

# Problem 2 Report

## Azure CosmosDB–Backed Movie Analytics Dashboard

Mallekedi Hari Babu

November 23, 2025

### Abstract

This report presents the development of a cloud-supported interactive analytics dashboard using Azure Cosmos DB (MongoDB API) and Streamlit. A document-based movie dataset (Sample MFlix) was uploaded and managed on Azure, enabling remote access and scalable querying. The dashboard visualizes key insights such as movie trends, user engagement, rating dynamics, and word clouds derived from document-level text fields. This work demonstrates how cloud-hosted document databases can support analytical applications in real-world business and media intelligence contexts.

## 1 Introduction

Document databases such as MongoDB and Azure Cosmos DB are widely used in modern data-driven applications due to their flexibility, scalability, and ability to store semi-structured JSON-like data. In this exercise, a full cloud-to-application workflow was implemented:

- Upload a document-based dataset to Azure Cosmos DB (MongoDB API).
- Connect a Streamlit dashboard directly to the cloud database.
- Analyze and visualize key metrics such as ratings, genres, comments, and text patterns.
- Explain the purpose and business value of the dashboard in a realistic decision-making context.

The final deployed dashboard link is: <https://sample-mflix-analysis-7dxkw8fnehrjaxensvgx8a.st>

## 2 Dataset Upload to Azure Cosmos DB

A full document-based movie dataset (Sample MFlux) containing collections for movies, comments, users, and theaters was uploaded to Azure Cosmos DB under the MongoDB API compatibility mode.

### Collections in Azure Cosmos DB

The following screenshot shows the online collections hosted on Azure:

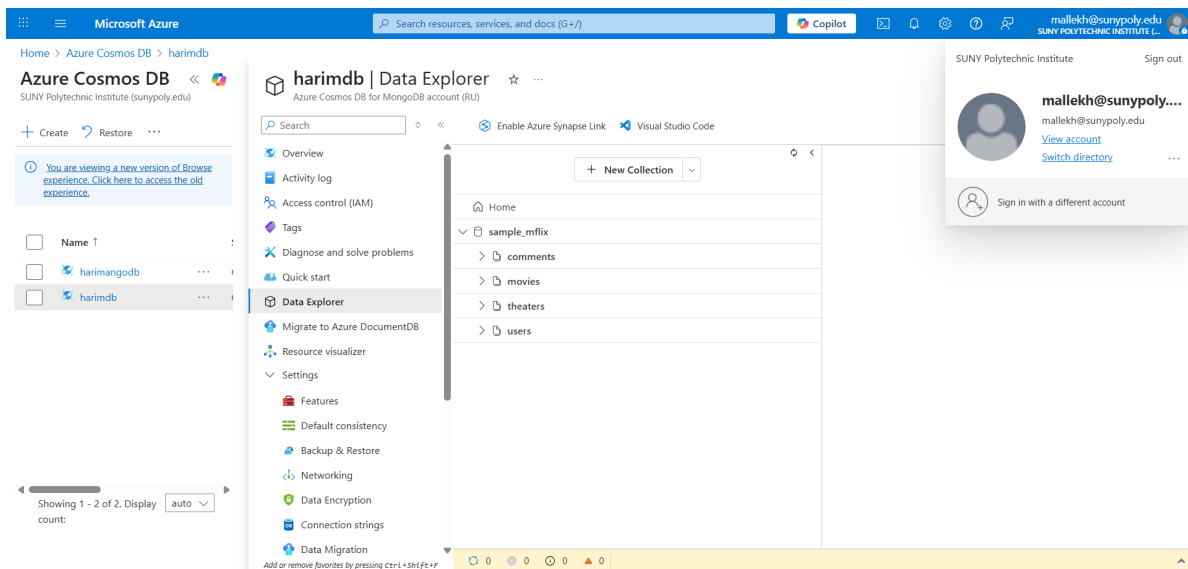


Figure 1: CosmosDB Data Explorer showing uploaded collections

These collections are now accessible online using a secure connection string managed via Streamlit Secrets.

## 3 Streamlit Dashboard Design

A complete Streamlit dashboard was developed to provide insight into movie content, ratings, user interactions, and textual patterns.

