Part 1: Retrieval-Augmented Generation (RAG) Model for QA Bot

- **OpenAI's embedding models** (e.g., text-embedding-ada-002).
- **Vector Database**:Chromadb (I have used openai hosted on azure which is paid plan in pinecone so went with chromadb)
- **Generative Model**:OpenAI’s GPT-4o-mini

2. Approach to Retrieval

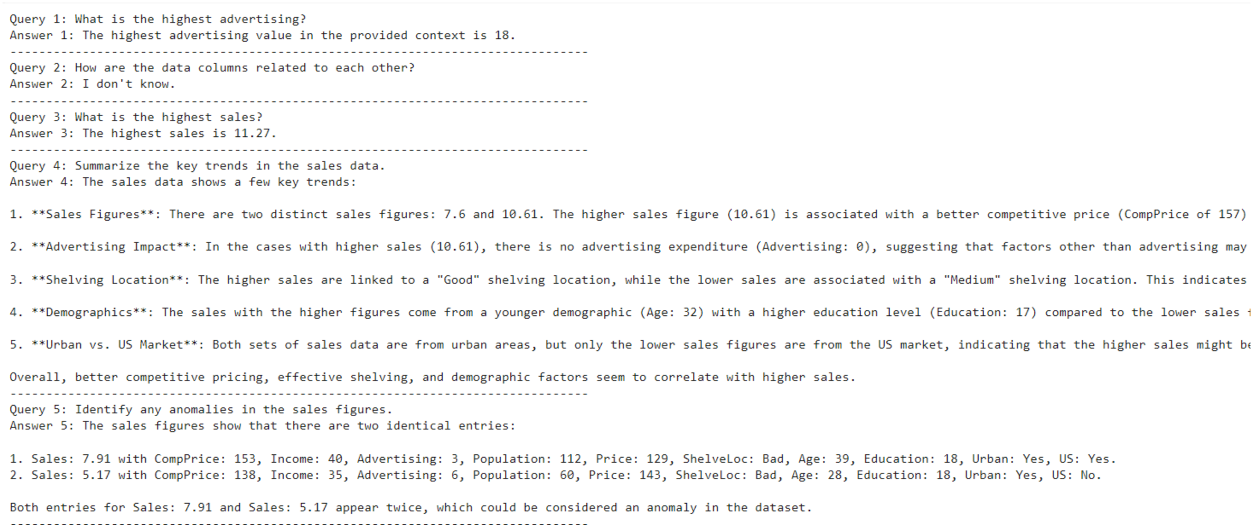
The retrieval process consists of several steps:

- **Document Loading**:
- Documents or datasets are loaded and pre-processed into a format suitable for embedding (e.g., chunks of text).
- **Document Embedding**:
- Each document is transformed into a fixed-size vector using an embedding model. This allows for semantic representation of the documents.
- **Storage in Vector Database**:
- The embeddings are stored in a vector database, where they can be efficiently searched. The database allows for fast nearest neighbor searches using techniques like cosine similarity.-
- **QueryProcessing**:
- When a user submits a query, it is also embedded into a vector representation using the same embedding model.

- The retrieved documents are combined with the original query to form a prompt for the generative model. This can be done in several ways, such as concatenating the query and document text.

Output generated:

