in /usr/local/lib/python3.11/dist-packages (from langchain-core<1.0.0,>=0.3.66->langcha
Requirement already satisfied: httpx<1,>=0.23.0 in /usr/local/lib/python3.11/dist-packages (from langsmith>=0.1.125->langchain-communi
Requirement already satisfied: orjson>=3.9.14 in /usr/local/lib/python3.11/dist-packages (from langsmith>=0.1.125->langchain-community
Requirement already satisfied: requests-toolbelt>=1.0.0 in /usr/local/lib/python3.11/dist-packages (from langsmith>=0.1.125->langchain
Requirement already satisfied: zstandard>=0.23.0 in /usr/local/lib/python3.11/dist-packages (from langsmith>=0.1.125->langchain-commun
Requirement already satisfied: python-dotenv>=0.21.0 in /usr/local/lib/python3.11/dist-packages (from pydantic-settings<3.0.0,>=2.4.0-
Requirement already satisfied: typing-inspection>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from pydantic-settings<3.0.0,>=2.4
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain-com
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain-community) (3.1
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain-community
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain-community
Requirement already satisfied: greenlet>=1 in /usr/local/lib/python3.11/dist-packages (from SQLAlchemy<3,>=1.4->langchain-community) (
Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages (from httpx<1,>=0.23.0->langsmith>=0.1.125->langchain-
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-packages (from httpx<1,>=0.23.0->langsmith>=0.1.125->la
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.11/dist-packages (from httpcore==1.*->httpx<1,>=0.23.0->langsmith>=
Requirement already satisfied: jsonpointer>=1.9 in /usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33->langchain-core<
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4->langcha
Requirement already satisfied: pydantic-core==2.33.2 in /usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4->langchai
Requirement already satisfied: mypy-extensions>=0.3.0 in /usr/local/lib/python3.11/dist-packages (from typing-inspect<1,>=0.4.0->datac
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-packages (from anyio->httpx<1,>=0.23.0->langsmith>=0.1.1
Requirement already satisfied: PyPDF2 in /usr/local/lib/python3.11/dist-packages (3.0.1)
```

Upload File

```
1 from google.colab import files
2 uploaded = files.upload()
```

[Choose Files] No file chosen

Load the PDF file using PyPDFLoader

```
1 from langchain.document_loaders import PyPDFLoader
2
```

```
1 loader = PyPDFLoader('/content/GenAI_Basics_Refrence (1).pdf')
2 documents = loader.load()
```

Split the document into chunks

```
1 from langchain.text_splitter import RecursiveCharacterTextSplitter
2
3 text_splitter = RecursiveCharacterTextSplitter(
4     chunk_size=1000,
5     chunk_overlap=200
6 )
7
8 chunks = text_splitter.split_documents(documents)
```

Print the total number of chunks created

```
1 print(f"Total number of document chunks created: {len(chunks)}")
```

Total number of document chunks created: 25