# Dashboard Home View

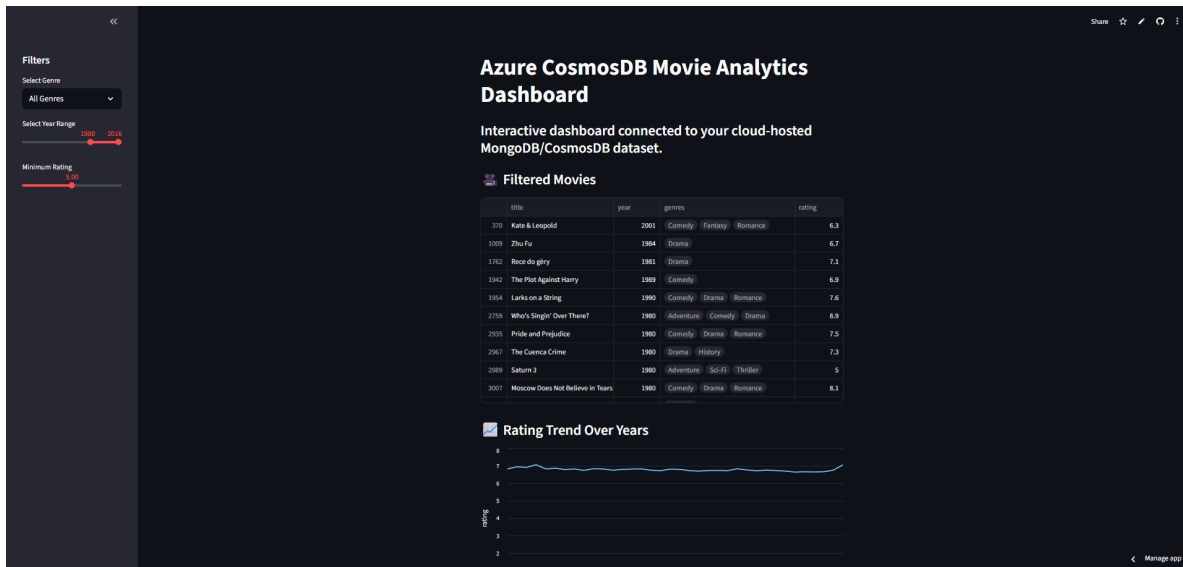


Figure 2: Main interface of the Streamlit dashboard

The dashboard includes:

- Dynamic movie filtering by genre, year, and rating
- Rating trend visualization
- Commenter activity analytics
- Word cloud from movie plots



## Rating Trend Over the Years

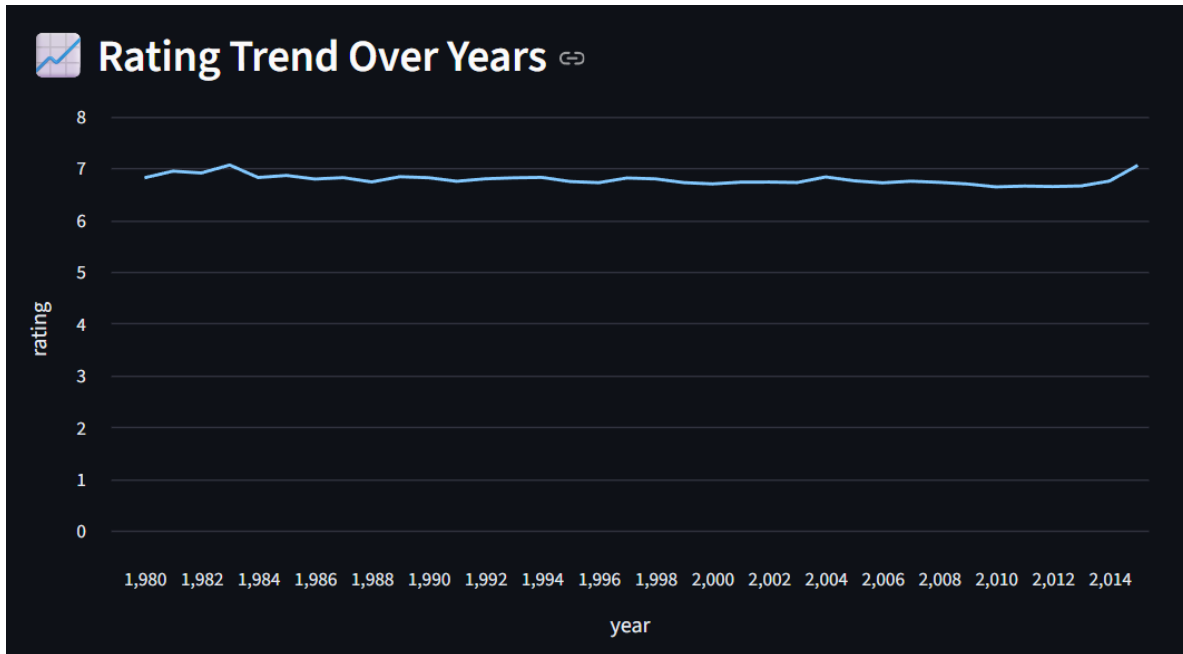


Figure 5: Trend of IMDb ratings over movie release years

## 5 Narrative and Business Value

The dashboard serves as an analytical tool demonstrating how cloud-hosted document data can support real-world decision making. In a media or streaming platform context:

- **Content Strategy:** Identifying genres with high engagement or strong rating trends can guide future acquisition decisions.
- **User Engagement Insights:** Commenter behavior reveals active user clusters and helps design targeted community features.
- **Quality Monitoring:** Rating trends across decades detect periods of strong or weak movie performance.
- **Text Mining Insights:** Word clouds and text trends uncover narrative themes that resonate with audiences.

The cloud architecture enables:

- consistent online availability,

- scalable data access,
- secure integration via secrets management,
- low-latency retrieval for interactive dashboards.

## 6 Conclusion

This project successfully demonstrates the integration of a cloud-hosted document database with an interactive analytics dashboard. Azure Cosmos DB ensures reliable data availability, while Streamlit provides an intuitive interface for exploring insights. The combination of cloud data storage and real-time visualization is directly applicable to business domains such as entertainment analytics, retail platforms, IoT monitoring, and educational dashboards.