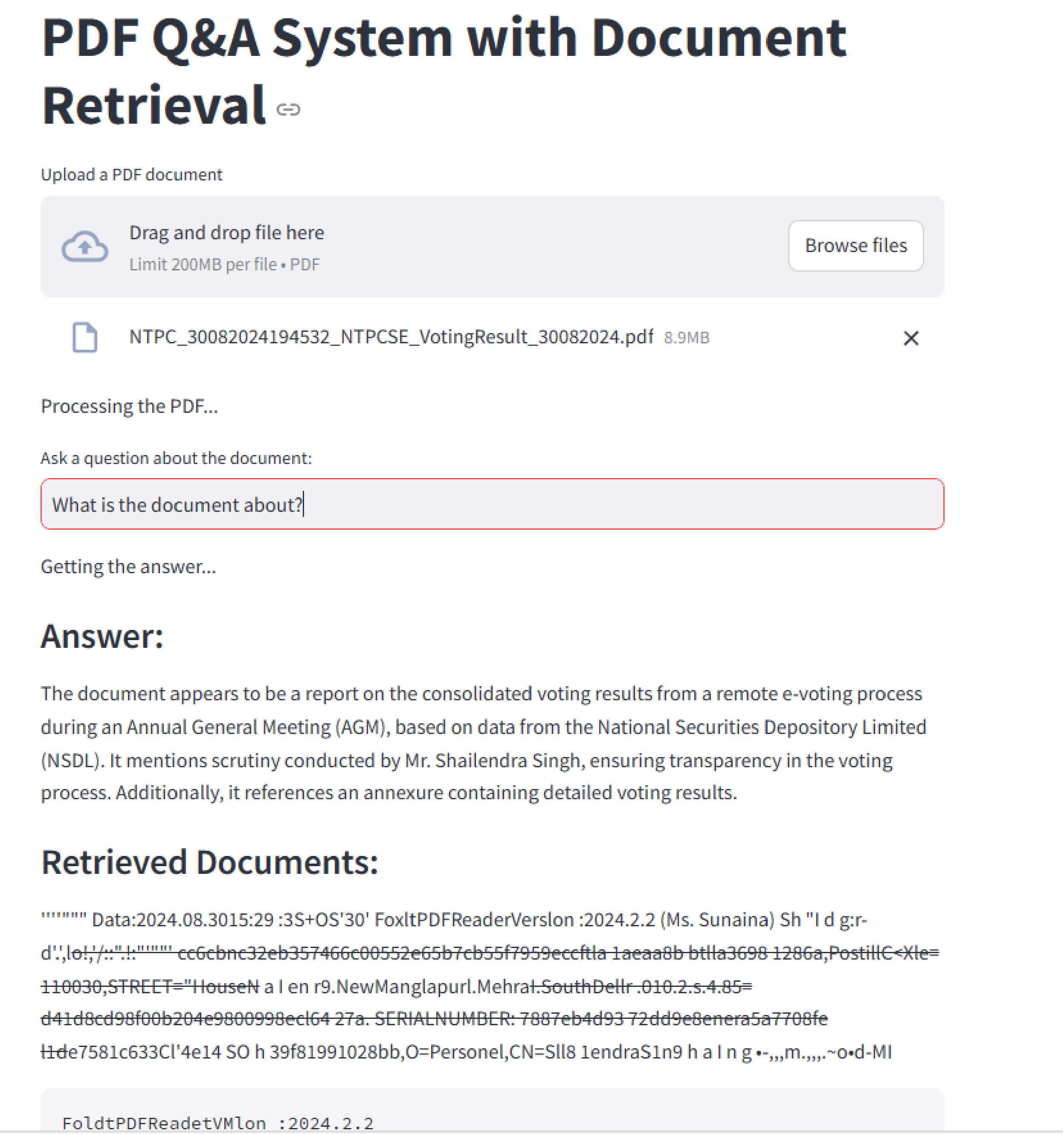
-



```
Query 1: What is the highest advertising?
Answer 1: The highest advertising value in the provided context is 18.
--------------------------------------------------------------------------------
Query 2: How are the data columns related to each other?
Answer 2: I don't know.
--------------------------------------------------------------------------------
Query 3: What is the highest sales?
Answer 3: The highest sales is 11.27.
--------------------------------------------------------------------------------
Query 4: Summarize the key trends in the sales data.
Answer 4: The sales data shows a few key trends:

1. **Sales Figures**: There are two distinct sales figures: 7.6 and 10.61. The higher sales figure (10.61) is associated with a better competitive price (CompPrice of 157)

2. **Advertising Impact**: In the cases with higher sales (10.61), there is no advertising expenditure (Advertising: 0), suggesting that factors other than advertising may

3. **Shelving Location**: The higher sales are linked to a "Good" shelving location, while the lower sales are associated with a "Medium" shelving location. This indicates

4. **Demographics**: The sales with the higher figures come from a younger demographic (Age: 32) with a higher education level (Education: 17) compared to the lower sales

5. **Urban vs. US Market**: Both sets of sales data are from urban areas, but only the lower sales figures are from the US market, indicating that the higher sales might be

Overall, better competitive pricing, effective shelving, and demographic factors seem to correlate with higher sales.
--------------------------------------------------------------------------------
Query 5: Identify any anomalies in the sales figures.
Answer 5: The sales figures show that there are two identical entries:

1. Sales: 7.91 with CompPrice: 153, Income: 40, Advertising: 3, Population: 112, Price: 129, ShelveLoc: Bad, Age: 39, Education: 18, Urban: Yes, US: Yes.
2. Sales: 5.17 with CompPrice: 138, Income: 35, Advertising: 6, Population: 60, Price: 143, ShelveLoc: Bad, Age: 28, Education: 18, Urban: Yes, US: No.

Both entries for Sales: 7.91 and Sales: 5.17 appear twice, which could be considered an anomaly in the dataset.
--------------------------------------------------------------------------------
```

Part 2: Interactive QA Bot Interface

- **Locate the Upload Section**: On the main interface, you will find a section labeled "Upload a PDF document."
- **Locate the Question Input Box**: After the PDF is processed, a text box will appear labeled "Ask a question about the document."
- The system will process your question and provide an answer below the input box.
- Example:



PDF Q&A System with Document Retrieval

Upload a PDF document

NTPC_30082024194532_NTPCSE_VotingResult_30082024.pdf 8.9MB

Processing the PDF...

Ask a question about the document:

What is the document about?

Getting the answer...

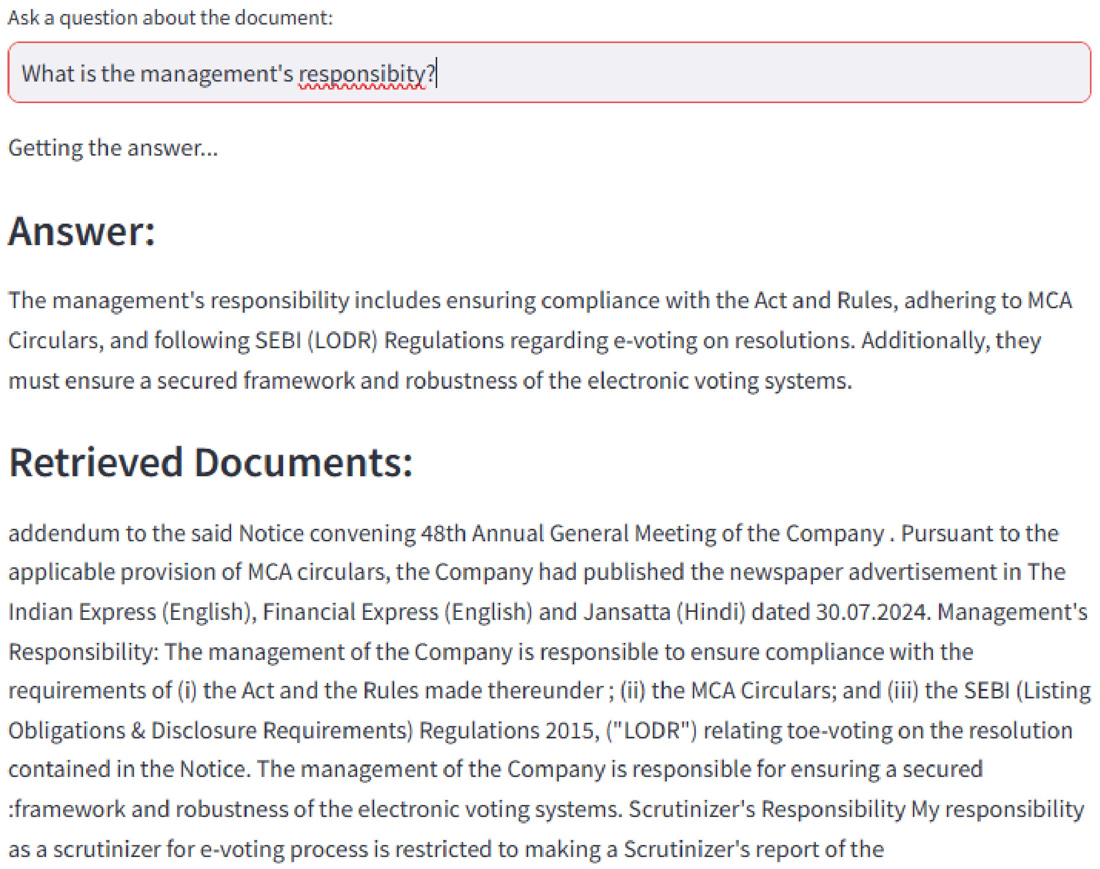
Answer:

The document appears to be a report on the consolidated voting results from a remote e-voting process during an Annual General Meeting (AGM), based on data from the National Securities Depository Limited (NSDL). It mentions scrutiny conducted by Mr. Shailendra Singh, ensuring transparency in the voting process. Additionally, it references an annexure containing detailed voting results.

Retrieved Documents:

"""""" Data:2024.08.3015:29 :3S+OS'30' FoxltPDFReaderVerslon :2024.2.2 (Ms. Sunaina) Sh "I d g:r-d'.,~~lo!,'/::".!:"""" cc6cbnc32eb357466c00552e65b7cb55f7959eccftla 1aeaa8b btlla3698 1286a,PostillC<Xle= 110030,STREET="HouseN~~ a I en r9.NewManglapurl.Mehra~~l.SouthDellr .010.2.s.4.85= d41d8cd98f00b204e9800998ecl64 27a. SERIALNUMBER: 7887eb4d93 72dd9e8enera5a7708fe 11d~~e7581c633Cl'4e14 SO h 39f81991028bb,O=Personel,CN=Sll8 1endraS1n9 h a I n g •-,,,m.,,,.~o•d-MI

```
FoldtPDFReadetVMlon :2024.2.2
```



Ask a question about the document:

What is the management's responsibity?

Getting the answer...

Answer:

The management's responsibility includes ensuring compliance with the Act and Rules, adhering to MCA Circulars, and following SEBI (LODR) Regulations regarding e-voting on resolutions. Additionally, they must ensure a secured framework and robustness of the electronic voting systems.

Retrieved Documents:

addendum to the said Notice convening 48th Annual General Meeting of the Company . Pursuant to the applicable provision of MCA circulars, the Company had published the newspaper advertisement in The Indian Express (English), Financial Express (English) and Jansatta (Hindi) dated 30.07.2024. Management's Responsibility: The management of the Company is responsible to ensure compliance with the requirements of (i) the Act and the Rules made thereunder ; (ii) the MCA Circulars; and (iii) the SEBI (Listing Obligations & Disclosure Requirements) Regulations 2015, ("LODR") relating toe-voting on the resolution contained in the Notice. The management of the Company is responsible for ensuring a secured :framework and robustness of the electronic voting systems. Scrutinizer's Responsibility My responsibility as a scrutinizer for e-voting process is restricted to making a Scrutinizer's report of the